On Bayesian data assimilation for PDEs with ill-posed forward problems

OPEN ACCESS
IOP Publishing
Inverse Problems
Inverse Problems **38** (2022) 085012 (44pp)
https://doi.org/10.1088/1361-6420/ac7acd

On Bayesian data assimilation for PDEs with ill-posed forward problems

S Lanthaler[1,*], **S Mishra**[1] and **F Weber**[2]

[1] Seminar for Applied Mathematics, ETH Zürich Rämistrasse 101, 8092 Zürich, Switzerland
[2] Department of Mathematical Sciences, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, United States of America

E-mail: samuel.lanthaler@math.ethz.ch, siddhartha.mishra@math.ethz.ch and franzisw@andrew.cmu.edu

Received 19 July 2021, revised 15 May 2022
Accepted for publication 21 June 2022
Published 12 July 2022

Abstract

We study Bayesian data assimilation (filtering) for time-evolution Partial differential equations (PDEs), for which the underlying forward problem may be very unstable or ill-posed. Such PDEs, which include the Navier–Stokes equations of fluid dynamics, are characterized by a high sensitivity of solutions to perturbations of the initial data, a lack of rigorous global well-posedness results as well as possible non-convergence of numerical approximations. Under very mild and readily verifiable general hypotheses on the forward solution operator of such PDEs, we prove that the posterior measure expressing the solution of the Bayesian filtering problem is stable with respect to perturbations of the noisy measurements, and we provide quantitative estimates on the convergence of approximate Bayesian filtering distributions computed from numerical approximations. For the Navier–Stokes equations, our results imply uniform stability of the filtering problem even at arbitrarily small viscosity, when the underlying forward problem may become ill-posed, as well as the compactness of numerical approximants in a suitable metric on time-parametrized probability measures.

Keywords: Bayesian inversion, data assimilation, fluid dynamics, numerical discretization

(Some figures may appear in colour only in the online journal)

*Author to whom any correspondence should be addressed.

1. Introduction

Partial differential equations (PDEs) are ubiquitous as mathematical models in the sciences and engineering. A time-dependent PDE takes the following generic form,

$$\begin{aligned} \partial_t u + \mathcal{D}\big(f, u, \nabla_x u, \nabla_x^2 u, \ldots\big) &= 0, \qquad \forall\ x \in D, t \in (0,T) \\ \mathcal{B}u &= \hat{u}, \qquad \forall\ x \in \partial D, t \in (0,T), \\ u(x,0) &= \bar{u}, \qquad \forall\ x \in D. \end{aligned} \tag{1.1}$$

Here, $\mathcal{D}$ is a differential operator that depends on the solution u and its spatial derivatives, as well as on a coefficient (source term) f. The PDE is supplemented with initial conditions and with boundary conditions, imposed through a boundary operator $\mathcal{B}$. The inputs to the PDE are given by $\mathbf{v} = [\bar{u}, \hat{u}, f]$, which constitute the initial data, boundary data and coefficients (source terms). These inputs are related to the solution u of the PDE (1.1) through the so-called *data-to-solution operator*,

$$\mathcal{S}^\dagger : X \mapsto Y, \qquad \mathbf{v} \mapsto \mathcal{S}^\dagger(\mathbf{v}) = u, \tag{1.2}$$

with u solving the PDE (1.1). X and Y are suitable (subsets of) Banach spaces.

Often, one is interested, not just in the solution field u of (1.1), but rather in finite-dimensional *quantities of interest* or *observables*, which are given in the generic form,

$$\mathcal{L}^\dagger : X \to \mathbb{R}^d, \qquad \mathbf{v} \mapsto \mathcal{L}^\dagger(\mathbf{v}). \tag{1.3}$$

The observable $\mathcal{L}^\dagger$ can be written as a composition $\mathcal{L}^\dagger(\mathbf{v}) = \mathcal{G}^\dagger(\mathcal{S}^\dagger(\mathbf{v}))$, with $\mathcal{G}^\dagger : Y \to \mathbb{R}^d$ a functional (e.g. given by point evaluations or local averages). Thus, the so-called *forward problem* for the PDE (1.1), is to evaluate the solution operator $\mathcal{S}^\dagger$ or the observables $\mathcal{L}^\dagger$, given the inputs $\mathbf{v}$.

However, it is not always possible to exactly *know* the inputs $\mathbf{v}$ (initial and boundary data, coefficients, sources etc). Rather in practice, one has to infer information about the inputs $\mathbf{v}$, and consequently the solution u, from *measurements* of the observables in (1.3). Moreover in general, these measurements are *noisy*. Thus one has to solve the so-called *inverse problem* for a PDE, i.e., *determine the input* $\mathbf{v}$ *(and solution* u*) for the PDE* (1.1)*, given measurements of the form,*

$$y = \mathcal{L}^\dagger(\mathbf{v}) + \eta, \qquad \eta \sim \rho(y)\mathrm{d}y, \tag{1.4}$$

with the noise sampled from a probability measure on $\mathbb{R}^d$, defined by its density ρ.

It is well known that, in general, a *deterministic formulation* of the afore-mentioned inverse problem can be *ill-posed*. Although different regularization procedures have been developed over the last few decades to deal with this ill-posedness, it is now well-established that a statistical formulation of the inverse problem, based on a Bayesian framework, is very suitable in this context [18, 36, 39].

Within a Bayesian formulation of the inverse problem, associated with the mapping (1.3) and measurements (1.4), one encodes statistical information about the system (say inputs **v** in (1.1)) in terms of a *prior probability measure*. The additional information from the measurements (1.4) can be used to improve the prior by an application of the well-known Bayes' theorem [36]. This results in a so-called *posterior* probability measure, on the inputs **v**, which represents the conditional probability of the underlying inputs, given the measurements (1.4). Thus, the *Bayesian inverse problem* (BIP) can be interpreted as a *mapping from the measurements* (1.4) *to the posterior measure*.

In contrast to the generic situation for deterministic inverse problems, it has been shown that the corresponding BIP for PDEs is often *well-posed*, i.e., the posterior measures exists, is unique and *depends continuously* (in suitable metrics) on the measurements (1.4) [25, 35, 36]. Furthermore, BIPs can incorporate the deterministic formulation of regularized ill-posed inverse problems: as shown in [36], the latter can often be viewed as the maximum *a posteriori* (MAP) estimator of a BIP with a suitable choice of the underlying prior.

However, these remarkable well-posedness results for BIPs for PDEs rely on the well-posedness of the underlying *forward problem*, often requiring that the mapping $\mathcal{L}^\dagger$ in (1.3) is *Lipschitz continuous* in suitable metrics and converting this Lipschitz continuity into stability results for the posterior measure with respect to perturbations in the measurements, see [36] for a survey of these results and their applications to a variety of PDEs. More recently in [25, 35], these Lipschitz continuity assumptions on the forward map $\mathcal{L}^\dagger$ in (1.3), have been considerably relaxed. In particular, under suitable assumptions on the measurement noise η in (1.4), mere *existence and measurability* of the forward map suffices for the well-posedness of the underlying BIP [25].

For models describing the *temporal evolution* $t \mapsto u(t)$ of a system, the forward operator $\mathcal{S}^\dagger$ can either be interpreted as a mapping from the given data to a space of time-dependent solutions $t \mapsto u(t)$, or equivalently, as a time-parametrized operator $\mathcal{S}_t^\dagger : X \to X$, such that $t \mapsto u(t) = \mathcal{S}_t^\dagger(\bar{u}; \hat{u}, f)$ describes the evolution of the system. For such systems, which include the PDE (1.1), Bayesian inversion can be used to estimate the initial state $\bar{u}$, the boundary conditions $\hat{u}$, or the source term f, under very general conditions on the measurement operator $\mathcal{L}^\dagger$. However, for many problems of practical importance, one is ultimately interested in an estimate of the *underlying state* $u(t)$ at the present or a future time t. The resulting *data assimilation* (DA) or *filtering* problem thus seeks to blend measurement data with the underlying evolution model to make predictions about the future state. Besides its intrinsic interest, one motivation for studying a statistical viewpoint of DA based on a Bayesian approach [1] comes from the fact that many popular methods for DA, such as the three-dimensional variational filter (3DVAR) [5], the four-dimensional variational filter [33] or the ensemble Kalman filter (e.g. [8] and references therein) can suitably be interpreted as arising from MAP estimators or Gaussian approximations of this Bayesian approach [1]. In fact, the Bayesian formulation has been proposed as a 'gold-standard' against which other methods can be evaluated [27]. A mathematically rigorous introduction to DA from this Bayesian perspective is presented in [26], where attention is restricted to finite-dimensional models and Gaussian noise. To the best of the authors' knowledge, a systematic investigation of the well-posedness of Bayesian DA for *infinite-dimensional models* arising from PDEs, and the extension of the corresponding theory on BIPs of [36] to the DA setting, has so far been outstanding.

DA is of particular importance in the context of fluid flows. For many fluid models, it is well-known that predictions of future states can depend very sensitively on small perturbations of the initial data (or boundary data, source terms, etc) [12, 32]. This sensitivity to small changes can render the forward evolution (effectively) *ill-posed*. Prototypical examples for such ill-posed PDEs are provided by the fundamental equations of fluid dynamics, such as

the incompressible Navier–Stokes or the compressible Euler equations. For the incompressible Navier–Stokes equations, there are currently no global well-posedness results in three space dimensions. Although admissible weak solutions exist [28], the uniqueness, stability and regularity of such solutions are outstanding open problems. Even for the two-dimensional Navier–Stokes equations, for which existence and uniqueness results have been obtained [21], the known stability estimates for the forward problem exhibit a very unfavourable, exponential dependence $\sim \exp(t/\nu)$ on the viscosity[3] ν, reflecting the high sensitivity to small perturbations of the initial data. As the viscosity $\nu \ll 1$ is often a very small number in applications, this can render even the two-dimensional Navier–Stokes equations so unstable, as to be effectively ill-posed. Similar remarks apply to the compressible Euler equations, which are canonical examples of *hyperbolic systems of conservation laws* [7]. In this case, there are no rigorous global-in-time well-posedness results in either two or three space dimensions, reflecting a lack of stability of the underlying forward solution operator $\mathcal{S}_t^\dagger$, or of (numerical) approximations thereof $\mathcal{S}_t^\Delta \approx \mathcal{S}_t^\dagger$. Due to the lack of stability of the forward problem for these fundamental equations of fluid dynamics, it is thus not clear to what extent the well-posedness results of [25, 35, 36] obtained for BIPs can be extended to this time-varying setting. Indeed, Bayesian inversion apparently only yields estimates on the *initial state* in this setting, whereas DA involves an additional prediction step to estimate *future states*.

This lack of stability of many fluids with respect to perturbations of the initial data motivates the following question: does Bayesian DA suffer from a similar sensitivity to perturbations in the measurement data? I.e., is the well-posedness of the Bayesian DA problem contingent on the well-posedness of the corresponding forward problem? The Bayesian framework has been remarkably successful in the context of weather forecasting, climate modeling and oceanography [34]. Given that the underlying models include the incompressible Navier–Stokes and the compressible Euler equations as the core governing PDEs, how does one reconcile the empirical success of the Bayesian framework with the lack of stability of the underlying forward problem?

This dichotomy sets the stage for the current article where we investigate the well-posedness of Bayesian DA for PDEs where the forward evolution operator may be ill-posed. Besides investigating the well-posedness of Bayesian DA for the exact solution operator $\mathcal{S}_t^\dagger$, we also consider approximations to the forward map, $\mathcal{S}_t^\Delta \approx \mathcal{S}_t^\dagger$, which may stem from numerical approximations of the underlying PDE (1.1). Such approximations lead to a family of *approximate posteriors* for the Bayesian DA problems. In this article, we will prove, under very general hypotheses, that

- The Bayesian filtering problem is well-posed under mild assumptions, even if the forward problem may be *ill-posed*; in particular, the mapping from measurements to posterior is uniformly Lipschitz continuous, *independently* of the stability of the forward problem.
- Under mild conditions on the convergence of approximate solution operators $\mathcal{S}_t^\Delta \to \mathcal{S}_t^\dagger$, the corresponding approximate Bayesian posteriors are consistent, in the sense that they converge in a suitable metric to the exact posterior as $\Delta \to 0$, and with the same convergence rate.
- Suitable families of approximate posteriors for the Navier–Stokes equations (and related equations) are compact in an appropriate metric, as $\Delta \to 0$, with uniformly continuous dependence on the measurements $\boldsymbol{y}$. This allows us to define a non-empty set of candidate

[3] Physically, the non-dimensional quantity to consider is the Reynolds number $1/\mathrm{Re} \propto \nu$, obtained after suitable normalization of the equation. We will assume that the equations are suitably scaled, and will not distinguish between Reynolds number and the viscosity.

solutions for the limiting Bayesian DA problem, as $\Delta \to 0$, even for models for which there are no known convergence guarantees, $\mathcal{S}_t^\Delta \xrightarrow{??} \mathcal{S}_t^\dagger$, for the forward problem, such as the *three-dimensional* Navier–Stokes equations.

Although uniqueness of the posterior is not necessarily guaranteed with these compactness arguments, our construction could pave the way for proposing additional selection criteria on the set of approximate posteriors to recover uniqueness.

1.1. Organization

This work is organized as follows: in section 2, we introduce the precise mathematical setting. To this end, we first formalize Bayesian DA in the infinite-dimensional setting considered in the present work, and provide a formal definition of well-posedness and consistency, following similar considerations as for the BIP in [25, 35, 36]. We also briefly review key elements of the well-posedness theory for the Navier–Stokes equations in section 2.3, which serves as our main prototypical model, motivating the present work. In section 3, we point out the precise connection between Bayesian inversion and Bayesian DA (filtering), before stating our main results regarding the well-posedness, consistency and uniform stability of Bayesian filtering (cp section 3.2). The technical details of the mathematical derivation of these main results are collected in section 4, where we also comment on related results for hyperbolic conservation laws. Conclusions are provided in section 5. Some mathematical background is summarized in the appendix A.

2. Mathematical setting and notation

In the present section, we introduce notation that is employed throughout this work, and provide definitions for the Bayesian DA problems of interest. Besides setting the background for our main results, summarized in the subsequent section 3, we will also review some key results on the well-posedness and numerical approximation of the Navier–Stokes equations which have largely motivated the present work on the well-posedness of the corresponding Bayesian DA problem, and the convergence of approximate posteriors obtained by discretization.

Throughout this work, we follow the convention that constants C appearing in estimates may change their value from line to line. The dependency of the constant C on the given data (e.g. parameters α, β, γ) should usually be clear from the context and will be indicated by writing $C = C(\alpha, \beta, \gamma)$.

2.1. Bayesian data assimilation

DA seeks to provide an estimate for the underlying state of a system, by combining available measurements with a model of the system. The temporal evolution of the system's state can often be described by a forward solution operator $\mathcal{S}_t^\dagger : X \to X$ depending on time $t \in [0, T]$ and mapping the initial data $\bar{u} \in X$ to the solution at time t. Here, we assume X to be a Banach space, equipped with a norm $\|\cdot\|_X$. The evolution of the system starting from initial state $\bar{u}$ is thus given by $t \mapsto \mathcal{S}_t^\dagger(\bar{u})$. In view of the application to ill-posed problems, we will make essentially no assumptions on the regularity of $\mathcal{S}_t^\dagger$; in fact, unless otherwise stated, we will merely assume that:

($\mathcal{S}$.1) The solution operator defines a *Borel measurable* mapping

$$\mathcal{S}^\dagger : [0, T] \times X \to X, \quad (t, \bar{u}) \mapsto \mathcal{S}_t^\dagger(\bar{u}), \tag{$\mathcal{S}$.1}$$

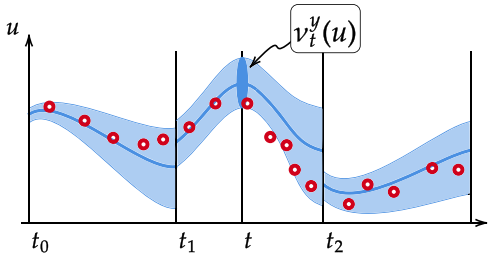

Figure 1. Schematic illustration of the DA problem: measurements (red circles) are used at times $t = t_0, t_1, \ldots,$ to periodically update the posterior measure ν_t^y (indicated by its credibility interval in blue), combining all available information from the deterministic evolution and noisy measurements.

with $t \mapsto \mathcal{S}_t(\bar{u})$ and $\bar{u} \mapsto \mathcal{S}_t(\bar{u})$ measurable for all $t \in [0, T]$, $\bar{u} \in X$.
($\mathcal{S}.2$) There exists a constant $B_{\mathcal{S}} > 0$, such that

$$\|\mathcal{S}_t^\dagger(\bar{u})\|_X \leq B_{\mathcal{S}} \|\bar{u}\|_X, \tag{$\mathcal{S}.2$}$$

for all $\bar{u}, \bar{u}' \in X$, and $t \in [0, T]$.

The Bayesian DA problem can then be stated as follows: given a prior probability measure $\mu_{\text{prior}} \in \mathcal{P}(X)$, we consider the initial state as a random variable $\bar{u} \sim \mu_{\text{prior}}$. Given a time interval $[0, T]$ and a sequence $0 = t_0 < t_1 < \cdots < t_N$, we assume that noisy measurements $y_1, \ldots, y_N$ are made, where y_j depends only on the underlying state during the time interval $[t_{j-1}, t_j]$ and is of the form:

$$y_j = \mathcal{G}_j(\mathcal{S}^\dagger(\bar{u})) + \eta_j. \tag{2.1}$$

These measurements (2.1) are defined in terms of certain measurement functionals $\mathcal{G}_j$ and random variables $\eta_j \sim \mu_{\text{noise}}$ modeling (additive) measurement noise, both of which are further specified next (figure 1).

2.1.1. Assumptions on observables. The (potentially non-linear) functionals $\mathcal{G}_j : L^1([0, T]; X) \to \mathbb{R}^d$, $j = 1, \ldots, N$, will be referred to as *observables*, and are assumed to depend only on the values $\mathcal{S}_t^\dagger(\bar{u})$ for $t \in [t_{j-1}, t_j]$. We will say that the observables $\mathcal{G}_j$ are Lipschitz continuous, if there exists $L_{\mathcal{G}} > 0$, such that

$$|\mathcal{G}_j(u) - \mathcal{G}_j(u')| \leqslant L_{\mathcal{G}} \int_{t_{j-1}}^{t_j} \|u - u'\|_X \, \mathrm{d}t, \quad \forall\, u, u' \in L^1([0, T]; X), \tag{2.2}$$

for all $j = 1, \ldots, N$. Here, we recall that the space $L^1([0, T];\ X)$ consists of all measurable mappings $u : [0, T] \to X$, such that $\int_0^T \|u(t)\|_X \, \mathrm{d}t < \infty$. We note that by ($\mathcal{S}.1$) and ($\mathcal{S}.2$), we have $t \mapsto \mathcal{S}_t^\dagger(\bar{u}) \in L^1([0, T]; X)$ for any $\bar{u} \in X$, and hence, we have a well-defined mapping $\mathcal{S}^\dagger : X \to L^1([0, T]; X)$, $\bar{u} \mapsto \mathcal{S}^\dagger(\bar{u})$. In particular, the composition $\mathcal{G}_j(\mathcal{S}^\dagger(\bar{u}))$ in (2.1) is well-defined.

Example 2.1 (Eulerian observables). If $X = L^2(D; \mathbb{R}^n)$ consists of square-integrable functions on a bounded domain $D \subset \mathbb{R}^m$, then a specific class of (Eulerian) observables are

functionals $\mathcal{G}_j : L^1([0,T];X) \to \mathbb{R}^d$ of the following form:

$$\mathcal{G}_j(u) = \int_{t_{j-1}}^{t_j} \int_D \phi(x,t) g(u(x,t)) \mathrm{d}x \, \mathrm{d}t,$$

where $\phi \in L^\infty(D \times [t_{j-1}, t_j])$, and $g : \mathbb{R}^n \to \mathbb{R}^d$ is a Lipschitz continuous function.

2.1.2. Assumptions on measurement noise. The measurement noise is modeled by random variables $\eta_1, \ldots, \eta_N \sim \mu_{\text{noise}}$ which, for simplicity, we assume to be iid and independent of $\bar{u} \sim \mu_{\text{prior}}$. Fix a symmetric, positive definite matrix $\Gamma \in \mathbb{R}^{d\times d}$, and denote by $|\cdot|_\Gamma$ the corresponding norm on $\mathbb{R}^d$ given by

$$|y|_\Gamma = \sqrt{\langle y, y\rangle_\Gamma}, \qquad \langle y, y'\rangle_\Gamma = \langle y, \Gamma^{-1} y'\rangle, \tag{2.3}$$

with $\langle \cdot, \cdot \rangle$ the standard Euclidean inner product on $\mathbb{R}^d$. We assume that the noise $\eta_j \sim \mu_{\text{noise}} = \rho(y)\mathrm{d}y$ in (2.1) possesses a distribution that is absolutely continuous with respect to Lebesgue measure $\mathrm{d}y$ on $\mathbb{R}^d$ with probability density $\rho(y) > 0$, satisfying the following assumptions:

(N.1) *Regularity:* $y \mapsto \rho(y)$ is Lipschitz continuous with respect to $|\cdot|_\Gamma$,[4], i.e. there exists $L_\rho > 0$, such that

$$|\rho(y) - \rho(y')| \leqslant L_\rho |y - y'|_\Gamma, \quad \forall\, y, y' \in \mathbb{R}^d. \tag{N.1}$$

(N.2) *Boundedness:* $y \mapsto \rho(y)$ is bounded from above, i.e. there exists $C > 0$, such that

$$\sup_{y\in\mathbb{R}^d} \rho(y) \leqslant C. \tag{N.2}$$

(N.3) *Tail-condition:* there exists a constant $C > 0$, such that

$$\rho(y) \geqslant \frac{\exp\left(-\frac{1}{2}|y|_\Gamma^2\right)}{C}, \quad \forall\, y \in \mathbb{R}^d. \tag{N.3}$$

Remark 2.2. Note that if, instead of (N.3), $\rho(y)$ satisfies a tail-condition of the form $\rho(y) \geqslant \exp(-C|y|_\Gamma^2)/C$, then upon simply rescaling $\widetilde{\Gamma} := \sqrt{2/C}\,\Gamma$, we have $\rho(y) \geqslant \exp(-\frac{1}{2}|y|_{\widetilde{\Gamma}}^2)/C$. Hence $\rho(y)$ satisfies assumptions (N.1)–(N.3) with a rescaled matrix $\Gamma \to \widetilde{\Gamma}$ in this case. Therefore, the precise constant $\frac{1}{2}$ in the tail-condition (N.3) can be assumed without loss of generality. The factor of $1/2$ turns out to be particularly convenient.

Assumptions (N.1)–(N.3) are clearly fulfilled for normally distributed measurement noise $\eta_j \sim \mathcal{N}(0,\Gamma)$. This is the main application we have in mind. However, it is worth pointing out that the assumption is satisfied for a much wider class of measurement noise: in particular, since the tail-condition requires only a lower bound, our results apply to situations in which one encounters noise *with a heavy tail*.

[4] Although all norms on the finite-dimensional space $\mathbb{R}^d$ are equivalent, measurement noise such as Gaussian noise is naturally associated with the norm $|\cdot|_\Gamma$ induced by the covariance matrix Γ.

2.1.3. Posterior/conditional probability. Given a time $t \geqslant 0$ and a given subset of measurements $y_{1:k} = (y_1, \dots, y_k)$, we are interested in the conditional probability

$$\nu_t^{y_{1:k}}(\,\cdot\,) := \mathbb{P}\Big[\mathcal{S}_t^\dagger(\bar{u}) \in \,\cdot\, \Big|\, \mathcal{G}_j(\mathcal{S}^\dagger(\bar{u})) + \eta_j = y_j,\ \forall\ j = 1, \dots, k\Big], \tag{2.4}$$

providing a Bayesian estimate of the underlying state $u(t) = \mathcal{S}_t^\dagger(\bar{u})$ at time $t \geqslant 0$ given the prior distribution $\mu_{\text{prior}}(\mathrm{d}\bar{u})$ at $t = 0$ and the measurements $y_1, \dots, y_k$. If all available measurements at past times $t_j \leqslant t$ are taken into account, this estimate is referred to as the *filtering distribution*; if the estimate also takes into account measurements obtained at times $t_j \geqslant t$, i.e. the state is estimated in hindsight, the posterior is referred to as the *smoothing distribution*. In the case of filtering, the set of available measurements $y_{1:k}$ will itself vary with time t, i.e. $k = k(t)$. For concreteness, we will mostly focus on the filtering problem in the following; given all measurements $\boldsymbol{y} = (y_1, \dots, y_N)$ over a time interval $[0, T]$, the filtering distribution

$$\nu_t^{\boldsymbol{y}}(\,\cdot\,) = \mathbb{P}[\mathcal{S}_t^\dagger(\bar{u}) \in \,\cdot\, | y_j \text{ with } t_j \leqslant t], \tag{2.5}$$

provides the best-estimate at time t given *only the past measurements*; the filtering distribution can be written in terms of the conditional probabilities (2.4):

$$\nu_t^{\boldsymbol{y}} := \begin{cases} \nu_t^{\emptyset}, & t \in [0, t_1), \\ \nu_t^{y_1}, & t \in [t_1, t_2), \\ \vdots & \\ \nu^{y_{1:(N-1)}}, & t \in [t_{N-1}, t_N), \\ \nu^{y_{1:N}}, & t \geqslant t_N. \end{cases} \tag{2.6}$$

Here we have formally defined $\nu_t^{\emptyset} := \mathcal{S}_{t,\#}^\dagger \mu_{\text{prior}}$, corresponding to the best prediction in the absence of any measurements. The filtering distribution (2.6) thus defines a mapping from measurements $\boldsymbol{y} = (y_1, \dots, y_N) \in \mathbb{R}^{d\times N}$ to time-parametrized probability measures $\nu_t^{\boldsymbol{y}}$.

2.1.4. Definition of well-posedness. We are interested in the well-posedness of the filtering problem, as defined next:

Definition 2.3 (well-posedness of Bayesian-DA). Given a forward operator $\mathcal{S}_t^\dagger$, a prior μ_{prior}, a noise distribution μ_{noise}, and a space $(\boldsymbol{\Sigma}, d_T)$ of time-parametrized probability measures $t \mapsto \nu_t$, we say that the Bayesian-DA problem is **well-posed**, provided that the following properties are satisfied:

(a) *Existence:* for any $\boldsymbol{y} \in \mathbb{R}^{d\times N}$, the posterior filtering distribution (2.4) exists in $\boldsymbol{\Sigma}$,
(b) *Uniqueness:* the filtering distribution is unique,
(c) *Stability:* the measurement-to-posterior mapping

$$\mathbb{R}^{d\times N} \to \boldsymbol{\Sigma}, \qquad \boldsymbol{y} \mapsto \nu_t^{\boldsymbol{y}},$$

is locally Lipschitz continuous wrt. d_T, i.e. for any $R > 0$, there exists $C(R) > 0$, such that

$$d_T(\nu_t^{\boldsymbol{y}}, \nu_t^{\boldsymbol{y}'}) \leqslant C|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma,$$

for all $\boldsymbol{y}, \boldsymbol{y}' \in \mathbb{R}^{d\times N}$, such that $|\boldsymbol{y}|_\Gamma, |\boldsymbol{y}'|_\Gamma \leqslant R$, and where we define $|\boldsymbol{y}|_\Gamma := \sqrt{\sum_{j=1}^N |y_j|_\Gamma^2}$.

One possible choice for the metric space $(\mathbf{\Sigma}, d_T)$ will be discussed below (cp section 2.2). At this point, we would like to point out that the above definition is a direct analogue of the corresponding definition of well-posedness for the BIP [25, definitions 2.7 and 3.4], as well as the notion of well-posedness for the forward problem:

Remark 2.4 (well-posedness of the forward problem). If $\mathcal{S}_t^\dagger : X \to X$ is the solution operator associated with a time-evolution PDE on a time interval $[0, T]$, then the well-posedness of the forward problem is usually defined as the existence, uniqueness and stability of $\mathcal{S}_t^\dagger$ on X, where the stability requires $\mathcal{S}_t^\dagger$ to be *continuous* as a mapping $\bar{u} \to \mathcal{S}_t^\dagger(\bar{u})$ for any $t \in [0, T]$. In fact, $\mathcal{S}_t^\dagger$ is often required to be *Lipschitz continuous*, i.e. there exists a constant $L_{\mathcal{S}} \geqslant 0$, such that

$$\|\mathcal{S}_t^\dagger(\bar{u}) - \mathcal{S}_t^\dagger(\bar{u}')\|_X \leqslant L_{\mathcal{S}} \|\bar{u} - \bar{u}'\|_X,$$

for any $\bar{u}, \bar{u}' \in X$. Thus, the well-posedness of the forward problem is reflected in the regularity of $\mathcal{S}_t^\dagger$. In the present work, we will study the well-posedness of the associated Bayesian filtering problem (2.5) *in the absence of such regularity*, thus formally allowing for $L_{\mathcal{S}} = \infty$.

2.1.5. Numerical discretization and consistency. As the true forward operator $\mathcal{S}_t^\dagger$ is usually not computable in practice, one often needs to replace $\mathcal{S}_t^\dagger$ by a numerical approximation $\mathcal{S}_t^\Delta \approx \mathcal{S}_t^\dagger$, depending on a parameter $\Delta > 0$. The parameter $\Delta > 0$ may reflect the grid size in a numerical discretization, or may represent more general modeling errors; in the following, we will usually refer to Δ as the 'grid size', and will focus on errors due to discretization of a given PDE. Upon discretization, the exact posterior (2.4) is replaced by the following conditional probability:

$$\nu_t^{\Delta, y_{1:k}}(\,\cdot\,) := \mathbb{P}\big[\mathcal{S}_t^\Delta(\bar{u}) \in \,\cdot\, \big|\, \mathcal{G}_j(\mathcal{S}^\Delta(\bar{u})) + \eta_j = y_j,\ \forall\ j = 1, \ldots, k\big]. \tag{2.7}$$

The corresponding filtering distribution $\nu_t^{\Delta, \boldsymbol{y}}$ is defined as in (2.6), but with $\nu_t^{\Delta, y_{1:k}}$ replacing $\nu_t^{y_{1:k}}$. Given such a discretization, a fundamental question concerns the *consistency* of the approximate posteriors $\nu_t^{\Delta, \boldsymbol{y}}$ with the limiting posterior $\nu_t^{\boldsymbol{y}}$:

Definition 2.5 (consistency). The sequence of approximate posteriors $\nu_t^{\Delta, \boldsymbol{y}}$ is *consistent* with the limiting posterior $\nu_t^{\boldsymbol{y}}$, with respect to a space of time-parametrized probability measures $(\mathbf{\Sigma}, d_T)$, if $\nu_t^{\Delta, \boldsymbol{y}} \to \nu_t^{\boldsymbol{y}}$ converges locally uniformly in $\boldsymbol{y} = (y_1, \ldots, y_N) \in \mathbb{R}^{d \times N}$; i.e., if for any $R > 0$, we have

$$\lim_{\Delta \to 0} \sup_{|\boldsymbol{y}|_\Gamma \leqslant R} d_T(\nu_t^{\Delta, \boldsymbol{y}}, \nu_t^{\boldsymbol{y}}) = 0.$$

2.2. Time-parametrized probability measures

Given our definition of well-posedness and consistency of the Bayesian DA problem, and the solution (2.6) of the filtering problem, we need to define a suitable space $(\mathbf{\Sigma}, d_T)$ of time-parametrized probability measures, $t \mapsto \nu_t$. To this end, we follow [24], and introduce the following space $L_t^1(\mathcal{P}) = L_t^1([0, T]; \mathcal{P}_1(X))$:

Definition 2.6. Let X be a separable Banach space with norm $\|\cdot\|_X$, and let $\mathcal{P}_1(X)$ denote the set of Borel probability measures μ on X, with finite first-moment $\int_X \|u\|_X \, \mu(\mathrm{d}u) < \infty$. We recall that $\mathcal{P}_1(X)$ is metrized by the one-Wasserstein metric $W_1(\cdot,\cdot)$ (cp appendix A.2 for definitions). Given a time $T > 0$, we define $L_t^1([0,T];\mathcal{P}_1(X))$ to be the set of weak-$*$ measurable[5] mappings $[0,T] \to \mathcal{P}_1(X)$, $t \mapsto \nu_t$, such that

$$\int_0^T \int_X \|u\|_X \, \nu_t(\mathrm{d}u)\mathrm{d}t < \infty,$$

and we introduce the following metric on $L_t^1([0;T];\mathcal{P}_1(X))$:

$$d_T(\nu_t, \nu_t') \coloneqq \int_0^T W_1(\nu_t, \nu_t')\mathrm{d}t, \quad \forall\ \nu_t, \nu_t' \in L_t^1([0;T];\mathcal{P}_1(X)).$$

We shall usually employ the simpler notation $L_t^1(\mathcal{P}) = L_t^1([0,T];\mathcal{P}_1(X))$, when the temporal domain $[0,T]$ and the underlying Banach space X are clear from the context. Following [24, proposition 2.1], we also recall

Proposition 2.7. *Let X be a separable Banach space. Then $L_t^1(\mathcal{P}) = L_t^1([0,T];\mathcal{P}_1(X))$ is a complete metric space under the norm $d_T(\,\cdot\,,\,\cdot\,) = \int_0^T W_1(\,\cdot\,,\,\cdot\,)\,\mathrm{d}t$.*

The motivation for considering this particular metric on time-parametrized probability measures is two-fold: firstly, this metric and closely related quantities have been shown to be relevant empirically as well as analytically for the convergence of numerical approximations to so-called 'statistical solutions', for several fundamental equations of fluid dynamics including the incompressible Navier–Stokes [2], incompressible Euler [23, 24] as well as the compressible Euler equations [10]. Secondly, metrics other than the Wasserstein W_1-metric, such as the Hellinger and total variation distances or the Kullback–Leibler divergence, which have been considered in the context of BIPs [25, 35], may be less suitable in the context of Bayesian DA, since these latter distances require absolute continuity of the involved measures. While this requirement of absolute continuity is often not an issue for BIPs [25, 35, 36], the filtering distributions ν_t^y, $\nu_t^{\Delta,y}$ considered in the present work are generally singular with respect to each other (due to the additional prediction step). Hence, we focus on distances which allow for *disjoint supports* of the underlying measures, such as the Wasserstein distance.

2.3. Navier–Stokes equations

To illustrate ill-posed forward problems arising in fluid mechanics, we next review some elements of the stability theory for the incompressible Navier–Stokes equations. The Navier–Stokes equations are here viewed as a prototypical model of fluid flows, given by the following system of PDEs:

$$\begin{cases} \partial_t u + \mathrm{div}(u \otimes u) + \nabla p = \nu \Delta u, \\ \quad \mathrm{div}(u) = 0, \quad u(\,\cdot\,,0) = \overline{u}. \end{cases} \tag{2.8}$$

These equations describe the evolution of the flow vector field $u : D \times [0,T] \to \mathbb{R}^n$, $u = (u_1, \ldots, u_n)$, of a fluid in n dimensions. The parameter $\nu > 0$ denotes the viscosity of

[5] For any $\Phi \in C_b(X)$, the mapping $t \mapsto \int_X \Phi(u)\nu_t(\mathrm{d}u)$ is measurable.

the fluid. The divergence term $\mathrm{div}(u \otimes u)$ has components $[\mathrm{div}(u \times u)]_i = \sum_{j=1}^n \partial_j (u_j u_i) (i = 1, \ldots, n)$, $\nabla p = (\partial_1 p, \ldots, \partial_n p)$ is the gradient of the pressure p, and $\Delta u = \sum_{j=1}^n \partial_j^2 u$ on the right-hand side denotes the Laplacian applied to u. For simplicity we shall focus on the case of periodic boundary conditions.

Remark 2.8 (setting for Navier–Stokes equations). For $n \in \{2, 3\}$, we consider initial data $\bar{u} \in L^2_x := L^2(\mathbb{T}^n; \mathbb{R}^n)$, consisting of 2π-periodic L^2-integrable vector fields defined on the periodic torus $\mathbb{T}^n \simeq [0, 2\pi]^n$ in n dimensions. In addition, any initial data is required to be divergence-free, $\mathrm{div}(\bar{u}) = 0$. For such initial data, we seek weak solutions $u \in L^\infty([0, T]; L^2_x)$ of (2.8). We note that physically, the quantity $\frac{1}{2}\|u(t)\|^2_{L^2_x}$ corresponds to the kinetic energy of the underlying fluid, and hence the requirement that $\operatorname{ess\,sup}_{t\in[0,T]} \|u(t)\|_{L^2_x} < \infty$ is natural.

2.3.1. Theoretical results. It is well-known that for very small values of the viscosity $\nu \ll 1$, solutions of the Navier–Stokes equations can exhibit turbulent behaviour, characterized by a high sensitivity to perturbations to the initial data and involving dynamics across a wide range of spatial and temporal scales [12, 32]. This empirically observed turbulent behaviour at small viscosity is mathematically reflected by a strong ν-dependence in the available *a priori* stability results for the solution operator $\mathcal{S}_t^\dagger$ of the system (2.8). This is summarized in the following well-known theorem for the two-dimensional case:

Theorem 2.9 (stability of N–S in 2d, see [21, p 170, chap. 6, theorem 11]). *Let $\bar{u} \in L^2_x$ be initial data for the incompressible Navier–Stokes equations (2.8) for $n = 2$. There exists a unique solution $u(t) = \mathcal{S}_t^\dagger(\bar{u})$ of the Navier–Stokes equations with initial data $\bar{u}$. This solution satisfies the energy bound $\|u(t)\|_{L^2_x} = \|\mathcal{S}_t^\dagger(\bar{u})\|_{L^2_x} \leqslant \|\bar{u}\|_{L^2_x}$. Furthermore, for any initial data $\bar{u}, \bar{u}' \in L^2_x$, we have*

$$\|\mathcal{S}_t^\dagger(\bar{u}) - \mathcal{S}_t^\dagger(\bar{u}')\|_{L^2_x} \leqslant \|\bar{u} - \bar{u}'\|_{L^2_x} \exp\left(\frac{2}{\nu}\int_0^t \|\nabla u(\tau)\|^2_{L^2} \,\mathrm{d}\tau\right).$$

Thus, even if the solution $u(t) = \mathcal{S}_t^\dagger(\bar{u})$ is assumed to be Lipschitz continuous, theorem 2.9 provides at best a stability estimate of the form $\|\mathcal{S}_t^\dagger(\bar{u}) - \mathcal{S}_t^\dagger(\bar{u}')\|_{L^2_x} \leqslant \mathrm{e}^{Ct/\nu}\|\bar{u} - \bar{u}'\|_{L^2_x}$, which exhibits an *exponential dependence on* $1/\nu$.

In the three-dimensional case, a global existence and stability result such as theorem 2.9 remains unknown; it is well-known that solutions exist *locally* in time and when starting from sufficiently regular initial data. For general initial data in L^2_x, it has been shown in the celebrated work of Leray [28] that energy admissible solutions satisfying $\|u(t)\|_{L^2_x} \leqslant \|\bar{u}\|_{L^2_x}$ exist, but their uniqueness remains an open problem; in particular, *there are no guarantees on the stability of a solution operator $\mathcal{S}_t^\dagger : L^2_x \to L^2_x$ for the three-dimensional Navier–Stokes equations.*

2.3.2. Numerical discretization. Popular numerical discretizations of the forward problem for the incompressible Navier–Stokes equations, especially on periodic domains, are spectral methods [4, 6, 13, 16, 19]. Spectral methods are based on the following ansatz $u^\Delta(x, t) = \sum_{|k|_\infty \leqslant N} \widehat{u}_k^\Delta(t) \mathrm{e}^{\mathrm{i}k\cdot x}$, where now and in the following we shall consistently denote $\Delta = 1/N$, and $|k|_\infty := \max_{i=1,\ldots,d}|k_i|$. A straight-forward spectral approximation of the Navier–Stokes equations is based on a Galerkin projection onto this ansatz space:

$$\begin{cases} \partial_t u^\Delta + \mathcal{P}_N \,\mathrm{div}(u^\Delta \otimes u^\Delta) + \nabla p^\Delta = \nu \Delta u^\Delta, \\ \qquad\qquad \mathrm{div}(u^\Delta) = 0, \quad u^\Delta|_{t=0} = \mathcal{P}_N \bar{u}. \end{cases} \tag{2.9}$$

Here $\mathcal{P}_N$ is the spatial Fourier projection operator, mapping an arbitrary function $f(x,t)$ onto the first N Fourier modes: $\mathcal{P}_N f(x,t) = \sum_{|k|_\infty \leqslant N} \widehat{f}_k(t) \mathrm{e}^{\mathrm{i}k\cdot x}$. The scheme (2.9) can be equivalently written in terms of a system of ordinary differential equaitons (ODEs) for the Fourier coefficients $\widehat{u}_k(t)$, $|k|_\infty \leqslant N$. Hence, combined with a suitable (e.g. Runge–Kutta) time-stepping, (2.9) provides a numerical discretization of the Navier–Stokes equations. In the following proposition, we summarize some basic observations on the numerical approximations (2.9):

Proposition 2.10. *For any $\Delta = 1/N > 0$, let $\mathcal{S}_t^\Delta : L^2_x \to L^2_x$ denote the solution operator associated with the numerical scheme (2.9). Then for initial data $\bar{u} \in L^2_x$, the numerical solution $u^\Delta(t) = \mathcal{S}_t^\Delta(\bar{u})$ satisfies:*

(Δ.1) *Energy bound:* $\|\mathcal{S}_t^\Delta(\bar{u})\|_{L^2_x} \leqslant \|\bar{u}\|_{L^2_x}$,
(Δ.2) *Coercivity:* $\int_0^T \|\nabla u^\Delta(t)\|^2_{L^2_x} \,\mathrm{d}t \leqslant \nu^{-1}\|\bar{u}\|^2_{L^2_x}$,
(Δ.3) *Weak time-regularity: there exist constants $C = C(\|\bar{u}\|_{L^2_x}), L > 0$, such that $\|u^\Delta(t) - u^\Delta(s)\|_{H^{-L}_x} \leqslant C|t-s|$. In particular, we have $u^\Delta \in \mathrm{Lip}([0,T]; H^{-L}_x)$ uniformly in $\Delta > 0$.*

The basic properties summarized above will form the basis for the well-posedness results of the present work. Additional control on the numerical approximations for the forward problem, especially for the 3D Navier–Stokes equations and/or rough initial data, can be achieved by refining the scheme (2.9) through the addition of suitable numerical diffusion; this can provide additional control on the small scale behaviour, resulting e.g. in spectral viscosity schemes as proposed in [16, 22, 37, 38]. The basic properties of proposition 2.10 will, however, suffice for the purposes of the present work.

3. Main results

As pointed out in the last section, for many PDEs encountered in the context of fluid dynamics (such as the Navier–Stokes equations in 3D), the current mathematical understanding does not allow to rigorously *prove* the existence, uniqueness and stability of the corresponding forward problem, i.e. a unique forward solution operator $\mathcal{S}_t^\dagger : X \to X$ is not known to exist, and even if it exists there may be *no stability* in the sense that we could have $\mathrm{Lip}(\mathcal{S}_t^\dagger : X \to X) = \infty$. Furthermore, even in those special cases, where the solution operator $\mathcal{S}_t^\dagger$ can be shown to exist, stability estimates may exhibit a very unfavourable (exponential) dependence on small physical parameters $\nu \ll 1$, e.g. yielding $\mathrm{Lip}(\mathcal{S}_t^\dagger) \sim \mathrm{e}^{Ct/\nu}$ as for the Navier–Stokes equations in 2D. Such an exponential dependence on $1/\nu$ renders the forward evolution *effectively ill-posed* for small values of $\nu \ll 1$.

In view of these observations, in this section, we will summarize our results on

- the **well-posedness of Bayesian DA** in the sense of definition 2.3, even when the forward problem may be ill-posed,
- the **consistency of approximate posteriors** $\nu_t^{\Delta,y}$ with the limiting distribution ν_t^y in the sense of definition 2.5, including **convergence rates**, when a unique solution operator $\mathcal{S}_t^\dagger$ of the limiting problem exists,
- **compactness** and **(uniform) stability properties** of the approximate filtering distributions (2.7), in the limit $\Delta \to 0$, even when no rigorous convergence guarantees $\mathcal{S}_t^\Delta \xrightarrow{??} \mathcal{S}_t^\dagger$ to a unique limiting forward solution operator are known.

In section 3.1, we first start with several remarks on the connection between inverse problems, as considered in [25, 35, 36], and the Bayesian DA (filtering) problem of the present work. We will also indicate the main mathematical difficulty encountered when considering *ill-posed* forward problems, for which $\mathrm{Lip}(\mathcal{S}_t^\dagger) = \infty$ or $\mathrm{Lip}(\mathcal{S}_t^\dagger) \gg 1$ (cp proposition 3.4, below).

3.1. Connection with Bayesian inverse problems

Closely related to the Bayesian DA problem is the corresponding BIP, which can be used to determine the following posterior probability on the *initial data* $\bar{u}$ at time $t = 0$, for $\mathcal{L}_j^\dagger(\bar{u}) := \mathcal{G}_j(\mathcal{S}^\dagger(\bar{u}))$:

$$\mu^{y_{1:k}}(\,\cdot\,) = \mathbb{P}\left[\bar{u} \in \,\cdot\, \middle| \mathcal{L}_j^\dagger(\bar{u}) + \eta_j = y_j,\ \forall\ j = 1, \ldots, k\right], \tag{3.1}$$

or, upon discretization with $\mathcal{L}_j^\Delta(\bar{u}) := \mathcal{G}_j(\mathcal{S}^\Delta(\bar{u}))$,

$$\mu^{\Delta, y_{1:k}}(\,\cdot\,) = \mathbb{P}\left[\bar{u} \in \,\cdot\, \middle| \mathcal{L}_j^\Delta(\bar{u}) + \eta_j = y_j,\ \forall\ j = 1, \ldots, k\right]. \tag{3.2}$$

A detailed analysis of the posterior (3.1) and (3.2) has been provided for infinite-dimensional problems in [36]. An extended discussion of the well-posedness of the BIP under minimal assumptions on the forward problem has been given in [25, 35]; as follows from [25, theorem 2.5], under the present assumptions on the solution operator ($\mathcal{S}$.1) and ($\mathcal{S}$.2) and the (strictly positive) noise distribution (N.1)–(N.3), the solution $\mu^{y_{1:k}}$ of the BIP exists and is explicitly given by

$$\mu^{y_{1:k}}(\mathrm{d}\bar{u}) = \frac{1}{\mathcal{Z}_k^\dagger(y_{1:k})} \exp\left(-\sum_{j=1}^k \Phi_j^{\dagger, y_j}(\bar{u})\right) \mu_{\mathrm{prior}}(\mathrm{d}\bar{u}), \tag{3.3}$$

where

$$\Phi_j^{\dagger, y_j}(\bar{u}) := -\log\, \rho(\mathcal{L}_j^\dagger(\bar{u}) - y_j), \tag{3.4}$$

denotes the log-likelihood function, and

$$\mathcal{Z}_k^\dagger(y_{1:k}) = \int_X \exp\left(-\sum_{j=1}^k \Phi_j^{\dagger, y_j}(\bar{u})\right) \mu_{\mathrm{prior}}(\mathrm{d}\bar{u}), \tag{3.5}$$

is the required normalization constant, which depends on $y_{1:k} = (y_1, \ldots, y_k)$. We note that the condition that $\rho(y) > 0$ implies that the log-likelihood $\Phi^{\dagger, y_{1:k}}$ is finite, i.e., $\Phi^{\dagger, y_{1:k}}(\bar{u}) < \infty$ for all $\bar{u} \in X$.

Remark 3.1 (Gaussian noise). If the noise $\eta \sim \mathcal{N}(0, \Gamma)$ is normally distributed (Gaussian), then (up to an unimportant additive constant)

$$\Phi^{\dagger, y_j}(\bar{u}) = \frac{1}{2} |y_j - \mathcal{L}_j^\dagger(\bar{u})|_\Gamma^2,$$

where the natural Γ-norm is given by (2.3). In this case, we have

$$\frac{\mathrm{d}\mu^{y_{1:k}}}{\mathrm{d}\mu_{\text{prior}}}(\bar{u}) = \frac{1}{\mathcal{Z}_k^\dagger(y_{1:k})} \exp\left(-\frac{1}{2}\sum_{j=1}^{k}\left|y_j - \mathcal{L}_j^\dagger(\bar{u})\right|_\Gamma^2\right). \tag{3.6}$$

3.1.1. Push-forward of BIP and stability. In the next simple proposition, we note an explicit expression for the distribution (2.4) in terms of the solution of the corresponding BIP (3.1):

Proposition 3.2. *Let $\mathcal{S}_t^\dagger : X \to X$ be a Borel measurable mapping. Let $\nu_t^{y_{1:k}}$, $\mu^{y_{1:k}}$ be given by (2.4) and (3.1), respectively, for observables $\mathcal{G}_j$ and with measurement noise satisfying (N.1)–(N.3). Then*

$$\nu_t^{y_{1:k}} = \mathcal{S}_{t,\#}^\dagger \mu^{y_{1:k}}. \tag{3.7}$$

Proof. Let $A \subset X$ be a Borel measurable set. We have

$$\begin{aligned}
\nu_t^{y_{1:k}}(A) &= \mathbb{P}\left[\mathcal{S}_t^\dagger(\bar{u}) \in A | y_1, \ldots, y_k\right] \\
&= \mathbb{P}\left[\bar{u} \in [\mathcal{S}_t^\dagger]^{-1}(A) | y_1, \ldots, y_k\right] \\
&= \mu^{y_{1:k}}([\mathcal{S}_t^\dagger]^{-1}(A)) \\
&= \mathcal{S}_{t,\#}^\dagger \mu^{y_{1:k}}(A).
\end{aligned}$$

As A was arbitrary, it follows that $\nu_t^{y_{1:k}} = \mathcal{S}_{t,\#}^\dagger \mu^{y_{1:k}}$. □

Proposition 3.2 specifies the relation between the Bayesian-DA problem and the corresponding BIP, via the push-forward under the solution operator.

Remark 3.3. Based on proposition 3.2, we also remark that the posteriors $\nu_t^{y_{1:k}}$ are indeed elements of $L_t^1(\mathcal{P})$, under the assumptions of the present work: if $\mathcal{S}_t^\dagger$ satisfies ($\mathcal{S}$.1) and ($\mathcal{S}$.2), then the push-forward $\mu \mapsto \mathcal{S}_{t,\#}^\dagger \mu$ defines a well-defined map $\mathcal{P}_1(X) \to L_t^1(\mathcal{P}_1)$, and hence $\mathcal{P}[\mathcal{S}_t^\dagger \in \cdot \,|\, y_1, \ldots, y_k] = \mathcal{S}_{t,\#}^\dagger \mu^{y_{1:k}} \in L_t^1(\mathcal{P})$, if $\mu_{\text{prior}} \in \mathcal{P}_1(X)$.

Furthermore, we note that proposition 3.2 immediately yields the following representation of the filtering distribution (2.6) in terms of the push-forward $\mathcal{S}_{t,\#}^\dagger$, and the solutions of the BIP for the initial data $\mu^{y_{1:k}}$ in (3.3):

$$\nu_t^{\boldsymbol{y}} = \begin{cases} \mathcal{S}_{t,\#}^\dagger \mu^{y_{1:(j-1)}}, & t \in [t_{j-1}, t_j),\ j = 1, \ldots, N \\ \mathcal{S}_{t,\#}^\dagger \mu^{y_{1:N}}, & t \geqslant t_N. \end{cases} \tag{3.8}$$

At this point, we recall that the well-posedness of the BIP has been studied under very mild conditions on the forward operators and in a variety of metrics between probability measures in [25, 35], including the Wasserstein distance. It is therefore natural to ask whether the results of [25, 35] can be used to obtain corresponding results also for the Bayesian DA problem, based on their relationship (3.7) via the push-forward under $\mathcal{S}_t^\dagger$? The following proposition indicates that, in general, bounds on the Wasserstein distance for the BIP (as obtained in e.g. [35]) do

not automatically translate to corresponding bounds for Bayesian DA, if the solution operator is not sufficiently regular. For the straight-forward proof, we refer to appendix A.2, page 37.

Proposition 3.4. *Let $L_{\mathcal{S}} := \mathrm{Lip}(\mathcal{S}_t^\dagger : X \to X)$ denote the Lipschitz constant of the forward operator. Then*

$$W_1\left(\mathcal{S}_{t,\#}^\dagger \mu, \mathcal{S}_{t,\#}^\dagger \mu'\right) \leqslant L_{\mathcal{S}} W_1(\mu, \mu'), \quad \forall\ \mu, \mu' \in \mathcal{P}_1(X), \tag{3.9}$$

and $L_{\mathcal{S}}$ is optimal: if $L > 0$ is any other constant such that $W_1\left(\mathcal{S}_{t,\#}^\dagger \mu, \mathcal{S}_{t,\#}^\dagger \mu'\right) \leqslant L W_1(\mu, \mu')$ for all $\mu, \mu' \in \mathcal{P}_1(X)$, then $L \geqslant L_{\mathcal{S}}$.

As shown in proposition 3.4, the Wasserstein distance between two probability measures is amplified by the Lipschitz constant of the forward solution operator (for general measures $\mu,\ \mu'$). For the ill-posed problems considered in the present work, this Lipschitz constant is either very large or even infinite, $L_{\mathcal{S}} = \infty$, and hence W_1-estimates on the BIP—even under the minimal assumptions of [25, 35]—do not suffice to conclude similarly robust well-posedness results for Bayesian filtering problem. In view of applications to such ill-posed problems, it would be highly desirable to obtain estimates which are *independent of the stability of the forward problem*. This is one of the main goals of the present work. Overcoming the seemingly fundamental limitations imposed by proposition 3.4 will rely on the specific form of the posteriors $\nu_t^{\boldsymbol{y}}, \nu_t^{\boldsymbol{y}'} \in \mathcal{P}_1(X)$, which are *not* arbitrary probability measures but are instead constrained by a common prior and potentially different measurements $\boldsymbol{y} \neq \boldsymbol{y}'$.

3.1.2. Alternative representation of the filtering distribution. We finally point out a different, recursive formulation of the Bayesian DA problem, which is closer in spirit to filtering schemes such as the ensemble Kalman filter [8] or 3DVAR [5].

Remark 3.5 (recursive filtering). Fix a prior measure μ_{prior} at the initial time, and define

$$\widetilde{\nu}_{t_0}^{\dagger,y_0} := \mathcal{S}_{0,\#}^\dagger \mu_{\text{prior}}. \tag{3.10}$$

Given times $0 = t_0 < t_1 < \cdots < t_N = T$ and measurements $y_1, \ldots, y_N$, carry out the following two recursive steps.

(a) **Correction step:** given $\widetilde{\nu}_{t_{j-1}}^{y_{1:(j-1)}}$ as a prior at time t_{j-1}, solve the BIP with new measurement $y_j = \mathcal{G}_j(\mathcal{S}_{t-t_{j-1}}^\dagger(u)) + \eta_j$, for $t \in [t_{j-1}, t_j]$, to obtain a corrected Bayesian estimate

$$\widetilde{\nu}_{t_{j-1}}^{y_{1:j}}(\mathrm{d}u) = \frac{1}{Z_j^\dagger(y_j)} \exp\left(-\widetilde{\Phi}_j^{\dagger,y_j}(u)\right) \widetilde{\nu}_{t_{j-1}}^{y_{1:(j-1)}}(\mathrm{d}u), \tag{3.11}$$

where $\widetilde{\Phi}_j^{\dagger,y_j}(u) := -\log\left(\mathcal{G}_j(\mathcal{S}_{t-t_{j-1}}^\dagger(u)) - y_j\right)$.

(b) **Prediction step:** based on this corrected estimate, predict the probability distribution at time t_j, as the push-forward:

$$\widetilde{\nu}_{t_j}^{y_{1:j}} = \mathcal{S}_{\delta t_j,\#}^\dagger \widetilde{\nu}_{t_{j-1}}^{y_{1:j}}, \tag{3.12}$$

where $\delta t_j = t_j - t_{j-1}$.

We note that,

$$\Phi_j^{\dagger,y_j}(\bar{u}) = \widetilde{\Phi}_j^{\dagger,y_j}(\mathcal{S}_{t_{j-1}}^{\dagger}(\bar{u})), \tag{3.13}$$

by definition (3.4) of $\Phi_j^{\dagger,y_j}$.

We next observe that this recursive formulation is indeed equivalent to (2.4) (see section 4.1, p 19 for a proof):

Proposition 3.6. *Assume that the solution operator satisfies $\mathcal{S}_s^{\dagger} \circ \mathcal{S}_t^{\dagger} = \mathcal{S}_{s+t}^{\dagger}$ for all $s, t \geqslant 0$, in addition to ($\mathcal{S}$.1) and ($\mathcal{S}$.2). Then the sequence $\widetilde{\nu}_{t_j}^{y_{1:j}}$ obtained by the recursive prediction-correction procedure of remark 3.5 agrees with the filtering distribution (2.5), i.e. we have $\widetilde{\nu}_{t_j}^{y_{1:j}} = \nu_{t_j}^{\boldsymbol{y}}$, for all $j = 1, \ldots, N$.*

In the present section, we have discussed the precise relation between the BIP and the Bayesian DA problem, showing that the DA problem is a combination of a suitably formulated BIP for the initial data, followed by a prediction step. We finally point out that BIPs can be thought of as a special instance of the Bayesian DA (with trivial forward solution operator). This will allow us to translate certain results on DA to the context of BIPs.

Remark 3.7 (BIP as a special case of Bayesian DA). Set $\mathcal{S}_t^{\dagger}(\bar{u}) := \bar{u}$ for all $t \in [0, T]$, and assume that all measurements are obtained at $t = 0$. Then

$$\mu^{y_{1:N}}(\mathrm{d}u) = \nu_t^{y_{1:N}}(\mathrm{d}u) = \nu_t^{\boldsymbol{y}}(\mathrm{d}u),$$

for all $t \in [0, T]$. Furthermore, we have for the discretized posterior

$$W_1(\mu^{y_{1:N}}, \mu^{\Delta,y_{1:N}}) = \frac{1}{T}\, d_T\left(\nu_t^{\boldsymbol{y}}, \nu_t^{\Delta,\boldsymbol{y}}\right),$$
$$W_1(\mu^{y_{1:N}}, \mu^{y'_{1:N}}) = \frac{1}{T}\, d_T\left(\nu_t^{\boldsymbol{y}}, \nu_t^{\boldsymbol{y}'}\right).$$

Hence, all results regarding the well-posedness, stability and consistency obtained for the Bayesian filtering setting in the present work, should readily imply corresponding results for the BIP setting, under the Wasserstein W_1-distance.

3.2. Well-posedness results for Bayesian DA

3.2.1. General well-posedness result. We can now state the following general well-posedness result for the Bayesian filtering problem, which shows that the filtering problem is well-posed under very mild boundedness assumptions, even if the corresponding forward problem is *ill-posed*. Before stating our result, we recall that (cp notation defined in appendix A.1), $\|\bar{u}\|_{L^1(\mu_{\text{prior}})} := \int_X \|\bar{u}\|_X \,\mathrm{d}\mu_{\text{prior}}(\bar{u})$. We then have:

Theorem 3.8 (filtering well-posedness). *Let $\mathcal{S}_t^{\dagger} : [0, T] \times X \to X$, $(t, \bar{u}) \mapsto \mathcal{S}_t^{\dagger}(\bar{u})$ be a Borel measurable solution operator, such that $\|\mathcal{S}_t^{\dagger}(\bar{u})\|_X \leqslant B_{\mathcal{S}} \|\bar{u}\|_X$ for all $t \in [0, T]$. Let $\mu_{\text{prior}} \in \mathcal{P}_1(X)$ be a prior with finite first moment. Then the Bayesian filtering problem is well-posed: more precisely, the conditional probability $\nu_t^{\boldsymbol{y}}$ in (2.6) exists for any measurements*

$\boldsymbol{y} = (y_1, \dots, y_N) \in \mathbb{R}^{d\times N}$, $\nu_t^{\boldsymbol{y}}$ belongs to $L_t^1(\mathcal{P})$, and furthermore $\boldsymbol{y} \mapsto \nu_t^{\boldsymbol{y}}$ is stable, in the sense that for any $R > 0$, there exists $C = C(R, \rho, N, B_{\mathcal{S}}, \|\bar{u}\|_{L^1(\mu_{\text{prior}})}, T) > 0$, such that

$$W_1\left(\nu_t^{\boldsymbol{y}}, \nu_t^{\boldsymbol{y}'}\right) \leqslant C|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma, \quad \forall\, t \in [0, T], \tag{3.14}$$

and

$$\int_0^T W_1\left(\nu_t^{\boldsymbol{y}}, \nu_t^{\boldsymbol{y}'}\right) \mathrm{d}t \leqslant C|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma, \tag{3.15}$$

for all $\boldsymbol{y}, \boldsymbol{y}'$ such that $|\boldsymbol{y}|_\Gamma, |\boldsymbol{y}'|_\Gamma \leqslant R$.

The proof of theorem 3.8 is provided in section 4.3 on page 25.

Example 3.9 (filtering well-posedness for 2D Navier–Stokes). As an immediate consequence of theorem 3.8, we conclude that if $\mathcal{S}_t^\dagger : L_x^2 \to L_x^2$ is the solution operator of the incompressible Navier–Stokes equations (2.8) in two-dimensions, then the corresponding filtering problem is well-posed for *any* viscosity $\nu > 0$, and the mapping $\boldsymbol{y} \mapsto \nu_t^{\boldsymbol{y}}$ from measurements to the solution is locally Lipschitz stable, with a constant that is *independent* of the viscosity ν. Indeed, it follows from theorem 2.9 that the Navier–Stokes equations satisfy the bound $\|\mathcal{S}_t^\dagger(\bar{u})\|_{L_x^2} \leqslant \|\bar{u}\|_{L_x^2}$ (with $B_{\mathcal{S}} = 1$), and hence the constant $C = C(R, \rho, N, B_{\mathcal{S}}, \|\bar{u}\|_{L^1(\mu_{\text{prior}})}, T)$ of theorem 3.8 is independent of ν. In contrast, we emphasize that the Lipschitz constant for the corresponding forward problem depends *exponentially* on $1/\nu$ (cp theorem 2.9).

Example 3.10 (filtering well-posedness for 3D Navier–Stokes). Similarly, for the three-dimensional Navier–Stokes equations we obtain a short-time well-posedness result if the prior μ_{prior} is supported on sufficiently smooth initial data: e.g. if $\mu_{\text{prior}}(\{\bar{u} \in H_x^s \mid \|\bar{u}\|_{H_x^s} \leqslant M\}) = 1$ for some $s > 3/2$, and if the time-interval is sufficiently short $T < T^*(s, M)$, then the corresponding filtering problem $t \mapsto \nu_t^{\boldsymbol{y}}$ is well-posed on $t \in [0, T]$. Here, $H_x^s = H^s(\mathbb{T}^3; \mathbb{R}^3)$ denotes the well-known Sobolev space consisting of vector fields with square-integrable derivatives of order s.

Example 3.11 (uniform well-posedness for numerical discretizations in 2D and 3D). Finally, we note that theorem 3.8 (with $\mathcal{S}_t^\dagger$ replaced by $\mathcal{S}_t^\Delta$) also implies the well-posedness of the filtering problem for numerical approximations, such as the spectral method introduced in section 2.3.2, for any fixed $\nu, \Delta > 0$ and in *both two and three dimensions*. Furthermore, the stability constant $C > 0$ in (3.15) can be chosen *uniformly*, for all values of $\nu, \Delta > 0$.

3.2.2. Consistency. Next, we discuss the consistency of approximate filtering based on a discretized solution operator $\mathcal{S}_t^\Delta$, and the limiting filtering problem with solution operator $\mathcal{S}_t^\dagger$. More precisely, we show that if $\mathcal{S}_t^\Delta(\bar{u}) \to \mathcal{S}_t^\dagger(\bar{u})$ converges in a suitable sense, then $\nu_t^{\Delta,\boldsymbol{y}} \to \nu_t^{\boldsymbol{y}}$ in $L_t^1(\mathcal{P})$ also converges:

Theorem 3.12 (filtering consistency). *Let $\mu_{\text{prior}} \in \mathcal{P}_1(X)$ be a prior with finite second moments, $\|\bar{u}\|_{L^2(\mu_{\text{prior}})} < \infty$. Assume that $\mathcal{S}_t^\Delta, \mathcal{S}_t^\dagger : X \to X$ satisfy* (Δ.1)–(Δ.3) *and* ($\mathcal{S}$.1)

and ($\mathcal{S}$.2)*, respectively. Then there exists a constant* $C = C(R, T, N, \rho, \|\bar{u}\|_{L^2(\mu_{\text{prior}})}, \mathcal{G}) > 0$*, independent of* Δ*, such that*

$$\int_0^T W_1\left(\nu_t^{\Delta,\boldsymbol{y}}, \nu_t^{\boldsymbol{y}}\right) \mathrm{d}t \leqslant C \int_0^T \|\mathcal{S}_t^{\Delta}(\bar{u}) - \mathcal{S}_t^{\dagger}(\bar{u})\|_{L^2(\mu_{\text{prior}})} \, \mathrm{d}t. \tag{3.16}$$

In particular, if $\mathcal{S}_t^{\Delta}(\bar{u}) \to \mathcal{S}_t^{\dagger}(\bar{u})$ *in* $L^1([0,T];\ L^2(\mu_{\text{prior}}))$ *at a certain convergence rate, then* $\nu_t^{\Delta,\boldsymbol{y}} \to \nu_t^{\boldsymbol{y}}$ *in* $L_t^1(\mathcal{P})$ *converges at the same rate.*

The proof of theorem 3.12 is provided in section 4.4 below, on page 26.

Example 3.13. Based on the analysis of [3], the solutions computed by the numerical scheme in section 2.3.2 are expected to converge spectrally for the two-dimensional Navier–Stokes equations: if $\mu_{\text{prior}}(\{\bar{u} \in H^s \mid \|\bar{u}\|_{H^s} \leqslant M\}) = 1$ for some $M > 0$, then $\|\mathcal{S}_t^{\Delta}(\bar{u}) - \mathcal{S}_t^{\dagger}(\bar{u})\|_{L_x^2} \leqslant C\Delta^s$. In particular, by theorem 3.12, this implies a similar convergence rate also for the Bayesian filtering problem, i.e.

$$\int_0^T W_1(\nu_t^{\Delta,\boldsymbol{y}}, \nu_t^{\boldsymbol{y}}) \mathrm{d}t \leqslant C\Delta^s.$$

Remark 3.14 (surrogate models). Recently, surrogate models based on novel neural network architectures have been proposed to speed up many-query problems such as Bayesian DA (see e.g. [30] for first results in this direction). These neural network-based methods provide an approximation $\mathcal{S}_t^{\Delta} \approx \mathcal{S}_t^{\dagger}$ of the underlying solution operator based on the minimization of an empirical loss function, which is chosen as a Monte-Carlo approximation of

$$\widehat{\mathcal{L}}_{\text{loss}} = \int_0^T \|\mathcal{S}_t^{\Delta}(\bar{u}) - \mathcal{S}_t^{\dagger}(\bar{u})\|_{L^2(\mu_{\text{prior}})}^2 \, \mathrm{d}t.$$

Theorem 3.12 provides a first step towards a more detailed estimate on the approximation error of the underlying filtering problem $\nu_t^{\Delta,\boldsymbol{y}} \approx \nu_t^{\boldsymbol{y}}$, in terms of the loss $\widehat{\mathcal{L}}_{\text{loss}}$. Indeed, the upper bound (3.16) implies the estimate $d_T(\nu_t^{\Delta,\boldsymbol{y}}, \nu_t^{\boldsymbol{y}}) \leqslant C\sqrt{T\widehat{\mathcal{L}}_{\text{loss}}}$, on the time-integrated Wasserstein-distance between $\nu_t^{\Delta,\boldsymbol{y}}$ and $\nu_t^{\boldsymbol{y}}$!

3.2.3. Compactness and uniform stability. We finally turn our attention to the approximate filtering problem for the particular case of the Navier–Stokes equations in 3D, in the *absence* of *a priori* well-posedness for the forward problem. In contrast, the approximate solutions obtained from numerical discretizations, such as the spectral scheme outlined in section 2.3.2 are well-defined for any given grid size $\Delta > 0$; hence we focus our attention on the behaviour of these numerical discretizations $\mathcal{S}_t^{\Delta} : X \to X$, with $X = L_x^2$ the space of square-integrable vector fields $\bar{u} : \mathbb{T}^3 \to \mathbb{R}^3$ on the three-dimensional, 2π-periodic torus $\mathbb{T}^3$, satisfying $\mathrm{div}(\bar{u}) = 0$. As pointed out in section 2.3.1, in this case, the uniqueness and stability of the forward problem for the Navier–Stokes equations is not known for general input data $\bar{u} \in L_x^2$. Nevertheless, the corresponding numerical approximations computed by the scheme (2.9) are well-defined for any $\Delta > 0$. Such approximations allow us to compute approximate filtering distributions $\nu_t^{\Delta,\boldsymbol{y}}$ for a given discretization parameter $\Delta > 0$. Even though stability of the corresponding forward problem is not known, and we could have $\mathrm{Lip}(\mathcal{S}_t^{\Delta}) \to \infty$ as $\Delta \to 0$, the results of the present work nevertheless allow us to prove *uniform stability* and *compactness* for the corresponding approximate filtering distributions.

Before stating the next theorem, we recall that a probability measure $\mu_{\text{prior}} \in \mathcal{P}_1(L^2_x)$ is said to have bounded support, if there exists $M > 0$, such that

$$\mu_{\text{prior}}\left(\left\{\bar{u} \in L^2_x \mid \|\bar{u}\|_{L^2_x} \leqslant M\right\}\right) = 1. \tag{3.17}$$

We can now state the following compactness result:

Theorem 3.15 (filtering compactness for Navier–Stokes). *Assume that the prior $\mu_{\text{prior}} \in \mathcal{P}_1(L^2_x)$ has bounded support (3.17) for some $M > 0$, where $L^2_x := L^2(\mathbb{T}^d; \mathbb{R}^d)$ denotes the space of square integrable, periodic vector fields. Assume that μ_{prior} is concentrated on divergence-free vector fields. Let $0 = t_0 < t_1 < \cdots < t_N = T$ be a strictly increasing sequence for fixed $N \in \mathbb{N}$. Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_N) \in \mathbb{R}^{d\times N}$ be a sequence of measurements. Let $\mathcal{S}^\Delta_t : L^2_x \to L^2_x$ for $\Delta > 0$ be approximate solution operators satisfying (Δ.1)–(Δ.3) of proposition 2.10, and let $\nu^{\Delta,\boldsymbol{y}}_t$ be the solution of the associated filtering problem. Then the sequence $\nu^{\Delta,\boldsymbol{y}}_t$ is pre-compact in $C_{\text{loc}}(\mathbb{R}^{d\times N}; L^1_t(\mathcal{P}))$, as $\Delta \to 0$. In fact, for any $R > 0$, there exists a constant $C = C(R, \rho, N, M) > 0$, such that*

$$\sup_{|\boldsymbol{y}|_\Gamma, |\boldsymbol{y}'|_\Gamma \leqslant R} W_1(\nu^{\Delta,\boldsymbol{y}}_t, \nu^{\Delta,\boldsymbol{y}'}_t) \leqslant C|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma, \quad \forall\, t \in [0, T], \tag{3.18}$$

and there exists a subsequence $\Delta_k \to 0$, and $\nu^{*,\boldsymbol{y}}_t$ such that for any $R > 0$,

$$\sup_{|\boldsymbol{y}|_\Gamma \leqslant R} \int_0^T W_1\left(\nu^{\Delta_k,\boldsymbol{y}}_t, \nu^{*,\boldsymbol{y}}_t\right) \mathrm{d}t \to 0,$$

converges locally uniformly in $\boldsymbol{y}$. Any such limit satisfies the stability estimate (3.18) in $\boldsymbol{y}$.

For the details of the proof of theorem 3.15, we refer to section 4.5, page 31.

Remark 3.16 (stability of expectations). The stability estimate (3.18) in theorem 3.15 is a consequence of the continuity properties of the noise distribution ρ, and is *independent of any continuity properties of the observable $\mathcal{L}^\Delta(u)$*. One implication of (3.18) is that for any Lipschitz continuous $\Phi \in \text{Lip}(X)$ and $t \in [0, T]$, the mapping

$$\boldsymbol{y} \mapsto \mathbb{E}^{\Delta,\boldsymbol{y}}_t[\Phi] := \int_X \Phi(u) \mathrm{d}\nu^{\Delta,\boldsymbol{y}}_t(u),$$

is locally Lipschitz continuous under the assumptions of theorem 3.15: by Kantorovich duality, we have

$$W_1\left(\nu^{\Delta,\boldsymbol{y}}_t, \nu^{\Delta,\boldsymbol{y}'}_t\right) = \sup_{\|\Phi\|_{\text{Lip}}=1} \int_X \Phi(u)\left[\nu^{\Delta,\boldsymbol{y}}_t(\mathrm{d}u) - \nu^{\Delta,\boldsymbol{y}'}_t(\mathrm{d}u)\right].$$

Thus, by (3.18) for any $R > 0$, there exists $C(R, \rho, N, M) > 0$, such that

$$\begin{aligned}\left|\mathbb{E}^{\Delta,\boldsymbol{y}}[\Phi] - \mathbb{E}^{\Delta,\boldsymbol{y}'}[\Phi]\right| &= \|\Phi\|_{\text{Lip}} \left| \int_X \frac{\Phi(u)}{\|\Phi(u)\|_{\text{Lip}}} \left[\nu^{\Delta,\boldsymbol{y}}_t(\mathrm{d}u) - \nu^{\Delta,\boldsymbol{y}'}_t(\mathrm{d}u)\right]\right| \\ &\leqslant \|\Phi\|_{\text{Lip}} W_1\left(\nu^{\Delta,\boldsymbol{y}}_t, \nu^{\Delta,\boldsymbol{y}'}_t\right) \\ &\leqslant C\|\Phi\|_{\text{Lip}}|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma. \end{aligned} \tag{3.19}$$

Remark 3.17 (real-analyticity of expectations). Under the assumptions of theorem 3.15 and assuming additionally that the noise $\mu_{\text{noise}} = \rho(y)\mathrm{d}y$ is *Gaussian noise*, then the Lipschitz continuity (3.19) of remark 3.16 can be considerably strengthened to show that, for any $\phi \in L^\infty(\mu_{\text{prior}})$ and $t \in [0, T]$, the mapping

$$\mathbb{R}^{d\times N} \to \mathbb{R}, \qquad \boldsymbol{y} \mapsto \mathbb{E}_t^{\Delta,\boldsymbol{y}}[\phi],$$

is *real analytic*: indeed, for the corresponding BIP (estimating the initial data) it follows from [17, lemma 4.5] that the mapping

$$\mathbb{R}^{d\times N} \to \mathbb{R}, \qquad \boldsymbol{y} = (y_1, \ldots, y_N) \mapsto \int_X \psi(\bar{u})\mathrm{d}\mu^{\Delta,y_{1:k}}(\bar{u}), \tag{3.20}$$

is real-analytic for any $\psi \in L^1(\mu_{\text{prior}})$ and $k \in \{1, \ldots, N\}$. By (3.8), we have

$$\nu_t^{\Delta,\boldsymbol{y}} = \sum_{k=0}^{N} 1_{[t_k,t_{k+1})}(t)\ \mathcal{S}_{t,\#}^{\Delta}\mu^{\Delta,y_{1:k}},$$

where we formally set $t_{N+1} = \infty$, and hence for fixed $t \in [0, T]$, there exists k such that $\nu_t^{\Delta,\boldsymbol{y}} = \mathcal{S}_{t,\#}^{\Delta}\mu^{\Delta,y_{1:k}}$. Thus, for any $\phi \in L^\infty(\mu_{\text{prior}})$, we see that

$$\begin{aligned}\mathbb{E}_t^{\Delta,\boldsymbol{y}}[\phi] &= \int_X \phi(u)\nu_t^{\Delta,\boldsymbol{y}}(\mathrm{d}u) = \int_X \phi(u)\big[\mathcal{S}_{t,\#}^{\Delta}\mu^{\Delta,y_{1:k}}\big](\mathrm{d}u)\\ &= \int_X \phi\big(\mathcal{S}_t^{\Delta}(\bar{u})\big)\mu^{\Delta,y_{1:k}}(\mathrm{d}\bar{u}),\end{aligned}$$

is of the form (3.20) with $\psi(\bar{u}) := \phi\big(\mathcal{S}_t^{\Delta}(\bar{u})\big) \in L^\infty(\mu_{\text{prior}}) \subset L^1(\mu_{\text{prior}})$. Hence $\boldsymbol{y} \mapsto \mathbb{E}_t^{\Delta,\boldsymbol{y}}[\phi]$ is real-analytic by the results of [17]. In particular, this conclusion is independent of any regularity properties of $\bar{u} \mapsto \mathcal{S}_t^{\Delta}(\bar{u})$.

Remark 3.18. Theorem 3.15 shows that even though the forward problem for the three-dimensional Navier–Stokes equations may be *ill-posed*, we can nevertheless assign a set of candidate solutions for the Bayesian DA problem to a family of approximate posteriors $\nu_t^{\Delta,\boldsymbol{y}}$ at resolution $\Delta > 0$. This set of candidate solutions in the limit $\Delta \to 0$ is given by

$$\mathcal{M} = \left\{ \nu_t^{*,\boldsymbol{y}} \in L_t^1(\mathcal{P}) \,\middle|\, \exists \Delta_k \to 0, \text{ s.t. } \nu_t^{*,\boldsymbol{y}} = \lim_{k\to\infty} \nu_t^{\Delta_k,\boldsymbol{y}} \right\},$$

or equivalently, we can write

$$\mathcal{M} = \bigcap_{\overline{\Delta}>0} \mathrm{cl}\Big(\Big\{ (t,\boldsymbol{y}) \mapsto \nu_t^{\Delta,\boldsymbol{y}} \,|\, \Delta \leqslant \overline{\Delta} \Big\} \Big),$$

where cl denotes the closure in $C_{\text{loc}}(\mathbb{R}^d; L_t^1(\mathcal{P}))$. We note that the set $\mathcal{M}$ is non-empty: this follows from the fact that any finite intersections are clearly non-empty and that each of the sets is a compact subset of $C_{\text{loc}}(\mathbb{R}^d; L_t^1(\mathcal{P}))$. It then follows from the finite intersection property of compact sets that also their intersection $\mathcal{M} \neq \emptyset$, i.e. there always exists at least one candidate solution.

The last remark can be interpreted as an *existence result* for solutions of the Bayesian DA problem. This is an analogue of corresponding existence results for the forward problem of the Navier–Stokes equations [28]. However, in contrast to the existence result of the forward problem, which implies the existence of suitable limits $\mathcal{S}_t^{\Delta_k}(\bar{u}) \to \mathcal{S}_t^{\dagger}(\bar{u})$ for fixed $\bar{u}$ and which may exhibit *no stability in* $\bar{u}$, limits obtained for the filtering problem do not only exist, but are also *uniformly stable* with respect to $\boldsymbol{y}$, giving rise to limits $\nu_t^{*,\boldsymbol{y}}$ with continuous dependence on $\boldsymbol{y}$. This remarkable stability of the DA problem is in stark contrast with the corresponding forward problem, even though both problems involve the prediction of a future state.

We also note that, following the connection between BIPs and Bayesian DA pointed out in remark 3.7, we can readily obtain a corresponding existence result for BIPs, which we state in passing:

Theorem 3.19 (compactness and stability for BIP). *Let* $\mu_{\text{prior}} \in \mathcal{P}_1(X)$. *Let* $\mathcal{L}^{\Delta} : X \to \mathbb{R}^d$ *be a sequence of approximate numerical functionals, and let* $\mu^{\Delta,y}$ *be the solution of the associated BIP. If there exists* $M > 0$, *such that* $\sup_{\Delta} \|\mathcal{L}^{\Delta}(\bar{u})\|_{L^2(\mu_{\text{prior}})} \leqslant M$, *then the sequence* $\mu^{\Delta,y}$ *is pre-compact in* $C_{\text{loc}}(\mathbb{R}^d; \mathcal{P}_1(X))$, *as* $\Delta \to 0$*: in fact, for any* $R > 0$, *there exists* $C = C(R, M, \rho) > 0$, *such that*

$$W_1(\mu^{\Delta,y}, \mu^{\Delta,y'}) \leqslant C|y - y'|_{\Gamma}, \quad \forall\ |y|_{\Gamma}, |y'|_{\Gamma} \leqslant R.$$

Furthermore, there exists a subsequence $\Delta_k \to 0$, *such that for any* $R > 0$,

$$\sup_{|y|_{\Gamma} \leqslant R} W_1\left(\mu^{\Delta_k,y}, \mu^{*,y}\right) \to 0,$$

converges locally uniformly in y. Any such limit $\mu^{*,y}$ *is locally Lipschitz continuous with respect to y, and can be represented in the form* $\mathrm{d}\mu^{*,y} = \frac{1}{Z^*(y)} \exp(-\Phi^*(\bar{u}))\mathrm{d}\mu_{\text{prior}}$ *for a suitable functional* $\Phi^* : X \to \mathbb{R}$.

We emphasize that in this case, the mere uniform boundedness of the mappings $\mathcal{L}^{\Delta} : X \to \mathbb{R}^d$ is sufficient to obtain uniform stability and compactness.

4. Derivation of the main results

In this section, we provide the detailed mathematical derivation of the main results stated in the previous section.

4.1. Recursive filtering

We begin by providing a proof of the equivalence between the recursive filtering scheme of remark 3.5 and (2.4).

Proof of proposition 3.6. We proceed by induction on j. The case $j = 0$ is trivial, since

$$\widetilde{\nu}_{t_0}^{\emptyset} \;=\; \mathcal{S}_{t_0,\#}^{\dagger}\mu_{\text{prior}} \;=\; \mathcal{S}_{t_0,\#}^{\dagger}\mu^{\emptyset}.$$

For $j \geqslant 1$, we integrate against an arbitrary, integrable (cylindrical) test function $\Psi(u)$ to find, with $\delta t_j = t_j - t_{j-1}$:

$$\int_{L_x^2} \Psi(u)\widetilde{\nu}_{t_j}^{y_{1:j}}(\mathrm{d}u) = \int_{L_x^2} \Psi(u)\left[\mathcal{S}_{\delta t_j,\#}^{\dagger}\widetilde{\nu}_{t_{j-1}}^{y_{1:j}}\right](\mathrm{d}u) = \int_{L_x^2} \Psi\left(\mathcal{S}_{\delta t_j}^{\dagger}(u)\right)\widetilde{\nu}_{t_{j-1}}^{y_{1:j}}(\mathrm{d}u).$$

Substitution of the correction step (3.11), yields

$$\int_{L^2_x} \Psi(u)\, \widetilde{\nu}^{y_{1:j}}_{t_{j-1}}(\mathrm{d}u) = \int_{L^2_x} \Psi\left(\mathcal{S}^\dagger_{\delta t_j}(u)\right) q^\dagger_j(u)\, \widetilde{\nu}^{y_{1:(j-1)}}_{t_{j-1}}(\mathrm{d}u),$$

where

$$q^\dagger_j(u) = \frac{1}{Z^\dagger_j(y_j)} \exp\left(-\widetilde{\Phi}^{\dagger,y_j}_j(u)\right).$$

By the induction hypothesis, the measure $\widetilde{\nu}^{y_{1:(j-1)}}_{t_{j-1}}$ can be written as a push-forward:

$$\widetilde{\nu}^{y_{1:(j-1)}}_{t_{j-1}} = \mathcal{S}^\dagger_{t_{j-1},\#}\mu^{y_{1:(j-1)}}.$$

Thus, substituting above, we find

$$\begin{aligned}\int_{L^2_x} \Psi(u)\, \widetilde{\nu}^{y_{1:j}}_{t_j}(\mathrm{d}u) &= \int_{L^2_x} \Psi(\mathcal{S}^\dagger_{\delta t_j}(u))\, q^\dagger_j(u) \left[\mathcal{S}^\dagger_{t_{j-1},\#}\mu^{y_{1:(j-1)}}(u)\right](\mathrm{d}u) \\ &= \int_{L^2_x} \Psi(\mathcal{S}^\dagger_{t_j}(\bar{u}))q^\dagger_j(\mathcal{S}^\dagger_{t_{j-1}}(\bar{u}))\mu^{y_{1:(j-1)}}(\mathrm{d}\bar{u}),\end{aligned}$$

where we have used that $\mathcal{S}^\dagger_{t_{j-1}} \circ \mathcal{S}^\dagger_{\delta t_j} = \mathcal{S}^\dagger_{t_j}$ to simplify the argument of Ψ in the last step. We now note that, by our definition of $q^\dagger_j$ and $\mu^{y_{1:(j-1)}}$, we have

$$\begin{aligned}q^\dagger_j(\mathcal{S}^\dagger_{t_{j-1}}(\bar{u}))\, \mu^{y_{1:(j-1)}}(\mathrm{d}\bar{u}) &\propto \exp\left(-\widetilde{\Phi}^{\dagger,y_j}_j\left(\mathcal{S}^\dagger_{t_{j-1}}(\bar{u})\right)\right) \\ &\quad \times \exp\left(-\sum_{k=1}^{j-1} \Phi^{\dagger,y_k}_k(\bar{u})\right)\mu_{\mathrm{prior}}(\mathrm{d}\bar{u}) \\ &= \exp\left(-\sum_{k=1}^{j} \Phi^{\dagger,y_k}_k(\bar{u})\right)\mu_{\mathrm{prior}}(\mathrm{d}\bar{u}),\end{aligned}$$

where we have taken into account the identity $\widetilde{\Phi}^{\dagger,y_j}_j\left(\mathcal{S}^\dagger_{t_{j-1}}(\bar{u})\right) = \Phi^{\dagger,y_j}_j(\bar{u})$ in the last step (cp equation (3.13)). The proportionality constant can be determined by normalization. The last expression is equal to $\mu^{y_{1:j}}$, and hence

$$\int_{L^2_x} \Psi(u)\widetilde{\nu}^{y_{1:j}}_{t_j}(\mathrm{d}u) = \int_{L^2_x} \Psi(\mathcal{S}^\dagger_{t_j}(u))\mu^{y_{1:j}}(\mathrm{d}u) = \int_{L^2_x} \Psi(u)\left[\mathcal{S}^\dagger_{t_j,\#}\mu^{y_{1:j}}\right](\mathrm{d}u).$$

Since Ψ was an arbitrary (cylindrical) test function, the claimed identity follows. □

4.2. Stability results for the BIP

The goal of the present section is to derive general stability results for the BIP. Combined with the push-forward representation of the Bayesian filtering and smoothing distributions of proposition 3.2, these estimates form the basis of our analysis of the well-posedness of the Bayesian DA problem.

While the temporal nature of the measurement data is important for the filtering distribution $\nu^{\mathbf{y}}_t$, the solutions to the BIP of interest take the form $\mu^{y_{1:k}}$ of (3.3) for some

given k, corresponding to a combined observable $\widetilde{\mathcal{L}}^\dagger : X \to \mathbb{R}^{\widetilde{d}}$, $\widetilde{d} = d \times k$, of the form $\widetilde{\mathcal{L}}^\dagger(\bar{u}) = (\mathcal{L}_1^\dagger(\bar{u}), \dots, \mathcal{L}_k^\dagger(\bar{u}))$, and with combined measurement $\widetilde{y} = (y_1, \dots, y_k)$ of the form $\widetilde{y} = \widetilde{\mathcal{L}}^\dagger(\bar{u}) + \widetilde{\eta}$ with $\widetilde{\eta} = (\eta_1, \dots, \eta_k)$. Hence, resulting in a standard BIP. Dropping the tildes in the following, we will thus study the general properties of posteriors for the BIP for measurable observables

$$\mathcal{L}^\dagger : X \to \mathbb{R}^d, \qquad \bar{u} \mapsto \mathcal{L}^\dagger(\bar{u}), \tag{4.1}$$

and with $\mathbb{R}^d$-valued noise $\eta \sim \rho(y)\mathrm{d}y$ satisfying the assumptions (N.1)–(N.3). The solution of the BIP for measurement operator (4.1) is then given [25] by the posterior

$$\mu^y(\mathrm{d}\bar{u}) = \frac{1}{Z^\dagger(y)} \exp\big(-\Phi^{\dagger,y}(\bar{u})\big)\, \mu_{\text{prior}}(\mathrm{d}\bar{u}), \tag{4.2}$$

where

$$\Phi^{\dagger,y}(\bar{u}) := -\log\, \rho\big(y - \mathcal{L}^\dagger(\bar{u})\big) \tag{4.3}$$

denotes the log-likelihood function, and

$$Z^\dagger(y) = \int_X \exp\big(-\Phi^{\dagger,y}(\bar{u})\big)\mu_{\text{prior}}(\mathrm{d}\bar{u}), \tag{4.4}$$

is the required normalization constant. As is customary, we will denote the Radon–Nikodym derivative of μ^y with respect to μ by $\mathrm{d}\mu^y/\mathrm{d}\mu_{\text{prior}}$, i.e.

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_{\text{prior}}}(\bar{u}) = \frac{1}{Z^\dagger(y)} \exp\big(-\Phi^{\dagger,y}(\bar{u})\big). \tag{4.5}$$

Given an approximation $\mathcal{L}^\Delta \approx \mathcal{L}^\dagger$, we similarly define $\mu^{\Delta,y}$, $\Phi^{\Delta,y}(\bar{u})$ and $Z^\Delta(y)$, with $\mathcal{L}^\Delta(\bar{u})$ replacing $\mathcal{L}^\dagger(\bar{u})$, in equations (4.2)–(4.4).

While the existence of a solution to the BIP is ensured by the non-negativity of the noise distribution $\rho(y)$, the stability and compactness results of the present work will be based on assumptions (N.1)–(N.3) on the noise. We begin our discussion by noting the following immediate observations from these assumptions:

Lemma 4.1. *Let $\mathcal{L}^\dagger : X \to \mathbb{R}^d$ be any map. If the noise $\eta \sim \rho(y)\mathrm{d}y$ satisfies assumptions (N.1)–(N.3), then we have for all $y, y' \in \mathbb{R}^d$*

$$\left|\mathrm{e}^{-\Phi^{\dagger,y}(\bar{u})} - \mathrm{e}^{-\Phi^{\dagger,y'}(\bar{u})}\right| \leqslant L_\rho |y - y'|_\Gamma. \tag{4.6}$$

The log-likelihood $\Phi^{\Delta,y}$ is bounded from below, uniformly in $\Delta > 0$ and $y \in \mathbb{R}^d$: there exists a constant $C_\rho \geqslant 0$ depending only on $\sup_{y\in\mathbb{R}^d} \rho(y) < \infty$, such that

$$\operatorname*{ess\,inf}_{\bar{u}\in X} \Phi^{\dagger,y}(\bar{u}) \geqslant -C_\rho, \quad \forall\, y \in \mathbb{R}^d. \tag{4.7}$$

There exists a constant $B_\rho \geqslant 0$, such that

$$\Phi^{\dagger,y}(\bar{u}) \leqslant B_\rho + \frac{1}{2}|y - \mathcal{L}^\dagger(\bar{u})|_\Gamma^2. \tag{4.8}$$

In particular, we have

$$\Phi^{\dagger,y}(\bar{u}) \leqslant B_\rho + |y|_\Gamma^2 + |\mathcal{L}^\dagger(\bar{u})|_\Gamma^2. \tag{4.9}$$

Furthermore, if $\mathcal{L}^\Delta$ is an approximation of $\mathcal{L}^\dagger$, then for the same constant L_ρ as above:

$$\left|\mathrm{e}^{-\Phi^{\dagger,y}(\bar{u})} - \mathrm{e}^{-\Phi^{\Delta,y}(\bar{u})}\right| \leqslant L_\rho |\mathcal{L}^\dagger(\bar{u}) - \mathcal{L}^\Delta(\bar{u})|_\Gamma. \tag{4.10}$$

We now discuss the stability of the posterior μ^y for the BIP with respect to the measurement y. We note that our discussion of stability for the BIP overlaps in part with a similar discussion contained in [25, 35]. In particular, [35] contains a general discussion of the stability of posteriors with respect to both the log-likelihood and priors, and with respect to a number of distance metrics between probability measures. Since some necessary estimates have not appeared in [25, 35], at least in the precise form needed for our purposes, we have decided to include detailed proofs in this manuscript.

We begin our discussion of the stability properties of the BIP with the following lemma, proving that the sequence of densities $\mathrm{d}\mu^{\Delta,y}/\mathrm{d}\mu_{\text{prior}}$ is uniformly bounded in $L^\infty(\mu_{\text{prior}})$.

Recalling that $\|\mathcal{L}^\dagger\|_{L^2(\mu_{\text{prior}})} := \left(\int_X \|\mathcal{L}^\dagger(\bar{u})\|_Y^2 \mu_{\text{prior}}(\mathrm{d}\bar{u})\right)^{1/2}$, we now state the following

Lemma 4.2. *Let $\mathrm{d}\mu^y/\mathrm{d}\mu_{\text{prior}}$ be given by (4.5), and $Z^\dagger(y)$ be defined as in (4.4). Then*

$$Z^\dagger(y) \geqslant \exp\left(-\int_X \Phi^{\dagger,y}(\bar{u})\mu_{\text{prior}}(\mathrm{d}\bar{u})\right), \tag{4.11}$$

and

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_{\text{prior}}}(\bar{u}) \leqslant \exp\left(\int_X \Phi^{\dagger,y}(\bar{u})\mu_{\text{prior}}(\mathrm{d}\bar{u}) - \operatorname*{ess\,inf}_{u\in X} \Phi^{\dagger,y}(\bar{u})\right), \quad \forall\ \bar{u} \in X, \tag{4.12}$$

In particular, if the noise $\eta \sim \rho(y)\mathrm{d}y$ satisfies (N.1)–(N.3), then there exists a constant $C > 0$ depending only on the noise distribution $\rho(y)$, such that

$$Z^\dagger(y) \geqslant \frac{1}{C}\exp\left(-|y|_\Gamma^2 - \|\mathcal{L}^\dagger\|_{L^2(\mu_{\text{prior}})}^2\right), \tag{4.13}$$

and

$$\frac{\mathrm{d}\mu^{\dagger,y}}{\mathrm{d}\mu_{\text{prior}}}(u) \leqslant C\,\exp\left(|y|_\Gamma^2 + \|\mathcal{L}^\dagger\|_{L^2(\mu_{\text{prior}})}^2\right), \quad \forall\ u \in X. \tag{4.14}$$

Proof. Since the exponential (Gaussian-like) factor in the definition of $\mathrm{d}\mu^{\dagger,y}/\mathrm{d}\mu_{\text{prior}}$, equation (4.5), is bounded from above by $\exp(-\operatorname{ess\,inf}_{u\in X} \Phi^{\dagger,y}(u))$, it suffices to prove the lower bound on $Z^\dagger(y)$. From the convexity of $z \mapsto \exp(-z)$ and Jensen's inequality, we obtain

$$\exp\left(-\int_X \Phi^{\dagger,y}(\bar{u})\mu_{\text{prior}}(\mathrm{d}\bar{u})\right) \leqslant \int_X \exp(-\Phi^{\dagger,y}(\bar{u}))\mu_{\text{prior}}(\mathrm{d}\bar{u}) = Z^\dagger(y).$$

This implies the first two estimates (4.11) and (4.12) of this lemma.

Under the noise assumptions (N.1)–(N.3), and by (4.9), there exists $B_\rho > 0$ depending only on the noise distribution $\rho(y)$, such the last term can be bounded from below, yielding

$$Z^\dagger(y) \geqslant \exp\left(-B_\rho - |y|_\Gamma^2 - \int_X |\mathcal{L}^\dagger(\bar{u})|_\Gamma^2 \mu_{\text{prior}}(\mathrm{d}\bar{u})\right),$$

and thus the claimed inequality (4.13) for $Z^\dagger(y)$ with $C = \exp(B_\rho)$. Furthermore, by (4.7), there exists C_ρ, such that

$$\operatorname*{ess\,inf}_{\bar{u}\in X} \Phi^{\dagger,y}(\bar{u}) \geqslant -C_\rho.$$

Thus the claimed inequality (4.14) holds with $C = \exp(B_\rho + C_\rho)$. □

We next discuss the stability of $\mathrm{d}\mu^{\dagger,y}/\mathrm{d}\mu_{\text{prior}}$ with respect to y. The following lemma shows that the map $y \mapsto \mathrm{d}\mu^{\dagger,y}/\mathrm{d}\mu_{\text{prior}}$ is locally Lipschitz continuous with respect to the L^∞-norm.

Lemma 4.3. *Assume that the noise $\eta \sim \rho(y)\mathrm{d}y$ satisfies (N.1)–(N.3). Let $\mathcal{L}^\dagger(\bar{u}) \in L^2(\mu_{\text{prior}})$. There exists a constant $C > 0$, depending only on the noise distribution $\rho(y)$, such that*

$$\left\| \frac{\mathrm{d}\mu^{\dagger,y}}{\mathrm{d}\mu_{\text{prior}}} - \frac{\mathrm{d}\mu^{\dagger,y'}}{\mathrm{d}\mu_{\text{prior}}} \right\|_{L^\infty(\mu_{\text{prior}})} \leqslant C|y-y'|_\Gamma \exp\left(|y|_\Gamma^2 + |y'|_\Gamma^2 + 2\|\mathcal{L}^\dagger\|_{L^2(\mu_{\text{prior}})}^2\right). \tag{4.15}$$

Proof. Fix $\bar{u} \in X$ for the moment. Denote $e(y) := e(y;\bar{u}) = \exp(-\Phi^{\dagger,y}(\bar{u}))$, so that

$$\begin{aligned}\frac{\mathrm{d}\mu^{\dagger,y}}{\mathrm{d}\mu_{\text{prior}}} - \frac{\mathrm{d}\mu^{\dagger,y'}}{\mathrm{d}\mu_{\text{prior}}} &= \frac{e(y)}{Z^\dagger(y)} - \frac{e(y')}{Z^\dagger(y')} \\ &= \frac{e(y)-e(y')}{Z^\dagger(y)} + \frac{e(y')}{Z^\dagger(y')}\frac{(Z^\dagger(y') - Z^\dagger(y))}{Z^\dagger(y)}.\end{aligned}$$

By (4.6), we can estimate $|e(y) - e(y')| \leqslant C|y-y'|_\Gamma$. Next, we note that this bound for $e(y)$ also implies that

$$\begin{aligned}|Z^\dagger(y) - Z^\dagger(y')| &\leqslant \int_X |e(y;\bar{u}) - e(y';\bar{u})|\mu_{\text{prior}}(\mathrm{d}\bar{u}) \\ &\leqslant C|y-y'|_\Gamma \underbrace{\int_X 1\mu_{\text{prior}}(\mathrm{d}\bar{u})}_{=1}.\end{aligned}$$

Hence,

$$\left|\frac{\mathrm{d}\mu^{y}}{\mathrm{d}\mu_{\text{prior}}} - \frac{\mathrm{d}\mu^{y'}}{\mathrm{d}\mu_{\text{prior}}}\right| \leqslant \frac{C|y-y'|_\Gamma}{Z^\dagger(y)} + \frac{e(y')}{Z^\dagger(y')}\frac{C|y-y'|_\Gamma}{Z^\dagger(y)}.$$

We proceed to estimate the factors multiplying $|y-y'|$ in the last two terms: from lemma 4.2, we can estimate

$$\frac{1}{Z^\dagger(y)} \leqslant C\,\mathrm{e}^{|y|_\Gamma^2 + \|\mathcal{L}^\dagger\|_{L^2(\mu_{\text{prior}})}^2} \leqslant C\,\mathrm{e}^{|y|_\Gamma^2 + |y'|_\Gamma^2 + 2\|\mathcal{L}^\dagger\|_{L^2(\mu_{\text{prior}})}^2},$$

and

$$\frac{e(y')}{Z^\dagger(y')}\frac{1}{Z^\dagger(y)} \leqslant C\,\mathrm{e}^{|y|_\Gamma^2 + |y'|_\Gamma^2 + 2\|\mathcal{L}^\dagger\|_{L^2(\mu_{\text{prior}})}^2}.$$

Combining these estimates, we conclude that

$$\left|\frac{\mathrm{d}\mu^{y}}{\mathrm{d}\mu_{\text{prior}}}-\frac{\mathrm{d}\mu^{y'}}{\mathrm{d}\mu_{\text{prior}}}\right| \leqslant 2C|y-y'|_{\Gamma}\ \exp\Big(|y|_{\Gamma}^{2}+|y'|_{\Gamma}^{2}+2\|\mathcal{L}^{\dagger}\|_{L^{2}(\mu_{\text{prior}})}^{2}\Big).$$

Since $\bar{u}\in X$ was arbitrary, the claimed inequality follows by taking the supremum over $\bar{u}\in X$ on the left. □

Let us also remark in passing the following lemma, whose proof is analogous to the proof of lemma 4.3.

Lemma 4.4. *Assume that the noise $\eta\sim\rho(y)\mathrm{d}y$ satisfies (N.1)–(N.3). Let $\mathcal{L}^{\dagger}(\bar{u}),\mathcal{L}^{\Delta}(\bar{u})\in L^{2}(\mu_{\text{prior}})$, and $y\in\mathbb{R}^{d}$, and denote the resulting (approximate) solution of the BIP by μ^{y}, $\mu^{\Delta,y}$, respectively. There exists a constant $C>0$, depending only on the noise distribution $\rho(y)$, such that for any $p\in[1,\infty]$, we have*

$$\left\|\frac{\mathrm{d}\mu^{\Delta,y}}{\mathrm{d}\mu_{\text{prior}}}-\frac{\mathrm{d}\mu^{y}}{\mathrm{d}\mu_{\text{prior}}}\right\|_{L^{p}(\mu_{\text{prior}})} \leqslant C\big\|\mathcal{L}^{\Delta}(\bar{u})-\mathcal{L}^{\dagger}(\bar{u})\big\|_{L^{p}(\mu_{\text{prior}})}\ \exp\Big(2|y|_{\Gamma}^{2} + \|\mathcal{L}^{\Delta}\|_{L^{2}(\mu_{\text{prior}})}^{2}+\|\mathcal{L}^{\dagger}\|_{L^{2}(\mu_{\text{prior}})}^{2}\Big).$$

Proof. The proof is an almost verbatim repetition of the proof of lemma 4.3, with the roles of y,y' and $\mathcal{L}^{\Delta}(\bar{u}),\mathcal{L}^{\dagger}(\bar{u})$ interchanged. □

4.3. Stability results for Bayesian DA

In this section, we investigate the stability properties of the solution of the filtering problem with respect to the measurements $y_1,\dots,y_N$. Our analysis will be based on the push-forward representation (3.7) of the previous section and the stability results for the BIP in section 4.2. We recall that the space $L_t^1(\mathcal{P})=L^1([0,T];\mathcal{P}(X))$ defined in section 2.2, consists of all weak-$*$ measurable mappings $[0,T]\to\mathcal{P}(X)$, $t\mapsto\nu_t$, such that $\int_0^T\|u\|_{L_x^2}\,\mathrm{d}\nu_t(u)\mathrm{d}t<\infty$, with metric $d_T(\nu_t,\nu_t'):=\int_0^T W_1(\nu_t,\nu_t')\mathrm{d}t$.

We can now state the following lemma

Lemma 4.5. *Let $T>0$. Let $\mu_{\text{prior}}\in\mathcal{P}_1(X)$ be a prior with finite first moments. Assume that the solution operator $\mathcal{S}_t^{\dagger}:X\to X$ satisfies ($\mathcal{S}$.1) and ($\mathcal{S}$.2), and let $\nu_t^{y_{1:j}}$ be given by (2.4) for $t\in[0,T]$. Then for any $R>0$, there exists $C=C(R,\rho,N,B_{\mathcal{S}},\|\bar{u}\|_{L^1(\mu_{\text{prior}})})>0$, such that for all $t,\delta t\geqslant 0$, with $t+\delta t\in[0,T]$, we have*

$$W_1\left(\nu_\tau^{y_{1:j}},\nu_\tau^{y'_{1:j}}\right)\leqslant C\left(\sum_{k=1}^{j}|y_k-y_k'|_{\Gamma}^{2}\right)^{1/2} \tag{4.16}$$

and

$$\int_t^{t+\delta t}W_1\left(\nu_\tau^{y_{1:j}},\nu_\tau^{y'_{1:j}}\right)\mathrm{d}\tau\leqslant C\delta t\left(\sum_{k=1}^{j}|y_k-y_k'|_{\Gamma}^{2}\right)^{1/2}, \tag{4.17}$$

for all $y_{1:j} = (y_1, \dots, y_j)$, $y'_{1:j} = (y'_1, \dots, y'_j)$ *such that*

$$\left(\sum_{k=1}^{j} |y_k|_{\Gamma_k}^2\right)^{1/2} \leqslant R, \qquad \left(\sum_{k=1}^{j} |y'_k|_{\Gamma_k}^2\right)^{1/2} \leqslant R.$$

Proof. To simplify the notation in the following, we set

$$|y_{1:j}|_\Gamma := \left(\sum_{k=1}^{j} |y_k|_\Gamma^2\right)^{1/2}.$$

By proposition 3.2, (3.7), we have $\nu_t^{y_{1:j}} = \mathcal{S}_{t,\#}^\dagger \mu^{y_{1:j}}$, where $\mu^{y_{1:j}}$ solves a BIP and is given by (3.3). In fact, $\mu^{y_{1:j}}$ is the solution of a standard BIP with noise $\widetilde{\eta} = (\eta_1, \dots, \eta_j)$. The noise distribution of $\widetilde{\eta}$ satisfies assumption (N.1)–(N.3) with the distribution $\rho : \mathbb{R}^d \to \mathbb{R}$ replaced by $\widetilde{\rho} : \mathbb{R}^{d\times j} \to \mathbb{R}$, where $\widetilde{\rho}(y_1, \dots, y_j) := \prod_{k=1}^{j} \rho(y_k)$, and with altered constants in (N.1)–(N.3) depending now also on $j \leqslant N$ in addition to ρ. Thus, by lemma 4.3, there exists a constant $C = C(\rho, N, R) > 0$, depending only on the noise distribution, the total number of measurements N and on $R > 0$, such that we obtain

$$\left\| \frac{\mathrm{d}\mu^{y_{1:j}}}{\mathrm{d}\mu_{\text{prior}}} - \frac{\mathrm{d}\mu^{y'_{1:j}}}{\mathrm{d}\mu_{\text{prior}}} \right\|_{L^\infty(\mu_{\text{prior}})} \leqslant C|y_{1:j} - y'_{1:j}|_\Gamma. \tag{4.18}$$

Let $\Phi(u) \in \mathrm{Lip}(X)$ be a function with Lipschitz constant $\leqslant 1$. Then there exists $g(u)$ such that

$$\Phi(u) - \Phi(0) = g(u)\|u\|_X, \quad |g(u)| \leqslant 1.$$

Now note that

$$\begin{aligned}
&\int_X \Phi(u) \left[\nu_t^{y_{1:j}}(\mathrm{d}u) - \nu_t^{y'_{1:j}}(\mathrm{d}u)\right] \\
&\quad = \int_X [\Phi(u) - \Phi(0)] \left[\nu_t^{\Delta, y_{1:j}}(\mathrm{d}u) - \nu_t^{\Delta, y'_{1:j}}(\mathrm{d}u)\right] \\
&\quad = \int_X g(u)\|u\|_X \mathcal{S}_{t,\#}^\dagger \left[\mu^{y_{1:j}} - \mu^{y'_{1:j}}\right](\mathrm{d}u) \\
&\quad = \int_X g(\mathcal{S}_t^\dagger(\bar{u}))\|\mathcal{S}_t^\dagger(\bar{u})\|_X \left[\frac{\mathrm{d}\mu^{y_{1:j}}}{\mathrm{d}\mu_{\text{prior}}} - \frac{\mathrm{d}\mu^{y'_{1:j}}}{\mathrm{d}\mu_{\text{prior}}}\right] \mu_{\text{prior}}(\mathrm{d}\bar{u}) \\
&\quad \leqslant \int_X \left|g(\mathcal{S}_t^\dagger(\bar{u}))\right| \|\mathcal{S}_t^\dagger(\bar{u})\|_X \left|\frac{\mathrm{d}\mu^{y_{1:j}}}{\mathrm{d}\mu_{\text{prior}}} - \frac{\mathrm{d}\mu^{y'_{1:j}}}{\mathrm{d}\mu_{\text{prior}}}\right| \mu_{\text{prior}}(\mathrm{d}\bar{u}) \\
&\quad \overset{\substack{|g(u)|\leqslant 1, \\ \|\mathcal{S}_t^\dagger(\bar{u})\|_X \leqslant B_{\mathcal{S}}\|\bar{u}\|_X \\ \downarrow \\ \leqslant}}{} B_{\mathcal{S}} \int_X \|\bar{u}\|_X \left|\frac{\mathrm{d}\mu^{y_{1:j}}}{\mathrm{d}\mu_{\text{prior}}} - \frac{\mathrm{d}\mu^{y'_{1:j}}}{\mathrm{d}\mu_{\text{prior}}}\right| \mu_{\text{prior}}(\mathrm{d}\bar{u}) \\
&\quad \leqslant B_{\mathcal{S}} \left(\int_X \|\bar{u}\|_X \mu_{\text{prior}}(\mathrm{d}\bar{u})\right) \left\| \frac{\mathrm{d}\mu^{y_{1:j}}}{\mathrm{d}\mu_{\text{prior}}} - \frac{\mathrm{d}\mu^{y'_{1:j}}}{\mathrm{d}\mu_{\text{prior}}} \right\|_{L^\infty(\mu_{\text{prior}})}.
\end{aligned}$$

Taking the supremum over all $\Phi(u)$ such that $\|\Phi\|_{\mathrm{Lip}} \leqslant 1$ on the left, and noting the upper bound (4.18) on the last term, we find

$$W_1\left(\nu_t^{y_{1:j}}, \nu_t^{y'_{1:j}}\right) \leqslant C|y_{1:j} - y'_{1:j}|_\Gamma,$$

where the constant $C = C(R, \rho, N, B_{\mathcal{S}}, \|\bar{u}\|_{L^1(\mu_{\mathrm{prior}})}) > 0$ depends on R, the noise distribution ρ, the number of measurements N, the first moment of the prior μ_{prior} and on the boundedness constant $B_{\mathcal{S}}$ of the forward operator $\mathcal{S}_t^\dagger$ (cp ($\mathcal{S}$.2)). In particular, C is independent of $y_{1:j}$, $y'_{1:j}$. This shows the upper bound (4.16). Integrating in time, we obtain the claimed inequality (4.17)

$$\int_t^{t+\delta t} W_1\left(\nu_t^{y_{1:j}}, \nu_t^{y'_{1:j}}\right) \mathrm{d}t \leqslant C\delta t|y_{1:j} - y'_{1:j}|_\Gamma.$$

□

While the above estimate is applicable to the smoothing problem, i.e. with a fixed set of measurements for all $t \in [0, T]$, we will next prove a corresponding stability theorem for the solution of the *filtering* problem; more precisely, we prove that if $\nu_t^{\boldsymbol{y}}$ denotes the solution of the filtering problem with prior $\mu_{\mathrm{prior}} \in \mathcal{P}_1(X)$, for a solution operator $\mathcal{S}_t^\dagger : X \to X$ satisfying ($\mathcal{S}$.1) and ($\mathcal{S}$.2), and measurements $\boldsymbol{y} = (y_1, \ldots, y_N)$, then for any $R > 0$, there exists $C = C(R, \rho, N, B_{\mathcal{S}}, \|\bar{u}\|_{L^1(\mu_{\mathrm{prior}})}) > 0$, such that

$$W_1\left(\nu_t^{\boldsymbol{y}}, \nu_t^{\boldsymbol{y}'}\right) \leqslant C|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma, \quad \forall\, t \in [0, T], \tag{4.19}$$

and

$$\int_0^T W_1\left(\nu_t^{\boldsymbol{y}}, \nu_t^{\boldsymbol{y}'}\right) \mathrm{d}t \leqslant CT|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma, \tag{4.20}$$

for all $\boldsymbol{y}, \boldsymbol{y}'$ such that $|\boldsymbol{y}|_\Gamma, |\boldsymbol{y}'|_\Gamma \leqslant R$.

Proof of theorem 3.8. The claimed stability estimate follows readily from lemma 4.5: indeed, $\nu_t^{\boldsymbol{y}}$ is defined piece-wise in time, for $t \geqslant 0$, as

$$\nu_t^{\boldsymbol{y}} = \sum_{k=0}^{N} 1_{[t_k, t_{k+1})}(t)\, \nu_t^{y_{1:k}},$$

where we formally define $t_{N+1} := \infty$. By lemma 4.5, equation (4.16), this implies that for any $|\boldsymbol{y}|_\Gamma, |\boldsymbol{y}'|_\Gamma \leqslant R$, we have

$$\begin{aligned}
\sup_{t\in[0,T]} W_1\left(\nu_t^{\boldsymbol{y}}, \nu_t^{\boldsymbol{y}'}\right) &= \max_{k=0,\ldots,N} W_1\left(\nu_t^{y_{1:k}}, \nu_t^{y'_{1:k}}\right) \\
&\leqslant C \max_{k=0,\ldots,N} |y_{1:k} - y'_{1:k}|_\Gamma \\
&\leqslant C|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma,
\end{aligned}$$

where we recall that the constant $C > 0$ in lemma 4.5 depends on R, ρ, N, $B_{\mathcal{S}}$ and $\|\bar{u}\|_{L^1(\mu_{\text{prior}})}$. This last estimate immediately implies

$$\int_0^T W_1\left(\nu_t^{\boldsymbol{y}}, \nu_t^{\boldsymbol{y}'}\right) \mathrm{d}t \leqslant CT|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma.$$

□

4.4. Consistency results for Bayesian DA

In the present subsection, we discuss the consistency of the approximate filtering problems based on the discretized solution operator $\mathcal{S}_t^\Delta$, and the limiting filtering problem with solution operator $\mathcal{S}_t^\dagger$ (cp theorem 3.12). More precisely, we show that if $\mu_{\text{prior}} \in \mathcal{P}_1(X)$ has finite second moments, if the noise distribution $\rho(y)\mathrm{d}y$ satisfies (N.1)–(N.3) and if the observables $\mathcal{G}_j : L^1([0,T];X) \to \mathbb{R}^d$ are Lipschitz continuous, then for any (approximate) solution operators $\mathcal{S}_t^\dagger, \mathcal{S}_t^\Delta : X \to X$ satisfying a uniform estimate

$$\|\mathcal{S}_t^\dagger(\bar{u})\|_X, \|\mathcal{S}_t^\Delta(\bar{u})\|_X \leqslant B_{\mathcal{S}}\|\bar{u}\|_X, \tag{4.21}$$

we have

$$\int_0^T W_1\left(\nu_t^{\Delta,\boldsymbol{y}}, \nu_t^{\boldsymbol{y}}\right) \mathrm{d}t \leqslant C\int_0^T \|\mathcal{S}_t^\Delta(\bar{u}) - \mathcal{S}_t^\dagger(\bar{u})\|_{L^2(\mu_{\text{prior}})} \mathrm{d}t. \tag{4.22}$$

Before coming to the proof of this claim, we note that, under assumption ($\mathcal{S}$.2) and for Lipschitz continuous observables $\mathcal{G}_j : L^1([0,T];X) \to \mathbb{R}^d$ (cp equation (2.2)), we have for $\mathcal{L}_j^\Delta(\bar{u}) = \mathcal{G}_j(\mathcal{S}^\Delta(\bar{u}))$, $\mathcal{L}_j^\dagger(\bar{u}) = \mathcal{G}_j(\mathcal{S}^\dagger(\bar{u}))$:

$$\begin{aligned} \|\mathcal{L}_j^\Delta(\bar{u})\|_{L^2(\mu_{\text{prior}})}, \|\mathcal{L}_j^\dagger(\bar{u})\|_{L^2(\mu_{\text{prior}})} &\leqslant C\left(1 + \|\bar{u}\|_{L^2(\mu_{\text{prior}})}\right), \\ \|\mathcal{L}_j^\Delta(\bar{u}) - \mathcal{L}_j^\dagger(\bar{u})\|_{L^2(\mu_{\text{prior}})} &\leqslant C\int_{t_{j-1}}^{t_j} \|\mathcal{S}_t^\Delta(\bar{u}) - \mathcal{S}_t^\dagger(\bar{u})\|_{L^2(\mu_{\text{prior}})} \mathrm{d}t, \end{aligned} \tag{4.23}$$

for all $j = 1, \ldots, N$, where $C = C(\mathcal{G}, B_{\mathcal{S}}, T) > 0$ depends only on T, the (Lipschitz continuous) observables $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_N)$, and the boundedness constant $B_{\mathcal{S}}$ in (4.21).

Proof of theorem 3.12. By proposition 3.2 and (3.8), we have

$$\left.\begin{aligned} \nu_t^{\boldsymbol{y}} &= \mathcal{S}_{t,\#}^\dagger \mu^{y_{1:j}}, \\ \nu_t^{\Delta,\boldsymbol{y}} &= \mathcal{S}_{t,\#}^\Delta \mu^{\Delta,y_{1:j}}, \end{aligned}\right\} \quad \forall\ t \in [t_j, t_{j+1}).$$

Given $\Phi \in \mathrm{Lip}$, with $\|\Phi\|_{\mathrm{Lip}} \leqslant 1$ and $\Phi(0) = 0$, we find

$$\begin{aligned}
&\int_X \Phi(u)\Big[\nu_t^{\Delta,y}(\mathrm{d}u) - \nu_t^{y}(\mathrm{d}u)\Big] \\
&\quad = \int_X \Phi(u)\Big[\mathcal{S}^{\Delta}_{t,\#}\mu^{\Delta,y_{1:j}}(\mathrm{d}u) - \mathcal{S}^{\dagger}_{t,\#}\mu^{y_{1:j}}(\mathrm{d}u)\Big] \\
&\quad = \int_X \Phi(u)\big[\mathcal{S}^{\Delta}_{t,\#}\mu^{\Delta,y_{1:j}}(\mathrm{d}u) - \mathcal{S}^{\Delta}_{t,\#}\mu^{y_{1:j}}(\mathrm{d}u)\big] \\
&\qquad + \int_X \Phi(u)\Big[\mathcal{S}^{\Delta}_{t,\#}\mu^{y_{1:j}}(\mathrm{d}u) - \mathcal{S}^{\dagger}_{t,\#}\mu^{y_{1:j}}(\mathrm{d}u)\Big] \\
&\quad =: \mathrm{(I)} + \mathrm{(II)}.
\end{aligned}$$

We can estimate the two last terms individually as follows: for the first term, we obtain

$$\begin{aligned}
\mathrm{(I)} &= \int_X \Phi(u)\big[\mathcal{S}^{\Delta}_{t,\#}\mu^{\Delta,y_{1:j}}(\mathrm{d}u) - \mathcal{S}^{\Delta}_{t,\#}\mu^{y_{1:j}}(\mathrm{d}u)\big] \\
&= \int_X \Phi(\mathcal{S}^{\Delta}_t(\bar{u}))\left[\frac{\mathrm{d}\mu^{\Delta,y_{1:j}}}{\mathrm{d}\mu_{\mathrm{prior}}} - \frac{\mathrm{d}\mu^{y_{1:j}}}{\mathrm{d}\mu_{\mathrm{prior}}}\right]\mu_{\mathrm{prior}}(\mathrm{d}\bar{u}) \\
&\leqslant C\int_X \|\bar{u}\|_X \left|\frac{\mathrm{d}\mu^{\Delta,y_{1:j}}}{\mathrm{d}\mu_{\mathrm{prior}}} - \frac{\mathrm{d}\mu^{y_{1:j}}}{\mathrm{d}\mu_{\mathrm{prior}}}\right|\mu_{\mathrm{prior}}(\mathrm{d}\bar{u}) \\
&\leqslant C\|\bar{u}\|_{L^2(\mu_{\mathrm{prior}})}\left\|\frac{\mathrm{d}\mu^{\Delta,y_{1:j}}}{\mathrm{d}\mu_{\mathrm{prior}}} - \frac{\mathrm{d}\mu^{y_{1:j}}}{\mathrm{d}\mu_{\mathrm{prior}}}\right\|_{L^2(\mu_{\mathrm{prior}})}.
\end{aligned}$$

The last term can be estimated using lemma 4.4, recalling that $\mu^{\Delta,y_{1:j}}$ is defined as the posterior with prior μ_{prior} and given the measurements $\mathcal{L}^{\Delta}_{1:j} = (\mathcal{L}^{\Delta}_1, \ldots, \mathcal{L}^{\Delta}_j)$, with $y_{1:j} = \mathcal{L}^{\Delta}_{1:j} + \eta_{1:j}$ and $\eta_{1:j} = (\eta_1, \ldots, \eta_j) \sim \rho(y_1)\mathrm{d}y_1 \otimes \cdots \otimes \rho(y_j)\mathrm{d}y_j$. Lemma 4.4 therefore yields

$$\begin{aligned}
\left\|\frac{\mathrm{d}\mu^{\Delta,y_{1:j}}}{\mathrm{d}\mu_{\mathrm{prior}}} - \frac{\mathrm{d}\mu^{y_{1:j}}}{\mathrm{d}\mu_{\mathrm{prior}}}\right\|_{L^2(\mu_{\mathrm{prior}})} &\leqslant C\|\mathcal{L}^{\Delta}_{1:j}(\bar{u}) - \mathcal{L}^{\dagger}_{1:j}(\bar{u})\|_{L^2(\mu_{\mathrm{prior}})} \\
&= C\left(\sum_{\ell=1}^{j}\|\mathcal{L}^{\Delta}_{\ell}(\bar{u}) - \mathcal{L}^{\dagger}_{\ell}(\bar{u})\|^2_{L^2(\mu_{\mathrm{prior}})}\right)^{1/2},
\end{aligned}$$

for some constant $C = C(R, T, N, \rho, \|\bar{u}\|_{L^2(\mu_{\mathrm{prior}})}, \mathcal{G}, T) > 0$ depending on R, T, the noise distribution $\rho(y)$, the prior μ_{prior} and on the observables $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_N)$; here, we have used the fact that $y_{1:j}$ is fixed, and that $\|\mathcal{L}^{\Delta}_{\ell}(\bar{u})\|_{L^2(\mu_{\mathrm{prior}})}, \|\mathcal{L}^{\dagger}_{\ell}(\bar{u})\|_{L^2(\mu_{\mathrm{prior}})} \leqslant C(1 + \|\bar{u}\|_{L^2(\mu_{\mathrm{prior}})}) < \infty$ are bounded independently of $\Delta > 0$ for a constant $C = C(\mathcal{G}, T) > 0$ (cp equation (4.23)). The latter observation allows us to bound the additional exponential factor in lemma 4.4 uniformly in Δ. Continuing, we note that the observables are Lipschitz continuous by assumption (4.23); we have

$$\|\mathcal{L}^{\Delta}_{\ell}(\bar{u}) - \mathcal{L}^{\dagger}_{\ell}(\bar{u})\|_{L^2(\mu_{\mathrm{prior}})} \leqslant C\int_{t_{\ell-1}}^{t_\ell}\|\mathcal{S}^{\Delta}_t(\bar{u}) - \mathcal{S}^{\dagger}_t(\bar{u})\|_{L^2(\mu_{\mathrm{prior}})}\,\mathrm{d}t,$$

where $C = C(\mathcal{G}) > 0$. It follows that

$$\text{(I)} \leqslant C\left(\sum_{\ell=1}^{j}\left[\int_{t_{\ell-1}}^{t_\ell}\left\|\mathcal{S}_t^{\Delta}(\bar{u}) - \mathcal{S}_t^{\dagger}(\bar{u})\right\|_{L^2(\mu_{\text{prior}})}\mathrm{d}t\right]^2\right)^{1/2}.$$

Denoting $F(t,\ell) := 1_{[t_{\ell-1},t_\ell)}(t)\|\mathcal{S}_t^{\Delta}(\bar{u}) - \mathcal{S}_t^{\dagger}(\bar{u})\|_{L^2(\mu_{\text{prior}})}$, we can estimate the last term as follows, using Minkowski's integral inequality:

$$C\left(\sum_{\ell=1}^{j}\left[\int_0^T F(t,\ell)\,\mathrm{d}t\right]^2\right)^{1/2} \leqslant C\int_0^T\left(\sum_{\ell=1}^{j}|F(t,\ell)|^2\right)^{1/2}\mathrm{d}t.$$

Finally, recalling that all $F(t,\ell)$, $\ell = 1,\ldots,j$, have *disjoint* supports in t, we conclude that

$$\begin{aligned}\text{(I)} &\leqslant C\int_0^T\left(\sum_{\ell=1}^{j}|F(t,\ell)|^2\right)^{1/2}\mathrm{d}t = C\sum_{\ell=1}^{j}\int_{t_{\ell-1}}^{t_\ell}|F(t,\ell)|\,\mathrm{d}t\\ &\leqslant C\int_0^T\|\mathcal{S}_t^{\Delta}(\bar{u}) - \mathcal{S}_t^{\dagger}(\bar{u})\|_{L^2(\mu_{\text{prior}})}\,\mathrm{d}t.\end{aligned}$$

To estimate the second term, we note that

$$\begin{aligned}&\int_X \Phi(u)\left[\mathcal{S}_{t,\#}^{\Delta}\mu^{y_{1:j}}(\mathrm{d}u) - \mathcal{S}_{t,\#}^{\dagger}\mu^{y_{1:j}}(\mathrm{d}u)\right]\\ &\quad= \int_X\left[\Phi(\mathcal{S}_t^{\Delta}(\bar{u})) - \Phi(\mathcal{S}_t^{\dagger}(\bar{u}))\right]\mu^{y_{1:j}}(\mathrm{d}\bar{u})\\ &\quad\leqslant \int_X\left\|\mathcal{S}_t^{\Delta}(\bar{u}) - \mathcal{S}_t^{\dagger}(\bar{u})\right\|_X\mu^{y_{1:j}}(\mathrm{d}\bar{u})\\ &\quad\leqslant C\int_X\left\|\mathcal{S}_t^{\Delta}(\bar{u}) - \mathcal{S}_t^{\dagger}(\bar{u})\right\|_X\mu_{\text{prior}}(\mathrm{d}\bar{u})\\ &\quad\leqslant C\left\|\mathcal{S}_t^{\Delta}(\bar{u}) - \mathcal{S}_t^{\dagger}(\bar{u})\right\|_{L^2(\mu_{\text{prior}})}.\end{aligned}$$

Thus, employing the above estimates for (I) and (II), we conclude that for any $\Phi \in \text{Lip}$, $\|\Phi\|_{\text{Lip}} \leqslant 1$, and for any $t \in [0,T]$, we have

$$\begin{aligned}\int_X \Phi(u)\left[\nu_t^{\Delta,\boldsymbol{y}}(\mathrm{d}u) - \nu_t^{\boldsymbol{y}}(\mathrm{d}u)\right] &\leqslant C\left\|\mathcal{S}_t^{\Delta}(\bar{u}) - \mathcal{S}_t^{\dagger}(\bar{u})\right\|_{L^2(\mu_{\text{prior}})}\\ &\quad + C\int_0^T\left\|\mathcal{S}_t^{\Delta}(\bar{u}) - \mathcal{S}_t^{\dagger}(\bar{u})\right\|_{L^2(\mu_{\text{prior}})}\mathrm{d}t.\end{aligned}$$

Taking the supremum over all such Φ on the left, and integrating over $t \in [0,T]$, it follows that

$$\int_0^T W_1\left(\nu_t^{\Delta,\boldsymbol{y}}, \nu_t^{\boldsymbol{y}}\right)\mathrm{d}t \leqslant C\int_0^T\left\|\mathcal{S}_t^{\Delta}(\bar{u}) - \mathcal{S}_t^{\dagger}(\bar{u})\right\|_{L^2(\mu_{\text{prior}})}\mathrm{d}t,$$

where $C = C(R,T,N,\rho,\|\bar{u}\|_{L^2(\mu_{\text{prior}})},\mathcal{G}) > 0$ is independent of Δ. □

4.5. Compactness results for Bayesian DA

In the present section, we will prove a compactness result, which applies in particular to the numerical approximations of the Navier–Stokes equations introduced in section 2.3.2. We recall that, by proposition 2.10, numerical solutions computed e.g. by suitable spectral schemes satisfy the following properties:

- ($\Delta.1'$) *Energy bound:* there exists $C > 0$, independent of $\Delta > 0$, such that $\|\mathcal{S}_t^\Delta(\bar{u})\|_{L_x^2} \leqslant \bar{B}_\mathcal{S}\|\bar{u}\|_{L_x^2}$,
- ($\Delta.2$) *Coercivity:* $\int_0^T \|\nabla u^\Delta(t)\|_{L_x^2}^2 \,\mathrm{d}t \leqslant \nu^{-1}\|\bar{u}\|_{L_x^2}^2$,
- ($\Delta.3$) *Weak time-regularity:* there exists $C = C(\|\bar{u}\|), L > 0$, such that $\|u^\Delta(t) - u^\Delta(s)\|_{H_x^{-L}} \leqslant C|t-s|$. In particular, we have $u^\Delta \in \mathrm{Lip}([0,T]; H_x^{-L})$ uniformly in $\Delta > 0$.

When re-stating the first property, we have slightly relaxed ($\Delta.1$), allowing for a uniformly bounded constant $C > 0$ in ($\Delta.1'$).

Our compactness result is motivated by the study of statistical solutions of the compressible and incompressible Euler equations in [10, 23, 24]. There, it is shown that under a suitable average regularity condition, the sequence of discretized approximate solutions $\mu_t^\Delta := \mathcal{S}_{t,\#}^\Delta \mu$ (push-forward by the discretized solution operator) is compact in $\mathcal{P}_1(L_x^2)$, provided that the following measure of average two-point correlations

$$\mathscr{S}_2^T(\mu_t^\Delta; r) := \left(\int_0^T \int_{L_x^2} \mathscr{S}_2(u; r)^2 \, \mu_t^\Delta(\mathrm{d}u)\,\mathrm{d}t \right)^{1/2}, \tag{4.24}$$

are uniformly bounded as $\Delta \to 0$, where

$$\mathscr{S}_2(u; r) := \left(\int_D \fint_{B_r(0)} |u(x+h) - u(x)|^2 \,\mathrm{d}h\,\mathrm{d}x \right)^{1/2}, \tag{4.25}$$

measures the average of two-point correlations of u. The quantity $r \mapsto \mathscr{S}_2^T(\mu_t^\Delta; r)$ is referred to as the (time-integrated) structure function of μ_t^Δ. For simplicity, we will state the following results in the periodic setting with domain $D = \mathbb{T}^n \simeq [0, 2\pi]^n$ the 2π-periodic torus, and with $L_x^2 := L^2(\mathbb{T}^n; \mathbb{R}^n)$ the space of square-integrable vector fields on $\mathbb{T}^n$. More precisely, we state the following proposition, which follows from [24, theorem 2.2]:

Proposition 4.6. *Let μ_t^Δ be of the form $\mu_t^\Delta = \mathcal{S}_{t,\#}^\Delta \mu_{\text{prior}}$, where $\mathcal{S}_t^\Delta : L_x^2 \to L_x^2$ satisfies assumptions ($\Delta.1'$), ($\Delta.3$), and where $\mu_{\text{prior}} \in \mathcal{P}_1(L_x^2)$ has bounded support, i.e. there exists $M > 0$, such that $\mu_{\text{prior}}\left(\left\{\bar{u} \in L_x^2 \mid \|\bar{u}\|_{L_x^2} \leqslant M\right\}\right) = 1$. If there exists a modulus of continuity $\phi(r)$, such that*

$$\sup_\Delta \mathscr{S}_2^T(\mu_t^\Delta; r) \leqslant \phi(r), \quad \forall\ r \geqslant 0,$$

then μ_t^Δ is compact in $L_t^1(\mathcal{P})$.

Numerical evidence for the uniform boundedness of these structure functions $\mathscr{S}_2^T(\mu_t^\Delta; r)$ has been presented for a variety of initial probability measures $\mu_{\text{prior}} \in \mathcal{P}_1(L_x^2)$ for the incompressible Euler equations (i.e. in the zero-viscosity limit of the Navier–Stokes equations) in [23, 24], and in the context of hyperbolic conservation laws in [10]. While these results of [10, 23, 24] were mostly based on empirical observations, in the present case of the Navier–Stokes

equations, we will *rigorously prove* that the structure functions (4.24) are uniformly bounded as $\Delta \to 0$ (cp lemma 4.8, below). We will furthermore extend the compactness result summarized in proposition 4.6 to the Bayesian filtering context, when the underlying model is combined with additional measurements.

We formulate the numerically observed [10, 23, 24] bounds on $\mathscr{S}_2^T(\mu_t^\Delta; r) \leqslant \phi(r)$ abstractly as the following assumption:

Assumption 4.7. The prior $\mu_{\text{prior}} \in \mathcal{P}_1(L_x^2)$ has bounded support, i.e. there exists $M > 0$, such that

$$\mu_{\text{prior}}\left(\left\{\bar{u} \in L_x^2 \,|\, \|\bar{u}\|_{L_x^2} \leqslant M\right\}\right) = 1,$$

and there exists a modulus of continuity $\phi(r)$, such that

$$\mathscr{S}_2^T(\mathcal{S}_{t,\#}^\Delta \mu_{\text{prior}}; r) \leqslant \phi(r), \quad \forall\ r > 0,\ t \in [0, T], \tag{4.26}$$

uniformly for all $\Delta > 0$. Here $\mathcal{S}_{t,\#}^\Delta \mu_{\text{prior}}$ denotes the push-forward measure of the prior μ_{prior} by the discretized solution operator $\mathcal{S}_t^\Delta$.

The next proposition shows that the structure function bound (4.26) is automatically satisfied for numerical schemes satisfying (Δ.2):

Lemma 4.8. *Let $\mu_{\text{prior}} \in \mathcal{P}_1(L_x^2)$ be a prior for the incompressible Navier–Stokes equations (2.8), such that $\int_{L_x^2} \|\bar{u}\|_{L_x^2}^2 \mu_{\text{prior}}(\mathrm{d}\bar{u}) < \infty$. Let $\mathcal{S}_t^\Delta : L_x^2 \to L_x^2$ denote the approximate solution operator obtained from the spectral scheme (2.9). Then we have the following* a priori *structure function bound:*

$$\mathscr{S}_2^T(\mathcal{S}_{t,\#}^\Delta \mu_{\text{prior}}; r) \leqslant \frac{r}{\sqrt{2\nu}} \left(\int_{L_x^2} \|\bar{u}\|_{L_x^2}^2 \mu_{\text{prior}}(\mathrm{d}\bar{u}) \right)^{1/2}.$$

In particular, $\mathscr{S}_2^T(\mathcal{S}_{t,\#}^\Delta \mu_{\text{prior}}; r) \leqslant Cr$ is uniformly bounded by a modulus of continuity as $\Delta \to 0$.

Proof. By definition, we have

$$\mathscr{S}_2^T(\mathcal{S}_{t,\#}^\Delta \mu_{\text{prior}}; r)^2 = \int_{L_x^2} \left(\int_0^T \mathscr{S}_2(\mathcal{S}_t^\Delta(\bar{u}); r)^2 \,\mathrm{d}t \right) \mu_{\text{prior}}(\mathrm{d}\bar{u}),$$

where, setting $u^\Delta = \mathcal{S}_t^\Delta(\bar{u})$,

$$\begin{aligned}
\mathscr{S}_2(u^\Delta; r)^2 &= \int_{\mathbb{T}^d} \fint_{B_r(0)} |u^\Delta(x+h) - u^\Delta(x)|^2 \,\mathrm{d}h \,\mathrm{d}x \\
&= \fint_{B_r(0)} \|u^\Delta(\cdot + h) - u^\Delta(\cdot)\|_{L_x^2}^2 \,\mathrm{d}h \\
&\leqslant \fint_{B_r(0)} \|\nabla u^\Delta\|_{L_x^2}^2 |h|^2 \,\mathrm{d}h \\
&\leqslant \|\nabla u^\Delta\|_{L_x^2}^2 r^2.
\end{aligned}$$

By ($\Delta.2$), with $u^\Delta(t) = \mathcal{S}_t^\Delta(\bar{u})$, we have

$$\int_0^T \|\nabla u^\Delta\|_{L_x^2}^2 \,\mathrm{d}t \leqslant \frac{1}{2\nu}\|\bar{u}\|_{L_x^2}^2.$$

Thus, from the above estimates, we conclude that

$$\int_0^T \mathscr{S}_2(\mathcal{S}_t^\Delta(\bar{u}); r)^2 \,\mathrm{d}t = \int_0^T \mathscr{S}_2(u^\Delta; r)^2 \,\mathrm{d}t \leqslant r^2 \int_0^T \|\nabla u^\Delta\|_{L_x^2}^2 \,\mathrm{d}t \leqslant \frac{r^2}{2\nu}\|\bar{u}\|_{L_x^2}^2.$$

Integration against $\mu_{\text{prior}}(\mathrm{d}\bar{u})$ yields

$$\mathscr{S}_2^T(\mathcal{S}_{t,\#}^\Delta \mu_{\text{prior}}; r)^2 \leqslant \frac{r^2}{2\nu}\int_{L_x^2} \|\bar{u}\|_{L_x^2}^2 \,\mu_{\text{prior}}(\mathrm{d}\bar{u}),$$

as claimed. □

Conditional on assumption 4.7, we can now prove a compactness result for the filtering problem.

Lemma 4.9. *Let $\mathcal{S}_t^\Delta : L_x^2 \to L_x^2$ be a sequence of approximate solution operators satisfying ($\Delta.1'$)–($\Delta.3$). Let $\nu_t^{\Delta,\mathbf{y}}$ be the solution of the associated filtering problem with prior $\mu_{\text{prior}} \in \mathcal{P}_1(L_x^2)$ having bounded support, and measurements $\mathbf{y} = (y_1, \ldots, y_N)$. If assumption 4.7 holds, then $\nu_t^{\Delta,\mathbf{y}}$ is a compact sequence in $L_t^1(\mathcal{P})$, as $\Delta \to 0$.*

Proof. We observe that the structure function can be written as

$$\begin{aligned}
\mathscr{S}_2^T\left(\nu_t^{\Delta,\mathbf{y}}; r\right)^2 &= \int_0^T \int_{L_x^2} \mathscr{S}_2(u; r)^2 \nu_t^{\Delta,\mathbf{y}}(\mathrm{d}u)\mathrm{d}t \\
&= \sum_{k=1}^N \int_{t_{k-1}}^{t_k} \int_{L_x^2} \mathscr{S}_2(u; r)^2 \nu_t^{\Delta,\mathbf{y}}(\mathrm{d}u)\mathrm{d}t \\
&= \sum_{k=1}^N \int_{t_{k-1}}^{t_k} \int_{L_x^2} \mathscr{S}_2(u; r)^2 \left[\mathcal{S}_{t,\#}^\Delta \mu^{\Delta,y_{1:(k-1)}}\right](\mathrm{d}u)\mathrm{d}t \\
&= \sum_{k=1}^N \int_{t_{k-1}}^{t_k} \int_{L_x^2} \mathscr{S}_2(\mathcal{S}_t^\Delta(\bar{u}); r)^2 \mu^{\Delta,y_{1:(k-1)}}(\mathrm{d}\bar{u})\mathrm{d}t.
\end{aligned}$$

We recall that on the last line, $\mu^{\Delta,y_{1:(k-1)}}$ is the solution of the BIP (3.3). Using the uniform boundedness result for such BIP, lemma 4.2, we conclude that

$$\begin{aligned}
\frac{\mathrm{d}\mu^{\Delta,y_{1:(k-1)}}}{\mathrm{d}\mu_{\text{prior}}} &\leqslant C \exp\left(\sum_{j=1}^{k-1} |y_j|_\Gamma^2 + \sum_{j=1}^{k-1} \|\mathcal{L}_j^\Delta(\bar{u})\|_{L^2(\mu_{\text{prior}})}^2\right) \\
&\leqslant C \exp\left(|\mathbf{y}|_\Gamma^2 + \sum_{j=1}^{N} \|\mathcal{L}_j^\Delta(\bar{u})\|_{L^2(\mu_{\text{prior}})}^2\right).
\end{aligned}$$

By our boundedness assumption ($\Delta.1'$), and the upper bound on observables (cp equation (4.23)), it follows that there exists a constant $C = C(\mathrm{Lip}(\mathcal{G}_j), B) > 0$, such that

$$\|\mathcal{L}_j^{\Delta}(\bar{u})\|_{L^2(\mu_{\text{prior}})} = C\Big(1 + \|\bar{u}\|_{L^2(\mu_{\text{prior}})}\Big) < \infty,$$

is uniformly bounded. Since $\boldsymbol{y}$ is fixed, we conclude that there exists a constant $C > 0$, such that

$$\frac{\mathrm{d}\mu^{\Delta,y_{1:(k-1)}}}{\mathrm{d}\mu_{\text{prior}}} \leqslant C,$$

and hence

$$\begin{aligned}
\mathscr{S}_2^T\Big(\nu_t^{\Delta,\boldsymbol{y}}; r\Big)^2 &= \sum_{k=1}^{N} \int_{t_{k-1}}^{t_k} \int_{L_x^2} \mathscr{S}_2(\mathcal{S}_t^{\Delta}(\bar{u}); r)^2 \mu^{\Delta,y_{1:(k-1)}}(\mathrm{d}\bar{u})\,\mathrm{d}t \\
&\leqslant \sum_{k=1}^{N} \int_{t_{k-1}}^{t_k} \int_{L_x^2} \mathscr{S}_2(\mathcal{S}_t^{\Delta}(\bar{u}); r)^2 C\mu_{\text{prior}}(\mathrm{d}\bar{u})\mathrm{d}t \\
&= C\sum_{k=1}^{N} \int_{t_{k-1}}^{t_k} \int_{L_x^2} \mathscr{S}_2(u; r)^2 \big[\mathcal{S}_{t,\#}^{\Delta}\mu_{\text{prior}}\big](\mathrm{d}u)\,\mathrm{d}t \\
&= C\int_0^T \int_{L_x^2} \mathscr{S}_2(u; r)^2 \big[\mathcal{S}_{t,\#}^{\Delta}\mu_{\text{prior}}\big](\mathrm{d}u)\mathrm{d}t \\
&= C\mathscr{S}_2^T\big(\mathcal{S}_{t,\#}^{\Delta}\mu_{\text{prior}}; r\big)^2 \\
&\leqslant C\phi(r)^2.
\end{aligned}$$

The last estimate follows from assumption 4.7. Thus, $\mathscr{S}_2^T(\nu_t^{\Delta,\boldsymbol{y}}; r) \leqslant \sqrt{C}\phi(r)$ is uniformly bounded by a modulus of continuity, implying compactness in $L_t^1(\mathcal{P})$, by proposition 4.6. □

Combining the uniform stability result of theorem 3.8 (applied to the numerical approximations $\mathcal{S}_t^{\Delta}$) with lemma 4.9 and the *a priori* structure function estimate of lemma 4.8, we can now prove theorem 3.15 on the compactness of the approximate filtering distributions for the Navier–Stokes equations.

Proof of theorem 3.15. We first note that by ($\Delta.1'$), there exists a constant $\bar{B}_{\mathcal{S}} > 0$, such that *uniformly in* $\Delta > 0$, we have $\|\mathcal{S}_t^{\Delta}(\bar{u})\|_{L_x^2} \leqslant \bar{B}_{\mathcal{S}}\|\bar{u}\|_{L_x^2}$, for all $\bar{u} \in L_x^2$. By theorem 3.8, applied to the forward operator $\mathcal{S}_t^{\Delta}$, the mapping

$$\mathbb{R}^{d\times N} \ni \boldsymbol{y} \mapsto \nu_t^{\Delta,\boldsymbol{y}} \in L_t^1(\mathcal{P}),$$

is uniformly bounded on any compact subsets $K \subset \mathbb{R}^{d\times N}$ and uniformly equicontinuous on K; indeed, there exists $R > 0$, such that any $\boldsymbol{y} \in K$ satisfies $|\boldsymbol{y}|_{\Gamma} \leqslant R$. By theorem 3.8, there exists a constant $C = C(R, \rho, N, \bar{B}_{\mathcal{S}}, \|\bar{u}\|_{L^1(\mu_{\text{prior}})}) > 0$, independent of $\Delta > 0$, such that for any $t \in [0, T]$, we have

$$W_1(\nu_t^{\Delta,\boldsymbol{y}}, \nu_t^{\Delta,\boldsymbol{y}'}) \leqslant C|\boldsymbol{y} - \boldsymbol{y}'|_{\Gamma}, \quad \forall\ |\boldsymbol{y}|_{\Gamma}, |\boldsymbol{y}'|_{\Gamma} \leqslant R. \tag{4.27}$$

We note that in the present case, under the assumptions of theorem 3.15, we actually have $\bar{B}_{\mathcal{S}} = 1$, and $\|\bar{u}\|_{L^1(\mu_{\text{prior}})} \leqslant M$. Hence the constant $C = C(R, \rho, N, M)$ depends only on R, the noise distribution ρ, the number of measurements N and on the upper bound on the support M. An immediate consequence of (4.27) is that

$$\int_0^T W_1(\nu_t^{\Delta,\boldsymbol{y}}, \nu_t^{\Delta,\boldsymbol{y}'})\mathrm{d}t \leqslant CT|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma, \quad \forall\ |\boldsymbol{y}|_\Gamma, |\boldsymbol{y}'|_\Gamma \leqslant R.$$

Furthermore, by lemma 4.9, the sets

$$\left\{\nu_t^{\Delta,\boldsymbol{y}} \,|\, \Delta > 0\right\} \subset L_t^1(\mathcal{P}),$$

are pre-compact for any fixed $\boldsymbol{y} \in \mathbb{R}^{d\times N}$ (pointwise compactness). By the Arzelá–Ascoli theorem A.1, the claimed compactness result follows. □

We finally would like to emphasize that while we have chosen the incompressible Navier–Stokes equations as our main prototypical example of ill-posed problems arising in fluid mechanics, the results of the present section apply to more general models, as will be briefly indicated next.

4.6. Comment on related models: hyperbolic systems of conservation laws

The results of this work apply, for example, to the numerical approximation of BIPs for hyperbolic systems of conservation laws. Again, we take as our domain $D = \mathbb{T}^n = [0, 2\pi]^n$ with periodic boundary conditions. We recall that a system of conservation laws is a PDE of the form

$$\partial_t u^i + \sum_{j=1}^n \partial_j\big(F^{ij}(u)\big) = 0, \tag{4.28}$$

describing the temporal evolution of m conserved quantities $u^1, \ldots, u^m : D \times [0, T] \to \mathbb{R}$, and $F^{ij} : \mathbb{R}^m \to \mathbb{R}$ are the fluxes. It is convenient to write the system (4.28) in the succinct form

$$\partial_t u + \mathrm{div}(F(u)) = 0,$$

where $u = (u^1, \ldots, u^m) : D \times [0, T] \to \mathbb{R}^m$ and $F = (F^{ij}) : \mathbb{R}^m \to \mathbb{R}^{m\times n}$. The system of conservation laws (4.28) is called hyperbolic, provided that the Jacobian $D_u(F \cdot n)$ possesses real eigenvalues for all unit vectors $n \in \mathbb{R}^n$ with $|n| = 1$. A great variety of systems in continuum mechanics can be formulated as hyperbolic systems of conservation laws, including the compressible Euler equations of gas dynamics, the shallow water equations of oceanography, the magneto-hydro-dynamics equations of plasma physics, and the equations of nonlinear elastodynamics [7].

As it is well-known, even in the special case of a scalar conservation law ($m = 1$), weak solutions to (4.28) are not necessarily unique. It is therefore necessary to augment hyperbolic conservation laws (4.28) with additional entropy, or admissibility conditions. These entropy conditions are based on the existence of an entropy/entropy-flux pair (η, q) consisting of a convex function $\eta : \mathbb{R}^m \to \mathbb{R}$ and a flux $q : \mathbb{R}^m \to \mathbb{R}^n$, such that

$$D_u q = D_u \eta \cdot D_u F.$$

Here, $D_u q$, $D_u \eta$ denote the Jacobian matrices of $q(u)$ and $\eta(u)$. A weak solution u of (4.28) is called an *entropy* weak solution, provided that, in addition to (4.28), also

$$\partial_t \eta(u) + \operatorname{div}(q(u)) \leqslant 0, \tag{4.29}$$

holds in the sense of distributions.

In the following we will restrict our attention to hyperbolic systems of conservation laws for which

$$\|D^2 F\|_{L^\infty} < \infty,$$

and which admit a coercive, smooth flux function $\eta(u) \in C^2(\mathbb{R}^m)$, in the sense that there exist constants $c, C > 0$, such that

$$c \leqslant \left(D^2\eta(u)v, v\right) \leqslant C, \quad \forall\, u, v \in \mathbb{R}^m, \text{ with } |v| = 1.$$

Note that in this case, the entropy admissibility condition (4.29) implies, upon integration over D, an *a priori* bound of the form

$$\|u(t)\|_{L^2} \leqslant B_{\mathcal{S}} \|\overline{u}\|_{L^2},$$

for any admissible weak solution u with initial data $u(t=0) = \overline{u}$.

4.6.1. Numerical methods. In the context of systems of conservation laws, a popular method of choice are finite volume and finite difference methods, as e.g. employed in the numerical experiments for statistical solutions of [10]. We briefly review the form of these numerical schemes, following [10, section 4.1]. For a more complete review, we refer to e.g. [15, 29].

The computational spatial domain is discretized by a collection of cells

$$\{(x^1_{i^1-1/2}, x^1_{i^1+1/2}) \times \cdots \times (x^n_{i^n-1/2}, x^n_{i^n+1/2})\}_{(i^1,\ldots,i^n)},$$

with corresponding cell midpoints

$$x_{i^1,\ldots,i^n} = \left(\frac{x^1_{i^1+1/2} + x^1_{i^1-1/2}}{2}, \ldots, \frac{x^n_{i^n+1/2} + x^n_{i^n-1/2}}{2},\right).$$

We assume that the mesh is equidistant, i.e. for some $\Delta > 0$ we have

$$x^k_{i^k+1/2} - x^k_{i^k-1/2} \equiv \Delta, \quad \forall\, k = 1, \ldots, n.$$

For $\mathbf{i} = (i^1, \ldots, i^n)$, we denote the averaged value in the cell at time $t \geqslant 0$ by $u^\Delta_{\mathbf{i}}(t) = u^\Delta_{(i^1,\ldots,i^n)}(t)$. We consider the following semi-discrete scheme

$$\frac{\mathrm{d}}{\mathrm{d}t} u^\Delta_{i^1,\ldots,i^n}(t) + \sum_{k=1}^{n} \frac{1}{\Delta}\left(F^{k,\Delta}(u^\Delta_{\mathbf{i}-(q-1)\mathbf{e}_k}(t), \ldots, u^\Delta_{\mathbf{i}+q\mathbf{e}_k}(t)) \right. \tag{4.30}$$

$$\left. - F^{k,\Delta}(u^\Delta_{\mathbf{i}-q\mathbf{e}_k}(t), \ldots, u^\Delta_{\mathbf{i}+(q-1)\mathbf{e}_k}(t))\right) = 0, \tag{4.31}$$

and $u^\Delta_{i^1,\ldots,i^n}(0) = \overline{u}(x_{i^1,\ldots,i^n})$. Here, $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are the canonical unit vectors in $\mathbb{R}^n$. $F^{k,\Delta}$ denotes the numerical flux function in direction $k = 1, \ldots, n$, and $\overline{u}_{\mathbf{i}} = \fint_{C_{\mathbf{i}}} u_0(x)\mathrm{d}x$ is the average of the initial data over the $\mathbf{i}$th cell.

As in [10], we make the following assumptions on the discretization (4.31):

Assumption 4.10. We assume that the finite volume scheme (4.31) is consistent in the sense that there exists a constant $C > 0$ such that for $k = 1, \dots, d$,

$$|F^{k,\Delta}(u^{\Delta}_{\mathbf{i}-(q-1)\mathbf{e}_k}, \dots, u^{\Delta}_{\mathbf{i}+q\mathbf{e}_k}) - f^k(u^{\Delta}_{\mathbf{i}})| \leqslant C \sum_{j=-q+1}^{q} |u^{\Delta}_{\mathbf{i}}(t) - u^{\Delta}_{\mathbf{i}+j\mathbf{e}_k}(t)|$$

and the discretized solutions satisfy

(a) L^2 bound: there exists $C > 0$ such that

$$\Delta^d \sum_{\mathbf{i}} |u^{\Delta}_{\mathbf{i}}(t)|^2 \leqslant C\Delta^d \sum_{\mathbf{i}} |\overline{u}_{\mathbf{i}}|^2,$$

(b) weak BV bound: there exists $s \geqslant 2$, such that

$$\Delta^d \int_0^T \sum_{k=1}^{d} \sum_{\mathbf{i}} |u^{\Delta}_{\mathbf{i}+\mathbf{e}_k}(t) - u^{\Delta}_{\mathbf{i}}(t)|^s \, \mathrm{d}t \leqslant C\Delta,$$

with the constant $C = C(\|\overline{u}\|_{L^2_x})$ depending only on the L^2-norm of the initial data.

Remark 4.11. It is not difficult to see that, under assumption 4.10, the discrete solution operator $S^{\Delta}_t : L^2_x \to L^2_x$, which is obtained by locally constant reconstruction (or suitable higher-order variants),

$$S^{\Delta}_t(\overline{u}) := \sum_{\mathbf{i}} u^{\Delta}_{\mathbf{i}}(t) 1_{C_{\mathbf{i}}},$$

satisfies the assumptions ($\Delta.1'$) and ($\Delta.3$) of section 4.5. Furthermore, as pointed out in [10, remark 4.2], many examples of finite volume/difference schemes can be shown to satisfy assumption 4.10. Examples include the so-called entropy stable Lax–Wendroff schemes and the TeCNO schemes of [11].

Based on the results of sections 4.2 and 4.5, we immediately obtain:

Result 4.12. Assume that the FV scheme (4.31) satisfies assumption 4.10. If $\mu_{\text{prior}} \in \mathcal{P}_1(L^2_x)$, then the posteriors $\mu^{\Delta,y}$ of the BIP (4.2) for the hyperbolic conservation law are uniformly stable in y, for any $\Delta > 0$; i.e., for any $R > 0$ there exists a constant $C > 0$, independent of Δ, such that

$$W_1(\mu^{\Delta,y}, \mu^{\Delta,y'}) \leqslant C|y - y'|_{\Gamma}, \quad \forall\ |y|_{\Gamma}, |y'|_{\Gamma} \leqslant R.$$

Furthermore, the posteriors $\mu^{\Delta,y}$ form a compact sequence in $\mathcal{P}_1(L^2_x)$.

This result complements recent work in [31], where (unconditional) convergence was shown for the case of scalar conservation laws ($m = 1$). For the filtering problem, we have:

Result 4.13. Assume that the FV scheme (4.31) satisfies assumption 4.10. Then the approximate solutions $\nu^{\Delta,\boldsymbol{y}}_t$ of the filtering problem computed by the FV scheme are uniformly stable with respect to the measurements $\boldsymbol{y}$, in the sense of (3.15), for any $\Delta > 0$. In addition, if

the prior $\mu_{\text{prior}} \in \mathcal{P}_1(L^2_x)$ satisfies assumption 4.7, then the posteriors $\nu_t^{\Delta,\boldsymbol{y}}$ form a compact sequence in $L^1_t(\mathcal{P})$.

Remark 4.14. The validity of assumption 4.7 has been investigated for a diverse set of initial priors in [10]. The numerical evidence, presented in [10] strongly suggest that it is fulfilled for the cases considered there, and we conjecture that the structure functions (4.26) are uniformly bounded for a wide range of priors of practical relevance.

Remark 4.15. The current section has been formulated for hyperbolic systems of convergence laws with a strictly convex entropy. The main reason for this restriction is that the results of section 4.5, and the compactness proof of [24] are based on the L^2-framework that is natural in the context of the incompressible Euler equations. However, there should be no essential difficulty in extending these results to L^p-based spaces for $p \neq 2$.

5. Discussion

The Bayesian framework has been well-established as a suitable formulation of inverse problems arising in the context of PDEs (1.1) [36]. The well-posedness of the BIP has been demonstrated under very mild assumptions on the well-posedness of the *forward problem* for the underlying PDE, requiring essentially only the existence and uniqueness of solutions defined on an infinite-dimensional Banach space [25, 35]. Corresponding well-posedness results for DA have focused mostly on finite-dimensional problems with Gaussian noise and when the solution operator is continuous [26].

However, for a large numbers of PDEs, such as the fundamental equations of fluid dynamics, the forward problem may not be well-posed. Existence, uniqueness or stability of solutions are either not true or can not be established rigorously. This issue is further exacerbated by the fact that for many of these PDEs, numerical approximations either may not converge on mesh refinement or converge too slowly to be useful. This is often a result of the sensitivity of solutions to small perturbations and the appearances of structures at smaller and smaller scales, as the grid is refined [9, 10, 14].

Our main aim in this paper was to investigate Bayesian DA (filtering) for such PDEs with a very unstable or even *ill-posed forward problem*. Our main results, summarized in section 3, concern the properties of the time-dependent filtering distribution $\nu_t^{\boldsymbol{y}}$ (exact posterior based on the underlying 'ground truth' map $\mathcal{S}_t^\dagger$) and numerical approximations $\nu_t^{\Delta,\boldsymbol{y}}$ based on a approximate solution operator $\mathcal{S}_t^\Delta$, with $\Delta > 0$ a discretization parameter (grid size) and $\boldsymbol{y} = (y_1, \dots, y_N)$ being finite-dimensional (noisy) measurements acquired over time $t \in [0, T]$. We were able to show:

- **(Well-posedness).** We prove that Bayesian filtering is *well-posed* under very general assumptions on the forward solution operator $\mathcal{S}_t^\dagger : X \to X$; in particular, we show that the measurement-to-posterior mapping $\boldsymbol{y} \mapsto \nu_t^{\boldsymbol{y}}$ is locally Lipschitz continuous, even if the forward mapping $\bar{u} \mapsto \mathcal{S}_t^\dagger(\bar{u})$ is discontinuous (cp theorem 3.8).
- **(Consistency).** We prove that approximations $\nu_t^{\Delta,\boldsymbol{y}} \approx \nu_t^{\boldsymbol{y}}$ of the filtering distribution, e.g., arising from numerical discretization of the underlying PDE at mesh size $\Delta > 0$, *converge* to the exact posterior as $\Delta \to 0$, provided that the approximate solution operators $\mathcal{S}_t^\Delta \to \mathcal{S}_t^\dagger$ converge only in a mean-square sense relative to the prior, (cp theorem 3.12).

- **(Compactness and stability).** We demonstrate that even in the absence of any convergence-guarantees of the approximate solution operators $\mathcal{S}_t^\Delta : X \to X$, the corresponding approximate Bayesian filtering distributions $\{\mu^{\Delta,y}\}_{\Delta>0}$ form a *compact sequence* under mild conditions (satisfied e.g. by the Navier–Stokes equations), and hence possess limit points, (cp theorem 3.15).

The well-posedness results in the context of Bayesian DA presented in this work, even for models for which the forward problem may be ill-posed, have been derived under mild assumptions and are applicable to a wide range of models encountered in practice. The stability results should be of particular significance to practitioners, as they demonstrate that under readily verifiable conditions on the numerical scheme, the approximate solutions of the DA problem are stable with respect to perturbations of the measurements, *independently of the numerical resolution and physical parameters such as the viscosity*.

Our consistency results do not only imply the convergence of the filtering distributions $\nu_t^{\Delta,y} \to \nu_t^y$ given convergent approximations of the forward problem $\mathcal{S}_t^\Delta \to \mathcal{S}_t^\dagger$, but they also provide *quantitative error bounds* on the (Wasserstein-) distance between the exact and approximate filtering distributions. These upper bounds are obtained in terms of the mean-square distance between the exact and approximate solution operators with respect to the prior. Such quantitative estimates are not only of importance in studying the convergence of Bayesian filtering based on traditional numerical discretizations, but also open up the possibility of deriving similar quantitative bounds for Bayesian DA based on novel neural network-based operator learning frameworks such as [30], extending the work [20] to the Bayesian filtering context (cp remark 3.14).

Finally, the general compactness properties uncovered in the present work allow us to define a set of candidate solutions to the filtering problem, generated by suitable numerical schemes. As this set can be shown to be non-empty *a priori*, this potentially opens up the possibility of identifying the correct solution among these candidates by a suitable selection criterion to single out a 'canonical' posterior amongst the set of candidate solutions. We propose to further investigate these questions in forthcoming work.

Acknowledgment

The research of SM and SL is partially supported by the European Research Council Consolidator Grant ERC-COG 770880 COMANFLO. FW is partially supported by NSF DMS 2042454 and NSF OIA-DMR 2021019.

Data availability statement

No new data were created or analysed in this study.

Appendix A. Mathematical complements

A.1. L^p norms

Let X be a separable Banach space, and let $\mu_{\text{prior}} \in \mathcal{P}(X)$ be a probability measure. We introduce the $L^p(\mu_{\text{prior}})$-norm ($p \in [1, \infty]$) of a Borel measurable mapping $\mathcal{F} : X \to Y$, $\bar{u} \mapsto \mathcal{F}(\bar{u})$, with

X, Y Banach space, as follows:

$$\|\mathcal{F}\|_{L^p(\mu_{\text{prior}})} := \begin{cases} \left(\int_X \|\mathcal{F}(\bar{u})\|_Y^p \mu_{\text{prior}}(\mathrm{d}\bar{u}) \right)^{1/p}, & (p < \infty), \\ \operatorname*{ess\,sup}_{\bar{u} \sim \mu_{\text{prior}}} \|\mathcal{F}(\bar{u})\|_Y, & (p = \infty), \end{cases} \tag{A.1}$$

where we recall that for $p = \infty$, the essential supremum is defined by

$$\operatorname*{ess\,sup}_{\bar{u} \sim \mu_{\text{prior}}} \|\mathcal{F}(\bar{u})\|_Y := \sup\{M > 0 \,|\, \mu_{\text{prior}}(\{\bar{u} | \|\mathcal{F}(\bar{u})\|_X > M\}) > 0\}.$$

In particular, if $Y = X$ and if $X \to X$, $\bar{u} \mapsto \mathcal{F}(\bar{u}) = \bar{u}$ is given by the identity mapping, then we have (for $p < \infty$):

$$\|\bar{u}\|_{L^p(\mu_{\text{prior}})} = \left(\int_X \|\bar{u}\|_X^p \mu_{\text{prior}}(\mathrm{d}\bar{u}) \right)^{1/p}.$$

A.2. Wasserstein distance

In this section, we introduce the notation for the rest of the paper and recall some preliminaries that are necessary to define the BIP in a mathematically precise manner.

Given a separable Hilbert space X, we denote by $\mathcal{P}(X)$ the space of Borel probability measures on X. The term 'measurable' will always refer to Borel measurability. A sequence $\mu^\Delta \in \mathcal{P}(X)$ is said to converge weakly to a limit μ, denoted $\mu^\Delta \rightharpoonup \mu$, if

$$\int_X \phi(u) \mu^\Delta(\mathrm{d}u) \to \int_X \phi(u) \mu(\mathrm{d}u), \qquad \forall\ \phi \in C_{\text{b}}(X),$$

where $C_{\text{b}}(X)$ denotes the space of bounded, continuous functions on X. We denote by $\mathcal{P}_p(X)$ the space of Borel probability measures $\mu \in \mathcal{P}(X)$, possessing finite pth moments, $\int_X \|u\|_X^p \mu(\mathrm{d}u) < \infty$, metrized by the p-Wasserstein distance W_p:

$$W_p(\mu, \nu) := \inf_{\pi \in \Gamma(\mu,\nu)} \left(\int_{X \times X} \|u - v\|_X^p \pi(\mathrm{d}u, \mathrm{d}v) \right)^{1/p}. \tag{A.2}$$

Here, $\Gamma(\mu, \nu)$ is the set of couplings between μ and ν, i.e. probability measures π on $X \times X$, with projections $(\text{Proj}_1)_\# \pi = \mu$, $(\text{Proj}_2)_\# \pi = \nu$. Given a map $F : X \to Y$, we denote by $F_\# \mu \in \mathcal{P}(Y)$ the push-forward of a probability measure $\mu \in \mathcal{P}(X)$ by F; the push-forward measure satisfies the relation

$$\int_Y \phi(v) \left(F_\# \mu\right)(\mathrm{d}v) = \int_X (\phi \circ F)(u) \mu(\mathrm{d}u),$$

for all measurable functions $\phi : Y \to \mathbb{R}$ such that $\phi \circ F \in L^1(\mu)$. We recall that the one-Wasserstein distance $W_1(\mu, \nu)$ between measures $\mu, \nu \in \mathcal{P}_1(X)$ can also be determined via the Kantorovich duality:

$$W_1(\mu, \nu) = \sup_{\Phi} \int_X \Phi(u)[\mu(\mathrm{d}u) - \nu(\mathrm{d}u)], \tag{A.3}$$

where the supremum is taken over all Lipschitz continuous $\Phi \in \mathrm{Lip}(X)$, with $\|\Phi\|_{\mathrm{Lip}} \leqslant 1$, and we define the semi-norm $\|\cdot\|_{\mathrm{Lip}}$ by

$$\|\Phi\|_{\mathrm{Lip}} := \sup_{u \neq v} \frac{|\Phi(u) - \Phi(v)|}{\|u - v\|_X}. \tag{A.4}$$

We also recall that for a sequence of measures $\mu^\Delta \in \mathcal{P}_1(X)$, $\Delta \to 0$, and $\mu \in \mathcal{P}_1(X)$, we have

$$\lim_{\Delta \to 0} W_1(\mu^\Delta, \mu) = 0 \;\Leftrightarrow\; \left\{ \begin{array}{l} \mu^\Delta \rightharpoonup \mu \text{ converges weakly and} \\ \displaystyle\int_X \|u\|_X \mu^\Delta(\mathrm{d}u) \to \int_X \|u\|_X \mu(\mathrm{d}u). \end{array} \right\} \tag{A.5}$$

We finally prove that if $\mathcal{F} : X \to X$ is a Lipschitz continuous map and $\mathcal{F}_\# : \mathcal{P}_1(X) \to \mathcal{P}_1(X)$ denotes the push-forward under $\mathcal{F}$, then $\mathrm{Lip}(\mathcal{F} : X \to X) = \mathrm{Lip}(\mathcal{F}_\# : \mathcal{P}_1(X) \to \mathcal{P}_1(X))$, which implies the claim of proposition 3.4, for $\mathcal{F} = \mathcal{S}_t^\dagger$.

Proof of proposition 3.4. Let $\mathcal{F} : X \to X$ be any Lipschitz continuous map (in particular, the following applies to $\mathcal{F} = \mathcal{S}_t^\dagger : X \to X$). By definition, we have

$$\begin{aligned} W_1(\mathcal{F}_\#\mu, \mathcal{F}_\#\mu') &= \inf_{\pi \in \Gamma(\mathcal{F}_\#\mu, \mathcal{F}_\#\mu')} \int_{X\times X} \|u - u'\|_X \, \pi(\mathrm{d}u, \mathrm{d}u') \\ &\leqslant \inf_{\pi \in \Gamma(\mu,\mu')} \int_{X\times X} \|u - u'\|_X \, (\mathcal{F} \times \mathcal{F})_\# \pi(\mathrm{d}u, \mathrm{d}u') \\ &= \inf_{\pi \in \Gamma(\mu,\mu')} \int_{X\times X} \|\mathcal{F}(u) - \mathcal{F}(u')\|_X \, \pi(\mathrm{d}u, \mathrm{d}u') \\ &\leqslant \mathrm{Lip}(\mathcal{F}) \inf_{\pi \in \Gamma(\mu,\mu')} \int_{X\times X} \|u - u'\|_X \, \pi(\mathrm{d}u, \mathrm{d}u') \\ &= \mathrm{Lip}(\mathcal{F}) W_1(\mu, \mu'). \end{aligned}$$

To see the optimality of $\mathrm{Lip}(\mathcal{F})$, we note that for any $\widetilde{L} > 0$, such that $W_1(\mathcal{F}_\#\mu, \mathcal{F}_\#\mu') \leqslant \widetilde{L} W_1(\mu, \mu')$ holds for all $\mu,\ \mu'$, we have

$$\begin{aligned} \widetilde{L} &\geqslant \sup_{\mu \neq \mu' \in \mathcal{P}_1(X)} \frac{W_1(\mathcal{F}_\#\mu, \mathcal{F}_\#\mu')}{W_1(\mu, \mu')} \geqslant \sup_{u \neq u' \in X} \frac{W_1(\mathcal{F}_\#\delta_u, \mathcal{F}_\#\delta_{u'})}{W_1(\delta_u, \delta_{u'})} \\ &= \sup_{u \neq u' \in X} \frac{\|\mathcal{F}(u) - \mathcal{F}(u')\|_X}{\|u - u'\|_X} = \mathrm{Lip}(\mathcal{F}). \end{aligned}$$

□

A.3. Compactness

We recall the Arzela–Ascoli theorem, characterizing compactness in $C_{\mathrm{loc}}(X, Y)$:

Theorem A.1 (Arzela–Ascoli). *Let X be a locally compact Hausdorff space. Let Y be a complete metric space. A subset $F \subset C_{\mathrm{loc}}(X, Y)$ is relatively compact iff it is equi-continuous and for all $x \in X$, the set $\{f(x) \,|\, f \in F\}$ is relatively compact in Y.*

ORCID iDs

S Lanthaler https://orcid.org/0000-0003-1911-246X
S Mishra https://orcid.org/0000-0002-2665-5385
F Weber https://orcid.org/0000-0002-7159-0170

References

[1] Apte A, Jones C K R T, Stuart A M and Voss J 2008 Data assimilation: mathematical and statistical perspectives *Int. J. Numer. Methods Fluids* **56** 1033–46
[2] Bansal P 2021 Numerical approximation of statistical solutions of the incompressible Navier–Stokes equations (arXiv:2107.06073)
[3] Bardos C and Tadmor E 2015 Stability and spectral convergence of Fourier method for nonlinear problems: on the shortcomings of the de-aliasing method *Numer. Math.* **129** 749–82
[4] Chorin A J 1968 Numerical solution of the Navier–Stokes equations *Math. Comput.* **22** 745–62
[5] Courtier P, Andersson E, Heckley W, Pailleux J, Vasiljevic D, Hamrud M, Hollingsworth A, Rabier F and Fisher M 1998 The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: formulation *Q. J. R. Meteorol. Soc.* **124** 1783–807
[6] Gottlieb Y M H D and Orszag S 1984 Theory and application of spectral methods *Spectral Methods for PDEs* (Philadelphia, PASIAM) pp 1–54
[7] Dafermos C M 2005 *Hyperbolic Conservation Laws in Continuum Physics* vol 3 (Berlin: Springer)
[8] Evensen G 2009 *Data Assimilation: The Ensemble Kalman Filter* (Berlin: Springer)
[9] Fjordholm U S, Käppeli R, Mishra S and Tadmor E 2017 Construction of approximate entropy measure-valued solutions for hyperbolic systems of conservation laws *Found. Comput. Math.* **17** 763–827
[10] Fjordholm U S, Lye K, Mishra S and Weber F 2020 Statistical solutions of hyperbolic systems of conservation laws: numerical approximation *Math. Models Methods Appl. Sci.* **30** 539–609
[11] Fjordholm U S, Mishra S and Tadmor E 2012 Arbitrarily high-order accurate entropy stable essentially nonoscillatory schemes for systems of conservation laws *SIAM J. Numer. Anal.* **50** 544–73
[12] Frisch U 1995 *Turbulence: The Legacy of A. N. Kolmogorov* (Cambridge: Cambridge University Press)
[13] Ghoshal S 1996 An analysis of numerical errors in large eddy simulations of turbulence *J. Comput. Phys.* **125** 187–206
[14] Glimm J, Grove J W, Li X L, Oh W and Sharp D H 2001 A critical analysis of Rayleigh–Taylor growth rates *J. Comput. Phys.* **169** 652–77
[15] Godlewski E and Raviart P-A 2013 *Numerical Approximation of Hyperbolic Systems of Conservation Laws* vol 118 (New York: Springer)
[16] Guermond J-L and Prudhomme S 2003 Mathematical analysis of a spectral hyperviscosity LES model for the simulation of turbulent flows *ESAIM: Math. Modelling Numer. Anal.* **37** 893–908
[17] Herrmann L, Schwab C and Zech J 2020 Deep neural network expression of posterior expectations in Bayesian PDE inversion *Inverse Problems* **36** 125011
[18] Kaipio J and Somersalo E 2006 *Statistical and Computational Inverse Problems* vol 160 (New York: Springer)
[19] Karamanos G-S and Karniadakis G E 2000 A spectral vanishing viscosity method for large-eddy simulations *J. Comput. Phys.* **163** 22–50
[20] Kovachki N, Lanthaler S and Mishra S 2021 On universal approximation and error bounds for Fourier neural operators *J. Mach. Learn. Res.* **22** 1–76
[21] Ladyzhenskaya O A 1969 *The Mathematical Theory of Viscous Incompressible Flow* (*Mathematics and its Applications* vol 2) (London: Gordon and Breach)
[22] Lanthaler S and Mishra S 2019 On the convergence of the spectral viscosity method for the two-dimensional incompressible Euler equations with rough initial data *Found. Comput. Math.* **20** 1309–62
[23] Lanthaler S, Mishra S and Parés-Pulido C 2021 On the conservation of energy in two-dimensional incompressible flows *Nonlinearity* **34** 1084–135

[24] Lanthaler S, Mishra S and Parés-Pulido C 2021 Statistical solutions of the incompressible Euler equations *Math. Models Methods Appl. Sci.* **31** 223–92
[25] Latz J 2020 On the well-posedness of Bayesian inverse problems *SIAM/ASA J. Uncertain. Quantification* **8** 451–82
[26] Law K, Stuart A and Zygalakis K 2015 *Data Assimilation* (Cham: Springer) p 214
[27] Law K J H and Stuart A M 2012 Evaluating data assimilation algorithms *Mon. Weather Rev.* **140** 3757–82
[28] Leray J 1934 Sur le mouvement d'un liquide visqueux emplissant l'espace *Acta Math.* **63** 193–248
[29] LeVeque R J 1992 *Numerical Methods for Conservation Laws* vol 3 (Basel: Springer)
[30] Li Z, Kovachki N B, Azizzadenesheli K, Liu B, Bhattacharya K, Stuart A and Anandkumar A 2021 Fourier neural operator for parametric partial differential equations *Int. Conf. on Learning Representations*
[31] Mishra S, Ochsner D, Ruf A M and Weber F 2021 Well-posedness of Bayesian inverse problems for hyperbolic conservation laws (arXiv:2107.09701)
[32] Pope S B 2001 *Turbulent Flows* (Cambridge: Cambridge University Press)
[33] Rabier F, Järvinen H, Klinker E, Mahfouf J-F and Simmons A 2000 The ECMWF operational implementation of four-dimensional variational assimilation. I: experimental results with simplified physics *Q. J. R. Meteorol. Soc.* **126** 1143–70
[34] Schneider T, Lan S, Stuart A and Teixeira J 2017 Earth system modeling 2.0: a blueprint for models that learn from observations and targeted high-resolution simulations *Geophys. Res. Lett.* **44** 12396–417
[35] Sprungk B 2020 On the local Lipschitz stability of Bayesian inverse problems *Inverse Problems* **36** 055015
[36] Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559
[37] Tadmor E 1989 Convergence of spectral methods for nonlinear conservation laws *SIAM J. Numer. Anal.* **26** 30
[38] Tadmor E 2004 Burgers' equation with vanishing hyper-viscosity *Commun. Math. Sci.* **2** 317–24
[39] Tarantola A 2005 *Inverse Problem Theory and Methods for Model Parameter Estimation* (Philadelphia, PA: SIAM)