Assessing the performance of data assimilation algorithms which employ linear error feedback

Article

Accepted Version

Assessing the Performance of Data Assimilation Algorithms which employ Linear Error Feedback

Noeleene Mallia-Parfitt[1] and Jochen Bröcker[1]

School of Mathematical, Physical and Computational Sciences, University of Reading, Whiteknights, PO BOX 220, Reading, RG6 6AX, United Kingdom

(Dated: 4 October 2016)

Data assimilation means to find an (approximate) trajectory of a dynamical model that (approximately) matches a given set of observations. A direct evaluation of the trajectory against the available observations is likely to yield a too optimistic view of performance, since the observations were already used to find the solution. A possible remedy is presented which simply consists of estimating that optimism, thereby giving a more realistic picture of the 'out of sample' performance. Our approach is inspired by methods from statistical learning employed for model selection and assessment purposes in statistics. Applying similar ideas to data assimilation algorithms yields an operationally viable means of assessment. The approach can be used to improve the performance of models or the data assimilation itself. This is illustrated by optimising the feedback gain for data assimilation employing linear feedback.

Data assimilation means to find an (approximate) trajectory of a dynamical model that (approximately) matches a given set of observations. A fundamental problem of data assimilation experiments in atmospheric contexts is that there is no possibility of replication, that is, truly "out of sample" observations from the same underlying flow pattern but with independent observational errors are typically not available. A direct evaluation against the available observations is likely to yield unrealistic results though, since the observations were already used to find the solution. A possible remedy is presented which simply consists of estimating that optimism, thereby giving a more realistic picture of the 'out of sample' performance. The approach is particularly simple when applied to data assimilation algorithms employing linear error feedback. A realistic performance assessment is obtained by comparing with the true trajectory. In addition this method provides a simple and efficient means to determine the optimal feedback gain operationally since it only requires known quantities to be calculated. The optimality of this gain is verified numerically. Further, we illustrate theoretical results which demonstrate that in linear systems with gaussian perturbations, the feedback thus determined will approach the optimal (Kalman) gain in the limit of large observational windows (the proof will be given elsewhere).

I. INTRODUCTION

Data Assimilation involves the incorporation of observational data into a numerical model to produce a model state that accurately describes the observed reality. This procedure uses an explicit dynamical model for the time evolution of the observed reality. The results produced by data assimilation must satisfy two requirements. Firstly they must be close to the observations up to a certain degree of accuracy and secondly they should be consistent with the dynamical model to a certain degree of accuracy. In other words, the trajectory produced by data assimilation must be close to the observations and it must be close to being an orbit of the model.

Once the observations have been used to estimate these trajectories, they should not be used to evaluate the performance of the model (at least not without precaution) as this might give unrealistic results. Simply comparing the observations with the output of the data assimilation scheme will provide an overly optimistic picture of performance. Moreover, assessing the performance using this tracking error could easily be cheated. An example is taking the output to be the observations themselves.

As we will see in Section II, a more realistic evaluation of the performance needs to take into account that the output and the observation errors are correlated. To this end, we investigate the concept of out-of-sample error from statistics and adapt it to the problem of data assimilation. In statistics, estimates of the out-of-sample error are used to measure how well a statistical model, after fitting it to observations, generalises to unseen data[1,2]. Although the concept of the out-of-sample error is a very general one, actual implementations differ considerably depending on the structure of the estimation problem. Further, a fundamental assumption often made in statistics is that the observations (conditionally on the explanatory variables) are independent and identically distributed. In the case of linear regression models, a popular statistic for model selection in statistical learning is the Cp statistic[3,4]. Other examples are Akaike's Information Criterion (AIC) or the Bayesian Information Criterion (BIC). These concepts differ in terms of precise interpretation and range of applicability.

The aim of this paper is to provide similar tools in the context of data assimilation. The underlying problem is essentially the same as in statistics. Suppose a time series of observations has been assimilated into a dynamical model. Then the output should be close to hypothetical observations from the same flow patterns but with independent errors. If the results are not close to these hypothetical observations, then this can only mean that the model is in fact not able to explain the dynamics underlying the observations. The out-of-sample error should be a measure of how close the output will be to such hypothetical observations. Although observations from the same flow pattern but with independent errors are typically not available in practice, we show that the out-of-sample error can be estimated using terms that are operationally available. Specifically we show that the out-of-sample error is the sum of the tracking error and a term which we call the optimism. This optimism gives us a representation of how the model and observations depend on each other and it quanties how much the tracking error misestimates the out-of-sample error. The derived expression is reminiscent of the Cp statistic used in model selection in statistical learning[3,4]. We show that the optimism takes a very simple form if we assume that the model employs a linear error feedback. There are many data assimilation algorithms that implement such a feedback[5]. More details and references concerning such algorithms can be found in section II.

Wahba *et al.*[6] apply the ideas of out-of-sample performance to data assimilation for linear systems. In this publication they use generalised cross validation to get an estimate of the true performance. The key equation in this paper is equation (2.11) which is similar to equation (7.46) in Hastie, Tibshirani, and Friedman[3] with the new aspect being the stochastic approximation to the denominator. The results presented in Wahba *et al.*[6] however, apply only in a linear context. As it will be shown, the analysis presented in our paper does not require linear models but merely linear error feedback.

We stress that although in terms of the problem we are addressing there is a strong similarity between statistics and data assimilation, our analysis will be different. For instance, although the data assimilation uses linear error feedback, the dependence of the output on the observations as a whole is nonlinear, due to the nonlinearity of the dynamic model. Further, the observations are not independent. The derivation of the Cp statistic, AIC, BIC and many other related concepts used in statistics however assumes either linearity, independence or both (see Hastie, Tibshirani, and Friedman[3], Sec 7.4).

We demonstrate the usefulness of our approach with three numerical examples. In all three cases, we consider a simple data assimilation scheme by means of filtering with a linear error feedback. A persistent problem in practice is to find a suitable feedback. The feedback acts as a coupling between the true dynamics and the model. If the coupling is too weak the stability of the system cannot be guaranteed while if the coupling is too strong, results deteriorate because the noise will be overly attenuated. Striking the right balance requires a reliable assessment of the performance which is provided by our estimate of the out-of-sample performance. Note that this is relevant even in the case of linear systems with gaussian perturbations as computing the theoretically optimal Kalman Gain requires knowledge of the dynamical noise which is usually not available in practice. Our experiments demonstrate that the technique can be used in situations where the feedback gain matrix is completely unspecified and also in situations where it has a pre-determined structure but contains unknown parameters.

In section II we define the tracking error, out-of-sample error and the optimism. These considerations are valid for any data assimilation algorithm in the case of additive observational noise. We also consider general data assimilation algorithms which employ linear error feedback and determine an analytical expression for the optimism. Section III contains several numerical experiments. In Section III A we apply the methodology to a linear system with gaussian perturbations. We minimise an estimate of the out-of-sample error to determine a feedback gain. We then compare this with the asymptotic Kalman Gain which is known to be optimal in this situation. Our experiments suggest that the gain determined numerically agrees with the optimal Kalman Gain in the limit of large observation windows. We discuss a theoretical result which confirms this finding. Next we consider a situation in which the data assimilation algorithm is constrained to have poles in certain locations which determines the gain up to a single parameter. This parameter is determined by minimising an estimate of the out-of-sample error.

The remaining experiments consider non linear systems. In Section III B we consider a system in Lur'e form. These systems are special in that, despite being non linear, they permit observers with linear error dynamics. Again a linear feedback is used and we show how an estimate of the out-of-sample error can be used to determine the feedback. The performance of this feedback is assessed numerically by considering the error between the reconstructed and the true orbit. Our results indicate that this strategy of choosing the feedback gives close to optimal performance. Repeating the experiment with the Lorenz '96 system in Section III C confirm the results.

II. TRACKING ERROR, OUTPUT ERROR AND OPTIMISM IN DATA ASSIMILATION

Data assimilation is the procedure by which trajectories $\{z_n \in \mathbb{R}^D, n = 1, \ldots, N\}$ (in some state space which we take to be $\mathbb{R}^D$) are computed with the help of a dynamical model and observations, $\{\eta_n, n = 1, \ldots, N\}$. These trajectories should reproduce the observations up to some degree of accuracy for all $n = 1, \ldots, N$. We express this latter part of the procedure formally as: The output $y_n = h(z_n)$ is close to the observations $\{\eta_n, n = 1, \ldots, N\}$ up to some degree of accuracy, where $h : \mathbb{R}^D \to \mathbb{R}^d$ is a function which maps the model's state space into the observation space. This function is usually part of the problem specification. The exact structure of the model and of h is not important at this stage.

Suppose we have observations $\{\eta_n \in \mathbb{R}^d, n = 1, \ldots, N\}$ from some real world dynamical phenomenon. We assume η_n can be written as

$$\eta_n = \zeta_n + \sigma r_n \tag{1}$$

where $\{\zeta_n, n = 1, \ldots, N\}$ are unknown quantities representing the desired signal, and $\sigma \in \mathbb{R}^{d\times d}$ is the observational error standard deviation. We assume that $\{\zeta_n, n = 1, \ldots, N\}$ can be modelled as some stochastic process. The observation errors or noise, $\{r_n, n = 1, \ldots, N\}$ are assumed to be independent with mean $\mathbb{E}r_n = 0$ and variance $\mathbb{E}r_n r_n^T = \mathbb{1}$ and they are independent of $\{\zeta_n, n = 1, \ldots, N\}$.

Deviation of the output from the observations can be quantified by means of the tracking error,

$$E_T = \mathbb{E}[y_n - \eta_n]^2. \tag{2}$$

The tracking error though is not a very useful performance measure of data assimilation approaches. It is not difficult to design algorithms which achieve zero tracking error by simply using the observations as output, that is any DA algorithm which satisfies $y_n = \eta_n$, $n = 1, \ldots, N$ achieves optimal performance with respect to E_T as a performance measure.

A performance measure which is much harder to hedge is the output error

$$E_O = \mathbb{E}[y_n - \zeta_n]^2. \tag{3}$$

A useful relation between E_O and E_T can be established. Substituting the expression (1) for the observations into (2) and expanding, we get

$$E_T = \mathbb{E}[y_n - \eta_n]^2 = \mathbb{E}[y_n - \zeta_n]^2 + \mathrm{tr}(\sigma^T\sigma) - 2\mathrm{tr}(\sigma\mathbb{E}[r_n y_n^T]) \tag{4}$$

since ζ_n and r_n are independent. The notation 'tr' denotes the trace of the matrix.

We re-write this as

$$E_O + \mathrm{tr}(\sigma^T\sigma) = \mathbb{E}[y_n - \eta_n]^2 + 2\mathrm{tr}(\sigma\mathbb{E}[r_n y_n^T]). \tag{5}$$

The term $2\sigma\mathbb{E}[r_n y_n^T]$ is called the *optimism*. The optimism should be understood as a correlation between r_n and y_n, where y_n depends on $\{r_k, k = 1, \ldots, N\}$. It is a measure of how much the tracking error misestimates the output error. We will argue that both the optimism and the tracking error (i.e the first term on the right hand side of (5) can be estimated using operationally available quantities. This will give us a handle on the output error which is, as we have argued, directly related to the true performance of the data assimilation.

The quantity $E_O + \sigma^2$ can be interpreted as an "Out-of-sample error" as follows: Define hypothetical observations

$$\eta'_n = \zeta_n + r'_n, \quad n = 1, \ldots, N \tag{6}$$

where $\{\zeta_n, n = 1, \dots, N\}$ is as before, $\{r'_n, n = 1, \dots, N\}$ is a process with the same distribution as $\{r_n, n = 1, \dots, N\}$ but independent from it. Then the out-of-sample error is the error between $\{y_n, n = 1 \dots, N\}$ and $\{\eta'_n, n = 1, \dots, N\}$, which can be written as

$$\mathbb{E}[y_n - \eta'_n]^2 = E_O + \sigma^2. \tag{7}$$

The key difference between the tracking error and the out-of-sample error is the absence of correlation between $\{y_n, n = 1 \dots, N\}$ and $\{r'_n, n = 1, \dots, N\}$ in the latter, which is precisely the optimism.

Equation (5) shows that the tracking error augmented with further terms, can be a useful measure of performance. Further the tracking error and optimism are relatively easy to estimate. In our experiments we will estimate the tracking error through an empirical average, namely

$$\hat{E}_T = \frac{1}{N} \sum_{k=1}^{N} (y_k - \eta_k)^2. \tag{8}$$

Estimates of the optimism will be discussed next.

We will first calculate a general expression for the optimism for data assimilation schemes which employ a linear error feedback. Most operational data assimilation schemes work in cycles over time. The *background field*, $\hat{z}_n$, is computed at the start of each cycle and usually it is based on information from previous cycles. Since any cycle uses observations available up to that point, the background field at time n only depends on $\eta_1, \dots, \eta_{n-1}$. Nonetheless, the background field $\hat{z}_n$ is supposed to be a first guess of the the state of the system at time n.

In this paper we consider data assimilation algorithms which combine the new observation and background through a relationship of the form

$$z_n = \hat{z}_n + \mathbf{K}_n(\eta_n - h(\hat{z}_n)) \tag{9}$$

where $\mathbf{K}_n$ is a $D \times d$ matrix and can depend on $\eta_1, \dots, \eta_{n-1}$ but not on η_n. As before, the mapping $h : \mathbb{R}^D \to \mathbb{R}^d$, maps points from model state space to observation space. The modified background, z_n, is referred to as the *analysis*.

The matrix $\mathbf{K}_n$ is the error feedback gain. Equation (9) tells us that the analysis has a linear dependence on the current observation, η_n and it depends on the previous observations through $\mathbf{K}_n$ and $\hat{z}_n$. Data assimilation schemes that fall into the presented approach include

Successive Correction Method (SCM)[7,8]; Optimal Interpolation (OI)[9]; 3D-Var[10,11]; Kalman Filter variants,[12] and certain Synchronisation approaches. Synchronisation between dynamical systems has been studied for some time, see for example Pikovsky, Rosenblum, and Kurths[13]; Huijberts, Nijmeijer, and Pogromsky[14]; Boccaletti *et al.*[15]. Synchronisation in the setting of data assimilation has also been studied, see Bröcker and Szendro[16]; Szendro, Rodrìguez, and Lopez[17]; Yang, Baker, and Li[18]. These methods differ only on the approach they take to calculate the background $\hat{z}_n$ and the matrix $\mathbf{K}_n$[5].

We now consider the optimism as in (5) in the context of DA scheme with linear feedback as in (9). We assume that the function $h(x_n)$ is linear so that $h(x_n) = \mathbf{H}x_n$, where $\mathbf{H}$ is a $d \times D$ matrix. Then,

$$
\begin{aligned}
\mathbb{E}[r_n y_n^T] &= \mathbb{E}[r_n(\mathbf{H}z_n)^T] = \mathbb{E}[r_n z_n^T]\mathbf{H}^T && (10)\\
&= \mathbb{E}[r_n\{(\mathbb{1} - \mathbf{K}_n\mathbf{H})\hat{z}_n + \mathbf{K}_n(\zeta_n + \sigma r_n)\}^T]\mathbf{H}^T && (11)\\
&= \mathbb{E}[r_n((\mathbb{1} - \mathbf{K}_n\mathbf{H})\hat{z}_n)^T]\mathbf{H}^T \\
&\quad + \mathbb{E}[r_n(\mathbf{K}_n\zeta_n)^T]\mathbf{H}^T + \mathbb{E}[r_n(\mathbf{H}\mathbf{K}_n\sigma r_n)^T] && (12)\\
&= \mathbb{E}[r_n r_n^T \sigma^T \mathbf{K}_n^T]\mathbf{H}^T && (13)\\
&= \mathrm{tr}(\mathbb{E}[r_n r_n^T]\sigma^T \overline{\mathbf{K}}_n^T \mathbf{H}^T) && (14)
\end{aligned}
$$

where $\overline{\mathbf{K}}_n = \mathbb{E}[\mathbf{K}_n]$. The first two equalities, (10) and (11), are obtained by substituting the relevant information while (12) is obtained by simply expanding the previous equation. The derivation from (12) to (13) requires some explanation. Notice first that only the third term of (12) survives. The first term is equal to zero because $\hat{z}_n$ and $\mathbf{K}_n$ are uncorrelated with r_n. The second term is also equal to zero because ζ_n is independent of r_n and because the coupling matrix $\mathbf{K}_n$ depends on the observations $(\eta_1 \ldots \eta_{n-1})$ and thus is uncorrelated with r_n.

Therefore, we are only left with the third term of (12) in (13). Since $\mathbb{E}(r_n r_n^T) = \mathbb{1}$, (14) implies that

$$2\mathrm{tr}(\sigma\mathbb{E}[r_n y_n^T]) = 2\mathrm{tr}(\sigma \cdot \sigma^T \overline{\mathbf{K}}_n^T \mathbf{H}^T). \tag{15}$$

In the case when $d = 1$, which is the case we consider in the numerical experiments later, this reduces to

$$2\sigma\mathbb{E}[y_n r_n] = 2\mathbf{H}\overline{\mathbf{K}}_n\sigma^2. \tag{16}$$

We recall that the assumptions necessary to derive this formula are a linear observation operator, r_n is independent of $\{\eta_1, \ldots, \eta_{n-1}\}$, $\mathbb{E}r_n = 0$, $\mathbb{E}r_n r_n^T = \mathbb{1}$ and $\mathbf{K}_n$ depends only on the observations $(\eta_1, \ldots, \eta_{n-1})$.

In our numerical experiments we approximate the expected value of a random variable by the empirical mean. In particular E_T is replaced by its empirical average in (5), resulting in the following estimate for E_O for all subsequent numerical experiments (in which $\mathbf{K_n}$ is in fact constant):

$$\hat{E}_O = \hat{E}_T + \frac{1}{N}\sum_{n=1}^{N} 2\sigma^2 \mathrm{tr}(\overline{\mathbf{K}}_n^T \mathbf{H}^T) - \sigma^2. \tag{17}$$

Let us briefly digress on how the background $\hat{z}_n$ and $\mathbf{K}_n$ might be calculated in the context of synchronisation, although this is in fact irrelevant for the optimism. Suppose that the reality is given by the non linear dynamical system

$$\begin{aligned} x_{n+1} &= \tilde{f}(x_n) \\ \zeta_n &= \tilde{h}(x_n) \\ \eta_n &= \zeta_n + \sigma r_n \end{aligned} \tag{18}$$

where $x_n \in \mathbb{R}^D$ is referred to as the state and $\zeta_n \in \mathbb{R}^d$ are the true observations. For this non linear dynamical system we construct a sequential scheme

$$\begin{aligned} \hat{z}_{n+1} &= f(z_n) \\ z_{n+1} &= \hat{z}_{n+1} - \mathbf{K}_n(h(\hat{z}_{n+1}) - \eta_{n+1}) \\ y_n &= h(z_n) \end{aligned} \tag{19}$$

where $\mathbf{K}_n$ is a $D \times d$ coupling matrix which depends on the observations $\eta_1, \ldots \eta_n$ but not on η_{n+1}; and y_n is the model output where we hope that $y_n \cong \zeta_n$. Here f and h are approximations to the functions $\tilde{f}$ and $\tilde{h}$, respectively. The coupling introduced in this scheme creates a linear feedback, in the sense that the error between $y_n = h(\hat{z}_n)$ and the observations η_n is fed back into the model.

Synchronisation refers to a situation in which, due to coupling, the error $y_n - \eta_n$ becomes small asymptotically irrespective of the initial conditions for the model[13]. Often a control theoretic approach is taken to determine conditions which guarantee the model output, $y_n = h(z_n)$, converging to the observations, η_n or even z_n converging to x_n (strictly speaking, the difference converging to zero; note that this can only be expected in case of noise free observations).

It has been highlighted above that the tracking error is not an ideal measure of performance; however the output error is and moreover, it can be calculated using terms that are readily available. An important question that arises in operational practice is to how to choose the gain matrix $\mathbf{K}$. The numerical experiments detailed below consider different conditions under which to select the appropriate coupling matrix to use in the assimilation. For the first linear experiment we consider arbitrary candidates for the gain matrix, while for the second linear experiment we consider gains that guarantee a certain structure of the system matrix (or more specifically the poles thereof).

III. NUMERICAL EXPERIMENTS

We now demonstrate the usefulness of our approach with three numerical examples. In Section III A we present the methodology for a linear system with gaussian perturbations. We minimise an estimate of the out-of-sample error to determine a feedback gain and compare this with the asymptotic Kalman Gain which is known to be optimal in this situation.

The remaining two experiments concern nonlinear systems. In Section III B we present numerical results for the Hénon Map and in Section III C results are established for the Lorenz'96 System. Again a linear feedback is used and we show how an estimate of the out-of-sample error can be used to determine the feedback.

There is some repetition in the obtained results, however this repetition validates our approach across different experiments. The three systems we consider all use a data assimilation scheme that employs linear error feedback. However the underlying systems in each are different; one is linear, one is in Lur'e form and one is nonlinear. The similarities in the results confirm that our methodology applies to many different dynamical systems.

A. Numerical Experiment 1: Linear Map

In this first linear example the following experimental setup was used: The reality is given by

$$x_{n+1} = \underbrace{\begin{bmatrix} -1 & 10 \\ 0 & 0.5 \end{bmatrix}}_{\mathbf{A}} x_n + \rho q_{n+1} \tag{20}$$

with corresponding observations

$$\eta_n = \mathbf{H}x_n + \sigma r_n \tag{21}$$

where $\mathbf{H} = [1 \quad 0]$, $\zeta_n = \mathbf{H}x_n$ and $\rho \in \mathbb{R}^{D\times D}$ is the model error standard deviation. We assume that the model and observations are corrupted by random noise. For these experiments we have $x_n \in \mathbb{R}^2$ and $\eta_n \in \mathbb{R}$. The model errors, q_n, are assumed to be serially independent errors with mean $\mathbb{E}q_n = 0$ and variance $\mathbb{E}q_n q_n^T = \mathbb{1}$.

We set up an observer analogous to our sequential scheme (19),

$$z_{n+1} = \hat{z}_{n+1} + \mathbf{K}_n(\eta_{n+1} - \mathbf{H}\hat{z}_{n+1}), \qquad y_n = \mathbf{H}z_n \tag{22}$$

where

$$\hat{z}_{n+1} = \underbrace{\begin{bmatrix} -1 & 10 \\ 0 & 0.5 \end{bmatrix}}_{\text{A}} z_n. \tag{23}$$

In this case the model is coupled to the observations through a linear coupling term which is dependent on the difference between the actual output and the expected output value based on the next estimate of the state. For these experiments we will take the coupling matrix $\mathbf{K}_n$ to be constant so from here on we write $\mathbf{K}_n = \mathbf{K}$.

The error dynamics in this linear example are given by

$$\begin{aligned} e_{n+1} &= x_{n+1} - z_{n+1} \\ &= (\mathbf{A} - \mathbf{KHA})e_n + \mathbf{K}r_{n+1} - (\mathbb{1} - \mathbf{KH})q_{n+1}. \end{aligned} \tag{24}$$

Since the noisy part of the error dynamics (Eq. 24) is stationary, synchronisation can be guaranteed if the eigenvalues of the matrix $(\mathbf{A} - \mathbf{KHA})$ all lie within the unit circle. Synchronisation here means that the error dynamics is asymptotically stationary with finite covariance. To achieve this, we use a result from control theory, for which we need a few definitions. Let $\mathbf{HA} = \mathbf{C}$ so that the error dynamics are described by the system matrix $(\mathbf{A} - \mathbf{KC})$. A pair of matrices $(\mathbf{A}, \mathbf{C})$ is called *observable* if the observability matrix

$$\mathcal{O} = [\mathbf{C} \quad \mathbf{CA} \quad \mathbf{CA}^2 \quad \ldots \quad \mathbf{CA}^{D-1}]^T \tag{25}$$

has full rank. If this condition holds then the poles of the matrix $(\mathbf{A} - \mathbf{KC})$ can be placed anywhere in the complex plane by proper selection of $\mathbf{K}$. In particular they can be placed within the unit circle[19].

In our example, $x_n \in \mathbb{R}^2$ so our observability matrix is

$$\mathcal{O} = [\mathbf{HA} \quad \mathbf{HA}^2]^T. \tag{26}$$

It is straightforward to check that the linear system we are working with here is observable even though $\mathbf{A}$ is not stable.

It is well known in Kalman Filter theory (see for example Anderson and Moore [20]) that the optimal gain matrix $\boldsymbol{\kappa}_n$ for a linear filter (in the sense of giving least error covariance) is the Kalman Gain which is defined by

$$\boldsymbol{\kappa}_n = \Sigma_n \mathbf{H}^T (\mathbf{H}\Sigma_n \mathbf{H}^T + \sigma^2)^{-1} \tag{27}$$

where Σ_n is the error covariance matrix defined by $\Sigma_n = \mathbb{E}[(\hat{z}_n - x_n)(\hat{z}_n - x_n)^T]$ and expressed by the following recursive equation,

$$\Sigma_n = \mathbf{A}(\Sigma_n - \Sigma_n \mathbf{H}^T (\mathbf{H}\Sigma_n \mathbf{H}^T + \sigma^2)^{-1} \mathbf{H}\Sigma_n)\mathbf{A}^T + \rho^2 \cdot \mathbb{1}. \tag{28}$$

Kalman Filter theory states that for n large, the error covariance Σ_n converges to Σ_∞ which is the solution to

$$\Sigma_\infty = \mathbf{A}[\Sigma_\infty - \Sigma_\infty \mathbf{H}^T (\mathbf{H}\Sigma_\infty \mathbf{H}^T + \sigma^2)^{-1} \mathbf{H}\Sigma_\infty]\mathbf{A}^T + \rho^2 \cdot \mathbb{1}. \tag{29}$$

This in turn implies that the Kalman Gain (27) converges to the asymptotic gain which is defined by

$$\boldsymbol{\kappa}_\infty = \Sigma_\infty \mathbf{H}^T (\mathbf{H}\Sigma_\infty \mathbf{H}^T + \mathbf{R})^{-1} \tag{30}$$

The asymptotic gain, $\boldsymbol{\kappa}_\infty$, is obtained by solving the Discrete Algebraic Riccati Equation (DARE) given by (29) and using the solution to calculate (30). Using Maple's inbuilt DARE solver we were able to find the solution to this equation for the experimental setup described above. The Algebraic Riccati Equation is solved using the method described in Arnold III and Laub [21].

The aim of this experiment is to estimate the optimal gain matrix, $\boldsymbol{\kappa}_\infty$ without referring to the DARE, in particular without knowledge of ρ. We do this by minimising the empirical out-of-sample error with respect to $\mathbf{K}$. In other words, our estimate of $\boldsymbol{\kappa}_\infty$ is the minimiser of $\hat{E}_O$ for a large (but finite) set of observations (paragraph a. below). This strategy is motivated by our previous discussion about the out-of-sample error being an adequate measure of performance. In fact, in the context of linear systems, we can prove (see Appendix A for details) that the out-of-sample error is equivalent (in a certain sense) to the asymptotic covariance of e_n as a measure of performance. We also stress that estimating the optimism only requires knowledge of $\mathbf{A}, \mathbf{H}, \sigma$ but not ρ, the model noise. This is the term that is difficult to determine operationally, so estimating the optimism in an operational situation is possible as all the required terms are readily available. In paragraph b. we discuss a variant of this experiment where the gain matrix is supposed to be optimal under the constraint that the characteristic polynomial has a certain shape.

a. Estimating optimal gain matrix The results obtained in this first experiment are shown in Figure 1. The model noise is iid with $\mathbb{E}q_n = 0$, $\mathbb{E}q_n q_n^T = 1$ and $\rho = 0.01$ while for the observational noise, which was also iid with mean zero and variance one, we used $\sigma = 0.1$. We let n vary between zero and 3.5×10^5. For each n the empirical out-of-sample error was minimised and the minimiser was recorded as an estimate of $\boldsymbol{\kappa}_\infty$. The experiment was repeated for 100 realisations of the observational noise, r_n so that the estimates were different every time. As a measure of accuracy, 90% confidence intervals were constructed. We expect that the estimates converge to the asymptotic gain $\boldsymbol{\kappa}_\infty$ given by the solution of (29,30).

The results obtained are shown in Figure 1. Figure 1(a) shows a plot in blue squares of the quantity $\|\mathbf{K} - \boldsymbol{\kappa}_\infty\| \,/\, \|\boldsymbol{\kappa}_\infty\|$ against n. The figure shows that the gain matrix that minimises the out-of-sample error converges exponentially to the asymptotic gain. Moreover, it is illustarted in Figure 1(c) that the eigenvalues of the matrix $(\mathbf{A} - \mathbf{KHA})$ for each gain minimising the out-of-sample error, converge to the eigenvalues of the matrix $(\mathbf{A} - \boldsymbol{\kappa}_\infty\mathbf{HA})$. Figure 1(c) shows the quantity $\|\lambda - \lambda_\infty\| \,/\, \|\lambda_\infty\|$ against n in blue diamonds, where λ represents the eigenvalues of the matrix $(\mathbf{A} - \mathbf{KHA})$. The convergence of the eigenvalues is also exponential. The values of these eigenvalues confirm that the minimising gains stabilise the system since all of then are within the unit circle.

The remaining two figures in Figure 1 show a log plot of the same information outlined above. Figure 1(b) represents the convergence of the gain matrices while Figure 1(d) shows the same information for the eigenvalues. Both plots are almost straight lines as expected since the convergence has already been noted to be exponential. The addition to these plots are the 90% confidence intervals. As previously stated, the experiment was repeated for 100 realisations of the observational noise and the plotted confidence intervals represents the uncertainty in the numerical experiment. The lower limit of the error bars was taken at the

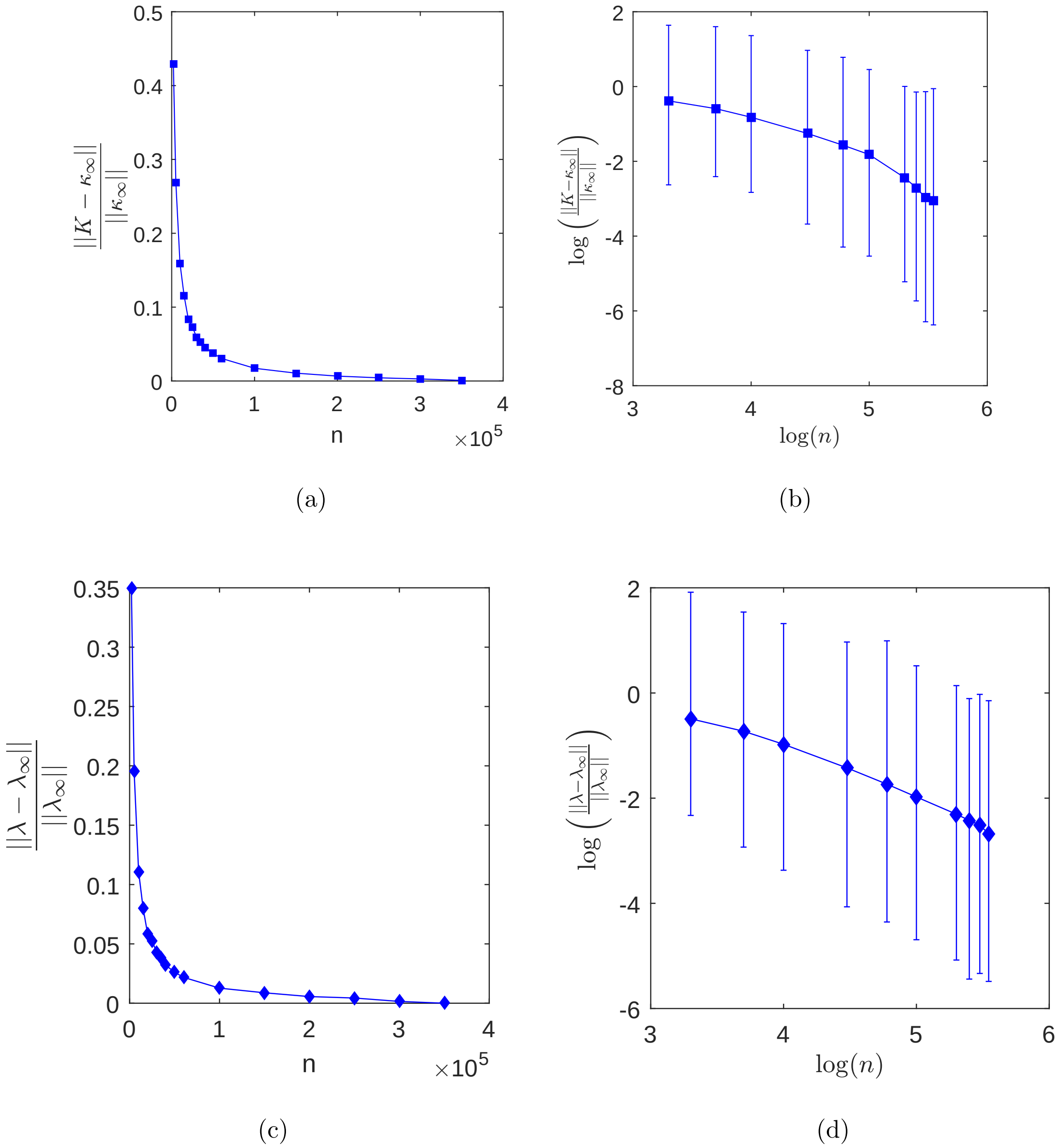
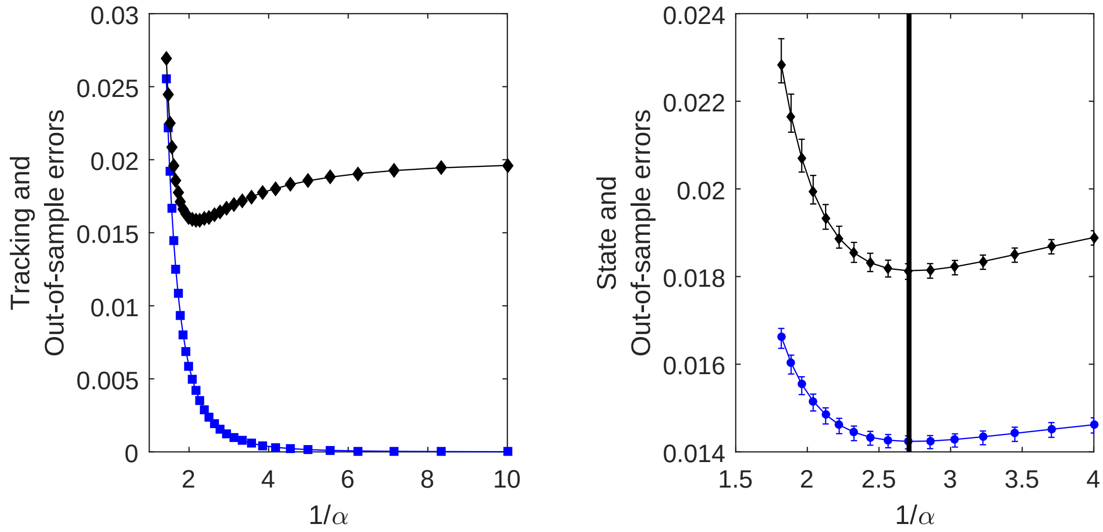


FIG. 1. Figure 1(a) shows the convergence of the gain minimising the out-of-sample error to the asymptotic gain for increasing n. We plot the quantity $\|\mathbf{K} - \boldsymbol{\kappa}_\infty\| \,/\, \|\boldsymbol{\kappa}_\infty\|$ against n in blue squares. Figure 1(b) shows a log plot of the same information with 90% confidence intervals. Figure 1(c) shows the quantity $\|\lambda - \lambda_\infty\| \,/\, \|\lambda_\infty\|$ against n in blue diamonds, where $\lambda = (\lambda_1, \lambda_2)$ represents the eigenvalues of the matrix $(\mathbf{A} - \mathbf{KHA})$. It is evident that the eigenvalues of the matrix $(\mathbf{A} - \mathbf{KHA})$ for each gain minimising the out-of-sample error, converge to the eigenvalues of the matrix $(\mathbf{A} - \boldsymbol{\kappa}_\infty\mathbf{HA})$, with n increasing. Figure 1(d) shows a log plot of the same information with 90% confidence intervals.



(a) (b)

FIG. 2. Figure 2(a) shows a plot of the tracking error in blue squares and the out-of-sample error in black diamonds. The errors are plotted against the inverse of α for $\sigma = 0.1$ and $\rho = 0.01$. Figure 2(b) shows a plot of the out-of-sample error in black diamonds for 100 realisations of the noise r_n with $\sigma = 0.1$ as well as the state error in blue circles. They are displayed for the range of α where the minimum occurs. The error bars in both curves represent 90% confidence intervals. The black vertical line draws attention to the minimum of the out-of-sample error which coincides with the minimum of the state error.

fifth percentile while the upper limit was taken at the 95th percentile thus creating the 90% confidence intervals.

b. Gain Matrix with Symmetric Poles In this part of the linear numerical experiment, we want $(\mathbf{A} - \mathbf{KHA})$ to have a certain characteristic polynomial. Suppose that the desired characteristic equation is given by

$$q(\lambda) = (\lambda + \alpha)(\lambda - \alpha) \tag{31}$$

so that $\lambda_1 = -\lambda_2$ and $|\lambda_1| = |\lambda_2| = \alpha$. The appropriate $\mathbf{K}$ for a desired characteristic polynomial, $q(\lambda)$ of the matrix $(\mathbf{A} - \mathbf{KHA})$ follows from Ackermann's Formula[19] which is given by

$$\mathbf{K} = q(A)\mathcal{O}^{-1}[0\ldots1]^T \tag{32}$$

where $\mathcal{O}$ is the observability matrix defined in (26).

The results obtained from our numerical experiment to test the validity of (16) are shown in Figure 2. Figure 2(a) shows a plot of the tracking error in blue squares and the out-of-sample error in black diamonds. The out-of-sample error calculated via (16) is equivalent to calculating the out-of-sample error explicitly using the output error. We can see that the tracking error tends to zero with decreasing α. This is what we expected and is confirmed by using our analytical expression for the optimism.

It is clear from Figure 2(a) that while the tracking error tends to zero, the out-of-sample error initially decreases and then increases resulting in a well-defined minimum. This is because as the coupling strength increases, the observations are tracked too closely and thus the output adapts too closely to the observations resulting in an increase of the out-of-sample error. On the other hand when α is large and the coupling strength is weak, the observations are tracked poorly resulting in large tracking and out-of-sample errors. In these experiments α was varied between 0 and 1 with the assimilation window taken to be $N = 10000$.

The well defined minimum of the out-of-sample error is also shown in Figure 2(b). Figure 2(b) shows the out-of-sample error in black diamonds for the range of α where the minimum occurs. The figure shows the out-of-sample error for 100 realisations of the observation noise r_n with $\sigma = 0.1$ so that the sample estimate is different each time. The error bars in the plot represent 90% confidence intervals for each value of α. The lower limit of the error bars is taken at the fifth percentile, while the upper limit is taken at the 95th percentile, hence obtained 90% confidence intervals as a measure of accuracy. Some further experiments using different values of σ where carried out however the results are not included here. The results produced were the same as the ones presented in this paper; the only difference was the size of the error bars produced. A smaller value of σ resulted in smaller error bars.

To quantify the variation of the parameter α in this experiment, we considered the following calculation. The mean value of the optimal α plus/minus one standard deviation in this case is

$$\bar{\alpha}^* \pm \sqrt{(\alpha^* - \bar{\alpha}^*)^2} = 0.3698 \pm 0.028. \tag{33}$$

The second plot in Figure 2(b) illustrates the state error. This estimate of the state error is defined by

$$\hat{E}_S = \frac{1}{N} \sum_{n=1}^{N} (z_n - x_n)^2. \tag{34}$$

This is the error that ultimately wants to be analysed and minimised in data assimilation experiments. However, because the model noise (ρq_n) is difficult to determine, we cannot explicitly analyse the state error which is why we consider errors we can calculate, namely the tracking, output or out-of-sample errors. We can plot the state error $\hat{E}_S$ in this example because we have access to it, however in general this is not possible. The vertical line in Figure 2(b) draws attention to the minimum of the out-of-sample error. It is evident that the state error also has a minimum and the plot suggests that the minima of the out-of-sample and the state error are the same. Again, we ran the experiment for 100 realisations and plotted the error bars with 90% confidence intervals.

B. Numerical Experiment 2: Hénon Map

In this experiment, the reality is given by

$$x_{n+1} = \underbrace{\begin{bmatrix} a & b \\ 1 & 0 \end{bmatrix}}_{\mathbf{A}} x_n + c \begin{bmatrix} (\mathbf{H}x_n)^2 \\ 0 \end{bmatrix} + d \tag{35}$$

which for the values $a = 0$, $b = 0.3$, c$=-1.4$, $d = [1 \quad 0]^T$ is the chaotic Hénon Map with corresponding observations

$$\eta_n = \mathbf{H}x_n + \sigma r_n \tag{36}$$

where $\mathbf{H} = [1 \quad 0]$, and $\zeta_n = \mathbf{H}x_n$. The model describing the reality is completely deterministic and we assume that the observations are corrupted by random noise. Notice that we now have a non linear term in the dynamical system. Such systems are said to be in Lur'e form.

Once again we consider data assimilation by means of synchronisation so we set up an observer roughly analogous to our sequential scheme (19) with certain differences,

$$z_{n+1} = \hat{z}_{n+1} + \mathbf{K}_n(\eta_{n+1} - \mathbf{H}\hat{z}_{n+1}), \qquad y_n = \mathbf{H}z_n \tag{37}$$

where

$$\hat{z}_{n+1} = \underbrace{\begin{bmatrix} a & b \\ 1 & 0 \end{bmatrix}}_{\mathrm{A}} z_n + c \begin{bmatrix} \eta_n^2 \\ 0 \end{bmatrix} + d \tag{38}$$

where a, b, c, d are the same as for the reality. In this case as in the first example, the model is coupled to the observations through a linear coupling term which is dependent on the difference between the actual output and the output value expected based on the next estimate of the state. However there is also a non linear coupling introduced here by the presence of η_n^2 in the background term. Note that (16) is still valid nonetheless because $\hat{z}_{n+1}$ is still uncorrelated with r_{n+1}. For these experiments we will take the coupling matrix $\mathbf{K}_n$ to be constant so from here on in we write $\mathbf{K}_n = \mathbf{K}$.

We need to choose the matrix $\mathbf{K}$ appropriately so that we can vary the coupling strength. For illustration purposes consider the error dynamics for the noise-free situation so that $\eta_n = \mathbf{H}x_n$. The error dynamics in this case are given by

$$\begin{aligned} e_{n+1} &= x_{n+1} - z_{n+1} \\ &= x_{n+1} - \hat{z}_{n+1} - \mathbf{KH}(x_{n+1} - \hat{z}_{n+1}) \\ &= (\mathbb{1} - \mathbf{KH})(x_{n+1} - \hat{z}_{n+1}) \\ &= (\mathbf{A} - \mathbf{KHA})(x_n - z_n) \\ &= (\mathbf{A} - \mathbf{KHA})e_n. \end{aligned} \tag{39}$$

The matrix $(\mathbf{A} - \mathbf{KHA})$ is stable even if $\mathbf{K} = \mathbf{0}$. This means that synchronisation occurs even if there is no linear coupling between the model output and observations because of the non linear coupling introduced in the model (38). The eigenvalues for such a case are $\lambda_{1,2} = \pm\sqrt{b}$, where b is as in the matrix $\mathbf{A}$. However, it might be that with noise, the out-of-sample error is not optimal for this coupling and can be improved by some additional linear coupling.

It is straightforward to check that the system we are working with here is observable provided that $b \neq 0$. The appropriate $\mathbf{K}$ for a desired characteristic polynomial, $q(\lambda)$ of the matrix $(\mathbf{A} - \mathbf{KHA})$ again follows from Ackermann's Formula (32). Suppose that the desired characteristic equation is given by

$$q(\lambda) = (\lambda + \alpha)(\lambda - \alpha) \tag{40}$$

so that $\lambda_1 = -\lambda_2$ and $|\lambda_1| = |\lambda_2| = \alpha$. Then by Ackermann's formula we get

$$\mathbf{K} = \begin{bmatrix} 1 - \alpha^2/b \\ a\alpha^2/b^2 \end{bmatrix} \quad \Rightarrow \quad \mathbf{HK} = 1 - \frac{\alpha^2}{b} \tag{41}$$

where $a = 0$ and $b = 0.3$ as in the matrix $\mathbf{A}$. From (41) we see that $\mathbf{HK} = 1$ if $\alpha = 0$. Thus,

$$y_n = \mathbf{H}z_n = (\mathbb{1} - \mathbf{HK})\mathbf{H}\hat{z}_n + \mathbf{HK}\eta_n \to \eta_n, \tag{42}$$

meaning that our data assimilation scheme simply replaces y_n with η_n, implying that the tracking error is zero. In other words, in this example, it is possible to render the eigenvalues of the error dynamics exactly zero and also to obtain zero tracking error. However, the data assimilation is not perfect and the out-of-sample and state errors will not necessarily be small.

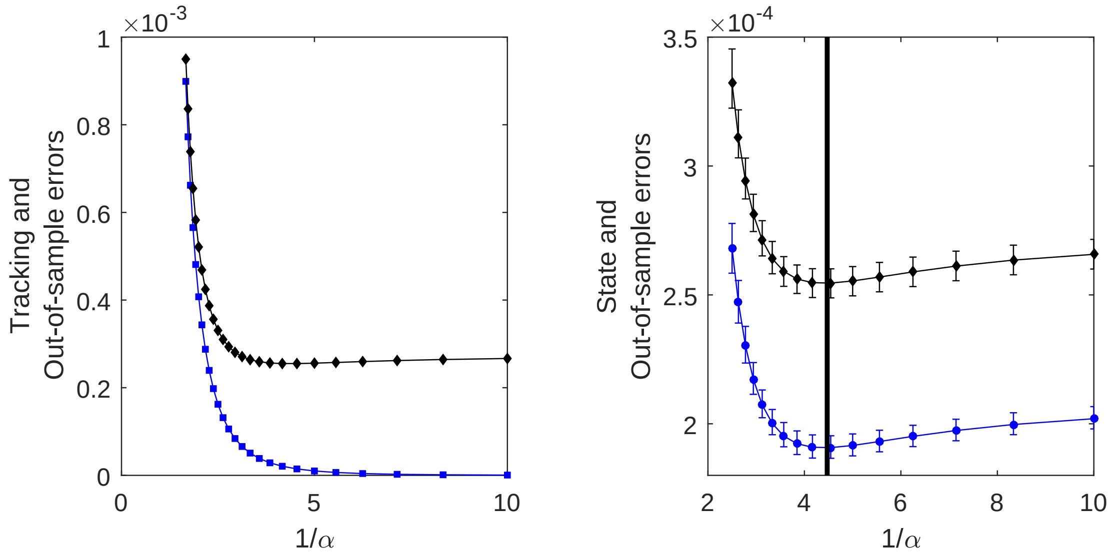
Therefore, from (16) we know that

$$\hat{E}_O = \hat{E}_T - 2\sigma^2\left(1 - \frac{\alpha^2}{b}\right) - \sigma^2. \tag{43}$$

Recall that the aim of this work is to find a way to estimate the out-of-sample error to get a more realistic picture of model performance. We have already determined that when there is no linear coupling (i.e. $\mathbf{K} = \mathbf{0}$) the system is stable and synchronisation occurs. We can see from (43) that this happens when $\alpha = \pm\sqrt{b}$. There are two further cases to consider. When $\alpha^2 > b$ the feedback, due to the linear coupling, is negative. Therefore, in this case we will not be able to improve the out-of-sample error. However as α tends to zero the optimism will increase and be bounded by $2\sigma^2$. Therefore when $\alpha^2 < b$ it may be possible to improve the out-of-sample error and determine a coupling matrix $\mathbf{K} \neq \mathbf{0}$, that minimises the out-of-sample error, to be used in the model. We calculate the errors as we did for the linear numerical example in Section III A.

The results obtained from our numerical experiment to test the validity of (16) are shown in Figure 3. Figure 3(a) shows the tracking error in blue squares and the out-of-sample error in black diamonds. We can see that the tracking error tends to zero with decreasing α. This is what we expected and is confirmed by using our analytical expression for the optimism. In these experiments α was varied between 0 and 1 with the assimilation window taken to be $N = 10000$.

By analysing the expression for the optimism in this case, we see that there is a point where the tracking and out-of-sample errors meet. This happens when $\alpha^2 = b$. To the left of this, when $\alpha^2 > b$, the tracking error is greater than the out-of-sample error. To the right, when $\alpha^2 < b$, the tracking error is smaller than the out-of-sample error. In fact the tracking error tends to zero while the out-of-sample error decreases and then starts to increase again resulting in a well defined minimum.



(a) (b)

FIG. 3. Figure 3(a) shows a plot of the tracking error in blue squares and the out-of-sample error in black diamonds. The errors are plotted against the inverse of α for $\sigma = 0.01$. Figure 3(b) shows a plot of the out-of-sample error in black diamonds for 100 realisations of the noise r_n with $\sigma = 0.01$. It is displayed for the range of α where the minimum occurs. The error bars represent 90% confidence intervals. The state error is show in blue circles also for 100 realisations of the observation noise with 90% confidence intervals. The vertical line draws attention to the minimum of both curves.

The well defined minimum of the out-of-sample error is shown more clearly in Figure 3(b). Figure 3(b) shows the out-of-sample error in black diamonds for the range of α where the minimum occurs. The figure shows the out-of-sample error for 100 realisations of the noise r_n for $\sigma = 0.01$. The error bars represent 90% confidence intervals for each α. Once again we would like to quantify the variation of the parameter α. The mean value of the optimal α plus/minus one standard deviation in this case is

$$\bar{\alpha}^* \pm \sqrt{(\alpha^* - \bar{\alpha}^*)^2} = 0.2238 \pm 0.0079. \tag{44}$$

Figure 3(b) also shows a plot of the state error in blue circles for 100 realisations. The black, vertical line draws attention to the minimum of both curves. We can see that the minimising gain is the same for both errors. When running data assimilation schemes, the state error is the error we are interested in minimising, however we only have access to the error in observation space. Even though this is the case, we have shown numerically that the minimising gain is the same for both errors, even in this non linear situation.

As with the linear numerical experiment presented in Section III A, further experiments using different values of σ where carried out. The results produced were the same as the ones presented here; the only difference was the size of the error bars produced. A smaller value of σ resulted in smaller error bars much like it did for the linear numerical example.

What is particularly of interest here is that even though the dynamical system included a non linear term, the methodology still applies, provided that the matrix $(\mathbf{A} - \mathbf{KHA})$ is stable. As an aside, the experiment suggests that the eigenvalues of the linear part of the error dynamics have to be $< 1 - \epsilon$ with some small but non-zero ϵ in order to stabilise the error dynamics.

C. Numerical Experiment 3: Lorenz '96

For this third numerical experiment, the reality is given by the Lorenz'96 model which is governed by the following equations

$$\dot{x}_i = -x_{i-1}(x_{i-2} - x_{i+1}) - x_i + F \tag{45}$$

and exhibits chaotic behaviour for $F = 8$. By integrating the above differential equation with a time step $\delta = 1.5 \times 10^{-2}$, we obtain a discrete model for our reality which we denote by

$$x_{n+1} = \Phi(x_n). \tag{46}$$

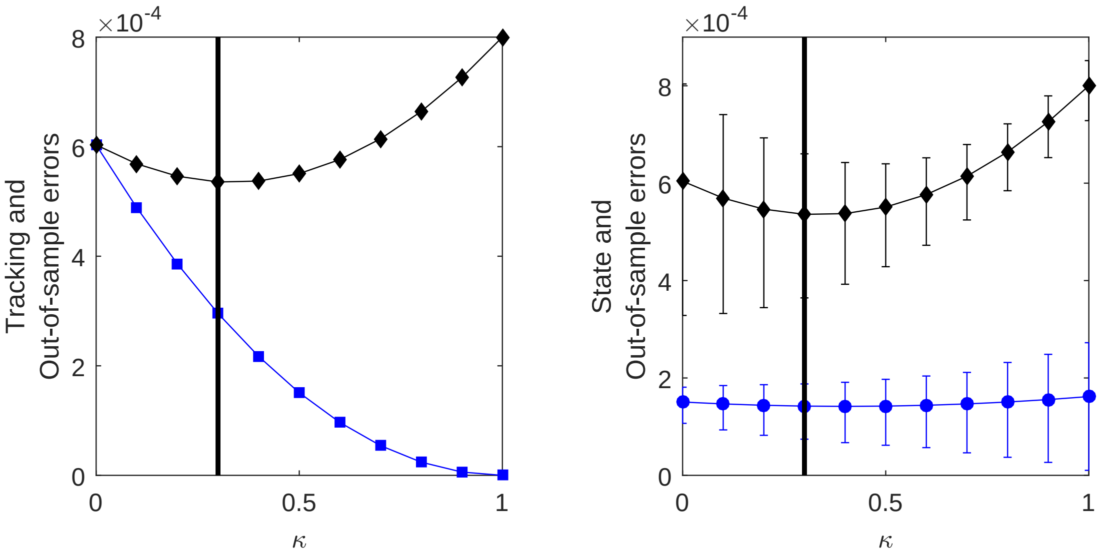
We take corresponding observations of the form

$$\eta_n = \mathbf{H}x_n + \sigma r_n \tag{47}$$

where $\mathbf{H}$ is the observation operator and r_n is iid noise. We shall take the state dimension to be $D = 12$, the observation space to be $d = 4$ and we define the observation operator so that we observe every third element of the state; that is (x_1, x_4, x_7, x_{10}). The system we construct here is fully non-linear with linear observations.

The assimilating model will use the Lorenz'96 model coupled to the observations through a simple linear coupling term, as done in the the previous numerical experiments. We set the coupling matrix $\mathbf{K}$, to be defined by

$$\mathbf{K} = \kappa \mathbf{H}^T \tag{48}$$



(a) (b)

FIG. 4. Figure 4(a) presents the out-of-sample error (black diamonds) and the tracking error (blue squares). Figure 4(b) illustrates the out-of-sample error (black diamonds) and the state error (blue circles) with the error bars representing 90% confidence intervals. The black vertical line draws attention to the minimum of the out-of-sample error.

where κ is a coupling parameter taken to be between 0 and 1. With this information, the assimilating model is defined by the following equations

$$\hat{z}_{n+1} = \Phi(z_n); \quad z_{n+1} = \hat{z}_{n+1} + \kappa \mathbf{H}^T(\eta_{n+1} - \mathbf{H}\hat{z}_{n+1}). \tag{49}$$

Once again we will vary the coupling strength in the observer by adjusting the coupling parameter κ. If the coupling is too strong, the observations will be tracked too rigorously and so the observational noise will not be filtered out. If the coupling is too weak the observations are tracked poorly; so once again we expect the out-of-sample error to take a minimum at some non-trivial value of κ.

As always we are interested in the behaviour of the state error and, ultimately, this is the error we want to be minimal. We saw in Section III B that the minimiser for the out-of-sample error was the same as for the state error. We investigate this here too.

The results obtained are shown in Figure 4. Once again the observational noise is iid with $\mathbb{E}r_n = 0$, $\mathbb{E}r_n r_n^T = 1$ and $\sigma = 0.01$. Since the gain is given by equation (48), the optimism reduces to $8\sigma^2\kappa$. To see this note that the observation operator, $\mathbf{H}$, was defined so that every third element of the state was observed. It follows then that $\mathbf{H}\mathbf{H}^T = \mathbb{1}$, the identity matrix. Since we are observing four states, the trace of $\mathbf{HH}^T$ is equal to four. Thus, since the optimism is defined by $2\sigma^2\text{tr}(\mathbf{HK})$ and $\mathbf{K}$ is given by equation (48), it follows that the optimism reduces to $8\sigma^2\kappa$.

To calculate the the errors, a transient time was ignored to give the system time to synchronise. In Figure 4(a) the out-of-sample error (black diamonds) is presented together with the tracking error (blue squares). The black vertical line draws the eye to the minimum of the out-of-sample error. As in the previous experiments, the tracking error reduces to zero while the out-of-sample error increases eventually with increasing coupling strength.

Figure 4(b) presents the out-of-sample error (black diamonds) and the state error (blue circles). The figure shows the errors for 100 realisations of the observational noise, r_n. The error bars represent 90% confidence intervals for each value of κ with the lower limit of the error bars taken at the fifth percentile and the upper limit taken at the 95th. The mean value of the optimal κ plus/minus one standard deviation in this case is

$$\bar{\kappa}^* \pm \sqrt{(\kappa^* - \bar{\kappa}^*)^2} = 0.3050 \pm 0.1184. \tag{50}$$

The black line draws attention to the minimum of the out-of-sample error and we once again see that the minima of the state and out-of-sample errors coincide. It is evident here that these results support the results determined previously in the numerical experiments. Further experiments using different values of σ where also carried out for this non linear system. The results produced were the same as the ones presented here; the only difference was the size of the error bars produced. Again, as with the results in the previous two experiments, a smaller value of σ resulted in smaller error bars.

The flatness of the curves and the uncertainty shown in the figures are rather deceptive in the plots presented in this paper. By looking at these figures, one might expect that the errors in the estimate of $\boldsymbol{\kappa}^*$ are in fact quite large. However this is not the case as it is the correlation between the errors in the plots that matters.

IV. CONCLUSIONS

A fundamental problem of data assimilation experiments in atmospheric contexts is that there is no possibility of replication, that is, truly "out of sample" observations from the same underlying flow pattern but with independent observational errors are typically not available. A direct evaluation of assimilated trajectories against the available observations is likely to yield optimistic results though, since the observations were already used to find the solution.

A possible remedy was presented which simply consists of estimating that optimism, thereby giving a more realistic picture of the 'out of sample' performance. The optimism represents the correlation between the observations and the output of the data assimilation scheme. This estimate depends on the observational noise, the observation operator and the feedback gain matrix but not on the underlying dynamics or dynamical noise parameters. The model noise is the term that is difficult to determine operationally, so estimating the optimism in an operational situation is possible as all the required terms are readily available. In this paper, this approach was applied to data assimilation algorithms employing linear error feedback. Several numerical experiments concerning both linear and non-linear systems give evidence to the success of this method as it provides more realistic assessment of performance. This was demonstrated by comparing the out-of-sample performance with the true state error of the algorithm which was available in these numerical simulations.

The approach outlined above also provides a simple and efficient means to determine the optimal feedback gain by optimising the out-of-sample error with respect to the gain matrix. Further, theoretical results demonstrate that in linear systems with gaussian perturbations, the feedback thus determined will approach the optimal (Kalman) gain in the limit of large observational windows. The numerical experiments presented in this paper support this result for linear systems.

We cannot deduce the same thing for the non-linear systems since firstly, we do not have a candidate for the asymptotic error or gain since the Kalman Filter equations do not hold in these cases. Secondly, even if the existence of an optimal asymptotic gain could be proved, the sequence of minimisers might not converge to it.

As an outlook for future work, it seems that the presence of dynamical noise in the underlying system is important when considering the convergence of the optimal gain matrix for non-linear systems. (Even in the linear case, the presence of nondegenerate dynamical noise is essential for the proof to work). If there is no model noise present, then we cannot expect the gain matrix to converge in a meaningful way as the optimal asymptotic gain may not be well defined. For example it is possible that the dynamics of both the underlying system and model enter a region of stability, resulting in a reduction of the error. In this case it would make sense to reduce or completely eliminate the feedback gain matrix. This would need the gain matrix to be adaptive in some way; a concept not considered here.

ACKNOWLEDGMENTS

This paper was prepared with the support of the Engineering and Physical Sciences Research Council for Jochen Bröcker under first grant agreement EP/L012669/1. The authors wish to thank Peter Jan van Leeuwen for helpful discussions and constructive suggestions which motivated some of the work in this paper.

Appendix A

In this appendix, we want to clarify the relationship between the output error

$$E_{O,n} = \mathbb{E}[(\mathbf{H}(x_n - z_n))^2] \tag{A1}$$

(which we give an index n here as it depends on n) and the error covariance matrix

$$\Gamma_n = \mathbb{E}[(x_n - z_n)(x_n - z_n)^T] \tag{A2}$$

in the context of linear systems (Section III A). Re-writing the output error we obtain

$$\begin{aligned} E_{O,n} &= \mathbb{E}\{(\mathbf{H}(x_n - z_n))^T(\mathbf{H}(x_n - z_n))\} \\ &= \mathbb{E}\mathrm{tr}\{(\mathbf{H}(x_n - z_n))^T\mathbf{H}(x_n - z_n)\} \\ &= \mathbb{E}\mathrm{tr}\{\mathbf{H}(x_n - z_n)(x_n - z_n)^T\mathbf{H}^T\} \\ &= \mathrm{tr}\{\mathbf{H}\Gamma_n\mathbf{H}^T\} \end{aligned} \tag{A3}$$

and if we assume real values observations (i.e $d = 1$), we get $E_{O,n} = \mathbf{H}\Gamma_n\mathbf{H}^T$. This does not mean that $E_{O,n}$ carries the same information as Γ_n since $\mathbf{H}$ is not invertible.

To investigate this further, introduce the mappings $F : \mathbb{R}^D \times \mathbb{R}^{D\times D} \to \mathbb{R}^{D\times D}$, $(\mathbf{K}, \mathbf{M}) \to (\mathbf{A} - \mathbf{KHA})\mathbf{M}(\mathbf{A} - \mathbf{KHA})^T$ and $G : \mathbb{R}^D \to \mathbb{R}^{D\times D}$; $\mathbf{K} \to \sigma^2\mathbf{KK}^T + \rho^2(\mathbb{1} - \mathbf{KH})(\mathbb{1} - \mathbf{KH})^T$ and $\Phi(\mathbf{K}, \mathbf{M}) = F(\mathbf{K}, \mathbf{M}) + G(\mathbf{K})$. Note that F is linear in $\mathbf{M}$, and we will write $F(\mathbf{K}) \cdot \mathbf{M}$ to emphasize this. It follows from linear filter theory that

$$\begin{aligned} \Gamma_{n+1} &= (\mathbf{A} - \mathbf{KHA})\Gamma_n(\mathbf{A} - \mathbf{KHA})^T + \sigma^2\mathbf{KK}^T + \rho^2(\mathbb{1} - \mathbf{KH})(\mathbb{1} - \mathbf{KH})^T \\ &= F(\mathbf{K}) \cdot \Gamma_n + G(\mathbf{K}) = \Phi(\mathbf{K}, \Gamma_n). \end{aligned} \tag{A4}$$

Suppose that $\mathbf{K}$ is stabilising, then $\Gamma_n \to \Gamma(\mathbf{K})$ which is a fixed point of (A4), i.e $\Gamma(\mathbf{K}) = F(\mathbf{K}) \cdot \Gamma(\mathbf{K}) + G(\mathbf{K})$. Note that $\Gamma(\mathbf{K})$ describes the asymptotic error performance of the feedback $\mathbf{K}$.

We will now show that the output error is able to distinguish (asymptotically) between better and worse feedbacks. For any two symmetric matrices $\mathbf{M}_1, \mathbf{M}_2$, we write $\mathbf{M}_1 \geq \mathbf{M}_2$ if $\mathbf{M}_1 - \mathbf{M}_2$ is positive semi-definite but not zero. Let $\mathbf{K}_1, \mathbf{K}_2$ be two stabilising feedbacks so that $\Gamma(\mathbf{K}_1) \geq \Gamma(\mathbf{K}_2)$; that is $\mathbf{K}_2$ performs better than $\mathbf{K}_1$. Further, assume $(\mathbb{1} - \mathbf{H}\mathbf{K}_1) \neq 0$ which implies that $(\mathbf{A} - \mathbf{K}_1\mathbf{H}\mathbf{A}, \mathbf{H})$ is observable. (This condition might seem artificial but we will see later that it is in fact rather natural). We will now show that $\mathbf{H}\Gamma(\mathbf{K}_1)\mathbf{H}^T > \mathbf{H}\Gamma(\mathbf{K}_2)\mathbf{H}^T$. Note that because $\Gamma(\mathbf{K}_1) \geq \Gamma(\mathbf{K}_2)$ we have

$$\mathbf{M}_n = F^n(\mathbf{K}_1)\{\Gamma(\mathbf{K}_1) - \Gamma(\mathbf{K}_2)\} \geq 0 \tag{A5}$$

for any n since $F(\mathbf{K}_1)$ preserves positive and negative semi-definiteness. Further, the sequence $\mathbf{M}_n$ is decreasing. To see this, note that it must be monotone since

$$\mathbf{M}_{n+1} - \mathbf{M}_n = F(\mathbf{K}_1)\{\mathbf{M}_n - \mathbf{M}_{n-1}\} \tag{A6}$$

and again $F(\mathbf{K}_1)$ preserves definiteness. It cannot be increasing though since $\mathbf{K}_1$ is stabilising and hence $\mathbf{M}_n \to 0$. Therefore $\mathbf{H}\mathbf{M}_n\mathbf{H}^T \geq 0$ and decreasing.

Assuming $\mathbf{H}\Gamma(\mathbf{K}_1)\mathbf{H}^T = \mathbf{H}\Gamma(\mathbf{K}_2)\mathbf{H}^T$ would then imply

$$\begin{aligned} 0 = \mathbf{H}\mathbf{M}_n\mathbf{H}^T &= \mathbf{H}F^n(\mathbf{K}_1)\{\Gamma(\mathbf{K}_1) - \Gamma(\mathbf{K}_2)\}\mathbf{H}^T \\ &= \mathbf{H}(\mathbf{A} - \mathbf{K}_1\mathbf{H}\mathbf{A})^n(\Gamma(\mathbf{K}_1) - \Gamma(\mathbf{K}_2))\left(\mathbf{A} - \mathbf{K}_1\mathbf{H}\mathbf{A}\right)^{n\,T}\mathbf{H}^T \end{aligned} \tag{A7}$$

for all n. Now using the spectral decomposition of $\mathbf{M}_0 = \Gamma(\mathbf{K}_1) - \Gamma(\mathbf{K}_2)$,

$$\mathbf{M}_0 = \sum_{i=1}^{d} \lambda_i v_i v_i^T \tag{A8}$$

where λ_i are the eigenvalues of $\mathbf{M}_0$ and v_i are the corresponding eigenvectors, we see that

$$0 = \mathbf{H}\mathbf{M}\mathbf{H}^T = \sum_{i=1}^{d} \lambda_i \left(\mathbf{H}(\mathbf{A} - \mathbf{K}_1\mathbf{H}\mathbf{A})^n v_i\right)^2 \tag{A9}$$

for all n. Since $\mathbf{M}_0 \neq 0$, there is a $\lambda_j > 0$ and hence

$$\mathbf{H}(\mathbf{A} - \mathbf{K}_1\mathbf{H}\mathbf{A})^n v_j = 0 \quad \forall n \tag{A10}$$

which contradicts the observability of $(\mathbf{H}, \mathbf{A} - \mathbf{K}_1\mathbf{H}\mathbf{A})$. This shows that $\mathbf{M}_0 = 0$ finishing the proof.

From the preceding arguments, it follows that any minimiser of the output error must be the asymptotic Kalman gain. To see this, assume $\mathbf{K}_2$ is the Kalman gain while $\mathbf{K}_1$ optimises the output error $\mathbf{H}\Gamma(\mathbf{K})\mathbf{H}^T$. By definition of the kalman gain, $\Gamma(\mathbf{K}_1) \geq \Gamma(\mathbf{K}_2)$, and the preceding discussion shows that $\Gamma(\mathbf{K}_1) = \Gamma(\mathbf{K}_2)$ if $(\mathbb{1} - \mathbf{H}\mathbf{K}_1) \neq 0$.

To check that this is true, use that the asymptotic output error satisfies

$$\mathbf{H}\Gamma(\mathbf{K})\mathbf{H}^T = (\mathbb{1} - \mathbf{H}\mathbf{K})^2 \left\{ \mathbf{H}\Gamma(\mathbf{K})\mathbf{H}^T + \rho^2 \mathbf{H}\mathbf{H}^T \right\} + \sigma^2 (\mathbf{H}\mathbf{K})^2. \tag{A11}$$

Taking the derivative with respect to $\mathbf{K}$ at $\mathbf{K}_1$ and using the optimality yields the condition

$$\mathbf{H}\mathbf{K}_1 = \frac{\mathbf{H}\Gamma(\mathbf{K}_1)\mathbf{H}^T + \mathbf{H}\mathbf{H}^T\rho^2}{\mathbf{H}\Gamma(\mathbf{K}_1)\mathbf{H}^T + \mathbf{H}\mathbf{H}^T\rho^2 + \sigma^2} \tag{A12}$$

so $\mathbb{1} = \mathbf{H}\mathbf{K}_1 > 0$. As a final remark, $\mathbb{1} - \mathbf{H}\mathbf{K} = 0$ implies that $y_n = \eta_n$ (check example (22) for constant $\mathbf{K}$), that is the data assimilation simply reports back the observations.

REFERENCES

[1] C. M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press Inc., 1995).

[2] B. Efron, "How biased is the apparent error rate of a prediction rule?" Journal of the American Statistical Association **81**, 461–470 (1986).

[3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction (Second Edition)* (Springer-Verlag, 2009).

[4] B. Efron, "The estimation of prediction error: Covariance penalties and cross-validation," Journal of the American Statistical Association **99** (2004).

[5] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*, 1st ed. (Cambridge University Press, 2001).

[6] G. Wahba, D. R. Johnson, F. Gao, and J. Gong, "Adaptive tuning of numerical weather prediction models: Randomized GCV in three- and Four-Dimensional data assimilation," Monthly Weather Review (1995).

[7] G. P. Cressman, "An operational objective analysis system," Monthly Weather Review **87**, 367–374 (1959).

[8]S. L. Barnes, "A technique for maximizing details in numerical weather map analysis," Journal of Applied Meteorology **3**, 396–409 (1964).

[9]W. Lahoz, B. Khattatov, and R. Menard, *Data Assimilation: Making Sense of Observations* (Springer-Verlag, 2010).

[10]Y. Sasaki, "Some basic formalisms in numerical variational analysis," Monthly Weather Review **98**, 875–883 (1970).

[11]A. Lorenc, "A global three-dimensional multivariate statistical interpolation scheme," Monthly Weather Review **109**, 701–721 (1981).

[12]A. H. Jazwinski, *Stochastic Processes and Filtering Theory Volume 64* (Academic Press Inc., 1970).

[13]A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Univeral Concept in Nonlinear Sciences* (Cambridge University Press, 2001).

[14]H. J. C. Huijberts, H. Nijmeijer, and A. Y. Pogromsky, "Discrete-time observers and synchronization," Controlling chaos and bifurcations in engineering systems , 439–455 (1999).

[15]S. Boccaletti, J. Kurths, G. Osipov, D. Valladares, and C. Zhou, "The synchronization of chaotic systems," Physics Reports **366**, 1–101 (2002).

[16]J. Bröcker and I. G. Szendro, "Sensitivity and Out-Of-Sample Error in Continuous Time Data Assimilation," Quarterly Journal of the Royal Meteorological Society **138**, 1785–801 (2012).

[17]I. G. Szendro, M. A. Rodrìguez, and J. M. Lopez, "On the problem of data assimilation by means of synchronization," Journal Of Geophysical Research **114**, D20109 (2009).

[18]S.-C. Yang, D. Baker, and H. Li, "Data Assimilation as Synchronization of Truth and Model: Experiments with the Three-Variable Lorenz System," Journal of the Atmospheric Sciences **63**, 2340–2354 (2006).

[19]R. Dorf and R. Bishop, *Modern Control Systems Tenth Edition* (Pearson Education Inc., 2005).

[20]B. Anderson and J. Moore, *Optimal Filtering* (Dover Publications Inc, 1979).

[21]W. F. Arnold III and A. J. Laub, "Generalized eigenproblem algorithms and software for algebraic riccati equations," Proceedings of the IEEE **72**, 1746–1754 (1984).