On posterior consistency of data assimilation with Gaussian process priors: the $2D$ Navier-Stokes equations

Richard Nickl * and Edriss S. Titi †

July 16, 2023

Abstract

We consider a non-linear Bayesian data assimilation model for the periodic two-dimensional Navier-Stokes equations with initial condition modelled by a Gaussian process prior. We show that if the system is updated with sufficiently many discrete noisy measurements of the velocity field, then the posterior distribution eventually concentrates near the ground truth solution of the time evolution equation, and in particular that the initial condition is recovered consistently by the posterior mean vector field. We further show that the convergence rate can in general not be faster than inverse logarithmic in sample size, but describe specific conditions on the initial conditions when faster rates are possible. In the proofs we provide an explicit quantitative estimate for backward uniqueness of solutions of the two-dimensional Navier-Stokes equations.

1 Introduction

High- and infinite-dimensional Bayesian methods have been increasingly popular in statistical inference problems arising with partial differential equations (PDEs) and related uncertainty quantification tasks, we selectively mention the contributions [41, 11, 21, 20, 12, 5]. Recent years have seen substantial progress in our theoretical understanding of the performance of such algorithms in non-linear settings, see [34] for an overview and many references. The results so far have covered a variety of prototypical examples ranging from basic steady state elliptic equations [32, 18, 1, 36, 37] to X-ray-type problems [30, 31, 7, 6, 40] and diffusion models [35, 19, 24, 8, 33, 22] where

*Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge CB3 0WB, UK. email: nickl@maths.cam.ac.uk

†Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK; and Department of Mathematics, Texas A&M University, College Station, TX 77840, USA. email: Edriss.Titi@maths.cam.ac.uk and titi@math.tamu.edu

posterior distributions are obtained from updating a Gaussian process prior given noisy measurements of the solution of a PDE or SDE.

A particularly important and active application area of the Bayesian inference paradigm in PDE settings is the field of *data assimilation* [27, 39, 14]. For example in geophysical sciences, non-linear dynamical systems are used to model the atmosphere [23], oceans [3], turbulence [29] and fluid flow [10]. A Bayesian model for the initial conditions is then updated whenever new measurements are taken, and posteriors can be approximately computed by Monte Carlo and filtering methods. We refer to the recent monograph [14], especially its Part II, for an overview of a variety of concrete scientific application areas in the context of data assimilation. Beyond the setting of linear systems (e.g., [25, 26, 38]), the *statistical validity* of such posterior based inferences remains largely an open question. In the present article we study non-linear data assimilation problems arising with the periodic two-dimensional *Navier-Stokes equations* as a paradigm for the underlying dynamics of the system. This non-linear PDE provides the physical description of viscous flow in fluid mechanics and forms the mathematical foundation for the abovementioned applications in geophysical sciences. It also constitutes one of the key PDE examples for the Bayesian approach to data assimilation and inverse problems, see [10] and [41].

In the literature often reduced or approximate models (e.g., the Lorenz model) are used, see, e.g., [23, 29, 39, 27, 14]. To develop a general understanding we avoid such reductions as much as possible. The only substantial simplification we make is that we restrict to a two-dimensional state space – this is simply because our theoretical development relies on PDE theory for the well-posedness of global solutions to Navier-Stokes equations which is not (yet!) known to be valid in dimensions higher than 2. We also only consider *periodic* boundary conditions to streamline the exposition and to make the proofs accessible to a wider audience, but this restriction is not essential.

To introduce the setting, recall that the (incompressible) Navier-Stokes equations on some domain Ω postulate the evolution in time $t \in [0,T]$ of a divergence free velocity vector field $u : [0,T] \times \Omega \to \mathbb{R}^2$ solving the non-linear PDE

$$\begin{aligned} \frac{du}{dt} + \nu A u + B(u,u) &= f \\ u(0) &= u_0 \end{aligned} \tag{1}$$

where $u(0)$ is the initial condition of the system. Here A is the (linear) Stokes operator, $\nu > 0$ a viscosity parameter, f a forcing term, and B is the bi-linear form modelling the non-linearity. Roughly speaking (in the periodic case, or up to a canonical Leray-Helmholtz projection step recalled below) we can think of A as the negative Laplacian and of $B(u,u) = u \cdot \nabla u$. We refer to [9] for a classical reference on this material.

The task of data assimilation begins with the specification of an initial condition u_0. In absence of specific background knowledge in a given experimental setting, and in part to aid the computational tasks that follow, such initial conditions are often modelled by a Gaussian random field $u_0 = (\theta(x) : x \in \Omega)$ over the domain Ω – see [41, 39, 27, 14] and also specifically [10] in the setting of the Navier-Stokes model. The law $\mathcal{L}(\theta) \equiv \Pi$ of this field in an appropriate function space plays the role of the prior in (infinite-dimensional) Bayesian statistics, and its covariance structure parallels the choices of penalty norms in 'variational (optimisation based) data assimilation'. Under appropriate assumptions the solutions to (1) are unique and the random initial condition $\theta = u(0)$ determines a complete stochastic *forward model*, that is, a probability distribution on the states

$$\big(u_\theta(t,x) : t > 0, x \in \Omega\big), \text{ that solve (1) with initial condition } u(0) = \theta \sim \Pi,$$

of the velocity fields at all points in space Ω and time $[0,T]$. Even though the initial condition follows a Gaussian distribution, due to the non-linearity of the Navier-Stokes system, the implied stochastic model for $u_\theta(t,\cdot)$ at times $t > 0$ is *not* Gaussian any longer. Nevertheless, this 'forward model' can be updated via Bayes' rule after measurements at discrete points (t_i, X_{ij}) in $(0,T] \times \Omega$ are taken. We follow here the 'Eulerian measurement' scheme relevant in fluid mechanics (see Sec.3 in [10]) where noisy 'regression' type measurements of the velocity field of the form

$$Y_{ij} = u_\theta(t_i, X_{ij}) + \varepsilon_{i,j}, \quad \varepsilon_{ij} \sim N(0, I_{\mathbb{R}^2}),\ i = 1,\dots,m,\ j = 1,\dots,n, \tag{2}$$

are collected. If Π denotes the prior probability measure induced by the Gaussian process model, the posterior distribution of the initial state θ is then given by

$$d\Pi(\theta|(Y_{ij}, X_{ij}, t_i)_{1\le i\le m,\ 1\le j\le n}) \propto \exp\Big\{ -\frac{1}{2}\sum_{i,j} |Y_{ij} - u_\theta(t_i, X_{ij})|^2_{\mathbb{R}^2}\Big\} d\Pi(\theta),$$

where u_θ is the solution of (1) corresponding to initial condition $u_0 = \theta$. Approximate computation of such non-Gaussian posterior distributions is possible using Monte Carlo methods, and we can then use PDE forward solvers or filtering methods to retrieve posterior inferences for u_θ at times $t > 0$ and points $x \in \Omega$ – see [27, 39, 11, 5, 14]. Of key importance is that – in contrast to standard non-parametric regression techniques – such posterior estimates u_θ are themselves solutions of a Navier-Stokes system so that the algorithmic outputs retain physical interpretation (e.g., incompressibility of the flow, $\nabla \cdot u_\theta = 0$).

Such Bayesian methodology naturally addresses all main tasks of data assimilation which are to provide estimates for

$$\begin{aligned} u(t,\cdot),\ t > T, &\quad \textbf{prediction} \\ u(t,\cdot),\ t = t_i, &\quad \textbf{filtering} \\ u(t,\cdot),\ 0 < t \le T, &\quad \textbf{smoothing} \\ u(0,\cdot), &\quad \textbf{inversion}; \end{aligned}$$

cf. p.173 in [39] and also [27] for this terminology. The question we address here is whether the *posterior distribution* of all the states

$$\big(u_\theta(t,x) : t \ge 0, x \in \Omega\big), \quad \theta \sim \Pi(\cdot|(Y_{ij}, X_{ij}, t_i)_{1\le i\le m,\ 1\le j\le n}),$$

is *statistically consistent*, that is, whether it places almost all of its mass (measured in a suitable norm in function space, and with high probability under the law of

$(Y_{ij}, X_{ij}, t_i)_{1\le i\le m,\ 1\le j\le n})$, near the ground truth state $(u_{\theta_0}(t,x) : t \ge 0, x \in \Omega)$ of the system generated by the actual (unobserved) initial condition $u_0 = \theta_0$. Such results validate Bayesian data assimilation algorithms in a scientifically desirable 'objective', that is, *prior-independent* or 'frequentist' way, cf. [16]. To the best of our knowledge, no results of this type are known in the literature at the moment. We will show in Theorem 3 that posterior consistency indeed occurs for a flexible class of infinite-dimensional Gaussian process prior models for θ, and as sample size $N = mn \to \infty$. Our proof strategy resembles the Bayesian forward model and is based on solving the hardest of the above four problems (inversion) first, and then uses forward Lipschitz continuity in time of strong solutions of the (two-dimensional) Navier-Stokes equations. We further show that the 'logarithmic' convergence rates we obtain for the inversion problem cannot be improved in general, and discuss a set of (strong) restrictions on the initial condition in terms of the spectrum of the Stokes operator where faster rates are possible.

Our proofs are based on general statistical theory for Bayesian non-linear inverse problems developed in [34], on classical results from the PDE analysis of $2D$ Navier-Stokes equations (e.g., [9]), and on an explicit quantitative stability (inverse continuity) estimate for the forward map $\theta \mapsto u_\theta$ given in Theorem 1 below. The latter is inspired by old work on backward in time uniqueness of solutions to the $2D$ Navier-Stokes equations in [2], who gave explicit estimates on the difference between the initial data in terms of the difference of the final state of the strong solutions of the system. This implies the desired backward uniqueness, but further allows one to obtain a quantitative stability estimate (unlike alternative non-constructive proofs using the analyticity in time of strong solutions, cf. [9]).

This paper is organised as follows: the main analytical results on the $2D$ Navier-Stokes equations will be given in Subsection 2.1, while the statistical theory for data assimilation with Gaussian process priors is developed in Subsection 2.2. Proofs can be found in Section 3.

2 Main results

2.1 Forward and inverse stability in the $2D$ Navier-Stokes equations

Throughout we denote by $\Omega = [0, 2\pi]^2$ the two-dimensional flat torus, i.e., opposite endpoints are identified and all functions are periodic: $u(\cdot + 2\pi e_i) = u(\cdot + e_i)$ for $i = 1, 2$ where $e_1 = (1, 0), e_2 = (0, 1)$ are the canonical orthogonal basis vectors of the plane. We define $C^\infty(\Omega)$ as the space of infinitely differentiable periodic functions with fundamental periodic domain Ω. We also require the usual $L^2(\Omega)$ spaces of square integrable functions for Lebesgue measure dx, as well as the Sobolev spaces $H^m(\Omega), m \in \mathbb{N}$, of functions $f \in L^2(\Omega)$ whose partial derivatives up to order m lie in $L^2(\Omega)$. When considering two-dimensional vector fields $v = (v_1, v_2) : \Omega \to \mathbb{R}^2$ with components v_1, v_2 lying in some function space $\mathscr{X}$, we will write $v \in \mathscr{X}^2$ – or sometimes even only $v \in \mathscr{X}$

when no confusion may arise. The divergence operation

$$\nabla \cdot v = \frac{\partial}{\partial x_1} v_1 + \frac{\partial}{\partial x_2} v_2$$

for smooth vector fields extends to a linear operation $\nabla\cdot$ in the sense of (periodic Schwartz) distributions. We can then define spaces of vector fields

$$H = \Big\{ u \in L^2(\Omega)^2 : \nabla \cdot u = 0, \int_\Omega u = 0 \Big\}, \tag{3}$$

as well as

$$V = \Big\{ u \in (H^1(\Omega))^2 : \nabla \cdot u = 0, \int_\Omega u = 0 \Big\}, \tag{4}$$

and equip these spaces with inner products $\langle \cdot, \cdot \rangle_H \equiv \langle \cdot, \cdot \rangle_{L^2}$ and

$$\langle u, v \rangle_V \equiv \langle \nabla u, \nabla v \rangle_{L^2} = \sum_{i,j=1}^{2} \int_\Omega \frac{\partial u_i(x)}{\partial x_j} \frac{\partial v_i(x)}{\partial x_j} dx,$$

where ∇ is the gradient operator. The resulting norms are denoted by $\| \cdot \|_H, \| \cdot \|_V$, respectively. One can show that H and V arise as the completion of

$$\mathcal{V} = \Big\{ u = (u_1, u_2), u_i \text{ a trigonometric polynomial on } \Omega : \nabla \cdot u = 0, \int_\Omega u = 0 \Big\}, \tag{5}$$

for the norms of $L^2(\Omega)$ and $H^1(\Omega)$, respectively.

We now introduce some standard notation for general Navier-Stokes equations – while this is not strictly necessary in the periodic case, we include it here as our proofs do extend in this notation to general two-dimensional domains Ω with smooth boundary by appealing to the theory developed in [9]. The 'Helmholtz-Leray' L^2-projector is given by $P : L^2(\Omega)^2 \to H$ and on

$$\mathcal{D}(A) \equiv H^2(\Omega)^2 \cap V$$

we then define the *Stokes operator*

$$A = -P\Delta, \quad \Delta = \nabla \cdot \nabla \text{ the Laplacian}, \tag{6}$$

which on $\mathcal{D}(A)$ has 'graph norm' $\|u\|_{\mathcal{D}(A)} \equiv \|Au\|_{L^2} \simeq \|u\|_{H^2}$ (as in Prop 4.7 in [9]) and is self-adjoint on its domain. Notably, we observe that in the case of periodic boundary conditions one in fact has $A = -\Delta$. We also define the bilinear form

$$B(u, v) = P[v \cdot \nabla u], \quad B : V \times V \to V', \tag{7}$$

where V' is the topological dual space of V, with the usual dual pairing $\langle \cdot, \cdot, \rangle_{V,V'}$. In this notation we have for all $u \in V$, from integration by parts and since P is a L^2-projector,

$$\|u\|_V^2 = \langle \nabla u, \nabla u \rangle_{L^2} = \langle -\Delta u, u \rangle_{V,V'} = \langle Au, u \rangle_{V,V'}, \tag{8}$$

and also (as in (6.17), (6.18) in [9]), for $u, v, w \in V$,

$$\langle B(u,v), w\rangle_{V,V'} = -\langle B(u,w), v\rangle_{V,V'}, \text{ and thus } \langle B(u,v), v\rangle_{V,V'} = 0. \tag{9}$$

Now let $\nu > 0$ be a fixed viscosity term and $f \in H^1(\Omega)$ a forcing such that $\int_\Omega f = 0$. To expedite proofs we take f to be time-independent but this is not necessary. We consider spatially periodic solutions $u = u_\theta = (u_\theta(t,x) : t \in (0,T], x \in \Omega)$ of the incompressible Navier-Stokes equations which represent the system of non-linear partial differential equations given by

$$\frac{\partial}{\partial t} u - \nu\Delta u + u \cdot \nabla u = f - \nabla p \quad \text{ on } \Omega \times (0,T], \tag{10}$$

$$\nabla \cdot u = 0 \quad \text{ on } \Omega \times (0,T],$$

$$\int_\Omega u(t,\cdot) = 0 \quad \text{for all } t \in (0,T],$$

$$u(0,\cdot) = u_0 \equiv \theta \quad \text{ on } \Omega, \tag{11}$$

where $u(0) = \theta \in V$ is an initial condition and ∇p is a pressure term.

Using the preceding notation and applying the Helmhotz-Leray projector P to the last set of equations, one can alternatively study solutions $u \in V$ of the non-linear evolution equation in Hilbert space H given by

$$\frac{du}{dt} + \nu A u + B(u,u) = f \tag{12}$$

$$u(0) = u_0$$

where $u_0 = \theta \in V$. [See Ch.5 in [9] for details. Note that this eliminates the pressure term ∇p – but as is well known, the original equations can be recovered from the solution of the 'projected' equations by solving a standard elliptic problem.]

The regularity of solutions will be expressed in terms of function spaces

$$L^p((0,T],\mathscr{X}) \equiv \left\{u : (0,T] \times \Omega \to \mathbb{R}^2 : \int_0^T \|u(t,\cdot)\|_{\mathscr{X}}^p dt < \infty\right\}, \quad 1 \le p < \infty,$$

with corresponding Bochner-integral norm for $\mathscr{X}$-valued maps, where $\mathscr{X}$ is a normed linear space of vector fields over Ω to be specified. Similarly we define the spaces $L^\infty((0,T],\mathscr{X})$ and $C([0,T],\mathscr{X})$ of time-bounded or -continuous $\mathscr{X}$ valued maps. The following result can be deduced without difficulty from existing theory for the two-dimensional periodic Navier-Stokes equations.

Proposition 1. *Let $T > 0$ and let $u_0 \in V$ satisfy $\|u_0\|_V \le U$ for some $U > 0$.*

A) The two-dimensional periodic Navier-Stokes equations (12) have a unique strong solution $u \in C([0,T],V) \cap L^2((0,T),\mathcal{D}(A))$ with $du/dt \in L^2((0,T],H)$. There exists a constant $c_U \equiv c(U, \|f\|_{L^2}, \nu, T) < \infty$ such that

$$\sup_{0\le t\le T} \|u(t)\|_V + \int_0^T \|u(t)\|_{\mathcal{D}(A)}^2 dt \le c_U. \tag{13}$$

Moreover, for every $m > 0$ there exists $c = c(m, \|f\|_{H^1}, \nu, T) > 0$ such that we have

$$\sup_{u_0 \in \mathcal{D}(A):\ \|Au_0\|_{L^2} \le m} \ \sup_{0 \le t \le T, x \in \Omega} |u(t,x)| \le c < \infty. \tag{14}$$

B) If $v_0 \in V$ is another initial condition, then we have

$$\sup_{0 \le t \le T} \|u(t) - v(t)\|_{L^2(\Omega)} \le K \|u_0 - v_0\|_{L^2(\Omega)} \tag{15}$$

for some constant $K = K(U, \|f\|_{L^2}, \nu, T) < \infty$.

Proof. The main steps of the proof are given for convenience of the reader in Section 3.5 below. □

Let us now turn to the inverse problem of solving for u_0 from $u(t)$. It is known that under natural regularity hypotheses, the Navier-Stokes solution map $\theta = u(0) \mapsto u(t) = u_\theta(t)$ is analytic in time and as a consequence injective for any $t > 0$ (and $d = 2$), see Theorem 12.2 in [9]. Therefore the solutions $u(t)$ determine their initial conditions $u(0)$, and a Bayesian approach to solve this inverse problem with discrete noisy data will be studied in the next section. A main contribution of this article is the following explicit stability estimate for this forward map $\theta \mapsto u_\theta(t)$, to be used in the proofs below. It is inspired by backward uniqueness results [2] for general non-linear parabolic equations in Hilbert space.

Theorem 1. *Let $T > 0$, and for initial conditions $u(0), v(0) \in V$ such that $\|u(0)\|_V + \|v(0)\|_V \le U < \infty$, consider the corresponding strong solutions $u, v \in C([0,T], V)$, to the 2-dimensional periodic Navier-Stokes equations (12).*

A) There exist constants c_0, c_1 depending only on $U, \nu, T, \|f\|_{L^2}$ such that

$$\|u(0) - v(0)\|_{L^2(\Omega)} \le c_0 \Big(\log \frac{c_1}{\|u(t) - v(t)\|_{L^2}} \Big)^{-1/2}, \quad \text{for every } t \in [0,T], \tag{16}$$

where $\|u(t) - v(t)\|_{L^2} < c_1$.

B) Suppose further that there exists a fixed ('inverse Poincaré') constant $0 < c_P < \infty$ such that

$$\frac{\|u(0) - v(0)\|_V}{\|u(0) - v(0)\|_{L^2}} \le c_P. \tag{17}$$

Then there exists a constant $c_2 = c_2(U, \nu, T, \|f\|_{L^2})$ such that

$$\|u_0 - v_0\|_{L^2(\Omega)} \le e^{c_2 c_P} \|u(t) - v(t)\|_{L^2(\Omega)}, \qquad \text{for every } t \in [0,T]. \tag{18}$$

Versions of these inequalities where $\|u(t) - v(t)\|_{L^2(\Omega)}$ is replaced by its quadratic time average over $(0,T]$ hold as well – see Corollary 1 below. The proof of Theorem 1 extends to the case of bounded smooth domains $\Omega \subset \mathbb{R}^2$, if the spaces V, H are appropriately defined (with Dirichlet boundary conditions, cf. [9]), after only minor technical modifications.

The growth of the constants in the preceding stability estimates is exponential in the time horizon T, as is not unexpected due do the limited time predictability of chaotic dynamical systems of even much simpler nature than the Navier-Stokes equations, cf. the classical contribution by Lorenz [28] and also more generally the discussion in [23] about limitations of forecasting geophysical systems. The previous stability estimates hence should be interpreted as informative in 'moderate' time horizons T.

Remark 1 (Inverse Poincaré inequality and the Stokes spectrum)**.** The ratio in (17) is lower bounded by a fixed positive constant in view of the Poincaré inequality (p.292 in [13]), but we are asking here for a similar *uniform upper bound* (for the class of initial conditions considered). While one cannot in general expect that such an 'inverse Poincaré constant' exists, examples can be given where it holds. For instance suppose both $u_0, v_0 \in V$ have a *finite* expansion in the H-orthonormal eigen-basis $(e_j : j \geq 0)$ of the Stokes operator A (see (22) below), up to 'frequency' J. Then

$$\frac{\|u_0 - v_0\|_V^2}{\|u_0 - v_0\|_{L^2}^2} = \frac{\sum_{j\leq J} \lambda_j \langle e_j, u_0 - v_0\rangle_{L^2}^2}{\sum_{j\leq J} \langle e_j, u_0 - v_0\rangle_{L^2}^2} \leq \lambda_J \equiv c_P.$$

Observe that one has the asymptotic distribution $\lambda_j \simeq j$ as $j \to \infty$, of the eigenvalues of the Stokes operator A in $d = 2$, see Proposition 4.14 in [9]. This will permit the creation of 'Stokes band-limited' models of initial conditions for which we can obtain 'fast' (better than logarithmic) convergence rates in Theorem 5 below. Other sets of initial conditions could be conceived for which (17) holds with a uniform constant c_P, but this is beyond the scope of the present paper.

Without an a priori upper bound on the terms required to represent u_0, v_0 in the eigen-basis of the Stokes operator, the stability estimate (16) is sharp in the sense that the inverse modulus of continuity is attained for particular sets of motions of the Navier-Stokes equation – at least up to the power of the logarithm. Inspection of the proof of (20) shows that the power of log in the lower bound could be made to approach $1/2$ if we consider initial conditions bounded in H^1 only, but we give a result under the stronger H^2-hypothesis relevant in the statistical theorems that follow.

Theorem 2. *There exists a sequence of initial conditions* $u_j(0) \in C^\infty(\Omega)^2 \cap V$, $j \in \mathbb{N}$, *with corresponding strong solutions* $u_j(t)$ *to the periodic Navier-Stokes equations (12) on* Ω *with* $\nu = 1/2, f = 0$, *such that*

$$\|u_j(0)\|_{H^2} \lesssim 1, \quad \|u_j(0)\|_{L^2} \simeq j^{-2}, \quad \|u_j(t)\|_{L^2} \simeq e^{-j^2 t} j^{-2}, \tag{19}$$

for all $t > 0$. *In particular, setting* $v(0) = 0$ *and hence* $v(t) \equiv 0$, *we have for some* $c' = c'(c, t) > 0$ *that*

$$\|u_j(0) - v(0)\|_{L^2(\Omega)} \geq c' \frac{1}{\log\left(\frac{1}{\|u_j(t) - v(t)\|_{L^2}}\right)}, \quad \text{all } j \in \mathbb{N}, t > 0. \tag{20}$$

The exponential instability arises from a (linear, scalar) heat equation that can be 'planted' within the set of solutions of Navier-Stokes equations for a specific set of initial conditions for which the non-linearity vanishes at all times. At least when $f = 0$ we conjecture that similar phenomena persist even when the non-linear term does not vanish, by employing the (non-linear) spectral manifolds constructed in [15] instead of the eigenfunctions of the Laplacian from the previous proof.

2.2 Non-linear data assimilation with Gaussian process priors

We now consider data assimilation tasks arising with discrete observations from the Navier-Stokes equations and give statistical guarantees for Bayesian methodology proposed and developed for such problems in [10] – see also [41, 39, 27, 14]. For $u = u_\theta$ a solution of the PDE (12) with unknown initial condition $u(0) = \theta$, the statistical observations are assumed to consist of the random vectors $Z^{(N)} = (Y_i, t_i, X_i)_{i=1}^N$

$$Y_i = u_\theta(t_i, X_i) + \varepsilon_i, \quad \varepsilon_i \sim^{i.i.d.} N(0, I_{\mathbb{R}^2}), \quad i = 1, \dots, N, \tag{21}$$

with $(t_i, X_i)_{i=1}^N$ drawn iid from the uniform distribution λ on $(0, T] \times \Omega$, independently of the Gaussian noise vectors ε_i. The law on $(\mathbb{R}^2 \times (0, T] \times \Omega)^N$ of the data vector $Z^{(N)} = (Y_i, t_i, X_i)_{i=1}^N$ when u_θ arises from the initial condition θ will be denoted by P_θ^N. One could (as in (2)) consider distinct sample sizes for time and space measurements, in particular a single time measurement t would be sufficient for the results that follow, but we abstain from this to keep the exposition straightforward. Let us emphasise that the 'white' noise ε_i in (21) is purely of measurement error type (arising from the discretisation of u_θ) and that we do not model explicit stochasticity in the Navier-Stokes dynamics itself, in particular we do not consider a SPDE model (whose transition densities and hence likelihood functions would be numerically inconvenient for the implementation of the Bayesian approach).

We assume the viscosity $\nu > 0$ and forcing f to be given, possibly determined beforehand by independent experiments (theory for these distinct inverse problems could be developed as well, for ν for instance following ideas in [18, 33]). In contrast, and following common practice in data assimilation, we assume that the initial condition $\theta = u(0)$ of the system is *unknown*. Inferences on the state $u_\theta(t, \cdot)$ of the system based on observations $Z^{(N)}$ need to incorporate this uncertainty. One systematic way to do this is to adopt a Bayesian approach and to model the initial condition $u(0) = \theta$ by a Gaussian random field over Ω – see [10] in the setting of Navier-Stokes equations and fluid mechanics specifically; and also [41, 39, 27, 14] more generally. For the theory that follows we will employ the following assumption – for standard notions of and background on Gaussian processes we refer the reader to [17] and [34].

Condition 1. *Consider a Borel probability measure Π' on $V \cap H^2(\Omega)^2$ arising as the law of the centred Gaussian random vector field $(\theta'(x) = (\theta'_1(x), \theta'_2(x)) : x \in \Omega)$ with reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ continuously imbedded into $V \cap H^\alpha(\Omega)^2$ for some $\alpha \geq 2$. Then take as prior $\Pi = \Pi_N$ for θ the law of the rescaled random vector field $\theta = \theta'/N^{1/(2\alpha+2)}$.*

Examples of 'base priors' Π' with RKHS $\mathcal{H} = (V \cap H^\alpha, \|\cdot\|_{H^\alpha})$ for any $\alpha \geq 2$ can be easily constructed: for instance one starts with two independent periodic α-regular

Gaussian random fields over Ω (e.g., expanded in a basis of periodic wavelets with independent Gaussian coefficients, as in [35]) and then applies the (linear) Helmholtz-Leray-projector P to the Gaussian vector to enforce the H-constraint. Alternatively we could define the prior for θ' immediately as a Gaussian series expansion (e.g., (B.1) in [34]) for the H-orthonormal eigenfunctions

$$e_k \propto (k_2, -k_1)e^{2\pi i k.(\cdot)}, \quad k \in \mathbb{Z}^2 \setminus \{(0,0)\}, \tag{22}$$

of the Stokes operator. The N-dependent rescaling of θ' follows ideas in [30] and provides an (in the proofs essential) increase of the amount of regularisation provided by the prior. Finally, if information on θ is available from past observations ('training samples'), we could center the prior at such a 'trained' mean vector, but for the theory we only consider generic mean *zero* Gaussian process priors.

As the random vector fields θ', θ lie almost surely in the space $V \cap H^2 = \mathcal{D}(A)$, by Proposition 1 such a prior postulates a complete stochastic 'forward' model of uniformly bounded solutions $u_\theta(t, \cdot)$ of the Navier-Stokes equations drawn from the Gaussian initial condition θ. Given observations $Z^{(N)}$ we can update this model via Bayes' rule to produce the best 'posterior' fore- and hind-casts for the solution $u_\theta(x,t)$ at (potentially unobserved) times t and points $x \in \Omega$. Algorithmically we first compute the posterior distribution for the initial state θ, which in the model (21) is of the form

$$d\Pi(\theta|Z^{(N)}) \propto e^{\ell_N(\theta)} d\Pi(\theta); \quad \ell_N(\theta) = -\frac{1}{2}\sum_{i=1}^{N} |Y_i - u_\theta(X_i, t_i)|^2, \ \theta \in V, \tag{23}$$

where $|\cdot| = |\cdot|_{\mathbb{R}^2}$ is the Euclidean norm. Even though the prior is Gaussian, the non-linearity of the map $\theta \to u_\theta$ renders $\Pi(\cdot|Z^{(N)})$ a non-Gaussian random probability measure in the function space V. Nevertheless, posterior draws $\theta \sim \Pi(\cdot|Z^{(N)})$ and then also an estimate for the posterior mean $E^\Pi[\theta|Z^{(N)}]$ can be calculated from Markov chain Monte Carlo (MCMC) techniques, for instance by the pCN, ULA or MALA algorithm (see [11] and specifically in the context of data assimilation also Ch.3 in [27]; as well as [21, 37, 34] for results towards computational guarantees). This approach requires numerical solutions of the forward PDE at each iterate ϑ_k of the Markov chain – but no inversion step, or backward solution of the PDE, is required. We can then compute an estimate $u_{\bar\theta}$ for $u_\theta(x,t)$ at any given point (x,t) by computing the solution of (12) with the initial condition $\bar\theta = E^\Pi[\theta|Z^{(N)}]$ (itself approximated by ergodic MCMC averages $\sum_{k=1}^K \vartheta_k/K$ on a suitable discretisation space for θ).

The absence of an explicit inversion step in this algorithm is attractive for applications but also triggers the question whether guarantees can be given that it will recover the true physical state of the system. Our goal here is to prove that this method indeed will be statistically *consistent*, that is, that it will recover the 'correct' solution of the Navier-Stokes equations at any point in time and space, arising from the 'ground truth' initial condition θ_0 that has actually generated the data, at least if we take sufficiently many measurements $N \to \infty$. Mathematically this means that we study the statistical behaviour of the posterior distribution under the law $P^N_{\theta_0}$, following the usual paradigm of frequentist analysis of Bayes procedures, see [16] or also Ch.7.3 in [17]. In our proofs we combine Proposition 1 and Theorem 1 with recent techniques from the theory of Bayesian non-linear inversion with Gaussian process priors [30, 34] to show asymptotic concentration properties of this posterior around the true states $(u_{\theta_0}(t,x), t \geq 0, x \in \Omega)$ of the non-linear system.

By convention we regard the first N measurements as the training sample and the additional pair (X_{N+1}, t) as the prediction sample, where X_{N+1} is drawn at random from λ_Ω and $t \in (0, T_p], T_p \geq T$, is a (deterministic) time we wish to fore- or hind-cast. If we denote by θ_0 the ground truth initial condition that generated the data (21) and by $\theta \sim \Pi(\cdot|Z^{(N)})$ a draw from the posterior distribution, then this leads us to consider the posterior quadratic 'prediction risk'

$$E_{X_{N+1}}\Big(\big|u_\theta(t, X_{N+1}) - u_{\theta_0}(t, X_{N+1})\big|^2\Big) = \|u_\theta(t,\cdot) - u_{\theta_0}(t,\cdot)\|^2_{L^2(\Omega)^2},\ t > 0, \tag{24}$$

which measures how well we predict on average the state of the system at time t and at a 'generic' position $X_{N+1} \sim \lambda_\Omega$.

Theorem 3. *Consider a Gaussian process prior as in Condition 1 with $\alpha \geq 2$, RKHS $\mathcal{H}$, and resulting posterior distribution (23) arising from observations (21) in the 2-dimensional periodic Navier-Stokes equations (12). Suppose the ground truth initial condition θ_0 lies in $\mathcal{H}$. Then for every $T_P \geq T$ there exists a sequence $\eta_N = O(1/\sqrt{\log N})$ as $N \to \infty$ (with constants uniform in $\|\theta_0\|_\mathcal{H} \leq U$) such that*

$$\Pi\Big(\theta \in V : \sup_{0<t\leq T_p} \|u_\theta(t,\cdot) - u_{\theta_0}(t,\cdot)\|_{L^2(\Omega)^2} < \eta_N |Z^{(N)}\Big) \to^{P^N_{\theta_0}} 1 \tag{25}$$

as well as

$$\Pi\big(\theta \in V : \|\theta - \theta_0\|_{L^2(\Omega)^2} < \eta_N|Z^{(N)}\big) \to^{P^N_{\theta_0}} 1. \tag{26}$$

Moreover, if

$$\bar\theta_N = E^\Pi[\theta|Z^{(N)}] \in V$$

is the posterior ('Bochner-') mean and $u_{\bar\theta_N}$ the solution of the Navier-Stokes equation (12) with initial condition $\bar\theta_N$, then

$$\|\bar\theta_N - \theta_0\|_{L^2(\Omega)^2} + \sup_{0<t\leq T_p} \|u_{\bar\theta_N}(t,\cdot) - u_{\theta_0}(t,\cdot)\|_{L^2(\Omega)^2} = O_{P^N_{\theta_0}}(\eta_N). \tag{27}$$

The logarithmic rates obtained may not be sharp at 'observed' times $t > 0$. For instance in 'average prediction' loss where one takes the quadratic time average of (24) over $[0, T]$, our proofs imply much faster rates, see (41). But inference for fixed possibly unobserved times t constitutes a nonlinear inverse problem which the Bayesian posterior distribution arising from a prior on the initial condition θ solves 'implicitly' by first inferring θ and then updating the resulting forward model prediction for $u_\theta(t)$.

We now show that for the underlying key step of recovery of the initial condition, the posterior convergence rates obtained are essentially optimal in an information theoretic 'minimax' sense in the family of two-dimensional periodic Navier-Stokes equations arising from H^2-initial conditions.

Theorem 4. *Consider periodic solutions of the Navier-Stokes equations (12) with viscosity $\nu = 1/2$, forcing $f = 0$, and observations $Z^{(N)}$ arising as in (21). Then there exists $c = c(U,T) > 0$ such that*

$$\liminf_{N\to\infty} \inf_{\tilde{\theta}_N} \sup_{\theta\in V:\|\theta\|_{H^2}\le U} P_\theta^N\Big(\|\tilde{\theta}_N - \theta\|_{L^2} > \frac{c}{\log N}\Big) > 1/4, \tag{28}$$

where the infimum extends over all estimators $\tilde{\theta}_N$ of θ (i.e., all measurable functions of $Z^{(N)}$ taking values in the space V).

This lower bound uses Theorem 2 and is reminiscent of similar 'logarithmic' minimax rates in the *linear* inverse problem of recovering the initial condition from an observed solution of the scalar heat equation (see, e.g., [26, 38]). As with the much simpler case of such heat equations, one can ask if faster rates can be obtained for 'super-smooth' initial conditions. If we use a Gaussian prior that has a slowly growing expansion in the eigenfunctions from (22), and if the true initial condition θ_0 is 'band-limited' in the Stokes spectrum, then we can indeed obtain convergence rates that approach the 'parametric' rate $1/\sqrt{N}$ of finite-dimensional models as we increase the regularity of the prior, $\alpha \to \infty$.

Theorem 5. *Denote by $(e_j : j \in \mathbb{N}) \subset V$ an enumeration of the L^2-orthonormal basis of H arising from the eigenfunctions of the Stokes operator $A = -P\Delta$ from (22), ordered by increasing eigenvalues. Let the prior Π be as in Condition 1 and project it onto the span $E_J = \{e_j : j \le J\}$ with $J = J_N = O(\log\log N)$. Suppose the ground truth initial condition θ_0 lies in $\mathcal{H} \cap E_{J_0}$ for some arbitrary fixed $J_0 \in \mathbb{N}$. Then the conclusions of Theorem 3 remain true with convergence rate*

$$\eta_N = (\log N)^\beta \times N^{-\alpha/(2\alpha+2)}, \quad \textit{some} \quad \beta > 0.$$

3 Proofs

3.1 Proof of Theorem 1

The idea of the proof is based on [2]. We can assume $w(0) \equiv u_0 - v_0 \neq 0$ in $V \subset L^2(\Omega)$ and by forward uniqueness therefore also $w(t) \equiv u(t) - v(t) \neq 0$ in $V \subset L^2(\Omega)$ for all t. The 'Dirichlet ratio' at time t is defined as

$$\Phi(t) = \frac{\|w(t)\|_V^2}{\|w(t)\|_{L^2}^2} = \frac{\langle Aw(t), w(t)\rangle_{L^2}}{\|w(t)\|_{L^2}^2}, \quad t \in [0,T], \tag{29}$$

where we recall (8) and where we now write, unless specified otherwise, $L^2 = L^2(\Omega)$. If we set $\bar{u} = (u+v)/2$ then we see from (7) and an elementary calculation that

$$B(u,u) - B(v,v) = B(\bar{u}, w) + B(w, \bar{u}). \tag{30}$$

Then since $u(t), v(t)$ solve the Navier-Stokes equations for the respective initial conditions we see that $w(t)$ solves the inhomogeneous non-linear parabolic equation in H given by

$$\frac{dw}{dt} + \nu Aw = g, \quad \text{where } g(t) = -B(\bar{u}(t), w(t)) - B(w(t), \bar{u}(t)) \tag{31}$$

with initial condition $w(0) = u(0) - v(0)$. The following bounds for the Dirichlet ratio (29) will be the key to the proof of Theorem 1.

Lemma 1. *We have*

$$\frac{d}{dt}\Phi(t) \leq \frac{\|g(t)\|_{L^2}^2}{\nu\|w(t)\|_{L^2}^2} \quad \forall t \in (0, T]. \tag{32}$$

Moreover,

$$\|g(t)\|_{L^2} \leq k(t)\|w(t)\|_V \ \ \textit{for all} \quad t \in (0, T] \tag{33}$$

for some $k \in L^4((0,T))$ whose L^4-norm is bounded by a fixed constant that depends on $T, \nu, \|f\|_{L^2}$ and on the initial conditions u_0, v_0 via the upper bound $U \geq \|u_0\|_V + \|v_0\|_V$.

Proof. We take the L^2-inner product of equation (31) with w and Aw, respectively, which gives

$$\frac{1}{2}\frac{d}{dt}\|w(t)\|_{L^2}^2 + \nu\|w(t)\|_V^2 = \langle g, w\rangle_{L^2},$$
$$\frac{1}{2}\frac{d}{dt}\|w(t)\|_V^2 + \nu\|Aw(t)\|_{L^2}^2 = \langle g, Aw\rangle_{L^2}$$

where we have also used (8) in the second identity. Therefore we have

$$\begin{aligned}
\frac{d}{dt}\Phi(t) &= 2\frac{\|w\|_{L^2}^2\big(\langle g, Aw\rangle_{L^2} - \nu\|Aw\|_{L^2}^2\big) - \|w\|_V^2\big(\langle g, w\rangle_{L^2} - \nu\|w\|_V^2\big)}{\|w\|_{L^2}^4} \\
&= 2\nu\frac{\|w\|_V^4 - \|w\|_V^2\langle g/\nu, w\rangle_{L^2} - \|w\|_{L^2}^2\|Aw\|_{L^2}^2 + \|w\|_{L^2}^2\langle g/\nu, Aw\rangle_{L^2}}{\|w\|_{L^2}^4} \\
&= 2\nu\frac{\big(\|w\|_V^2 - \langle g/2\nu, w\rangle_{L^2}\big)^2 - \langle g/2\nu, w\rangle_{L^2}^2 - \|w\|_{L^2}^2\|Aw\|_{L^2}^2 + \|w\|_{L^2}^2\langle g/\nu, Aw\rangle_{L^2}}{\|w\|_{L^2}^4} \\
&= 2\nu\frac{\langle Aw - g/2\nu, w\rangle_{L^2}^2 - \langle g/2\nu, w\rangle_{L^2}^2 - \|w\|_{L^2}^2\|Aw\|_{L^2}^2 + \|w\|_{L^2}^2\langle g/\nu, Aw\rangle_{L^2}}{\|w\|_{L^2}^4}.
\end{aligned}$$

using again (8) in the last step. By the Cauchy-Schwarz inequality, the last ratio is bounded from above by

$$\frac{2\nu}{\|w\|_{L^2}^4}\Big[\Big\|Aw - \frac{g}{2\nu}\Big\|_{L^2}^2\|w\|_{L^2}^2 + \frac{\|g\|_{L^2}^2\|w\|_{L^2}^2}{4\nu^2} - \|w\|_{L^2}^2\|Aw\|_{L^2}^2 + \|w\|_{L^2}^2\Big\langle\frac{g}{\nu}, Aw\Big\rangle_{L^2}\Big] = \frac{\|g\|_{L^2}^2}{\nu\|w\|_{L^2}^2}$$

and the first claim of the lemma follows. Using again the Cauchy-Schwarz and standard interpolation inequalities (Ch.6 in [4]) as well as the Poincaré inequality (p.292 in [13])

$$\frac{\|w(t)\|_{L^2}}{\|\nabla w(t)\|_{L^2}} \leq c,$$

we can bound $\|g(t)\|_{L^2}$ by

$$
\begin{aligned}
&\|B(\bar u(t), w(t))\|_{L^2} + \|B(w(t), \bar u(t))\|_{L^2} \\
&\le \|w(t)\|_{L^4(\Omega)} \|\nabla \bar u(t)\|_{L^4(\Omega)} + \|\bar u(t)\|_\infty \|\nabla w(t)\|_{L^2} \\
&\le c_0 \|w(t)\|_{L^2(\Omega)}^{1/2} \|\nabla w(t)\|_{L^2(\Omega)}^{1/2} \|\nabla \bar u(t)\|_{L^4(\Omega)} + \|\bar u(t)\|_\infty \|\nabla w(t)\|_{L^2(\Omega)} \\
&\le \|w(t)\|_V \big(c_1 \|\bar u(t)\|_{H^1}^{1/2} \|\bar u(t)\|_{H^2}^{1/2} + c_2 \|\bar u(t)\|_{L^2}^{1/2} \|\bar u(t)\|_{H^2}^{1/2} \big) \equiv k(t) \|w(t)\|_V,
\end{aligned}
$$

with appropriate universal constants c_i. Now by (13) the norms $\|\bar u(t)\|_{L^2}, \|\bar u(t)\|_V$ are bounded by an fixed constant and $\|\bar u(t)\|_{H^2} \lesssim \|u(t)\|_{\mathcal{D}(A)} + \|v(t)\|_{\mathcal{D}(A)}$ lies in $L^2((0,T))$ for strong solutions of the Navier-Stokes equations. In particular the L^4-norms of k are bounded as required, completing the proof of the lemma. □

Combining the two inequalities from the last lemma we obtain the differential inequality

$$
\frac{d}{dt}\Phi(t) \le \frac{k^2(t)}{\nu}\Phi(t) \quad \forall t \in (0,T], \tag{34}
$$

which after integrating (i.e., Gronwall's inequality, p.711 in [13]) implies

$$
\Phi(t) \le \Phi(0) e^{\frac{1}{\nu}\int_0^t k^2(s)ds} \equiv \Phi(0) K(t) \quad \forall t \in (0,T], \tag{35}
$$

where we notice that $\Phi(0)$ is bounded by a constant from below by the Poincaré inequality (p.292 in [13]), and also finite as $w(0) \neq 0$ and $u(0), v(0) \in V$ by hypothesis.

Now taking inner products of (31) with w and noting that $\langle B(\bar u, w), w\rangle_{L^2} = 0$ from (9) we arrive at

$$
\frac{1}{2}\frac{d}{dt}\|w(t)\|_{L^2}^2 + \nu \|w(t)\|_V^2 + \langle B(w, \bar u), w\rangle_{L^2} = 0, \quad 0 < t \le T.
$$

Just as in the proof of the previous lemma we can bound for some $C < \infty$

$$
|\langle B(w, \bar u), w\rangle_{L^2(\Omega)}| \le C \|w\|_{L^2} \|w\|_V \|\bar u\|_V
$$

and we obtain for all $0 < t \le T$ that

$$
\begin{aligned}
\frac{1}{2}\frac{d}{dt}\|w(t)\|_{L^2}^2 &\ge -\nu \|w(t)\|_V^2 - C\|w(t)\|_{L^2} \|w(t)\|_V \|\bar u(t)\|_V \\
&\ge \Big(-\nu \Phi(t) - C \|\bar u(t)\|_V \Phi^{1/2}(t) \Big) \|w(t)\|_{L^2}^2 \\
&\ge -\phi_T \|w(t)\|_{L^2}^2
\end{aligned}
$$

where

$$
\phi_T \equiv \sup_{0 \le t \le T} [\nu \Phi(t) + C \|\bar u(t)\|_V \Phi^{1/2}(t)].
$$

Again from this we can deduce the inequality

$$
\|w(t)\|_{L^2}^2 \ge \|w(0)\|_{L^2}^2 e^{-2T\phi_T} \quad \forall t \in (0,T]. \tag{36}
$$

We can also integrate the last inequality with respect to t over the interval $[0,T]$ to obtain similarly

$$\frac{1}{T}\int_0^T \|w(t)\|_{L^2}^2 dt \geq \|w(0)\|_{L^2}^2 e^{-2T\phi_T}, \tag{37}$$

as will be relevant later, in Corollary 1 below. It remains to examine the constant ϕ_T, and we distinguish the two cases A) and B) now.

B) Assume first the simpler case where an apriori bound (17) on the 'inverse Poincaré constant' $\Phi(0) \leq c_P$ is available. Then assuming $c_P \geq 1$ without loss of generality we have from (35) and Proposition 1 that

$$\phi_T \leq \sup_t [\nu\Phi(t) + C\|\bar{u}(t)\|_V \Phi^{1/2}(t)] \leq \mathcal{K} c_P$$

for some constant $\mathcal{K} = \mathcal{K}(U, \|f\|_{L^2}, \nu, T)$ and so (36) becomes

$$\|w(0)\|_{L^2} \leq e^{c_2 c_P}\|w(t)\|_{L^2}, \quad c_2 = 2T\mathcal{K},$$

which is the desired stability estimate.

A) We can proceed as in the last step but use the estimate $\Phi(0) \leq U/\|u(0)-v(0)\|_{L^2}$ instead of appealing to an inverse Poincaré constant. By the usual Poincaré inequality, $\Phi(0) \geq c > 0$ and hence $\sqrt{\Phi(0)} \lesssim \Phi(0)$. Combining this with (35) and (36) gives

$$\|u(0)-v(0)\|_{L^2}^2 \exp\Big\{-\frac{c'}{\|u(0)-v(0)\|_{L^2}^2}\Big\} \leq \|u(t)-v(t)\|_{L^2}^2 \tag{38}$$

for a constant $c' = c'(T, U, \nu, \|f\|_{L^2})$. Since $e^{-c'/x^2} \leq x^2/c'$ for all $x > 0$ we deduce

$$\exp\Big\{-\frac{2c'}{\|u(0)-v(0)\|_{L^2}^2}\Big\} \leq \|u(t)-v(t)\|_{L^2}^2/c' \equiv Z.$$

If we set $c_1^2 = c'$ then by hypothesis the right hand side $Z < 1$ and hence $\log Z < 0$, so taking logarithms in the previous display gives

$$-\frac{2c'}{\|u(0)-v(0)\|_{L^2}^2} \leq -\log(1/Z) \iff \|u(0)-v(0)\|_{L^2}^2 \leq \frac{2c'}{\log(1/Z)}$$

from which it follows that

$$\|u(0)-v(0)\|_{L^2} \leq \sqrt{2c'}\Big(\log\frac{c_1^2}{\|u(t)-v(t)\|_{L^2}^2}\Big)^{-1/2},$$

completing the proof upon setting $c_0 = \sqrt{c'}$.

3.2 Proof of Theorem 2

Proof. For $j = 1, 2, \dots$ we define univariate trigonometric polynomials on $[0, 2\pi]$ as $\phi_j(x) = e^{ijx}/\sqrt{2\pi}, x \in [0, 2\pi]$. Then for initial conditions $w_j(0) = \phi_j/j^2$ the (scalar) heat equation

$$\frac{\partial w}{\partial t} - \frac{\partial^2 w}{\partial x^2} = 0 \text{ on } [0, 2\pi] \times [0, T]$$

has unique solutions

$$w_j(t, x) = \frac{-e^{-j^2 t}}{j^2} \phi_j(x), \ x \in [0, 2\pi], t \in [0, T].$$

Now following an idea in Remark 7 in [15], consider the periodic Navier-Stokes equation (12) for initial conditions

$$u_j(0, x) = j^{-2}(\phi_j(x_1 - x_2), \phi_j(x_1 - x_2)), \quad x = (x_1, x_2) \in \Omega,$$

which clearly satisfy $\nabla \cdot u_j(0) = 0$ and hence lie in $C^\infty(\Omega) \cap V$. The vector fields

$$u_j(t, x) = (w_j(t, x_1 - x_2), w_j(t, x_1 - x_2)), \quad x = (x_1, x_2) \in \Omega,$$

are also divergence free, have *vanishing non-linear term* $u_j \cdot \nabla u_j = 0$, and solve the Navier-Stokes equations (12) for $\nu = 1/2, f = 0$ and the given initial conditions. To compute the quantities in (19), notice that the vector fields

$$(\bar{e}_{lk}, \bar{e}_{l'k'}), \text{ with } \bar{e}_{lk}(x_1, x_2) = \phi_l(x_1)\phi_k(x_2), x_i \in [0, 2\pi], \ k, k', l, l' \in \mathbb{Z},$$

form an orthonormal tensor basis of the Hilbert space $L^2([0, 2\pi]^2)^2$. Also, for any fixed x_i, by standard properties of trigonometric polynomials (and for δ_{jl} the Kronecker-δ)

$$\langle \phi_j(\cdot - x_i), \phi_l \rangle_{L^2([0,2\pi])} \propto e^{-ijx_i} \langle \phi_j, \phi_l \rangle_{L^2([0,2\pi])} = \delta_{jl}$$

and hence using Parseval's identity (and with universal constants in $\simeq$)

$$\|u_j(0)\|_{L^2([0,2\pi]^2)^2} \simeq j^{-2}, \quad \|u_j(t)\|_{L^2([0,2\pi]^2)^2} \simeq e^{-j^2 t}/j^2.$$

Similarly, the Sobolev norms of these initial conditions are of order

$$\|u_j(0)\|_{H^2}^2 \lesssim \sum_{l,k} (|l|^2 + |k|^2) \langle u_j(0), \bar{e}_{lk} \rangle_{L^2([0,2\pi]^2)^2}^2 \lesssim \frac{j^2}{j^2} \leq const.$$

The inequality (20) now follows from (19), $v(0) = v(t) = 0$ and elementary properties of the exponential/logarithm map. □

3.3 Proof of Theorems 3 and 5

We will apply the general theory for non-linear Bayesian inverse problems from [34] in conjunction with Proposition 1 and the stability estimate Theorem 1. In fact we will use the following corollary to this theorem, proved by just replacing (36) by (37) in the last step of its proof.

Corollary 1. *In the setting of the Theorem 1 B), for any $T > 0$, there exists a constant c_2 depending on $U, \nu, T, \|f\|_{L^2}$ such that*

$$\|u(0) - v(0)\|_{L^2(\Omega)} \leq e^{c_2 c_P} \Big(\frac{1}{T} \int_0^T \|u(t, \cdot) - v(t, \cdot)\|^2_{L^2(\Omega)} dt\Big)^{1/2}, \tag{39}$$

while in the setting of Theorem 1 A), there exist constants c_0, c_1 depending on $U, \nu, T, \|f\|_{L^2}$ such that

$$\|u(0) - v(0)\|_{L^2(\Omega)} \leq c_0 \Big(\log \frac{c_1}{\int_0^T \|u(t, \cdot) - v(t, \cdot)\|^2_{L^2(\Omega)} dt}\Big)^{-1/2} \tag{40}$$

where $\|u - v\|^2_{L^2((0,T]\times\Omega)} < c_1$.

Now in the notation of Section 1.2.1 in [34], the parameter space Θ of initial conditions is chosen as the subspace

$$\Theta \equiv V \cap H^2(\Omega)^2 \ \text{ of } L^2(\mathcal{Z}, W)$$

with V from (4), and where we choose $\mathcal{Z} = \Omega$ and $W = \mathbb{R}^2$. The forward map

$$\theta \mapsto \mathscr{G}(\theta) = u_\theta, \ \mathscr{G} : \Theta \to L^2(\mathcal{X}, \mathbb{R}^2),$$

is the solution map of the PDE (12) with initial condition $u_0 = \theta$, where we set $\mathcal{X} = (0, T] \times \Omega$ and the finite-dimensional vector space V in [34] (different from V in our context) is identified with $\mathbb{R}^2$ here. Then on $\mathcal{X}$ we have the uniform probability measure λ while on $\mathcal{Z}$ we can just take ζ equal to Lebesgue measure, so that our statistical model (21) co-incides precisely with the one in eq. (1.9) in [34] with 'random design' $(t_i, X_i)_{i=1}^N$ equal to the $(X_i)_{i=1}^N$ there. By hypothesis, the prior is supported in $V \cap \mathcal{D}(A) \equiv \mathcal{R}$ almost surely and we will apply Theorem 2.2.2 in [34] with regularisation space $\mathcal{R}$, norm $\|\cdot\|_{\mathcal{R}} \equiv \|\cdot\|_{H^2}$, and choices

$$\kappa = 0, \delta_N = N^{-\alpha/(2\alpha+2)}, \ d = 2.$$

The Condition 2.1.1 in [34] is then verified in view of Proposition 1, and we obtain that for some large enough constant $M > 0$, as $N \to \infty$,

$$\Pi\big(\theta \in V : \|\theta\|_{H^2} \leq M, \|u_\theta - u_{\theta_0}\|_{L^2((0,T]\times\Omega)} \leq M\delta_N | Z^{(N)}\big) \to^{P^N_{\theta_0}} 1. \tag{41}$$

By the stability estimate Theorem 1 with $u = u_\theta, v = u_{\theta_0}$, specifically (40), this gives contraction rate η_N for $\|\theta - \theta_0\|_{L^2}$ and proves (26). The bound (25) then follows from what precedes and the forward estimate from Proposition 1, Part 2) (now with T_p replacing T). The last claim of Theorem 3 (convergence rate of the posterior mean) follows from the preceding bounds and a uniform integrability argument given in Theorem 2.3.2 in [34], and again Proposition 1, and details are left to the reader.

The proof of Theorem 5 follows the same pattern (cf. also Exercise 2.4.3 in [34]), noting that the RKHS is now $E_J \cap \mathcal{H}$ were E_J is the linear span of the Stokes-eigenfunctions up to order J. The preceding arguments go through since $\theta_0 \in E_J \cap \mathcal{H}$ for all N (and thus J) large enough. We can then use the stronger stability estimate (39) with inverse Poincare constant now growing at most as $c_P = O(\log \log N)$ as explained in Remark 1, so that the result follows by introducing the additional log-factor $e^{c_2 c_P} = O((\log N)^\beta)$ in the contraction rate.

3.4 Proof of Theorem 4

We use a standard lower bound proof technique from nonparametric statistics, e.g., Theorem 6.3.2 in [17] (or since we will only use a two-hypotheses case, we can also argue as in the the proof of Theorem 2.2 in [1]). The Kullback-Leibler divergence in our measurement model (21) is

$$KL(P_\theta^N, P_{\theta_0}^N) = \frac{N}{2T} \|u_\theta - u_{\theta_0}\|^2_{L^2((0,T]\times\Omega)^2}, \quad \theta \in \Theta,$$

see Proposition 1.3.1 in [34] (and with choices of $\Theta, \mathcal{X}, V, \lambda$ etc. as described after Corollary 1). Using the base hypothesis $\theta_0 = 0$ and alternative hypothesis $\theta_j = u_j(0)$ from Theorem 2, we can for every $\mu > 0$ choose $j = j_N = \sqrt{L \log N}$ such that

$$KL(P_{\theta_j}^N, P_{\theta_0}^N) \lesssim N \|u_j(t)\|^2_{L^2(\Omega\times(0,T))} \lesssim N e^{-c j_N^2 T} \le \mu$$

for $L = L(\mu)$ large enough. At the same time $\|\theta_j - \theta_0\|_{L^2} \gtrsim \log N$ from Theorem 2, which verifies (6.100) in [17] with $r_n \simeq \log N$, and so the result follows from Theorem 6.3.2 in [17] with $M = 2$ and μ small enough (cf. also (6.99) to obtain the 'in probability' version of the lower bound).

3.5 Proof of Proposition 1

We start with the following basic a priori estimate for solutions $u = u(t) \in V$ of (12): Taking the L^2-inner product of (12) with u and using (8), (9) we obtain

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|u\|^2_{L^2} + \nu\|\nabla u\|^2_{L^2} &= \langle f, u\rangle_{L^2} \le \lambda^{-1}\|f\|_{L^2}\|\nabla u\|_{L^2} \\ &\le \frac{1}{2\nu\lambda}\|f\|^2_{L^2} + \frac{\nu}{2}\|\nabla u\|^2_{L^2}\end{aligned}$$

where we have used the Cauchy-Schwarz, the Poincaré (with constant λ) and the Young inequalities. This readily implies,

$$\frac{d}{dt}\|u\|^2_{L^2} + \nu\|\nabla u\|^2_{L^2} \le \frac{\|f\|^2_{L^2}}{\nu\lambda},$$

which can be integrated to give

$$\int_0^T \|\nabla u(t)\|_{L^2}^2 dt \leq \frac{\|u_0\|_{L^2}^2}{\nu} + T\frac{\|f\|_{L^2}^2}{\nu^2\lambda} \leq K_1 \equiv K_1(U,\nu,T,\|f\|_{L^2}). \tag{42}$$

We now turn to the proof of A). We will show

$$\sup_{u_0\in V:\|u_0\|_{H^2}\leq m} \sup_{0\leq t\leq T} \|u(t)\|_{H^2(\Omega)} \leq c \tag{43}$$

for a constant $c = c(m, \|f\|_{H^1}, T, \nu)$, which implies (14) by the Sobolev imbedding $H^2 \subset L^\infty$ and which also implies (13) for initial conditions bounded in H^2 (rather than in H^1). The proof of (13) under weaker H^1-conditions follows from similar arguments, but we omit it here as we only require $\theta = u(0) \in H^2$ elsewhere in this paper.

Applying the curl operation $\nabla\times$ to (11) we obtain the vorticity formulation of the Navier-Stokes equations

$$\frac{\partial}{\partial t}\omega - \nu\Delta\omega + u\cdot\nabla\omega = \nabla\times f \tag{44}$$

where the vorticity $\omega = \nabla^\perp\cdot u$. Notice that since $\nabla\cdot u = 0$ one has $\|\nabla\omega\|_{L^2} = \|\Delta u\|_{L^2}$. Therefore, we can prove (43) by bounding the H^1-norms of ω uniformly in time $t \in [0,T]$. Taking the L^2-inner product of the last equation with $-\Delta\omega$ and using (8) we obtain

$$\frac{1}{2}\frac{d}{dt}\|\nabla\omega\|_{L^2}^2 + \nu\|\Delta\omega\|_{L^2}^2 - \int_\Omega (u\cdot\nabla\omega)\Delta\omega = -\int_\Omega (\nabla\times f)\Delta\omega. \tag{45}$$

The Cauchy-Schwarz inequality implies

$$\Big| -\int_\Omega (\nabla\times f)\Delta\omega\Big| \leq \|\nabla f\|_{L^2}\|\Delta\omega\|_{L^2}$$

and we can further estimate, by integration by parts

$$\begin{aligned}\Big| -\int_\Omega (u\cdot\nabla\omega)\Delta\omega\Big| = \Big| -\sum_{l=1}^2\int_\Omega (u\cdot\nabla\omega)\partial_l^2\omega\Big| = \Big|\sum_{l=1}^2\int_\Omega \left[(\partial_l u\cdot\nabla\omega)\partial_l\omega\right]\Big| \\ \leq \|\nabla u\|_{L^2}\|\nabla\omega\|_{L^4}^2 \leq c_0\|\nabla u\|_{L^2}\|\nabla\omega\|_{L^2}\|\Delta\omega\|_{L^2}\end{aligned}$$

since $\sum_l\int_\Omega(u\cdot\nabla\partial_l\omega)\partial_l\omega = 0$ as in (9), and where we have used, for every $\varphi\in H^1$ with $\int_\Omega\phi = 0$, the interpolation inequality $\|\phi\|_{L^4}^2 \leq c_0\|\phi\|_{L^2}\|\nabla\phi\|_{L^2}$. Substituting the previous bounds into (45) gives

$$\frac{1}{2}\frac{d}{dt}\|\nabla\omega\|_{L^2}^2 + \nu\|\Delta\omega\|_{L^2}^2 \leq \|\nabla f\|_{L^2}\|\Delta\omega\|_{L^2} + c_0\|\nabla u\|_{L^2}\|\nabla\omega\|_{L^2}\|\Delta\omega\|_{L^2}.$$

By Young's inequality for products we deduce

$$\frac{1}{2}\frac{d}{dt}\|\nabla\omega\|_{L^2}^2 + \nu\|\Delta\omega\|_{L^2}^2 \leq \frac{\|\nabla f\|_{L^2}^2}{\nu} + \frac{\nu}{4}\|\Delta\omega\|_{L^2}^2 + \frac{c_0^2}{\nu}\|\nabla u\|_{L^2}^2\|\nabla\omega\|_{L^2}^2 + \frac{\nu}{4}\|\Delta\omega\|_{L^2}^2$$

and rearranging we obtain

$$\frac{d}{dt}\|\nabla\omega\|_{L^2}^2 \leq \frac{2}{\nu}\|\nabla f\|_{L^2}^2 + \frac{2c_0^2}{\nu}\|\nabla u\|_{L^2}^2\|\nabla\omega\|_{L^2}^2.$$

We can apply Gronwall's inequality (p.711 in [13]) to deduce

$$\|\nabla\omega(t)\|_{L^2}^2 \leq e^{\frac{2c_0^2}{\nu}\int_0^t \|\nabla u(s)\|_{L^2}^2 ds}\Big[\|\nabla\omega(0)\|_{L^2}^2 + \frac{2t}{\nu}\|\nabla f\|_{L^2}^2\Big], \quad 0 < t \leq T. \tag{46}$$

Now we use (42) to bound the constants in the preceding integrals and deduce

$$\|u\|_{H^2} \lesssim \|\Delta u(t)\|_{L^2}^2 = \|\nabla\omega(t)\|_{L^2}^2 \lesssim \big(\|\Delta u(0)\|_{L^2}^2 + \|f\|_{H^1}\big) \leq K_2, \quad 0 \leq t \leq T, \tag{47}$$

and $K_2 = K_2(T, \nu, \|f\|_{H^1}, m)$, which implies (43) as desired.

To conclude the proof of the proposition, we note that what precedes are *formal* estimates for solutions to Navier-Stokes equations. The existence and uniqueness of such solutions then follows from these formal estimates and standard compactness arguments applied to Galerkin approximations of solutions of (12), in complete analogy to those given in [9], Ch.9, and are left to the reader. Finally, the estimate (15) in B) also follows from simple variations of the preceding arguments, see the steps leading to eq. (10.6) in [9], and is also left to the reader.

Acknowledgements. RN was supported by EPSRC grant EP/V026259. The research of EST was made possible by NPRP grant #S-0207-200290 from the Qatar National Research Fund (a member of Qatar Foundation), and is based upon work supported by King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-2020-CRG9-4336. The work of EST has also benefited from the inspiring environment of the CRC 1114 "Scaling Cascades in Complex Systems", Project Number 235221301, Project A02, funded by Deutsche Forschungsgemeinschaft (DFG).

References

[1] Abraham, K., and Nickl, R. On statistical Calderón problems. *Math. Stat. Learn. 2*, 2 (2019), 165–216.

[2] Bardos, C., and Tartar, L. Sur l'unicité rétrograde des équations paraboliques et quelques questions voisines. *Arch. Rational Mech. Anal. 50* (1973), 10–25.

[3] Bennett, A. F. *Inverse modeling of the ocean and atmosphere.* Cambridge University Press, Cambridge, 2002.

[4] Bergh, J., and Löfström, J. *Interpolation spaces. An introduction.* Grundlehren der Mathematischen Wissenschaften, No. 223. Springer-Verlag, Berlin-New York, 1976.

[5] Beskos, A., Girolami, M., Lan, S., Farrell, P. E., and Stuart, A. M. Geometric MCMC for infinite-dimensional inverse problems. *J. Comput. Phys. 335* (2017), 327–351.

[6] Bohr, J. Stability of the non-abelian X-ray transform in dimension ≥ 3. *J. Geom. Anal. 31*, 11 (2021).

[7] Bohr, J., and Nickl, R. On log-concave approximations of high-dimensional posterior measures and stability properties in non-linear inverse problems. *Ann. Inst. H. Poincaré (Probab. Stat.), to appear* (2023).

[8] Borggaard, J., Glatt-Holtz, N., and Krometis, J. On Bayesian consistency for flows observed through a passive scalar. *Ann. Appl. Probab. 30*, 4 (2020), 1762–1783.

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

[10] Cotter, S. L., Dashti, M., Robinson, J. C., and Stuart, A. M. Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems 25*, 11 (2009), 115008, 43.

[11] Cotter, S. L., Roberts, G. O., Stuart, A. M., and White, D. MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci. 28*, 3 (2013), 424–446.

[12] Cui, T., Law, K. J. H., and Marzouk, Y. M. Dimension-independent likelihood-informed MCMC. *J. Comput. Phys. 304* (2016), 109–137.

[13] Evans, L. C. *Partial differential equations*, second ed., vol. 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010.

[14] Evensen, G., Vossepoel, F., and van Leeuwen, J. *Data assimilation fundamentals*. Springer textbooks in Earth sciences, Geography and Environment. Springer, Cham, 2022.

[15] Foias, C., and Saut, J.-C. Asymptotic behavior, as $t \to +\infty$, of solutions of Navier-Stokes equations and nonlinear spectral manifolds. *Indiana Univ. Math. J. 33*, 3 (1984), 459–477.

[16] Ghosal, S., and van der Vaart, A. W. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, New York, 2017.

[17] Giné, E., and Nickl, R. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press, New York, 2016.

[18] Giordano, M., and Nickl, R. Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. *Inverse Problems 36*, 8 (2020).

[19] Giordano, M., and Ray, K. Nonparametric bayesian inference for reversible multi-dimensional diffusions. *Ann. Statist.* (2022, to appear).

[20] Graham, I. G., Kuo, F. Y., Nichols, J. A., Scheichl, R., Schwab, C., and Sloan, I. H. Quasi-Monte Carlo finite element methods for elliptic PDEs with lognormal random coefficients. *Numer. Math. 131*, 2 (2015), 329–368.

[21] Hairer, M., Stuart, A. M., and Vollmer, S. J. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Ann. Appl. Probab. 24*, 6 (2014), 2455–2490.

[22] Hoffmann, M., and Ray, K. Bayesian estimation in a multidimensional diffusion model with high frequency data. *arXiv* (2022).

[23] Kalnay, E. *Atmospheric modelling, data assimilation, and predictability*. Cambridge University Press, Cambridge, UK, 2003.

[24] Kekkonen, H. Consistency of bayesian inference with gaussian process priors for a parabolic inverse problem. *Inverse Problems* (2021).
[25] Knapik, B., van der Vaart, A. W., and van Zanten, J. H. Bayesian inverse problems with Gaussian priors. *Ann. Statist. 39*, 5 (2011), 2626–2657.
[26] Knapik, B. T., van der Vaart, A. W., and van Zanten, J. H. Bayesian recovery of the initial condition for the heat equation. *Comm. Statist. Theory Methods 42*, 7 (2013), 1294–1313.
[27] Law, K., Stuart, A., and Zygalakis, K. *Data assimilation*, vol. 62 of *Texts in Applied Mathematics*. Springer, Cham, 2015. A mathematical introduction.
[28] Lorenz, E. N. Deterministic nonperiodic flow. *J. Atmospheric Sci. 20*, 2 (1963), 130–141.
[29] Majda, A. J., and Harlim, J. *Filtering complex turbulent systems*. Cambridge University Press, Cambridge, 2012.
[30] Monard, F., Nickl, R., and Paternain, G. P. Consistent inversion of noisy non-Abelian X-ray transforms. *Comm. Pure Appl. Math. 74*, 5 (2021), 1045–1099.
[31] Monard, F., Nickl, R., and Paternain, G. P. Statistical guarantees for Bayesian uncertainty quantification in nonlinear inverse problems with Gaussian process priors. *Ann. Statist. 49*, 6 (2021), 3255–3298.
[32] Nickl, R. Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation. *J. Eur. Math. Soc. (JEMS) 22*, 8 (2020), 2697–2750.
[33] Nickl, R. Inference for diffusions from low frequency measurements. *arXiv* (2022).
[34] Nickl, R. *Bayesian non-linear statistical inverse problems.* Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS) press, Berlin, 2023.
[35] Nickl, R., and Ray, K. Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Ann. Statist. 48*, 3 (2020), 1383–1408.
[36] Nickl, R., van de Geer, S., and Wang, S. Convergence rates for penalized least squares estimators in PDE constrained regression problems. *SIAM/ASA J. Uncertain. Quantif. 8*, 1 (2020), 374–413.
[37] Nickl, R., and Wang, S. On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms. *J. Eur. Math. Soc. (JEMS), to appear* (2022).
[38] Ray, K. Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat. 7* (2013), 2516–2549.
[39] Reich, S., and Cotter, C. *Probabilistic forecasting and Bayesian data assimilation.* Cambridge University Press, New York, 2015.
[40] St-Amant, S. Stability estimate for the broken non-abelian X-ray transform in Minkowski space. *Inverse Problems 38*, 10 (2022), Paper No. 105007, 36.
[41] Stuart, A. M. Inverse problems: a Bayesian perspective. *Acta Numer. 19* (2010), 451–559.