Fast state estimation in nonlinear parametric time dependent systems using Tensor Train

Damiano Lombardi

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Fast state estimation in nonlinear parametric time dependent systems using Tensor Train.

Lombardi Damiano*

[1]COMMEDIA, Inria Paris, 2 rue Simone Iff, 75012, Paris, France

Correspondence
*Lombardi Damiano. Email: damiano.lombardi@inria.fr

Present Address
Inria Paris.

Summary

In the present work we propose a reduced-order method to solve the state estimation problem when nonlinear parametric time-dependent systems are at hand. The method is based on the approximation of the set of system solutions by means of a Tensor Train format. The particular structure of Tensor Train makes it possible to set up both a variational and a sequential method. Several numerical experiments are proposed to assess the behaviour of the method.

KEYWORDS:
State Estimation, Reduced-order modelling, Tensor Train.

1 | INTRODUCTION

State estimation is one of the fundamental problems in data assimilation, and we refer to [1,2,3,4,5] for a general introduction to this topic. As many problems in the realm of data assimilation, state estimation often involves a heavy computational burden, making it a challenging or a prohibitive task.

The classical formulation of state estimation problems is the so called variational formulation: the problem is recast as an optimisation problem in which the cost function is a norm of the discrepancy between the measurements and the state observations (a comprehensive overview can be found in [1]). In order to mitigate the computational cost of variational approaches, several works proposed to replace the costly full-order model by a reduced-order one [6,7,8,9,10,11,12,13,14]. Equivalently, in a number of works, a low rank constraint has been taken into account [15,16].

Another way of speeding up the variational data assimilation consists in trying to directly reduce the inverse problem: this is what is achieved, for instance, by the Parametrized Background Data Weak (PBDW) method [17,18,19,20,21,22]. Synthetically, we

could say that PBDW recasts the state estimation problem as an optimal recovery problem. PBDW has been originally proposed for static problems. It has been adapted to time dependent problems in [23].

An alternative route to data assimilation is represented by sequential methods [24,25,26,27,28,29]. These are suitable in cases in which the data are progressively acquired, consisting often in long time series, and when data assimilation is a step towards control tasks in quasi real time. From a computational point of view, we can see sequential methods as a viable way to decrease substantially the computational cost of variational data assimilation.

In several works the authors proposed methods to further reduce the computational cost of sequential methods by using reduced-order models [30,31,32,33,34].

Recently, tensor methods have been taken into consideration to solve inverse problems and data assimilation problems in an efficient way. In [35] tensor methods are used to speed-up the bayesian inverse problem arising in electromyography. Tensor methods have been proposed to mitigate the high complexity of the Kalman filter with the state dimension in [36] and in combination with a Kalman filter strategy to identify multilinear forms in [37]. In [38] a tensor Henkel completion problem is introduced in order to perform data assimilation for traffic speed estimation.

In the present work we investigate the possibility of exploiting a tensor approximation of a set of parametric solutions of a non-linear time dependent dynamical system to set up an efficient framework for data assimilation. The contributions of the present work are the following:

1. By exploiting a Tensor Train (TT) approximation of a set of solutions (we refer to [39]) of a parametric dynamical system, we propose a fast variational method to solve the state estimation problem when partial noisy observations of the state are available. The proposed approach is similar, in the spirit, to the time adaptation of PBDW (presented in [23]), in the sense that we can perform a state estimation without performing a parameter estimation. An analysis of the problem makes it possible to gain some insight on the observability.

2. The interpretation of the Tensor Train as a dynamical basis approximation makes it possible to derive a sequential method. Since the state estimation problem to be solved is linear and we restrict to linear observation operators, the resulting approach is a particularisation of the Kalman filter.

The structure of the work is as follows. In Section 2 we introduce the notation used throughout the whole manuscript. In Section 3 we present the main idea of the method, the variational and the sequential formulation as well as the error analysis. In Section 4 we present three numerical experiments in order to assess the behaviour of the method.

2 | NOTATION.

Let $d, p \in \mathbb{N}^*$, let $\Omega \subset \mathbb{R}^d$ be an open bounded set, the physical domain, let $\Theta \subset \mathbb{R}^p$ be the parameters domain. In the following we denote a point $x \in \Omega$, a parameter vector $\vartheta \in \Theta$ and the time by $t \in [0, T]$. The system state is a real valued function $u \in \mathcal{U}(\Omega \times [0, T] \times \Theta)$, where $\mathcal{U}$ is a suitable Banach space:

$$u(x, t, \vartheta) : \begin{cases} \Omega \times [0, T] \times \Theta \ \rightarrow \ \mathbb{R} \\ (x, t, \vartheta) \ \mapsto u(x, t, \vartheta) \end{cases} \tag{1}$$

For sake of simplicity we consider here a scalar function, the extension to vector valued functions being straightforward (Some cases will indeed be presented in Section 4).

The system state is in general governed by an equation of the form $\mathcal{F}(u, x, t, \vartheta) = 0$. In the present work we make the following assumptions:

1. The system state $u \in \mathcal{U} \subseteq L^2(\Omega \times [0, T] \times \Theta)$. It holds $L^2(\Omega \times [0, T] \times \Theta) = L^2(\Omega) \otimes L^2([0, T]) \otimes L^2(\Theta)$. We denote the L^2 scalar product of two elements $u, v \in L^2(\Omega)$ as $\langle u, v \rangle_{L^2(\Omega)}$.

2. We introduce an approximation of the state in a Tensor Train format[39]. We consider here a continuous formulation, and, with a slight abuse of notation we introduce the TT format for a multivariate function for which the function domain is the cartesian product of domains. Tensor train in a continuous setting has been presented in[40,41]. Concerning the system state, we choose here to set a priori the variable index order as to be $x - t - \vartheta$. This choice is motivated by the problem at hand, in which we would like to estimate a space-time field, resulting from a linear combination of few space-time modes.

Let $r_1, r_2 \in \mathbb{N}^*$ be the TT-ranks and $i_1 = 1, \dots, r_1$, $i_2 = 1, \dots, r_2$.

The space modes are defined as $\varphi_{i_1} \in L^2(\Omega)$:

$$\varphi_{i_1}(x) : \begin{cases} \Omega \ \rightarrow \ \mathbb{R} \\ x \ \mapsto \varphi_{i_1}(x) \end{cases} \tag{2}$$

The time-dependent TT-core is defined as a matrix function whose entries $G_{i_1 i_2} \in L^2(0, T)$ read:

$$G_{i_1 i_2}(t) : \begin{cases} [0, T] \ \rightarrow \ \mathbb{R} \\ t \ \mapsto G_{i_1 i_2}(t) \end{cases} \tag{3}$$

The parameter modes are real valued functions $s_{i_2}(\vartheta) \in L^2(\Theta)$ defined as:

$$s_{i_2}(\vartheta) : \begin{cases} \Theta \ \rightarrow \ \mathbb{R} \\ \vartheta \ \mapsto s_{i_2}(\vartheta) \end{cases} \tag{4}$$

The TT approximation of the system state is:

$$u(x,t,\vartheta) \approx \sum_{i_1=1}^{r_1}\sum_{i_2=1}^{r_2} \varphi_{i_1}(x) G_{i_1 i_2}(t) s_{i_2}(\vartheta). \tag{5}$$

This is the way in which the model will be taken into account in the data assimilation problem, and it is efficient provided that the ranks r_1, r_2 are not too large. The system state can be observed through a process, which provides, in general, partial and noisy measurements. In the present work we make the following assumptions:

1. A measurement is an application of a linear form on the system state.
2. The noise is additive, unbiased and independent.
3. The noise has a finite covariance.

Let $m \in \mathbb{N}^*$ denote the number of measurements and $j = 1, \dots, m$. There exists a set of elements ω_j such that the measurement is the scalar product between the state and the elements ω_j. The observable is a vector valued function $y \in [L^2([0,T]\times\Theta)]^m$, whose components are defined as:

$$y_j(t,\vartheta) : \begin{cases} [0,T]\times\Theta \ \to \ \mathbb{R} \\ (t,\vartheta) \ \mapsto y_j(t,\vartheta) = \langle u, \omega_j \rangle_{L^2(\Omega)} \end{cases} \tag{6}$$

A system instance is featured by a specific value ϑ_* and when it is observed at time t_* this produces a real vector $y_* \in \mathbb{R}^m$. The second hypothesis on the nature of the measurements simply means that we can model them as:

$$y^* = y(t_*, \vartheta_*) + \nu. \tag{7}$$

The third hypothesis translates into the fact that we can introduce a norm based on the covariance $W \in \mathbb{R}^{m\times m}$, defined as follows:

$$v \in \mathbb{R}^m, \quad \|v\|_W^2 = v^T W^{-1} v. \tag{8}$$

In the present work we will make use of some terms and notation related to tensor scientific computing, in particular fibres and slices. When translated to the context of the present work, a fibre is a function depending upon one single variable (e.g. we fix $t = t_0, \vartheta = \vartheta_0$ and we consider $u(x, t_0, \vartheta_0)$). The slices are functions depending upon two variables, for instance $u(x, t, \vartheta_0)$.

3 | THE METHOD.

The main idea which is investigated in the present contribution is the following. We would like to estimate the system state without explicitly computing the hidden parameters ϑ. Under the assumptions presented in the previous section, by leveraging the properties of the TT approximation, we can achieve this goal by solving a linear problem.

Let us assume that the 'true' state of the system is $u^*(x,t)$ and there exists at least one parameter value $\vartheta_* \in \Theta$ such that the distance between the TT-model and the true state is less than a prescribed accuracy. In general, the existence of a unique parameter value such that the TT-model is the closest possible to the true state is an open question (identifiability), and it is not investigated in this work.

When introducing a value $\vartheta_* \in \Theta$ in the parameters modes of the TT expression we get a real vector $\beta^* \in \mathbb{R}^{r_2}$ whose entries are:

$$\beta^*_{i_2} = s_{i_2}(\vartheta_*).$$

In this context, performing a state estimation amounts to finding an approximation $\beta \in \mathbb{R}^{r_2}$ of β^* by exploiting a set of measurements at different times. Let us call $\hat{u}$ the resulting state approximation:

$$\hat{u}(x,t) = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \varphi_{i_1}(x) G_{i_1 i_2}(t) \beta_{i_2}. \tag{9}$$

Instead of solving a heavily non-linear problem to find ϑ, we directly try to estimate the state u by solving a linear problem for β. Several remarks are in order.

Remark.

For sake of simplicity in the presentation, we have considered here the case in which the solution variability is purely related to the parameters. Of course in general there can be also a variability in the initial and boundary conditions. This can be accounted for in a similar way. Indeed, once the TT approximation of the set of solutions is available, the state estimation problem simply translates into a linear problem for β. We would like to estimate a linear combination of space-time modes in such a way that the discrepancy between measurements and model observations is low in norm.

Remark.

The TT approximation introduced above can be interpreted as a dynamical basis approach in view of solving the data assimilation problem, indeed:

$$\hat{u}(x,t) = \sum_{i_2=1}^{r_2} \left(\sum_{i_1=1}^{r_1} \varphi_{i_1}(x) G_{i_1 i_2}(t) \right) \beta_{i_2}. \tag{10}$$

The actual approximation of the state is obtained as a linear combination with constant in time weights β of the time-varying modes determined by the contraction of the space modes φ_{i_1} with the time dependent core $G_{i_1 i_2}$. For every time t we can consider $\sum_{i_1=1}^{r_1} \varphi_{i_1}(x) G_{i_1 i_2}(t)$ as a set of i_2 space modes built by means of a linear combination of i_1 space ambient modes φ_{i_1}. When $i_2 < i_1$ we are in a case in which the ambient space is bigger and we consider a linear subspace embedded in it and evolving in it. When considering the case $i_2 > i_1$ we are performing a lifting thanks to the core G, introducing a redundant dictionary of linear dependent space modes.

Remark.

In the present work, we will construct a database of solutions of the system for different values of the parameters and then we will determine the Tensor Train by using the TT-SVD algorithm (detailed in[39]). Given the system equation $\mathcal{F}(u, \vartheta) = 0$ we could directly construct an approximation in Tensor Train format of the set of solutions[42,43,44,45,46,47,48]. This is an appealing possibility in view of realistic applications.

3.1 | Variational data assimilation.

The variational data assimilation approach consists in recasting the state estimation as an optimisation problem. Let $N_t \in \mathbb{N}^*$ be the number of time observations. Let $k = 1, \dots, N_t$ and the observation times be $t_k \in [0, T]$. The set of observations is $y_k^* \in \mathbb{R}^m$. The problem consists in finding $\beta \in \mathbb{R}^{r_2}$ such that a misfit functional, $\mathcal{J}$, is minimised.

$$\mathcal{J}(\beta) : \begin{cases} \mathbb{R}^{r_2} \rightarrow \mathbb{R}^+ \\ \beta \mapsto \mathcal{J}(\beta) = \frac{1}{2} \sum_{k=1}^{N_t} \| y_k^* - \langle \hat{u}(x, t_k), \boldsymbol{\omega} \rangle_{L^2(\Omega)} \|_W^2 \end{cases}$$

A regularisation by means of a prior can be introduced. Let $P_0 \in \mathbb{R}^{r_2 \times r_2}$ be a symmetric positive definite matrix. A norm can be induced by the scalar product:

$$\beta \in \mathbb{R}^{r_2}, \quad \|\beta\|_{P_0}^2 = \beta^T P_0^{-1} \beta.$$

The minimisation problem to be solved reads:

$$\hat{\beta} = \arg \inf_{\beta \in \mathbb{R}^{r_2}} \mathcal{J}(\beta) + \frac{1}{2} \|\beta - \beta_0\|_{P_0}^2. \tag{11}$$

3.1.1 | Existence and uniqueness of the solution.

The problem presented in Eq.(11), with the regularisation term $\|\beta - \beta_0\|_{P_0}^2$ admits a unique solution. This can be deduced by the expression of the function $\mathcal{J}$. Indeed, provided that P_0 is symmetric and positive definite, it is a strictly convex function, such that $\lim_{\|\beta\|_{\ell^{2,m}} \rightarrow \infty} \mathcal{J}(\beta) = +\infty$. This implies that there exists, unique, the minimiser of the problem. In this case, there exists a unique stationary point, which is also the global minimum, solution of the Euler-Lagrange equations. Let us introduce some notation. Let $G^x \in \mathbb{R}^{m \times r_1}$ be the Grammian matrix corresponding to the observation of the space modes $\varphi_{i_1}(x)$:

$$G^x_{j i_1} = \langle \varphi_{i_1}, \omega_j \rangle_{L^2(\Omega)}.$$

Let $\Psi \in \mathbb{R}^{m \times N_t \times r_2}$ be a third order tensor whose expression reads:

$$\Psi_{j k i_2} = \sum_{i_1=1}^{r_1} G^x_{j i_1} G_{i_1 i_2}(t_k).$$

It represents the observation of the time evolving r_2 modes that are used to approximate the model state. Henceforth, the observation of the model outcome reduces to the linear combination of these:

$$\hat{y}_j(t_k) = \sum_{i_2=1}^{r_2} \Psi_{jki_2}\beta_{i_2},$$

which is a contraction of the third order tensor by β. The slices of Ψ with respect to the index $k = 1, \dots, N_t$ are considered, consisting in matrices: $\Psi^{(k)} \in \mathbb{R}^{m\times N_t}$. The Euler-Lagrange equations read:

$$\left[P_0^{-1} + \sum_{k=1}^{N_t} [\Psi^{(k)}]^T W^{-1} [\Psi^{(k)}]\right] \beta = P_0^{-1}\beta_0 + \sum_{k=1}^{N_t} [\Psi^{(k)}]^T W^{-1} y_k^*. \tag{12}$$

By inspection, we see that the matrix on the left hand side is actually invertible (this condition being ensured by the regularisation), providing a unique solution.

It is of interest to consider the case of no regolarisation, namely $P_0^{-1} = 0$. In this case, the condition for the system matrix on the right-hand side to be invertible depends on the sequence of matrices $\Psi^{(k)}$. Let us consider the unfolding of Ψ consisting in stacking vertically the slices $\Psi^{(k)}$. This provides a matrix $O \in \mathbb{R}^{m\cdot N_t\times r_2}$:

$$O = \begin{bmatrix} \Psi^{(1)} \\ \Psi^{(2)} \\ \dots \\ \Psi^{(N_t)} \end{bmatrix}$$

We introduce $\hat{W}^{-1} \in \mathbb{R}^{m\cdot N_t\times m\cdot N_t}$ which is a block diagonal matrix in which all the N_t blocks are equal to W^{-1}:

$$\hat{W}^{-1} = \begin{bmatrix} W^{-1} & 0 & \dots & 0 \\ 0 & W^{-1} & \dots & 0 \\ \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & W^{-1} \end{bmatrix}$$

It holds:

$$\sum_{k=1}^{N_t} [\Psi^{(k)}]^T W^{-1} [\Psi^{(k)}] = O^T \hat{W}^{-1} O.$$

The matrix $O^T\hat{W}^{-1}O \in \mathbb{R}^{r_2\times r_2}$ is invertible provided that: $\mathrm{rank}(O) = r_2$. This is an Observability condition (perfectly analogue to the Kalman observability[49]). Remark that an approximation of β can be obtained by repeated observations in time even for cases in which $m < r_2$, which is a necessary condition for the observability of static systems.

A way to assess the observability is to estimate the lowest eigenvalue of $O^T\hat{W}^{-1}O \in \mathbb{R}^{r_2\times r_2}$, denoted λ_{r_2}.

Remark.

In cases in which r_2 is not large the smallest eigenvalue can be directly computed. As an alternative, an analytical estimate can be obtained as follows: let $\gamma = \frac{\mathrm{tr}([O^T\hat{W}^{-1}O])}{r_2}$ and $\eta = \frac{\mathrm{tr}([O^T\hat{W}^{-1}O]^2)}{r_2} - \gamma^2$. In[50] Wolkowicz and Styan proved a bound for the lowest and the largest singular values (respectively λ_{r_2} and λ_1) for a complex square matrix:

$$\gamma - \eta(r_2-1)^{1/2} \leq \lambda_{r_2} \leq \gamma - \frac{\eta}{(r_2-1)^{1/2}},$$
$$\gamma + \frac{\eta}{(r_2-1)^{1/2}} \leq \lambda_1 \leq \gamma + \eta(r_2-1)^{1/2}.$$

3.1.2 | Error analysis.

In this section, we study the error in the prediction of a state $u^*(x,t,\vartheta)$. The following hypotheses are made:

(H_1) Let the best TT approximation of the system at time t and parameter ϑ be denoted by $\hat{u}^*$. The $L^2(\Omega_x)$ error of the system state satisfies:

$$\sup_{t\in[0,T]}\sup_{\vartheta\in\Theta}\|u^* - \hat{u}^*\|^2_{L^2(\Omega_x)} \leq \varepsilon^2,$$

which is the TT truncation error.

(H_2) The noise ν is additive, independent, unbiased, with covariance $W \in \mathbb{R}^{m\times m}$.

Proposition 1. Let the hypotheses $H_1 - H_2$ hold and let us consider no regularisation ($P_0^{-1} = 0$). Let β and β^* be the representation of the best TT state and of the estimated state on the space-time basis. Due to the presence of noise β is a random variables. Let $M = [O^T\hat{W}^{-1}O]$ denote the observability matrix, for sake of compactness in the notation. It holds:

1. $\|\mathbb{E}[(\beta - \beta^*)]\|_{\ell^{2,r_2}} \leq \frac{\varepsilon}{\lambda_{r_2}^{1/2}}\|\omega\|_{[L^2(\Omega)]^m}$.
2. $\mathbb{E}[\|\beta - \beta_*\|^2_{\ell^{2,r_2}}] =\leq \mathrm{tr}(M^{-1}) \leq \frac{\lambda_{r_2}r_2\mathrm{tr}(M) - \mathrm{tr}(M)^2 + r_2\|M\|_F^2 - \lambda_{r_2}^2 r_2}{\lambda_{r_2}\|M\|_F^2 - \lambda_{r_2}^2\mathrm{tr}(M)}$.

Proof. The equation for β is obtained by measuring the true state:

$$[O^T\hat{W}^{-1}O]\beta = \sum_{k=1}^{N_t}[\Psi^{(k)}]^T W^{-1}(\langle\omega, u^*(t_k)\rangle_{L^2(\Omega_x)} + \nu^{(k)}).$$

Let the TT truncation error (the model error in the present case) be denoted by z. By introducing the relationship: $u^* = \hat{u}^* + z$, we get:

$$[O^T\hat{W}^{-1}O](\beta - \beta^*) = \sum_{k=1}^{N_t}[\Psi^{(k)}]^T W^{-1}(\langle\omega, z(t_k,\vartheta^*)\rangle_{L^2(\Omega)} + \nu^{(k)}).$$

There are two contributions to the error: the first one is deterministic and it is related to the observation of the model error (the measurement of the TT truncation error z), the second one is stochastic and it is related to the measurement noise. Since the

noise is unbiased, the expectation of the error reads:

$$\mathbb{E}[(\beta - \beta^*)] = [O^T \hat{W}^{-1} O]^{-1} \sum_{k=1}^{N_t} [\Psi^{(k)}]^T W^{-1} \langle \omega, z(t_k, \vartheta^*) \rangle_{L^2(\Omega)}.$$

The ℓ^2 norm of the mean error, obtained by considering the singular value decomposition of the matrix O, satisfies:

$$\|\mathbb{E}[(\beta - \beta^*)]\|_{\ell^{2,r_2}} \leq \frac{\varepsilon}{\lambda_{r_2}^{1/2}} \|\omega\|_{[L^2(\Omega)]^m}. \tag{13}$$

This proves the first point of the proposition.

The covariance of the error is given by:

$$C = \mathbb{E}[((\beta - \beta^*) - \mathbb{E}[\beta - \beta^*]) \otimes ((\beta - \beta^*) - \mathbb{E}[\beta - \beta^*])].$$

For sake of compactness let us introduce $O^\dagger = [O^T \hat{W}^{-1} O]^{-1} O^T \hat{W}^{-1}$. The covariance of the error, by virtue of the hypotheses on the noise structure, satisfies:

$$C_{ih} = \sum_{k,p=1}^{N_t} \sum_{j,s=1}^{m} O^\dagger_{ikj} W_{js} \delta_{kp} O^\dagger_{hps} = [O^T \hat{W}^{-1} O]^{-1}_{ih}. \tag{14}$$

The expected value of the ℓ^{2,r_2} norm squared of the error satisfies:

$$\mathbb{E}[\|\beta - \beta_*\|^2_{\ell^{2,r_2}}] = \mathrm{tr}(C) - \mathbb{E}[\beta - \beta^*]^2 \leq \mathrm{tr}(C),$$

where the equality holds for the cases in which the model error is not observable, *i.e.* $\langle \omega, z \rangle_{L^2(\Omega)} = 0$.

In order to bound the trace of the error covariance, we use a result proved by Bai and Golub in[51]:

$$\mathrm{tr}([O^T \hat{W}^{-1} O]^{-1}) \leq \begin{bmatrix} \mathrm{tr}(O^T \hat{W}^{-1} O) & r_2 \end{bmatrix} \begin{bmatrix} \|O^T \hat{W}^{-1} O]\|_F^2 & \mathrm{tr}(O^T \hat{W}^{-1} O) \\ \lambda_{r_2}^2 & \lambda_{r_2} \end{bmatrix}^{-1} \begin{bmatrix} r_2 \\ 1 \end{bmatrix}. \tag{15}$$

This leads to the conclusion. □

The further bound on the trace of the inverse of the observability matrix, obtained by computing explicitly the bound proposed by Bai and Golub, might be computationally appealing in view of estimating the error propagation in large systems. Indeed, it makes it possible to get an estimation without computing and storing the inverse of the observability matrix.

Proposition 2. Let the hypotheses $H_1 - H_2$ hold and let us consider no regularisation terms ($P_0^{-1} = 0$). Let the error norm squared be denoted by:

$$e^2 = \|u_* - \hat{u}\|^2_{L^2(\Omega \times [0,T])}. \tag{16}$$

Due to the presence of noise, this is a random variable. Its expected value verifies:

$$\mathbb{E}[e^2] \leq T\varepsilon^2 + \mathrm{tr}([O^T \hat{W}^{-1} O]^{-1}). \tag{17}$$

Proof. Let the TT state approximation be denoted by $\hat{u}$. The first step to prove the error bound consists in noticing that the error can be written as follows:

$$\|u^* - \hat{u}\|^2_{L^2(\Omega\times[0,T])} = \int_0^T \|u^* - \hat{u}\|^2_{L^2(\Omega_x)}\, dt,$$

which can be rewritten as:

$$\int_0^T \|u^* - \hat{u}^* + \hat{u}^* - \hat{u}\|^2_{L^2(\Omega)}\, dt \leq T\varepsilon^2 + \int_0^T \|\hat{u}^* - \hat{u}\|^2_{L^2(\Omega)}.$$

This holds true because of the orthogonality properties of the TT approximation and by virtue of the hypothesis (H_1).

Let us focus on the second term on the right hand side:

$$\int_0^T \|\hat{u}^* - \hat{u}\|^2_{L^2(\Omega)} = \int_0^T \sum_{i_1,j_1}^{r_1} \langle \varphi_{i_1}, \varphi_{j_1}\rangle_{L^2(\Omega)} \sum_{i_2,j_2}^{r_2} G_{i_1 i_2} G_{j_1 j_2} (\beta^*_{i_2} - \beta_{i_2})(\beta^*_{j_2} - \beta_{j_2})\, dt.$$

Under the hypothesis that the TT approximation has been obtained or rounded by TT-SVD, it holds:

$$\langle \varphi_{i_1}, \varphi_{j_1}\rangle_\Omega = \delta_{i_1,j_1},$$
$$\int_0^T \sum_{i_1}^{r_1} G_{i_1 i_2} G_{i_1 j_2}\, dt = \delta_{i_2,j_2},$$

leading to:

$$\int_0^T \|\hat{u}^* - \hat{u}\|^2_{L^2(\Omega)} = \|\beta^* - \beta\|^2_{\ell^{2,r_2}}. \tag{18}$$

The result of the previous proposition makes it possible to conclude. □

3.2 | Sequential data assimilation.

If the model solution can be represented in a TT format, a model instance (*i.e.* a state evolution $u(x,t)$) is simply identified by a vector $\beta \in \mathbb{R}^{r_2}$. This consideration is appealing in order to introduce a model for the evolution of the state: this is the starting point to set up a sequential data assimilation approach. The cost of solving a sequential data assimilation in this particular setting is not per se smaller than solving a variational data assimilation. Despite this, the sequential approach is of the utmost interest, since it is the most natural to deal with applications in which measurements are sets of long time series, acquired progressively, or applications in which the observer is a mean to set up a control.

Consider two subsequent time steps t_k, t_{k+1}. The states are given by:

$$u(x, t_k, \vartheta_*) = \sum_{i_1=1}^{r_1}\sum_{i_2=1}^{r_2} \varphi_{i_1}(x) G_{i_1 i_2}(t_k)\beta_{i_2} \tag{19}$$

$$u(x, t_{k+1}, \vartheta_*) = \sum_{i_1=1}^{r_1}\sum_{i_2=1}^{r_2} \varphi_{i_1}(x) G_{i_1 i_2}(t_{k+1})\beta_{i_2}. \tag{20}$$

The time evolution is accounted for by the time dependent core G and henceforth the vector $\beta \in \mathbb{R}^{r_2}$ is constant. The model on the evolution of β reduces to:

$$d_t \beta = 0.$$

When interpreting TT as a dynamical basis approach we see that this condition derives from the fact that the basis evolve, not the representation coefficients of the state on it. When such an approximation is considered, the evolution model is linear, the observations are linear, a Kalman filter is probably the best way to set up a sequential state estimation. Let $\beta^{(0)}$, $P^{(0)}$ be the initial state estimation (a pure *a priori* value) and its associated covariance. The Kalman filter reads as follows:

$$\beta^{(k+1,k)} = \beta^{(k)}, \tag{21}$$

$$P^{(k+1,k)} = P^{(k)} + M, \tag{22}$$

$$d = y^*(t_{k+1}) - \Psi^{(k+1)}\beta^{(k+1,k)}, \tag{23}$$

$$S = W + \Psi^{(k+1)} P^{(k+1,k)} [\Psi^{(k+1)}]^T, \tag{24}$$

$$K = P^{(k+1,k)} [\Psi^{(k+1)}]^T S^{-1} \tag{25}$$

$$\beta^{(k+1)} = \beta^{(k+1,k)} + Kd, \tag{26}$$

$$P^{(k+1)} = [I - K\Psi^{(k+1)}] P^{(k+1,k)} \tag{27}$$

3.2.1 | Error analysis

The error of the sequential approach is studied. In particular, we prove that, in absence of model error ($M^{(k)} = 0,\ \forall k$), the estimate provided at final time by the Kalman filter equals the variational one. This is a know result in a standard setting. Here we simply prove that the proposed formulation respects this.

Proposition 3. Let $\beta^{(0)}$, $P^{(0)}$ be the initial guess of the estimation. The final covariance matrix of the Kalman filter estimation satisfies:

$$[P^{(N_t)}]^{-1} = [P^{(0)}]^{-1} + O^T \hat{W}^{-1} O.$$

Moreover:

$$\beta^{(N_t)} = \beta^{(var)},$$

where $\beta^{(var)}$ satisfies (the reader is referred to Eq.12):

$$[P^{(0)}]^{-1} + O^T \hat{W}^{-1} O] \beta^{(var)} = [P^{(0)}]^{-1} \beta^{(0)} + \sum_{i=1}^{N_t} [\Psi^{(k)}]^T W^{-1} y_*^{(k)}.$$

Proof. Let us write the Kalman filter as an observer:

$$K = P^{(k)}[\Psi^{(k+1)}]^T(W + \Psi^{(k+1)}P^{(k)}[\Psi^{(k+1)}]^T)^{-1} \tag{28}$$

$$\beta^{(k+1)} = [I - K\Psi^{(k+1)}]\beta^{(k)} + Ky_*^{(k+1)}. \tag{29}$$

$$P^{(k+1)} = [I - K\Psi^{(k+1)}]P^{(k)} \tag{30}$$

By making use of the Woodbury matrix identity, this iteration can be rewritten as follows:

$$[P^{(k+1)}]^{-1} = [P^{(k)}]^{-1} + [\Psi^{(k+1)}]^T W^{-1}\Psi^{(k+1)}, \tag{31}$$

$$\left[P^{(k+1)}\right]^{-1}\beta^{(k+1)} = [P^{(k)}]^{-1}\beta^{(k)} + [\Psi^{(k+1)}]^T W^{-1}y_*^{(k+1)}. \tag{32}$$

This leads to:

$$\beta^{(k+1)} = P^{(k+1)}[P^{(k)}]^{-1}\beta^{(k)} + P^{(k+1)}[\Psi^{(k+1)}]^T W^{-1}y_*^{(k+1)}, \tag{33}$$

and iterating for $k, k-1, \dots, k=1$, we obtain:

$$\beta^{(N_t)} = P^{(N_t)}\left([P^{(0)}]^{-1}\beta^{(0)} + \sum_{i=1}^{N_t}[\Psi^{(k)}]^T W^{-1}y_*^{(k)}\right). \tag{34}$$

The estimate of $P^{(N_t)}$ can be obtained by making use of the equation for the inverse of the covariance. It holds:

$$[P^{(N_t)}]^{-1} = [P^{(0)}]^{-1} + \sum_{i=1}^{N_t}[\Psi^{(k)}]^T W^{-1}\Psi^{(k)}, \tag{35}$$

which concludes the proof, since it delivers the same result as the corresponding variational estimation. □

4 | NUMERICAL EXPERIMENTS.

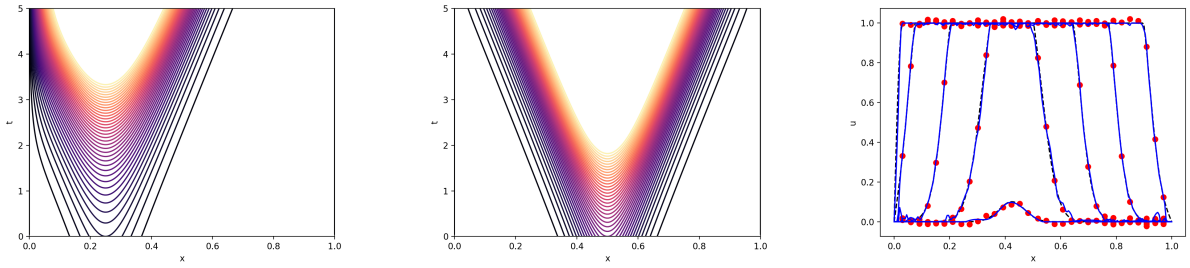
Several numerical experiments are proposed to illustrate the method and assess its properties and performances.

In all the test cases presented in this section, we will consider, as a measure of quality of the reconstruction, the relative L^2 error of the state reconstruction:

$$e = \frac{\int_0^T \int_\Omega (u^* - \hat{u})^2 \, dxdt}{\int_0^T \int_\Omega (u^*)^2 \, dxdt} \tag{36}$$

In particular, the methods proposed in the present work will be tested on $n_s \in \mathbb{N}$ random instances, and the mean and the standard deviation of e will be computed, denoted by $\overline{e}$ and σ respectively. The subscripts *var* and *seq* will be used to denote the results obtained by the variational and the sequential approaches.

Moreover, to assess the observability of the system in different configurations, we will compute the smallest eigenvalue of the observability matrix λ_r and the Bai and Golub bound of the trace of its inverse (denoted by *BG*).



(a) (b) (c)

FIGURE 1 Test case presented in Section 4.1: (a-b) contour plots of two different solutions, for two different values of parameters. (c) Reconstruction at different time steps: the red dots are the observations, the black dashed line is the true state and the blue line is the solution of the variational data assimilation.

4.1 | 1D-1D Fisher-Kolmogorov-Petrovski-Piskunov equation.

The first test proposed is on the Fisher-Kolmogorov-Petrovski-Piskunov (FKPP) equation. Let $x \in [0, 1]$ and $t \in [0, 5]$. Let $\nu = 10^{-3}$, $c \in [0.5, 5]$ be the reaction coefficient, $A \in [0.1, 0.6]$, $\kappa \in [100, 250]$, $\mu \in [0.25, 0.75]$ be the parameters determining the initial condition:

$$u_0(x) = A \exp\left(-\kappa(x - \mu)^2\right). \tag{37}$$

The equation reads:

$$\partial_t u = \nu \partial_x^2 u + cu(1 - u), \tag{38}$$

$$u(0, t) = u(1, t) = 0, \tag{39}$$

$$u(x, 0) = u_0. \tag{40}$$

The parameters are $(c, A, \kappa, \mu) \in \Theta$. The set is sampled taking a cartesian grid with $n = 6$ values uniformly spaced, in each direction, leading to $N_s = 1296$ solution samples. Two solutions examples are shown in Fig. 1.a-b). The TT approximation is extracted from this set by using the TT-SVD algorithm.

The measurements are point values of u, uniformly distributed on the domain. The configurations we are going to test are the following:

- $m = \{8, 16, 32, 64\}$, corresponding to $\{4\%, 8\%, 16\%, 32\%\}$ of the space domain degrees of freedom, spanning the spectrum from scarce to rich observations.
- $W^{-1} = \left\{10^4 I, 10^2 I\right\}$ corresponding to low and high levels of noise.
- The value of $\varepsilon_{TT} = \left\{10^{-2}, 10^{-3}\right\}$. For the case considered the TT ranks are $(r_1 = 40, r_2 = 193)$ and $(r_1 = 141, r_2 = 932)$ respectively.

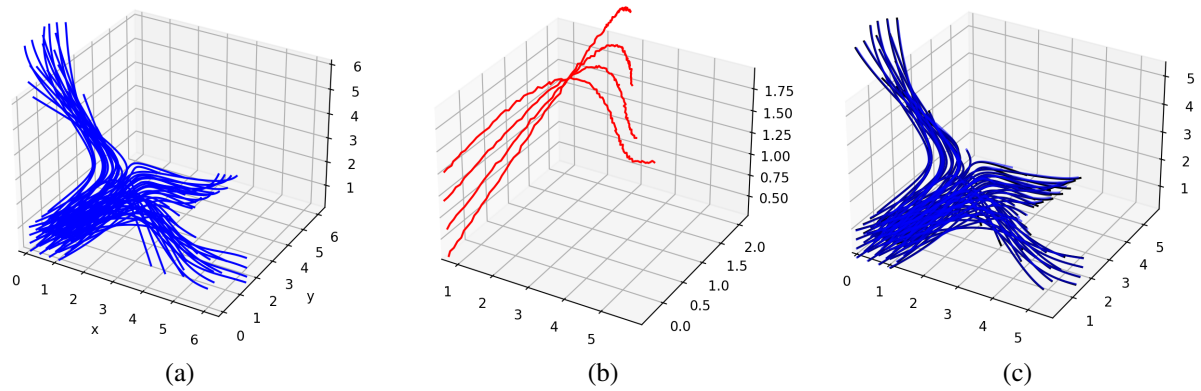
ε_{TT}	m	W^{-1}	λ_r	BG	$\overline{e}_{var}$	σ_{var}	$\overline{e}_{seq}$	σ_{seq}
10^{-2}	8	10^{4}	1.70	30.67	$4.3\ 10^{-2}$	$1.94\ 10^{-2}$	$4.7\ 10^{-2}$	$1.45\ 10^{-2}$
10^{-2}	8	10^{2}	$1.7\ 10^{-2}$	3067	$1.0\ 10^{-1}$	$2.79\ 10^{-2}$	$2.0\ 10^{-1}$	$4.2\ 10^{-2}$
10^{-2}	16	10^{4}	25.8	0.726	$2.4\ 10^{-2}$	$9.10\ 10^{-3}$	$2.7\ 10^{-2}$	$8.91\ 10^{-3}$
10^{-2}	16	10^{2}	$2.58\ 10^{-1}$	72.60	$4.0\ 10^{-2}$	$1.10\ 10^{-2}$	$1.43\ 10^{-1}$	$3.3\ 10^{-2}$
10^{-2}	32	10^{4}	96.70	$1.39\ 10^{-1}$	$2.1\ 10^{-2}$	$8.0\ 10^{-3}$	$2.2\ 10^{-2}$	$8.1\ 10^{-3}$
10^{-2}	32	10^{2}	$9.67\ 10^{-1}$	13.96	$3.1\ 10^{-2}$	$8.6\ 10^{-3}$	$1.0\ 10^{-1}$	$2.2\ 10^{-2}$
10^{-2}	64	10^{4}	2082	$5.9\ 10^{-2}$	$2.09\ 10^{-2}$	$8.9\ 10^{-3}$	$2.10\ 10^{-2}$	$8.2\ 10^{-3}$
10^{-2}	64	10^{2}	20.82	5.94	$2.7\ 10^{-2}$	$7.7\ 10^{-3}$	$7.0\ 10^{-2}$	$1.7\ 10^{-2}$
10^{-3}	8	10^{4}	$8.76\ 10^{-5}$	$5.0\ 10^{6}$	$5.71\ 10^{-1}$	$9.0\ 10^{-2}$	$6.7\ 10^{-2}$	$1.45\ 10^{-2}$
10^{-3}	8	10^{2}	$8.76\ 10^{-7}$	$5.0\ 10^{8}$	$8.10\ 10^{-1}$	$1.2\ 10^{-1}$	$2.1\ 10^{-1}$	$4.5\ 10^{-2}$
10^{-3}	16	10^{4}	1.39	187.76	$6.40\ 10^{-2}$	$3.7\ 10^{-2}$	$2.9\ 10^{-2}$	$6.7\ 10^{-3}$
10^{-3}	16	10^{2}	$1.39\ 10^{-2}$	$1.878\ 10^{4}$	$1.40\ 10^{-1}$	$2.8\ 10^{-2}$	$1.45\ 10^{-1}$	$3.1\ 10^{-2}$
10^{-3}	32	10^{4}	10.75	9.23	$2.30\ 10^{-2}$	$1.1\ 10^{-2}$	$2.2\ 10^{-2}$	$8.1\ 10^{-3}$
10^{-3}	32	10^{2}	$1.075\ 10^{-1}$	92.3	$2.04\ 10^{-2}$	$5.26\ 10^{-3}$	$1.04\ 10^{-1}$	$2.36\ 10^{-2}$
10^{-3}	64	10^{4}	208	0.42	$1.37\ 10^{-2}$	$5.9\ 10^{-3}$	$1.67\ 10^{-2}$	$5.1\ 10^{-3}$
10^{-3}	64	10^{2}	2.08	4.20	$3.7\ 10^{-2}$	$7.50\ 10^{-3}$	$7.6\ 10^{-2}$	$1.73\ 10^{-2}$

TABLE 1 Result of the test case described in Section 4.1, for the different configurations of the data assimilation problem.

- The tests are performed by considering the solutions obtained by simulating the system when $n_s = 100$ random samples of the parameters are drawn from a uniform distribution in the range of the parameters considered for the database. We will consider the average and the standard deviation of the relative error.

An example of the reconstruction obtained by the variational method is shown in Fig. 1.c). The results of the numerical experiments are summarised in Table 1. In general, we see that a better observability of the system (encoded in the eigenvalue and the theoretical bound) translates into better results in terms of error on the state reconstruction. In this respect, having a more precise model does not guarantee to have a better result: indeed, consider the case of scarce observations $m = 8$; when $\varepsilon_{TT} = 10^{-3}$ a large number of TT-modes implies a poor observability (a lack of regularisation), and an error which is around 60% (obtained by means of a variational approach) with a moderate level of noise, contrary to the case $\varepsilon_{TT} = 10^{-2}$, for which the error is around 4%. Of course, when rich observations are available $m = 64$, a less precise model results in a saturation of error. We see that, with a variational approach the error is around 1.3% for $\varepsilon_{TT} = 10^{-3}$ and it stays at around 2% for $\varepsilon_{TT} = 10^{-2}$, as when taking only $m = 32$ observations into account. This is due to the fact that, for a less precise model, the TT truncation error starts polluting the reconstruction result.

Some differences can be seen between the variational and the sequential methods. In general, we observe that, for rich observations and low level of noise, the result of the variational estimation is better, and this coherent with the fact that the Kalman filter reaches the same estimate than the variational method only at the final time, whereas the error accounts for all the time history. In the cases in which the noise is large, or in cases in which the observability is poor, the sequential method is less prone to error propagation and the result is more robust.



(a) (b) (c)

FIGURE 2 Test case presented in Section 4.2: (a) Lagrangian trajectories of 64 fluid particles. (b) Observed trajectories, affective by a low level noise. (c) Reconstruction performed by a variational method, in black, the true Lagrangian trajectories, in blue, the reconstructed ones.

4.2 | Arnold-Beltrami-Childress flow.

The Arnold-Beltrami-Childress (ABC) flow is an example of chaotic dynamical system, whose solution is also a solution of the Euler equations written in a Lagrangian framework. Let $\mathbf{x} = (x, y, z) \in \mathbb{R}^3$, $t \in [0, 4]$, $(A, B, C) \in [0.9, 1.1]^3$. The system of equations reads:

$$\dot{x} = A\sin(z) + C\cos(y), \tag{41}$$

$$\dot{y} = B\sin(x) + A\cos(z), \tag{42}$$

$$\dot{z} = C\sin(y) + B\cos(x). \tag{43}$$

The initial condition for the system is a set of $N_p = 64$ points, $\mathbf{x}_0^{(i)}$, $1 \leq i \leq 64$, uniformly distributed in $[0, 1]^3$. We chose a final time $T = 4$ to discretise the equations with a $\Delta t = 0.04$. The problem we solve is the following: we observe m trajectories our of 64, with some noise. Based on these observations, we try to reconstruct the whole set of 64 Lagrangian trajectories. An example of solution is presented in Fig. 2.a). We solved the data assimilation problems in the following configurations:

- $m = \{2, 4, 8, 16\}$, which corresponds to observations ranging from 3.1% to 25% of the number of trajectories (space degrees of freedom).
- We considered two different levels of noise: $W^{-1} = \left\{10^4 I, 10^2 I\right\}$
- The value of $\varepsilon_{TT} = \left\{10^{-2}, 10^{-3}\right\}$. For the case considered the TT ranks are $(r_1 = 20, r_2 = 10)$ and $(r_1 = 48, r_2 = 31)$ respectively.
- The tests are performed by considering the solutions obtained by simulating the system when $n_s = 100$ random samples of the parameters are drawn from a uniform distribution in the range of the parameters considered for the database.

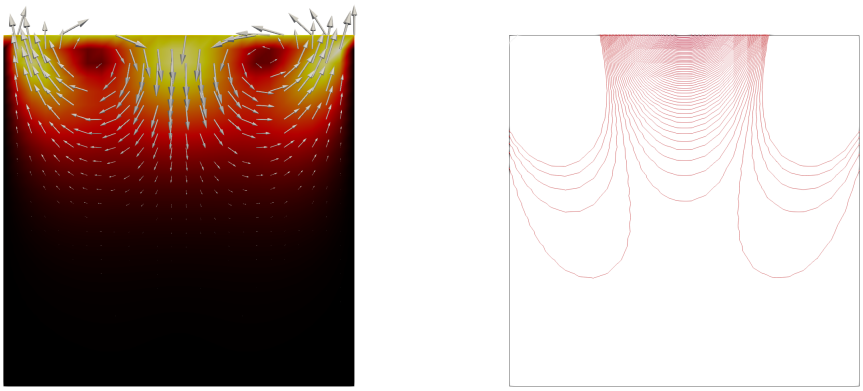
ε_{TT}	m	W^{-1}	λ_r	BG	$\overline{e}_{var}$	σ_{var}	$\overline{e}_{seq}$	σ_{seq}
10^{-2}	2	10^4	1.07	6.32	$1.75\ 10^{-2}$	$4.60\ 10^{-3}$	$1.89\ 10^{-2}$	$4.06\ 10^{-3}$
10^{-2}	2	10^2	$1.07\ 10^{-2}$	632	$4.45\ 10^{-2}$	$1.73\ 10^{-2}$	$1.90\ 10^{-1}$	$1.89\ 10^{-2}$
10^{-2}	4	10^4	1.94	3.72	$1.51\ 10^{-2}$	$3.93\ 10^{-3}$	$1.59\ 10^{-2}$	$4.35\ 10^{-3}$
10^{-2}	4	10^2	$1.94\ 10^{-2}$	372	$3.47\ 10^{-2}$	$1.16\ 10^{-2}$	$1.75\ 10^{-1}$	$1.75\ 10^{-2}$
10^{-2}	8	10^4	8.02	$8.10\ 10^{-1}$	$1.08\ 10^{-2}$	$2.88\ 10^{-3}$	$1.23\ 10^{-2}$	$3.75\ 10^{-3}$
10^{-2}	8	10^2	$8.02\ 10^{-2}$	81.10	$1.75\ 10^{-2}$	$5.87\ 10^{-3}$	$9.47\ 10^{-2}$	$1.52\ 10^{-2}$
10^{-2}	16	10^4	22.01	$2.26\ 10^{-1}$	$1.09\ 10^{-2}$	$3.39\ 10^{-3}$	$1.16\ 10^{-2}$	$3.14\ 10^{-3}$
10^{-2}	16	10^2	$2.20\ 10^{-1}$	22.60	$1.37\ 10^{-2}$	$3.98\ 10^{-3}$	$6.54\ 10^{-2}$	$1.12\ 10^{-2}$
10^{-3}	2	10^4	$1.28\ 10^{-4}$	$2.90\ 10^{4}$	$3.51\ 10^{-1}$	$1.50\ 10^{-1}$	$2.1\ 10^{-1}$	$1.24\ 10^{-2}$
10^{-3}	2	10^2	$1.28\ 10^{-6}$	$2.90\ 10^{6}$	2.87	1.01	$3.4\ 10^{-1}$	$1.70\ 10^{-2}$
10^{-3}	4	10^4	$1.37\ 10^{-2}$	1969	$4.20\ 10^{-2}$	$1.10\ 10^{-2}$	$8.88\ 10^{-2}$	$1.06\ 10^{-2}$
10^{-3}	4	10^2	$1.37\ 10^{-4}$	$1.97\ 10^{5}$	$3.90\ 10^{-1}$	$1.15\ 10^{-1}$	$2.66\ 10^{-1}$	$1.99\ 10^{-2}$
10^{-3}	8	10^4	$9.12\ 10^{-2}$	271.80	$1.56\ 10^{-2}$	$5.95\ 10^{-3}$	$2.77\ 10^{-2}$	$5.82\ 10^{-3}$
10^{-3}	8	10^2	$9.12\ 10^{-4}$	$2.71\ 10^{4}$	$1.52\ 10^{-1}$	$4.95\ 10^{-2}$	$1.87\ 10^{-1}$	$1.60\ 10^{-2}$
10^{-3}	16	10^4	0.32	65.0	$8.46\ 10^{-3}$	$3.0\ 10^{-3}$	$1.26\ 10^{-2}$	$4.37\ 10^{-3}$
10^{-3}	16	10^2	$3.20\ 10^{-3}$	$6.50\ 10^{3}$	$7.27\ 10^{-2}$	$2.63\ 10^{-2}$	$1.12\ 10^{-1}$	$1.29\ 10^{-2}$

TABLE 2 Result of the test case described in Section 4.2, for the different configurations of the data assimilation problem.

An example of the solution of the data assimilation problem is presented in Fig. 2: at the center we see $m = 4$ observed trajectories, with a low level of noise; on the right, we see the reconstruction of the 64 trajectories performed by means of a variational method. The results of the numerical experiments are summarised in Table 2. In general, the behaviour of the method is very similar to the one observed for the previous case. When we consider $\varepsilon_{TT} = 10^{-2}$ the model is less precise, but the lower dimensional space provides a regularisation which helps, in terms of observability, when having scarce observations. In this case, we observed that the variational approach always outperforms the sequential one and that the error is less than 5% on average. When having rich observations, the model error starts polluting the estimation, so that we cannot reach less than 1% of accuracy. When we consider a better approximation of the solutions set (taking $\varepsilon_{TT} = 10^{-3}$), we observe that the data assimilation is particularly sensitive to the case of scarce observations ($m = 2$). In this case, the sequential method provides a better results, with an error that stays below 35%. When increasing the number of observed trajectories, the variational approach can reach an error which is below 1%.

4.3 | 2D unsteady Stokes flow.

The last example we propose is a Stokes flow. It is a linear system, whose solution is a vector values function. Let $\overline{\Omega}_x = [0, 1]^2$, $T = [0, 1]$, and the viscosity be $\nu = 0.05$. Let $u(x, t)$ be the velocity field, let p be the pressure. We consider the following system



(a) (b)

FIGURE 3 Example of Stokes flow, Section 4.3, for $a_1 = 0.4$ and $a_2 = 0.35$. In (a) we show the velocity field and in (b) the contour lines of the pressure field, at time $t = 0.25$.

of equations:

$$\partial_t u = \nu \Delta u - \nabla p, \tag{44}$$

$$\nabla \cdot u = 0, \tag{45}$$

$$u_1 = a_1 \sin(2\pi x), u_2 = a_2 \cos(2\pi x), \text{ on } \Gamma_1 = \{(x, y) \in \partial\Omega | y = 1\}, \tag{46}$$

$$u = 0 \text{ on } \Gamma_0 = \partial\Omega \setminus \Gamma_1, \tag{47}$$

This is a classical 2D lid driven cavity setup. In the present work we consider, as parameters, the amplitude of the velocity field prescribed on the top of the cavity: $(a_1, a_2) \in [0.25, 1]^2$. The observation consists of the set of the average velocity in subsets of the domain. Let Ω be partitioned in m of non-overlapping squares Ω_i (such that $\bigcup_{1 \le i \le m} \Omega_i = \Omega$), the i-th observation at time $\bar{t}$ is the vector:

$$y_i(\bar{t}) = \frac{1}{|\Omega_i|} \int_{\Omega_i} u(x, \bar{t})\, dx, \tag{48}$$

that is to say, we measure the average of the velocity on certain regions. The system was simulated by P2-P1 finite elements (with a number of degrees of freedom of, roughly 2600 for the velocity and 680 for the pressure). In time, we considered an implicit Euler, with $\Delta t = 1/20$.

An example of the solution is shown in Figure 3.

We considered the following configurations for the data assimilation problem:

- $m = \left\{3^2, 4^2, 5^2\right\}$, regions, which corresponds to different degrees of space resolution in the observation.
- We considered two different levels of noise: $W^{-1} = \left\{10^3 I, 10^1 I\right\}$

ε_{TT}	m	W^{-1}	λ_r	BG	$\overline{e}_{var} \cdot 10^{-2}$	$\sigma_{var} \cdot 10^{-2}$	$\overline{e}_{seq} \cdot 10^{-2}$	$\sigma_{seq} \cdot 10^{-2}$
10^{-3}	3^2	10^3	0.39	4.57	(2.56, 1.57)	(1.65, 0.99)	(17.0, 18.0)	(1.29, 1.13)
10^{-3}	3^2	10^1	$3.9\ 10^{-3}$	457	(24.2, 15.4)	(16.0, 11.5)	(95.0, 95.0)	(0.72, 0.66)
10^{-3}	4^2	10^3	0.89	1.85	(1.48, 0.95)	(1.01, 0.81)	(9.70, 10.0)	(0.95, 0.90)
10^{-3}	4^2	10^1	$8.9\ 10^{-3}$	185	(13.3, 8.8)	(8.0, 0.66)	(91.0, 91.0)	(12.0, 0.87)
10^{-3}	5^2	10^3	1.46	1.03	(1.06, 0.75)	(0.60, 0.55)	(6.10, 6.50)	(0.71, 0.69)
10^{-3}	5^2	10^1	$1.46\ 10^{-2}$	103	(10.10, 7.37)	(7.02, 5.70)	(85.8, 87.3)	(1.76, 1.16)
10^{-5}	3^2	10^3	0.39	4.57	(2.23, 1.51)	(1.50, 1.06)	(17.4, 18.4)	(1.29, 1.13)
10^{-5}	3^2	10^1	$3.9\ 10^{-3}$	457	(22.3, 14.3)	(15.9, 10.5)	(95.4, 95.7)	(0.72, 0.66)
10^{-5}	4^2	10^3	0.89	1.85	(1.47, 1.0)	(1.03, 0.67)	(9.68, 10.0)	(0.97, 0.91)
10^{-5}	4^2	10^1	$8.9\ 10^{-3}$	185	(13.50, 9.10)	(9.46, 7.07)	(91.2, 91.8)	(0.99, 0.81)
10^{-5}	5^2	10^3	1.46	1.03	(0.99, 0.69)	(0.66, 0.53)	(5.93, 6.29)	(0.66, 0.64)
10^{-5}	5^2	10^1	146	103	(10.9, 7.97)	(7.39, 5.89)	(85.8, 87.3)	(1.69, 1.03)

TABLE 3 Result of the test case described in Section 4.3, for the different configurations of the data assimilation problem. Due to the graphical layout, we factorised the results for u and p and their values in terms of mean and variance have to be multiplied by 10^{-2}.

- The value of $\varepsilon_{TT} = \left\{10^{-3}, 10^{-5}\right\}$. For the case considered the TT ranks are $(r_1 = 7, r_2 = 2)$ and $(r_1 = 12, r_2 = 2)$ respectively.

- The tests are performed by considering the solutions obtained by simulating the system when $n_s = 100$ random samples of the parameters are drawn from a uniform distribution in the range of the parameters considered for the database.

In Table 3 we report the results in terms of mean and variance for the velocity and the pressure reconstruction. For this case, we observe that the theoretical indicators of the system observability are roughly the same when we consider $\varepsilon_{TT} = 10^{-3}$ or $\varepsilon_{TT} = 10^{-5}$, the differences being in the further digits. Moreover, the time history considered in relatively short. Since the number of measurements in time is low, the sequential method has a performance which is less satisfactory (the error being large in the first time steps). Concerning the other properties of the system such as the behaviour with respect to the model error and the amount of measurements m considered, we observe analogous trends with respect to the ones commented for the two previous numerical experiments.

General remarks on the results.

The results commented in this section show that, albeit the fact that the systems considered are quite different in terms of nature (kind of non-linearities, dynamics of the solution), the method has a uniform, stable behaviour. The differences observed between the variational and the sequential method are expected. Furthermore, the theoretical quantities based on the spectrum of the Observability matrix provide a good indication on the error made in the state estimation.

5 | CONCLUSION

In the present work we presented both a variational and a sequential formulations of the state estimation problem, which leverage the Tensor Train format representation of a set of parametric solutions of a system. Given the expression of the tensor train solution, it is possible to construct the observability matrix, whose spectrum provides an insight on the stability of the state estimation problem. The interpretation of the Tensor Train format used in the present work as a dynamical basis approximation makes it possible to naturally derive a sequential state estimation formulation. Several numerical results were proposed to assess the behaviour of the method. The main perspectives consist in extending the present approach to stochastic inverse problems, and to apply it to more realistic large scale systems.

Acknowledgments:

The author acknowledges support from ANR grant ADAPT 18-CE46-0001.

References

1. Lahoz BKW, Menard R. *Data assimilation*. Springer . 2010.

2. Chernousko FL. *State estimation for dynamic systems*. CRC Press . 1993.

3. Wouwer AV, Zeitz M. State estimation in distributed parameter systems. *Control Systems, Robotics and Automation–Volume XIV: Nonlinear, Distributed, and Time Delay Systems-III* 2009: 92.

4. Simon D. *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons . 2006.

5. Lei B, Xu G, Feng M, et al. *Classification, Parameter Estimation and State Estimation*. Wiley Online Library . 2017.

6. Bui-Thanh T, Damodaran M, Willcox K. Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition. *AIAA journal* 2004; 42(8): 1505–1516.

7. Robert C, Durbiano S, Blayo E, Verron J, Blum J, Le Dimet FX. A reduced-order strategy for 4D-Var data assimilation. *Journal of Marine Systems* 2005; 57(1-2): 70–82.

8. Buffoni M, Camarri S, Iollo A, Lombardi E, Salvetti MV. A non-linear observer for unsteady three-dimensional flows. *Journal of Computational Physics* 2008; 227(4): 2626–2643.

9. Kaercher M, Boyaval S, Grepl MA, Veroy K. Reduced basis approximation and a posteriori error bounds for 4D-Var data assimilation. *Optimization and Engineering* 2018; 19(3): 663–695.

10. Aretz-Nellesen N, Grepl MA, Veroy K. 3D-VAR for parameterized partial differential equations: a certified reduced basis approach. *Advances in Computational Mathematics* 2019; 45(5): 2369–2400.

11. Stefanescu R, Sandu A, Navon IM. POD/DEIM Strategies for reduced data assimilation systems. *Journal of Computational Physics* 2014.

12. Fang F, Pain C, Navon I, et al. A POD reduced-order 4D-Var adaptive mesh ocean modelling approach. *International Journal for Numerical Methods in Fluids* 2009; 60(7): 709–732.

13. Du J, Navon I, Zhu J, Fang F, Alekseev AK. Reduced order modeling based on POD of a parabolized Navier–Stokes equations model II: Trust region POD 4D VAR data assimilation. *Computers & Mathematics with Applications* 2013; 65(3): 380–394.

14. Zainib Z, Ballarin F, Fremes S, Triverio P, Jiménez-Juan L, Rozza G. Reduced order methods for parametric optimal flow control in coronary bypass grafts, toward patient-specific data assimilation. *International Journal for Numerical Methods in Biomedical Engineering* 2020: e3367.

15. Freitag MA, Green DL. A low-rank approach to the solution of weak constraint variational data assimilation problems. *Journal of Computational Physics* 2018; 357: 263–281.

16. Fisher M, Gratton S, Gürol S, Trémolet Y, Vasseur X. Low rank updates in preconditioning the saddle point systems arising from data assimilation problems. *Optimization Methods and Software* 2018; 33(1): 45–69.

17. Maday Y, Anthony T, Penn JD, Yano M. PBDW state estimation: Noisy observations; configuration-adaptive background spaces; physical interpretations. *ESAIM: Proceedings and Surveys* 2015; 50: 144–168.

18. Gong H, Maday Y, Mula O, Taddei T. PBDW method for state estimation: error analysis for noisy data and nonlinear formulation. *arXiv preprint arXiv:1906.00810* 2019.

19. Maday Y, Taddei T. Adaptive PBDW approach to state estimation: noisy observations; user-defined update spaces. *SIAM Journal on Scientific Computing* 2019; 41(4): B669–B693.

20. Binev P, Cohen A, Dahmen W, DeVore R, Petrova G, Wojtaszczyk P. Data assimilation in reduced modeling. *SIAM/ASA Journal on Uncertainty Quantification* 2017; 5(1): 1–29.

21. Binev P, Cohen A, Mula O, Nichols J. Greedy algorithms for optimal measurements selection in state estimation using reduced models. *SIAM/ASA Journal on Uncertainty Quantification* 2018; 6(3): 1101–1126.

22. Taddei T. An adaptive parametrized-background data-weak approach to variational data assimilation. *ESAIM: Mathematical Modelling and Numerical Analysis* 2017; 51(5): 1827–1858.

23. Benaceur A. A time-dependent Parametrized Background Data-Weak approach. In: Springer. 2021 (pp. 125–133).

24. Haykin S, deFreitas N. Special issue on sequential state estimation. *Proceedings of the IEEE* 2004; 92(3): 399–400.

25. Brasseur P, Verron J. The SEEK filter method for data assimilation in oceanography: a synthesis. *Ocean Dynamics* 2006; 56(5): 650–661.

26. Evensen G. *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media . 2009.

27. Collin A, Chapelle D, Moireau P. Sequential state estimation for electrophysiology models with front level-set data using topological gradient derivations. In: Springer. ; 2015: 402–411.

28. Bertoglio C, Chapelle D, Fernández MA, Gerbeau JF, Moireau P. State observers of a vascular fluid–structure interaction model through measurements in the solid. *Computer Methods in Applied Mechanics and Engineering* 2013; 256: 149–168.

29. Moireau P. A discrete-time optimal filtering approach for non-linear systems as a stable discretization of the Mortensen observer. *ESAIM: Control, Optimisation and Calculus of Variations* 2018; 24(4): 1815–1847.

30. Hoteit I, Triantafyllou G, Korres G. Using low-rank ensemble Kalman filters for data assimilation with high dimensional imperfect models. *JNAIAM* 2007; 2(1-2): 67–78.

31. Pagani S, Manzoni A, Quarteroni A. Efficient state/parameter estimation in nonlinear unsteady PDEs by a reduced basis ensemble Kalman filter. *SIAM/ASA Journal on Uncertainty Quantification* 2017; 5(1): 890–921.

32. Leroux R, Chatellier L, David L. Time-resolved flow reconstruction with indirect measurements using regression models and Kalman-filtered POD ROM. *Experiments in Fluids* 2018; 59(1): 1–27.

33. Xiao D, Du J, Fang F, Pain C, Li J. Parameterised non-intrusive reduced order methods for ensemble Kalman filter data assimilation. *Computers & Fluids* 2018; 177: 69–77.

34. Popov AA, Mou C, Sandu A, Iliescu T. A multifidelity ensemble Kalman filter with reduced order control variates. *SIAM Journal on Scientific Computing* 2021; 43(2): A1134–A1162.

35. Rörich A, Werthmann TA, Göddeke D, Grasedyck L. Bayesian inversion for electromyography using low-rank tensor formats. *Inverse Problems* 2021; 37(5): 055003.

36. Batselier K, Chen Z, Wong N. A Tensor Network Kalman filter with an application in recursive MIMO Volterra system identification. *Automatica* 2017; 84: 17–25.

37. Dogariu LM, Paleologu C, Benesty J, Stanciu CL, Oprea CC, Ciochină S. A Kalman Filter for Multilinear Forms and Its Connection with Tensorial Adaptive Filters. *Sensors* 2021; 21(10): 3555.

38. Wang X, Wu Y, Zhuang D, Sun L. Low-Rank Hankel Tensor Completion for Traffic Speed Estimation. *arXiv preprint arXiv:2105.11335* 2021.

39. Oseledets IV. Tensor-train decomposition. *SIAM Journal on Scientific Computing* 2011; 33(5): 2295–2317.

40. Bigoni D, Engsig-Karup AP, Marzouk YM. Spectral tensor-train decomposition. *SIAM Journal on Scientific Computing* 2016; 38(4): A2405–A2439.

41. Gorodetsky A, Karaman S, Marzouk Y. A continuous analogue of the tensor-train decomposition. *Computer methods in applied mechanics and engineering* 2019; 347: 59–84.

42. Dolgov SV, Khoromskij BN, Oseledets IV. Fast solution of parabolic problems in the tensor train/quantized tensor train format with initial application to the Fokker–Planck equation. *SIAM Journal on Scientific Computing* 2012; 34(6): A3016–A3038.

43. Zhang Z, Yang X, Oseledets IV, Karniadakis GE, Daniel L. Enabling high-dimensional hierarchical uncertainty quantification by ANOVA and tensor-train decomposition. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 2014; 34(1): 63–76.

44. Dolgov S, Khoromskij BN, Litvinenko A, Matthies HG. Polynomial chaos expansion of random coefficients and the solution of stochastic partial differential equations in the tensor train format. *SIAM/ASA Journal on Uncertainty Quantification* 2015; 3(1): 1109–1135.

45. Nouy A. Low-rank tensor methods for model order reduction. *arXiv preprint arXiv:1511.01555* 2015.

46. Bachmayr M, Dahmen W. Adaptive low-rank methods: problems on Sobolev spaces. *SIAM Journal on Numerical Analysis* 2016; 54(2): 744–796.

47. Corona E, Rahimian A, Zorin D. A tensor-train accelerated solver for integral equations in complex geometries. *Journal of Computational Physics* 2017; 334: 145–169.

48. Loukrezis D, Römer U, Casper T, Schöps S, De Gersem H. High-dimensional uncertainty quantification for an electrothermal field problem using stochastic collocation on sparse grids and tensor train decompositions. *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields* 2018; 31(2): e2222.

49. Ham FM, Brown RG. Observability, eigenvalues, and Kalman filtering. *IEEE Transactions on Aerospace and Electronic Systems* 1983(2): 269–273.

50. Wolkowicz H, Styan GP. Bounds for eigenvalues using traces. *Linear algebra and its applications* 1980; 29: 471–506.

51. Bai Z, Golub GH. Bounds for the trace of the inverse and the determinant of symmetric positive definite matrices. *Annals of Numerical Mathematics* 1996; 4: 29–38.

□