Data Assimilation using Time-Delay Nudging in the Presence of Gaussian Noise

Emine Celik[1] and Eric Olson[2]

May 18, 2023

[1]*Department of Mathematics, Sakarya University*
54050 Sakarya, Turkey
[2]*Department of Mathematics and Statistics, University of Nevada, Reno*
Reno, NV 89557, USA

Email addresses: `eminecelik@sakarya.edu.tr`, `ejolson@unr.edu`

Abstract

We study a discrete-in-time data-assimilation algorithm based on nudging through a time-delayed feedback control in which the observational measurements have been contaminated by a Gaussian noise process. In the context of the two-dimensional incompressible Navier–Stokes equations we prove the expected value of the square-error between the approximating solution and the reference solution over time is proportional to the variance of the noise up to a logarithmic correction. The qualitative behavior and physical relevance of our analysis is further illustrated by numerical simulation.

Keywords: Discrete data assimilation, two dimensional Navier-Stokes equations.

AMS subject classifications: 35Q30, 37C50, 76B75, 93C20.

1 Introduction

The goal of data assimilation is to combine incomplete and possibly noisy observational measurements taken over an interval of time with dynamical knowledge about the system being observed to obtain a more accurate estimate of the current state. The time-delay nudging algorithm was introduced in the context of the two-dimensional incompressible Navier–Stokes equations by Foias, Mondaini and Titi [16] as a sequential data-assimilation technique to processes a discrete time series of incomplete noisy observations. In this paper we extend the results of that work by adding a mechanism that removes outliers from the data. This eliminates the requirement that the noise be bounded and allows us to treat observations contaminated by a Gaussian process. We further remove some additional L^2-boundedness assumptions on the interpolant observables and follow [9] by applying a spectral filter to the feedback term.

The resulting analysis then provides bounds on $\mathbf{E}\big[\|U - u\|_{H^1}^2\big]$ where U is the reference solution and u is the approximating solution obtained through the modified delay-nudging method just mentioned. Our main result consists of mathematically rigorous conditions under which bounds on the expectation naturally depend on the variance of the Gaussian noise process up to a logarithmic correction.

Before proceeding we remark on the importance of removing outliers from the observations. Foremost, this allows our analysis to treat Gaussian noise processes. Intuitively with such a noise process there is a small chance that any observation is contaminated by an error that dwarfs the size of all possible trajectories lying on the global attractor. Therefore, upon assuming the unknown reference solution comes from a long term evolution prior to the observations, we can identify conditions for when an observation contains an unreasonably large amount of noise. Removing arbitrarily large outliers then prevents them from damaging the approximation obtained by data assimilation and provides rigorous bounds on the expected quality of that approximation.

From a practical point of view, any instrumentation which provides measurements of a physical process has a limited numerical range of possible observations. For example, the output of an anemometer simply can not display arbitrarily large velocities nor can it spin at arbitrarily large speeds. An individual device may include features to eliminate obviously wrong measurements, while signal processing techniques coupled with common sense are typically used to aggregate and reconcile data from many sources. Thus, even if the underlying noise in a physical observation is Gaussian, the resulting datasets do not include arbitrarily large errors. In this way the mathematical techniques described in this paper for removing outliers can also be seen to mimic realistic data collection.

As in [16], see also [1], [2], [5], [18] and [20] the model problem used for our study is the two-dimensional incompressible Navier–Stokes equations. In particular, the reference solution U used for our data-assimilation problem satisfies

$$\frac{\partial U}{\partial t} + (U \cdot \nabla)U - \nu \Delta U + \nabla p = f, \qquad \nabla \cdot U = 0 \tag{1.1}$$

on the domain $\mathbf{T} = [0, 2\pi]^2$ equipped with 2π-periodic boundary conditions. Here ν is the kinematic viscosity, f a time-independent body force, p the pressure and U the Eulerian velocity field of the fluid.

The use of the two-dimensional Navier–Stokes as a computational tool to study data assimilation can be traced to a 1998 report of Browning, Henshaw and Kreiss [6] while the main ideas behind the rigorous mathematical analysis trace their origin to a 1967 paper on determining modes by Foias and Prodi [14]. We note the rich analytic theory behind these equations, their computational tractability and the dynamical properties similar to more complicated physical models such as the primitive equations which govern the atmosphere. For these reasons we hope that our current research will be seen as both mathematically interesting and physically relevant.

For our computations we introduce a physically-motivated interpolant based on observations of the velocity field that consist of local averages taken at a set of points in space. From a theoretical point of view these local averages provide the regularity needed to obtain a type-I interpolant to which our theory applies. Aside from these theoretical advantages, we emphasize from a physical point of view that such local averages approximate nodal measurements

of the velocity field while at the same time more realistically represent the characteristic averaging properties of physical observational devices. For completeness, a proof that the interpolant described above is type I appears in Appendix A. Nudging with noise was investigated computationally with a focus on parameter recovery for the Lorenz equations by Carlson, Hudson, Larios, Martinez and Ng [7] and in the context of the Rayleigh-Bénard equations by El Rahman Hammoud, Le Maître, Titi, Hoteit and Knio [19].

The organization of this paper is as follows. Section 2 recalls some properties of the two-dimensional incompressible Navier–Stokes equations, some statistical results and then fixes our notation. We describe the process of removing outliers from the observational data and further state our main theoretical result as Theorem 2.6 which bounds the expectation in the presence of Gaussian noise. Section 3 obtains pathwise estimates similar to those appearing in [16] in the form needed for our main result which is then proved in Section 4. The paper finishes in Section 5 with a set of computations to test the physical relevance of the theory along with conclusions and directions for future work.

2   Preliminaries

We begin by introducing the functional notation used by the theoretical study of the Navier–Stokes equations and then stating some *a priori* bounds on long-time solutions.

Let $\mathcal{V}$ be the set of all $\mathbf{R}^2$-valued divergence-free 2π-periodic trigonometric polynomials with zero spatial averages, V the closure of $\mathcal{V}$ in $H^1(\mathbf{T})$ where $\mathbf{T} = [0, 2\pi]^2$ is the fundamental domain of the 2π-periodic torus, V^* be the dual of V and let P_H be the orthogonal projection of $L^2(\mathbf{T})$ onto H where H is the closure of $\mathcal{V}$ in $L^2(\mathbf{T})$. For simplicity we will write L^2 and H^1 without specifying the domain $\mathbf{T}$ when there is no chance of confusion.

Due to the periodic boundary conditions, the spaces V and H can also be characterized in terms of Fourier series. In particular,

$$H = \left\{ \sum_{k \in \mathbf{Z}^2 \setminus \{0\}} u_k e^{ik \cdot x} : \sum_{k \in \mathbf{Z}^2 \setminus \{0\}} |u_k|^2 < \infty, \quad k \cdot u_k = 0 \quad \text{and} \quad u_{-k} = \overline{u_k} \right\}$$

while

$$V = \left\{ \sum_{k \in \mathbf{Z}^2 \setminus \{0\}} u_k e^{ik \cdot x} : \sum_{k \in \mathbf{Z}^2 \setminus \{0\}} |k|^2 |u_k|^2 < \infty, \quad k \cdot u_k = 0 \quad \text{and} \quad u_{-k} = \overline{u_k} \right\}.$$

Here $u_k \in \mathbf{C}^2$ are the Fourier coefficients for the velocity field u.

Let $A \colon V \to V^*$ and $B \colon V \times V \to V^*$ be the continuous extensions for $u, v \in \mathcal{V}$ of the operators given by

$$Au = -P_H \Delta u \qquad \text{and} \qquad B(u, v) = P_H(u \cdot \nabla v).$$

Note A is a positive operator with smallest eigenvalue $\lambda_1 = 1$. This dimensional constant is carried throughout our analysis for consistency. We further write the L^2 norm of $u \in H$ as $|u|$, the H^1 norm of $u \in V$ as $\|u\|$ and note that $|Au|$ is equivalent to the H^2 norm on the domain $\mathcal{D}(A)$ of A into H.

Recall also the orthogonality property

$$\big(B(v,v), Av\big) = 0 \tag{2.1}$$

which holds for periodic two-dimensional divergence-free vector fields.

As shown in Constantin and Foias [11], Foias, Manley, Rosa and Temam [15], Robinson [28] or Temam [29], given $f \in H$ and $U_0 \in V$, the two-dimensional incompressible Navier–Stokes equations have a unique strong solution $U(t) \in V$ for $t \geq 0$ which depends continuously on the initial condition U_0 with respect to the V norm. In particular, given any $T > 0$ we have that

$$U \in L^\infty\big([0,T); V\big) \cap L^2\big([0,T); \mathcal{D}(A)\big) \qquad \text{and} \qquad \frac{dU}{dt} \in L^2\big([0,T); H\big). \tag{2.2}$$

We may then express (1.1) in functional form as

$$\frac{dU}{dt} + \nu AU + B(U,U) = f \tag{2.3}$$

with initial condition $U_0 \in V$.

Under the conditions mentioned above, it is well known that (2.3) possess a unique global attractor $\mathcal{A}$. To set our notation we follow [28] and denote the bounds on $\mathcal{A}$ by

Theorem 2.1. *Let $\mathcal{A}$ be the global attractor of (1.1) the two-dimensional incompressible Navier–Stokes equations. There exists a priori bounds ρ_H, ρ_V and ρ_A depending only on ν and f such that*

$$|U| \leq \rho_H, \qquad \|U\| \leq \rho_V \qquad and \qquad |AU| \leq \rho_A \tag{2.4}$$

for every $U \in \mathcal{A}$.

We now characterize the linear operation J_h that will be used to interpolate the observational measurements of the solution U.

Definition 2.2. A linear operator $J_h \colon V \to H$ is said to be a *type-I interpolant observable* if there exists $c_1 > 0$ such that

$$|\Phi - J_h(\Phi)|^2 \leq c_1 h^2 \|\Phi\|^2 \qquad \text{for all} \qquad \Phi \in V. \tag{2.5}$$

Sometimes interpolants $I_h \colon V \to L^2(\mathbf{T})$ which satisfy

$$|\Phi - I_h(\Phi)|^2_{L^2(\mathbf{T})} \leq c_1 h^2 \|\Phi\|^2 \tag{2.6}$$

are considered. In such cases taking $J_h = P_H I_h$ implies

$$|\Phi - J_h(\Phi)|^2 = |P_H(\Phi - I_h(\Phi))|^2_{L^2(\mathbf{T})} \leq |\Phi - I_h(\Phi)|^2_{L^2(\mathbf{T})}$$

and results in an interpolant which satisfies (2.5). Thus, no generality is lost by assuming the range of J_h is H in the first place.

Appendix A describes the exact type-I interpolant used for our numerics. In particular, Theorem A.1 shows the interpolant I_h given by (5.1) satisfies (2.6) for every $\Phi \in V$. More

information, other examples of type-I interpolants as well as the definition of a type-II interpolant may be found in [1], [2] and references therein. We remark that type-II interpolants would involve the use of stronger Sobolev norms in the analysis and are outside the scope of the present research. It is an interesting question, however, whether similar results as presented here also hold for type-II interpolants.

To model the effects of measurement errors, we set

$$\widetilde{J_h}U(t_n) = J_h U(t_n) + \eta_n$$

where η_n is sequence of independent H-valued Gaussian random variables such that

$$\mathbf{E}[\eta_n] = 0 \qquad \text{and} \qquad \mathbf{V}[\eta_n] = \mathbf{E}\big[|\eta_n|^2\big] = \sigma^2.$$

Specifically, we suppose each η_n is distributed as η where

$$\eta = \sum_{i=1}^{2N} \sigma_i Y_i \psi_i, \qquad \sum_{i=1}^{2N} \sigma_i^2 = \sigma^2, \qquad \psi_i \in H \quad \text{with} \quad |\psi_i| = 1 \tag{2.7}$$

and Y_i are independent standard normal random variables. Note the finite degrees of freedom reflected by $2N$ is based on the physical notion that noise in observations of the state of U arise from a finite number of independent noisy $\mathbf{R}^2$-valued measurements of the velocity field. Thus, N represents the number of measurements taken at each instance in time and in the two-dimensional setting considered here is inversely proportional to h^2.

Note that $|\eta|^2$ is the generalized χ^2 distribution given by

$$|\eta|^2 = \sum_{i,j} \sigma_i Y_i (\psi_i, \psi_j) \sigma_j Y_j = X^T \Psi X \quad \text{with} \quad \Psi_{ij} = (\psi_i, \psi_j) \quad \text{and} \quad X_i = \sigma_i Y_i.$$

In our analysis we shall make use of the exponential bound proved as Lemma 1 by Laurent and Massert in [24] and stated here for reference as

Lemma 2.3 (Laurent and Massert). *Let Y_i for $i = 1, \ldots, d$ be independent identically-distributed standard normal random variables and $a_i \geq 0$. Set*

$$|a|_\infty = \sup \big\{ |a_i| : i = 1, \ldots, d \big\}, \quad |a|_2 = \Big(\sum_{i=1}^d a_i^2 \Big)^{1/2} \quad \text{and} \quad Z = \sum_{i=1}^d a_i(Y_i^2 - 1).$$

Then for $x > 0$ holds
$$\mathbf{P}\big\{ Z \geq 2|a|_2 \sqrt{x} + 2|a|_\infty x \big\} \leq e^{-x}.$$

Note by setting $a_i = \sigma_i^2$ and $d = 2N$ we obtain

$$|a|_\infty \leq \sum_{i=1}^{2N} |a_i| \leq \sigma^2 \qquad \text{and} \qquad |a|_2 \leq \Big(|a|_\infty \sum_{i=1}^{2N} |a_i| \Big)^{1/2} \leq \sigma^2$$

in which case Lemma 2.3 implies

$$\mathbf{P}\big\{ \|X\|^2 - \sigma^2 \geq 2\sigma^2 \sqrt{x} + 2\sigma^2 x \big\} \leq e^{-x}. \tag{2.8}$$

With this framework in place, we describe in details the data assimilation method which appears in [16] that constitutes the beginning of our analytical and numerical study.

Definition 2.4. The *delay-nudging method* constructs an approximation u of the reference solution U by setting $u(t_0) = 0$ and then evolving u continuously as

$$\frac{du}{dt} + \nu A u + B(u, u) = f + \mu\big(\widetilde{J}_h U(t_n) - J_h u(t_n)\big) \quad \text{for} \quad t \in [t_n, t_{n+1}). \tag{2.9}$$

Here μ is a relaxation parameter which affects the strength of the feedback term and may be tuned based on the resolution and the noise present in the measurements.

Before proceeding it is critical to check that equations (2.9) are well posed and for any fixed realization of the noise process η_n uniquely determine an approximating solution $u(t)$. Since the reference solution $U(t) \in V$ for all $t \geq 0$, then $J_h U(t_n) \in H$ for $n = 0, 1, \dots$ by the definition of a type-I interpolant. Note first that the dynamics governing $u(t)$ on the interval $[t_0, t_1)$ are identical to the two-dimensional incompressible Navier–Stokes equations (2.3) with initial condition $0 \in V$ at $t = t_0$ and time-independent body forcing

$$f + \mu \widetilde{J}_h U(t_0) = f + \mu J_h U(t_0) + \eta_0 \in H.$$

Here we have used that $\eta_n \in H$. The theory of the two-dimensional incompressible Navier–Stokes equations now implies there exists a unique strong solution $u \in C([t_0, t_1]; V)$. To evolve u further in time consider (2.9) with $n = 1$ and initial condition $u(t_1) \in V$ at $t = t_1$. The argument now follows by induction.

Suppose $u \in C([t_0, t_n]; V)$. Then $u(t_n) \in V$ implies $J_h u(t_n) \in H$. It follows that

$$f + \mu\big(\widetilde{J}_h U(t_n) - J_h u(t_n)\big) = f + \mu J_h U(t_n) - \mu J_h u(t_n) + \eta_n \in H.$$

Consequently there exists a unique strong solution $u \in C([t_n, t_{n+1}]; V)$. This combined with the induction hypothesis yields that $u \in C([t_0, t_{n+1}]; V)$.

The delay-nudging method given by Definition 2.4 was originally described and analyzed by Foias, Mondaini and Titi [16]. That work includes a pathwise treatment of noisy observations for the case where the noise process is bounded. Namely, one has

Theorem 2.5 (Foias, Mondaini and Titi). *Let u be the approximating solution obtained by delay-nudging on $[t_0, \infty)$ satisfying $u(t_0) = 0$. Assume η_n is a noise process such that*

$$\|\eta_n\|_{H^1} \leq \mathcal{E}_1 \qquad \text{for all} \qquad n \in \mathbf{N} \tag{2.10}$$

and that the type-I interpolant observable $I_h \colon L^2(\mathbf{T}) \to L^2(\mathbf{T})$ further satisfies

$$\|I_h \Phi\|_{L^2} \leq c_2 \|\Phi\|_{L^2} \qquad \text{for all} \qquad \Phi \in L^2(\mathbf{T}). \tag{2.11}$$

Here c_2 is a positive constant. Suppose μ, h and δ satisfy

$$\mu \geq c \frac{(\rho_V + \mathcal{E}_1)^2}{\nu}\left(1 + \log\left(\frac{\rho_V + \mathcal{E}_1}{\nu \lambda_1^{1/2}}\right)\right), \qquad h \leq \frac{1}{2c_0}\left(\frac{\nu}{\mu}\right)^{1/2}$$

and

$$\delta \leq \frac{c}{\mu} \min\left\{1, \frac{\nu^{3/2}\mu^{1/2}}{\rho_H \rho_V}, \frac{\nu^2 \lambda_1^{1/2}}{\rho_H \rho_V}, \frac{\nu^2 \lambda_1}{(\rho_V + \mathcal{E}_1)^2}, \frac{(\nu \lambda_1)^{1/2}}{\mu^{1/2}}, \frac{(\nu \lambda_1)^2}{\mu^2}\right\}.$$

Then

$$\limsup_{t \to \infty} \|U - u\|_{H^1} \leq c\mathcal{E}_1.$$

This paper extends the above result to the case when η_n is an H-valued Gaussian noise processes that is not bounded by $\mathcal{E}_1$. We further remove the continuity condition (2.11) and the requirement that $\eta_n \in V$.

To these ends we consider a modified nudging method that removes outliers along with a bootstrapping argument to obtain a theorem that applies when the noise process is Gaussian. In particular, by defining the modified interpolant observable

$$\widetilde{J}_h^o U(t_n) = \begin{cases} \widetilde{J}_h U(t_n) & \text{for } |\widetilde{J}_h U(t_n)| \leq 2M \\ 0 & \text{otherwise} \end{cases} \tag{2.12}$$

we obtain a new noise process $\eta_n^o = \widetilde{J}_h^o U(t_n) - J_h U(t_n) \in H$ which is bounded and an interpolant that filters outliers that correspond to points outside the known absorbing ball of the global attractor when M is large enough. Replacing $\widetilde{J}_h$ by $\widetilde{J}_h^o$ in (2.9) subsequently leads to the *modified nudging method* and our main theoretical result.

Theorem 2.6. *Let u be an approximating solution obtained by the modified nudging method with suitable relaxation parameter μ. If $\delta > 0$ and $h > 0$ are small enough, then there exists a constant C_0 independent of σ and a logarithmic correction $f(\sigma)$ such that*

$$\limsup_{t \to \infty} \mathbf{E}\big[\|U - u\|^2\big] \leq C_0 \sigma^2 f(\sigma)$$

holds for all σ sufficiently small.

We remark that Theorem 2.6 forgoes pathwise bounds in order to handle the case where the measurement errors are distributed according to a Gaussian distribution. Since in the Gaussian case there is no finite $\mathcal{E}_0$ such that $|\eta_n| \leq \mathcal{E}_0$ holds for all n no matter how small the variance σ^2, with non-zero probability there will be arbitrarily long sequences of consecutive observations such that $|\eta_n|$ is large. It follows there exists $C > 0$ such that the approximating solutions obtained by the original nudging method satisfy

$$\limsup_{t \to \infty} |U - u| \geq C$$

almost surely no matter how small σ. For this reason Theorem 2.6 does not provide pathwise bounds, but instead bounds the expectation in a way that shows the corresponding approximating solutions asymptotically recover the exact solution when σ goes to zero.

We now recall some inequalities that will be used in the subsequent analysis. Writing the smallest eigenvalue of the Stokes operator A as $\lambda_1 = (2\pi/L)^2$ we have the Poincaré inequalities

$$\lambda_1 |U|^2 \leq \|U\|^2 \qquad \text{for} \qquad U \in V \tag{2.13}$$

and

$$\lambda_1^2 |U|^2 \leq \lambda_1 \|U\|^2 \leq |AU|^2 \qquad \text{for} \qquad U \in \mathcal{D}(A). \tag{2.14}$$

Recall also the combination of Agmon's inequality [31] with (2.14) given by

$$\|U\|_{L^\infty} \leq C|U|^{1/2}|AU|^{1/2} \leq C\lambda_1^{-1/2}|AU| \tag{2.15}$$

7

and Ladyzhenskaya's inequality

$$\|U\|_{L^4} \le C|U|^{1/2}\|U\|^{1/2}. \tag{2.16}$$

In both (2.15) and (2.16) the constant C is dimensionless and does not depend on U.

Applying Hölder followed by the Ladyzhenskaya and Agmon inequalities to the nonlinear terms in the two-dimensional Navier–Stokes equations yields

$$|B(u,v)| \le \|u\|_{L^4}\|A^{1/2}v\|_{L^4} \le C|u|^{1/2}\|u\|^{1/2}\|v\|^{1/2}|Av|^{1/2} \tag{2.17}$$

and

$$|B(u,v)| \le \|u\|_{L^\infty}\|A^{1/2}v\|_{L^2} \le C|u|^{1/2}|Au|^{1/2}\|v\|. \tag{2.18}$$

Finally, we shall make use of the logarithmic bound on the non-linear term proved in Titi [30], further employed in [16] and stated here as

Proposition 2.7. *If U and w are in $\mathcal{D}(A)$ then*

$$\left|\big(B(w,U) + B(U,w), Aw\big)\right| \le C\|w\|\|U\|\Big(1 + \log\frac{|AU|}{\lambda_1^{1/2}\|U\|}\Big)^{1/2}|Aw|,$$

where C is a non-dimensional constant depending only on the domain.

3 Bounds on the Paths

In this section we adapt the proof of Theorem 2.5 that appears in [16] to obtain a similar pathwise bound under the weaker hypothesis that the L^2 norm of η_n is bounded and without the requirement that J_h also satisfy (2.11). Unlike the original proof we limit our attention to periodic domains **T** which later form the context of our computational setting. Since the plan is to use this result to obtain bounds on the expectation in the next section, we explicitly track the rate of approximate synchronization over time.

Write $w = U - u$ where U is a free running solution to the two-dimensional Navier–Stokes equations given by (2.3) and u is an approximating solution obtained from the delay nudging method (2.9). It has already been shown that U and u are strong solutions of the two-dimensional incompressible Navier–Stokes equations on each interval $[t_n, t_{n+1})$, though each with a different initial condition and body force. Consequently w enjoys the same regularity properties. It follows that

$$\frac{dw}{dt} + \nu Aw + B(w,U) + B(U,w) + B(w,w) = -\mu J_h w(t_n) - \mu\eta_n \tag{3.1}$$

for $t \in [t_n, t_{n+1})$. To treat the time delay resulting from the first term on the right in the above equation we first prove

Lemma 3.1. *Assuming $|\eta_n| \le \alpha$ then for $t \in [t_n, t_{n+1})$ holds*

$$|w(t) - w(t_n)|^2 \le 4(t-t_n)^2\mu^2\big\{(c_1h^2 + \lambda_1^{-1})\|w(t_n)\|^2 + \alpha^2\big\}$$
$$+ 4(t-t_n)\int_{t_n}^t \big\{\nu + C\lambda_1^{-1/2}\big(2\rho_V + \|w(s)\|\big)\big\}^2|Aw(s)|^2ds.$$

8

Proof. Since the Cauchy–Schwarz inequality implies

$$|w(t) - w(t_n)|^2 \le \left(\int_{t_n}^t \left| \frac{dw(s)}{ds} \right| ds \right)^2 \le (t - t_n) \int_{t_n}^t \left| \frac{dw(s)}{ds} \right|^2 ds,$$

it is enough to estimate

$$\left| \frac{dw}{dt} \right| \le \nu|Aw| + |B(w, U)| + |B(U, w)| + |B(w, w)| + \mu|J_h w(t_n)| + \mu|\eta_n|.$$

Applying (2.17) and (2.18) to the nonlinear terms followed by the Poincaré inequality yields

$$|B(w, U)| \le C|w|^{1/2}|Aw|^{1/2}\|U\| \le C\lambda_1^{-1/2} \rho_V |Aw|$$

$$|B(U, w)| \le C|U|^{1/2}\|U\|^{1/2}\|w\|^{1/2}|Aw|^{1/2} \le C\lambda_1^{-1/2} \rho_V |Aw|$$

$$|B(w, w)| \le C|w|^{1/2}\|w\||Aw|^{1/2} \le C\lambda_1^{-1/2}\|w\||Aw|.$$

Note also

$$|J_h w(t_n)| \le |w(t_n) - J_h w(t_n)| + |w(t_n)| \le (c_1^{1/2}h + \lambda_1^{-1/2})\|w(t_n)\|.$$

Therefore

$$\left| \frac{dw}{dt} \right| \le \left\{ \nu + C\lambda_1^{-1/2}(2\rho_V + \|w\|) \right\}|Aw| + \mu(c_1^{1/2}h + \lambda_1^{-1/2})\|w(t_n)\| + \mu\alpha$$

and so

$$\left| \frac{dw}{dt} \right|^2 \le 4\left\{ \nu + C\lambda_1^{-1/2}(2\rho_V + \|w\|) \right\}^2|Aw|^2 + 4\mu^2(c_1 h^2 + \lambda_1^{-1})\|w(t_n)\|^2 + 4\mu^2\alpha^2.$$

Consequently,

$$|w(t) - w(t_n)|^2 \le 4(t - t_n)^2 \mu^2 \left\{ (c_1 h^2 + \lambda_1^{-1})\|w(t_n)\|^2 + \alpha^2 \right\}$$
$$+ 4(t - t_n) \int_{t_n}^t \left\{ \nu + C\lambda_1^{-1/2}(2\rho_V + \|w(s)\|) \right\}^2 |Aw(s)|^2 ds,$$

which was to be shown. $\qquad \square$

With the above lemma in hand we now turn our attention to a version of Theorem 2.5 that will be used to prove our main theoretical result. In addition to the already mentioned differences, Proposition 3.2 below provides an explicit separation of the dependency of the exponential rate θ on the maximum bound $\mathcal{E}_0$ from the estimate on $\|w(t)\|$ which results when $|\eta_n|$ is further bounded for a finite time by α. Note this separation of dependencies is required for the probabilistic estimates which appear in Section 4.

Proposition 3.2. *Let $\mathcal{M}_0$ and $\mathcal{E}_0$ be fixed. There are positive constants c, δ, θ, h and μ with $\theta < 1$ such that upon taking $t_n = t_0 + n\delta$ the conditions $\|w(t_k)\| \le \mathcal{M}_0$ and $|\eta_n| \le \alpha$ where $\alpha \le \mathcal{E}_0$ for $n = k, \ldots, k + p$ imply*

$$\|w(t)\|^2 \le \theta^{n-k}\|w(t_k)\|^2 + c\alpha^2 \quad \text{for all} \quad t \in [t_n, t_{n+1}] \tag{3.2}$$

and $n = k, \ldots, k + p$.

Proof. Taking the inner product of (3.1) with Aw and using the orthogonality (2.1) we have

$$\frac{1}{2}\frac{d\|w\|^2}{dt} + \nu|Aw|^2 + \big(B(w,U) + B(U,w), Aw\big) = -\mu\big(J_h w(t_n) + \eta_n, Aw\big). \qquad (3.3)$$

First, use Proposition 2.7 to estimate the non-linear terms as

$$\big|\big(B(w,U) + B(U,w), Aw\big)\big| \le L\|w\||Aw| \le \frac{2L^2}{\nu}\|w\|^2 + \frac{\nu}{8}|Aw|^2,$$

where

$$L = C\rho_V\Big(1 + \log\frac{\rho_A}{\lambda_1^{1/2}\rho_V}\Big)^{1/2}.$$

The last term in the inner product on the right side of (3.3) may be estimated as

$$\mu\big|(\eta_n, Aw)\big| \le \mu|\eta_n||Aw| \le \mu\alpha|Aw| \le \frac{2\mu^2}{\nu}\alpha^2 + \frac{\nu}{8}|Aw|^2.$$

Now write

$$-\mu\big(J_h w(t_n), Aw\big) = \mu\big(w(t_n) - J_h w(t_n), Aw\big) + \mu\big(w - w(t_n), Aw\big) - \mu\|w\|^2.$$

Since the Cauchy-Schwarz inequality and (2.5) implies

$$\mu\big|\big(w(t_n) - J_h w(t_n), Aw\big)\big| \le \frac{2\mu^2}{\nu}c_1 h^2\|w(t_n)\|^2 + \frac{\nu}{8}|Aw|^2$$

and

$$\mu\big|\big(w - w(t_n), Aw\big)\big| \le \frac{2\mu^2}{\nu}|w - w(t_n)|^2 + \frac{\nu}{8}|Aw|^2,$$

it follows that

$$\frac{d\|w\|^2}{dt} + 2\Big(\mu - \frac{2L^2}{\nu}\Big)\|w\|^2 + \nu|Aw|^2 \le \frac{4\mu^2}{\nu}\big(|w - w(t_n)|^2 + c_1 h^2\|w(t_n)\|^2 + \alpha^2\big).$$

Choosing μ large enough that $\mu\nu \ge 4L^2$ and applying Lemma 3.1 we have

$$\frac{d\|w\|^2}{dt} + \mu\|w\|^2 + \nu|Aw|^2 - \frac{16\mu^2(t - t_n)}{\nu}\int_{t_n}^t \big\{\nu + C\lambda_1^{-1/2}\big(2\rho_V + \|w(s)\|\big)\big\}^2|Aw(s)|^2 ds$$
$$\le R_1(t - t_n)\|w(t_n)\|^2 + R_2(t - t_n)\alpha^2$$

where

$$R_1(\tau) = \frac{4\mu^2}{\nu}\big\{4\tau^2\mu^2(c_1 h^2 + \lambda_1^{-1}) + c_1 h^2\big\} \qquad \text{and} \qquad R_2(\tau) = \frac{4\mu^2}{\nu}(4\tau^2\mu^2 + 1).$$

Consequently

$$\frac{d}{dt}\big(\|w\|^2 e^{\mu t}\big) + e^{\mu t_n}\nu|Aw|^2$$
$$- e^{\mu t_{n+1}}\frac{16\mu^2(t - t_n)}{\nu}\int_{t_n}^t \big\{\nu + C\lambda_1^{-1/2}\big(2\rho_V + \|w(s)\|\big)\big\}^2|Aw(s)|^2 ds$$
$$\le e^{\mu t}R_1(t - t_n)\|w(t_n)\|^2 + e^{\mu t}R_2(t - t_n)\alpha^2.$$

Integrate over the interval $[t_n, t]$. Estimate the resulting double integral as

$$\int_{t_n}^{t} \frac{16\mu^2(\tau - t_n)}{\nu} \int_{t_n}^{\tau} \left\{ \nu + C\lambda_1^{-1/2}(2\rho_V + \|w(s)\|) \right\}^2 |Aw(s)|^2 ds \, d\tau$$

$$\leq \frac{16\mu^2(t - t_n)^2}{\nu} \int_{t_n}^{t} \left\{ \nu + C\lambda_1^{-1/2}(2\rho_V + \|w(s)\|) \right\}^2 |Aw(s)|^2 ds.$$

Then use the fact that R_1 and R_2 are increasing functions to obtain

$$\|w\|^2 e^{\mu t} + e^{\mu t_n} \int_{t_n}^{t} \left(\nu - e^{\mu\delta} \frac{16\mu^2(t - t_n)^2}{\nu} \left\{ \nu + C\lambda_1^{-1/2}(2\rho_V + \|w(s)\|) \right\}^2 \right) |Aw(s)|^2 ds$$

$$\leq \|w(t_n)\|^2 e^{\mu t_n} + \frac{e^{\mu t} - e^{\mu t_n}}{\mu} \left\{ R_1(t - t_n) \|w(t_n)\|^2 + R_2(t - t_n)\alpha^2 \right\}.$$

Therefore,

$$\|w\|^2 + e^{-\mu(t - t_n)} \int_{t_n}^{t} \left(\nu - e^{\mu\delta} \frac{16\mu^2(t - t_n)^2}{\nu} \left\{ \nu + C\lambda_1^{-1/2}(2\rho_V + \|w(s)\|) \right\}^2 \right) |Aw(s)|^2 ds$$

$$\leq \varphi(t - t_n) \|w(t_n)\|^2 + \frac{1 - e^{-\mu(t - t_n)}}{\mu} R_2(t - t_n)\alpha^2 \tag{3.4}$$

where

$$\varphi(\tau) = e^{-\mu\tau} + \frac{1 - e^{-\mu\tau}}{\mu} R_1(\tau).$$

Since

$$\varphi'(\tau) = -\mu e^{-\mu t} + e^{-\mu\tau} R_1(\tau) + \frac{1 - e^{-\mu\tau}}{\mu} R_1'(\tau)$$

choosing h small enough implies

$$\varphi'(0) = -\mu + R_1(0) = -\mu + \frac{4\mu^2}{\nu} c_1 h^2 < 0.$$

As $\varphi(0) = 1$ and $R_1(0) < \mu$ it follows there is $\varepsilon > 0$ such that

$$\varphi(\tau) < 1 \qquad \text{and} \qquad R_1(\tau) < \mu \qquad \text{for} \qquad \tau \in (0, \varepsilon].$$

We emphasize at this point that ε is independent of α, $\mathcal{M}_0$ and $\mathcal{E}_0$ but depends on h, μ, ν and other parameters of the system.

Now set

$$S^2 = \max \left\{ \mathcal{M}_0^2, \frac{1 - e^{-\mu\varepsilon}}{\mu} R_2(\varepsilon)\mathcal{E}_0^2, \frac{R_2(\varepsilon)}{\mu - R_1(\varepsilon)} \mathcal{E}_0^2 \right\}$$

and choose $\delta \leq \varepsilon$ so small that

$$\nu - e^{\mu\delta} \frac{16\mu^2\delta^2}{\nu} \left\{ \nu + C\lambda_1^{-1/2}(2\rho_V + 2^{1/2}S) \right\}^2 \geq 0.$$

By hypothesis $\|w(t_k)\| \leq \mathcal{M}_0 \leq S < 2^{1/2}S$. Under the assumption on δ we obtain

$$\nu - e^{\mu\delta} \frac{16\mu^2(t - t_n)^2}{\nu} \left\{ \nu + C\lambda_1^{-1/2}(2\rho_V + \|w(s)\|) \right\}^2 > 0$$

11

for $n = k$ at $s = t_k$ and any $t \in [t_k, t_{k+1}]$. Therefore, by continuity the integral term on the left-hand side of (3.4) drops out over some non-trivial maximal interval. But then

$$\|w(t)\|^2 \le \varphi(t - t_k)\|w(t_k)\|^2 + \frac{1 - e^{-\mu(t - t_k)}}{\mu} R_2(t - t_k)\alpha^2$$

$$\le \|w(t_k)\|^2 + \frac{1 - e^{-\mu\varepsilon}}{\mu} R_2(\varepsilon)\mathcal{E}_0^2 \le 2S^2$$

shows this interval is at least as large as $[t_k, t_{k+1}]$.

Taking $t = t_{k+1}$ and $\theta = \varphi(\delta)$ yields

$$\|w(t_{k+1})\|^2 \le \theta\|w(t_k)\|^2 + \frac{1 - e^{-\mu\delta}}{\mu} R_2(\delta)\alpha^2.$$

Moreover, since

$$\theta = e^{-\mu\delta} + \frac{1 - e^{-\mu\delta}}{\mu} R_1(\delta) \qquad \text{implies} \qquad 1 - \theta = \frac{1 - e^{-\mu\delta}}{\mu}\big(\mu - R_1(\delta)\big)$$

then substituting further obtains

$$\|w(t_{k+1})\|^2 \le \theta\|w(t_k)\|^2 + (1 - \theta)\frac{R_2(\delta)}{\mu - R_1(\delta)}\alpha^2.$$

Noting $R_2(\tau)/\big(\mu - R_1(\tau)\big)$ is an increasing function of τ and that $\alpha \le \mathcal{E}_0$ results in

$$\|w(t_{k+1})\|^2 \le \theta\|w(t_k)\|^2 + (1 - \theta)\frac{R_2(\varepsilon)}{\mu - R_1(\varepsilon)}\mathcal{E}_0^2 \le S^2.$$

By induction it follows that $\|w(t_n)\|^2 \le S^2$ holds for all $n = k, \ldots, k+p$ and consequently

$$\|w(t_{n+1})\|^2 \le \theta\|w(t_n)\|^2 + (1 - \theta)\frac{R_2(\delta)}{\mu - R_1(\delta)}\alpha^2 \quad \text{for all} \quad n = k, \ldots, k + p.$$

Another induction immediately yields that

$$\|w(t_n)\|^2 \le \theta^{n-k}\|w(t_k)\|^2 + (1 - \theta)\frac{R_2(\delta)}{\mu - R_1(\delta)}\alpha^2 \sum_{j=0}^{n-k-1} \theta^j$$

$$\le \theta^{n-k}\|w(t_k)\|^2 + \frac{R_2(\delta)}{\mu - R_1(\delta)}\alpha^2$$

and also for $t \in [t_n, t_{n+1}]$ that

$$\|w(t)\|^2 \le \varphi(t - t_n)\|w(t_n)\|^2 + \frac{1 - e^{-\mu(t - t_n)}}{\mu} R_2(t - t_n)\alpha^2.$$

Finally, combining the above two inequalities we obtain

$$\|w(t)\|^2 \le \|w(t_n)\|^2 + \frac{R_2(\delta)}{\mu - R_1(\delta)}\alpha^2 \le \theta^{n-k}\|w(t_k)\|^2 + c\alpha^2.$$

where $c = 2R_2\big(\delta\big)/(\mu - R_1(\delta))$. Noting that c, δ, θ, h and μ depend on $\mathcal{E}_0$ and $\mathcal{M}_0$ but are independent of α and $\|w(t_k)\|$ finishes the proof. $\qquad\square$

4 Bounds on the Expectation

In this section we employ the modification (2.12) to the delay-nudging method of [16] to handle the outliers that result from Gaussian noise processes. Our criterion for the identification of outliers in the observational data is based on known *a priori* bounds of the global attractor for the reference solution.

Intuitively, if M is very large then $\widetilde{J}_h^o$ is the same as $\widetilde{J}_h$ most of the time. On the other hand, the few cases where these interpolants are different is sufficient to ensure the probability distribution behind the modified noise process η_n^o satisfies the hypothesis of Proposition 3.2 and is bounded. We remark that the η_n^o do not form a sequence of independent identically distributed random variables since the removal of outliers is relative to the size of $U(t_n)$ at different points of time.

Fortunately, the pathwise bounds obtained in the previous section do not require the noise be independent or identically distributed; therefore, upon noting $\widetilde{J}_h^o U(t_n) = J_h U(t_n) + \eta_n^o$ and $|\eta_n^o| \leq \mathcal{E}_0$ for suitable choice of M and $\mathcal{E}_0$, Proposition 3.2 can be applied directly to the modified nudging method. To find a suitable M and $\mathcal{E}_0$ we first prove

Lemma 4.1. *If* $M \geq \rho_H + 2\pi c_1^{1/2} \rho_V$, *then for all* $h \leq 2\pi$ *holds* $|\eta_n^o| \leq \mathcal{E}_0$ *where* $\mathcal{E}_0 = 3M$.

Proof. First note that

$$|J_h U(t_n)| \leq |U(t_n)| + |U(t_n) - J_h U(t_n)| \leq \rho_H + c_1^{1/2} h \|U(t_n)\| \leq M. \tag{4.1}$$

Now, if $|\widetilde{J}_h U(t_n)| > 2M$ then $\widetilde{J}_h^o U(t_n) = 0$. Therefore

$$|\eta_n^o| = |\widetilde{J}_h^o U(t_n) - J_h U(t_n)| = |J_h U(t_n)| \leq M.$$

If $|\widetilde{J}_h U(t_n)| \leq 2M$ then $\eta_n^o = \eta_n$. Therefore

$$|\eta_n^o| = |\eta_n| = |\widetilde{J}_h U(t_n) - J_h U(t_n)| \leq |\widetilde{J}_h U(t_n)| + |J_h U(t_n)| \leq 2M + M = 3M.$$

Since $|\eta_n^o| \leq \mathcal{E}_0$ in either case, the bound follows. $\square$

We remark that since $|\eta_n^o| \leq \mathcal{E}_0$ for all n it immediately follows from Proposition 3.2 that there exists c, δ, h and μ such that

$$\limsup_{t \to \infty} \|U - u\|^2 \leq c\mathcal{E}_0^2. \tag{4.2}$$

The drawback of (4.2) is the upper bound does not depend on the variance of the Gaussian noise process. Moreover, as there is always a chance of obtaining a long sequence of outliers in a row, it appears not possible to obtain better pathwise bounds.

For example, if the observational errors were such that $|\widetilde{J}_h U(t_n)| > 2M$ for all t_n over an interval of length T, then for those t_n the modified nudging method would evolve the approximating solution independently of the observations as

$$\frac{du}{dt} = \nu Au + B(u, u) = f - J_h u(t_n) \quad \text{for} \quad t \in [t_n, t_{n+1}).$$

13

Although these dynamics are different than those which govern the reference solution, the main difficulty is lack of any coupling along with sensitive dependence on initial conditions would lead U and u to become decorrelated when T is large. Thus, no pathwise bounds qualitatively better than (4.2) are available because a sequence of such outliers, though unlikely, will repeatedly occur with probability one over any infinite period of time.

We now proceed to the proof of Theorem 2.6, our main theoretical result, which provides a bound on $\mathbf{E}\big[\|U - u\|^2\big]$ that vanishes as σ^2 tends to zero.

Proof of Theorem 2.6. Fix $\mathcal{M}_0 = \rho_V$ and $\mathcal{E}_0 = 3M$ where $M = \rho_H + 2\pi c_1^{1/2}\rho_V$. Note that $u(t_0) = 0$ and $U(t_0) \in \mathcal{A}$ implies

$$\|w(t_0)\| = \|U(t_0) - u(t_0)\| = \|U(t_0)\| \leq \rho_V = \mathcal{M}_0.$$

Now, provided $h \leq 2\pi$ then Lemma 4.1 implies the noise present in the modified interpolant satisfies $\|\eta_n^o\|_{L^2} \leq \mathcal{E}_0$ for all n. It follows from Proposition 3.2 that there are positive constants c, δ, θ, h and μ such that (3.2) holds. Therefore, upon setting $\alpha = \mathcal{E}_0$ we obtain

$$\|w(t)\|^2 \leq cK \quad \text{for all} \quad t \geq t_0 \qquad \text{where} \qquad cK = \rho_V^2 + c\mathcal{E}_0^2.$$

We remark that the choice of h provided above fixes the interpolant J_h and consequently the exact form of the noise process η_n in $\widetilde{J}_h$. In particular N and the unit vectors $\psi_i \in H$ for $i = 1, \ldots, 2N$ should be considered fixed for the rest of the proof.

For each α such that $0 < \alpha \leq M$ let n_α be the value of n such that $\theta^n K \leq \alpha^2 < \theta^{n-1}K$. Since $K > \mathcal{E}_0^2 > \alpha^2$ it follows that $n_\alpha \geq 1$. Consequently, if

$$|\eta_n| \leq \alpha \qquad \text{for} \qquad n = k, \ldots, k + n_\alpha$$

then by (4.1) we have

$$|\tilde{J}_h U(t_n)| = |J_h U(t_n)| + |\eta_n| \leq M + \alpha \leq 2M.$$

This means $\eta_n = \eta_n^o$ and so $|\eta_n^o| \leq \alpha$. It again follows from (3.2) with the same choices of c, δ, θ, h and μ as before that

$$\|w(t)\|^2 \leq \theta^{n-k}\|w(t_k)\|^2 + c\alpha^2 \leq 2c\alpha^2 \quad \text{for} \quad t \in [t_n, t_{n+1}] \quad \text{where} \quad n = k + n_\alpha.$$

Therefore,

$$\mathbf{P}\big\{\|w(t)\|^2 \leq 2c\alpha^2\big\} \geq \mathbf{P}\big\{|\eta_n| \leq \alpha \text{ for } n = k, \ldots, k + n_\alpha\big\}$$
$$= \prod_{n=k}^{k+n_\alpha} \mathbf{P}\big\{|\eta_n| \leq \alpha\big\}. \tag{4.3}$$

Recall the η_n are distributed as in (2.7) where $|\eta|^2 = X^T \Psi X \leq \|\Psi\|\|X\|^2$. Here $\|\Psi\|$ denotes the spectral norm of the $2N \times 2N$ matrix Ψ and $\|X\|$ the Euclidean norm of the random vector X. Thus,

$$\mathbf{P}\big\{|\eta_n| \leq \alpha\big\} = \mathbf{P}\big\{|\eta| \leq \alpha\big\} \geq \mathbf{P}\big\{\|X\|^2 \leq \alpha^2/\|\Psi\|\big\}. \tag{4.4}$$

We emphasize that $\|\Psi\|$ is simply a constant at this point as we have already fixed N and ψ_i earlier in the proof. Note also that Ψ is independent of α and σ.

Substituting (4.4) into (4.3) yields

$$\mathbf{P}\big\{\|w(t)\|^2 \le 2c\alpha^2\big\} \ge \varphi \qquad \text{where} \qquad \varphi = \mathbf{P}\big\{\|X\|^2 \le \alpha^2/\|\Psi\|\big\}^{n_\alpha+1}.$$

Since

$$\mathbf{E}\big[\|w(t)\|^2\big] \le 2c\alpha^2 \mathbf{P}\big\{\|w(t)\|^2 \le 2c\alpha^2\big\} + \mathbf{P}\big\{\|w(t)\|^2 > 2c\alpha^2\big\}cK,$$

then $2c\alpha^2 \le 2cM^2 = 2c(\mathcal{E}_0/3)^2 < cK$ implies

$$\mathbf{E}\big[\|w(t)\|^2\big] \le \varphi(2c\alpha^2) + (1-\varphi)cK \quad \text{for} \quad t \in [t_n, t_{n+1}] \quad \text{where} \quad n = n_\alpha. \tag{4.5}$$

Now rewrite the bound in (4.5) in terms of σ by choosing a suitable α. By definition

$$n_\alpha \log\theta \le \log(\alpha^2/K) < (n_\alpha - 1)\log\theta$$

and so

$$n_\alpha - 1 \le \frac{\log(K/\alpha^2)}{\log(1/\theta)} < n_\alpha.$$

Therefore,

$$\varphi \ge \mathbf{P}\big\{\|X\|^2 \le \alpha^2/\|\Psi\|\big\}^{\log(K/\alpha^2)/\log(1/\theta)+2} = \mathbf{P}\big\{\|X\|^2 \le \alpha^2/\|\Psi\|\big\}^{(\log(K/\alpha^2)+\kappa)/\log(1/\theta)}$$

where $\kappa = 2\log(1/\theta)$.

Assuming $8\sigma^2 \le \alpha^2/\|\Psi\|$, then Lemma 2.3 used as (2.8) implies

$$\mathbf{P}\big\{\|X\|^2 \le \alpha^2/\|\Psi\|\big\} = 1 - \mathbf{P}\big\{\|X\|^2 \ge \alpha^2/\|\Psi\|\big\} \ge 1 - e^{-x}$$

where $x > 0$ is defined by $2\sigma^2\sqrt{x} + 2\sigma^2 x = \alpha^2/\|\Psi\| - \sigma^2$. Simple estimates then yield

$$x \ge \Big(\frac{\sqrt{15}-1}{\sqrt{32}}\Big)^2 \frac{\alpha^2}{\sigma^2\|\Psi\|} > \frac{\alpha^2}{4\sigma^2\|\Psi\|}.$$

We remark that although $\sigma^2 < \alpha^2/\|\Psi\|$ is sufficient to apply Lemma 2.3, the explicit lower bound on x scaling as α^2/σ^2 is used in our subsequent estimates. It follows that

$$\mathbf{P}\big\{\,\|X\|^2 \le \alpha^2/\|\Psi\|\,\big\} \ge 1 - e^{-\gamma\alpha^2/\sigma^2} \qquad \text{where} \qquad \gamma = \frac{1}{4\|\Psi\|}.$$

By Bernoulli's inequality

$$\varphi \ge \big(1 - e^{-\gamma\alpha^2/\sigma^2}\big)^{(\log(K/\alpha^2)+\kappa)/\log(1/\theta)} \ge 1 - e^{-\gamma\alpha^2/\sigma^2}\big(\log(K/\alpha^2)+\kappa\big)/\log(1/\theta)$$

so that

$$1 - \varphi \le e^{-\gamma\alpha^2/\sigma^2}\big(\log(K/\alpha^2)+\kappa\big)/\log(1/\theta).$$

Since $\mathcal{E}_0 = 3M$ and $\alpha \le M$ then $9\alpha^2 \le \mathcal{E}_0^2$. The definition $cK = \rho_V^2 + c\mathcal{E}_0^2$ then implies $9\alpha^2 < K$ or that $K/\alpha^2 \ge 9$. Moreover, $\kappa/\log(1/\theta) = 2$. Therefore

$$\log(K/\alpha^2) + \log\Big(\frac{\log(K/\alpha^2)+\kappa}{\log(1/\theta)}\Big) \ge \log 9 + \log 2 = \log(18).$$

Setting

$$\sigma^2 = \gamma\alpha^2\left\{\log(K/\alpha^2) + \log\left(\frac{\log(K/\alpha^2) + \kappa}{\log(1/\theta)}\right)\right\}^{-1} \tag{4.6}$$

implies

$$8\sigma^2 \le \frac{8\gamma\alpha^2}{\log(18)} = \frac{8\alpha^2}{4\|\Psi\|\log(18)} < \frac{\alpha^2}{\|\Psi\|}.$$

Thus, the previous assumption on the smallness of σ is satisfied. Moreover, it follows that

$$(1 - \varphi)K \le \alpha^2.$$

Consequently

$$\mathbf{E}\left[\|w(t)\|^2\right] \le \varphi(2c\alpha^2) + (1 - \varphi)cK \le 3c\alpha^2.$$

Finally, since $8\sigma^2 \le \alpha^2/\|\Psi\|$ then

$$\log(K/\alpha^2) \le \log(C_1/\sigma^2) \qquad \text{where} \qquad C_1 = \frac{K}{8\|\Psi\|}.$$

Substitute this inequality into (4.6) to obtain

$$\alpha^2 \le \frac{\sigma^2}{\gamma}\left\{\log(C_1/\sigma^2) + \log\left(\frac{\log(C_1/\sigma^2) + \kappa}{\log(1/\theta)}\right)\right\}.$$

Defining

$$f(\sigma) = \log(C_1/\sigma^2) + \log\left(\frac{\log(C_1/\sigma^2) + \kappa}{\log(1/\theta)}\right) \quad \text{and} \quad C_0 = \frac{3c}{\gamma} = 12c\|\Psi\|$$

then finishes the proof. $\qquad\qquad\square$

5   Comparison with Simulation

The purpose of this section is to compare the analytical results from Section 4 to numerical simulation, note areas of interest for future work and make some concluding remarks. To this end we consider a natural class of type-I interpolants obtained from noisy observational measurements given by local spatial averages taken near distinct points in space. We then choose numerical values of h, δ and μ that do not necessary satisfy the conditions of Proposition 3.2 but work in practice. Note that the existence of such optimistic choices for h, δ and μ is guided by previous computational experience, see for example [26], [27] and [18], and stems from the fact that even the simplest analytic bounds which determine ρ_H and ρ_V appear orders of magnitude too large. Moreover, the existing proof techniques rely essentially on the dissipation even though convection appears to play an important physical role in generating the small scales from the large.

Before beginning we remark that all computations were performed on a 512×512 spatial grid using a fully-dealiased pseudo-spectral method with time steps of size $dt = .015625$ carried out by means of the fourth-order exponential time integrator proposed by Cox and Matthews [12] and regularized as suggested by Kassam and Trefethen [22].

Let $d(x, y)$ be distance in the 2π-periodic domain $\mathbf{T}$ given by

$$d(x, y) = \min\left\{ |x - y + 2\pi m| : m \in \mathbf{Z}^2 \right\}.$$

Choose $x_i \in \mathbf{T}$ for $i = 1, \dots, N$ and for $r > 0$ fixed consider the noise-free measurements m_j at time t about those locations given by

$$m_j(t) = \frac{1}{\pi r^2} \int_{d(y,x_j)<r} U(t, y)dy.$$

We remark that the integrals over balls of radius r in the definition of m_j represent the fact that physical instrumentation always measures a local average of the velocity near a point rather than the exact velocity at a single point in space. Mathematically, this averaging provides a regularizing effect proven in Appendix A that leads to a type-I interpolant observable. In our numerics we take $r = 2\pi\sqrt{6}/512 \approx 0.0300598$. This corresponds to an average of 21 points from the spatial grid per measurement.

Figure 1: The left shows the Voronoi tessellation resulting from 81 points arranged in a rectangular grid; the right illustrates the same same number of points chosen at random. The circles depict the discs over which the local averages are taken that represent the observational measurements of the velocity field.
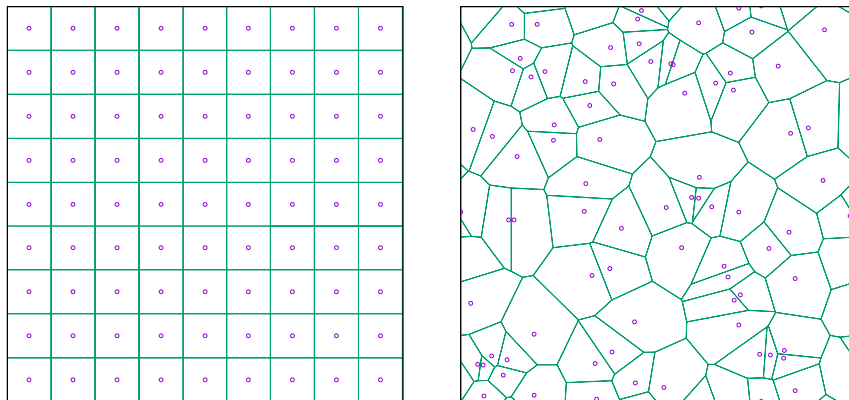
Interpolate the measurements m_j every δ units in time using simple a piecewise-constant interpolant I_h onto the Voronoi tessellation given by the points x_i. Thus,

$$I_h U(t_n, x) = \sum_{j=1}^{N} \chi_j(x) m_j(t_n) \qquad \text{where} \qquad t_n = \delta n \qquad (5.1)$$

and χ_j is the characteristic function given by

$$\chi_j(x) = \begin{cases} 1 & \text{for } d(x, x_j) \leq \min\left\{ d(x, x_i) : i = 1, \dots, N \right\} \\ 0 & \text{otherwise.} \end{cases}$$

On the left in Figure 1 is a depiction of the tessellation when the points x_i lie on a regular grid. The right illustrates points chosen randomly. In the present paper we take $\delta = 1$ and

focus on the regular 9×9 grid with $N = 81$ while noting when the points x_i do not lie on such a grid that convergence of the approximating solution to the free-running solution can be more erratic. Similar effects have been observed by Desamsetti, Dasari, Langodan, Titi, Knio and Hotiet [13] in the context of an operational mesoscale weather-prediction system and also by Carlson, Van Roekel, Godinez, Petersen and Larios [8] for an ocean model test case that simulates a wind-driven double gyre. Further study of observational data that exhibits areas of low-resolution spatial measurements along with localized areas of high-resolution measurements is planned for a future work. We remark that the related situation where all observations as well as the effects of the nudging are confined to a subdomain of $\mathbf{T}$ was considered by Biswas, Bradshaw and Jolly in [3].

Define the length scale h by

$$h = \max_{x \in \mathbf{T}} \min \left\{ d(x, x_i) : i = 1, \ldots, N \right\}$$

where intuitively $1/h$ is the minimum observational density. For a regular 9×9 grid one has $h = \pi\sqrt{2}/9 \approx 0.49365$; however, due to the 512×512 spatial discretization of the domain $\mathbf{T}$ our numerics actually satisfied $h = 56\pi\sqrt{2}/512 \approx 0.48594$.

Under the assumptions stated in Theorem A.1 there exists a constant c_0 such that

$$\|\Phi - I_h\Phi\|_{L^2}^2 \leq c_0 h^2 \|\Phi\|^2 \qquad \text{for all} \qquad \Phi \in V.$$

We remark for the values of r and h used in our simulations that $\gamma = r/h \approx 0.061859$ and $h = \tilde{h}$ implies taking $c_0 = 2 \cdot 3^6/(\pi\gamma^2) \approx 121284$ is sufficient. This notably large number demonstrates one of the limitations in using small spatial averages to obtain a type-1 interpolant observable and will play a role shortly when it comes to filtering the outliers.

Although setting $J_h = P_H I_h$ would satisfy Definition 2.2 and result in a type-I interpolant observable, we further include a spatial smoothing filter to remove the high-frequency spillover that would otherwise result from the discontinuities in I_h. The importance of such spatial smoothing was demonstrated computationally in [18] and further employed in [9] for the analysis of a different discrete-in-time data-assimilation algorithm.

In this paper we employ a smoothing filter given by the spectral projection

$$P_\lambda U = \sum_{|k|^2 \leq \lambda} U_k e^{ik \cdot x} \qquad \text{where} \qquad U = \sum_{k \in \mathbf{Z}^2} U_k e^{ik \cdot x}$$

and define $J_h = P_\lambda P_H I_h$. One advantage of using P_λ in our interpolant observable rather than a different spatial filter is the simple way the improved Poincaré inequality

$$\lambda |(I - P_\lambda)\Phi|^2 \leq \|(I - P_\lambda)\Phi\|^2$$

along with the Pythagorean theorem

$$|\Phi - J_h\Phi|^2 = |(I - P_\lambda)\Phi|^2 + |P_\lambda(\Phi - P_H I_h\Phi)|^2 \leq (\lambda^{-1} + c_0 h^2)\|\Phi\|^2,$$

and a fixed constant c_3 yields a smooth type-I interpolant observable with $c_1 = c_3 + c_0$ provided $\lambda^{-1} \leq c_3 h^2$. From a practical point of view it is reasonable for the resolution of the spectral filter and the interpolant observable to be comparable.

In our numerics we take $\lambda = 80$. After accounting for the conjugate symmetry $U_k = -\overline{U_{-k}}$ in the Fourier transform of a real vector field,

$$\mathrm{card}\{\, k \in \mathbf{Z}^2 : 0 < |k|^2 \leq 80 \,\} = 248$$

implies P_λ consists of 124 independent Fourier modes. This is similar in quantity to the 81 local averages which appear in the unfiltered interpolant I_h of the 9×9 spatial grid.

Having thus described J_h we now detail the noise process that leads to the noisy interpolant $\widetilde{J}_h U(t_n)$. Suppose at each time t_n the measurements $m_j(t_n) \in \mathbf{R}^2$ are contaminated by Gaussian errors such that

$$\widetilde{m}_j(t_n) = m_j(t_n) + \varepsilon_{2j-1} Y_{2j-1,n} + \varepsilon_{2j} Y_{2j,n} \qquad \text{for} \qquad j = 1, \ldots, N$$

where $\varepsilon_j \in \mathbf{R}^2$ and $Y_{j,n}$ form a family of independent standard normal random variables.

Upon setting

$$\widetilde{J}_h U(t_n) = P_\lambda P_H \widetilde{I}_h U(t_n) \qquad \text{where} \qquad \widetilde{I}_h U(t_n, x) = \sum_{j=1}^{N} \chi_j(x) \widetilde{m}_j(t_n)$$

we obtain that

$$\eta_n = \widetilde{J}_h U(t_n) - J_h U(t_n) = P_\lambda P_H \sum_{j=1}^{N} \chi_j (\varepsilon_{2j-1} Y_{2j-1,n} + \varepsilon_{2j} Y_{2j,n}) = \sum_{i=1}^{2N} \sigma_i Y_{i,n} \psi_i$$

where

$$\sigma_i = |P_\lambda P_H \chi_{\lfloor (i+1)/2 \rfloor} \varepsilon_i| \qquad \text{and} \qquad \psi_i = P_\lambda P_H \chi_{\lfloor (i+1)/2 \rfloor} \varepsilon_i / \sigma_i$$

for $i = 1, \ldots, 2N$ and $\lfloor (i+1)/2 \rfloor$ denotes the greatest integer less than or equal to $(i+1)/2$. Thus, η_n has the form hypothesized in (2.7).

In our numerics we take

$$\varepsilon_i = \frac{\varepsilon}{2\pi\sqrt{2}} \begin{cases} e_1 & \text{for } i \text{ even} \\ e_2 & \text{for } i \text{ odd} \end{cases} \tag{5.2}$$

where $e_1 = (1,0)$, $e_2 = (0,1)$ and $\varepsilon > 0$. Note the $2\pi\sqrt{2}$ normalization was chosen so the noise in unfiltered interpolant $\widetilde{I}_h$ satisfies $\mathbf{V}[\widetilde{I}_h] = \varepsilon^2$. Moreover,

$$\sigma^2 = \sum_{i=1}^{2N} \sigma_i^2 = \varepsilon^2 \sum_{j=1}^{N} \left(|P_\lambda P_H \chi_j e_1|^2 + |P_\lambda P_H \chi_j e_2|^2 \right) \approx (0.40058)\varepsilon^2$$

shows σ^2 is proportional to ε^2 which we shall vary as $\varepsilon = 10^{-\ell}$ for $\ell = 4, \ldots, 8$ as a test on the bounds in Theorem 2.6.

The time-independent body forcing f used for our simulations satisfies

$$f = \sum_{\lambda_m \leq |k|^2 \leq \lambda_M} f_k e^{ik \cdot x} \qquad \text{with} \qquad k \cdot f_k = 0 \tag{5.3}$$

Figure 2: The left shows level-curves of curl f spaced 0.03 apart with positive in red, negative in blue and zero in black where f is the time-independent body forcing used in all the simulations. the right shows level-curves of the vorticity curl U_0 spaced 0.2 apart where U_0 is the initial condition of the free-running solution.
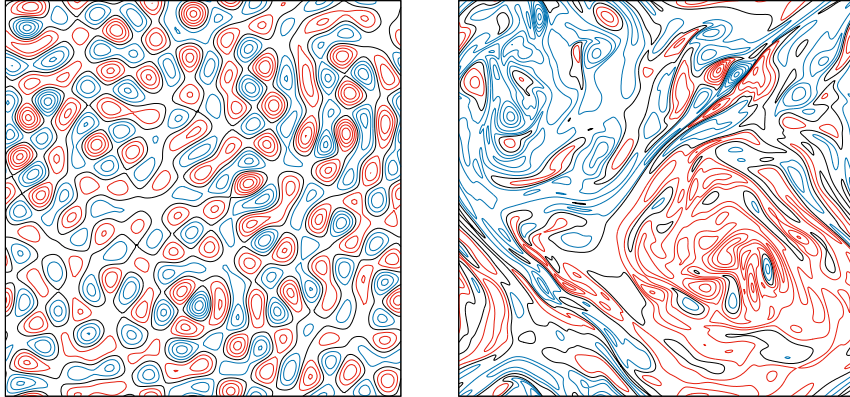
where $\lambda_m = 100$ and $\lambda_M = 142$. Here the Fourier coefficients f_k are time independent. They were obtained by projecting randomly chosen values onto H and then rescaling them to achieve the L^2 norm $|f| = 0.025$. Upon setting the viscosity $\nu = 0.0001$ we further have that the Grashof number of the free-running solution satisfies

$$\mathbf{Gr}(f) = \frac{1}{\lambda_1 \nu^2}|f| = 2.5 \times 10^6.$$

Here $\lambda_1 = 1$ due to the 2π-periodic boundary conditions. Recall also that $|f|$ refers to the L^2 norm, which in terms of the Fourier series may be computed as

$$|f| = 2\pi \left\{ \sum_{\lambda_m \leq |k|^2 \leq \lambda_M} |f_{k,1}|^2 + |f_{k,2}|^2 \right\}^{1/2} \qquad \text{where} \qquad f_k = (f_{k,1}, f_{k,2}).$$

Figure 2 illustrates the forcing function used in our simulations on the left. The exact values of the Fourier modes f_k corresponding to this force are given in [27] and are known to lead to a complicated time-dependent flow U.

The initial condition U_0 for the reference solution is theoretically assumed to lie on the global attractor. Numerically, we take U_0 as the final state of a long-time integration of (1.1) starting from zero at time $t = -20480$ in the past. Note the forcing function given by (5.3) is sufficient such that by time $t = 0$ all modes such that $k \neq 0$ which satisfy the 2/3 antialiasing condition are present in the solution. Furthermore, the energetics of the flow appear to represent the long-time statistical behavior that results from the forcing. The graph on the right in Figure 2 depicts the initial condition for the reference trajectory used for our data-assimilation experiments.

For comparison with our theory, we note the estimates in [26] applied to the forcing given

by (5.3) lead to the *a priori* bounds

$$\rho_H \leq \frac{1}{\lambda_m^{1/2}}\rho_V = 25 \qquad \text{and} \qquad \rho_V \leq \frac{1}{\lambda_1^{1/2}\nu}|f| = 250.$$

Therefore, a suitable theoretical bound to filter the outliers satisfies

$$M = \rho_H + c_1^{1/2}h\rho_V \approx 25 + c_1^{1/2}(121.49) \approx 42333.$$

Setting $M = 42333$ results in essentially no outliers being filtered by our modifications to the original time-delay nudging method. The size of M arises directly from our previous estimate on c_1 along with values of ρ_H and ρ_V that, as already mentioned, appear much larger than needed. The possibility of empirically tuning this cutoff to achieve better numerical results is interesting but outside the scope of the present paper.

To compute the expectation $\mathbf{E}\big[\|U - u\|^2\big]$ whose bounds are the subject of Theorem 2.6 we consider the time evolution of an ensemble of 500 approximating solutions obtained from observational measurements of the same reference solution U each contaminated by different realizations of the Gaussian noise η_n parameterized by $\omega_j \in \Omega$ where Ω is a suitable probability space. Thus, $u(t; w_j)$ is a random point in V and we approximate

$$\mathbf{E}\big[\|U(t) - u(t)\|^2\big] \approx \frac{1}{500}\sum_{j=1}^{500}\|U(t) - u(t;\omega_j)\|^2.$$

The statistics of the pathwise trajectories may further be characterized by the intervals $I_p = [a, b]$ for $p \in [0, 1]$ where $\mathbf{P}\big\{\|U - u\|^2 \in I_p\big\} = p$ by choosing a and b such that

$$\mathbf{P}\big\{\|U - u\|^2 \geq a\big\} = (1 + p)/2 \qquad \text{and} \qquad \mathbf{P}\big\{\|U - u\|^2 \leq b\big\} = (1 + p)/2.$$

At any point in time approximations for a and b are given by the thresholds below and above which the desired percentage of trajectories in the ensemble lie.

Parts of the time evolution of $\mathbf{E}\big[\|U - u\|^2\big]$ for selected values of $\sigma^2 = (0.40058)\varepsilon^2$ is depicted in Figure 3 along with a shaded region that indicates the intervals I_p corresponding to $p = 88, 60$ and 30. Upon characterizing the average large-eddy turnover time in the reference solution by

$$\tau = \frac{4\pi^2}{T}\int_0^T \|U\|_{H^{-1/2}(\mathbf{T})}^2 \Big/ \Big(\frac{1}{T}\int_0^T \|U\|_{L^2(\mathbf{T})}^2\Big)^{3/2} \approx 30.585,$$

we note the full trajectory of the ensemble was computed using 655360 time steps until time $T = 10240$ or equivalently for $T/\tau \approx 334.8$ large-eddy turnovers.

We remark by $t = 1500$ sufficient time has passed over which the initial condition $u_0 = 0$ is forgotten and after which the expected value of $\mathbf{E}\big[\|U - u\|^2\big]$ starts fluctuating about its mean. The maximum bound for $t \in [1500, T]$ is determined by the interval $t \in [2000, 2500]$ illustrated in Figure 3. Since the same reference solution U was used for each run, it is only a little surprising that this maximum occurs in the same place for the different choices of ε. More notable and further illustrated in Figure 4 is the fact that this maximum is about 100 times larger than the time-averaged value taken over the same interval.

Figure 3: Evolution of $\mathbf{E}\big[\|U - u\|^2\big]$ for a fixed solution U with approximating solutions u obtained by observational measurements contaminated by Gaussian noise where $\varepsilon \in \{10^{-4}, \ldots, 10^{-8}\}$. The shaded regions illustrate the bounds I_p on $p = 88$, 70 and 40 percent of the 500 paths in each ensemble. The $\varepsilon = 10^{-5}$ and 10^{-7} trajectories have been omitted for $t \in [2000, 2500]$ to avoid overlap.
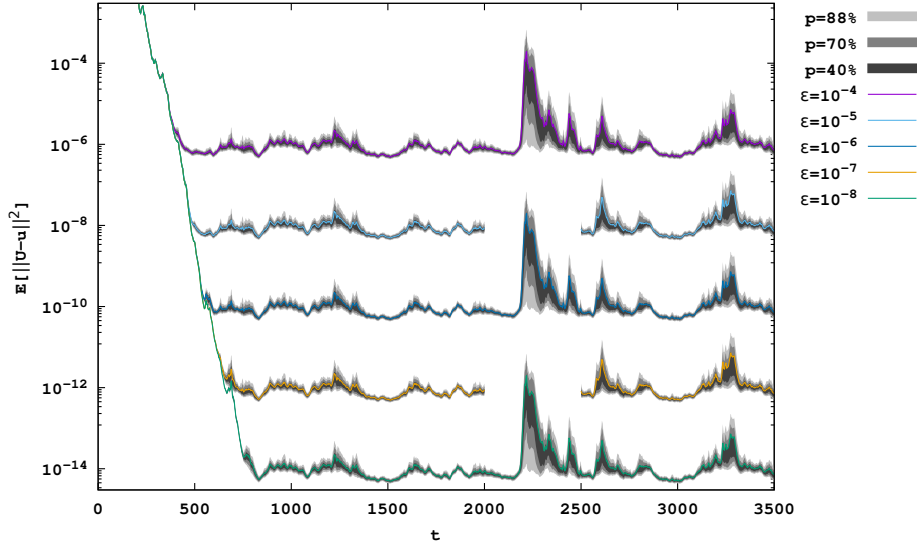

Computations not reported here suggest even greater excursions from the average can happen when the measurement points x_i are not given by a uniform grid. At the same time, simulations performed by Law, Sanz–Alonso, Shukla and Stuart [25] for the Lorenz 96 model suggest that a careful placement of the points x_i could yield more accurate synchronization between U and u than a uniform grid. Improved results for observations which scan the domain over time were obtained by Larios and Victor [23] for the Chafee–Infante equations, by Franz, Larios and Victor [17] for the two-dimensional Navier–Stokes equations and by Biswas, Bradshaw and Jolly [3] see also [4] for the two-dimensional Navier–Stokes equations. This leads us to an interesting question for future investigation: If the measurement points themselves are advected by the flow over time like buoys in the ocean, could the resulting distribution of the x_i be better than a uniform grid?

In either case, the 100-fold difference between the maximum of $\mathbf{E}\big[\|U - u\|^2\big]$ and the time-averaged value observed in the present simulations may be due to the fact that our numerical choices for h, δ and μ are significantly more optimistic than guaranteed by the theory. In particular, if h and δ were smaller with μ and perhaps r larger, it is reasonable to suppose such large fluctuations would not happen over time.

We conclude by noting the graph displayed as Figure 4 shows that the maximum, minimum and average values of $\mathbf{E}\big[\|U - u\|^2\big]$ are all proportional to σ^2 for over 10 decimal orders of magnitude. This suggests, except for the constants and the logarithmic correction, that the qualitative behavior of the analytic bound in Theorem 2.6 is physically reasonable.

Figure 4: Dependency of bounds on $\mathbf{E}\big[\|U-u\|^2\big]$ for $t \in [1500, 10240]$ as a function of σ^2. The values denoted by *max* are numerical versions of the analytic bounds provided by Theorem 2.6 and depend on σ^2 in a similar way. Values for *avg* and *min* are shown for comparison.
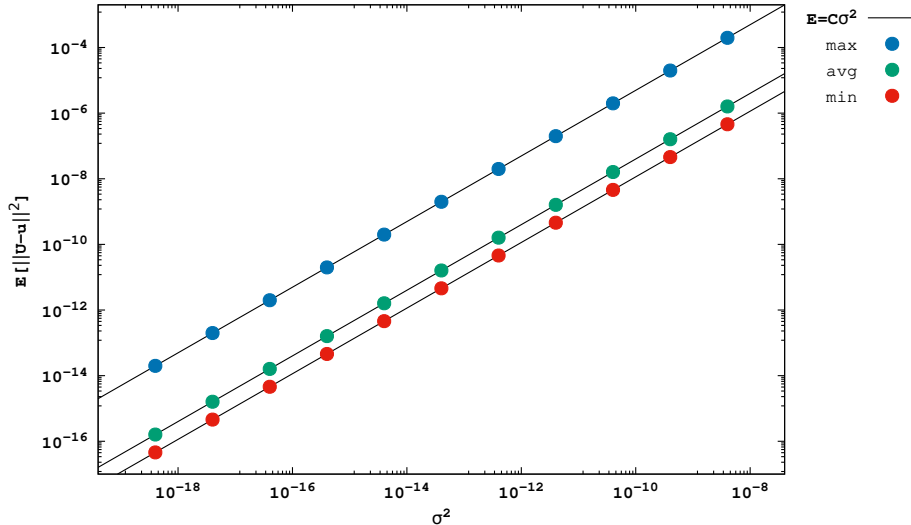
A Local Spatial Averages

In this appendix we show the interpolant observable I_h defined by (5.1) is a type-I interpolant observable. Mathematically this interpolant may be seen as a physically reasonable mix of the determining nodes and volume elements studied by Jones and Titi in [21]. Figure 1 illustrates the variety of interpolants treated by this construction. Note that the discs which represent the spatial averages in our observational measurements overlap in the tessellation on the right. Note also that the maximum distance h between any point in the domain and the nearest point of observation is much less for the regular grid.

Having examined two examples of the interpolants under consideration, we now prove

Theorem A.1. *Suppose r is proportional to h such that $r = \gamma h$ for some $\gamma \in (0,1)$. Then the piecewise constant interpolant $I_h \colon V \to L^2(\mathbf{T})$ defined by (5.1) satisfies*

$$\|\Phi - I_h(\Phi)\|_{L^2(\mathbf{T})}^2 \le c_0 h^2 \|\Phi\|_{H^1(\mathbf{T})}^2 \qquad \text{for all} \qquad \Phi \in V$$

where

$$c_0 = \frac{2^5 3^6}{\pi \gamma^2} \qquad and \qquad h = \sup_{x \in \mathbf{T}} \min\{\, d(x, x_i) : i = 1, \ldots, N \,\}. \qquad (\text{A.1})$$

Proof. Without loss of generality, we shall assume $h = 2\pi/\kappa$ for some $\kappa \in \mathbf{N}$. It this were not the case, there would be $\kappa \in \mathbf{N}$ such that

$$\frac{2\pi}{\kappa + 1} < h \le \frac{2\pi}{\kappa}.$$

Then, upon replacing h by $\tilde{h} = 2\pi/\kappa$ in the following proof and noting $\tilde{h} \le (\kappa + 1)h/\kappa \le 2h$, the desired result for any h could be obtained by a simple modification of the constant c_0.

We note before proceeding, that the value of c_0 stated in (A.1) has, in fact, already been modified by a factor of 16 to take the general case into account.

Now, divide the torus $\mathbf{T}$ into $M = \kappa^2$ equal squares with sides of length h denoted

$$R_{pq} = [(p-1)h, ph) \times [(q-1)h, qh) \qquad \text{for} \qquad p, q = 1, \ldots, \kappa.$$

Given $\Phi \colon \mathbf{T} \to \mathbf{R}^2$ extend Φ by periodicity to all of $\mathbf{R}^2$. For convenience, we shall continue to call the extended function Φ and assume that $\Phi(x + 2\pi m) = \Phi(x)$ for every $x \in \mathbf{R}^2$ and $m \in \mathbf{Z}^2$. Having done this, we define

$$\mathcal{U}_j = \left\{\, x \in \mathbf{R}^2 : |x - x_j| \le \min\{\, d(x, x_i) : i = 1, \ldots, N \,\} \,\right\}$$

and note that

$$I_h(\Phi)(\xi, \eta) = \frac{1}{\pi r^2} \int_{B_r(x_j)} \Phi(s, t)\, ds\, dt \qquad \text{for every} \qquad (\xi, \eta) \in \mathcal{U}_j.$$

Thus, after periodic extension, any integral over $\mathbf{T}$ is equal to the same integral over $\bigcup_{j=1}^m \mathcal{U}_j$.

Next, define $\mathcal{J}_{pq} = \{\, j : x_j \in R_{pq} \,\}$. Since the R_{pq} are disjoint, then $\mathcal{J}_{pq}$'s form a partition of $\{1, \ldots, n\}$. Expand the R_{pq} by h in each dimension to obtain

$$Q_{pq} = [(p-2)h, (p+1)h) \times [(q-2)h, (q+1)h) \quad \text{for} \quad p, q = 1, \ldots, \kappa.$$

It follows that $\mathcal{U}_j \times B_r(x_j) \subseteq Q_{pq}^2$ for each $j \in \mathcal{J}_{pq}$. Consequently,

$$\|\Phi - I_h\Phi\|_{L^2(\mathbf{T})}^2 = \int_{\mathbf{T}} \left|\Phi(\xi, \eta) - I_h(\Phi)(\xi, \eta)\right|^2 d\xi, d\eta$$

$$= \sum_{j=1}^n \int_{\mathcal{U}_j} \left|\Phi(\xi, \eta) - \frac{1}{\pi r^2} \int_{B_r(x_j)} \Phi(s, t)\, ds\, dy\right|^2 d\xi\, d\eta$$

$$= \frac{1}{\pi^2 r^4} \sum_{j=1}^n \int_{\mathcal{U}_j} \left|\int_{B_r(x_j)} \big(\Phi(\xi, \eta) - \Phi(s, t)\big)\, ds\, dt\right|^2 d\xi\, d\eta$$

$$\le \frac{1}{\pi r^2} \sum_{j=1}^n \int_{\mathcal{U}_j \times B_r(x_j)} \left|\Phi(\xi, \eta) - \Phi(s, t)\right|^2 ds\, dt\, d\xi\, d\eta$$

$$= \frac{1}{\pi r^2} \sum_{p,q=1}^\kappa \sum_{j \in \mathcal{J}_{pq}} \int_{\mathcal{U}_j \times B_r(x_j)} \left|\Phi(\xi, \eta) - \Phi(s, t)\right|^2 ds\, dt\, d\xi\, d\eta$$

$$\le \frac{1}{\pi r^2} \sum_{p,q=1}^\kappa \int_{Q_{pq}^2} \left|\Phi(\xi, \eta) - \Phi(s, t)\right|^2 ds\, dt\, d\xi\, d\eta.$$

We remark that the last inequality in the estimate above holds because the sets $\mathcal{U}_j$ are disjoint except for a set of measure zero, which implies the same for $\mathcal{U}_j \times B_r(x_j)$.

Since $(\xi, \eta) \in Q_{pq}$ and $(s, t) \in Q_{pq}$, we continue to estimate as

$$\left|\Phi(\xi, \eta) - \Phi(s, t)\right| \le \left|\Phi(\xi, \eta) - \Phi(s, \eta)\right| + \left|\Phi(s, \eta) - \Phi(s, t)\right|$$

$$= \left|\int_s^\xi \Phi_x(\alpha, \eta)\, d\alpha\right| + \left|\int_t^\eta \Phi_y(s, \beta)\, d\beta\right|$$

$$\le \int_{(p-2)h}^{(p+1)h} \left|\Phi_x(\alpha, \eta)\right| d\alpha + \int_{(q-2)h}^{(q+1)h} \left|\Phi_y(s, \beta)\right| d\beta.$$

This implies

$$\left|\Phi(\xi,\eta)-\Phi(s,t)\right|^2 \le 2\Big(\int_{(p-2)h}^{(p+1)h}\left|\Phi_x(\alpha,\eta)\right|d\alpha\Big)^2 + 2\Big(\int_{(q-2)h}^{(q+1)h}\left|\Phi_y(s,\beta)\right|d\beta\Big)^2$$
$$\le 6h\int_{(p-2)h}^{(p+1)h}\left|\Phi_x(\alpha,\eta)\right|^2 d\alpha + 6h\int_{(q-2)h}^{(q+1)h}\left|\Phi_y(s,\beta)\right|^2 d\beta.$$

It follows that

$$\int_{Q_{pq}^2}\left|\Phi(\xi,\eta)-\Phi(s,t)\right|^2 d\xi\,d\eta\,ds\,dt$$
$$\le \int_{(p-2)h}^{(p+1)h}\int_{(q-2)h}^{(q+1)h}\int_{(p-2)h}^{(p+1)h}\int_{(q-2)h}^{(q+1)h}\left|\Phi(\xi,\eta)-\Phi(s,t)\right|^2 d\xi\,d\eta\,ds\,dt$$
$$\le 6h\Big(\int_{(p-2)h}^{(p+1)h}\int_{(q-2)h}^{(q+1)h}\int_{(p-2)h}^{(p+1)h}\int_{(q-2)h}^{(q+1)h}\int_{(p-2)h}^{(p+1)h}\left|\Phi_x(\alpha,\eta)\right|^2 d\xi\,d\eta\,ds\,dt\,d\alpha$$
$$+\int_{(p-2)h}^{(p+1)h}\int_{(q-2)h}^{(q+1)h}\int_{(p-2)h}^{(p+1)h}\int_{(q-2)h}^{(q+1)h}\int_{(q-2)h}^{(q+1)h}\left|\Phi_y(s,\beta)\right|^2 d\xi\,d\eta\,ds\,dt\,d\beta\Big)$$
$$= 6h\Big(27h^3\int_{(q-2)h}^{(q+1)h}\int_{(p-2)h}^{(p+1)h}\left|\Phi_x(\alpha,\eta)\right|^2 d\eta\,d\alpha$$
$$+ 27h^3\int_{(p-2)h}^{(p+1)h}\int_{(q-2)h}^{(q+1)h}\left|\Phi_y(s,\beta)\right|^2 ds\,d\beta\Big)$$
$$\le 162h^4\big(\|\Phi_x\|^2_{L^2(Q_{pq})}+\|\Phi_y\|^2_{L^2(Q_{pq})}\big).$$

Therefore,

$$\|\Phi-I_h\Phi\|^2_{L^2(\mathbf{T})} \le \frac{164h^4}{\pi r^2}\sum_{p,q=1}^{\kappa}\big(\|\Phi_x\|^2_{L^2(Q_{pq})}+\|\Phi_y\|^2_{L^2(Q_{pq})}\big)$$
$$= \frac{164h^4}{\pi r^2}\sum_{p,q=1}^{\kappa}\big(9\|\Phi_x\|^2_{L^2(R_{pq})}+9\|\Phi_y\|^2_{L^2(R_{pq})}\big) = \frac{1458h^4}{\pi r^2}\|\Phi\|^2_V.$$

To identify the constant c_0 and finish the proof, substitute $r=h/\gamma$ and, as mentioned in the first paragraph, replace h by $2h$ to account for general h. $\qquad\square$

References

[1] A. Azouani, E. Olson, E.S. Titi. Continuous data assimilation using general interpolant observables. *J. Nonlinear Sci.*, 24(2), 277-304, 2014.

[2] H. Bessaih, E. Olson, E.S. Titi. Continuous assimilation of data with stochastic noise. *Nonlinearity*, 28, 2015, pp. 729–753.

[3] A. Biswas, Z. Bradshaw, M.S. Jolly, Data assimilation for the Navier–Stokes equations using local observables, *SIAM Journal on Applied Dynamical Systems*, Vol. 20, No. 4, 2021, pp. 2174–2203.

[4] A. Biswas, Z. Bradshaw, M.S. Jolly, Convergence of a mobile data assimilation scheme for the 2D Navier-Stokes equations, *arXiv*, 2023, pp. 1–25.

[5] D. Blömker, K. Law, A.M. Stuart, K.C. Zygalakis, Accuracy and stability of the continuous-time 3DVAR filter for the Navier–Stokes equations, *Nonlinearity*, Vol. 26, No. 8, 2013, pp. 2193–2219.

[6] G.L. Browning, W.D. Henshaw, H.O. Kreiss, A numerical investigation of the interaction between the large and small scales of the two-dimensional incompressible Navier–Stokes equations, *UCLA CAM Technical Report 98–23*, April 1998.

[7] E. Carlson, J. Hudson, A. Larios, V. R. Martinez, E. Ng, J. P. Whitehead, Dynamically learning the parameters of a chaotic system using partial observations, *Discrete and Continuous Dynamical Systems*, Vol. 48, No. 8, 2022, pp. 3809–3839.

[8] E. Carlson, L.P. Van Roekel, H.C. Godinez, M.R. Petersen, A. Larios, Exploring a New Computationally Efficient Data Assimilation Algorithm For Ocean Models, *ESS Open Archive*, June 23, 2021.

[9] E. Celik, E. Olson, E.S. Titi. *Spectral Filtering of Interpolant Observables for a Discrete-in-time Downscaling Data Assimilation Algorithm*, SIAM Journal on Applied Dynamical Systems, 2019.

[10] J. Charney, M. Halem, R. Jastrow, Use of incomplete historical data to infer the present state of the atmosphere, *J. Atomos. Sci.*, Vol. 26, 1969, pp.1160–1163.

[11] P. Constantin, C. Foias. *Navier-Stokes equations*. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1988.

[12] S.M. Cox, P.C. Matthews, Exponential Time Differencing for Stiff Systems, *Journal of Computational Physics*, 2001, pp. 430–455.

[13] S. Desamsetti, H.P. Dasari, S. Langodan, E.S. Titi, O. Knio, I. Hoteit, Efficient dynamical downscaling of general circulation models using continuous data assimilation, *Quarterly journal of the Royal Meteorological Society*, Vol. 145, 2019, pp. 3175–3194.

[14] C. Foias, G. Prodi, Sur le comportement global des solutions non stationnaries des équations de Navier–Stokes en dimension two, *Rend. Sem. Math. Univ. Padova*, Vol. 39, 1967, pp. 1–34.

[15] C. Foias, O. Manley, R. Rosa, R. Temam, *Navier–Stokes Equations and Turbulence*, Encyclopedia of Mathematics and Its Applications 83, Cambridge University Press, 2001.

[16] C. Foias, C. Mondaini, E.S. Titi. A discrete data assimilation scheme for the solutions of the 2D Navier-Stokes equations and their statistics. *SIAM Journal on Applied Dynamical Systems*, Vol. 15, No. 4, 2016, pp. 2109–2142.

[17] T. Franz, A. Larios, C. Victor, The bleeps, the sweeps, and the creeps: convergence rates for dynamic observer patterns via data assimilation for the 2D Navier-Stokes equations, *Comput. Methods Appl. Mech. Engrg.*, Vol. 392, No. 114673, 2022, pp. 1–19.

[18] M. Gesho, E. Olson, E. Titi, A Computational Study of a Data Assimilation Algorithm for the Two-dimensional Navier–Stokes Equations, *Communications in Computational Physics*, Vol. 19, No. 4, 2016, pp. 1094–1110.

[19] M.A. El Rahman Hammoud, O. Le Maître, E.S. Titi, I. Hoteit, O. Knio, Continuous and discrete data assimilation with noisy observations for the Rayleigh-Bénard convection: a computational study, *Compt. Geosci.*, Vol. 27, No. 1, 2023, pp.63–79.

[20] K. Hayden, E. Olson, E. S. Titi. Discrete data assimilation in the Lorenz and 2D Navier-Stokes equations. *Phys. D*, 240(18):1416–1425, 2011.

[21] D.A. Jones, E.S. Titi, Upper bounds on the number of determining modes, nodes and volume elements for the Navier–Stokes equations, *Indiana Univ. Math. J.*, Vol. 42, No. 3, 1993, pp. 875–887.

[22] A.K. Kassam, L.N. Trefethen, Fourth-order time stepping for stuff PDEs, *SIAM Journal on Scientific Computing*, Vol. 26, No. 4, 2005, pp. 1214–1233.

[23] A. Larios, C. Victor, Continuous data assimilation with a moving cluster of data points for a reaction diffusion equation: a computational study, *Commun. Comput. Phys.*, Vol. 29, No. 4, 2021, pp. 1273–1298.

[24] B. Laurent, P. Massart, Adaptive Estimation of a Quadratic Functional by Model Selection, *The Annals of Statistics*, Vol. 28, No. 5, 2000, pp. 1302–1338.

[25] K.J.H. Law, D. Sanz–Alonso, A. Shukla, A.M. Stuart, Filter accuracy for the Lorenz 96 model: fixed versus adaptive observation operators, *Physica D: Nonlinear Phenomena*, Vol. 325, 2016, pp. 1–13.

[26] E. Olson, E.S. Titi, Determining Modes for Continuous Data Assimilation in 2D Turbulence, *J. Statist. Phys.*, Vol. 113, No. 5.6, 2003, pp. 799–840.

[27] E. Olson, E.S. Titi, Determining Modes and Grashoff Number for Continuous Data Assimilation in 2D Turbulence, *Theoretical and Computational Fluid Dynamics*, Vol. 22, 2008, pp. 327–339.

[28] J. C. Robinson. *Infinite-dimensional dynamical systems*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2001.

[29] R. Temam, *Navier–Stokes Equations and Nonlinear Functional Analysis*, CBMS Regional Conference Series, No. 41, SIAM, Philadelphia, 1983.

[30] E.S. Titi, On a criterion for locating stable stationary solutions to the Navier–Stokes equations, *Nonlinear Anal.*, TMA, 1987, pp. 1085–1102.

[31] S. Agmon, *Lectures on elliptic boundary value problems*, Prepared by B. Frank Jones, Jr. with the assistance of George W. Batten, Jr., Van Nostrand Mathematical Studies, No. 2, D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London, 1965.