

## Research



**Cite this article:** Ahmed SE, San O, Lakshmivarahan S, Lewis JM. 2023 On the dual advantage of placing observations through forward sensitivity analysis. *Proc. R. Soc. A* **479**: 20220815.  
<https://doi.org/10.1098/rspa.2022.0815>

Received: 2 December 2022

Accepted: 15 May 2023

**Subject Areas:**

computational physics, applied mathematics, computational mathematics

**Keywords:**

optimal sensor placement, forward sensitivity analysis, dynamic data assimilation, inverse problems, estimation, control

**Author for correspondence:**

O. San

e-mail: [osan@okstate.edu](mailto:osan@okstate.edu)

# On the dual advantage of placing observations through forward sensitivity analysis

S. E. Ahmed<sup>1,2</sup>, O. San<sup>2</sup>, S. Lakshmivarahan<sup>3</sup>  
 and J. M. Lewis<sup>4,5</sup>

<sup>1</sup>Advanced Computing, Mathematics and Data Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA

<sup>2</sup>School of Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, OK 74078, USA

<sup>3</sup>School of Computer Science, University of Oklahoma, Norman, OK 73019, USA

<sup>4</sup>National Severe Storms Laboratory, Norman, OK 73072, USA

<sup>5</sup>Desert Research Institute, Reno, NV 89512, USA

OS, 0000-0002-2241-4648

The four-dimensional variational data assimilation methodology for assimilating noisy observations into a deterministic model has been the workhorse in forecasting centres for over three decades. While this method provides a computationally efficient framework for dynamic data assimilation, it is largely silent on the important question concerning the minimum number and placement of observations. To answer this question, we demonstrate the dual advantage of placing the observations where the square of the sensitivity of the model solution with respect to the unknown control variables, called forward sensitivities, attains its maximum. By following this approach, we can force the observability Gramian to be of full rank, which in turn guarantees efficient recovery of the optimal values of the control variables, which is the first of the two advantages of this strategy. We further show that the proposed strategy of placing observations has another inherent optimality: the square of the sensitivity of the optimal estimates of the control with respect to the observations (used to obtain these estimates) attains its minimum value,

a second advantage that is a direct consequence of the above strategy for placing observations. Our analytical framework and numerical experiments on linear and nonlinear systems confirm the effectiveness of our proposed strategy.

## 1. Introduction

Estimating the unknown parameters and initial/boundary conditions (collectively known as the control) of a dynamic model based on a finite set of noisy observations constitutes an important class of inverse problems of interest in a variety of disciplines [1–13]. At its core, this class of inverse problems gives rise to at least four question levels: (i) observability, (ii) algorithmic path to estimate the model control, (iii) observation count and placement of observation, and (iv) assessing the quality of the estimates. At the first level, the observability question may be stated as follows: Given the model dynamics and the observation operator, which relates the state of the dynamics to the observable, under what condition is it possible to estimate the unknowns based on a finite set of noisy observation? In a fundamental paper, Kalman [14] first provided a binary yes or no answer to this question based on the rank of the observability matrix derived from the linear model dynamics and the linear observation operator. Since then, this basic rank condition has been extended in several directions, including local and global observability of nonlinear systems [15,16]. Furthermore, building upon the traditions in stability theory [17], researchers have proposed the concept of the degree of observability and scalar measures to quantify this degree [18–21].

At the second level, given that a system is observable, the question is: What is an efficient algorithmic pathway to compute the estimates of the unknowns? Within the context of geosciences, a variational framework known as the four-dimensional variational data assimilation (four-dimensional VAR) provides an answer in two steps. In the first step, a cost functional is defined by the weighted sum of squared differences between the actual and model counterpart of the observation at the time of the observation. The gradient of this cost functional (also known as the adjoint gradient) is computed using a forward run of the model and a backward sweep of its adjoint model. This process is referred to as the adjoint method (refer to Chapters 22–25 in Lewis *et al.* [4] for details). In the second step, this gradient is used in a minimization algorithm (Chapters 9–12, Lewis *et al.* [4]) to find an update of the control. These two steps are repeated until the desired convergence is reached.

The four-dimensional VAR-based approach has been the workhorse of weather forecasting centres around the world since its introduction [22]. Despite its grand success in delivering forecast products for public consumption for decades, the four-dimensional VAR framework works with a given set of observations and is largely silent on the third level of inquiry, namely, what is the minimum number of observations required and their placement in the spatio-temporal domain to maximize the effectiveness of computing the estimates? At this juncture, it is useful to review the impact of the number of observations in general. Thanks to advances in the sensor, wireless communication, mass storage and powerful computing technologies, we are steadily moving away from data-sparse to data-rich regimes. The ability to sample spatio-temporal fields at very high frequency (resulting from ever decreasing sampling intervals) has resulted in truly large datasets that are several orders of magnitude greater than what was available a decade ago. This growth has resulted in two side effects. First, the data exhibit very high correlations, which implies that more data does not translate into more information. Second, it is computationally demanding to ingest all the available data into an assimilation algorithm. These latter considerations highlight the importance of identifying smaller and independent subsets of observations to be used in estimation, which has resulted in a growing body of literature on ‘thinning’ and creating of ‘super observations’ [23].

In Lakshmivarahan *et al.* [24,25], a strategy for answering this third level of question within the four-dimensional VAR framework was provided for the first time. An analysis based

on the forward sensitivity method (FSM) has shown that the observability Gramian,  $G$ , is a good approximation to the Hessian of the cost function (it is exact for the linear case). In addition, the observability Gramian admits an additive decomposition:  $G = \sum_{i=1}^N G_i$ , where  $G_i$  is the contribution to the overall Gramian at the  $i$ th observation time and  $N$  is the number of observations. While each of these components is symmetric and positive semi-definite, it is shown in [24,25] that we can indeed force the Gramian  $G$  to be positive definite by placing the  $N = (n + p)$  observations where the squares of the forward sensitivities of the model solution (with respect to  $n$  components of the initial condition and  $p$  parameters including the boundary conditions) attain their maxima. This strategy avoids flat patches in the cost functional by bounding the norm of the adjoint gradient away from zero. This is the first of the two advantages referred to in the title of this paper. The papers by Lewis *et al.* [26,27] and Ahmed *et al.* [28,29] contain several applications of this strategy.

At the fourth question level, the emphasis shifts to analysing the quality of the resulting estimates of the control variables. There are two primary ways to approach this question. The first is to theoretically quantify the asymptotic distribution of the estimates by letting the number of observations increase without bounds. Within the context of *time-series analysis* dealing with linear, discrete time and stochastic dynamic models of the *auto regressive integrated moving average* types, there is vast literature relating to the asymptotic analysis of the estimates of the parameters of these models [30,31]. Likewise, there is a large body of results relating to static nonlinear models [32,33]. But to our knowledge, there is no such theory of estimates for large-scale nonlinear dynamic models of interest in the geosciences. In the absence of such a statistical theory, we settle for the next best option—an analytical approach to quantify the sensitivity of the estimates with respect to the observations. This quantity is particularly important when the given dynamic model exhibits high sensitivity to errors in the control. In this paper, we prove that the sensitivity of the estimates with respect to the observations attains its minimum value when we place the observations where the squares of the forward sensitivities are maximized, as in [24,25]. The positive definite Gramian resulting from the proposed strategy for observation placement determines the control estimates that exhibit the smallest possible sensitivity to observations. This is the second advantage of setting the observation placement strategy using the forward sensitivity analysis.

## (a) Historical remarks

The four levels of inquiry described above have natural connections to many different areas of the data assimilation and parameter estimation in dynamical systems literature spanning several decades. First is the fundamental structural identifiability question examined by Bellman & Åström [34], where it was shown that certain types of model parametrization do not admit a unique solution to the parameter estimation problem. Furthermore, there is extensive literature on lumped parameter system identification [35] and adaptive control [36] dealing with parameter estimation. A recent paper by Villaverde [37] contains an extensive analysis of observability and identification of nonlinear systems of interest in mathematical biology.

Within the parlance of distributed parameter systems, especially in the context of process control in chemical engineering, there is an extensive body of results relating to parameter estimation, placement of observations and analysis of the sensitivity of parameter estimates with respect to observations. We refer to the monograph by Ucinski [38] and papers by Alaña [39] and Christopher & Fathalla [40] for more details. The results in the present paper share some connections with the developments in Christopher & Fathalla [40] and Alaña [39]. Nonetheless, these two papers deal only with parameter estimation for distributed parameter systems, while the present study addresses the mutual estimation of initial conditions and parameters (denoted as control) in the same footing for systems described by ordinary/partial differential equations (ODEs/PDEs). For general treatment of sensitivity-based methods, we refer to the foundational works by Cacuci and co-workers [41,42]. Some of the ideas developed in the current study can be also connected to the vast and growing literature on adjoint-based analysis of

targeted observations and adaptive observations [43–48] as well as ensemble-based approaches to observation impact analysis [49,50].

## (b) Significance

This paper addresses, for the first time, key questions related to the simultaneous estimate of unknown initial conditions and model parameters in variational data assimilation frameworks. The main questions that we answer correspond to observation count and optimal placement as well as the subsequent quality of the inverse problem solution. We present a mathematically concrete strategy for selecting the observation locations by tracking the model forward sensitivity metrics, namely the forecast sensitivity with respect to the initial condition and to the model parameters. Because these sensitivities can generally be positive or negative, we opt for placing the observations at points where the squares of the forward sensitivities attain their maximum. We support the proposed observation placement strategy with a clear theoretical analysis to understand the consequences of this strategy. In particular, we demonstrate both theoretically and empirically that the advocated observation placement methodology has two main advantages as follows.

First, by placing the observations where the forward sensitivities are maximized, we enforce the observability Gramian matrix to be positive definite. We illustrate that this Gramian is closely related to the adjoint gradient of the cost function. Specifically, a positive definite observability Gramian leads to non-zero values of the adjoint gradient norm, which accelerates the optimization algorithm that is used to estimate the unknown initial conditions and model parameters. We highlight that the choice of observation locations controls the shape of the resulting cost function (defined using a weighted sum of the squared differences between the collected measurements and the model counterpart of the observation). Placing the observations using the presented forward sensitivity analysis avoids flat regions of the cost function, which would negatively affect the algorithm convergence.

Second, we demonstrate that the optimal estimates of the initial conditions and model parameters are sensitive to the locations where measurements are collected. More specifically, we prove that the sensitivities of those estimates (with respect to the observations themselves) reach their minimum values when the forecast sensitivities (with respect to the control) reach their maximum values. In other words, by placing the observations where the squares of forecast sensitivities with respect to the unknown control are maximized, we guarantee that the solution of the inverse problem is most robust to small perturbations in observation values. This second result has significant merit in practice, considering the inevitably noisy sensor data with different levels of uncertainties. We back up the analytical framework with numerical experiments using representative systems exhibiting linear and nonlinear dynamics with unknown initial conditions and model parameters.

## (c) Organization of the paper

In §2, we provide a succinct version of the statement of the inverse problem of interest in this paper. In §3, we introduce an optimal observation placement strategy using the forward sensitivity analysis that links the observability Gramian and the adjoint gradient. We lay out the analytical framework for defining the dual advantages of the proposed observation placement strategy in §§3 and 4. In the latter, we provide a theoretical analysis for evaluating the sensitivity of the control estimates with respect to the observation. In §5, we illustrate the theory of observation placement through the forward sensitivity analysis using examples corresponding to scalar linear and nonlinear dynamical systems with two elements of control—the initial condition and model parameterization—as well as two systems governed by PDEs in one and two spatial dimensions. We draw concluding remarks about the derived optimality conditions for observation placement and the associated twin advantages in §6.

## 2. Mathematical preliminaries

We start with a description of the model. Our primary goal is to bring out the beauty and elegance of the underlying ideas and results using simple, easily understandable models. Therefore, we first derive the analytical framework and perform the theoretical analysis using a scalar (one-dimensional) dynamical system with a single parameter. After that, we provide extensions and insights regarding the higher dimensional cases.

### (a) Model

Let  $\mathcal{I} \subset \mathbb{R}$  and  $\mathcal{P} \subset \mathbb{R}$  be two subsets of the real line representing the set of all allowed values of the initial condition  $x(0)$  and parameter  $\alpha$  of a dynamic model, respectively. The state,  $x(t) \in \mathbb{R}$ , evolves according to the following nonlinear, time-invariant, ODE:

$$\dot{x} = f(x, \alpha), \quad (2.1)$$

where  $f: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,  $x_0 = x(0) \in \mathcal{I}$  and  $\alpha \in \mathcal{P}$ . Henceforth, the triplet  $(f, \mathcal{I}, \mathcal{P})$  denotes a class of models of interest. Let  $x(t) = x(t, x_0, \alpha)$  denote the solution of equation (2.1). Because the solution depends on  $x_0$  and  $\alpha$ , the vector  $c = (x_0, \alpha)^T \in \mathbb{R}^2$  is called the control and in the rest of the paper, we use  $x(t)$  and  $x(t, c)$  interchangeably.

For the FSM theoretical analysis that we carry out, in addition to the existence and uniqueness of the solution of equation (2.1), we are also interested in the smoothness of the solution regarding the existence of continuous (mixed) partial derivatives of  $x(t)$  with respect to  $x_0$  and  $\alpha$  of order  $K \geq 1$ . In this context, we recall that a function  $g: \mathbb{R} \rightarrow \mathbb{R}$  belongs to the class  $C^K$  if  $g$  and its first  $K$  derivatives are continuous. According to a theorem in Chapter 2, Section 7 in Arnold [51], the solution  $x(t)$  of equation (2.1) belongs to  $C^K$  for some integer  $K \geq 1$  if  $f(x, \alpha)$  belongs to the same class  $C^K$ . We define  $\mathcal{S}(f) = \{x(t, c) | c \in \mathcal{I} \times \mathcal{P}\}$  as the ensemble of all possible solutions of equation (2.1) for a given  $f(x, \alpha) \in C^K$ , where  $x(0)$  and  $\alpha$  are varied in  $\mathcal{I}$  and  $\mathcal{P}$ . For example, if  $\dot{x} = ax$ ,  $\mathcal{I} = \mathbb{R}^+$  (the positive real line), and  $\mathcal{P} = [-1, 1]$ , then  $\mathcal{S}(f = ax)$  is the set of all exponentials of the form  $x(t) = e^{at}x(0)$  where  $a \in [-1, 1]$  and  $x(0) \in \mathbb{R}^+$ .

The forward sensitivities of the solution at time  $t > 0$  with respect to the initial condition  $x_0$  and parameter  $\alpha$  can be defined as follows:

$$u(t) = \frac{\partial x(t)}{\partial x_0} \quad \text{and} \quad v(t) = \frac{\partial x(t)}{\partial \alpha}. \quad (2.2)$$

By differentiating equation (2.1) with respect to  $x(0)$  and  $\alpha$ , it can be verified that  $u(t)$  and  $v(t)$  evolve according to linear non-autonomous systems given by

$$\left. \begin{aligned} \dot{u}(t) &= D_f(x(t))u(t), \quad u(0) = 1 \\ \dot{v}(t) &= D_f(x(t))v(t) + D_f^\alpha(x(t)), \quad v(0) = 0, \end{aligned} \right\} \quad (2.3)$$

and

$$\text{where} \quad D_f(x) = \frac{\partial f}{\partial x} \quad \text{and} \quad D_f^\alpha(x) = \frac{\partial f}{\partial \alpha}. \quad (2.4)$$

We refer to Lakshmivarahan and co-workers [52,53] for the derivation of the FSM. By solving equations (2.1) and (2.3), we can compute the evolution of the solution  $x(t)$  and its sensitivities with respect to time.

### (b) Observations

It is assumed that there is an underlying physical process and that the model  $(f, \mathcal{I}, \mathcal{P})$  defined above is faithful to this process in the sense that all the large features of the process are captured by the model. Let  $\tilde{x}(t)$  for  $t \geq 0$  be the true (yet unknown) state of the system under consideration. It is often the case that we may not be able to observe  $\tilde{x}(t)$  but only a certain function of it at discrete points in time and space. In addition, observations get corrupted by device errors and

measurement noise that is usually modelled as additive Gaussian noise with zero mean and known variance  $\sigma^2$ .

Let  $h: \mathbb{R} \rightarrow \mathbb{R}$ , where  $h \in C^K$ , be the function that maps the true state  $\bar{x}(k)$  into the observable  $z(k)$  as follows:

$$z(k) = h(\bar{x}(k)) + \eta(k), \quad (2.5)$$

where  $k$  is the discrete time index and  $\eta(k)$  is white noise, meaning that the observation  $z(k)$  contains noisy information about the (unknown) true state. Finally, we define  $\mathcal{O} = \{z(k_i) | 1 \leq i \leq N\}$  as the set of  $N$  observations at discrete times given by  $0 \leq k_1 \leq k_2 \leq \dots \leq k_N$ .

### (c) Statement of the problem

Given the model, equation (2.1), and the set  $\mathcal{O}$  of observations, our goal is to find the initial condition  $x(0) \in \mathcal{I}$  and parameter  $\alpha \in \mathcal{P}$  such that the solution  $x(t)$  of equation (2.1) starting with  $c = (x_0, \alpha)^T$  minimizes the weighted sum of squared error between  $z(k)$  and  $h(x(k))$ . To this end, we define a cost functional as follows:

$$J(c) = \frac{1}{2\sigma^2} \sum_{i=1}^N [z(k_i) - h(x(k_i))]^2. \quad (2.6)$$

The problem of minimizing  $J(c)$  can be solved by computing the gradient  $\nabla J(c) \in \mathbb{R}^2$ , called the adjoint gradient, and the minimizer can be sought by using the adjoint gradient in an iterative minimization algorithm [4].

## 3. Observability Gramian and adjoint gradient

We first examine the fine structure of the adjoint gradient and its dependence on the observability Gramian using the forward sensitivity analysis [24,25]. Let  $x(t)$  and  $\bar{x}(t)$  be the solution of equation (2.1) starting from  $c = (x_0, \alpha)^T$  and  $\bar{c} = (\bar{x}_0, \bar{\alpha})^T$ , respectively, where the perturbations in the initial conditions and the parameter are given by

$$\delta x_0 = \bar{x}_0 - x_0, \quad \delta \alpha = \bar{\alpha} - \alpha, \quad \delta c = (\delta x_0, \delta \alpha)^T. \quad (3.1)$$

The induced first variation in the solution of equation (2.1) resulting from the initial perturbation  $\delta c$  in the control  $c$  is defined as follows:

$$\delta x(k) = \bar{x}(k) - x(k). \quad (3.2)$$

The forecast error, also known as the innovation, can be thus computed as follows:

$$e(k) = z(k) - h(x(k)) = h(\bar{x}(k)) - h(x(k)) + \eta(k). \quad (3.3)$$

Because  $h \in C^K$  and by substituting equation (3.2) in equation (3.3) and applying first-order Taylor expression, the deterministic component of  $e(k)$  can be written as follows:

$$e(k) = D_h(x(k))\delta x(k), \quad (3.4)$$

where  $D_h(x) = \partial h / \partial x$ . Also, recall from first principles  $\delta x(k) = u(k)\delta x_0 + v(k)\delta \alpha = F\delta c$ , where

$$F = [u, v] \in \mathbb{R}^{1 \times 2}. \quad (3.5)$$

From equations (3.4) and (3.5), we get:

$$e(k) = D_h(x(k))F\delta c. \quad (3.6)$$

Now, let  $\delta J$  be the induced variation in  $J(c)$  resulting from the initial perturbation  $\delta c$  in  $c$ . Then, it can be verified that [4]:

$$\delta J = -\frac{1}{\sigma^2} \sum_{i=1}^N e(k_i)D_h(x(k_i))\delta x(k) = -\frac{1}{\sigma^2} \sum_{i=1}^N e(k_i)D_h(x(k_i))F\delta c. \quad (3.7)$$



Substituting for  $e(k)$  using equation (3.6), we get an alternative expression for  $\delta J$  as follows:

$$\delta J = -\frac{1}{\sigma^2} \delta c^T G \delta c, \quad (3.8)$$

where

$$G = \sum_{i=1}^N G_i \in \mathbb{R}^{2 \times 2} \quad \text{and} \quad G_i = F^T D_h^2(x(k_i)) F \in \mathbb{R}^{2 \times 2}. \quad (3.9)$$

The  $2 \times 2$  matrix  $G$  is known as the observability Gramian, which is always symmetric and, in general, positive semi-definite. The problem of placing observations is determining the minimum number of observations required and where to place them in order to make  $G$  positive definite. To this end, we rewrite  $G$  as

$$G = \sum_{i=1}^N D_h^2(x(k_i)) \begin{bmatrix} u^2(k_i) & u(k_i)v(k_i) \\ v(k_i)u(k_i) & v^2(k_i) \end{bmatrix}. \quad (3.10)$$

Under the regularity assumption that  $D_h(x)$  does not vanish along the solution in the set  $\mathcal{S}(f)$  in §2a and that  $D_h^2(x(k_i))$  is positive, a good rule is to place the observations where  $u^2(k_i)$  and  $v^2(k_i)$  attain their maximum value. Clearly, in this case, we would need only  $N = 2$  observations: one at the maximum of  $u^2(k)$  and another at the maximum of  $v^2(k)$ . We note that by placing the observations where  $u^2$  and  $v^2$  attain their maximum values, the observability Gramian  $G$  becomes positive definite, which in turn bounds the norm of the adjoint gradient away from zero (see equation (3.8)). By doing so, flat patches of the cost functional are avoided and the convergence of the optimization algorithm is improved.

## 4. Natural consequence of the placement strategy: a theoretical analysis

The strategy of placing observations where the square of the forward sensitivities attains a maximum has another natural optimality property. To examine this inherent optimality, we seek an expression for the adjoint gradient starting from the fact that  $\delta J = \langle \nabla J(c), \delta c \rangle$  and at the extremum  $\nabla J(c) = 0$ , where  $\langle \cdot, \cdot \rangle$  denotes the inner product. Combining this with equation (3.7), we obtain the following conditions for the optimality:

$$g_1 = \frac{\partial J}{\partial x_0} = -\frac{1}{\sigma^2} \sum_{i=1}^N (z(k_i) - h(x(k_i))) D_h(x(k_i)) U(k_i) = 0 \quad (4.1)$$

and

$$g_2 = \frac{\partial J}{\partial \alpha} = -\frac{1}{\sigma^2} \sum_{i=1}^N (z(k_i) - h(x(k_i))) D_h(x(k_i)) V(k_i) = 0. \quad (4.2)$$

Because the maps  $f$  in equation (2.1) and  $h$  in equation (2.5) are fixed, it can be verified that  $g_i = g_i(z, x_0, \alpha)$  is a function of  $z, x_0, \alpha$ , where  $z = (z(k_1), \dots, z(k_N))^T \in \mathbb{R}^N$ . In the following analysis, we consider two cases corresponding to a single observation with a linear observation operator and multiple observations with a generic nonlinear observation operator. Although the former is a subset of the latter, we begin with the simpler case to define the main quantities that enable us to understand and analyse the second advantage of the given observation placement strategy.

### (a) Case 1: scalar dynamical systems with unknown initial condition

In the first case, we consider a single observation (i.e.  $N = 1$ ), known model parameter  $\alpha$  and unknown initial condition  $x_0$ . Moreover, we assume that the model state is directly observable (i.e.  $h(x) = x$  and  $D_h(x) = 1$ ). Thus, the optimality condition in equation (4.1) can be reduced to the

following:

$$-\sigma^2 g_1(z, k) = [z(k) - x(k)]u(k) = 0. \quad (4.3)$$

Because  $f$ ,  $h$  and  $\alpha$  are invariant, solving equation (4.3) yields an optimal value of  $x_0 = x_0(z, k)$  as a function of  $z(k)$  and  $k$ . We are interested in quantifying the sensitivity of the optimal estimate of  $x_0$  with respect to the observation  $z$  at time  $k$  as follows:

$$y(k) = \frac{\partial x_0(z, k)}{\partial z}. \quad (4.4)$$

To this end, by differentiating both sides of equation (4.3) with respect to  $z = z(k)$ , we obtain

$$0 = -\sigma^2 \frac{\partial g_1}{\partial z} = \left[1 - u(k) \frac{\partial x_0}{\partial z}\right] u(k) + [z(k) - x(k)] \frac{\partial u(k)}{\partial x_0} \frac{\partial x_0}{\partial z}. \quad (4.5)$$

Solving equation (4.5), we get the required expression:

$$y(k) = \frac{\partial x_0}{\partial z} = \frac{u(k)}{u^2(k) - e(k)(\partial u(k)/\partial x_0)}. \quad (4.6)$$

A little reflection reveals that when  $x(t)$  is the optimal trajectory,  $e(k) = z(k) - x(k)$  is small and can be set to zero in equation (4.6) to give the following approximation:

$$y(k) = \frac{\partial x_0}{\partial z} \approx \frac{u(k)}{u^2(k)} = \frac{\text{sign}(u(k))}{|u(k)|}. \quad (4.7)$$

Equation (4.7) implies that  $y(k)$  is minimum where  $u^2(k)$  attains its maximum value. In other words, placing the single observation at time  $k^*$ , where  $u^2(k)$  attains its maximum value, ensures that the sensitivity  $y(k^*)$  takes its minimum value.

For the complementary case when  $\alpha$  is not known but  $x_0$  is known, the optimality condition in equation (4.2) becomes:

$$0 = \sigma^2 g_2(z, k) = [z(k) - x(k)]v(k). \quad (4.8)$$

Through similar arguments, it can be verified that the optimal  $\alpha$  that satisfies equation (4.8) is a function of  $z$  and  $k$ . We define the sensitivity of the optimal  $\alpha$  with respect to the observation  $z(k)$  as follows:

$$w(k) = \frac{\partial \alpha(z, k)}{\partial z}. \quad (4.9)$$

Differentiating equation (4.8) with respect to  $z$  and simplifying using the same reasoning used in obtaining equation (4.7), we get:

$$w(k) = \frac{v(k)}{v^2(k) - e(k)(\partial v(k)/\partial \alpha)} \approx \frac{\text{sign}(v(k))}{|v(k)|}. \quad (4.10)$$

Thus,  $w(k)$  takes on its minimum value at  $k = k^*$  where  $v^2(k)$  attains its maximum value, implying that the proposed observation placement strategy (i.e. where  $v^2(k)$  attains its maximum value) leads to the minimum value of  $w(k)$ .

## (b) Case 2: scalar dynamical systems with unknown parameter and initial condition

Next, we extend our theoretical analysis to consider the simultaneous estimation of unknown initial conditions and model parameters in the general case of multiple observations and arbitrary observation operators. Solving equations (4.1) and (4.2), it follows that the optimal estimates  $x_0 = x_0(z)$  and  $\alpha = \alpha(z)$  are functions of the observations, where  $z = (z(k_1), \dots, z(k_N))^T \in \mathbb{R}^N$ . Let  $x = x(t, x_0(z), \alpha(z))$  be the optimal solution of equation (2.1), starting from the optimal control  $c(z) = (x_0(z), \alpha(z))^T$ . In addition, we suppose that  $u(t) = u(t, x_0(z), \alpha(z))$ ,  $v(t) = v(t, x_0(z), \alpha(z))$  and  $D_h(x(t))$  are evaluated along the optimal trajectory  $x(t)$ .

To bring out the key ideas and simplify the algebra, without loss of generality, we set  $N = 2$  and denote  $k_i$  by  $i$ . Then, the optimality conditions given by equations (4.1) and (4.2) for  $i = 1, 2$



can be rewritten as follows:

$$\left. \begin{aligned} \sigma^2 g_1 = H_1 = H_{11} + H_{12} = 0 \\ \sigma^2 g_2 = H_2 = H_{21} + H_{22} = 0, \end{aligned} \right\} \quad (4.11)$$

and

where

$$H_{11} = [z(1) - h(x(1))]D_h(x(1))u(1), \quad H_{12} = [z(2) - h(x(2))]D_h(x(2))u(2) \quad (4.12)$$

and

$$H_{21} = [z(1) - h(x(1))]D_h(x(1))v(1), \quad H_{22} = [z(2) - h(x(2))]D_h(x(2))v(2), \quad (4.13)$$

where  $z(i)$ ,  $x(i)$ ,  $u(i)$  and  $v(i)$  are the values of the respective quantities evaluated at time  $t = k_i = i$ ,  $1 \leq i \leq 2$ . We define the sensitivities of  $x_0(z)$  and  $\alpha(z)$  with respect to  $z_i$  for  $i = 1, 2$  as follows:

$$y_i = \frac{\partial x_0(z)}{\partial z_i} \quad \text{and} \quad w_i = \frac{\partial \alpha(z)}{\partial z_i}. \quad (4.14)$$

Our goal is to relate the sensitivities in equation (4.14) to the forecast sensitivities  $u(k)$  and  $v(k)$ . As illustrated in case 1 above, this can be accomplished by computing the derivatives of  $H_i = 0$  in equation (4.11) with respect to  $z_i$  for  $1 \leq i \leq 2$ , which, when simplified, gives the sought after relation. The details of this derivation are given in appendix A. From equations (A 12) and (A 13) of this appendix, we get the following:

$$G \begin{bmatrix} y_i \\ w_i \end{bmatrix} = D_h(i) \begin{bmatrix} u_i \\ v_i \end{bmatrix}, \quad (4.15)$$

for  $i = 1, 2$ , where  $G = G_1 + G_2$ ,  $F(i) = [u(i), v(i)] \in \mathbb{R}^{1 \times 2}$ . The observability Gramian decomposition  $G_i$  can be evaluated as follows:

$$G_i = F^T(i)D_h^2(i)F(i), \quad (4.16)$$

where it is assumed that  $D_h^2(i) = D_h^2(x(i)) > 0$  along the trajectory of  $x$  in equation (2.1). Consequently, the inverse of  $G$  controls the behaviour of  $y_i$  and  $w_i$ . Setting  $D_h^2(i) = d_i$  for simplicity in notation, it can be verified that

$$G = \begin{bmatrix} d_1^2 u_1^2 + d_2^2 u_2^2 & d_1^2 u_1 v_1 + d_2^2 u_2 v_2 \\ d_1^2 u_1 v_1 + d_2^2 u_2 v_2 & d_1^2 v_1^2 + d_2^2 v_2^2 \end{bmatrix}, \quad (4.17)$$

and its determinant is given by

$$|G| = d_1^2 d_2^2 (u_1 v_2 - u_2 v_1)^2. \quad (4.18)$$

Because the sensitivities can generally be positive or negative, for definiteness, our strategy is based on the square of the sensitivities. Therefore, placing the first observation  $z_1$  at the maximum value of  $u_1^2$  and the second observation  $z_2$  at the maximum value of  $v_2^2$  maximizes  $|G|$ . This in turn reduces the magnitudes of  $y_i$  and  $w_i$  as desired, per Cramer's rule. Table 1 provides expressions for  $y(i)$  and  $w(i)$  for  $i = 1, 2$ .

**Remark 4.1.** For an increased number of observations, it is crucial to guarantee that the observation placement strategy does not lead to a singular Gramian matrix. In appendix B, we prove that the forecast sensitivities with respect to the initial condition and parameter cannot be linearly dependent and that the resulting Gramian is, therefore, non-singular.

### (c) Extensions to high-dimensional systems

We consider the case for which the state variable  $\mathbf{x}$  is represented by a vector (i.e.  $\mathbf{x} \in \mathbb{R}^n$ , where  $n$  is the number of degrees of freedom of the system) and  $\boldsymbol{\alpha} \in \mathbb{R}^p$ . The discrete-time version of the

**Table 1.** Expressions for  $y(i)$  and  $w(i)$  for  $i = 1, 2$ .

	$y$	$w$
$i = 1$	$\frac{v_2(u_1 v_2 - u_2 v_1)}{ G }$	$\frac{u_2(u_1 v_1 - u_1 v_2)}{ G }$
$i = 2$	$\frac{v_1(u_2 v_1 - u_1 v_2)}{ G }$	$\frac{u_1(u_1 v_2 - u_2 v_1)}{ G }$

model in equation (2.1) for  $\mathbf{x} \in \mathbb{R}^n$  can be written as follows:

$$\mathbf{x}_{k+1} = \mathbf{M}(\mathbf{x}_k, \boldsymbol{\alpha}), \quad (4.19)$$

where  $\mathbf{M}: \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$  defines the one-time step mapping defined by applying a temporal integration scheme (e.g.  $\mathbf{M}(\mathbf{x}_k, \boldsymbol{\alpha}) = \mathbf{x}_k + (t_{k+1} - t_k)f(\mathbf{x}_k, \boldsymbol{\alpha})$  for a first-order Euler integrator). Equation (2.3) for the dynamics of  $\mathbf{u}(t) \in \mathbb{R}^{n \times n}$  and  $\mathbf{v}(t) \in \mathbb{R}^{n \times p}$  can be rewritten as follows:

$$\mathbf{u}_{k+1} = \mathbf{D}_\mathbf{M}(k)\mathbf{u}_k \quad \text{and} \quad \mathbf{v}_{k+1} = \mathbf{D}_\mathbf{M}(k)\mathbf{v}_k + \mathbf{D}_\mathbf{M}^\alpha(k), \quad (4.20)$$

where

$$\left[ \mathbf{D}_\mathbf{M}(k) \right]_{i,j} = \frac{\partial M_i}{\partial x_j} \bigg|_{\mathbf{x}=\mathbf{x}_k} \quad \text{and} \quad \left[ \mathbf{D}_\mathbf{M}^\alpha(k) \right]_{i,j} = \frac{\partial M_i}{\partial \alpha_j} \bigg|_{\mathbf{x}=\mathbf{x}_k}. \quad (4.21)$$

Without loss of generality, we focus our analysis on estimating the model's initial condition  $\mathbf{x}_0$  from a set of measurements,  $\mathbf{z} \in \mathbb{R}^m$  (i.e. the control is  $\mathbf{c} = \mathbf{x}_0$ ).

### (i) Linear dynamics (vector)

We first consider a linear dynamical system:

$$\mathbf{x}_{k+1} = \mathbf{M}\mathbf{x}_k, \quad (4.22)$$

where  $\mathbf{M} \in \mathbb{R}^{n \times n}$  is a square matrix with rank  $n$ . Furthermore, we suppose that the measurement vector  $\mathbf{z} \in \mathbb{R}^m$  is related to the state  $\mathbf{x}$  by a linear operator  $\mathbf{H} \in \mathbb{R}^{m \times n}$  as follows:

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \boldsymbol{\eta}_k, \quad (4.23)$$

where  $\boldsymbol{\eta}_k \in \mathbb{R}^m$  represents an additive measurement noise with zero mean and covariance matrix  $\mathbf{R} \in \mathbb{R}^{m \times m}$ . The cost function for the inverse problem can be written as

$$J(\mathbf{c}) = \frac{1}{2} \sum_{i=1}^N [\mathbf{z}_i - \mathbf{H}\mathbf{x}_i]^\top \mathbf{R}^{-1} [\mathbf{z}_i - \mathbf{H}\mathbf{x}_i]. \quad (4.24)$$

From equation (4.22), we can relate the prediction  $\mathbf{x}_i$  to the initial condition  $\mathbf{x}_0 = \mathbf{c}$  as  $\mathbf{x}_i = \mathbf{M}^i \mathbf{x}_0 = \mathbf{M}^i \mathbf{c}$  and  $\mathbf{x}_i^\top = \mathbf{c}^\top (\mathbf{M}^i)^\top = \mathbf{c}^\top (\mathbf{M}^\top)^i$ . We also note that the forward sensitivity of the model forecast with respect to its initial condition can be defined as  $\mathbf{u}_i = \mathbf{M}^i$  because the model Jacobian is  $\mathbf{D}_\mathbf{M} = \mathbf{M}$  and  $\mathbf{u}_0 = \mathbf{I}$  (the identity matrix). Therefore, equation (4.24) and its gradient can be rewritten as

$$J(\mathbf{c}) = \frac{1}{2} \sum_{i=1}^N \left[ \mathbf{z}_i^\top \mathbf{R}^{-1} \mathbf{z}_i - 2\mathbf{z}_i^\top \mathbf{R}^{-1} \mathbf{H} \mathbf{M}^i \mathbf{c} + \mathbf{c}^\top ((\mathbf{M}^\top)^i \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H} \mathbf{M}^i) \mathbf{c} \right] \quad (4.25)$$

and

$$\nabla_c J = \sum_{i=1}^N [(\mathbf{M}^\top)^i \mathbf{H}^\top \mathbf{R}^{-1} [\mathbf{H} \mathbf{M}^i \mathbf{c} - \mathbf{z}_i]]. \quad (4.26)$$

Setting the gradient  $\nabla_c J$  in equation (4.26) to zero, we get an optimal expression for  $\mathbf{c}$  as follows:

$$\sum_{i=1}^N (\mathbf{M}^\top)^i \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H} \mathbf{M}^i \mathbf{c} = \sum_{i=1}^N (\mathbf{M}^\top)^i \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{z}_i. \quad (4.27)$$

The observability Gramian is now defined as  $\mathbf{G} = \sum_{i=1}^N \mathbf{G}_i$ , where  $\mathbf{G}_i = (\mathbf{M}^T)^i \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{M}^i$ . Thus, the optimal  $\mathbf{c}$  is given by

$$\mathbf{c} = \mathbf{G}^{-1} \sum_{i=1}^N (\mathbf{M}^T)^i \mathbf{H}^T \mathbf{R}^{-1} \mathbf{z}_i. \quad (4.28)$$

Differentiating both sides of equation (4.28) with respect to  $\mathbf{z}_i$  yields the following expression for the sensitivity of the optimal estimate of the control  $\mathbf{c}$  to the measurements:

$$\frac{\partial \mathbf{c}}{\partial \mathbf{z}_i} = \mathbf{G}^{-1} (\mathbf{M}^T)^i \mathbf{H}^T \mathbf{R}^{-1}. \quad (4.29)$$

Maximizing the square of the forward sensitivity of the model forecast with respect to the control yields a Gramian matrix  $\mathbf{G}$  with the largest determinant, which minimizes the sensitivity of the optimal estimate of the control to the actual measurements.

We note that the minimum number of observation times  $N$  to yield a well-posed inverse problem is defined as  $N = n/m$ . However, we highlight that due to the existence of measurement noise and correlations between measurement components, a regularization is often imposed for the solution of the inverse problem. For the special case of  $m = n$ , we place a single measurement  $\mathbf{z}_i \in \mathbb{R}^n$  at arbitrary time  $i$ . Assuming  $\mathbf{H} = \mathbf{I}$  and  $\mathbf{R}^{-1} = (1/\sigma^2)\mathbf{I}$ , we get the following expressions for the Gramian and the sensitivity of the optimal estimate of the control  $\mathbf{c}$  with respect to the measurement  $\mathbf{z}_i$ :

$$\mathbf{G} = \frac{1}{\sigma^2} (\mathbf{M}^T)^i \mathbf{M}^i \quad \text{and} \quad \frac{\partial \mathbf{c}}{\partial \mathbf{z}_i} = \mathbf{M}^{-i}. \quad (4.30)$$

As noted above, the forecast sensitivity with respect to the initial conditions can be written as  $\mathbf{u}_i = \mathbf{M}^i$  in the linear dynamics case. Therefore, the sensitivity  $\partial \mathbf{c} / \partial \mathbf{z}_i$  is inversely proportional to the sensitivity of  $\mathbf{x}_i$  with respect to  $\mathbf{c}$ .

**Remark 4.2.** The system matrix  $\mathbf{M}$  in equation (4.22) is a real matrix and it may or may not be diagonalizable. For the latter, it is similar to the Jordan normal form. For the case where  $\mathbf{M}$  is diagonalizable, there exists a non-singular matrix  $\mathbf{P}$  such that

$$\Lambda = \mathbf{P}^{-1} \mathbf{M} \mathbf{P}, \quad (4.31)$$

where  $\Lambda$  is a diagonal matrix of the real eigenvalues of  $\mathbf{M}$ . Therefore,  $\mathbf{M}^i = \mathbf{P} \Lambda^i \mathbf{P}^{-1}$ , where  $\Lambda^i = \text{Diag}(\lambda_1^i, \lambda_2^i, \dots, \lambda_n^i)$ , where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ . In this case, equation (4.30) can be rewritten as follows:

$$\mathbf{P}^{-1} \left[ \frac{\partial \mathbf{c}}{\partial \mathbf{z}_i} \right] \mathbf{P} = \Lambda^{-i}. \quad (4.32)$$

Therefore, if  $|\lambda_k| > 1$ ,  $\lambda_k^i$  is large (as well as the square of forward sensitivities because  $\mathbf{u}_i = \mathbf{M}^i$ ) and  $\partial \mathbf{c} / \partial \mathbf{z}_i$  decreases. In addition, equation (4.32) shows that the basis  $\mathbf{P}$  that diagonalizes  $\mathbf{M}$  also diagonalizes the matrix  $[\partial \mathbf{c} / \partial \mathbf{z}_i]_{n \times n}$ , which implies that  $[\partial \mathbf{c} / \partial \mathbf{z}_i]_{n \times n}$  is similar to a diagonal matrix.

## (ii) Nonlinear dynamics (vector)

We consider the nonlinear case of equation (4.22) as follows:

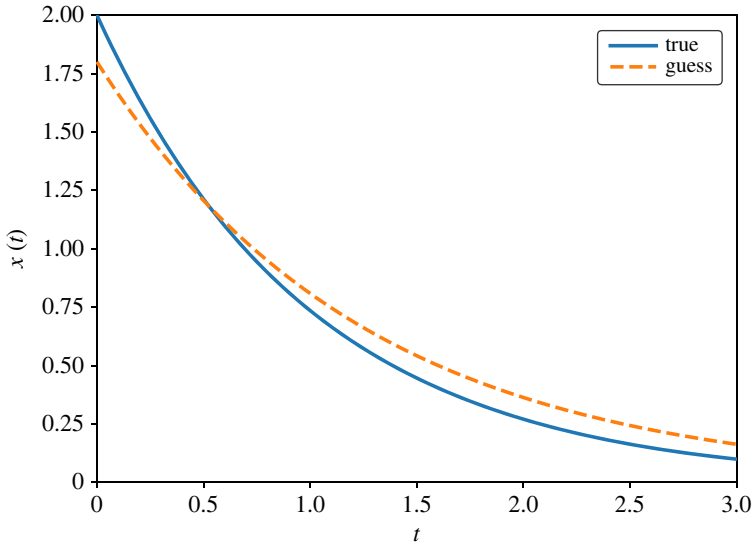
$$\mathbf{x}_{k+1} = \mathbf{M}(\mathbf{x}_k) \quad \text{and} \quad \mathbf{z}_k = \mathbf{h}(\mathbf{x}_k) + \boldsymbol{\eta}_k. \quad (4.33)$$

The forecast error is defined as  $\mathbf{e}_k = \mathbf{z}_k - \mathbf{h}(\mathbf{x}_k)$ , and the cost function can be written as follows:

$$J(\mathbf{c}) = \frac{1}{2} \sum_{i=1}^N \mathbf{e}_k^T \mathbf{R}^{-1} \mathbf{e}_k = \sum_{i=1}^N J_k(\mathbf{c}), \quad (4.34)$$

where  $J_k(\mathbf{c}) = (1/2) \mathbf{e}_k^T \mathbf{R}^{-1} \mathbf{e}_k$ . To simplify our analysis, we assume  $\mathbf{R} = \sigma^2 \mathbf{I}$ , which gives the following:

$$J_k(\mathbf{c}) = \frac{1}{2\sigma^2} \mathbf{e}_k^T \mathbf{e}_k = \frac{1}{2\sigma^2} \sum_{j=1}^m (z_{i,j} - h_j(\mathbf{x}_i))^2, \quad (4.35)$$



**Figure 1.** Solution of the decay model using the true and guess controls.

where  $z_{i,j}$  is the  $j$ th component of the measurement vector  $\mathbf{z}_i$  at time  $t_i$ . Thus, the cost function  $J(\mathbf{c})$  can be written as follows:

$$J(\mathbf{c}) = \frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{j=1}^m (z_{i,j} - h_j(\mathbf{x}_i))^2. \quad (4.36)$$

In other words,  $J(\mathbf{c})$  is additive in  $J_{k,i} := (1/2\sigma^2)(z_{i,j} - h_j(\mathbf{x}_i))^2$ . Therefore, it is sufficient to investigate  $J_{k,i}$  for the analysis of the dual advantages of the proposed forward sensitivity observation placement strategy. The analysis of  $J_{k,i}$  is similar to the analysis of  $J(\mathbf{c})$  for the scalar case presented in §4 and appendix A.

## 5. Numerical experiments

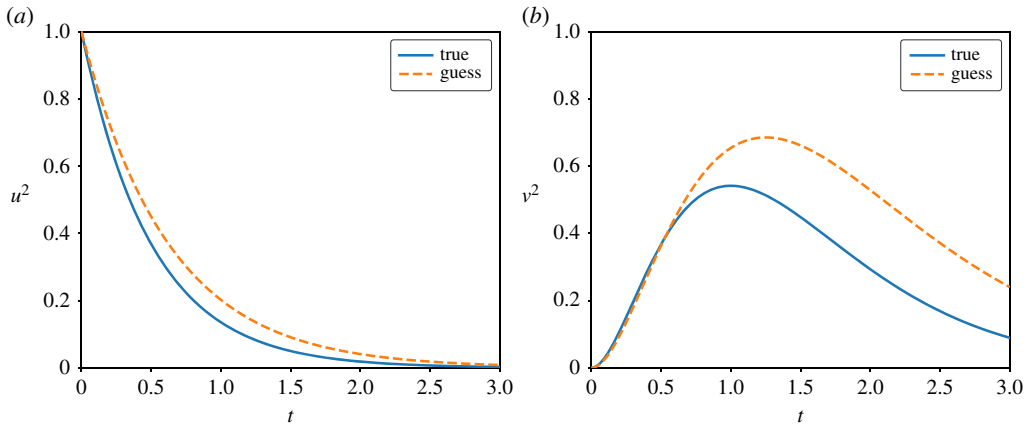
In this section, we consolidate the results related to the application of the theoretical analysis carried out in §§3 and 4 to the dynamical system introduced in §2. First, we explore the advantages of observation placement using the FSM with two test cases for linear and nonlinear scalar dynamical systems. In both cases, the initial condition and model parameter represent the control variable. We investigate the dependence of the optimal estimate of the sought control variable on the time instants at which measurements are collected. The forward sensitivities (i.e.  $u$  and  $v$ ) as well as the sensitivities of the control variable estimates with respect to the different measurements are computed to analyse the effect of observation placement on solving the inverse problem. After that, we showcase the success of the FSM-based observation placement strategy using two problems governed by PDEs, namely the one-dimensional Burgers and the two-dimensional advection diffusion test cases.

### (a) Linear dynamics (scalar)

The first example that we consider is an exponential decay model defined as follows:

$$\dot{x} = ax, \quad (5.1)$$

where  $x$  is the model state and  $a < 0$  is the decay parameter. It can be verified that the analytic solution of this ODE is  $x(t) = x_0 e^{at}$ , where  $x(0) = x_0$  is the initial condition. For our numerical experiments, we set the true controls as  $\bar{\mathbf{c}} = [\bar{x}_0, \bar{a}]^T = [2, -1]^T$ . These are assumed to be unknown,



**Figure 2.** Model sensitivity with respect to (a) initial condition  $x_0$  (left) and (b) the decay parameter  $a$  (right).

and instead the initial guessed values are given by  $c = [x_0, a]^T = [1.8, -0.8]^T$ . The dynamical evolution of the decay in response to the true and guess controls is shown in figure 1.

Synthetic observations are created from the true state with added noise of zero mean and a standard deviation 10% of the model state at the times when the observations are made. We follow the proposed methodology for placing the observation where the forward sensitivities attain their maximum values. For the considered exponential decay model, the forward sensitivities of the model predictions with respect to the control variables can be written as follows:

$$u(t) = e^{at} \quad \text{and} \quad v(t) = t e^{at} x_0 = tx(t). \quad (5.2)$$

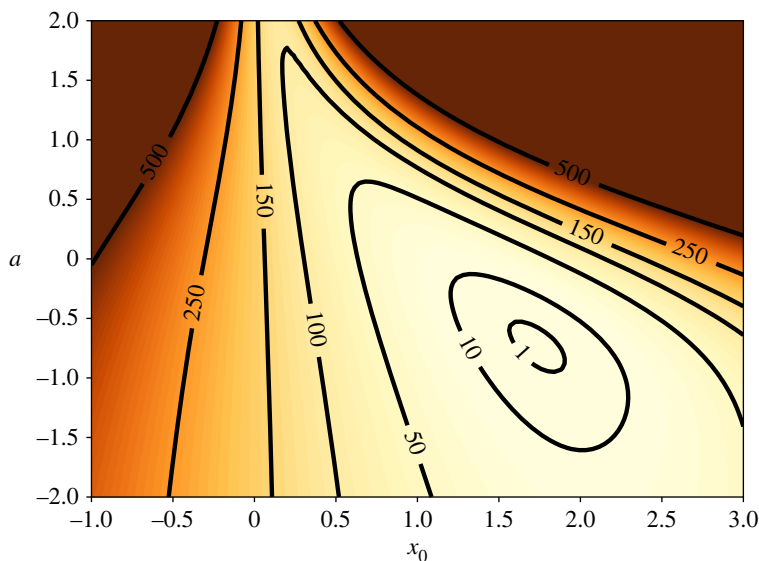
With  $a < 0$ ,  $u(t) = e^{-|a|t}$  attains the maximum at  $t = 0$  and  $v(t) = t e^{-|a|t} x_0$  attains the maximum at  $t^* = 1/|a|$ . To avoid a zero value of  $v(t)$ , we avoid placing observations at  $t = 0$ . Moreover, it can be verified from equation (4.18) that  $|G| = 0$  if  $t_1 = t_2$ . Therefore, we can place the two observations at  $t_1 = \epsilon > 0$  where  $\epsilon$  is a small value (as opposed to having the observation coincide with the initial time) and  $t_2 = t^* = 1/|a|$ . In this case, it can be shown that  $G$  is given by

$$G = e^{-2|a|\epsilon} \begin{bmatrix} 1 & \epsilon x_0 \\ \epsilon x_0 & \epsilon^2 x_0^2 \end{bmatrix} + e^{-2|a|t^*} \begin{bmatrix} 1 & t^* x_0 \\ t^* x_0 & t^{*2} x_0^2 \end{bmatrix}, \quad (5.3)$$

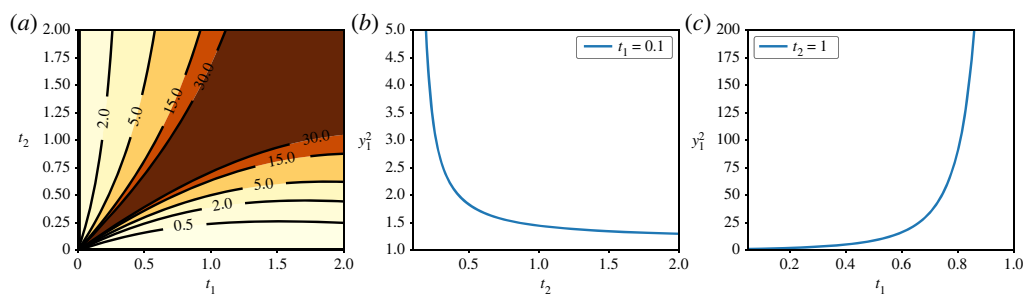
which is positive definite. The plots of forecast sensitivities to initial condition  $x_0$  and decay parameter  $a$  are shown in figure 2 for the true and guessed values of control  $c$ . In order to optimize the solution of the inverse problem, observation placements are chosen where the forecast sensitivities to control reach a maximum (i.e.  $u^2(t)$  and  $v^2(t)$  exhibit maxima). In particular, we place an observation at  $t_1 = k_1 = \epsilon = 0.1$  (to avoid singularity of the Gramian,  $G$ , in equation (4.18)) and at  $t_2 = k_2 = 1.0$ .

The resulting cost function is displayed in figure 3 for different control values where the minimum value occurs around the true values of the control. Numerically, the search for the minimum can be made using Newton's method, leading to optimal estimates of the control values as  $c = (x_0, a) = (1.7329, -0.7256)$ .

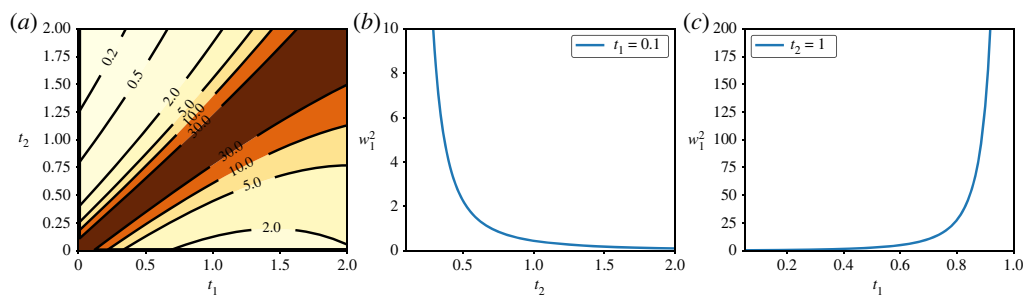
Plots of sensitivity of the optimal control estimates with respect to observations (i.e.  $y_1^2(t)$ ,  $w_1^2(t)$ ,  $y_2^2(t)$  and  $w_2^2(t)$ ) are given in figures 4–7. Figure 4 shows the contour plot of  $y_1^2(t)$  as well as the cross-sections of this contour at the observation times (i.e.  $t_1 = 0.1$  and  $t_2 = 1.0$ ). Similar plots for  $w_1^2(t)$  are covered in figure 5. It can be seen that maximum values of the sensitivity of the optimal estimate of  $x_0$  with respect to the first observation occur when  $t_1$  and  $t_2$  are close to each other. For example, by selecting  $t_1 = 0.1$ , placing the second observation at earlier times (e.g.  $t_2 < 0.5$ ) leads to very high sensitivity of the optimal estimate of  $x_0$  and  $a$  with respect to the first observation location. Thus, larger values of  $t_2$  are to be sought to minimize  $y_1^2$ .



**Figure 3.** Cost function for exponential decay model.

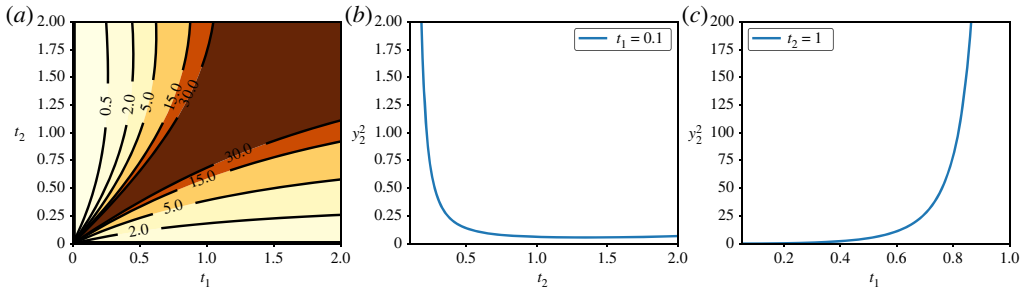


**Figure 4.** Sensitivity of the optimal estimates of the initial condition  $x_0$  to the first observation  $z_1$ . (a) Contour plot of  $y_1^2(t)$  at different values of  $t_1$  and  $t_2$ . (b)  $y_1^2(t)$  values at  $t_1 = 0.1$  and different values of  $t_2$ . (c)  $y_1^2(t)$  values at  $t_2 = 1.0$  and different values of  $t_1$ . Note that minimum values of  $y_1^2$  are obtained at  $t_1 = 0.1$  and  $t_2 = 1.0$ .

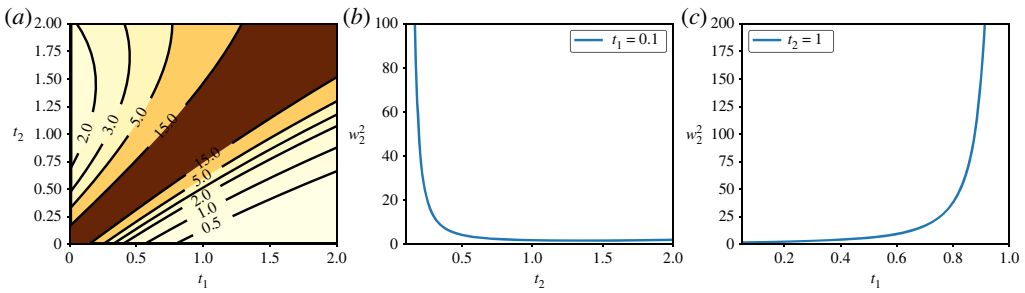


**Figure 5.** Sensitivity of the optimal estimates of the decay parameter  $a$  to the first observation  $z_1$ . (a) Contour plot of  $w_1^2(t)$  at different values of  $t_1$  and  $t_2$ . (b)  $w_1^2(t)$  values at  $t_1 = 0.1$  and different values of  $t_2$ . (c)  $w_1^2(t)$  values at  $t_2 = 1.0$  and different values of  $t_1$ . Note that minimum values of  $w_1^2$  are obtained at  $t_1 = 0.1$  and  $t_2 = 1.0$ .





**Figure 6.** Sensitivity of the optimal estimates of the initial condition  $x_0$  to the second observation  $z_2$ . (a) Contour plot of  $y_2^2(t)$  at different values of  $t_1$  and  $t_2$ . (b)  $y_2^2(t)$  values at  $t_1 = 0.1$  and different values of  $t_2$ . (c)  $y_1^2(t)$  values at  $t_2 = 1.0$  and different values of  $t_1$ . Note that minimum values of  $y_2^2$  are obtained at  $t_1 = 0.1$  and  $t_2 = 1.0$ .



**Figure 7.** Sensitivity of the optimal estimates of the decay parameter  $a$  to the second observation  $z_2$ . (a) Contour plot of  $w_2^2(t)$  at different values of  $t_1$  and  $t_2$ . (b)  $y_1^2(t)$  values at  $t_1 = 0.1$  and different values of  $t_2$ . (c)  $w_1^2(t)$  values at  $t_2 = 1.0$  and different values of  $t_1$ . Note that minimum values of  $w_2^2$  are obtained at  $t_1 = 0.1$  and  $t_2 = 1.0$ .

Likewise, figure 6 describes the properties of  $y_2^2(t)$  and figure 7 displays properties of  $w_2^2(t)$ . As expected from the theory, the sensitivity of the optimal control estimates with respect to the observations exhibit minimum values at the two times where the observation sites were chosen. For example, by selecting  $t_1 = 0.1$ ,  $y_2^2$  and  $w_2^2$  approach their minimum values at  $t_2 = 1$ . In a similar way, by selecting  $t_2 = 1$ ,  $y_2^2$  and  $w_2^2$  are minimized by placing the first observation at very small values of  $t_1$ .

## (b) Nonlinear dynamics (scalar)

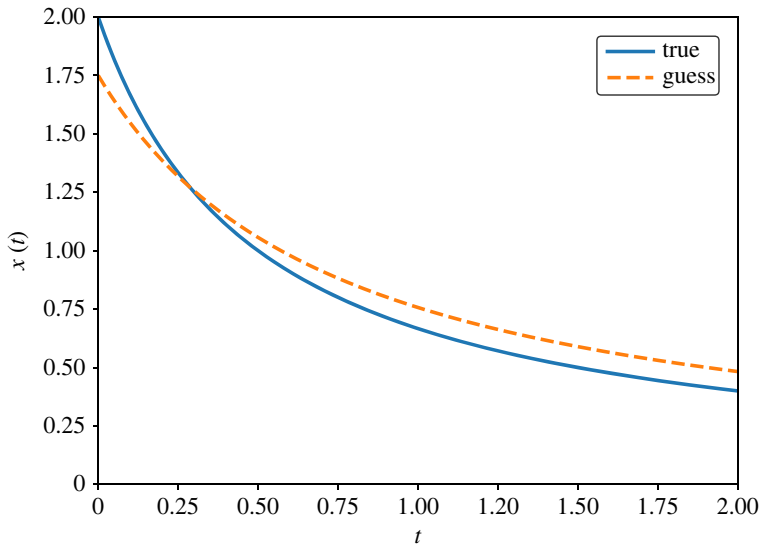
We extend our analysis and consider a system governed by nonlinear dynamics as follows:

$$\dot{x} = ax^2, \quad (5.4)$$

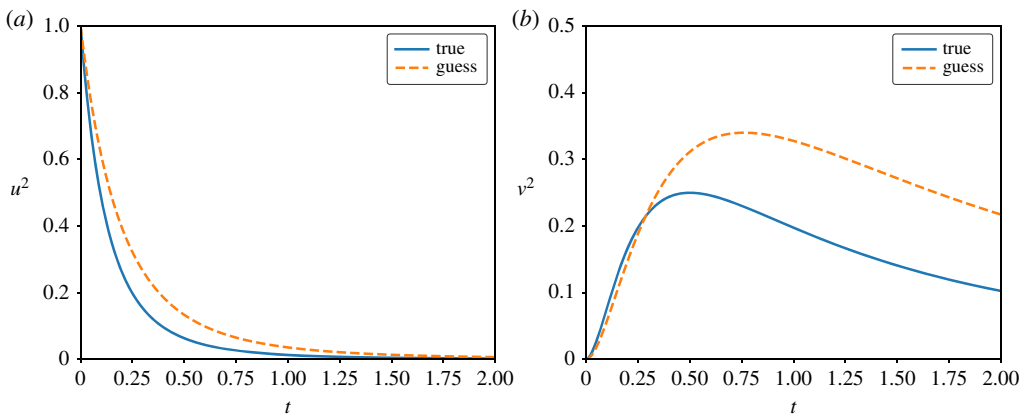
where  $x \in \mathbb{R}$  is the state variable and  $a \in \mathbb{R}$  is the model parameter. It can be further shown that the solution of this system is given by  $x(t) = x_0/(1 - ax_0t)$ . The true and guessed controls are defined as  $\bar{c} = [\bar{x}_0, \bar{a}]^T = [2, -1]^T$  and  $c = [x_0, a]^T = [1.75, -0.75]^T$ , respectively. The dynamical evolution of the nonlinear decay system in response to the true and guess controls is shown in figure 8.

The sensitivity of the model forecast with respect to the initial condition and model parameter can be written as follows:

$$u(t) = \frac{1}{(1 - ax_0t)^2} \quad \text{and} \quad v(t) = \frac{x_0^2 t}{(1 - ax_0t)^2}. \quad (5.5)$$



**Figure 8.** Solution of the nonlinear decay model using the true and guess controls.

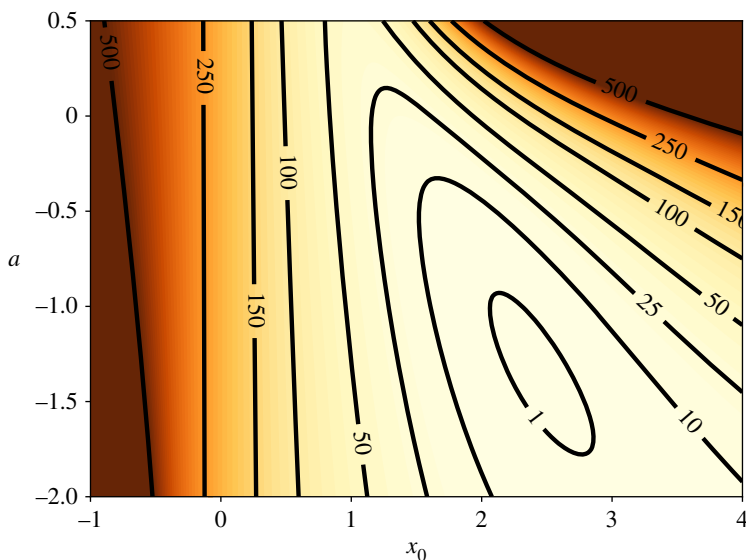


**Figure 9.** Model sensitivity with respect to (a) initial condition  $x_0$  (left) and (b) the decay parameter  $a$  (right) for  $\dot{x} = ax^2$ .

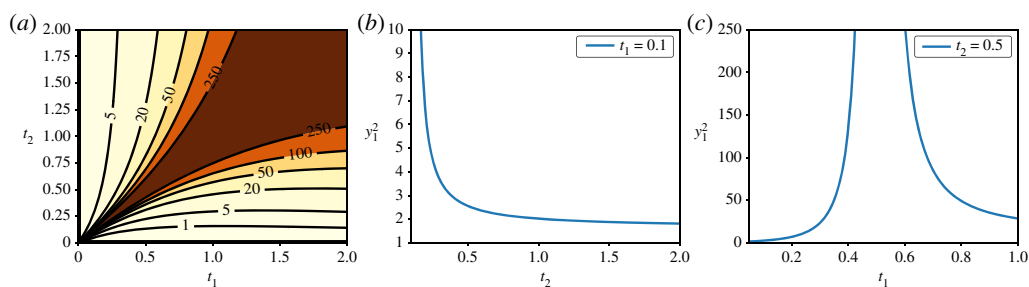
The evolution of these sensitivities with time is depicted in figure 9. We note that similar figures can be obtained by solving equation (2.3) numerically, where  $D_f(x) = 2ax$  and  $D_f^a(x) = x^2$ .

Although the maximum value of  $u^2$  occurs at  $t=0$ , it corresponds to a zero value of  $v^2$ . Thus, we avoid placing observations at the initial time and we select  $t_1 = \epsilon = 0.1 > 0$ . Moreover, we select  $t_2 = 0.5$  to maximize the value of  $v^2$ . We follow a twin experiment approach to generate synthetic observations by adding white Gaussian noise to the true system state (corresponding to the true controls). The resulting cost function is depicted in figure 10, where Newton's method can be followed to obtain the optimal values of  $x_0$  and  $a$ .

As highlighted before, we place the observations where the squares of the forward sensitivities of the model forecast with respect to the control are maximized. This guarantees that the observability Gramian is positive definite and subsequently keeps the adjoint gradient away from zero, which in turn accelerates the search for optimal estimates of the control. The second advantage of the proposed observation placement strategy is related to the robustness of the solution of the inverse problems against small perturbations to the collected measurement data. Figure 11 depicts the sensitivity of the optimal estimate of initial condition with respect to the



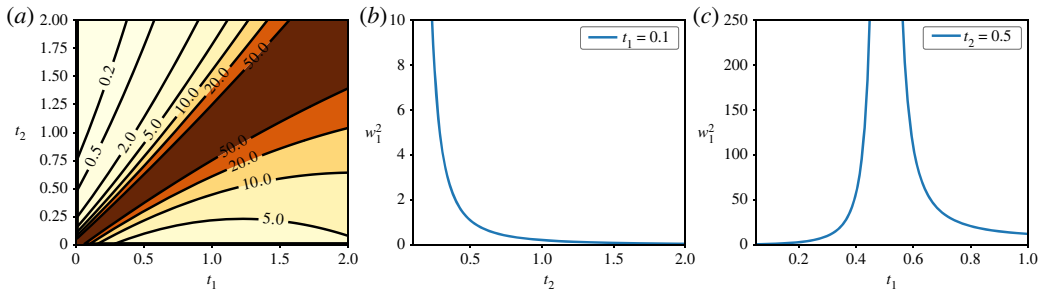
**Figure 10.** Cost function for nonlinear decay model:  $\dot{x} = \alpha x^2$ .



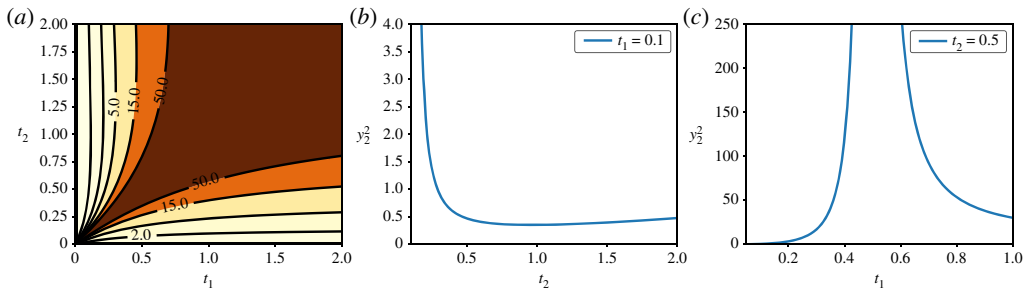
**Figure 11.** Sensitivity of the optimal estimates of the initial condition  $x_0$  to the first observation  $z_1$  for the nonlinear system  $\dot{x} = \alpha x^2$ . (a) Contour plot of  $y_1^2(t)$  at different values of  $t_1$  and  $t_2$ . (b)  $y_1^2(t)$  values at  $t_1 = 0.1$  and different values of  $t_2$ . (c)  $y_1^2(t)$  values at  $t_2 = 0.5$  and different values of  $t_1$ .

first observation, considering a total of two observations at  $t_1$  and  $t_2$ . From the contour plots of  $y_1^2$ , we see that collecting the two observations at close time instants results in higher sensitivity of the estimated initial condition to the measurement itself. Similar observations are obtained from figure 12 for  $w_1^2$ , where  $w_1$  denotes the sensitivity of the optimal estimate of the model parameter with respect to the first observation. This is in agreement with the fact that highly correlated measurements do not add valuable information to the inversion problem. Moreover, from equation (4.18), when  $t_1$  and  $t_2$  coincide, the determinant of the observability Gramian vanishes, resulting in extreme values of the sensitivities  $y_i$  and  $w_i$  for  $i = 1, 2$ .

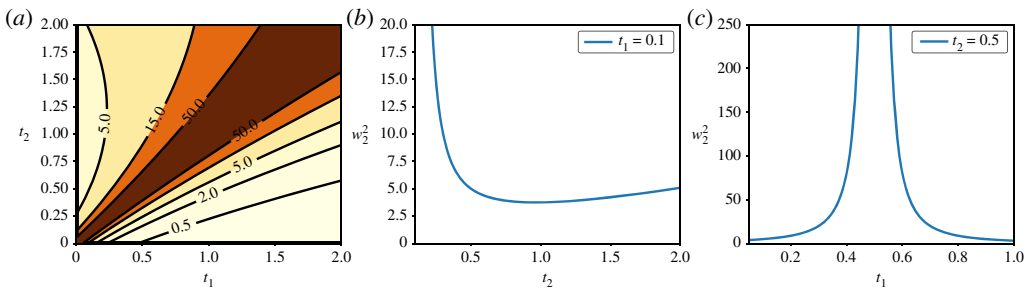
Figures 13 and 14 display the sensitivity of the inferred initial condition and model parameter, respectively, to the second observation value at  $t_2$ . Similar to  $y_1^2$  and  $w_1^2$  plots, we see that the maximum values of  $y_2^2$  and  $w_2^2$  occur when  $t_1$  and  $t_2$  are close to each other. In addition, looking at  $y_2^2$  and  $w_2^2$  values at  $t_1 = 0.1$ , it is clear that  $y_2^2$  and  $w_2^2$  exhibit very large values at small values of  $t_2$ , then decrease to attain their minima around  $t_2 = 0.5$  before slightly increasing again at larger values of  $t_2$ . In a similar fashion, through an investigation of  $y_2^2$  and  $w_2^2$  at  $t_2 = 0.5$  with varying  $t_1$ , we observe that maximum values occur when  $t_1 = t_2 = 0.5$ . On the other hand, minimum values of the sensitivity of the control estimate with respect to the second measurement  $z_2$  at  $t_2 = 0.5$  occur when the first measurement is collected earlier, at small values of  $t_1$ . The numerical investigations



**Figure 12.** Sensitivity of the optimal estimates of the decay parameter  $a$  to the first observation  $z_1$  for the nonlinear system  $\dot{x} = \alpha x^2$ . (a) Contour plot of  $w_1^2(t)$  at different values of  $t_1$  and  $t_2$ . (b)  $y_1^2(t)$  values at  $t_1 = 0.1$  and different values of  $t_2$ . (c)  $w_1^2(t)$  values at  $t_2 = 0.5$  and different values of  $t_1$ .

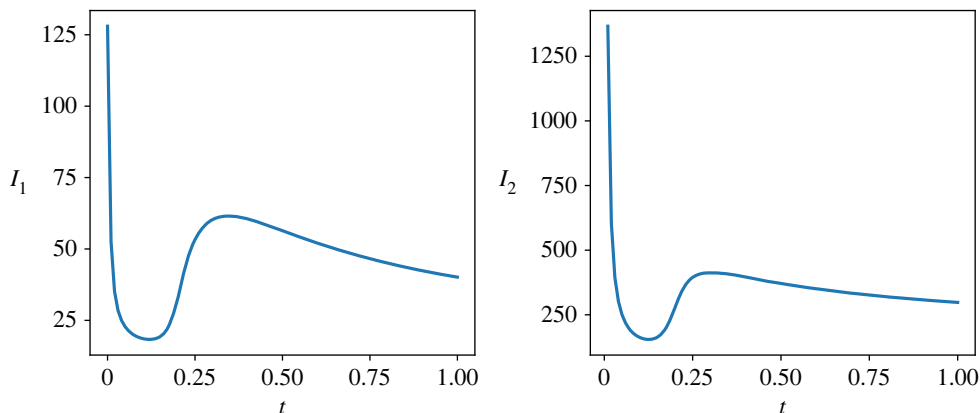


**Figure 13.** Sensitivity of the optimal estimates of the initial condition  $x_0$  to the second observation  $z_2$  for the nonlinear system  $\dot{x} = \alpha x^2$ . (a) Contour plot of  $y_2^2(t)$  at different values of  $t_1$  and  $t_2$ . (b)  $y_2^2(t)$  values at  $t_1 = 0.1$  and different values of  $t_2$ . (c)  $y_1^2(t)$  values at  $t_2 = 0.5$  and different values of  $t_1$ .



**Figure 14.** Sensitivity of the optimal estimates of the decay parameter  $a$  to the second observation  $z_2$  for the nonlinear system  $\dot{x} = \alpha x^2$ . (a) Contour plot of  $w_2^2(t)$  at different values of  $t_1$  and  $t_2$ . (b)  $y_1^2(t)$  values at  $t_1 = 0.1$  and different values of  $t_2$ . (c)  $w_1^2(t)$  values at  $t_2 = 0.5$  and different values of  $t_1$ .

of the sensitivity of the inverse problem solution to the observation values and locations confirm that they are minimized when the observations are placed using the proposed forward sensitivity analysis. By placing the observations at points where the squares of the forward sensitivities of the forecast are maximized, we (i) control the shape of the cost function in such a way that flat patches (with zero gradients) are avoided, which improves the convergence characteristics of solving the inverse problem, and (ii) guarantee that the sensitivities of the estimated control values (solution to the inverse problem) with respect to the measurement values are minimized, which enhances the robustness of the inversion framework against possible perturbations in the collected sensor data (figures 13 and 14).



**Figure 15.** The first two invariants for the model forecast sensitivities with respect to the initial condition for the one-dimensional viscous Burgers problem.

### (c) One-dimensional Burgers problem

Our third test case is an advection shock problem governed by the one-dimensional viscous Burgers equation as follows:

$$\frac{\partial q}{\partial t} + q \frac{\partial q}{\partial x} = \frac{1}{\text{Re}} \frac{\partial^2 q}{\partial x^2}, \quad (5.6)$$

where  $q(x, t)$  is the velocity field and  $\text{Re}$  is the dimensionless Reynolds number, defined as the ratio of inertial effects to viscous effects. We note that  $x$  here denotes the spatial variable/coordinate while the discretization of  $q$  represents the state vector. We perform spatial discretization by defining the velocity field at  $n$  discrete locations, equally spaced in the domain  $L$ . We apply a second-order centred finite difference scheme for the linear term and use the skew-symmetric formulation from Aref & Daripa's scheme [54] for the nonlinear term as follows:

$$\frac{\partial q_i}{\partial t} = -\frac{1}{3} \frac{(q_{i+1} + q_{i-1} + q_i)(q_{i+1} - q_{i-1})}{2\Delta x} + \frac{1}{\text{Re}} \frac{q_{i+1} - 2q_i + q_{i-1}}{\Delta x^2}, \quad (5.7)$$

where  $\Delta x$  is the grid spacing,  $\Delta x = x_{i+1} - x_i$ . By arranging the velocity field in a column vector, equation (5.7) is equivalent to equation (4.19). We define a domain of length  $L = 1$  and enforce zero Dirichlet boundary conditions,  $q(0, t) = q(1, t) = 0$ . A twin experiment is employed to generate observational data and assess the quality of the solution to inverse problems. In particular, the ground truth corresponds to  $\text{Re} = 500$  and the following initial condition [55]:

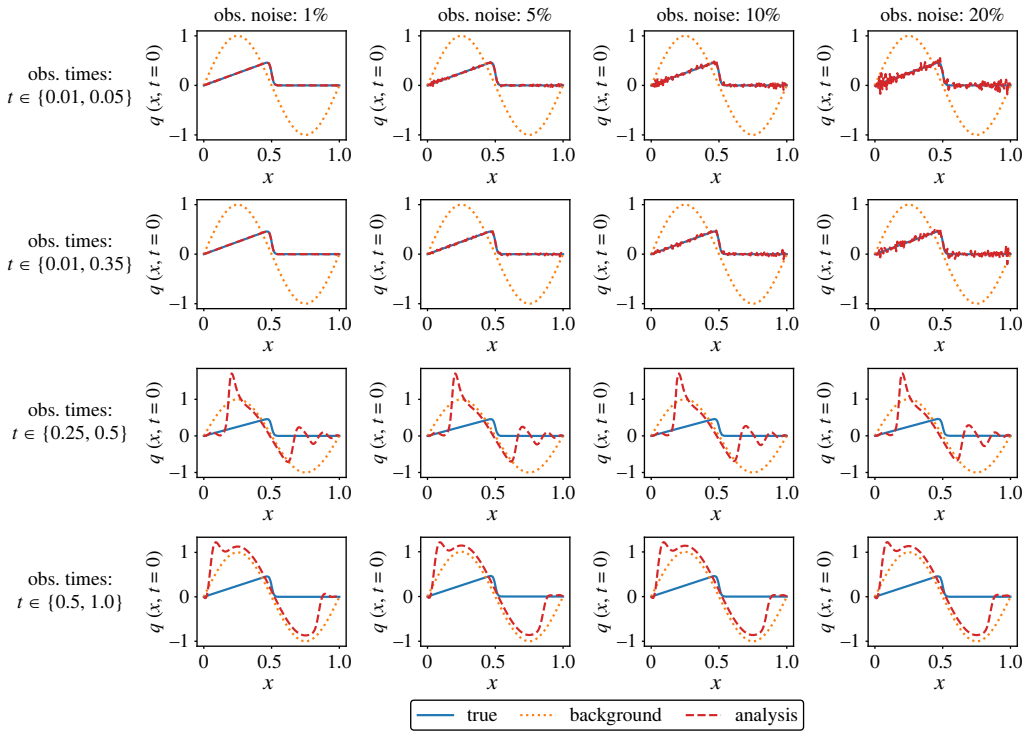
$$q(x, 0) = \frac{x}{1 + \exp((\text{Re}/16)(4x^2 - 1))}, \quad (5.8)$$

while the background solution corresponds to a sinusoidal wave as  $q(x, 0) = \sin(2\pi x/L)$ . Because it is unfeasible to track individual components of the forecast sensitivity matrices, we focus on the following two quantities related to the sensitivity matrix  $\mathbf{u}$ :

$$I_1 = \text{trace}(\mathbf{u}^T \mathbf{u}) \quad \text{and} \quad I_2 = \frac{1}{2} \left( \left( \text{trace}(\mathbf{u}^T \mathbf{u}) \right)^2 - \text{trace}(\mathbf{u}^T \mathbf{u}^2) \right). \quad (5.9)$$

Figure 15 shows the time variation of  $I_1$  and  $I_2$  for the one-dimensional Burgers problem. Both quantities attain their largest values near the initial time along with an additional *bump* around  $t = 0.35$ .

We empirically explore the advantages of the FSM-based observation placement strategy by considering different scenarios. The first two rows in figure 16 correspond to placing the observations by tracing the model forecast sensitivities. In particular, our first experiment corresponds to placing the observations near the initial time at  $t = 0.01$  and  $t = 0.05$ . The second



**Figure 16.** A comparison between inverse problem solutions (for initial conditions) at different observation placement times and varying levels of noise for the one-dimensional viscous Burgers problem. Background refers to the guessed (inaccurate) initial conditions while the analysis refers to the solution of the inverse problem after observations have been assimilated.

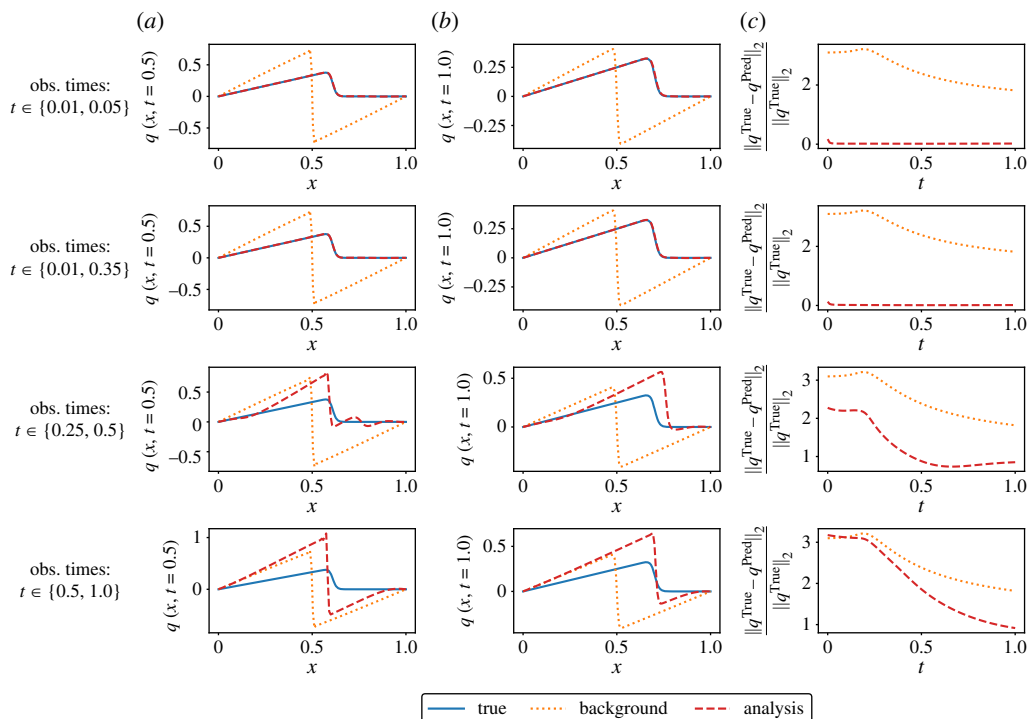
experiment corresponds to placing one observation at  $t = 0.01$  and another at  $t = 0.35$  near the second peak in figure 15. We test the quality of the inverse problem solution with different levels of measurement noise. Figure 16 demonstrates that the FSM placement strategy enables the accurate inference of the unknown initial condition even with large levels of noise. In this figure and the remaining discussion, we use standard terminology from data assimilation studies, where ‘background’ refers to the estimated guess of unknown initial conditions and/or parameters. This estimate could be based on previous model predictions, historical data or an intelligent guess. Thus, the background solution is the model forecast based on these supposedly inaccurate estimates. On the other hand, ‘analysis’ refers to the solution of the inverse problem (e.g. for unknown initial conditions and/or parameters) after assimilating observation data. We highlight that we apply a truncated singular value decomposition as a regularization technique for solving the inverse problem. Our third experiment corresponds to arbitrarily placing the observations at  $t = 0.25$  and  $t = 0.50$  while we place them at  $t = 0.50$  and  $t = 1.0$  in the fourth experiment. We can see that both cases result in an inaccurate solution of the inverse problem. It is worth noting that this inaccuracy is in part due to the off-target background information (the sinusoidal wave). However, this exaggeration is intentional to show the benefits of the proposed placement strategy even in challenging situations. Finally, the predicted velocity fields are depicted in figure 17, where placing the observations closer to the maximum values of the forecast sensitivities yields improved results.

#### (d) Two-dimensional advection diffusion problem

Our last test case is the two-dimensional advection diffusion equation as follows:

$$\frac{\partial q}{\partial t} + c_x \frac{\partial q}{\partial x} + c_y \frac{\partial q}{\partial y} = \nu \left( \frac{\partial^2 q}{\partial x^2} + \frac{\partial^2 q}{\partial y^2} \right). \quad (5.10)$$





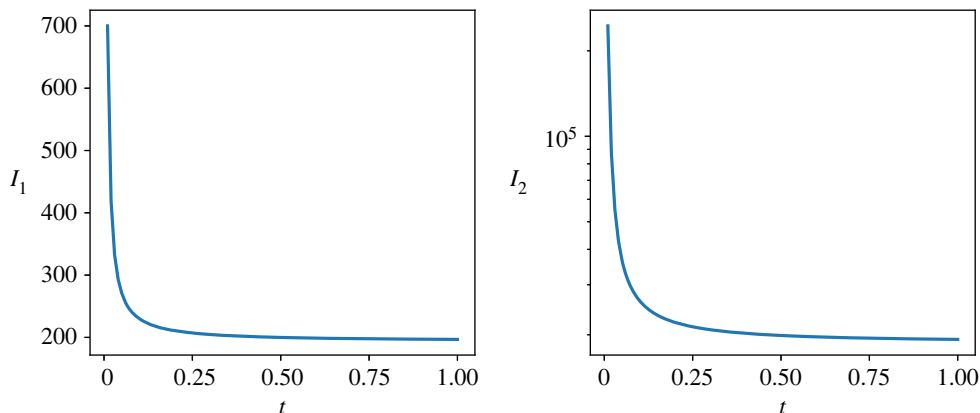
**Figure 17.** The predicted velocity field at  $t = 0.5$  (a) and  $t = 1.0$  (b) for the one-dimensional viscous Burgers problem with different observation placement times with 10% measurement noise along with the relative  $\ell_2$  error of the predicted velocity field at different times (c). Background solution refers to the forecast using the guessed (inaccurate) initial conditions while the analysis refers to the forecast after observations have been assimilated.

Equation (5.10) can be used to describe the evolution of the concentration of a substance under convection and diffusion effects, and the inversion of unknown initial conditions can be particularly important in the context of (contaminant) leak detection. Thus, variants of this test case have been used extensively in the literature of data assimilation and optimal experimental designs [56–58]. We use  $c_x = c_y = 0.5$  and  $\nu = 0.01$  and consider an initial condition of Gaussian distribution centred at  $(x_0, y_0)$  as  $q(x, y, 0) = e^{-(1/\nu)((x-x_0)^2 + (y-y_0)^2)}$ . The true field corresponds to  $(x_0, y_0) = (0.25, 0.25)$  while the background (guessed) solution corresponds to  $(x_0, y_0) = (0.5, 0.5)$ .

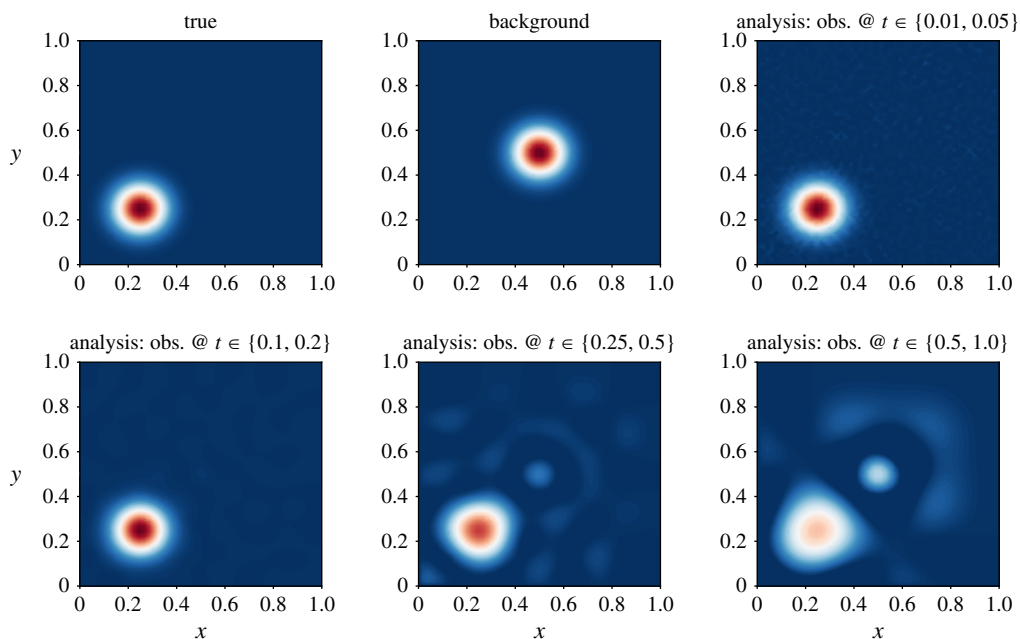
Figure 18 illustrates that the model forecast sensitivity with respect to the initial conditions exhibits maximum values near the initial time. In figure 19, we repeat the solution of the inverse problems while varying the times at which we place our observations. While it might not be possible in practice to place all observations near the initial time, figure 18 gives theoretical guidelines to prioritize observations at specific time instants over others. For instance, it is evident from figure 19 that placing observations at  $t = 0.5$  and  $t = 1.0$  (corresponding to minimum values of forecast sensitivities) produces erroneous initial conditions. On the other hand, collecting data at  $t = 0.1$  and  $t = 0.2$  (where the forward sensitivities are relatively larger) results in a better inference of the unknown initial condition. Finally, the predicted solution at  $t = 1$  is shown in figure 20 where observations are placed at  $t = 0.01$  and  $t = 0.05$ .

## 6. Conclusion

This paper systematically introduces a unified notion in the analysis and application of FSM to address various questions related to forecast error correction using dynamic data assimilation [53]. We propose a strategy based on maximizing the square of the forward sensitivities, which guarantees efficient recovery of control variables and minimizes the sensitivity of optimal control



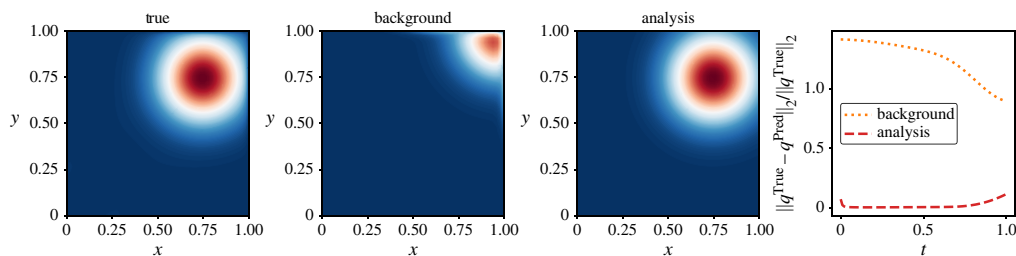
**Figure 18.** The first two invariants for the model forecast sensitivities with respect to initial conditions for the two-dimensional advection diffusion problem.



**Figure 19.** A comparison between inverse problem solutions (for initial conditions) at different observation placement times with 10% measurement noise for the two-dimensional advection diffusion problem. Background refers to the guessed (inaccurate) initial conditions while the analysis refers to the solution of the inverse problem after observations have been assimilated.

estimates with respect to observations. First and foremost, the contribution of this paper can be better understood when compared with our ongoing efforts dedicated to the development of the FSM framework. In Lakshmivarahan & Lewis [52], we first proved the equivalence between FSM and the four-dimensional VAR method. In Lakshmivarahan *et al.* [24,25], by relating the observability Gramian to the forward sensitivities, we derived an intrinsic expression for the adjoint sensitivity in terms of the observability Gramian. By exploiting the structure of the observability Gramian, we then derived a strategy to place observations where the squares of the forward sensitivities attain maximum values.

In this paper, we provide a detailed analysis of this observation placement strategy and show that it provides two crucial advantages in the context of inverse problems. First, we demonstrate



**Figure 20.** The predicted solution at  $t = 1$  for the two-dimensional advection diffusion problem following the proposed observation placement strategy (observations are placed at  $t = 0.01$  and  $t = 0.05$ ) with 10% measurement noise along with the relative  $\ell_2$  error of the predicted velocity field at different times (right). Background solution refers to the forecast using the guessed (inaccurate) initial conditions while the analysis refers to the forecast after observations have been assimilated.

that the FSM observation placement strategy avoids the occurrence of flat patches in the cost functional and improves the effectiveness of the minimization algorithm. Second, we also prove that it minimizes the sensitivity of the estimates (of the inverted quantities) with respect to the observations. The smallness of these quantitative measurements of the sensitivity of the estimates to observations supports determining observation sites using forward sensitivities. An advantage of the FSM analysis is that it does not require the adjoint code. However, computing the forward sensitivities is computationally demanding, but once done, we can solve two related problems. First, we can place the observations where the forward sensitivity attains a maximum value. Second, we can compute the optimal estimates of the control using FSM.

**Data accessibility.** There is no additional data.

**Authors' contributions.** S.E.A.: formal analysis, software, writing—review and editing; O.S.: formal analysis, supervision, writing—review and editing; S.L.: conceptualization, formal analysis, methodology, writing—original draft; J.M.L.: conceptualization, methodology.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

**Conflict of interest declaration.** We declare we have no competing interests.

**Funding.** O.S. was supported by the U.S. Department of Energy (DOE), Office of Science, Advanced Scientific Computing Research (ASCR) program under Award Number DE-SC0019290. The work of S.E.A is supported by the DOE ASCR program through the Pacific Northwest National Laboratory Distinguished Fellowship in Scientific Computing (Project no. 71268). Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for DOE under Contract DE-AC05-76RL01830.

**Acknowledgements.** We are grateful to the reviewers for their careful reading and suggestions that improved the readability of the paper.

## Appendix A. Proof of the second advantage of the FSM observation placement strategy: Scalar case

In this appendix, we derive an intrinsic relation that exists between the sensitivities of the estimates of the control with respect to the observations, namely  $y_j$  and  $w_j$  for  $1 \leq j \leq 2$ , and the forward sensitivities  $u(k)$  and  $v(k)$ .

— Step 1: Derivatives of  $x(t)$ ,  $u(t)$ ,  $v(t)$  and  $D_h(x(t))$  with respect to the observation  $z_j$  for  $1 \leq j \leq 2$  are given by

$$\begin{aligned} \frac{\partial x(t)}{\partial z_j} &= u(t)y_j + v(t)w_j, & \frac{\partial u(t)}{\partial z_j} &= \frac{\partial u(t)}{\partial x_0}y_j + \frac{\partial u(t)}{\partial \alpha}w_j, & \frac{\partial v(t)}{\partial z_j} &= \frac{\partial v(t)}{\partial x_0}y_j + \frac{\partial v(t)}{\partial \alpha}w_j, \\ \frac{\partial D_h(x(t))}{\partial z_j} &= D_h^{(2)}(x(t))\frac{\partial x(t)}{\partial z_j}, & \text{where } D_h^{(2)}(x) &= \frac{\partial^2 h}{\partial x^2} \neq \left(\frac{\partial h}{\partial x}\right)^2. \end{aligned} \quad (\text{A } 1)$$

— Step 2: From the definition of  $e(t)$  in equation (3.3), we get

$$\frac{\partial e(t)}{\partial z_j} = \delta_{tj} - D_h(x(t)) \frac{\partial x(t)}{\partial z_j}, \quad (\text{A } 2)$$

where  $\delta_{tj}$  is the Kronecker delta (i.e.  $\delta_{tj} = 1$  if  $t = j$  and  $\delta_{tj} = 0$  otherwise).

— Step 3: Differentiating  $H_{1j}$  in equation (4.12) with respect to  $z_p$ , we get the following:

$$(\delta_{jp} - D_h(x(j)) \frac{\partial x(j)}{\partial z_p}) D_h(x(j)) u(j) + e(j) \frac{\partial D_h(x(j))}{\partial z_p} u(j) + e(j) D_h(x(j)) \frac{\partial u(j)}{\partial z_p} = 0. \quad (\text{A } 3)$$

— Step 4: Differentiating  $H_{2j}$  in equation (4.13) with respect to  $z_p$ , we get the following:

$$(\delta_{jp} - D_h(x(j)) \frac{\partial x(j)}{\partial z_p}) D_h(x(j)) v(j) + e(j) \frac{\partial D_h(x(j))}{\partial z_p} v(j) + e(j) D_h(x(j)) \frac{\partial v(j)}{\partial z_p} = 0. \quad (\text{A } 4)$$

— Step 5: When  $x(t)$  is optimal, a little reflection reveals that  $e(k)$  along the optimal path is small, and henceforth it is set to zero in equations (A 3) and (A 4). Thus, the approximate expressions for  $\partial H_{ij}/\partial z_p$  can be given as listed in table 2.

— Step 6: The optimality conditions (i.e.  $\partial H_i/\partial z_p = \partial H_{i1}/\partial z_p + \partial H_{i2}/\partial z_p$ ) reduce to the following:

$$D_h(1)u(1) = D_h^2(1) \frac{\partial x(1)}{\partial z_1} u(1) + D_h^2(2) \frac{\partial x(2)}{\partial z_1} u(2), \quad (\text{A } 5)$$

$$D_h(2)u(2) = D_h^2(1) \frac{\partial x(1)}{\partial z_2} u(1) + D_h^2(2) \frac{\partial x(2)}{\partial z_2} u(2), \quad (\text{A } 6)$$

$$D_h(1)v(1) = D_h^2(1) \frac{\partial x(1)}{\partial z_1} v(1) + D_h^2(2) \frac{\partial x(2)}{\partial z_1} v(2) \quad (\text{A } 7)$$

$$\text{and} \quad D_h(2)v(2) = D_h^2(1) \frac{\partial x(1)}{\partial z_2} v(1) + D_h^2(2) \frac{\partial x(2)}{\partial z_2} v(2). \quad (\text{A } 8)$$

— Step 7: Substituting for  $\partial x(t)/\partial z_j$  using equation (A 1) in equations (A 5) and (A 7), we get

$$D_h(1)u(1) = D_h^2(1)(u^2(1) + u(1)v(1))y(1) + D_h^2(2)(u^2(2) + u(2)v(2))w(1) \quad (\text{A } 9)$$

and

$$D_h(1)v(1) = D_h^2(1)(u(1)v(1) + v^2(1))y(1) + D_h^2(2)(u(2)v(2) + v^2(2))w(1). \quad (\text{A } 10)$$

Referring to equations (3.5) and (3.9), we define  $F(i) = [u(i), v(i)]^T \in \mathbb{R}^{2 \times 1}$  and  $G(i) = F^T(i)D_h^2(i)F(i) \in \mathbb{R}^{2 \times 2}$ . Thus, equations (A 9) and (A 10) can be rewritten as

$$D_h(1) \begin{bmatrix} u(1) \\ v(1) \end{bmatrix} = G \begin{bmatrix} y_1 \\ w_1 \end{bmatrix}, \quad (\text{A } 11)$$

where  $G = G(1) + G(2)$ . Therefore,  $y_1$  and  $w_1$  can be obtained as

$$\begin{bmatrix} y_1 \\ w_1 \end{bmatrix} = G^{-1} D_h(1) \begin{bmatrix} u(1) \\ v(1) \end{bmatrix}. \quad (\text{A } 12)$$

Similarly, from equations (A 6) and (A 8) and using the same procedure, we get

$$\begin{bmatrix} y_2 \\ w_2 \end{bmatrix} = G^{-1} D_h(2) \begin{bmatrix} u(2) \\ v(2) \end{bmatrix}. \quad (\text{A } 13)$$

**Table 2.** Expressions for  $\partial H_{ij}/\partial z_p$  for  $i, j, p \in \{1, 2\}$ .

		$j = 1$	$j = 2$
$i = 1$	$p = 1$	$\left(1 - D_h(1) \frac{\partial x(1)}{\partial z_1}\right) D_h(1) u(1)$	$-D_h^2(2) \frac{\partial x(2)}{\partial z_1} u(2)$
	$p = 2$	$-D_h^2(1) \frac{\partial x(1)}{\partial z_2} u(1)$	$\left(1 - D_h(2) \frac{\partial x(2)}{\partial z_2}\right) D_h(2) u(2)$
$i = 2$	$p = 1$	$\left(1 - D_h(1) \frac{\partial x(1)}{\partial z_1}\right) D_h(1) v(1)$	$-D_h^2(2) \frac{\partial x(2)}{\partial z_1} v(2)$
	$p = 2$	$-D_h^2(1) \frac{\partial x(1)}{\partial z_2} v(1)$	$\left(1 - D_h(2) \frac{\partial x(2)}{\partial z_2}\right) D_h(2) v(2)$

## Appendix B. Linear independence of the forward sensitivities of the model forecast at different times with respect to the control

The discrete-time version of the model given in equation (2.1) can be written as

$$x_{k+1} = M(x_k, \alpha), \quad (\text{B } 1)$$

where  $M: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  defines the one-time step mapping defined by applying a temporal integration scheme. Equation (2.3) for the dynamics of  $u(t)$  and  $v(t)$  can be rewritten as follows:

$$\left. \begin{aligned} u_{k+1} &= D_M(k) u_k, & u_0 &= 1, \\ v_{k+1} &= D_M(k) v_k + D_M^\alpha(x_k), & v_0 &= 0, \end{aligned} \right\} \quad (\text{B } 2)$$

where

$$D_M(k) = \frac{\partial M}{\partial x} \Big|_{x=x_k} \quad \text{and} \quad D_M^\alpha(k) = \frac{\partial M}{\partial \alpha} \Big|_{x=x_k}. \quad (\text{B } 3)$$

For simplicity of notation, let  $a_k := D_M(k)$  and  $b_k := D_M^\alpha(k)$ . Thus, the sequence of  $u_k$  can be written as follows:

$$u_k = a_{k-1} a_{k-2} \dots a_1 a_0, \quad \text{for } k \geq 1. \quad (\text{B } 4)$$

Similarly, the sequence of  $v_k$  can be expanded as follows:

$$\begin{aligned} v_k &= (a_{k-1} a_{k-2} \dots a_1 b_0) + (a_{k-1} a_{k-2} \dots a_2 b_1) + (a_{k-1} a_{k-2} \dots a_3 b_2) + \dots \\ &\quad + (a_{k-1} a_{k-2} b_{k-3}) + (a_{k-1} b_{k-2}) + (b_{k-1}), \quad \text{for } k \geq 1. \end{aligned} \quad (\text{B } 5)$$

For the Gramian matrix  $G$  in equation (4.17) to be singular (i.e. the determinant in equation (4.18) equals zero), the vectors  $[u_1, u_2]^\top$  and  $[v_1, v_2]^\top$  ought to be linearly dependent. In what follows, we show that this situation cannot happen. For the vectors  $[u_1, u_2]^\top$  and  $[v_1, v_2]^\top$  to be linearly dependent, we get  $u_k = C v_k$  for all  $k$  where  $C$  is a non-zero constant, then we can write the following:

$$u_1 = C v_1 \iff a_0 = C b_0 \quad (\text{B } 6)$$

and

$$u_2 = C v_2 \iff a_1 a_0 = C a_1 b_0 + C b_1. \quad (\text{B } 7)$$

Because  $a_0 = C b_0$ , equation (B 7) can be rewritten as  $a_1 a_0 = a_1 a_0 + C b_1$ , leading to  $b_1 = 0$ . Similarly,

$$u_3 = C v_3 \iff a_2 a_1 a_0 = C a_2 a_1 b_0 + C a_2 b_1 + C b_2. \quad (\text{B } 8)$$

With  $a_0 = C b_0$  and  $b_1 = 0$ , we also get  $b_2 = 0$ . Following the same procedure, this leads to  $b_k = D_M^\alpha(k) = 0$  for all  $k \neq 0$ . In other words, this implies that the sensitivity of the model  $M$  with respect to the parameter  $\alpha$  equals zero along the trajectory except at  $t = 0$ . However, this cannot be the

case because a model cannot be sensitive to its parameters at  $t = 0$  and not sensitive at other times. Therefore, the vectors of  $u_k$  and  $v_k$  cannot be linearly dependent, and the proposed observation placement strategy thus does not lead to a singular observability Gramian.

## References

1. Tarantola A. 1987 *Inverse problem theory*. New York, NY: Elsevier.
2. Kalnay E. 2003 *Atmospheric modeling, data assimilation and predictability*. Cambridge, UK: Cambridge University Press.
3. Kaipio J, Somersalo E. 2006 *Statistical and computational inverse problems*, vol. 160. Berlin, Germany: Springer.
4. Lewis JM, Lakshmivarahan S, Dhall S. 2006 *Dynamic data assimilation: a least squares approach*, vol. 104. Cambridge, UK: Cambridge University Press.
5. Biros G *et al.* 2011 *Large-scale inverse problems and quantification of uncertainty*. Chichester, UK: John Wiley & Sons.
6. Edwards CA, Moore AM, Hoteit I, Cornuelle BD. 2015 Regional ocean data assimilation. *Annu. Rev. Mar. Sci.* **7**, 21–42. (doi:10.1146/annurev-marine-010814-015821)
7. Asch M, Bocquet M, Nodet M. 2016 *Data assimilation: methods, algorithms, and applications*. Philadelphia, PA: SIAM.
8. Vieli A, Payne AJ, Du Z, Shepherd A. 2006 Numerical modelling and data assimilation of the Larsen B ice shelf, Antarctic Peninsula. *Phil. Trans. R. Soc. A* **364**, 1815–1839. (doi:10.1098/rsta.2006.1800)
9. Houtekamer PL, Zhang F. 2016 Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **144**, 4489–4532. (doi:10.1175/MWR-D-15-0440.1)
10. Bannister R. 2017 A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. R. Meteorol. Soc.* **143**, 607–633. (doi:10.1002/qj.2982)
11. Carrassi A, Bocquet M, Bertino L, Evensen G. 2018 Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *Wiley Interdiscip. Rev.: Clim. Change* **9**, e535. (doi:10.1002/wcc.535)
12. Geer A. 2021 Learning earth system models from observations: machine learning or data assimilation?. *Phil. Trans. R. Soc. A* **379**, 20200089. (doi:10.1098/rsta.2020.0089)
13. Brajard J, Carrassi A, Bocquet M, Bertino L. 2021 Combining data assimilation and machine learning to infer unresolved scale parametrization. *Phil. Trans. R. Soc. A* **379**, 20200086. (doi:10.1098/rsta.2020.0086)
14. Kalman RE. 1960 On the general theory of control systems. In *Proc. First Int. Conf. on Automatic Control Congress, Moscow, USSR*, pp. 491–502.
15. Casti JL. 1985 *Nonlinear systems theory*. New York, NY: Academic Press.
16. Isidori A. 1985 *Nonlinear control systems*. Berlin, Germany: Springer.
17. Bellman R. 1953 *Stability of differential equations*. New York, NY: McGraw Hill.
18. Brown R. 1966 Not just observable, but how observable. In *Proc. of the 1966 National Electronics Conf., Volume 22, Chicago, IL*, pp. 709–714.
19. Kang W, Xu L. 2009 A quantitative measure of observability and controllability. In *Proc. of the 48th IEEE Conf. on Decision and Control, Shanghai, China*, pp. 6413–6418.
20. Kang W, Xu L. 2012 Optimal placement of mobile sensors for data assimilations. *Tellus A: Dyn. Meteorol. Oceanogr.* **64**, 17133. (doi:10.3402/tellusa.v64i0.17133)
21. King S, Kang W, Xu L. 2015 Observability for optimal sensor locations in data assimilation. *Int. J. Dyn. Control* **3**, 416–424. (doi:10.1007/s40435-014-0120-7)
22. Le Dimet FX, Talagrand O. 1986 Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A: Dyn. Meteorol. Oceanogr.* **38**, 97–110. (doi:10.3402/tellusa.v38i2.11706)
23. Ochotta T, Gebhardt C, Saupe D, Wergen W. 2005 Adaptive thinning of atmospheric observations in data assimilation with vector quantization and filtering methods. *Q. J. R. Meteorol. Soc.* **131**, 3427–3437. (doi:10.1256/qj.05.94)
24. Lakshmivarahan S, Lewis JM, Hu J. 2020 On controlling the shape of the cost functional in dynamic data assimilation: guidelines for placement of observations and application to Saltzman's model of convection. *J. Atmos. Sci.* **77**, 2969–2989.
25. Lakshmivarahan S, Lewis JM, Maryada SKR. 2022 Observability Gramian and its role in the placement of observations in dynamic data assimilation. In *Data assimilation for atmospheric*,



- oceanic and hydrologic applications*, vol. IV (eds SK Park, L Xu), pp. 215–257. Cham, Switzerland: Springer.
26. Lewis JM, Lakshmivarahan S, Hu J, Rabin R. 2020 Placement of observations to correct return-flow forecasts. *Electron. J. Sev. Storms Meteorol.* **15**, 1–20. (doi:10.55599/ejssm.v15i4.78)
  27. Lewis JM, Lakshmivarahan S, Maryada SKR. 2022 Placement of observations for variational data assimilation: application to Burgers' equation and Seiche phenomenon. In *Data assimilation for atmospheric, oceanic and hydrologic applications*, vol. IV (eds SK Park, L Xu), pp. 259–275. Cham, Switzerland: Springer.
  28. Ahmed SE, Bhar K, San O, Rasheed A. 2020 Forward sensitivity approach for estimating eddy viscosity closures in nonlinear model reduction. *Phys. Rev. E* **102**, 043302. (doi:10.1103/PhysRevE.102.043302)
  29. Ahmed SE, San O, Lakshmivarahan S. 2022 Forward sensitivity analysis of the FitzHugh–Nagumo system: parameter estimation. In *Advances in nonlinear dynamics* (eds W Lacarbonara, B Balachandran, MJ Leamy, J Ma, JA Tenreiro Machado, G Sepan), pp. 93–103. Cham, Switzerland: Springer.
  30. Fuller WA. 1976 *Introduction to statistical time series*. New York, NY: Wiley.
  31. Brockwell PJ, Davis RA. 2006 *Time series: theory and methods*, 2nd edn. New York, NY: Springer.
  32. Seber GA, Wild CJ. 1988 *Nonlinear regression*. New York, NY: Wiley Intersciences.
  33. Gallant AR. 1987 *Nonlinear statistical models*. New York, NY: Wiley.
  34. Bellman R, Åström KJ. 1970 On structural identifiability. *Math. Biosci.* **7**, 329–339. (doi:10.1016/0025-5564(70)90132-X)
  35. Ljung L. 1999 *System identification*. Englewood Cliff, NJ: Prentice Hall.
  36. Narendra KS, Annaswamy AM. 1989 *Stable adaptive systems*. Englewood Cliff, NJ: Prentice Hall.
  37. Villaverde AF. 2019 Observability and structural identifiability of nonlinear biological systems. *Complexity* **2019**, 8497093. (doi:10.1155/2019/8497093)
  38. Uciniski D. 2004 *Optimal measurement methods for distributed parameter system identification*. Baton Rouge, LA: CRC Press.
  39. Alaña JE. 2010 Optimal measurement locations for parameter estimation of non linear distributed parameter systems. *Braz. J. Chem. Eng.* **27**, 627–642. (doi:10.1590/S0104-66322010000400015)
  40. Christopher TH, Fathalla AR. 1999 Sensitivity analysis of parameters in modelling with delay-differential equations. *Manchester Center for Computational Mathematics, Numerical Analysis Reports, Report no. 349, The University of Manchester, UK*, pp. 180–199.
  41. Cacuci DG. 2003 *Sensitivity & uncertainty analysis, volume I: theory*, vol. 1. Boca Raton, FL: CRC Press.
  42. Cacuci DG, Ionescu-Bujor M, Navon IM. 2005 *Sensitivity and uncertainty analysis, volume II: applications to large-scale systems*, vol. 2. Boca Raton, FL: CRC Press.
  43. Baker NL, Daley R. 2000 Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Q. J. R. Meteorol. Soc.* **126**, 1431–1454. (doi:10.1002/qj.49712656511)
  44. Langland RH, Baker NL. 2004 Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus A: Dyn. Meteorol. Oceanogr.* **56**, 189–201. (doi:10.3402/tellusa.v56i3.14413)
  45. Daescu DN. 2008 On the sensitivity equations of four-dimensional variational (4D-Var) data assimilation. *Mon. Weather Rev.* **136**, 3050–3065. (doi:10.1175/2007MWR2382.1)
  46. Cardinali C. 2009 Monitoring the observation impact on the short-range forecast. *Q. J. R. Meteorol. Soc.* **135**, 239–250. (doi:10.1002/qj.366)
  47. Gelaro R, Zhu Y. 2009 Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus A: Dyn. Meteorol. Oceanogr.* **61**, 179–193. (doi:10.1111/j.1600-0870.2008.00388.x)
  48. Majumdar SJ. 2016 A review of targeted observations. *Bull. Am. Meteorol. Soc.* **97**, 2287–2303. (doi:10.1175/BAMS-D-14-00259.1)
  49. Ancell B, Hakim GJ. 2007 Comparing adjoint-and ensemble-sensitivity analysis with applications to observation targeting. *Mon. Weather Rev.* **135**, 4117–4134. (doi:10.1175/2007MWR1904.1)
  50. Liu J, Kalnay E. 2008 Estimating observation impact without adjoint model in an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.* **134**, 1327–1335. (doi:10.1002/qj.280)

51. Arnold VI. 1992 *Ordinary differential equations*. New York, NY: Springer Science & Business Media.
52. Lakshmivarahan S, Lewis JM. 2010 Forward sensitivity approach to dynamic data assimilation. *Adv. Meteorol.* **2010**, 375615. (doi:10.1155/2010/375615)
53. Lakshmivarahan S, Lewis JM, Jabrzemski R. 2017 *Forecast error correction using dynamic data assimilation*. New York, NY: Springer.
54. Aref H, Daripa P. 1984 Note on finite difference approximations to Burgers' equation. *SIAM J. Sci. Stat. Comput.* **5**, 856–864. (doi:10.1137/0905060)
55. Ahmed SE, San O, Rasheed A, Iliescu T. 2020 A long short-term memory embedding for hybrid uplifted reduced order models. *Physica D* **409**, 132471. (doi:10.1016/j.physd.2020.132471)
56. Petra N, Stadler G. 2011 Model variational inverse problems governed by partial differential equations. Technical report Texas University at Austin Institute for Computational Engineering and Sciences.
57. Attia A, Leyffer S, Munson TS. 2022 Stochastic learning approach for binary optimization: application to Bayesian optimal design of experiments. *SIAM J. Sci. Comput.* **44**, B395–B427. (doi:10.1137/21M1404363)
58. Attia A, Ahmed SE. 2023 PyOED: an extensible suite for data assimilation and model-constrained optimal design of experiments. (<http://arxiv.org/abs/2301.08336>)