

Combined State and Parameter Estimation in Level-Set Methods

Hans Yu^a, Matthew P. Juniper^a, Luca Magri^{a,b,*}

^a*Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK*

^b*Institute for Advanced Study, Technical University of Munich, Lichtenbergstrasse 2a, 85748 Garching, Germany (visiting fellowship)*

Abstract

Reduced-order models based on level-set methods are widely used tools to qualitatively capture and track the nonlinear dynamics of an interface. The aim of this paper is to develop a physics-informed, data-driven, statistically rigorous learning algorithm for state and parameter estimation with level-set methods. A Bayesian approach based on data assimilation is introduced. Data assimilation is enabled by the ensemble Kalman filter and smoother, which are used in their probabilistic formulations. The level-set data assimilation framework is verified in one-dimensional and two-dimensional test cases, where state estimation, parameter estimation and uncertainty quantification are performed. The statistical performance of the proposed ensemble Kalman filter and smoother is quantified by twin experiments. In the twin experiments, the combined state and parameter estimation fully recovers the reference solution, which validates the proposed algorithm. The level-set data assimilation framework is then applied to the prediction of the nonlinear dynamics of a forced premixed flame, which exhibits the formation of sharp cusps and intricate topological changes, such as pinch-off events. The proposed physics-informed statistical learning algorithm opens up new possibilities for making reduced-order models of interfaces quantitatively predictive, any time that reference data is available.

Keywords: Data assimilation, Ensemble Kalman filter, Level-set method, Parameter estimation, Uncertainty quantification

1. Introduction

A number of problems in computational physics involve the motion of interfaces, e.g. semiconductor manufacturing, multi-phase flows, crystal growth, groundwater flow, computer vision,

*Corresponding author

Email address: `lm547@cam.ac.uk` (Luca Magri)

grid generation and seismology [1]. In general, there are two approaches to calculating the motion of an interface [2]: In front-tracking methods, the interface is parameterized and discretized so that one follows the motion of the whole interface by tracking a sufficient number of points on the interface. In front-capturing methods, the interface is embedded into a function defined over the whole domain. One example of front-capturing methods are the so-called level-set methods, where the interface is embedded into a strictly monotonic function [3]. Both front-tracking and front-capturing methods have well-understood advantages, and hybrid methods exist to mitigate their respective disadvantages [4]. Nevertheless, it is worth mentioning that level-set methods provide a natural formulation for calculating the motion of an interface [1]: Level-set methods deal well with non-smooth features such as corners and cusps as well as topological merging and break-up. Furthermore, level-set methods are easily extended from two to three and higher dimensions. For more information on level-set methods, the reader is referred to many excellent expositions in the literature [5, 6, 1, 4, 2].

Despite the elegance of explaining physical phenomena by the motion of interfaces and calculating the motion of the interfaces by level-set methods, one has to remain aware that the assumption that manifolds are infinitely thin is often an asymptotic assumption. This is particularly true in fluid mechanics, which is governed by conservation laws [7]. One example is the kinematics of premixed flames [8]: Depending on the combustion regime, it may be assumed that a thin reactive-diffusive layer separates the burnt and unburnt gases. While the laminar flame speed, at which the premixed flame propagates from the burnt into the unburnt region, is a well-defined thermo-chemical property of the fuel-air mixture in a one-dimensional flow, it more generally depends on the balance between heat conduction and mass diffusion, whose effects vary with flame stretch and curvature. Moreover, the turbulent flame speed scales differently depending on the interactions between turbulence and combustion length scales. Despite the potential quantitative shortcomings in describing the kinematics of a premixed flame as the motion of an interface, it has been shown that this model successfully explains the linear and nonlinear dynamics observed in thermoacoustic instabilities of ducted premixed flames, which are relevant to the design of combustion chambers in jet and rocket engines [9, 10, 11, 12]. The discrepancies which arise when comparing to more faithful simulations or experiments are usually attributed to the unpredictable nature of turbulent flow or to uncertainties in the model and its parameters. Thus, it is relevant to assess the ability of a qualitative, physics-informed model to make quantitative, time-accurate predictions. The aim of this study is to develop a data-driven, statistically rigorous framework for state and parameter estimation in models using level-set methods, which

has not been done before, and to apply it to an oscillating flame.

Inference over interfaces based on level-set methods has been performed, e.g. in the context of shape optimization, either by directly taking the functional derivative of the objective functional with respect to the level-set function [13, 14], or, equivalently, by embedding the level-set method into shape calculus [15, 16, 17]. As an alternative to these variational approaches, the framework for state and parameter estimation in this study is based on data assimilation [18]. Data assimilation finds the statistically optimal combination of model predictions and observations. It combines concepts from control theory, probability theory and dynamic programming [19, 20, 21]. The data assimilation technique used in this study is the ensemble Kalman filter [22, 23]. In the ensemble Kalman filter, a Monte Carlo approach is used to represent the necessary statistics at every timestep [24], which makes it a computationally efficient technique in terms of storage requirements. Compared to other data assimilation techniques based on the Kalman filter, e.g. the extended Kalman filter [25, 26], the ensemble Kalman filter is found to be particularly robust with respect to larger nonlinearities [18]. This is relevant for level-set methods, due to strongly nonlinear events such as cusp formation, topological merging and break-up. A practical advantage of the ensemble Kalman filter is its non-intrusive implementation with little effort required for its parallelization. The ensemble Kalman filter has been successfully applied to a number of problems in fluid mechanics: turbulent near-wall flow [27]; transonic flows around airfoils and wings [28]; viscous flow around a cylinder [29]; model uncertainties in Reynolds-averaged Navier-Stokes (RANS) equations [30]; vortex models of separated flow [31]; and extinction and reignition dynamics in turbulent non-premixed combustion [32]. Relevant applications of data assimilation to level-set methods outside fluid mechanics include wildfire spread [33, 34, 35, 36, 37] and pollution on the surface of oceans and lakes [38, 39].

In this study, the ensemble Kalman filter is combined with a narrow-band level-set method and a fast marching method to form a computationally efficient level-set data assimilation framework [40, 41]. In analogy to the distinction between front-tracking and front-capturing methods in describing the motion of an interface, data can also be assimilated according to various paradigms, e.g. as demonstrated by [42], [43] and [44]. For the various paradigms, it is not a priori clear which will yield superior results. For this reason, the theory and the derivations behind our level-set data assimilation framework are worked out in detail. The paper is structured as follows: In Section 2, the ensemble Kalman filter and the ensemble Kalman smoother are derived within the context of Bayesian inference. The estimation problems in data assimilation and parameter estimation are then formulated in terms of probability distributions. In Section 3, various formu-

70 lations for the laws of motion are discussed, including the Hamilton-Jacobi equation. It is shown
that the solutions to the Hamilton-Jacobi equation, the so-called generating functions, form a
natural state space for data assimilation. The level-set data assimilation framework based on
the Hamilton-Jacobi equation is verified in one-dimensional and two-dimensional examples. In
Section 4, statistical inference using the level-set data assimilation framework is demonstrated
75 on the nonlinear dynamics of a ducted premixed flame. The results and insights of this study
are summarized in the conclusions (Section 5).

2. Data assimilation and parameter estimation

Data assimilation and parameter estimation are here treated as problems in statistical inference. Statistical inference quantifies the degree of belief (or confidence) in a physical model,
80 the parameters that it receives and the states that it predicts. Statistical inference follows a
probabilistic formulation, which provides precise definitions for the different tasks addressed in
inference, which include filtering, smoothing and prediction. Probabilistic formulations for sta-
tistical inference include frequentist inference and Bayesian inference. In frequentist inference, an
error functional is defined to measure the statistical distance between a candidate solution and
85 the available data. Inference becomes an optimization problem to minimize this error functional.
In Bayesian inference, existing knowledge is quantified in the form of a probability distribution
over candidate solutions. When data becomes available, the probability distribution is updated,
effectively combining the existing knowledge with the data. For normal distributions under lin-
ear dynamics, both formulations of statistical inference give equivalent results [18]. Under these
90 circumstances, frequentist inference may be considered more accessible because it allows the ap-
plication of familiar tools from convex optimization [45]. Nevertheless, Cox's axioms demonstrate
that probability theory, as used in Bayesian inference, gives a natural formulation for inference in
the general case [46]. All definitions and derivations in this section are given in terms of Bayesian
inference (Appendix A). For more details, the reader may refer to [18] or [47].

95 2.1. Probabilistic state space model

The state of a system at a timestep k is uniquely defined by the state vector x_k . The state
vector need not be a scalar or a finite-dimensional tuple, but may be taken from an infinite-
dimensional function space. The evolution of the state is governed by a physical model f and
its parameters θ . Thus, our belief in the state x_k depends exclusively on our belief in (i) the
previous state x_{k-1} as well as our choice of (ii) the physical model f and (iii) its parameters θ . At

the same time, the state vector x_k is compared to noisy observations y_k through a measurement operator M . Formulated in terms of a probabilistic state space model, we obtain

$$x_k = x_{k-1} + \int_{t_{k-1}}^{t_k} f(x(t), \theta) dt = G(x_{k-1}, \theta) \sim p(x_k | x_{k-1}, \theta, f) \quad , \quad (1)$$

$$y_k = M(x_k) \sim p(y_k | x_k) \quad . \quad (2)$$

In brief, the transition from one state to the next is governed by the operator G . The operator G is derived from the physical model f , and depends on the parameters θ . The states x_k and the observations y_k are considered realizations inside their respective probabilistic state spaces. The degrees of belief in each are subject to (“ \sim ”) the conditional probability distributions $p(x_k |$
100 $x_{k-1}, \theta, f)$ and $p(y_k | x_k)$ respectively. The degree of belief in the state x_k is conditional on the previous state x_{k-1} and the choices in the physical model f and its parameters θ . The degree of belief in the observation y_k is conditional on the true state x_k .

The rules of the probabilistic state space model may be formalized as follows [47]. Firstly, we assume that the physical model may be described by a Markov chain [24]. This means that
105 the belief in a state depends only on the belief in the previous state:

1. The future is independent of the past given the present:

$$p(x_k | x_{0:k-1}, y_{1:k-1}, \theta, f) = p(x_k | x_{k-1}, \theta, f) \quad , \quad (3)$$

where $x_{0:k-1}$ denotes the states at all timesteps from 0 to $k-1$, and $y_{1:k-1}$ denotes the observations at all timesteps from 1 to $k-1$.

2. The past is independent of the future given the present:

$$p(x_k | x_{k+1:N}, y_{k+1:N}, \theta, f) = p(x_k | x_{k+1}, \theta, f) \quad , \quad (4)$$

where $x_{k+1:N}$ and $y_{k+1:N}$ respectively denote the states and observations at all timesteps from $k+1$ to the final timestep N .

Secondly, observations are assumed to be conditionally independent in time. The probability of an observation depends only on the current state:

$$p(y_k | x_{0:N}, y_{1:k-1}, y_{k+1:N}, \theta, f) = p(y_k | x_k) \quad . \quad (5)$$

110 In Fig. 1, the relationship between states and observations is shown as well as the roles of models, parameters and measurement.

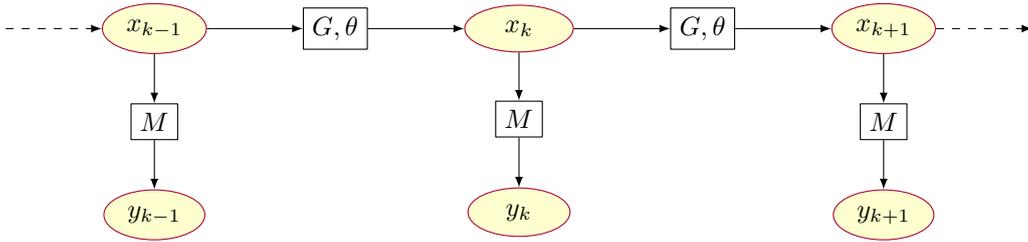


Figure 1: Probabilistic state space model as a Markov chain. The operator G and its parameters θ govern the transition from one state to the next. The state vector x_k is related to the observations y_k through a measurement operator M .

Table 1: Conditional probability distributions in data assimilation.

Task	PDF	Description
State estimation	$p(x_{0:N} y_{1:N}, \theta, f)$	Given a physical model and its parameters, what is our belief in a series of states?
Parameter estimation	$p(\theta y_{1:N}, f)$	Given a physical model, what is our belief in a set of parameters?
Model comparison	$p(f y_{1:N})$	Between two physical models, in which one do we believe more?

The goal of data assimilation is to find the joint probability distribution $p(x_{0:N}, y_{1:N}, \theta, f)$. The probabilistic state space spans all states x_k from timestep 0 to N and all observations y_k from timestep 1 to N , as well as the physical model f and its parameters θ . This joint probability distribution gives a complete statistical description. In principle, all probability distributions of interest may be derived from this joint probability distribution (Table 1). In practice, it is difficult to compute this probability distribution because of the high dimensionality of its probabilistic state space spanning multiple timesteps [48]. Therefore, state estimation focuses on the more direct computation of conditional probability distributions over a single timestep (Table 2).

In the probabilistic formulation, data assimilation is easily extended to account for parameters. In combined state and parameter estimation, the state is augmented by the parameters so that they become subject to the same inference (Eq. (1)):

$$\tilde{x}_k = \begin{pmatrix} x_k \\ \theta_k \end{pmatrix}, \quad \tilde{f}(\tilde{x}(t)) = \begin{pmatrix} f(x(t), \theta_k) \\ 0 \end{pmatrix}, \quad \tilde{x}_k = \tilde{x}_{k-1} + \int_{t_{k-1}}^{t_k} \tilde{f}(\tilde{x}(t)) dt \quad . \quad (6)$$

Table 2: Conditional probability distributions in state estimation.

Task	PDF	Description
Filtering	$p(x_k y_{1:k}, \theta, f)$	Given all observations from the past and now, what is our belief in the current state?
Smoothing	$p(x_k y_{1:N}, \theta, f)$	Given all observations from the past, the future and now, what is our belief in the current state?
Prediction	$p(x_{N+1} y_{1:N}, \theta, f)$	Given all observations from the past and now, what is our belief in a future state?

Table 3: Conditional probability distributions in combined state and parameter estimation.

Task	PDF	Description
Filtering	$p(x_k, \theta_k y_{1:k}, f)$	Given all observations from the past and now, what is our belief in the current state and set of parameters?
Smoothing	$p(x_k, \theta_k y_{1:N}, f)$	Given all observations from the past, the future and now, what is our belief in the current state and set of parameters?

120 The tasks in combined state and parameter estimation are given in Table 3. The filtered and smoothed distributions in the parameters θ_k , $p(\theta_k | y_{1:k}, f)$ and $p(\theta_k | y_{1:N}, f)$ respectively, are retrieved by marginalizing the states x_k (Appendix A). Note that the parameters θ_k are now time-dependent as the system traverses different regimes in state space. This turns the strongly constrained parameter estimation into a weakly constrained combined state and parameter estimation [18]. Thus, the results of marginalizing the probability distributions in combined state and parameter estimation (Table 3) are not strictly equivalent to the solutions of parameter estimation (Table 1).

125

2.2. Bayesian filtering and smoothing

For the filtering problem, Bayes' rule gives

$$p(x_k | y_{1:k}, \theta, f) = \frac{p(y_k | x_k, y_{1:k-1}, \theta, f)p(x_k | y_{1:k-1}, \theta, f)}{p(y_k | y_{1:k-1}, \theta, f)} \quad (7)$$

$$= \frac{p(y_k | x_k)p(x_k | y_{1:k-1}, \theta, f)}{p(y_k | y_{1:k-1}, \theta, f)} \quad (8)$$

The prediction $p(x_k | y_{1:k-1}, \theta, f)$ is given by the Chapman-Kolmogorov equation:

$$p(x_k | y_{1:k-1}, \theta, f) = \int p(x_k, x_{k-1} | y_{1:k-1}, \theta, f) dx_{k-1} \quad (9)$$

$$= \int p(x_k | x_{k-1}, y_{1:k-1}, \theta, f)p(x_{k-1} | y_{1:k-1}, \theta, f) dx_{k-1} \quad (10)$$

$$= \int p(x_k | x_{k-1}, \theta, f)p(x_{k-1} | y_{1:k-1}, \theta, f) dx_{k-1} \quad (11)$$

The Chapman-Kolmogorov equation requires the inverse from the previous timestep $k - 1$. In general, it is solved either numerically [49], analytically (Theorem 3) or by a Monte Carlo simulation (Theorem 4). The steps in the Bayesian filter may be summarized as follows:

Theorem 1 (Bayesian filter).

1. *Prediction step:*

$$p(x_k | y_{1:k-1}, \theta, f) = \int p(x_k | x_{k-1}, \theta, f)p(x_{k-1} | y_{1:k-1}, \theta, f) dx_{k-1} \quad (12)$$

2. *Update step:*

$$p(x_k | y_{1:k}, \theta, f) = \frac{p(y_k | x_k)p(x_k | y_{1:k-1}, \theta, f)}{p(y_k | y_{1:k-1}, \theta, f)} \quad (13)$$

At the timestep $k = N$, the filtered and smoothed distributions are identical (Tab. 2). If the smoothed distribution at a timestep $k + 1$ is known, the smoothed distribution at the previous timestep k is also known due to the Markov chain properties of the probabilistic state space model (Eq. (3), (4)). This may be formalized as follows:

$$p(x_k | y_{1:N}, \theta, f) = \int p(x_k, x_{k+1} | y_{1:N}, \theta, f) dx_{k+1} \quad (14)$$

$$= \int p(x_k | x_{k+1}, y_{1:N}, \theta, f)p(x_{k+1} | y_{1:N}, \theta, f) dx_{k+1} \quad (15)$$

$$= \int p(x_k | x_{k+1}, y_{1:k}, \theta, f)p(x_{k+1} | y_{1:N}, \theta, f) dx_{k+1} \quad (16)$$

$p(x_{k+1} | y_{1:N}, \theta, f)$ is the smoothed distribution from the subsequent timestep $k + 1$. $p(x_k | x_{k+1}, y_{1:k}, \theta, f)$ is computed via Bayes' rule:

$$p(x_k | x_{k+1}, y_{1:k}, \theta, f) = \frac{p(x_{k+1} | x_k, y_{1:k}, \theta, f)p(x_k | y_{1:k}, \theta, f)}{p(x_{k+1} | y_{1:k}, \theta, f)} \quad (17)$$

$$= \frac{p(x_{k+1} | x_k, \theta, f)p(x_k | y_{1:k}, \theta, f)}{p(x_{k+1} | y_{1:k}, \theta, f)} \quad (18)$$

Note that it involves the filtered distribution from the current timestep k . The steps in the Bayesian smoother may be summarized as follows:

135 **Theorem 2** (Bayesian smoother).

1. *Forward sweep: Bayesian filter (Theorem 1).*
2. *Backward step:*

$$p(x_k | y_{1:N}, \theta, f) = p(x_k | y_{1:k}, \theta, f) \int \frac{p(x_{k+1} | x_k, \theta, f)p(x_{k+1} | y_{1:N}, \theta, f)}{p(x_{k+1} | y_{1:k}, \theta, f)} dx_{k+1} \quad (19)$$

Note that both the Bayesian filter and smoother are sequential in nature. In the Bayesian filter, the predicted distribution $p(x_k | y_{1:k-1}, \theta, f)$ at a timestep k (Eq. (12)) mainly depends on the filtered distribution $p(x_{k-1} | y_{1:k-1}, \theta, f)$ from the previous timestep $k - 1$ (Eq. (13)). In the
 140 Bayesian smoother, the smoothed distribution $p(x_k | y_{1:N}, \theta, f)$ at a timestep k mainly depends on the smoothed distribution $p(x_{k+1} | y_{1:N}, \theta, f)$ from the subsequent timestep $k + 1$ (Eq. (19)). The existence of sequential algorithms for the computation of filtered and smoothed distributions significantly reduces the complexity of data assimilation [48].

2.3. The Kalman filter and the Rauch-Tung-Striebel smoother

Two additional assumptions are introduced to make the computation of filtered and smoothed distributions feasible. Firstly, the prior and the likelihood in the update step of the Bayesian filter are assumed to be normal (Eq. (13)):

$$p(x_k | y_{1:k-1}, \theta, f) = \mathcal{N}(x_k | \psi^f, C_{\psi\psi}^f) \quad (20)$$

$$p(y_k | x_k) = \mathcal{N}(y_k | Mx_k, C_{\epsilon\epsilon}) \quad (21)$$

145 where \mathcal{N} denotes a normal distribution with respective mean and covariance matrix. The mean of the prior is denoted by ψ^f , its covariance matrix by $C_{\psi\psi}^f$, and the covariance matrix of the likelihood, also known as the observation error, by $C_{\epsilon\epsilon}$. From Eq. (20) and (21), it follows that the filtered distribution $p(x_k | y_{1:k}, \theta, f)$ is normal (Eq. (13)). Secondly, the operator G (Eq. (1))

in the prediction step (Eq. (12)) is assumed to be linear in x . The result is the well-known
150 Kalman filter [50, 51]:

Theorem 3 (Kalman filter).

$$p(x_k | y_{1:k}, \theta, f) = \mathcal{N}(x_k | \psi^a, C_{\psi\psi}^a) \quad , \quad (22)$$

$$\psi^a = \psi^f + \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} (y_k - M\psi^f) \quad , \quad (23)$$

$$C_{\psi\psi}^a = C_{\psi\psi}^f - \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} \left(MC_{\psi\psi}^f \right) \quad , \quad (24)$$

where the superscript f denotes 'forecast' (everything pertaining to the prediction), and the superscript a denotes 'analysis' (everything pertaining to the update).

If the physical model is nonlinear, the prediction of the covariance matrix $C_{\psi\psi}^f$ for the Kalman filter may be generalized in several ways. In the extended Kalman filter, the predicted covariance matrix $C_{\psi\psi}^f$ is computed by linearizing f [20]. In strongly nonlinear dynamical systems, the predictions are found to be poor [18]. Higher-order extended Kalman filters are available [25]. Nevertheless, drawbacks include their significant storage requirements, which increase exponentially with the order of approximation. An alternative is the ensemble Kalman filter [22, 23]. Instead of a mean ψ and a covariance matrix $C_{\psi\psi}$, a distribution is represented by a sample $(\psi^j)_{j=1\dots n}$. During the prediction step, the ensemble members ψ^j evolve in time independently. Before the update step, the statistics may be recovered from the sample as follows:

$$\bar{\psi} \approx \frac{1}{n} \sum_{j=1}^n \psi^j \quad , \quad \Psi = \left(\psi^1 - \bar{\psi} \quad , \quad \psi^2 - \bar{\psi} \quad , \quad \dots \quad , \quad \psi^n - \bar{\psi} \right) \quad , \quad C_{\psi\psi} \approx \frac{1}{n-1} \Psi \Psi^T \quad . \quad (25)$$

The sample covariance matrix $C_{\psi\psi}$ involves division by $n-1$ instead of n in order to avoid a sample bias.

155 Various implementations of the ensemble Kalman filter exist, which differ in the update step. In the straightforward implementation of the ensemble Kalman filter [22], every ensemble member is individually updated (Eq. (23)). It can be shown that the observations must be randomly perturbed in order to guarantee a statistically consistent analysis scheme [23]. In order to avoid the introduction of randomly generated numbers, the square-root filter is used
160 here [52]. The square-root filter belongs to a larger family of ensemble Kalman filters called ensemble-transform Kalman filters [53, 54]. Unlike the straightforward implementation of the

ensemble Kalman filter, the mean and the deviations of the ensemble members are updated. This requires the singular value decomposition of a symmetric, positive, semi-definite matrix ($V\Sigma V^T$, where V is orthonormal and Σ diagonal), but no spurious errors due to the random perturbation of the observations are introduced.

Theorem 4 (Square-root filter).

$$(\psi^a)^j = \bar{\psi}^a + (\Psi^a)_j \quad , \quad (26)$$

$$\bar{\psi}^a = \bar{\psi}^f + \Psi^f (M\Psi^f)^T \left[(n-1)C_{\epsilon\epsilon} + M\Psi^f (M\Psi^f)^T \right]^{-1} (y_k - M\bar{\psi}^f) \quad , \quad (27)$$

$$\Psi^a = \Psi^f V [\mathbb{I} - \Sigma]^{\frac{1}{2}} V^T \quad , \quad V\Sigma V^T = (M\Psi^f)^T \left[(n-1)C_{\epsilon\epsilon} + M\Psi^f (M\Psi^f)^T \right]^{-1} M\Psi^f \quad , \quad (28)$$

where \mathbb{I} is the identity matrix.

Following Eq. (20) and (21), the Bayesian smoother (Theorem 2) becomes the Rauch-Tung-Striebel smoother, also known as Kalman smoother (Theorem 5). It can be shown that the smoothed distribution becomes a normal distribution with mean ψ^s and covariance matrix $C_{\psi\psi}^s$. In order to again avoid the shortcomings of assuming linearity, an ensemble Kalman smoother is presented here (Theorem 6).

Theorem 5 (Rauch-Tung-Striebel smoother).

$$p(x_k | y_{1:k}) = \mathcal{N}(x_k | \psi^s, C_{\psi\psi}^s) \quad , \quad (29)$$

$$\psi^s = \psi^a + \left((C_{\psi\psi}^f)_{k+1}^{-1} G C_{\psi\psi}^a \right)^T \left[\psi_{k+1}^s - \psi_{k+1}^f \right] \quad , \quad (30)$$

$$C_{\psi\psi}^s = C_{\psi\psi}^a - \left((C_{\psi\psi}^f)_{k+1}^{-1} G C_{\psi\psi}^a \right)^T \left[(C_{\psi\psi}^f)_{k+1} - (C_{\psi\psi}^s)_{k+1} \right] \left((C_{\psi\psi}^f)_{k+1}^{-1} G C_{\psi\psi}^a \right) \quad . \quad (31)$$

Theorem 6 (Ensemble Kalman smoother).

$$(\psi^s)^j = (\psi^a)^j + \Psi^a (\Psi^f)_{k+1}^{-1} \left[(\psi^s)^j_{k+1} - (\psi^f)^j_{k+1} \right] \quad . \quad (32)$$

3. Level-set methods

When data assimilation is applied to the motion of an interface, the question arises as to what constitutes its probabilistic state space. Refining the distinction between front-tracking and front-capturing methods, there are at least three ways to view the motion of an interface [1, 4]:

Geometric view. The interface is parameterized and discretized so that one follows the motion of the whole interface by solving the laws of motion for a sufficient number of points on the interface.

Set-theoretic view. A characteristic function is defined over the whole domain. The characteristic function assumes one of two values, depending on whether the point at the location in question is inside or outside the region enclosed by the interface.

Analytic view (also known as “analysis view” [4]). A level-set function is defined over the whole domain. The interface is reconstructed by identifying the position of a particular level set.

Beyond the description of motion, these three views also pertain to the assimilation of data. In fact, all three views have been employed in earlier level-set data assimilation frameworks: While using a level-set method when solving for the motion of the interface, [43] take each entry in the innovation vector $y_k - M\psi^f$ (Eq. (22)) as the distance between one point on the predicted interface and one point on the observed interface. This framework corresponds to the geometric view. [44] compute y_k and ψ^f in the innovation vector from a progress variable. The progress variable assumes values between zero and one, with most intermediate values assumed near the interface. This framework is an approximation of the set-theoretic view. [42] compute y_k and ψ^f from a level-set function. This framework corresponds to the analytic view, and comes closest to the level-set data assimilation framework presented in this study. The motion of an interface depends on the values of the level-set function in the immediate environment of the interface, whereas the values of the level-set function are not unique away from the interface. The level-set data assimilation framework presented in this study (i) introduces an additional constraint in agreement with the analytic view, (ii) validates the choice of the constraint for two analytical test cases in one and two dimensions respectively, and (iii) addresses shortcomings in the aforementioned frameworks.

In order to construct the probabilistic state space, the derivation of the level-set method is revisited in this section. The centerpiece of this derivation is the Hamilton-Jacobi equation. In

theory, the level-set method may alternatively be derived as the transport of a passive scalar quantity [8, Chapter 2]. Although the value of the level-set function is well defined at the interface, the choice of values for the level-set function away from the interface is in general not unique, and would require an ad-hoc constraint. It is demonstrated how the necessary constraint to the Hamilton-Jacobi equation naturally follows from the choice of phase space. Finally, it is shown that the solutions to the Hamilton-Jacobi equation, the so-called generating functions, form the appropriate probabilistic state space for data assimilation. Two complimentary level-set algorithms are combined to solve the Hamilton-Jacobi equation computationally efficiently: the narrow-band method [40] and the fast-marching method [41]. Along with the ensemble Kalman filter (Theorem 4), the three algorithms form the backbone of our level-set data assimilation framework.

3.1. Hamilton-Jacobi equation

The laws of motion of an interface shall be given by

$$\frac{d\mathbf{r}}{dt} = \mathbf{u}(\mathbf{r}) \quad , \quad (33)$$

where \mathbf{r} is the position of one point on the interface, and \mathbf{u} is the velocity; vector-valued variables and functions are denoted by boldface letters. In Hamiltonian mechanics, motion is described in phase space using generalized coordinates and momenta [55, Chapter 3]. With interfaces in mind, a natural choice for the phase space is to use the position \mathbf{r} as generalized coordinates and the normal vector \mathbf{n} as generalized momenta.

For the Hamilton-Jacobi equation, the generating function $G(\mathbf{r}, t)$ shall satisfy the following properties:

- The initial interface $\mathbf{r}(0)$ is given by a level set $G(\mathbf{r}(0), 0) = \text{const}$.
- The moving interface is identified as the level set $G(\mathbf{r}(t), t) = G(\mathbf{r}(0), 0)$.
- Huygens' principle [55, Chapter 9] states that $\nabla G = \mathbf{n}$. Given the choice of phase space, this is a necessary relationship between the generating function and the generalized coordinates and momenta.

The first two properties are present in every front-capturing method [8]. The third property, Huygens' principle, follows from Hamiltonian mechanics (Appendix B). Physically speaking, every point on the interface is interpreted as the source of a wave. As the waves propagate

230 through the domain, they connect each point on the interface to every other point in the domain. Geometrically speaking, Huygens' principle gives an eikonal field [56].

It remains to show how the generating function translates into a level-set method. The Lagrangian $\mathcal{L}(\mathbf{r}, \dot{\mathbf{r}}, t)$ is given by (Appendix B, Lemma 1)

$$\mathcal{L}(\mathbf{r}(t), \dot{\mathbf{r}}(t), t) = 0 \quad , \quad (34)$$

$$\text{s.t. } G(\mathbf{r}(t), t) - G(\mathbf{r}(0), 0) = 0 \quad , \quad (35)$$

$$\text{s.t. } \mathbf{n}(t) \cdot \mathbf{n}(t) - 1 = 0 \quad . \quad (36)$$

The Hamiltonian $\mathcal{H}(\mathbf{r}, \mathbf{n}, t)$, subject to the same constraints as the Lagrangian $\mathcal{L}(\mathbf{r}, \dot{\mathbf{r}}, t)$ (Eq. (35), (36)), is given by (Appendix B, Lemma 2)

$$\mathcal{H}(\mathbf{r}(t), \mathbf{n}(t), t) = \mathbf{u}(\mathbf{r}(t)) \cdot \mathbf{n}(t) \quad . \quad (37)$$

Finally, the Hamilton-Jacobi equation is given by (Appendix B, Theorem 9)

$$\frac{\partial G}{\partial t} + \mathbf{u}(\mathbf{r}(t)) \cdot \mathbf{n}(t) = 0 \quad . \quad (38)$$

The solution to the Hamilton-Jacobi equation, the generating function $G(\mathbf{r}, t)$, is defined over the whole domain. As such, it represents the state of the interface in the probabilistic state space. Eq. (38) is solved near the interface. Away from the interface, the generating function is constrained by (Eq. (36))

$$|\nabla G| = 1 \quad . \quad (39)$$

Eq. (39) is an eikonal equation [56]. The solutions to this eikonal equation are signed distance functions.

3.2. Level-set data assimilation framework

235 The level-set data assimilation framework presented in this study combines three algorithms: (i) the narrow-band method to solve the Hamilton-Jacobi equation near the interface [40], (ii) the fast-marching method to extend the eikonal field from the narrow band to the whole domain [41], and (iii) the ensemble Kalman filter and smoother to assimilate data [18].

The narrow-band method forms the backbone of the level-set data assimilation framework [40].
240 It solves the Hamilton-Jacobi equation (Eq. 38) in a narrow band near the interface. Using a

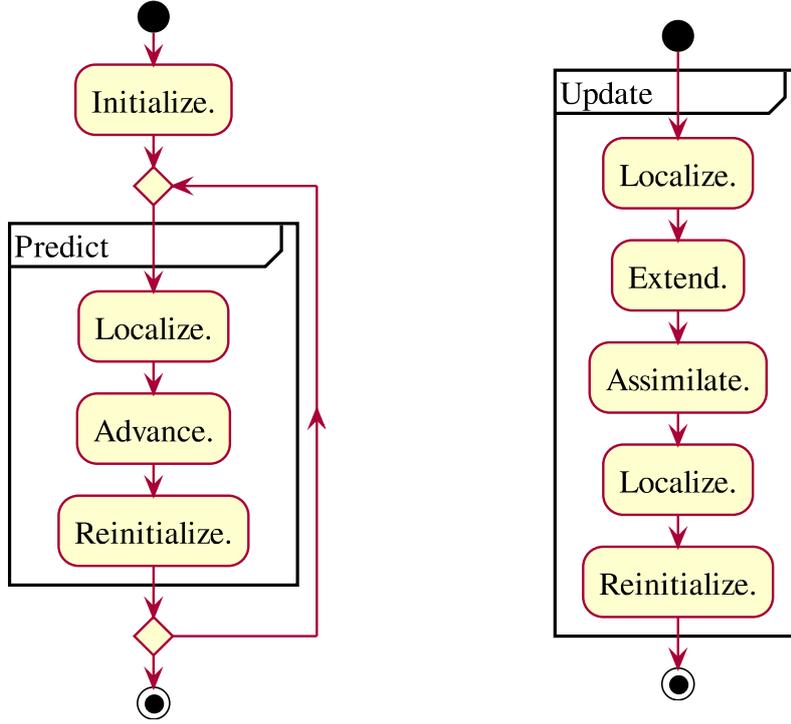


Figure 2: Activity diagrams of narrow-band method (left) and level-set data assimilation (right).

narrow band reduces the computational cost by one order of magnitude compared to finding the generating function over the whole domain [57]. In a numerical simulation, the spatial resolution shall be denoted by Δx and the temporal resolution by Δt . The temporal resolution has to satisfy the Courant-Friedrich-Lewy condition $u\Delta t < \Delta x$ at every point on the interface [40].

245 The narrow-band method consists of the following steps (Fig. 2, left):

Initialize. Before the first timestep, the signed distance field has to be known in a sufficiently wide neighborhood of the initial interface. It may be obtained from either a previously computed generating function or exact knowledge of the position of the initial interface.

250 **Localize.** At the timestep k , a band $T^k = \{\mathbf{r} : G(\mathbf{r}, t_k) < \gamma\Delta x\}$ is formed, where γ is a constant chosen according to the width of the narrow band. A second, enveloping band $N^k = \{\mathbf{r} : G(\mathbf{r} + \mathbf{r}', t_k) < \gamma\Delta x, \exists |\mathbf{r}'| < \Delta x\}$ is formed. It follows from the Courant-Friedrich-Lewy condition that the band N^k contains the band T^{k+1} formed at the next timestep $k + 1$.

Advance. The generating function over the band N^k is evolved by one timestep. Note that the solution $\tilde{G}^{k+1}(\mathbf{r})$ is not the generating function at the next timestep $k + 1$. While

255 the position of the interface correctly coincides with the level set of the solution $\tilde{G}^{k+1}(\mathbf{r})$, Huygens' principle is no longer satisfied [58].

Reinitialize. The following Hamilton-Jacobi equation is solved over the band N^k until steady state is reached [58]:

$$\frac{\partial \tilde{G}}{\partial \tilde{t}} + \text{sgn}(\tilde{G}(\mathbf{r}, \tilde{t})) \left(|\nabla \tilde{G}| - 1 \right) = 0 \quad , \quad (40)$$

$$\tilde{G}(\mathbf{r}, 0) = \tilde{G}^{k+1}(\mathbf{r}) \quad . \quad (41)$$

The steady-state solution $\tilde{G}(\mathbf{r}, \tilde{t} \rightarrow \infty)$ gives the generating function $G(\mathbf{r}, t_{k+1})$ over the band N^k at the timestep $k + 1$. Note that this Hamilton-Jacobi equation is a numerical continuation of the eikonal equation (Eq. (39)), where \tilde{t} denotes a pseudo-time [59].

260 When observations are available, the prediction step is followed by an update step (Fig. 2, right). Like the prediction step, the update step involves localization and reinitialization so that the narrow band follows the position of the interface. Instead of the advancement step, there is an extension step and an assimilation step:

Extend. While a narrow band is sufficient to capture the motion of an interface, data assimilation has to be performed in one probabilistic state space for all interfaces. In a numerical simulation, the state comprises the values of the generating function at every grid point. The fast-marching method extends the generating function from the narrow band to the whole domain, thus returning a well-defined state. The complexity of the fast-marching method is $\mathcal{O}(m^2 \log(m))$ in two dimensions or $\mathcal{O}(m^3 \log(m))$ in three dimensions, where m denotes the number of grid points in one dimension [41]. While the generating function over the whole domain is relatively expensive to compute, the fast-marching method is only called when observations are available. If combined with the narrow-band method (complexity $\mathcal{O}(m)$ or $\mathcal{O}(m^2)$ [40]), this level-set method becomes affordable overall.

Assimilate. The ensemble Kalman filter (Theorem 4) and/or smoother (Theorem 6) are applied. The assimilation step is followed by a localization step and a reinitialization step in order to prepare the narrow band for the next prediction step. Note that the result of data assimilation is in general not an eikonal field. While the band localized after data assimilation may in theory be too narrow, it can be shown that the slopes in the filtered solutions are in expectation less than or equal to unity (Appendix C).

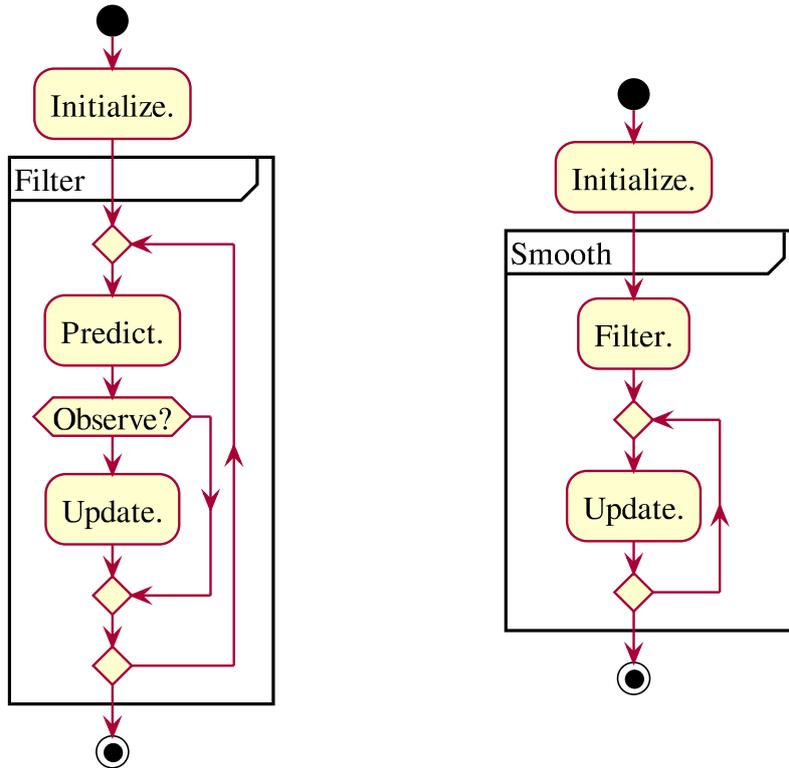


Figure 3: Activity diagrams of filtering (left) and smoothing (right) frameworks. The prediction and update steps are defined in Fig. 2. Note how the smoothing framework (right) builds on top of the filtering framework (left).

280 Combining the building blocks of the level-set method and data assimilation gives the level-set
 filtering and smoothing frameworks (Fig. 3). In the following subsections, the update step in the
 level-set filtering framework is verified using two analytical test cases: In a one-dimensional test
 case, the level-set filtering framework, which represents the analytic view of data assimilation, is
 contrasted with the geometric and set-theoretic views of data assimilation. It is shown that the
 285 analytic view of data assimilation works correctly, which sets the stage for its generalization to
 higher dimensions. Afterwards, the level-set filtering framework is applied to a two-dimensional
 test case, and the effect of the observations on the shape of the interface is discussed. It is shown
 how various parameters such as the number of observations and the observation error may affect
 the performance of the level-set data assimilation framework.

The predicted and observed positions x and y of an interface in one dimension are respectively given by (Fig. 4)

$$x \sim \mathcal{N}(x \mid \mu_\psi, C_{\psi\psi}) \quad , \quad \mu_\psi = 0 \quad , \quad C_{\psi\psi} = 0.64 \quad , \quad (42)$$

$$y = x \sim \mathcal{N}(x \mid \mu_\epsilon, C_{\epsilon\epsilon}) \quad , \quad \mu_\epsilon = 1 \quad , \quad C_{\epsilon\epsilon} = 0.36 \quad . \quad (43)$$

In one dimension, an interface reduces to a point. The normal vector on the interface is unique up to its orientation. Geometric data assimilation using the Kalman filter (Theorem 3) reduces to optimal interpolation of the positions μ_ψ and μ_ϵ (Eq. (23)):

$$x = \mu_\psi + \frac{C_{\psi\psi}}{C_{\epsilon\epsilon} + C_{\psi\psi}} (\mu_\epsilon - \mu_\psi) = 0.64 \quad . \quad (44)$$

The filtered position x is closer to the observation than to the prediction in accordance with the variances $C_{\psi\psi}$ and $C_{\epsilon\epsilon}$ (Fig. 4, left).

For set-theoretic data assimilation, the characteristic function χ is assumed to be bounded by ± 1 with the interface at $\chi = 0$:

$$\chi(x) = \begin{cases} -1 & x < c \\ 0 & x = c \\ 1 & x > c \end{cases} \quad , \quad c \sim \mathcal{N}(c \mid \mu_\psi, C_{\psi\psi}) \quad . \quad (45)$$

The characteristic function χ follows a Bernoulli distribution, which is the appropriate probability distribution for a random variable with two possible outcomes, at every location x [60, Chapter 6]:

$$p(\chi(x) = 1) = \Phi\left(\frac{x - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right) \quad , \quad p(\chi(x) = -1) = 1 - \Phi\left(\frac{x - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right) \quad , \quad (46)$$

where Φ denotes the cumulative density function of the normal distribution. The characteristic function χ does not follow a normal distribution whereas the Kalman filter requires the mean $\bar{\chi}$ and the covariance function $C_{\chi\chi}$:

$$\bar{\chi}(x) = 2\Phi\left(\frac{x - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right) - 1 \quad , \quad C_{\chi\chi}(x_1, x_2) = 4\Phi\left(\frac{\min(x_1, x_2) - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right) \left(1 - \Phi\left(\frac{\max(x_1, x_2) - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right)\right) \quad , \quad (47)$$

Note that the mean $\bar{\chi}$ is not a characteristic function. Set-theoretic data assimilation using the Kalman filter (Theorem 3) gives

$$\chi(x) = 2\Phi\left(\frac{x - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right) - 1 - \frac{4\Phi\left(\frac{\min(x, \mu_\epsilon) - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right) \left(1 - \Phi\left(\frac{\max(x, \mu_\epsilon) - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right)\right)}{1 + 4\Phi\left(\frac{\mu_\epsilon - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right) \left(1 - \Phi\left(\frac{\mu_\epsilon - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right)\right)} \left(2\Phi\left(\frac{\mu_\epsilon - \mu_\psi}{\sqrt{C_{\psi\psi}}}\right) - 1\right) \quad (48)$$

$$= 2\Phi(1.25x) - 1 - \frac{4\Phi(1.25 \min(x, 1)) (1 - \Phi(1.25 \max(x, 1)))}{1 + 4\Phi(1.25) (1 - \Phi(1.25))} (2\Phi(1.25) - 1) \quad . \quad (49)$$

The interface is localized at $x \approx 0.139$ (Fig. 4, middle). The filtered position x is closer to the prediction than to the observation although the prediction variance $C_{\psi\psi}$ is larger than the observation variance $C_{\epsilon\epsilon}$. Note that the filtered position x is independent of the variance $C_{\epsilon\epsilon}$.
295

For analytic data assimilation, the predicted and observed positions x and y are viewed as points on the level sets of the predicted and observed generating functions respectively. It follows from Eq. (42) that the predicted generating function is given by

$$G(x) \sim \mathcal{N}(G(x) \mid x - \mu_\psi, C_{\psi\psi}) \quad . \quad (50)$$

The mean \bar{G} and the covariance function \mathcal{C}_{GG} are given by

$$\bar{G}(x) = x - \mu_\psi \quad , \quad \mathcal{C}_{GG}(x_1, x_2) \equiv C_{\psi\psi} \quad . \quad (51)$$

Analytic data assimilation using the Kalman filter (Theorem 3) on the generating function gives

$$G(x) = x - \mu_\psi - \frac{C_{\psi\psi}}{C_{\epsilon\epsilon} + C_{\psi\psi}} (\mu_\epsilon - \mu_\psi) = x - 0.64 \quad . \quad (52)$$

The filtered position x is given by the level set of the filtered generating function G (Eq. (52)), which is identical to the result from geometric data assimilation (Fig. 4, right).

In summary, geometric, set-theoretic and analytic data assimilation are compared. Set-theoretic data assimilation gives implausible results in one dimension, and is thus discarded from any further consideration. Geometric and analytic data assimilation give consistent results in one dimension. Nevertheless, the generalization of geometric data assimilation to higher dimensions is not straightforward because it requires a one-to-one correspondence between the predicted interface and the observations. In the next test case, the generalization of analytic data assimilation, as the most feasible framework, to higher dimensions is investigated.
300

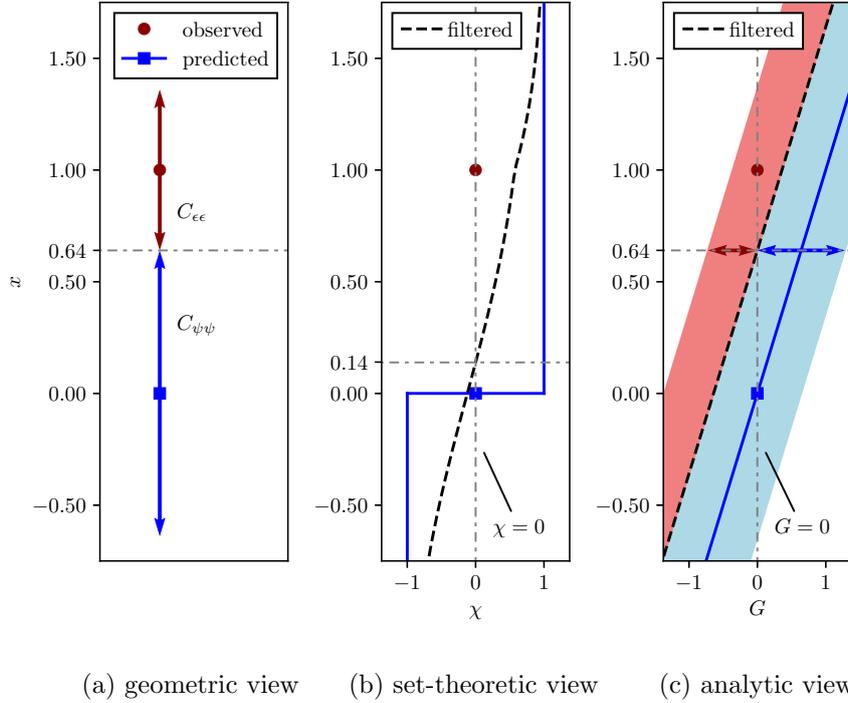


Figure 4: Data assimilation from geometric (left), set-theoretic (middle) and analytic view (right). The blue square at $x = 0$ marks the predicted position of the interface. The red circle at $x = 1$ marks the observed position of the interface. The grey, dash-dotted lines give the coordinates of the filtered position of the interface. In one dimension, data assimilation from the geometric view (left) reduces to optimal interpolation of the x coordinates, weighted by $C_{\psi\psi}$ and $C_{\epsilon\epsilon}$ respectively. The geometric solution may be considered the most intuitive one. In the middle figure, the Kalman filter is applied to the characteristic function χ (blue line). The set-theoretic solution (dashed line) localizes the interface at $x \approx 0.139$, significantly differently from the geometric solution. This shows that oft-used physical quantities, which exhibit behavior similar to a characteristic function in the vicinity of the interface of interest, are not suitable for data assimilation. In the right figure, the Kalman filter is applied to the generating function G (blue line). The analytic solution (dashed line) localizes the interface at $x = 0.64$, identically to the geometric solution.

The distance field S of a corner in (x_0, y_0) is given by

$$S(x, y; x_0, y_0) = \begin{cases} x - x_0 & \frac{\pi}{4} \leq \theta \leq \pi \\ y - y_0 & -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{4} \\ -\sqrt{(x - x_0)^2 + (y - y_0)^2} & \pi \leq \theta \leq \frac{3\pi}{2} \end{cases} . \quad (53)$$

For prediction, the coordinates of the corner (x_0, y_0) are taken to be independent and normally distributed:

$$x_0 \sim \mathcal{N}(x_0 | 0, 1) \quad , \quad y_0 \sim \mathcal{N}(y_0 | 0, 1) \quad . \quad (54)$$

A sketch of the corner and its median distance field $S(x, y; 0, 0)$ are given in Fig. 5.

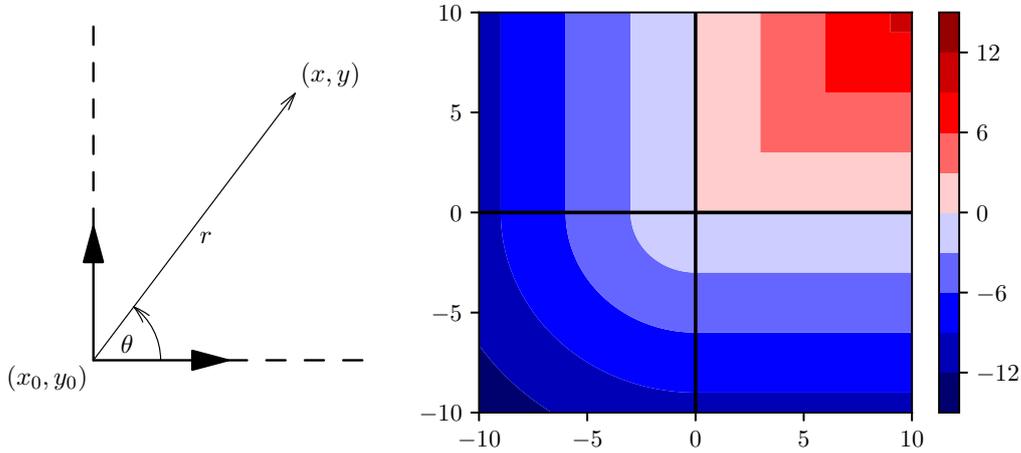


Figure 5: Sketch of a corner (left) and its median distance field $S(x, y; 0, 0)$ (right). The walls of the corner (left, dashed line), which are represented by the zero-level set in analytic data assimilation, are parallel to the positive x and y axes and share their orientations (right). The points with $0 < \theta < \frac{\pi}{2}$ are defined to have positive distance.

For analytic data assimilation, the predicted position x is again viewed as the level set of a predicted generating function. It follows from Eq. (53) and (54) that the predicted generating function is given by

$$G(x, y) \sim \mathcal{N}(G(x, y) | \bar{G}(x, y), C_{GG}(x_1, y_1, x_2, y_2)) \quad , \quad (55)$$

where the mean \overline{G} and the covariance function C_{GG} are given by

$$\overline{G}(x, y) = \mathbb{E}_{x_0, y_0}[S](x, y) \quad (56)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(x, y; \xi, \eta) p(x_0 = \xi, y_0 = \eta) d\xi d\eta \quad , \quad (57)$$

$$C_{GG}(x_1, y_1, x_2, y_2) = \mathbb{E}_{x_0, y_0} [(S - \mathbb{E}_{x_0, y_0}[S]) \times (S - \mathbb{E}_{x_0, y_0}[S])] (x_1, y_1, x_2, y_2) \quad (58)$$

$$= \mathbb{E}_{x_0, y_0} [S \times S] (x_1, y_1, x_2, y_2) - \mathbb{E}_{x_0, y_0}[S](x_1, y_1) \mathbb{E}_{x_0, y_0}[S](x_2, y_2) \quad (59)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(x_1, y_1; \xi, \eta) S(x_2, y_2; \xi, \eta) p(x_0 = \xi, y_0 = \eta) d\xi d\eta \\ - \mathbb{E}_{x_0, y_0}[S](x_1, y_1) \mathbb{E}_{x_0, y_0}[S](x_2, y_2) \quad , \quad (60)$$

where \mathbb{E}_{x_0, y_0} denotes the expected value over the random variables x_0 and y_0 , and the ' \times ' symbol denotes the Cartesian product of two random variables. The mean \overline{G} and the covariance function C_{GG} are computed from Eq. (57) and (60) using Gauss-Hermite quadrature [61]. Alternatively, the mean and the variance function can be computed from the following one-dimensional integrals on the unit circle:

$$\mathbb{E}_{x_0, y_0}[S](x, y) = \int_{\pi/4}^{\pi} \cos(\theta) g(\theta) d\theta + \int_{-\pi/2}^{\pi/4} \sin(\theta) g(\theta) d\theta - \int_{\pi}^{3\pi/2} g(\theta) d\theta \quad , \quad (61)$$

$$\mathbb{E}_{x_0, y_0}[S^2](x, y) = \int_{\pi/4}^{\pi} \cos(\theta)^2 h(\theta) d\theta + \int_{-\pi/2}^{\pi/4} \sin(\theta)^2 h(\theta) d\theta + \int_{\pi}^{3\pi/2} h(\theta) d\theta \quad . \quad (62)$$

The auxiliary functions g and h are given by

$$g(\theta) = \varphi(w) [(v^2 + 1) \Phi(v) + v\varphi(v)] \quad , \quad h(\theta) = \varphi(w) [(v^2 + 2) \varphi(v) + (3v + v^3) \Phi(v)] \quad , \quad (63)$$

where φ and Φ denote the probability density function and cumulative distribution function of the normal distribution. The natural coordinates (v, w) are given by

$$v = x \cos(\theta) + y \sin(\theta) \quad , \quad w = x \sin(\theta) - y \cos(\theta) \quad . \quad (64)$$

The mean and the standard deviation of the predicted generating function are shown in Fig. 6. Artefacts from the interaction between the nonlinearity of distance fields and the assumption of normal distributions are visible: The mean of the predicted generating function (Fig. 6, left) is rounded in the corner whereas distance fields are supposed to have right angles (Fig. 5). The standard deviation of the predicted generating function (Fig. 6, right) assumes non-unity values along $\theta = 45^\circ$ although the coordinates of the corner are independent and normally distributed (Eq. (54)). The effects of the artefacts in the predicted generating function are now examined in various observation scenarios.

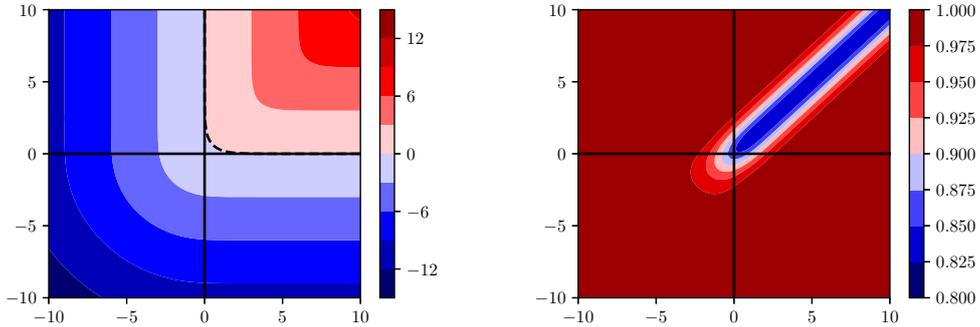


Figure 6: Mean (left) and standard deviation (right) of predicted generating function. The zero-level set in the mean is marked by a dashed line. Away from $\theta = 45^\circ$, the mean of the generating function comes close to its median (Fig. 5, right). Note that the distance field S (Eq. (53)) is smooth but not differentiable on $\theta = 45^\circ$. This explains why the right angles observed in the median are not preserved but are rounded in the mean. At the origin, the right angle is thus offset by $\frac{1}{2\sqrt{\pi}} + \frac{1}{4}\sqrt{\frac{\pi}{2}} \approx 0.5954$. Away from the origin, the right angles are asymptotically offset by $\frac{1}{\sqrt{\pi}} \approx 0.5642$. The same can be observed in the standard deviation. Away from $\theta = 45^\circ$, the standard deviation is close to unity, which is identical to the standard deviation in the coordinates of the corner (Eq. (54)). At the origin, the standard deviation assumes a value of $\sqrt{\frac{5}{4} - \frac{3}{4\pi} - \frac{\pi}{32} - \frac{1}{4\sqrt{2}}} \approx 0.8581$. Away from the origin, the standard deviation asymptotically assumes a value of $\sqrt{1 - \frac{1}{\pi}} \approx 0.8256$.

Unlike the one-dimensional example, where one point fully characterizes an interface, observations in two (and higher) dimensions vary in resolution, ranging from the position of one point on the interface to a full sampling of the interface. In either case, the observations y_k vanish according to the definition of a level set (Section 3.1). Eq. (23) simplifies to

$$\psi^a = \psi^f - \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} M \psi^f \quad . \quad (65)$$

The measurement operator M is defined by

$$(M\psi^f) = \begin{pmatrix} \psi^f(\mathbf{r}_1) \\ \vdots \\ \psi^f(\mathbf{r}_m) \end{pmatrix} \quad , \quad (66)$$

315 where $\mathbf{r}_1, \dots, \mathbf{r}_m$ are the positions of the m observed points on the interface.

In Fig. 7, the influence of one observation depending on its location is investigated. It is evident that one observation is insufficient to represent a corner. Nevertheless, analytic data assimilation gives qualitatively meaningful results in that the observation shifts the vertical and horizontal edges proportionally depending on the location of the observation. This is fundamen-
320 tally different from geometric data assimilation (Fig. 4, left), where every point on the interface

is required to correspond to an observation. This makes the presented (analytic) level-set data assimilation framework fully front-capturing (rather than front-tracking), both with respect to the description of the motion of the interface and the assimilation of data. As the number of observation points increases, the right angle in the corner is retrieved (Fig. 8).

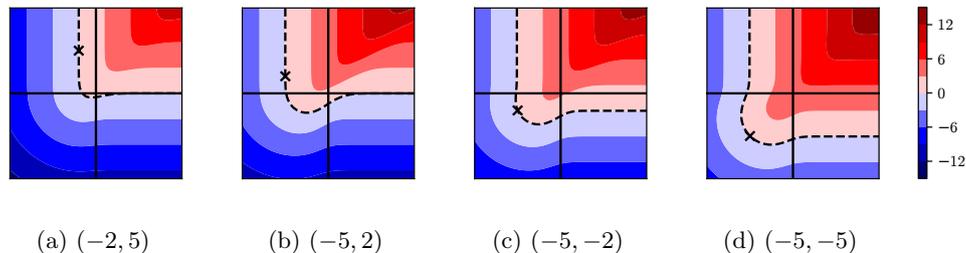
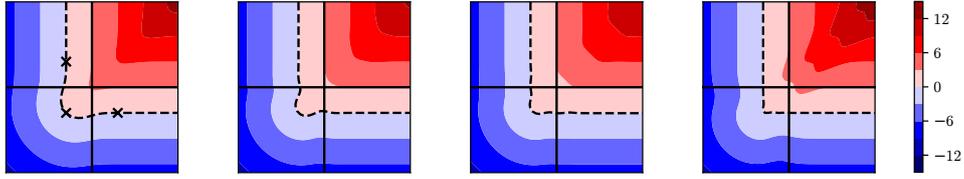


Figure 7: Means of filtered generating functions for one observation of the interface at various locations. The observed location is marked by a cross. Its coordinates are specified in each caption. The zero-level set is marked by a dashed line. Without loss of generality, the observation error is set to zero. (a) The observation is located significantly closer to the vertical than the horizontal edge of the predicted generating function (Fig. 6, left). Hence, the observation is primarily associated with points on the vertical edge. Data assimilation thus mainly shifts the points on the vertical edge while the vertex and the horizontal edge remain largely unaffected. (b) As the location of the observation moves downwards, the observation gets more strongly associated with the vertex of the corner while the horizontal edge remains largely unaffected. (c) As the location of the observation crosses the x -axis, data assimilation also shifts the points on the horizontal edge. (d) As the location of the observation reaches the diagonal, data assimilation equally affects the vertical and horizontal axes of the corner.

It is worth pointing out the probabilistic nature of the presented level-set data assimilation framework. Similar to soft clustering [60], the points on the predicted interface are not associated with distinct observations but with all observations to varying degrees. If we consider ψ^a to be a function of ψ^f , $C_{\psi\psi}^f$, y , $C_{\epsilon\epsilon}$ and M (Eq. (23)), the sensitivity to data is given by

$$\frac{\partial\psi^a}{\partial y} = \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1}, \quad (67)$$

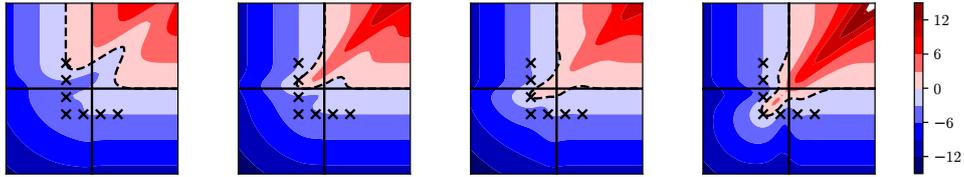
325 where the columns represent the sensitivities to individual observations. The shift in the zero-level set (Fig. 8) is the superposition of the contributions due to the individual observations. For example, while the points on the vertical edge far away from the vertex are mainly affected by the uppermost observation (Fig. 9a), the other observations have a vanishing effect on them as well (Fig. 9b-d). Nevertheless, the shape deformations are more localized compared to Fig. 7.
 330 While Eq. (67) does not formally depend on the observations y , it includes information about the number and the locations of the observations signified by the measurement operator M .



(a) 3 observations. (b) 7 observations. (c) 15 observations. (d) 31 observations.

Figure 8: Means of filtered generating functions for multiple observations of the interface at various resolutions. The observed locations are marked by crosses and the zero-level sets by dashed lines. The observations are equidistantly spaced between $(-3, 3)$ and $(3, -3)$ via $(-3, -3)$. The number of observations is specified in each caption. For better readability, the observations are not marked in (b)-(d). Without loss of generality, the observation error is set to zero. From (a) to (d), as the number of the observation increases, the observations give a clearer image of the observed corner. Hence, the zero-level set of the filtered generating function more closely resembles a corner.

Note that each observation of interest leaves the values of the generating function at the other observation locations intact.



(a) $(-3, 3)$ (b) $(-3, 1)$ (c) $(-3, -1)$ (d) $(-3, -3)$

Figure 9: Means of filtered generating functions due to individual observations. The observed locations are marked by crosses and the zero-level sets by dashed lines. The observations are equidistantly spaced between $(-3, 3)$ and $(3, -3)$ via $(-3, -3)$. The location of the observation of interest is specified in each caption. Without loss of generality, the observation error is set to zero.

In summary, analytic data assimilation is applied to a prototypical two-dimensional shape, a
 335 corner. Within the level-set data assimilation framework, the differences between distance fields,
 predicted generating functions and filtered generating functions are illustrated. In particular,
 the effects of the number of observation points and their locations are studied. In the following
 section, the level-set data assimilation framework is applied to a more realistic experiment. In
 general, the predicted generating function is taken as a composition of straight sections, which
 340 exhibit the behavior discussed for the one-dimensional test case, and strongly curved sections

insufficiently resolved by the observations, which exhibit the behavior discussed for the two-dimensional test case. The theoretical insights from this section will be referenced throughout the next section.

4. Twin experiments for a premixed flame inside a duct

345 Thermoacoustic instabilities are a persistent challenge in the design of jet and rocket engines: Velocity and pressure oscillations inside the combustion chamber interact with the flame and cause an unsteady heat release rate. If moments of higher heat release rate coincide with moments of higher pressure (and lower heat release rate with lower pressure), acoustic oscillations arise. This can lead to large-amplitude oscillations causing structural damage in the jet or rocket
350 engine [62, 63].

The so-called G -equation model is a reduced-order model to study the flame dynamics which lead to heat release rate perturbations [9, 10]. The premixed flame is modeled as an interface captured by a level set. The velocity of a point in the level set is the sum of the flame speed, which is normal to the interface and points towards the unburnt gas, and the underlying flow field. The
355 underlying flow field includes both hydrodynamic and acoustic contributions. The G -equation model is used to compute flame transfer functions (FTF) or flame describing functions (FDF), which give the heat release perturbation to a given velocity perturbation. Coupled with linear acoustics models, the G -equation model has been very successful in qualitatively characterizing the linear and nonlinear dynamics of self-excited thermoacoustic oscillations [11, 12, 64, 65].

360 We demonstrate our level-set data assimilation framework by performing twin experiments for the G -equation model applied to a ducted premixed Bunsen flame under acoustic forcing. In our twin experiments, both the model predictions and the observations come from G -equation simulations, but with different sets of parameters [66, Chapter 9]. This leads to uncertainties in the parameters and, as a result, in the states. In the absence of uncertainties in the model, the
365 twin experiment is an important benchmark in the quantitative assessment of a data assimilation framework because it becomes possible to compare filtered and smoothed solutions to a reference solution at all times [66, Chapter 8]. An example for the application of our level-set data assimilation framework to more realistic observations from a direct numerical simulation can be found in [67, 68].

The G -equation is given by [8, Chapter 2]

$$\frac{\partial G}{\partial t} + [\mathbf{u}(\mathbf{r}, t) - s_L \mathbf{n}] \cdot \nabla G = 0 \quad , \quad (68)$$

where the underlying flow field is denoted by $\mathbf{u}(\mathbf{r}, t)$ and the laminar flame speed by s_L . For $|\nabla G| = 1$, the G -equation is formally equivalent to the Hamilton-Jacobi equation (Eq. 38). The underlying flow field $\mathbf{u}(\mathbf{r}, t)$ is a superposition of the steady, inviscid flow field U and the time-dependent perturbation velocity field $\mathbf{u}'(\mathbf{r}, t)$. In the reduced-order model of $\mathbf{u}'(\mathbf{r}, t)$, the perturbation has an amplitude εU , and travels downstream at the phase speed U/K , where K and ε are model parameters [69]. In the axisymmetric case, the cylindrical components of the perturbation velocity field are given by

$$\mathbf{u}'(\mathbf{r}, t) = \begin{pmatrix} u'_\rho \\ u'_\theta \\ u'_z \end{pmatrix} = \begin{pmatrix} \frac{1}{2} k \rho \varepsilon U \sin(kz - \omega t) \\ 0 \\ \varepsilon U \cos(kz - \omega t) \end{pmatrix} \quad , \quad (69)$$

where the wavenumber k and the angular frequency ω satisfy the dispersion relation $\omega(k) = Uk/K$. The perturbation velocity field $\mathbf{u}'(\mathbf{r}, t)$ is, mathematically speaking, divergence-free, and, physically speaking, satisfies the continuity equation.

In Fig. 10, a sketch of the ducted premixed flame is shown. The ducted premixed flame is forced at an angular frequency of 277.78 rad/s. Tab. 4 shows the measurements of the burner and the tube. The G -equation is numerically solved using the narrow-band level-set method with distance reinitialization [40, 58]: The computational domain is discretized using a weighted essentially non-oscillatory (WENO) scheme in space and a total-variation diminishing (TVD) version of the Runge-Kutta scheme in time, which give up to fifth-order accuracy in space and third-order accuracy in time [70, 71, 72]. At the base of the flame, a rotating boundary condition is used [73]. The G -equation solver has been verified in a number of studies [74, 11, 12, 75, 65]. These studies showed that the G -equation reproduces the dynamics of premixed flames only qualitatively, not quantitatively. With our level-set data assimilation framework, significant improvements on quantitative predictions are achieved. In axial and radial coordinates, a uniform 401×401 Cartesian grid is used, which corresponds to $30 \mu\text{m} \times 180 \mu\text{m}$ grid cells. An advection Courant-Friedrich-Lewy (CFL) number of 0.02 is chosen, which corresponds to a timestep of $3.6 \mu\text{s}$ or approximately 6,283 timesteps over one period.

Snapshots of the reference solution to the G -equation for the ducted premixed flame are shown in Fig. 11. The flame is attached to the burner lip, while the perturbations travel from the base

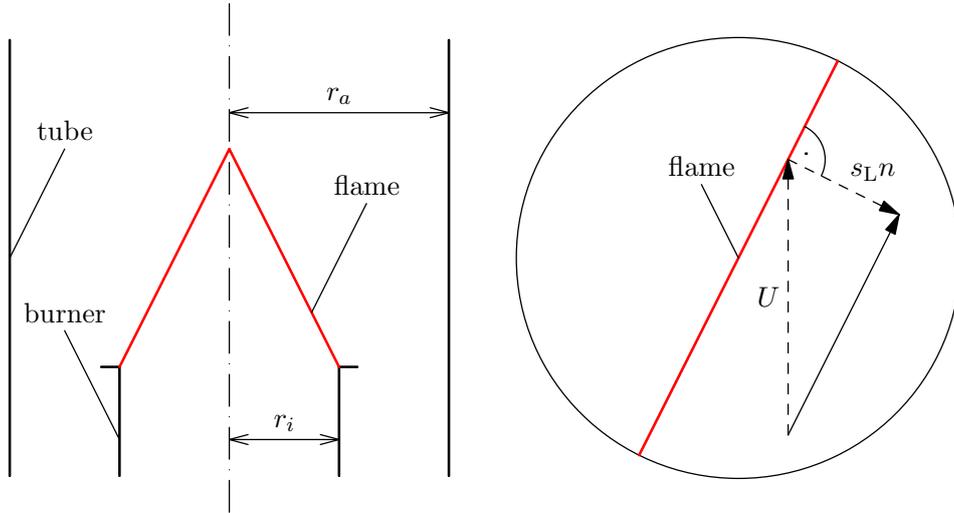


Figure 10: Sketch of the ducted premixed flame without perturbations (left). The superposition of the underlying flow field and the propagation of the premixed flame into the unburnt gas gives velocity vectors, which are locally tangential to the flame surface (right).

Table 4: G -equation parameters for the reference simulation.

Parameter	r_i [m]	r_a [m]	U [m/s]	s_L [m/s]	ω [rad/s]	K [-]	ε [-]
Parameter value	0.006	0.012	1.0	0.164	277.78	1.0	0.1

390 of the flame to the tip. When the perturbations are sufficiently large, a fuel-air pocket pinches off. In our reduced-order model, the perturbations are mainly governed by two non-dimensional parameters K and ε , which govern phase speed and amplitude respectively [69]. In practice, neither parameter is accurately known a priori, which is a major source of uncertainty.

4.2. Combined state and parameter estimation

395 Three variations of the twin experiment are performed:

Monte Carlo experiment. An ensemble of 32 G -equation simulations is performed. Each simulation has a different set of parameters K and ε . They are independently sampled from a normal distribution with a standard deviation of 20%.

State estimation. In addition to the procedure described for the Monte Carlo experiment, 400 the ensemble Kalman filter (Theorem 4) is applied every 1,000 timesteps to update the discretized generating function G while leaving each set of parameters K and ε unaltered. The observations are extracted from the reference solution (Section 4.1): The measurement

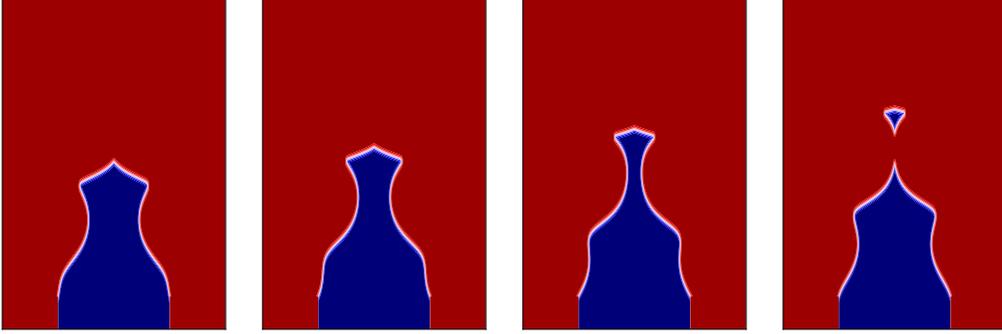


Figure 11: Snapshots of G -equation simulation over one period of harmonic forcing. The fuel-air mixture leaves the burner at the bottom of each frame. The infinitely thin flame surface separates the burnt gas (red) from the unburnt gas (blue). Each snapshot from left to right is distanced by a quarter of the forcing period.

operator M is an indicator matrix which identifies all grid points in the reference solution adjacent to its zero-level set. All aforementioned grid points are treated as observations of the flame surface. As such, all entries in the vector y_k are set to 0. The observation error σ_ϵ is set to $60 \mu\text{m}$, which corresponds to two grid cells.

Combined state and parameter estimation. In addition to the procedure described for the state estimation, the discretized generating function G is augmented by appending the parameters K and ϵ to the state vector (Eq. 6). Thus, both the state and the parameters are updated whenever data in the form of observations from the reference solution is assimilated.

In each variation of the twin experiment, the k -th entry in the state vector ψ is marginally distributed according to

$$\psi[k] \sim \mathcal{N}(\psi[k] \mid \bar{\psi}[k], C_{\psi\psi}[k, k]) \quad . \quad (70)$$

The mean $\bar{\psi}[k]$ and the variance $C_{\psi\psi}[k, k]$ are computed from Eq. (25). Explicitly, the likelihood for the flame surface to be found at the location of the k -th entry is given by

$$p[k] = \frac{1}{\sqrt{2\pi C_{\psi\psi}[k, k]}} \exp\left(-\frac{\bar{\psi}[k]^2}{2C_{\psi\psi}[k, k]}\right) \quad . \quad (71)$$

Alternatively, the logarithm of the normalized likelihood is given by [76]

$$\log\left(\frac{p[k]}{p_0[k]}\right) = -\frac{\bar{\psi}[k]^2}{2C_{\psi\psi}[k, k]} \quad . \quad (72)$$

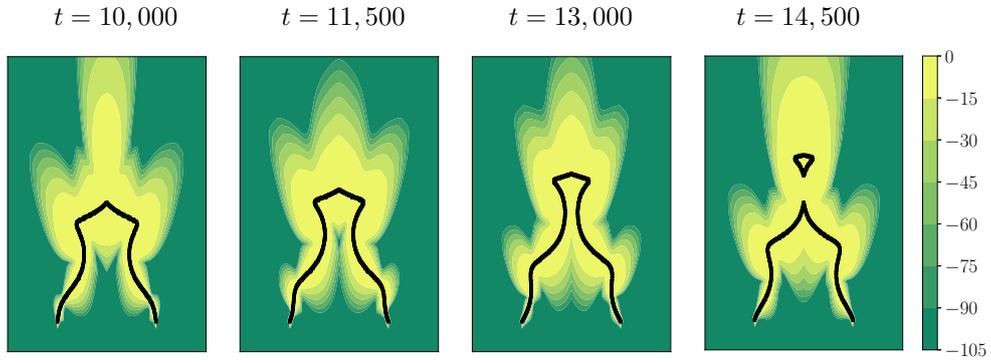
In Fig. 12, the logarithm of the normalized likelihood is shown for the three variations of the twin experiment. Its zero-level set gives the maximum-likelihood location of the flame surface.

The more negative the value at a location is, the less likely the flame surface is to be found there. In Fig. 12a, the logarithm of the normalized likelihood is shown for the Monto-Carlo experiment. As the perturbation travels from the base of the flame to the tip, the high-likelihood region for the location of the flame surface spreads out. The high-likelihood region is largest when fuel-air pockets pinch off, which represents maximal uncertainty. In Fig. 12b, the logarithm of the normalized likelihood is shown for the twin experiment with state estimation. A qualitative comparison to Fig. 12a shows that the high-likelihood region for the location of the flame surface is significantly tighter. While the high-likelihood region still grows as the perturbation travels, the regular assimilation of data suppresses it to the vicinity of the observed flame surface. In Fig. 12c, the logarithm of the normalized likelihood is shown for the twin experiment with combined state and parameter estimation. After few data assimilation cycles, the state-augmented ensemble Kalman filter has learnt the parameters K and ε almost exactly. Knowledge of the parameters enables highly precise predictions of the location of the flame surface, even during pinch-off events. Unlike in Fig. 12a and b, no growth in the high-likelihood region is qualitatively discernable.

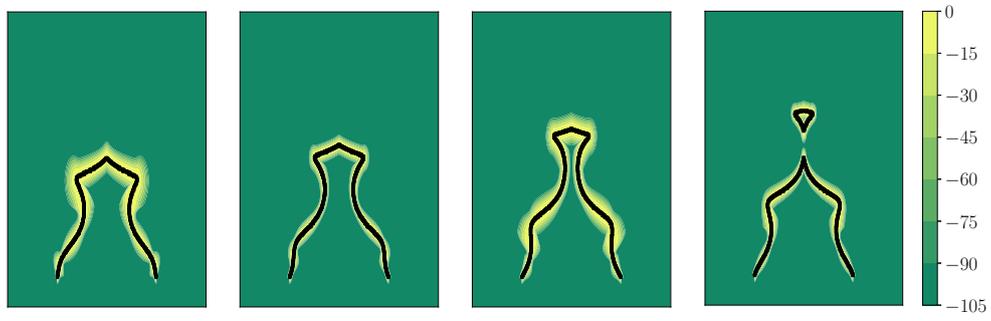
A global, more quantitative measure for the uncertainty in the location of the flame surface is the root mean square (RMS) error, which is defined as the square-root of the trace of the covariance matrix of the ensemble (Eq. (25)):

$$\text{RMS error} = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (\psi_j - \bar{\psi})^T (\psi_j - \bar{\psi})} \quad . \quad (73)$$

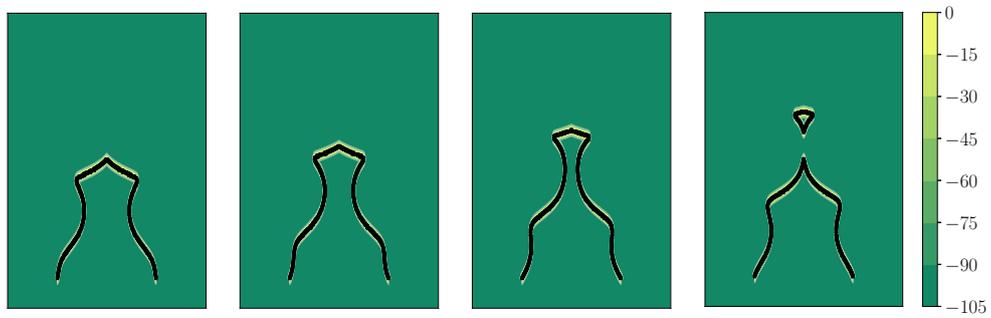
In Fig. 13, the RMS error is plotted over time for the three variations of the twin experiment from $t = 0$ for eight cycles. The RMS error is initially zero in all three twin experiments because all members in each ensemble share the same initial condition. In the Monte Carlo experiment (Fig. 13, blue line), the RMS error subsequently grows until it reaches a high-uncertainty plateau. Momentary spikes in the uncertainty occur approximately every 6000 timesteps, and coincide with the pinch-off events observed in the reference solution (Fig. 11). With state estimation (Fig. 13, orange line), the assimilation of data regularly suppresses the uncertainty as qualitatively observed in Fig. 12b. As a result, the predictions of the location of the flame surface significantly improve. With combined state and parameter estimation (Fig. 13, green line), the predictions further improve. Beginning with the first instance of data assimilation at timestep 1,000, the uncertainty steadily decreases until it reaches a low-uncertainty plateau after approximately 10,000 timesteps. At this point, the state-augmented ensemble Kalman filter has optimally calibrated the parameters, and the state has reached a statistically stationary state. The assessment based on the RMS error is supported by an analysis of the bias,



(a) Monte Carlo experiment.

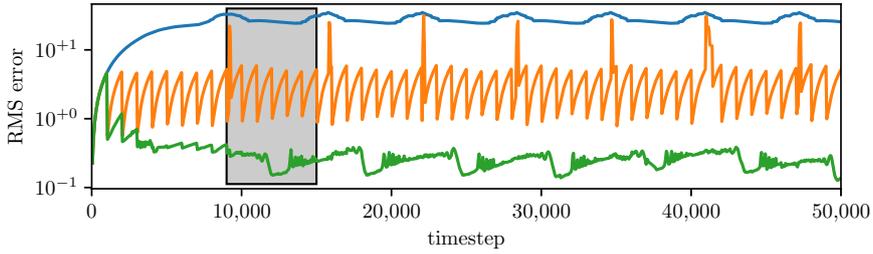


(b) State estimation.



(c) Combined state and parameter estimation.

Figure 12: Snapshots of the logarithm of the normalized likelihood (Eq. (72)) over one period for the three variations of the twin experiment respectively. The observations (black) are extracted from a reference solution to the G -equation (Fig. 11). High-likelihood (yellow) and low-likelihood (green) regions are shown. Each snapshot from left to right is distanced by a quarter of the forcing period.



— Monte Carlo experiment — state estimation — combined state and parameter estimation

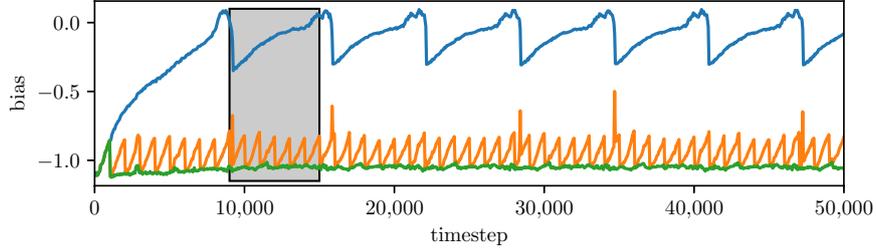
Figure 13: Logarithmic plot of root mean square (RMS) error over 50,000 timesteps for the Monte Carlo experiment (blue line) as well as the twin experiments with state estimation (orange line) and combined state and parameter estimation (green line). 6,283 timesteps correspond to one period of harmonic forcing. Data is assimilated every 1,000 timesteps. The black rectangle marks the time window depicted in Fig. 12.

which is plotted in Fig. 14 and defined as the difference between the mean prediction and the observations (Eq. (65)):

$$\text{bias} = \sqrt{(M\bar{\psi})^T (M\bar{\psi})} \quad . \quad (74)$$

Note that both the unknown parameters and the observation noise contribute to the bias. Hence, the bias is particularly useful as a quantitative measure in twin experiments, where the reference solution and thus the observation error are exactly known.

To gain insight into the effect of combined state and parameter estimation, it is instructive to analyze the probability distributions in the parameters. In Fig. 15, the marginal distributions in the normalized residuals of K and ε are shown, signified by their means and three-sigma confidence levels at every timestep. By timestep 15,000, no improvement is visible in either the estimates of the means or the uncertainties, which matches the evolution of the RMS error (Fig. 13). While both K and ε quickly converge to the values used in the reference simulation (Tab. 4), the low uncertainty in K compared to ε reflects the physical significance of this parameter: The nonlinear dynamics of a premixed flame are highly sensitive to the timings of pinch-off events [77], which in turn depend on the phase speed at which perturbations travel along with the flame. In our reduced-order model based on the G -equation, the phase speed is regulated by the parameter K , which makes it highly observable even in the presence of observation noise. Similar to sensitivity analysis [78, 79, 80], this physical insight is gained from the inspection of the uncertainties, which here exceed the residuals between the means and the set of reference values by several orders of magnitude.



— Monte Carlo experiment — state estimation — combined state and parameter estimation

Figure 14: Logarithmic plot of bias over 50,000 timesteps for the Monte Carlo experiment (blue line) as well as the twin experiments with state estimation (orange line) and combined state and parameter estimation (green line). 6,283 timesteps correspond to one period of harmonic forcing. Data is assimilated every 1,000 timesteps. The black rectangle marks the time window depicted in Fig. 12.

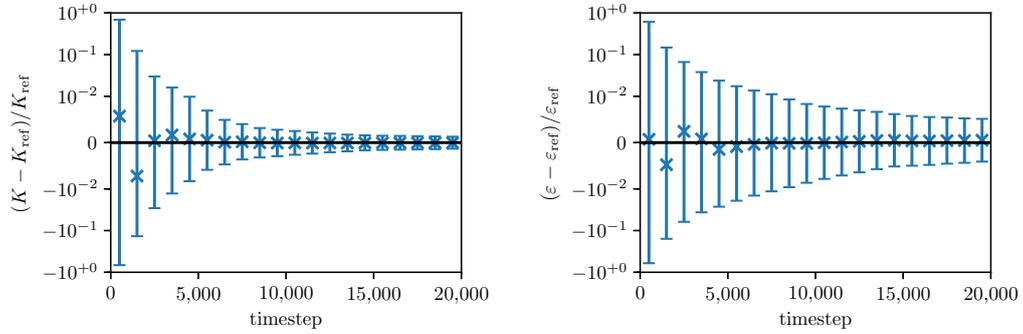


Figure 15: Logarithmic plots of normalized residuals of K (left) and ε (right) over 20,000 timesteps for combined state and parameter estimation; linear scales for normalized likelihoods between $\pm 10^{-2}$. The marginal distributions in K and ε are represented by their means (crosses) and three-sigma confidence levels (error bars).

445 Finally, combined state and parameter estimation using the ensemble Kalman smoother is
 performed (Theorem 6). In Fig. 15, the marginal distributions in the normalized residuals of K
 and ε are shown for both the ensemble Kalman filter and smoother. The effect of the ensemble
 Kalman filter has already been described in Fig. 15. In the forward-backward implementation,
 the ensemble Kalman smoother takes the last solution of the ensemble Kalman filter and works
 450 itself backwards in time (Theorem 2). In terms of information theory, the ensemble Kalman
 smoother takes at every timestep k all observations, from the past, present and future, into
 account compared to the ensemble Kalman filter, which only takes the past and the present
 into account. Statistically speaking, the smoothed distributions are more strictly conditioned
 than the filtered distributions, over $y_{1:N}$ compared to $y_{1:k}$ (Tab. 3). This surplus in information
 455 is evident in the form of lower error bars not just towards the end of the combined state and
 parameter estimation, but extending all the way to the very first instances of data assimilation.
 The evaluation of Eq. 6 does not rely on the solution of the governing equations, but instead relies
 on the predicted and filtered distributions in storage. The ensemble Kalman smoother is thus a
 computationally inexpensive tool to quantify the uncertainties especially at the beginning of the
 460 simulation, where little retrospective data is available. In analogy to direct-adjoint looping [81,
 82], combined state and parameter estimation based on filtering and smoothing can be used to
 obtain otherwise inevitably ad-hoc initial conditions and parameters.

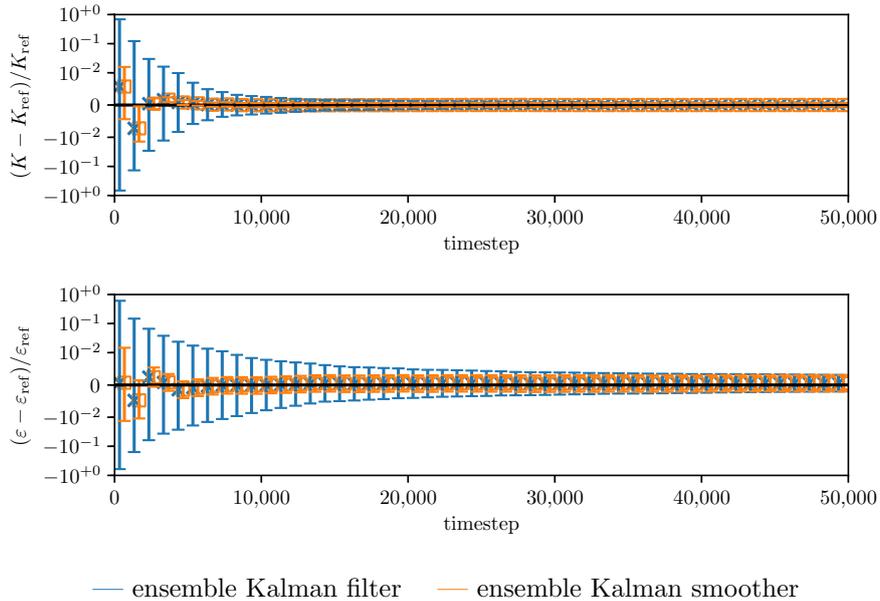


Figure 16: Logarithmic plots of normalized residuals of K (top) and ε (bottom) over 50,000 timesteps for combined state and parameter estimation using the ensemble Kalman filter (blue cross) and the ensemble Kalman smoother (orange square); linear scales for normalized likelihoods between $\pm 10^{-2}$.

4.3. Effect of observation noise

In a first parameter study, the observation error σ_ε is varied to investigate the effect of observation noise on the level-set data assimilation framework. In Section 4.2, it has been established that combined state and parameter estimation eventually leads to a statistically stationary state. In Fig. 17, the same RMS error is plotted over time for different observation errors σ_ε . This reveals that the sustained low uncertainties are epistemic in nature [83]: Every reduction in the observation error σ_ε by one order of magnitude reduces the RMS error by one order of magnitude. It is clear that the observation error poses a lower epistemic bound on how low the uncertainty can be reduced by combined state and parameter estimation.

In Fig. 18, the marginal distributions in the normalized residuals of K and ε are plotted over time for the different observation errors. The means of the residuals quickly vanish for all three observation errors σ_ε . For the lower observation errors, the error bars increasingly fail to contain the zero residual. Considering that three sigmas correspond to a confidence of 99.7% under the assumption of normal distributions (Section 2.3), the frequency at which the error bars fail leads to the conclusion that the level-set data assimilation framework underpredicts the uncertainties for low observation errors. This is in agreement with the theoretical analysis of the level-set data

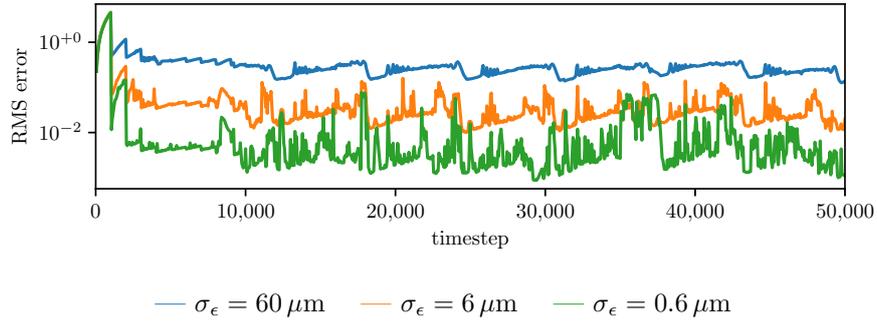


Figure 17: Logarithmic plot of root mean square (RMS) error over 50,000 timesteps for combined state and parameter estimation with different observation errors σ_ϵ ($\sigma_\epsilon = 60 \mu\text{m}$ blue line, $\sigma_\epsilon = 6 \mu\text{m}$ orange line, $\sigma_\epsilon = 0.6 \mu\text{m}$ green line).

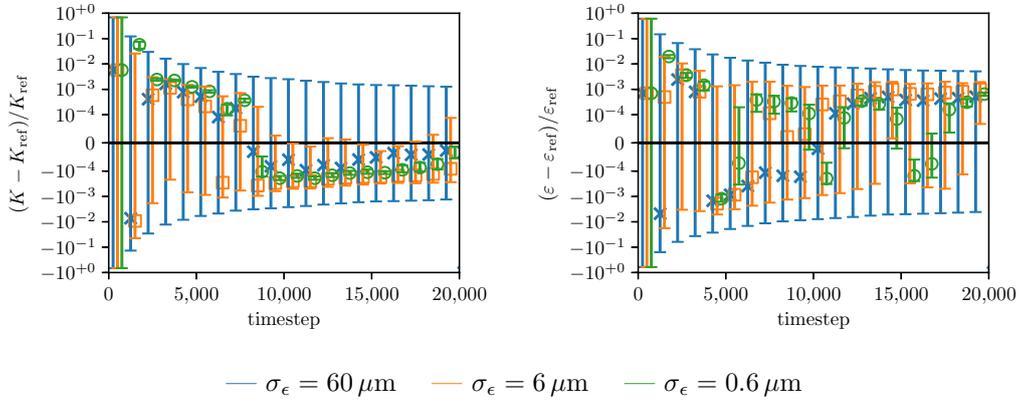


Figure 18: Logarithmic plots of normalized residuals of K (left) and ϵ (right) over 20,000 timesteps for combined state and parameter estimation with different observation errors ($\sigma_\epsilon = 60 \mu\text{m}$ blue cross, $\sigma_\epsilon = 6 \mu\text{m}$ orange square, $\sigma_\epsilon = 0.6 \mu\text{m}$ green circle); linear scales for normalized likelihoods between $\pm 10^{-4}$.

assimilation framework for the two-dimensional test case (Section 3.4): Although the individual
480 simulations predict the formation of sharp cusps as the perturbations travel on the respective
flame surfaces, the cusps are rounded in the mean of the ensemble (Fig. 6, left). Furthermore,
the observable sharpness of a cusp depends on the local number of observations of points on
the flame surface. A decrease in the observation error σ_ϵ without an increase in the number
of observations amounts to a relative decrease in the resolution of the observed cusp (Fig. 8).
485 Therefore, the local values of the eikonal fields of the individual simulations significantly deviate
from normal distributions (Fig. 6, right), which leads to incorrect distributions and uncertainties.

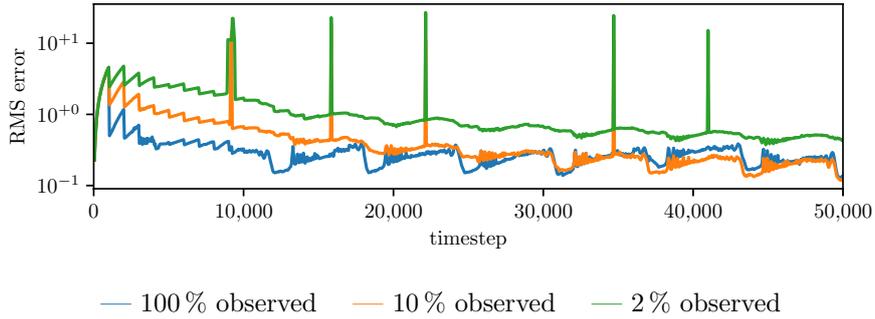


Figure 19: Logarithmic plot of root mean square (RMS) error over 50,000 timesteps for combined state and parameter estimation with different subsamples of observation points (100% blue line, 10% orange line, 2% green line).

4.4. Effect of number of observations

In a second parameter study, subsamples of the observation points are used to investigate the effect of the number of observations on the level-set data assimilation framework. Furthermore, this serves to illustrate the superiority of our analytic view over the geometric view in data
 490 assimilation (Section 3): In the geometric view, the observed interface must be parametric or at least as highly resolved as the grid used in the level-set method to allow the optimal interpolation of a sufficiently large number of points on the predicted interfaces. This restriction does not exist in the analytic view. Hence, the effect of dispersed observation points on the level-set data
 495 assimilation framework is also studied here.

In Fig. 19, the RMS error is plotted over time for different randomly sampled subsets of the observation points. In Section 4.2, it has been established that combined state and parameter estimation eventually leads to a statistically stationary state after approximately 10,000 timesteps. With 10% of the observation points, it takes more than 30,000 timesteps to reach a comparable
 500 low-uncertainty plateau. With 2% of the observation points, the RMS error is still decreasing after 50,000 timesteps while repeatedly failing to correctly predict the pinch-off events for the whole ensemble.

In Fig. 20, the marginal distributions in the normalized residuals of K and ε are plotted over time for the different subsamples. While the means of the residuals quickly vanish for all three
 505 subsamples, the confidence levels for the incomplete subsamples improve less rapidly than for the original sample of observation points. This is in agreement with the theoretical analysis of the level-set data assimilation framework for the two-dimensional test case (Section 3.4): In particular for the 2%-subsample, certain features in the observed interfaces remain underresolved at the

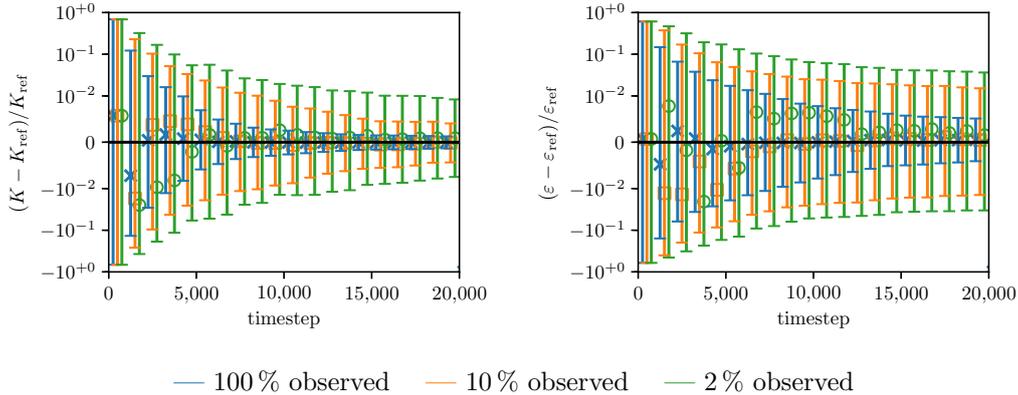


Figure 20: Logarithmic plots of normalized residuals of K (left) and ε (right) over 20,000 timesteps for combined state and parameter estimation with different subsamples of observation points (100 % blue cross, 10 % orange square, 2 % green circle); linear scales for normalized likelihoods between $\pm 10^{-2}$.

individual timesteps (Fig. 7). Nevertheless, the ensemble at a timestep k representing the filtered
510 probability distribution $p(x_k, \theta \mid y_{1:k}, f)$ (Tab. 3) reflects the knowledge of the observations
at all previous timesteps due to the Markov chain property of the probabilistic state space
model (Section 2.1). In theory, the lack of resolution in the observations is compensated by more
instances of data assimilation to accumulate the same amount of information. The result is a
similar statistically stationary state reached at a later time.

515 5. Conclusion

Reduced-order models based on level-set methods are widely used tools to qualitatively cap-
ture and track the nonlinear dynamics of an interface. In this paper, we enhance such a level-set
model with a statistical learning technique in order to make the model quantitatively predic-
tive. The statistical learning method is Bayesian, so the uncertainty of the outputs is naturally
520 included in this framework. The main ingredients of the level-set statistical learning are (i) a
reduced-order model of an interface based on a level-set method with a law of motion; (ii) the
ensemble Kalman filter and smoother; and (iii) external reference data with its uncertainty. The
ensemble Kalman filter uses Bayes' rule as its first principle while assuming normal distributions
525 when data is assimilated. The output of the algorithm is the combined state and parameters
estimation, which provides the most likely position of the interface and the model's parameters
given observations. These observations can originate from experimental data or high-fidelity
simulations.

To show and verify the algorithm, we perform twin experiments, where the observations are produced by the same code. The assimilation of data from experimental data or high-fidelity simulation requires virtually no modification to the proposed algorithm. The algorithm is applied to three test cases, which are of increasing complexity.

First, a one-dimensional case is presented. We choose the Hamilton-Jacobi generating functions over the whole domain to perform data assimilation, which is called the “analytic approach”. The one-dimensional example shows that the analytic approach overcomes the inaccuracy of the set-theoretic approach, which is based on characteristic functions. Moreover, it is argued that the analytic approach is more straightforward to implement than the geometric approach when moving to higher dimensions.

Second, a two-dimensional, time-independent case is presented, where the position of a sharp corner has to be predicted. It is shown that the level-set data assimilation framework proposed is fully front capturing. The effect that the number of observations has on capturing a sharp corner is investigated. The sensitivity of the analysis solution to each observation is calculated. The one- and two-dimensional cases show that the analytic approach we adopt is more robust than the geometric and set-theoretic approaches that have been used in the past.

Third, the time-dependent nonlinear dynamics of a conical forced premixed flame is studied. The most uncertain states are pinch-offs and the formation of sharp cusps, which are highly nonlinear topological changes of the interface. The two most uncertain parameters are the amplitude of the velocity perturbation at the flame’s base and the perturbation phase speed that wrinkles the flame. In the twin experiment, the combined state and parameter estimation fully recovers the reference solution, which validates the algorithm. Finally, the effect of uncertainty and number of observations is analysed. The uncertainty on the observation most influences the detection of the pinch-off events or the formation of sharp cusps. The data assimilation analytic approach is shown to be versatile because it can assimilate data any time it becomes available, avoiding the cumbersome parameterization required in a geometric approach.

The level-set data assimilation framework is applicable to a number of problems in computational physics involving the motion of an interface [1]. For example, a physical model coupling the dynamics of a premixed flame with an acoustic model of a duct may be used for a data-driven investigation of the Rijke tube [9, 10]. More generally, a number of computational fluid dynamics (CFD) models of premixed combustion, short of direct numerical simulations, combine the laws of motion with conservation laws for the inert flow, e.g. the Bray-Moss-Libby model or the flamelet model [8]. Although it is tempting to directly apply data assimilation to the primitive

variables in the Navier-Stokes equations, the discontinuities in density, temperature and species concentrations in light of the theoretical analysis of the set-theoretic approach to level-set data assimilation raise doubts as to whether the kinematics of a premixed flame, which are crucial to thermoacoustics, will be correctly predicted.

565 Future work will focus on extending the applications of the proposed level-set data assimilation framework. When applied to experiments, the effects of imperfect models will need to be mitigated [66], for example with techniques such as covariance inflation [84, 85] and localization [86]. Future work will also assess the uncertainties in the model, as opposed to uncertainties in the state and parameters.

570 Acknowledgments

The authors would like to thank the Data Assimilation Research Centre at the University of Reading for the useful discussions at their 2018 Data Assimilation Training. H. Yu is supported the Cambridge Commonwealth, European & International Trust under a Schlumberger Cambridge International Scholarship. L. Magri is supported by the Royal Academy of Engineering
575 under the RAEng Research Fellowships scheme and partially by the visiting fellowship at the Technical University of Munich – Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement n. 291763.

References

- 580 [1] J. Sethian, Evolution, Implementation, and Application of Level Set and Fast Marching Methods for Advancing Fronts, *Journal of Computational Physics* 169 (2001) 503–555.
- [2] F. Gibou, R. Fedkiw, S. Osher, A review of level-set methods and some recent applications, *Journal of Computational Physics* 353 (2018) 82–109.
- [3] S. Osher, J. A. Sethian, Fronts propagating with curvature-dependent speed: Algorithms
585 based on Hamilton-Jacobi formulations, *Journal of Computational Physics* 79 (1988) 12–49.
- [4] J. A. Sethian, P. Smereka, Level Set Methods for Fluid Interfaces, *Annual Review of Fluid Mechanics* 35 (2003) 341–372.
- [5] J. A. Sethian, Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science, number 3 in

- 590 Cambridge monographs on applied and computational mathematics, 2nd ed ed., Cambridge
University Press, Cambridge, U.K. ; New York, 1999.
- [6] S. Osher, R. P. Fedkiw, Level Set Methods: An Overview and Some Recent Results, *Journal
of Computational Physics* 169 (2001) 463–502.
- [7] W. Kollmann, *Fluid Mechanics in Spatial and Material Description*, University Readers,
595 2010.
- [8] N. Peters, *Turbulent Combustion*, Cambridge University Press, 2000. URL: [http://
ebooks.cambridge.org/ref/id/CB09780511612701](http://ebooks.cambridge.org/ref/id/CB09780511612701). doi:10.1017/CB09780511612701.
- [9] M. Fleifil, A. M. Annaswamy, Z. A. Ghoneim, A. F. Ghoniem, Response of a laminar pre-
mixed flame to flow oscillations: A kinematic model and thermoacoustic instability results,
600 *Combustion and Flame* 106 (1996) 487–510.
- [10] A. P. Dowling, A kinematic model of a ducted flame, *Journal of Fluid Mechanics* 394 (1999)
51–72.
- [11] K. Kashinath, I. C. Waugh, M. P. Juniper, Nonlinear self-excited thermoacoustic oscillations
of a ducted premixed flame: bifurcations and routes to chaos, *Journal of Fluid Mechanics*
605 761 (2014) 399–430.
- [12] I. C. Waugh, K. Kashinath, M. P. Juniper, Matrix-free continuation of limit cycles and their
bifurcations for a ducted premixed flame, *Journal of Fluid Mechanics* 759 (2014) 1–27.
- [13] H.-K. Zhao, T. Chan, B. Merriman, S. Osher, A Variational Level Set Approach to Multi-
phase Motion, *Journal of Computational Physics* 127 (1996) 179–195.
- 610 [14] S. J. Osher, F. Santosa, Level Set Methods for Optimization Problems Involving Geometry
and Constraints, *Journal of Computational Physics* 171 (2001) 272–288.
- [15] J. A. Sethian, A. Wiegmann, Structural Boundary Design via Level Set and Immersed
Interface Methods, *Journal of Computational Physics* 163 (2000) 489–528.
- [16] G. Allaire, F. Jouve, A.-M. Toader, Structural optimization using sensitivity analysis and
615 a level-set method, *Journal of Computational Physics* 194 (2004) 363–393.
- [17] F. Chantalat, C.-H. Bruneau, C. Galusinski, A. Iollo, Level-set, penalization and cartesian
meshes: A paradigm for inverse problems and optimal design, *Journal of Computational
Physics* 228 (2009) 6291–6315.

- [18] G. Evensen, *Data Assimilation*, Springer Berlin Heidelberg, 2009. URL: <http://link.springer.com/10.1007/978-3-642-03711-5>. doi:10.1007/978-3-642-03711-5.
- [19] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Dover Publications, 2007.
- [20] A. Gelb (Ed.), *Applied optimal estimation*, MIT Press, 1974.
- [21] R. F. Stengel, *Optimal control and estimation*, Dover Publications, 1994.
- [22] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research* 99 (1994) 10143.
- [23] G. Burgers, P. J. van Leeuwen, G. Evensen, Analysis scheme in the ensemble Kalman filter, *Monthly Weather Review* 126 (1998) 1719–1724.
- [24] A. Doucet, N. Freitas, N. Gordon (Eds.), *Sequential Monte Carlo Methods in Practice*, Springer New York, 2001. URL: <http://link.springer.com/10.1007/978-1-4757-3437-9>, doi: 10.1007/978-1-4757-3437-9.
- [25] R. N. Miller, M. Ghil, F. Gauthiez, Advanced data assimilation in strongly nonlinear dynamical systems, *Journal of the Atmospheric Sciences* 51 (1994) 1037–1056.
- [26] D. Rozier, F. Birol, E. Cosme, P. Brasseur, J. M. Brankart, J. Verron, A Reduced-Order Kalman Filter for Data Assimilation in Physical Oceanography, *SIAM Review* 49 (2007) 449–465.
- [27] C. H. Colburn, J. B. Cessna, T. R. Bewley, State estimation in wall-bounded flow systems. Part 3. The ensemble Kalman filter, *Journal of Fluid Mechanics* 682 (2011) 289–303.
- [28] H. Kato, A. Yoshizawa, G. Ueno, S. Obayashi, A data assimilation methodology for reconstructing turbulent flows around aircraft, *Journal of Computational Physics* 283 (2015) 559–581.
- [29] V. Mons, J.-C. Chassaing, T. Gomez, P. Sagaut, Reconstruction of unsteady viscous flows using data assimilation schemes, *Journal of Computational Physics* 316 (2016) 255–280.
- [30] H. Xiao, J.-L. Wu, J.-X. Wang, R. Sun, C. Roy, Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier-Stokes simulations: A data-driven, physics-informed Bayesian approach, *Journal of Computational Physics* 324 (2016) 115–136.

- [31] D. Darakananda, A. F. d. C. da Silva, T. Colonius, J. D. Eldredge, Data-assimilated low-order vortex modeling of separated flows, *Physical Review Fluids* 3 (2018).
- [32] J. W. Labahn, H. Wu, B. Coriton, J. H. Frank, M. Ihme, Data assimilation using high-speed measurements and LES to examine local extinction events in turbulent flames, *Proceedings of the Combustion Institute* (2018).
- [33] J. D. Beezley, J. Mandel, Morphing ensemble Kalman filters, *Tellus A: Dynamic Meteorology and Oceanography* 60 (2008) 131–140.
- [34] J. Mandel, L. S. Bennethum, J. D. Beezley, J. L. Coen, C. C. Douglas, M. Kim, A. Vodacek, A wildland fire model with data assimilation, *Mathematics and Computers in Simulation* 79 (2008) 584–606.
- [35] J. Mandel, J. D. Beezley, A. K. Kochanski, Coupled atmosphere-wildland fire modeling with WRF-Fire, *Geoscientific Model Development* 4 (2011) 591–610. ArXiv: 1102.1343.
- [36] M. Rochoux, A. Collin, C. Zhang, A. Trouvé, D. Lucor, P. Moireau, Front Shape Similarity Measure for Shape-Oriented Sensitivity Analysis and Data Assimilation for Eikonal Equation, *ESAIM: Proceedings and Surveys* 63 (2018) 258–279.
- [37] C. Zhang, A. Collin, P. Moireau, A. Trouvé, M. Rochoux, Front shape similarity measure for data-driven simulations of wildland fire spread based on state estimation: Application to the RxCADRE field-scale experiment, *Proceedings of the Combustion Institute* 37 (2019) 4201–4209.
- [38] L. Li, F.-X. Le Dimet, J. Ma, A. Vidard, A Level-Set-Based Image Assimilation Method: Potential Applications for Predicting the Movement of Oil Spills, *IEEE Transactions on Geoscience and Remote Sensing* 55 (2017) 6330–6343.
- [39] L. Li, A. Vidard, F.-X. Le Dimet, J. Ma, Topological data assimilation using Wasserstein distance, *Inverse Problems* 35 (2019) 015006.
- [40] D. Peng, B. Merriman, S. Osher, H. Zhao, M. Kang, A PDE-Based Fast Local Level Set Method, *Journal of Computational Physics* 155 (1999) 410–438.
- [41] J. A. Sethian, A fast marching level set method for monotonically advancing fronts, *Proceedings of the National Academy of Sciences of the United States of America* 93 (1996) 1591–1595.

- [42] D. L. Moreno, S. I. Aanonsen, Stochastic Facies Modelling Using the Level Set Method, in: Petroleum Geostatistics, Cascais, Portugal, 2007. URL: <http://www.earthdoc.org/publication/publicationdetails/?publication=7826>. doi:10.3997/2214-4609.201403056.
- 680 [43] M. C. Rochoux, B. Cuenot, S. Ricci, A. Trouvé, B. Delmotte, S. Massart, R. Paoli, R. Paugam, Data assimilation applied to combustion, *Comptes Rendus Mécanique* 341 (2013) 266–276.
- [44] X. Gao, Y. Wang, N. Overton, M. Zupanski, X. Tu, Data-assimilated computational fluid dynamics modeling of convection-diffusion-reaction problems, *Journal of Computational*
685 *Science* 21 (2017) 38–59.
- [45] S. P. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [46] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, 2003. URL: <http://ebooks.cambridge.org/ref/id/CB09780511790423>. doi:10.1017/CB09780511790423.
- 690 [47] S. Sarkka, *Bayesian Filtering and Smoothing*, Cambridge University Press, Cambridge, 2013. URL: <http://ebooks.cambridge.org/ref/id/CB09781139344203>. doi:10.1017/CB09781139344203.
- [48] R. Bellman, *Dynamic programming*, Dover Publications, 2003. OCLC: 834140800.
- [49] D. J. Higham, An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations, *SIAM Review* 43 (2001) 525–546.
695
- [50] R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering* 82 (1960) 35.
- [51] R. E. Kalman, R. S. Bucy, New Results in Linear Filtering and Prediction Theory, *Journal of Basic Engineering* 83 (1961) 95.
- 700 [52] J. S. Whitaker, T. M. Hamill, Ensemble Data Assimilation without Perturbed Observations, *Monthly Weather Review* 130 (2002) 1913–1924.
- [53] M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, J. S. Whitaker, Ensemble Square Root Filters, *Monthly Weather Review* 131 (2003) 1485–1490.

- [54] D. M. Livings, S. L. Dance, N. K. Nichols, Unbiased ensemble square root filters, *Physica D: Nonlinear Phenomena* 237 (2008) 1021–1028.
- [55] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, volume 60 of *Graduate Texts in Mathematics*, second ed., Springer New York, 1989. URL: <http://link.springer.com/10.1007/978-1-4757-2063-1>. doi:10.1007/978-1-4757-2063-1.
- [56] J. A. Sethian, *Fast Marching Methods*, *SIAM Review* 41 (1999) 199–235.
- [57] D. Adalsteinsson, J. A. Sethian, A Fast Level Set Method for Propagating Interfaces, *Journal of Computational Physics* 118 (1995) 269–277.
- [58] M. Sussman, P. Smereka, S. Osher, A Level Set Approach for Computing Solutions to Incompressible Two-Phase Flow, *Journal of Computational Physics* 114 (1994) 146–159.
- [59] U. M. Ascher, R. M. M. Mattheij, R. D. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*, Society for Industrial and Applied Mathematics, 1995. URL: <http://epubs.siam.org/doi/book/10.1137/1.9781611971231>.
- [60] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.
- [61] G. Szego, *Orthogonal Polynomials*, volume 23 of *Colloquium Publications*, American Mathematical Society, Providence, Rhode Island, 1939. URL: <http://www.ams.org/coll/023>. doi:10.1090/coll/023.
- [62] T. C. Lieuwen, V. Yang (Eds.), *Combustion Instabilities in Gas Turbine Engines: Operational Experience, Fundamental Mechanisms and Modeling*, Progress in astronautics and aeronautics, American Institute of Aeronautics and Astronautics, 2005. OCLC: ocm62221288.
- [63] F. E. C. Culick, *Unsteady motions in combustion chambers for propulsion systems*, Technical Report, AGARD, 2006.
- [64] A. Orchini, G. Rigas, M. P. Juniper, Weakly nonlinear analysis of thermoacoustic bifurcations in the Rijke tube, *Journal of Fluid Mechanics* 805 (2016) 523–550.
- [65] B. Semlitsch, A. Orchini, A. P. Dowling, M. P. Juniper, G-equation modelling of thermoacoustic oscillations of partially premixed flames, *International Journal of Spray and Combustion Dynamics* 9 (2017) 260–276.

- [66] S. Reich, C. Cotter, Probabilistic Forecasting and Bayesian Data Assimilation, Cambridge University Press, Cambridge, 2015. URL: <http://ebooks.cambridge.org/ref/id/CB09781107706804>. doi:10.1017/CB09781107706804.
- 735
- [67] H. Yu, T. Jaravel, M. Ihme, M. P. Juniper, L. Magri, Physics-informed data-driven prediction of premixed flame dynamics with data assimilation, Technical Report, Center for Turbulence Research, 2018. URL: <https://ctr.stanford.edu/proceedings-2018-summer-program>.
- [68] H. Yu, T. Jaravel, M. Juniper, M. Ihme, L. Magri, Data assimilation and optimal calibration in nonlinear models of flame dynamics, Journal of Engineering for Gas Turbines and Power, doi:10.17863/CAM.41434 (2019).
- 740
- [69] K. Kashinath, S. Hemchandra, M. P. Juniper, Nonlinear thermoacoustics of ducted premixed flames: The influence of perturbation convection speed, Combustion and Flame 160 (2013) 2856–2865.
- 745
- [70] X.-D. Liu, S. Osher, T. Chan, Weighted Essentially Non-oscillatory Schemes, Journal of Computational Physics 115 (1994) 200–212.
- [71] G.-S. Jiang, C.-W. Shu, Efficient Implementation of Weighted ENO Schemes, Journal of Computational Physics 126 (1996) 202–228.
- [72] C.-W. Shu, S. Osher, Efficient implementation of essentially non-oscillatory shock-capturing schemes, Journal of Computational Physics 77 (1988) 439–471.
- 750
- [73] I. Waugh, Methods for Analysis of Nonlinear Thermoacoustic Systems, Ph.D. thesis, University of Cambridge, 2013. URL: <https://doi.org/10.17863/CAM.36094>.
- [74] Preetham, H. Santosh, T. Lieuwen, Dynamics of Laminar Premixed Flames Forced by Harmonic Velocity Disturbances, Journal of Propulsion and Power 24 (2008) 1390–1402.
- 755
- [75] A. Orchini, M. P. Juniper, Linear stability and adjoint sensitivity analysis of thermoacoustic networks with premixed flames, Combustion and Flame 165 (2016) 97–108.
- [76] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning, MIT Press, 2006. OCLC: ocm61285753.
- [77] M. P. Juniper, R. Sujith, Sensitivity and Nonlinearity of Thermoacoustic Oscillations, Annual Review of Fluid Mechanics 50 (2018) 661–689.
- 760

- [78] L. Magri, M. Bauerheim, M. P. Juniper, Stability analysis of thermo-acoustic nonlinear eigenproblems in annular combustors. Part I. Sensitivity, *Journal of Computational Physics* 325 (2016) 395–410.
- 765 [79] L. Magri, M. Bauerheim, F. Nicoud, M. P. Juniper, Stability analysis of thermo-acoustic nonlinear eigenproblems in annular combustors. Part II. Uncertainty quantification, *Journal of Computational Physics* 325 (2016) 411–421.
- [80] L. Magri, Adjoint Methods as Design Tools in Thermoacoustics, *Applied Mechanics Reviews* 71 (2019) 020801.
- 770 [81] M. P. Juniper, Triggering in the horizontal Rijke tube: non-normality, transient growth and bypass transition, *Journal of Fluid Mechanics* 667 (2011) 272–308.
- [82] R. Kerswell, Nonlinear Nonmodal Stability Theory, *Annual Review of Fluid Mechanics* 50 (2018) 319–345.
- [83] A. D. Kiureghian, O. Ditlevsen, Aleatory or epistemic? Does it matter?, *Structural Safety* 31 (2009) 105–112.
- 775 [84] J. L. Anderson, S. L. Anderson, A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts, *Monthly Weather Review* 127 (1999) 2741–2758.
- [85] E. Ott, B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. Patil, J. A. Yorke, A local ensemble Kalman filter for atmospheric data assimilation, *Tellus A: Dynamic Meteorology and Oceanography* 56 (2004) 415–428.
- 780 [86] G. Gaspari, S. E. Cohn, Construction of correlation functions in two and three dimensions, *Quarterly Journal of the Royal Meteorological Society* 125 (1999) 723–757.
- [87] O. Forster, *Analysis 1*, Springer Fachmedien Wiesbaden, 2013. URL: <http://link.springer.com/10.1007/978-3-658-00317-3>. doi:10.1007/978-3-658-00317-3.
- 785 [88] G. Fischer, *Lineare Algebra*, Springer Fachmedien Wiesbaden, 2014. URL: <http://link.springer.com/10.1007/978-3-658-03945-5>. doi:10.1007/978-3-658-03945-5.

Appendix A. Bayesian Inference

Theorem 7 (Marginalization). *For a joint probability distribution $p(X, Y)$, the marginal probability distributions $p(X)$ and $p(Y)$ are given by*

$$p(X) = \int p(X, Y) \, dY \quad , \quad (\text{A.1})$$

$$p(Y) = \int p(X, Y) \, dX \quad . \quad (\text{A.2})$$

Theorem 8 (Bayes' rule). *The conditional probability distributions $p(X | Y)$ and $p(Y | X)$ are related to the joint probability distribution via*

$$p(X, Y) = p(X)p(Y | X) = p(Y)p(X | Y) \quad . \quad (\text{A.3})$$

The conditional probability distribution $p(X | Y)$ is sometimes referred to as the inverse of the conditional probability distribution $p(Y | X)$. The inverse is given by Bayes' rule:

$$p(X | Y) = \frac{p(X, Y)}{p(Y)} = \frac{p(X)p(Y | X)}{p(Y)} \quad . \quad (\text{A.4})$$

If $p(X)$ and $p(Y | X)$ are given, $p(Y)$ can be computed through marginalization.

790 Appendix B. Hamiltonian mechanics

Lemma 1 (Lagrangian).

$$\mathcal{L}(r(t), \dot{r}(t), t) = 0 \quad . \quad (\text{B.1})$$

Proof. The generating function is formally the same as the action integral [55, Chapter 9]. Thus, the Lagrangian $\mathcal{L}(r, \dot{r}, t)$ is the total derivative of the generating function:

$$\mathcal{L}(r(t), \dot{r}(t), t) = \frac{d}{dt}G(r(t), t) = \frac{d}{dt}G(r(0), 0) = 0 \quad . \quad (\text{B.2})$$

□

Lemma 2 (Hamiltonian).

$$\mathcal{H}(r(t), n(t), t) = u(r(t)) \cdot n(t) \quad . \quad (\text{B.3})$$

Proof. A Legendre transform of the Lagrangian $\mathcal{L}(r, \dot{r}, t)$ (Eq. (B.1)) gives the Hamiltonian $\mathcal{H}(r, n, t)$:

$$\mathcal{H}(r(t), n(t), t) = \dot{r}(t) \cdot n(t) - \mathcal{L}(r(t), \dot{r}(t), t) = u(r(t)) \cdot n(t) \quad . \quad (\text{B.4})$$

It remains to be shown that $\mathcal{H}(r, n, t)$ is indeed a Hamiltonian. The constraints, Eq. (35) and (36), are only functions in r and n respectively. Hence, a suitable choice of Lagrange multipliers λ_r and λ_n gives

$$\frac{\partial \mathcal{H}}{\partial n} = u + \lambda_n n = u = \frac{dr}{dt} \quad , \quad (\text{B.5})$$

$$\frac{\partial \mathcal{H}}{\partial r} = n \cdot \nabla u + n \times (\nabla \times u) + \lambda_r \frac{\partial G}{\partial r} = n \cdot \nabla u + n \times (\nabla \times u) - [n \cdot (n \cdot \nabla u)] n = -\frac{dn}{dt} \quad . \quad (\text{B.6})$$

Eq. (B.5) follows from the laws of motion (Eq. (33)). Eq. (B.6) follows from tensor calculus:

$$\frac{dn}{dt} = \frac{\partial n}{\partial t} + u \cdot \nabla n + \mu_n n \quad , \quad (\text{B.7})$$

where μ_n is the Lagrange multiplier for the constraint imposed by Eq. (36). The gradient of the total derivative of the generating function gives an expression for the partial derivative of the normal vector (Lemma 1):

$$\nabla \left(\frac{dG}{dt} \right) = \nabla \left(\frac{\partial G}{\partial t} + u \cdot \nabla G \right) = \frac{\partial n}{\partial t} + \nabla (u \cdot n) = 0 \quad . \quad (\text{B.8})$$

The dot product of the normal vector with the total derivative of the normal vector gives an expression for the Lagrange multiplier μ_n :

$$n \cdot \frac{dn}{dt} = \frac{1}{2} \frac{d}{dt} (n \cdot n) = 0 \quad , \quad (\text{B.9})$$

$$n \cdot \frac{dn}{dt} = n \cdot \left(\frac{\partial n}{\partial t} + u \cdot \nabla n + \mu_n n \right) \quad (\text{B.10})$$

$$= n \cdot \frac{\partial n}{\partial t} + \frac{1}{2} (u \cdot \nabla) (n \cdot n) + \mu_n n \cdot n \quad (\text{B.11})$$

$$= -n \cdot \nabla (u \cdot n) + \mu_n \quad . \quad (\text{B.12})$$

This gives:

$$\frac{dn}{dt} = -\nabla (u \cdot n) + u \cdot \nabla n + [n \cdot \nabla (u \cdot n)] n \quad (\text{B.13})$$

$$= -u \cdot \nabla n - n \cdot \nabla u - u \times (\nabla \times n) - n \times (\nabla \times u) + u \cdot \nabla n + [(n \cdot \nabla u) \cdot n] n + [(n \cdot \nabla n) \cdot u] n \quad (\text{B.14})$$

$$= -n \cdot \nabla u - n \times (\nabla \times u) + n [n \cdot (n \cdot \nabla u)] \quad . \quad (\text{B.15})$$

Note that $\nabla \times n = 0$ because $n = \nabla G$. This concludes the proof that $\mathcal{H}(r, n, t)$ is a Hamiltonian. \square

Theorem 9 (Hamilton-Jacobi equation).

$$\frac{\partial G}{\partial t} + u(r(t)) \cdot n(t) = 0 \quad . \quad (\text{B.16})$$

Proof. The Hamilton-Jacobi equation is given by [55, Chapter 9]

$$\frac{\partial G}{\partial t} + \mathcal{H} = \frac{\partial G}{\partial t} + u \cdot n = 0 \quad . \quad (\text{B.17})$$

□

795 **Appendix C. Level-set methods**

Lemma 3. *The covariance matrix of a gradient in a filtered generated function without reinitialization is given by*

$$\mathbb{E} \left[D\psi^a (D\psi^a)^T \right] = \mathbb{E} \left[D\psi^f (D\psi^f)^T \right] - D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \quad , \quad (\text{C.1})$$

where D denotes the gradient operator.

Proof. Eq. (23) gives

$$D\psi^a = D\psi^f + D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} (y_k - M\psi^f) \quad , \quad (\text{C.2})$$

$$\begin{aligned} D\psi^a (D\psi^a)^T &= D\psi^f (D\psi^f)^T \\ &+ D\psi^f (y_k - M\psi^f)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \\ &+ D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} (y_k - M\psi^f) (D\psi^f)^T \\ &+ D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} (y_k - M\psi^f) (y_k - M\psi^f)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \end{aligned} \quad (\text{C.3})$$

$$\begin{aligned} &= D\psi^f (D\psi^f)^T \\ &+ D\psi^f (y_k - Mx_k)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \\ &+ D\psi^f (x_k - \psi^f)^T M^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \\ &+ D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} (y_k - Mx_k) (D\psi^f)^T \\ &+ D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} M (x_k - \psi^f) (D\psi^f)^T \\ &+ D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} (y_k - Mx_k) (y_k - Mx_k)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \\ &+ D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} M (x_k - \psi^f) (x_k - \psi^f)^T M^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \\ &+ D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} (y_k - Mx_k) (x_k - \psi^f)^T M^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \\ &+ D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} M (x_k - \psi^f) (y_k - Mx_k)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \quad . \end{aligned} \quad (\text{C.4})$$

The probabilistic state space model (Eq. (1), (2)) gives

$$\mathbb{E} \left[(y_k - Mx_k) (y_k - Mx_k)^T \right] = C_{\epsilon\epsilon} \quad , \quad \mathbb{E} \left[(\psi^f - x_k) (\psi^f - x_k)^T \right] = C_{\psi\psi}^f \quad ,$$

$$\mathbb{E} \left[(\psi^f - x_k) (y_k - Mx_k)^T \right] = 0 \quad . \quad (\text{C.5})$$

Substituting Eq. C.5 into Eq. C.4 gives

$$\begin{aligned}
\mathbb{E} \left[D\psi^a (D\psi^a)^T \right] &= \mathbb{E} \left[D\psi^f (D\psi^f)^T \right] \\
&\quad - D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \\
&\quad - DC_{\psi\psi}^f M^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \\
&\quad + D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} C_{\epsilon\epsilon} \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \\
&\quad + D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f M^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T
\end{aligned} \tag{C.6}$$

$$\begin{aligned}
&= \mathbb{E} \left[D\psi^f (D\psi^f)^T \right] - D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \quad .
\end{aligned} \tag{C.7}$$

□

Lemma 4. *The expected value of a slope in a filtered generated function without reinitialization satisfies*

$$\mathbb{E} \left[|D\psi^a|^2 \right] = 1 - \text{tr} \left(D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \right) \quad . \tag{C.8}$$

Proof. In general, the expected value of a slope is related to the covariance matrix of the corresponding gradient in the following way:

$$\mathbb{E} \left[|D\psi|^2 \right] = \mathbb{E} \left[\text{tr} \left(D\psi (D\psi)^T \right) \right] = \text{tr} \left(\mathbb{E} \left[D\psi (D\psi)^T \right] \right) \quad . \tag{C.9}$$

Substituting Eq. (C.7) into Eq. (C.9) gives

$$\mathbb{E} \left[|D\psi^a|^2 \right] = \text{tr} \left(\mathbb{E} \left[D\psi^a (D\psi^a)^T \right] \right) \tag{C.10}$$

$$\begin{aligned}
&= \text{tr} \left(\mathbb{E} \left[D\psi^f (D\psi^f)^T \right] \right) - \text{tr} \left(D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \right)
\end{aligned} \tag{C.11}$$

$$\begin{aligned}
&= \mathbb{E} \left[|D\psi^f|^2 \right] - \text{tr} \left(D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \right) \quad .
\end{aligned} \tag{C.12}$$

In the first term, the slope $|D\psi^f|$ is equal to 1 due to Huygens' principle. This gives Lemma 4. □

Theorem 10.

$$\mathbb{E} \left[|D\psi^a| \right] < 1 \quad . \tag{C.13}$$

Proof. Applying the following two observations to Eq. (C.8) gives Theorem 10:

1. $\mathbb{E} [|D\psi^a|]^2 \leq \mathbb{E} [|D\psi^a|^2]$ due to Jensen's inequality [87].
2. $\text{tr} \left(D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T \right) \geq 0$ because $D \left(MC_{\psi\psi}^f \right)^T \left[C_{\epsilon\epsilon} + MC_{\psi\psi}^f M^T \right]^{-1} MC_{\psi\psi}^f D^T$ is symmetric and positive semi-definite [88].

805

□