Introduction

Key question: Does data assimilation lead to improvements in the estimate of initial state? Is the error in the observations contained?

We study data assimilation for the 2D viscous, incompressible Navier-Stokes (N-S) equations, with L-periodic boundary conditions as a simpler proxy for the full governing equations:

\[ \frac{\partial U}{\partial t} - \nu \Delta U + (U, \nabla) U + \nabla p = f, \]

where \( U \) is the velocity, \( \nu \) is the kinematic viscosity, \( f \) the time independent body forcing and \( p \) the pressure.

Navier-Stokes equations

We assume that the model is a perfect representation of the atmosphere, so that a solution \( U(t) \) represents the true weather. We can express \( U \) in terms of its Fourier series

\[ U = \sum_{k \in \mathbb{Z}^2} U_k e^{i k \cdot x}, \]

where \( k \) is a 2-D non-zero integer vector.

The N-S equations (1) can be reformulated in Fourier space, while the zero divergence condition and periodicity allows us to remove the pressure term. In this formulation, (1) becomes

\[ \frac{dU_k}{dt} + i \nu k^2 U_k + B(U, U) = f_k, \]

(5)

which is an ODE and \( A \) and \( B \) are operators s.t. \( (AU, v) = \int_{\Omega} -\Delta U \cdot v \, \text{d}x \), \( (B(U, v)) = \int_{\Omega} (U \cdot \nabla) v \, \text{d}x \).

A model for observations

An observation of the weather at time \( t_n \) is given by \( P_t U(t_n) + \sigma R_n \), where \( \sigma R_n \) models the error in the observation.

We suppose that we can observe Fourier modes of the solution and we define a projection \( P_t \) onto the observation space given by

\[ P_t U = \sum_{k \in \mathbb{Z}^2} U_k e^{i k \cdot x}, \]

where \( \lambda \) is a finite positive integer. We say that \( P_t \) is a projection onto the ‘low’ modes and \( \lambda \) is a measure of the size of the observed space or the largest wave number still observed.

A discrete data assimilation algorithm

Straightforward replacement of model values by observations.

The approximating solution of discrete data assimilation is obtained by inserting the observations at discrete times \( t_n \) such that

\[ u_0 = \eta + P_t U_0 + \sigma R_0 \]

and

\[ u_n = Q_t \left( t_n, U_{n-1}, u_{n-1} \right) + P_t U(t_n) + \sigma R_n \]

(2)

where \( \psi \) is the semi-flow of (1), \( \eta \) is the initial guess and \( Q_t \) is the projection onto the unobserved space. Then, the approximating solution \( u(t) \) is a piece-wise continuous in time function defined by

\[ u_n(t) = \psi \left( t; t_n, u_{n-1} \right) \quad \text{for} \quad t \in [t_n, t_{n+1}). \]

(3)

Data assimilation error

The data assimilation error \( \delta(t) \) is the difference between the true and approximating solution. It is a piece-wise continuous in time function defined by

\[ \delta(t) = U(t) - u(t) \]

in the interval \( [t_n, t_{n+1}) \).

Main Theorem

We are able to prove a rigorous estimate derived using analytical properties of the underlying dynamics.

Theorem Let \( U \) be as defined by (5) and \( \delta \) be the data assimilation error as defined by (4) and suppose that \( \mathbb{E}(\| R_n \|^2) < \infty \). Then, for any data assimilation interval \( h = t_{n+1} - t_n > 0 \) a finite \( \lambda \) such that for all \( \lambda > \lambda^* \),

\[ \limsup_{n \to \infty} \left( \| u(t_n) - \delta \| \right) \leq 0 \]

a.s., where \( B_n \) is a stationary a.s. finite process and \( \sigma^2 B_n \to 0 \) as \( \sigma \to 0 \) a.s.

Main assumptions

Noise: We assume that the observation error is random, unbounded and \( R_n \) is a stationary, tempered process with zero mean and \( \mathbb{E}(R_n^2) = 1 \). Therefore \( \mathbb{E}\left( \| R_n \|^2 \right) = \sigma^2 < \infty \) and \( \sigma^2 \) is the variance of observation noise.

We note that if \( R_n \) does not have zero mean, this would represent a systematic error, which would likely be corrected for and therefore we can make the simplifying assumption.

Model: There is no model error. That is, the dynamical system (1) is a perfect representation of the atmosphere and we use it for the forecasting.

Observations: We can observe the ‘low’ Fourier modes of the true solution.

Proof part I: The error equation

Using the bi-linearity of \( B \) and the fact that the data assimilated solution satisfies the equation in every interval \( [t_n, t_{n+1}) \) we obtain the equation:

\[ \frac{d\delta}{dt} + \nu k^2 \delta + B(U, \delta) + B(\delta, U) - B(\delta, \delta) = 0. \]

Using the above, as well as estimates as in [1] and iterating over the intervals \( [t_n, t_{n+1}) \), we obtain:

Lemma 1 \( \| \delta(t_{n+1}) \| \leq M_0(h) \| \delta(t_n) \| + \sigma^2 \| R_n \|^2 \),

(7)

where \( \| \cdot \| \) is the \( \ell^2 \) norm and \( M_0(h) \) are functions that depend on \( \delta(t_0) \).

Proof part II: Controlling the error

To obtain a meaningful bound on the error \( \| \delta(t_{n+1}) \| \), we would need that RHS of (7) is almost surely finite in the long term. This will be the case if we can make \( M_0 < 1 \) for all \( k \).

Since \( \delta(t_0) \) is stochastic, the \( M_0 \) are also stochastic and therefore it is not, in general, possible to guarantee that \( M_0 < 1 \) for all \( k \) for any value of \( h \). However, we are able to use the Ergodic Theorem to show that if \( \mathbb{E}(M_0) < 1 \), it ensures that \( M_0 \) is ‘often enough’.

That is, for almost all realizations of the sequence \( \{ M_0 \} \), the proportion of \( M_0 < 1 \) is sufficient to ensure that the error remains almost surely finite.

References