Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

On the well-posedness of Bayesian inversion for PDEs with ill-posed forward problems

S. Lanthaler and S. Mishra and F. Weber

Research Report No. 2021-22
July 2021

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Funding ERC: 770880 COMANFLO

ON THE WELL-POSEDNESS OF BAYESIAN INVERSION FOR PDES WITH ILL-POSED FORWARD PROBLEMS

S. LANTHALER, S. MISHRA, F. WEBER

Abstract. We study the well-posedness of Bayesian inverse problems for PDEs, for which the underlying forward problem may be ill-posed. Such PDEs, which include the fundamental equations of fluid dynamics, are characterized by the lack of rigorous global existence and stability results as well as possible non-convergence of numerical approximations. Under very general hypotheses on approximations to these PDEs, we prove that the posterior measure, expressing the solution of the Bayesian inverse problem, exists and is stable with respect to perturbations of the (noisy) measurements. Moreover, analogous well-posedness results are obtained for the data assimilation (filtering) problem in the time-dependent setting. Finally, we apply this abstract framework to the incompressible Euler and Navier-Stokes equations and to hyperbolic systems of conservation laws and demonstrate well-posedness results for the Bayesian inverse and filtering problems, even when the underlying forward problem may be ill-posed.

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,

$$\begin{aligned} \mathbf{v}_t + \mathcal{D}\left(\mathbf{f}, \mathbf{v}, \nabla_x \mathbf{v}, \nabla_x^2 \mathbf{v}, \cdots\right) &= 0, \quad \forall x \in D, t \in (0,T) \\ \mathcal{B}\mathbf{v} &= \hat{\mathbf{v}}, \quad \forall x \in \partial D, t \in (0,T), \\ \mathbf{v}(x,0) &= \bar{\mathbf{v}}, \quad \forall x \in D. \end{aligned} \tag{1.1}$$

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

$$\mathcal{G} : X \mapsto Y, \quad u \mapsto \mathcal{G}(u) = \mathbf{v}, \tag{1.2}$$

with $\mathbf{v}$ solving the PDE (1.1). Here, X and Y are suitable Banach spaces.

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

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

Thus, the so-called *forward problem* for a PDE (1.1), is to evaluate the solution operator $\mathcal{G}$ or the observables $\mathcal{L}$, given the inputs u.

However, it is not always possible to exactly *know* the inputs u (initial and boundary data, coefficients, sources etc). Rather in practice, one has to infer information about the inputs u, and consequently the solution $\mathbf{v}$, from *measurements* of the observables in (1.3). Moreover in general, these measurements are *noisy*. Thus

Date: July 19, 2021.

one has to solve the so-called *inverse problem* for a PDE i.e., *determine the input u (and solution $\mathbf{v}$) for the PDE* (1.1)*, given measurements of observables of the form,*

$$y = \mathcal{L}(u) + \eta, \qquad \eta \sim \rho(y)\, dy, \tag{1.4}$$

with the noise sample 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 aforementioned 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 [32, 15, 29].

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 u 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 [29]. This results in a so-called *posterior* probability measure, on the inputs u, which represents the conditional probability of the underlying inputs, given the measurements (1.4). Thus, the *Bayesian Inverse Problem* 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 Bayesian inverse problem 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) [29, 21, 28]. Furthermore, Bayesian inverse problems can incorporate the deterministic formulation of regularized ill-posed inverse problems: As shown in [29], the latter can often be viewed as the maximum a posteriori (MAP) estimator of a Bayesian inverse problem with a suitable choice of the underlying prior.

These remarkable well-posedness results for Bayesian inverse problems for PDEs rely on the well-posedness of the underlying *forward problem*, often requiring that the mapping $\mathcal{L}$ 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 [29] for a survey of these results and their applications to a variety of PDEs. More recently in [21, 28], these Lipschitz continuity assumptions on the forward map $\mathcal{L}$ 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 Bayesian inverse problem [21].

However, even these very minimal assumptions on the forward map may not be satisfied for a large number of PDEs of immense practical interest. These correspond to PDEs for which well-posedness (existence, uniqueness, continuous dependence, stability of the input to solution operator $\mathcal{G}$ (1.2) and the resulting map $\mathcal{L}$ (1.3)) of the forward problem is either not true or cannot be proved rigorously. We label such PDEs as those with an *ill-posed forward problem.*

Prototypical examples for such ill-posed PDEs are provided by the fundamental equations of fluid dynamics, i.e., the incompressible Euler and Navier-Stokes equations as well as the compressible Euler equations. In particular, for the incompressible Navier-Stokes equations, there are currently no global well-posedness results in three space dimensions [24]. Although admissible weak solutions exist, the uniqueness, stability and regularity of such solutions are outstanding open problems. Similarly for the incompressible Euler equations, well-posedness results are only available in two space dimensions and with regular enough initial data (bounded initial vorticity). Well-posedness results in 2-d with less regular yet physically relevant data such as vortex sheets or in 3-d are mostly unavailable. In fact, it is now known that admissible weak solutions of the incompressible Euler equations need not be unique [5]. The compressible Euler equations are canonical examples of *hyperbolic systems of conservation laws* [4]. Again, there are no rigorous well-posedness results for hyperbolic systems of conservation laws in either two or three space dimensions. Thus, one cannot apply the abstract framework of [29] (or the recent modifications of [21, 28]) to rigorously conclude that the Bayesian inverse problem for these fundamental PDEs of fluid dynamics is well-posed.

This lack of mathematically rigorous well-posedness results for the afore-mentioned PDEs is not merely of academic interest but impacts the practical computation of both forward and Bayesian inverse problems as in practice, one approximates the posterior measure by sampling it using Markov Chain Monte Carlo (MCMC) algorithms of the Metropolis-Hastings type or their variants [29]. This, in turn, entails evaluating the forward map $\mathcal{L}$ (1.3) multiple times. However, in general, one has to numerically simulate this forward map i.e., replace it with an approximation $\mathcal{L}^{\Delta} \approx \mathcal{L}$, with Δ being a numerical parameter, such as the mesh size. Related to the lack of well-posedness of the underlying solution operator, it has been shown in recent papers such as [8, 9, 18, 19] that standard numerical approximations $\mathcal{L}^{\Delta}$ may not necessarily converge on mesh refinement (as $\Delta \to 0$) or converge too slowly to be of any practical interest. Consequently, even the rationale for numerical approximation of Bayesian inverse problems for these *ill-posed* PDEs is completely unclear.

Nevertheless, the Bayesian framework has been remarkably successful in the context of weather forecasting, climate modeling and oceanography [27]. Given that the underlying models include the incompressible (compressible) Euler and Navier-Stokes equations as the core governing PDEs, how does one reconcile the empirical success of the Bayesian framework with the lack of mathematically rigorous well-posedness for the underlying forward problem (and non-convergence of numerical approximations to it)?

This dichotomy sets the stage for the current article where we investigate the well-posedness of the Bayesian inverse problem for PDEs where the forward map $\mathcal{L}$ may be ill-posed. To this end, we focus on approximations to the forward map, $\mathcal{L}^{\Delta}$, which could stem from numerical approximations or physics based regularizations of the underlying PDE (1.1). These approximations are well-defined and lead to a family of *approximate posteriors* for the Bayesian inverse problems. For these family of posteriors, we will prove, under very general hypotheses, that

- The measurement to (approximate) posterior map is stable with respect to perturbations in the measurement, independent of the regularization (mesh) parameter Δ.
- The family of approximate posteriors is compact, in an appropriate metric, as $\Delta \to 0$, and the limit points are *posteriors* solving the Bayesian inverse problem. Thus, we will show existence and continuous dependence for the solutions of the Bayesian inverse problem corresponding to possibly ill-posed PDEs.

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

Moreover, we also consider the data assimilation (filtering) problem, associated with time-dependent finite-dimensional measurements of the underlying dynamical system, corresponding to ill-posed PDEs, and prove analogous existence, continuous dependence and stability results for this setting, too.

Finally, we apply our abstract results to investigate the well-posedness of the Bayesian inverse problem and data assimilation (filtering) problem for the incompressible Euler and Navier-Stokes equations as well as hyperbolic systems of conservation laws. We will show the surprising result that *although the forward problem associated with these PDEs, and its numerical approximation, may be ill-posed, solutions to the corresponding Bayesian problems exist and depend continuously and stably on the underlying measurements*. Thus, we provide the *first* rigorous results and rationale for Bayesian inversion for the fundamental equations of fluid dynamics.

The rest of the paper is organized as follows, we start with some notation and preliminaries in Section 2. The Bayesian inverse problem, with an ill-posed forward map is considered in Section 3 and the corresponding data assimilation (filtering) problem is presented in Section 4. We apply the abstract results of sections 3 and 4 to the fundamental equations of fluid dynamics in Section 5.

2. Notation and Preliminaries

In this section, we introduce the notation for the rest of the paper and recall some preliminaries that are necessary to define the Bayesian inverse problem 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 $\mu_n \in \mathcal{P}(X)$ is said to converge weakly to a limit μ, denoted $\mu_n \rightharpoonup \mu$, if

$$\int_X \phi \, d\mu_n \to \int_X \phi \, d\mu, \qquad \forall \phi \in C_b(X),$$

where $C_b(X)$ 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 p-th moments, $\int_X \|u\|_X^p \, d\mu(u) < \infty$, metrized by the p-Wasserstein distance W_p:

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

Here, $\Gamma(\mu,\nu)$ is the set of couplings between μ and ν, i.e. probability measures π on $X \times X$, with projections $(\mathrm{Proj}_1)_\# \pi = \mu$, $(\mathrm{Proj}_2)_\# \pi = \nu$. 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

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

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

$$W_1(\mu,\nu) = \sup_{\Phi} \int \Phi(u) \left[d\mu(u) - d\nu(u) \right], \tag{2.1}$$

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

$$\|\Phi\|_{\mathrm{Lip}} := \sup_{u \neq v} \frac{|\Phi(u) - \Phi(v)|}{\|u - v\|_X}. \tag{2.2}$$

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

$$\lim_{\Delta\to 0} W_1(\mu^\Delta, \mu) = 0 \iff \left\{ \begin{array}{c} \mu^\Delta \rightharpoonup \mu \text{ converges weakly and} \\ \displaystyle\int_X \|u\|_X \, d\mu^\Delta(u) \to \int_X \|u\|_X \, d\mu(u). \end{array} \right\} \tag{2.3}$$

We will denote the Kullback-Leibler (KL) divergence of a measure $\nu \in \mathcal{P}(X)$ with respect to $\mu \in \mathcal{P}(X)$ by $\mathcal{D}_{\mathrm{KL}}(\nu||\mu)$; We recall that the Kullback-Leibler divergence is defined by

$$\mathcal{D}_{\mathrm{KL}}(\nu||\mu) := \begin{cases} \int_X \log\left(\frac{d\nu}{d\mu}\right) \, d\nu, & (\nu \ll \mu), \\ +\infty, & (\nu \not\ll \mu). \end{cases} \tag{2.4}$$

It is well-known that $\mathcal{P}(X) \to \mathbb{R}$, $\nu \mapsto \mathcal{D}_{\mathrm{KL}}(\nu||\mu)$ is a strictly convex, coercive and lower semi-continuous function. In particular, for any $\alpha > 0$ the set $\{\nu \in \mathcal{P}(X) \,|\, \mathcal{D}_{\mathrm{KL}}(\nu||\mu) \le \alpha\}$ is compact in the weak topology on $\mathcal{P}(X)$.

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 $C = C(\alpha, \beta, \gamma)$.

3. Bayesian inverse problem

The goal of the present section is to investigate the general stability, compactness and consistency of the Bayesian inverse problem (BIP) for PDEs for which the forward problem is potentially ill-posed.

As mentioned in the introduction, our main tool, in this regard, is to consider a sequence of *approximate* observables (approximations of $\mathcal{L}$ (1.3), generated, for instance, either by numerical methods for the underlying PDE with a mesh size (time step) $\Delta > 0$ or a (viscous) regularization with a regularization parameter Δ, resulting in a mapping,

$$\mathcal{L}^\Delta : X \to \mathbb{R}^d, \qquad u \mapsto \mathcal{L}^\Delta(u), \tag{3.1}$$

that is well-defined and measurable for any $\Delta > 0$.

We consider the Bayesian inverse problem of finding the probability distribution $\mathbb{P}[u|y]$ for the underlying data u, given a finite-dimensional measurement $y \in \mathbb{R}^d$ of the form

$$y = \mathcal{L}^\Delta(u) + \eta, \qquad \eta \sim \rho(y)\, dy. \tag{3.2}$$

The noise $\eta \in \mathbb{R}^d$ is here assumed to have a distribution $\rho(y)$ which is absolutely continuous with respect to Lebesgue measure dy on $\mathbb{R}^d$, $\int_{\mathbb{R}^d} \rho(y)\, dy = 1$, and $\rho(y) > 0$ for all $y \in \mathbb{R}^d$. As shown in [21, Thm. 2.5], under these conditions on $\rho(y)$, the measurability of $\mathcal{L}^\Delta(u)$ is sufficient to guarantee the existence of a solution to the BIP to (3.2) given an arbitrary prior $\mu \in \mathcal{P}(X)$. This solution is given by the posterior

$$d\mu^{\Delta,y}(u) = \frac{1}{Z^\Delta(y)} \exp\left(-\Phi^{\Delta,y}(u)\right) \, d\mu(u), \tag{3.3}$$

where

$$\Phi^{\Delta,y}(u) := -\log \rho\left(y - \mathcal{L}^\Delta(u)\right) \tag{3.4}$$

denotes the log-likelihood function, and

$$Z^\Delta(y) = \int_X \exp\left(-\Phi^{\Delta,y}(u)\right) \, d\mu(u), \tag{3.5}$$

is the required normalization constant. We note that the condition that $\rho(y) > 0$ implies that the log-likelihood $\Phi^{\Delta,y}$ is finite, i.e., $\Phi^{\Delta,y}(u) < \infty$ for all $u \in X$.

As is customary, we will denote the Radon-Nikodym derivative of $\mu^{\Delta,y}$ with respect to μ by $d\mu^{\Delta,y}/d\mu$, i.e.

$$\frac{d\mu^{\Delta,y}}{d\mu}(u) = \frac{1}{Z^{\Delta}(y)} \exp\left(-\Phi^{\Delta,y}(u)\right). \tag{3.6}$$

The solution of the BIP (3.3) can be characterized as the unique minimizer $\mu^{\Delta,y} = \operatorname{argmin}_{\nu\in\mathcal{P}(X)} J^{\Delta,y}(\nu)$ of the following functional $J^{\Delta,y} : \mathcal{P}(X) \to \mathbb{R}$ (cp. e.g. [7, Prop. 1.4.2]):

$$J^{\Delta,y}(\nu) := \mathcal{D}_{\mathrm{KL}}(\nu||\mu) + \int_X \Phi^{\Delta,y}(u)\, d\nu(u), \tag{3.7}$$

where $\mathcal{D}_{\mathrm{KL}}(\nu||\mu)$ denotes the Kullback-Leibler divergence (2.4). Furthermore, the minimum of $J^{\Delta,y}$ is explicitly given by [7, eq. (1.15)],

$$-\log\left(\int_X e^{-\Phi^{\Delta,y}(u)}\, d\mu(u)\right) = \inf_{\nu\in\mathcal{P}(X)} J^{\Delta,y}(\nu). \tag{3.8}$$

Taking into account (3.5), we can write the last equation equivalently as follows:

$$Z^{\Delta}(y) = \exp\left(-\inf_{\nu\in\mathcal{P}(X)} J^{\Delta,y}(\nu)\right). \tag{3.9}$$

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 following additional assumptions on the noise:

***Assumption* 3.1.** 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

$$|y|_\Gamma = \sqrt{\langle y,y\rangle_\Gamma}, \quad \langle y,y'\rangle_\Gamma = \langle \Gamma^{-1/2}y, \Gamma^{-1/2}y'\rangle = \langle y, \Gamma^{-1}y'\rangle, \tag{3.10}$$

with $\langle\,\cdot\,,\,\cdot\,\rangle$ the standard Euclidean inner product on $\mathbb{R}^d$. We assume that the noise $\eta \sim \rho(y)\,dy$ in (3.2) possesses a distribution that is absolutely continuous with respect to Lebesgue measure dy on $\mathbb{R}^d$ with probability density $\rho(y)$, satisfying the following conditions:

- [regularity] $y \mapsto \rho(y)$ is Lipschitz continuous with respect to $|\cdot|_\Gamma$,[1],
- [boundedness] $y \mapsto \rho(y)$ is bounded from above,
- [tail-condition] there exists a constant $C > 0$, such that

$$\rho(y) \geq \frac{\exp\left(-\frac{1}{2}|y|_\Gamma^2\right)}{C}, \quad \forall\, y \in \mathbb{R}^d. \tag{3.11}$$

***Remark* 3.2.** Note that if, instead of (3.11), $\rho(y)$ satisfies a tail-condition of the form $\rho(y) \geq \exp(-C|y|_\Gamma^2)/C$, then upon simply rescaling $\tilde{\Gamma} := \sqrt{2/C}\,\Gamma$, we have $\rho(y) \geq \exp(-\frac{1}{2}|y|_{\tilde{\Gamma}}^2)/C$. Hence $\rho(y)$ satisfies assumption 3.1 with a rescaled matrix $\Gamma \to \tilde{\Gamma}$ in this case. Therefore, the precise constant $\frac{1}{2}$ in the tail-condition (3.11) can be assumed without loss of generality. The factor of $1/2$ turns out to be particularly convenient.

Assumption (3.1) is clearly fulfilled for normally distributed measurement noise η. 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

[1] 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.

particular, since the tail-condition requires only a lower bound, our results apply to situations in which one encounters noise *with a heavy tail*.

***Remark* 3.3** (Gaussian noise)**.** If the noise $\eta \sim \mathcal{N}(0,\Gamma)$ is normally distributed (Gaussian), then (up to an unimportant additive constant)

$$\Phi^{\Delta,y}(u) = \frac{1}{2}|y - \mathcal{L}^\Delta(u)|_\Gamma^2,$$

where the natural Γ-norm is given by (3.10). In this case, we have

$$\frac{d\mu^{\Delta,y}}{d\mu}(u) = \frac{1}{Z^\Delta(y)} \exp\left(-\frac{1}{2}\left|y - \mathcal{L}^\Delta(u)\right|_\Gamma^2\right). \tag{3.12}$$

Let us note the following immediate observations from assumption 3.1:

***Lemma* 3.4.** If the noise $\eta \sim \rho(y)\,dy$ satisfies assumption 3.1, then there exists a constant $L > 0$, such that for all $y, y' \in \mathbb{R}^d$, and $\Delta, \Delta' > 0$

$$\left|e^{-\Phi^{\Delta,y}(u)} - e^{-\Phi^{\Delta,y'}(u)}\right| \le L|y - y'|_\Gamma, \tag{3.13}$$

and

$$\left|e^{-\Phi^{\Delta,y}(u)} - e^{-\Phi^{\Delta',y}(u)}\right| \le L|\mathcal{L}^\Delta(u) - \mathcal{L}^{\Delta'}(u)|_\Gamma. \tag{3.14}$$

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

$$\operatorname*{ess\,inf}_{u\in X} \Phi^{\Delta,y}(u) \ge -C, \quad \forall\, \Delta > 0,\, y \in \mathbb{R}^d. \tag{3.15}$$

There exists a constant $C' \ge 0$, such that

$$\Phi^{\Delta,y}(u) \le C' + \frac{1}{2}|y - \mathcal{L}^\Delta(u)|_\Gamma^2. \tag{3.16}$$

In particular, we have

$$\Phi^{\Delta,y}(u) \le C' + |y|_\Gamma^2 + |\mathcal{L}^\Delta(u)|_\Gamma^2. \tag{3.17}$$

Given a sequence of observables $\mathcal{L}^\Delta(u)$ ($\Delta \to 0$) arising for example from numerical discretizations at grid scale Δ, it is now natural to ask what can be said about the limiting behaviour of the corresponding sequence of posteriors $\mu^{\Delta,y}$. For many problems arising in the context of fluid dynamics very limited information is available on the stability and convergence of the observables $\mathcal{L}^\Delta(u) \to \mathcal{L}(u)$ to a well-defined limit. Indeed, even the existence of a limiting observable $\mathcal{L}(u)$ is often not guaranteed, due to the (potential) ill-posedness of the forward model. It is thus important to study the behaviour of the sequence $\mu^{\Delta,y}$ under minimal assumptions on the observables $\mathcal{L}^\Delta(u)$. We pose that these assumption should either be *rigorously provable* for models of practical interest, or at least numerically *verifiable and routinely observed* in numerical experiments. In the remainder of this section, we will follow this programme for abstract Bayesian inverse problems. We will in particular consider

- the **stability** of the posteriors $\mu^{\Delta,y}$ with respect to the measurements y with respect to the Wasserstein distance, obtaining estimates which hold uniformly as $\Delta \to 0$,
- the general **compactness** properties of the sequence $\mu^{\Delta,y}$ in the Wasserstein distance, and
- the **consistency** of $\mu^{\Delta,y}$ with the posterior μ^y corresponding to the limiting measurement $\mathcal{L}^\Delta(u) \to \mathcal{L}(u)$, provided that the latter exists.

In particular, as a consequence of our discussion, we will prove the existence of a set of candidate solutions of the BIP in the limit $\Delta \to 0$, under mild boundedness assumptions on the observables $\mathcal{L}^\Delta(u)$.

3.1. **Stability with respect to measurements.** We first discuss the stability of the posterior $\mu^{\Delta,y}$ with respect to the measurement y. As a natural measure of the distance between two posteriors $\mu^{\Delta,y}$, $\mu^{\Delta,y'}$, we consider the 1-Wasserstein distance $W_1(\mu^{\Delta,y}, \mu^{\Delta,y'})$. Our goal is to prove an explicit upper bound on $W_1(\mu^{\Delta,y}, \mu^{\Delta,y'})$ in terms of $|y - y'|_\Gamma$. We note that our discussion of stability for the BIP overlaps in part with a similar discussion contained in [21, 28]. In particular, [28] 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 needed estimates have not appeared in [21, 28], 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 $d\mu^{\Delta,y}/d\mu$ is uniformly bounded in $L^\infty(\mu)$, provided that $\sup_{\Delta>0} \|\mathcal{L}^\Delta(u)\|_{L^2(\mu)} < \infty$; here we define the $L^2(\mu)$-norm of the observables $\mathcal{L}^\Delta(u)$ as follows

***Remark* 3.5.** The $L^2(\mu)$-norm of $\mathcal{L}^\Delta(u)$ in Lemma 3.6 is defined by

$$\|\mathcal{L}^\Delta(u)\|_{L^2(\mu)} := \left(\int |\mathcal{L}^\Delta(u)|_\Gamma^2 \, d\mu(u), \right)^{1/2}$$

where Γ is the covariance matrix of the additive noise η.

We now state the following

***Lemma* 3.6.** Let $d\mu^{\Delta,y}/d\mu$ be given by (3.6), and $Z^\Delta(y)$ be defined as in (3.5). Then

$$Z^\Delta(y) \ge \exp\left(- \int_X \Phi^{\Delta,y}(u) \, d\mu(u) \right), \tag{3.18}$$

and

$$\frac{d\mu^{\Delta,y}}{d\mu}(u) \le \exp\left(\int_X \Phi^{\Delta,y}(u) \, d\mu(u) - \operatorname*{ess\,inf}_{u\in X} \Phi^{\Delta,y}(u) \right), \quad \forall\, u \in X, \tag{3.19}$$

In particular, if the noise $\eta \sim \rho(y)dy$ satisfies the standing assumption 3.1, then there exists a constant $C > 0$ depending only on the noise distribution $\rho(y)$, such that

$$Z^\Delta(y) \ge \frac{1}{C} \exp\left(-|y|_\Gamma^2 - \|\mathcal{L}^\Delta\|_{L^2(\mu)}^2 \right), \tag{3.20}$$

and

$$\frac{d\mu^{\Delta,y}}{d\mu}(u) \le C \exp\left(|y|_\Gamma^2 + \|\mathcal{L}^\Delta\|_{L^2(\mu)}^2 \right), \quad \forall\, u \in X. \tag{3.21}$$

Proof. Since the exponential (Gaussian-like) factor in the definition of $d\mu^{\Delta,y}/d\mu$, eq. (3.6), is bounded from above by $\exp(-\operatorname{ess\,inf}_{u\in X} \Phi^{\Delta,y}(u))$, it suffices to prove the lower bound on $Z^\Delta(y)$. We recall that by (3.9), we can write

$$Z^\Delta(y) = \exp\left(- \inf_{\nu\in\mathcal{P}(X)} J^{\Delta,y}(\nu) \right),$$

where $J^{\Delta,y}(\nu) = \mathcal{D}_{\mathrm{KL}}(\nu||\mu) + \int_X \Phi^{\Delta,y}(u)\,d\nu(u)$. In particular, it follows that

$$\inf_{\nu\in\mathcal{P}(X)} J^{\Delta,y}(\nu) \le J^{\Delta,y}(\mu) = \int_X \Phi^{\Delta,y}(u)\,d\mu(u).$$

Thus, we conclude that

$$Z^{\Delta}(y) \ge \exp\left(-\int_X \Phi^{\Delta,y}(u)\,d\mu(u)\right).$$

This implies the first two estimates (3.18) and (3.19) of this lemma.

Under the noise assumption 3.1, by (3.17), there exists $C' > 0$ depending only on the noise distribution $\rho(y)$, such the last term can be bounded from below, yielding

$$Z^{\Delta}(y) \ge \exp\left(-C' - |y|_\Gamma^2 - \int_X |\mathcal{L}^{\Delta}(u)|_\Gamma^2\,d\mu(u)\right),$$

and thus the claimed inequality (3.20) for $Z^{\Delta}(y)$ with $C = \exp(C')$. Furthermore, by (3.15), there exists C'', such that

$$\operatorname*{ess\,inf}_{u\in X} \Phi^{\Delta,y}(u) \ge -C''.$$

Thus the claimed inequality (3.21) holds with $C = \exp(C' + C'')$. □

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

***Lemma* 3.7.** Under assumption 3.1. Let $\mathcal{L}^{\Delta}(u) \in L^2(\mu)$. There exists a constant $C > 0$ (depending only on the noise distribution), such that

$$\left\| \frac{d\mu^{\Delta,y}}{d\mu} - \frac{d\mu^{\Delta,y'}}{d\mu} \right\|_{L^\infty(\mu)} \le C|y-y'|_\Gamma \exp\left(|y|_\Gamma^2 + |y'|_\Gamma^2 + 2\|\mathcal{L}^{\Delta}\|_{L^2(\mu)}^2\right). \tag{3.22}$$

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

$$\begin{aligned} \frac{d\mu^{\Delta,y}}{d\mu} - \frac{d\mu^{\Delta,y'}}{d\mu} &= \frac{e(y)}{Z^{\Delta}(y)} - \frac{e(y')}{Z^{\Delta}(y')} \\ &= \frac{e(y)-e(y')}{Z^{\Delta}(y)} + \frac{e(y')}{Z^{\Delta}(y')}\frac{(Z^{\Delta}(y')-Z^{\Delta}(y))}{Z^{\Delta}(y)}. \end{aligned}$$

By (3.13), we can estimate $|e(y) - e(y')| \le C|y - y'|_\Gamma$. Next, we note that this bound for $e(y)$ also implies that

$$|Z^{\Delta}(y) - Z^{\Delta}(y')| \le \int_X |e(y;u) - e(y';u)|\,d\mu(u) \le C|y-y'|_\Gamma \underbrace{\int_X 1\,d\mu(u)}_{=1}.$$

Hence,

$$\left|\frac{d\mu^{y}}{d\mu} - \frac{d\mu^{y'}}{d\mu}\right| \le \frac{C|y-y'|_\Gamma}{Z^{\Delta}(y)} + \frac{e(y')}{Z^{\Delta}(y')}\frac{C|y-y'|_\Gamma}{Z^{\Delta}(y)}.$$

Finally, from Lemma 3.6, we can estimate

$$\frac{1}{Z^{\Delta}(y)} \le Ce^{|y|_\Gamma^2 + \|\mathcal{L}^{\Delta}\|_{L^2(\mu)}^2} \le Ce^{|y|_\Gamma^2 + |y'|_\Gamma^2 + 2\|\mathcal{L}^{\Delta}\|_{L^2(\mu)}^2},$$

and

$$\frac{e(y')}{Z^{\Delta}(y')}\frac{1}{Z^{\Delta}(y)} \le Ce^{|y|_\Gamma^2 + |y'|_\Gamma^2 + 2\|\mathcal{L}^{\Delta}\|_{L^2(\mu)}^2}.$$

Combining these estimates, we conclude that

$$\left|\frac{d\mu^y}{d\mu} - \frac{d\mu^{y'}}{d\mu}\right| \le 2C|y-y'|_\Gamma \exp\left(|y|_\Gamma^2 + |y'|_\Gamma^2 + 2\|\mathcal{L}^\Delta\|_{L^2(\mu)}^2\right).$$

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

Let us also remark in passing the following Lemma, whose proof is analogous to the proof of Lemma 3.7.

***Lemma* 3.8.** Under assumption 3.1. Let $\mathcal{L}^\Delta(u), \mathcal{L}(u) \in L^2(\mu)$, and $y \in \mathbb{R}^d$. There exists a constant $C > 0$ (depending only on the noise distribution), such that for any $p \in [1,\infty]$, we have

$$\left\|\frac{d\mu^{\Delta,y}}{d\mu} - \frac{d\mu^y}{d\mu}\right\|_{L^p(\mu)} \le C\left\|\mathcal{L}^\Delta(u) - \mathcal{L}(u)\right\|_{L^p(\mu)} \exp\left(2|y|_\Gamma^2 + \|\mathcal{L}^\Delta\|_{L^2(\mu)}^2 + \|\mathcal{L}\|_{L^2(\mu)}^2\right),$$

for all u for which $\mathcal{L}^\Delta(u)$, $\mathcal{L}(u)$ is defined.

Proof. The proof is an almost verbatim repetition of the proof of Lemma 3.7, with the roles of y, y' and $\mathcal{L}^\Delta(u), \mathcal{L}(u)$ interchanged. □

Using Lemma 3.7, we can now state the following theorem on the stability of the measurement-to-posteriors map:

***Theorem* 3.9.** We make the assumption 3.1 on the noise $\eta \sim \rho(y)\,dy$. Fix a prior $\mu \in \mathcal{P}_1(X)$. Given a measurement $y \in \mathbb{R}^d$, $\Delta > 0$ with observable $\mathcal{L}^\Delta(u)$ and prior μ, let $\mu^{\Delta,y}$ denote the corresponding posterior (3.3). Assume that

$$M := \sup_{\Delta>0}\left\|\mathcal{L}^\Delta\right\|_{L^2(\mu)} < \infty.$$

Then the family of posteriors $\{\mu^{\Delta,y}\}$ is uniformly bounded in $\mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d;\mathcal{P}_1(X))$ and hence locally equicontinuous: There exists a constant $C = C(\rho,\Gamma,M,\mu)$, independent of $\Delta > 0$ and y, y', such that

$$W_1(\mu^{\Delta,y},\mu^{\Delta,y'}) \le C|y-y'|_\Gamma e^{|y|_\Gamma^2+|y'|_\Gamma^2}. \tag{3.23}$$

Proof. Fix $\Phi \in \mathrm{Lip}(X)$ with $\mathrm{Lip}(\Phi) \le 1$. Then

$$\begin{aligned}\int_X \Phi(u)(d\mu^{\Delta,y}(u) - d\mu^{\Delta,y'}(u)) &= \int_X [\Phi(u)-\Phi(0)](d\mu^{\Delta,y}(u) - d\mu^{\Delta,y'}(u)) \\ &\le \int_X \|u\|_X \left|\frac{d\mu^{\Delta,y}}{d\mu} - \frac{d\mu^{\Delta,y'}}{d\mu}\right| d\mu(u) \\ &\le \left\|\frac{d\mu^{\Delta,y}}{d\mu} - \frac{d\mu^{\Delta,y'}}{d\mu}\right\|_{L^\infty(\mu)} \left(\int_X \|u\|_X\,d\mu(u)\right).\end{aligned}$$

Estimating the last term using Lemma 3.7 and taking the supremum over all such Lipschitz continuous Φ on the left-hand side, we obtain by Kantorovich duality:

$$W_1(\mu^{\Delta,y},\mu^{\Delta,y'}) \le \overline{C}\|u\|_{L^1(\mu)}|y-y'|_\Gamma e^{|y|_\Gamma^2+|y'|_\Gamma^2},$$

where $\overline{C}$ is independent of Δ. In fact, we can choose

$$\overline{C} = \sup_{\Delta>0} Ce^{2\|\mathcal{L}^\Delta(u)\|_{L^2(\mu)}^2} = Ce^{2M^2},$$

with C the constant from Lemma 3.7. □

***Remark* 3.10.** The previous stability result only depends on the continuity properties of the noise distribution ρ, and is independent of any continuity properties of the observable $\mathcal{L}(u)$. In the same spirit, if $d\mu^y/d\mu = 1/Z(y)\,\exp\left(-\frac{1}{2}|y-\mathcal{L}(u)|_\Gamma^2\right)$ is a posterior with Gaussian noise, and if $\|\mathcal{L}(u)\|_{L^2(\mu)} < \infty$, then we can show that for any $\phi(u) \in L^1(\mu)$ (i.e. $\phi(u)$ is integrable with respect to the prior μ), we have that

$$\mathbb{R}^d \to \mathbb{R}, \quad y \mapsto \mathbb{E}^y[\phi] := \frac{1}{Z(y)} \int_X \phi(u)\, d\mu^y(u),$$

is real analytic; this follows from [14, Lemma 4.5]. In particular, this result is independent of any smoothness properties of $\mathcal{L}(u)$. In section 4, we will show that the conclusion remains true even for the time-dependent *data assimilation* (filtering) problem (cp. Remark 4.10).

3.2. **Compactness properties.** Having established the uniform equicontinuity of the measurement-to-posterior mapping, we next wish to show that the posteriors $\mu^{\Delta,y}$, for fixed $y \in \mathbb{R}^d$, form a compact sequence as $\Delta \to 0$ in $(\mathcal{P}_1, W_1)$, and that all limit points are absolutely continuous with respect to the prior μ. The proof of compactness of $\mu^{\Delta,y}$ will be based on the variational characterization of the posteriors to the BIP, in terms of the Kullback-Leibler divergence with respect to the prior.

We now show pointwise compactness of the posteriors $\mu^{\Delta,y}$ for fixed $y \in \mathbb{R}^d$:

***Theorem* 3.11.** Fix a prior $\mu \in \mathcal{P}_1(X)$. Fix $y \in \mathbb{R}^d$. Assume that the log-likelihood $\Phi^{\Delta,y} \geq -C$ is uniformly bounded from below, and that $\int_X \Phi^{\Delta,y}(u)\, d\mu(u) \leq C$ are uniformly bounded from above for $\Delta > 0$. Then the family of posteriors $\{\mu^{\Delta,y}\}_{\Delta>0}$ is pre-compact in $\mathcal{P}_1(X)$, and any limit point $\mu^{*,y} = \lim_{\Delta_k \to 0} \mu^{\Delta_k,y}$ is absolutely continuous with respect to the prior μ.

Proof. As remarked in the introduction to this section, the posterior $\mu^{\Delta,y}$ can be characterized as the unique minimizer $\mu^{\Delta,y} = \operatorname{argmin}_{\nu \in \mathcal{P}_1(X)} J^{\Delta,y}(\nu)$ of the functional $J^{\Delta,y}$ (3.7). In particular, this variational characterization implies that

$$\begin{aligned}
\mathcal{D}_{\mathrm{KL}}(\mu^{\Delta,y}||\mu) &= J^{\Delta,y}(\mu^{\Delta,y}) - \int_X \underbrace{\Phi^{\Delta,y}(u)}_{\geq -C}\, d\mu^{\Delta,y}(u) \\
&\leq J^{\Delta,y}(\mu^{\Delta,y}) + C \\
&\leq J^{\Delta,y}(\mu) + C \\
&= \int_X \Phi^{\Delta,y}(u)\, d\mu(u) + C \\
&\leq 2C.
\end{aligned}$$

It follows that

$$\{\mu^{\Delta,y}\}_{\Delta>0} \subset \{\nu \in \mathcal{P}_1(X) \,|\, \mathcal{D}_{\mathrm{KL}}(\nu||\mu) \leq 2C\}.$$

From the coercivity property of the Kullback-Leibler divergence $\mathcal{D}_{\mathrm{KL}}$, the sublevel set $\{\nu \in \mathcal{P}_1(X) \,|\, \mathcal{D}_{\mathrm{KL}}(\nu||\mu) \leq 2C\}$ is compact with respect to the topology of weak convergence of probability measures. Furthermore, any weak limit point $\mu^{*,y} = w{-}\lim_{\Delta_k \to 0} \mu^{\Delta_k,y}$ satisfies $\mathcal{D}_{\mathrm{KL}}(\mu^{*,y}||\mu) \leq 2C < \infty$, and hence is absolutely continuous with respect to μ. This shows that $\{\mu^{\Delta,y}\}_{\Delta>0}$ is precompact with respect to the weak topology on $\mathcal{P}(X)$. We finally want to show that if $\mu^{*,y} = w{-}\lim_{\Delta_k \to 0} \mu^{\Delta_k,y}$ is a weak limit of the family $\{\mu^{\Delta,y}\}_{\Delta>0}$, then in fact $W_1(\mu^{*,y}, \mu^{\Delta_k,y}) \to 0$ converges with respect to the 1-Wasserstein distance. As a consequence, we conclude that $\{\mu^{\Delta,y}\}_{\Delta>0}$ is also pre-compact in the metric space $(\mathcal{P}_1, W_1)$.

To this end, suppose we are given a weakly convergent subsequence $\mu^{\Delta_k,y} \rightharpoonup \mu^{*,y}$. By (2.3), in order to show that $W_1(\mu^{*,y}, \mu^{\Delta_k,y}) \to 0$, it suffices to prove that

$$\int_X \|u\|_X \, d\mu^{\Delta_k,y}(u) \to \int_X \|u\|_X \, d\mu^{*,y}(u).$$

Let $\epsilon > 0$ be arbitrary. We want to show that

$$\limsup_{k\to\infty} \left| \int_X \|u\|_X \, d\mu^{\Delta_k,y}(u) - \int_X \|u\|_X \, d\mu^{*,y}(u). \right| \le \epsilon.$$

By Lemma 3.6, and the assumed uniform upper bound on $\int \Phi^{\Delta,y}(u)\, d\mu(u)$, there exists a constant $C > 0$, such that

$$\frac{d\mu^{\Delta,y}}{d\mu} \le C, \quad \forall\, \Delta > 0.$$

As $\int_X \|u\|_X \, d\mu(u) < \infty$, we can choose $M > 0$ sufficiently large, so that

$$\int_{\|u\|_X \ge M} \|u\|_X \, d\mu(u) < \epsilon/(2C).$$

Then, clearly

$$\begin{aligned} \int_{\|u\|_X \ge M} \|u\|_X \, d\mu^{\Delta_k,y}(u) &= \int_{\|u\|_X \ge M} \|u\|_X \, \frac{d\mu^{\Delta_k,y}}{d\mu} \, d\mu(u) \\ &\le C \int_{\|u\|_X \ge M} \|u\|_X \, d\mu(u) < \epsilon/2, \end{aligned} \tag{3.24}$$

for all $k \in \mathbb{N}$, and by the lower semi-continuity of weak limits, a similar inequality holds for $\mu^{*,y}$:

$$\int_{\|u\|_X \ge M} \|u\|_X \, d\mu^{*,y}(u) \le \liminf_{k\to\infty} \int_{\|u\|_X \ge M} \|u\|_X \, d\mu^{\Delta_k,y}(u) \le \epsilon/2. \tag{3.25}$$

Define $F_M(u) := \min(\|u\|_X, M) \in C_b(X)$. Then,

$$\begin{aligned} \limsup_{k\to\infty} &\left| \int_X \|u\|_X \;\; d\mu^{\Delta_k,y} - \int_X \|u\|_X \, d\mu^{*,y} \right| \\ &\le \limsup_{k\to\infty} \left| \int_X F_M(u) \, [d\mu^{\Delta_k,y} - d\mu^{*,y}] \right| \\ &\quad + \limsup_{k\to\infty} \int_{\|u\|_X \ge M} \|u\|_X \, d\mu^{\Delta_k,y}(u) + \int_{\|u\|_X \ge M} \|u\|_X \, d\mu^{*,y}(u) \\ &\le 0 + \epsilon/2 + \epsilon/2 = \epsilon. \end{aligned}$$

To pass to the last line, we used the upper bounds (3.24), (3.25) and the fact that

$$\int_X F_M(u) \, d\mu^{\Delta_k,y}(u) \to \int_X F_M(u) \, d\mu^{*,y}(u),$$

since $F_M \in C_b(X)$ and $\mu^{\Delta_k,y} \rightharpoonup \mu^{*,y}$. Since $\epsilon > 0$ was arbitrary, we conclude that

$$\int_X \|u\|_X \, d\mu^{\Delta_k,y}(u) \to \int_X \|u\|_X \, d\mu^{*,y}(u),$$

and hence $W_1(\mu^{\Delta_k,y}, \mu^{*,y}) \to 0$ (cp. (2.3)). In particular, this shows that any weak limit point of $\{\mu^{\Delta,y}\}_{\Delta>0}$ is also a limit point in $\mathcal{P}_1(X)$ with respect to the 1-Wasserstein metric W_1. Since $\{\mu^{\Delta,y}\}_{\Delta>0}$ is weakly pre-compact, it follows that it is also pre-compact in $\mathcal{P}_1(X)$ with respect to the W_1-metric. □

Finally, we can combine the uniform equicontinuity result of Theorem 3.9 with the point-wise compactness established in Theorem 3.11 to prove the following general compactness theorem for posteriors, now considered as mappings $y \mapsto \mu^{\Delta,y}$:

***Theorem* 3.12.** We make assumption 3.1 on the noise distribution. Fix a prior $\mu \in \mathcal{P}_1(X)$. Let $\{\mathcal{L}^\Delta\}_{\Delta>0}$ be a uniformly $L^2(\mu_0)$-bounded family of measurable mappings $\mathcal{L}^\Delta : X \to \mathbb{R}^d$. Then the corresponding family of posterior measures $y \mapsto \mu^{\Delta,y}$ is pre-compact with respect to the topology of locally uniform convergence on $\mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d; \mathcal{P}_1(X))$: For any sequence $\Delta \to 0$, there exists a subsequence $\Delta_k \to 0$ and a y-parametrized probability measure $y \mapsto \mu^{*,y} \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d; \mathcal{P}_1(X))$, such that for any $R > 0$, there exists $C = C(R, \Gamma, \mu)$, such that

$$W_1(\mu^{*,y}, \mu^{*,y'}) \le C|y - y'|_\Gamma, \quad \forall\, y, y' \in B_R(0),$$

and we have

$$\sup_{|y|_\Gamma \le R} W_1\left(\mu^{\Delta_k,y}, \mu^{*,y}\right) \to 0, \quad \text{as } k \to \infty.$$

Furthermore, any such limit $\mu^{*,y}$ is absolutely continuous with respect to the prior μ, and can be written in the form $d\mu^{*,y}(u) = Z(y)^{-1} \exp(-\Phi^*(u; y))\, d\mu(u)$.

Proof. This theorem is a direct consequence of the Arzelà-Ascoli Theorem A.1, the pointwise compactness Theorem 3.11 and the uniform equicontinuity Theorem 3.9. □

***Remark* 3.13.** The last theorem shows that under quite general conditions, we can assign a set of "solutions" of a BIP (or at least candidate solutions) to a family of posteriors $\mu^{\Delta,y}$ solving the discretized BIP at resolution $\Delta > 0$. This set of candidate solutions of the BIP in the limit $\Delta \to 0$ is given by

$$\mathcal{S} = \left\{ \mu^{*,y} \,\middle|\, \exists \Delta_k \to 0, \text{ s.t. } \mu^{*,y} = \lim_{k\to\infty} \mu^{\Delta_k,y} \right\},$$

or equivalently, we can write

$$\mathcal{S} = \bigcap_{\overline{\Delta}>0} \mathrm{cl}\left(\left\{ y \mapsto \mu^{\Delta,y} \,\middle|\, \Delta \le \overline{\Delta} \right\} \right),$$

where cl denotes the closure in $\mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d; P_p(X))$. We note that the set $\mathcal{S}$ is non-empty: This follows from the fact that finite intersections are clearly non-empty and that each of the sets is a compact subset of $\mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d; P_p(X))$ (finite intersection property of compact sets). So under these very general assumptions, there always exists at least one candidate solution.

One possible selection criterion to find the "best" solution among the candidate solutions $\mathcal{S}$ of Remark 3.13 is by minimizing the Kullback-Leibler divergence with respect to the prior μ (with the idea of this being the most conservative estimate):

$$\mu^{*,y} = \operatorname*{argmin}_{\nu \in \mathcal{S}} \mathcal{D}_{\mathrm{KL}}(\nu || \mu).$$

3.3. **Consistency with the canonical posterior.** In the previous section, we have shown that under very general assumptions on the observables $\mathcal{L}^\Delta(u)$, we can define a set of candidate solutions $\mathcal{S}$ for the BIP in the limit $\Delta \to 0$. In this section, we show that if $\mathcal{L}^\Delta(u) \to \mathcal{L}(u)$ converges to a unique limit (even in an average sense), then $\mu^{\Delta,y} \to \mu^y$ converges to the unique solution of the BIP with measurement $\mathcal{L}(u)$ with respect to the Wasserstein distance W_1. In particular, the set of candidate solutions $\mathcal{S}$ identified in Remark 3.13 is in this case given by $\mathcal{S} = \{\mu^y\}$. We term this posterior μ^y as the *canonical posterior*.

***Theorem* 3.14.** Under the noise assumption 3.1. Fix a prior $\mu \in \mathcal{P}_2(X)$. Let $\mu^{\Delta,y}$ and μ^y denote the posteriors for the BIP with observables $\mathcal{L}^\Delta$ and $\mathcal{L}$, respectively. Assume that there exists a constant $M > 0$, such that

$$\left\|\mathcal{L}^\Delta(u)\right\|_{L^2(\mu)}, \ \|\mathcal{L}(u)\|_{L^2(\mu)} \leq M. \quad \forall \Delta > 0.$$

Then, we have the estimate

$$W_1\left(\mu^{\Delta,y}, \mu^y\right) \leq C\|\mathcal{L}^\Delta(u) - \mathcal{L}(u)\|_{L^2(\mu)},$$

where $C = C(\Gamma, \mu, y, M)$ depends on the prior μ, the measurement $y \in R^d$ and the upper bound M, but is independent of Δ.

Proof. For any $\Phi \in \mathrm{Lip}$, such that $\|\Phi\|_{\mathrm{Lip}} \leq 1$, we find

$$\begin{aligned}
\int_X \Phi(u)\left[d\mu^{\Delta,y}(u) - d\mu^y(u)\right] &= \int_X [\Phi(u) - \Phi(0)]\left[d\mu^{\Delta,y}(u) - d\mu^y(u)\right] \\
&= \int_X [\Phi(u) - \Phi(0)]\left[\frac{d\mu^{\Delta,y}}{d\mu} - \frac{d\mu^y}{d\mu}\right] d\mu(u) \\
&\leq \int_X \|u\|_X \left|\frac{d\mu^{\Delta,y}}{d\mu} - \frac{d\mu^y}{d\mu}\right| d\mu(u) \\
&\leq \|u\|_{L^2(\mu)} \left\|\frac{d\mu^{\Delta,y}}{d\mu} - \frac{d\mu^y}{d\mu}\right\|_{L^2(\mu)}.
\end{aligned}$$

By Lemma 3.8, we have

$$\begin{aligned}
\left\|\frac{d\mu^{\Delta,y}}{d\mu} - \frac{d\mu^y}{d\mu}\right\|_{L^2(\mu)} &\leq C\left\|\mathcal{L}^\Delta(u) - \mathcal{L}(u)\right\|_{L^2(\mu)} e^{|y|^2 + \|\mathcal{L}^\Delta(u)\|^2_{L^2(\mu)} + \|\mathcal{L}(u)\|^2_{L^2(\mu)}} \\
&\leq C\left\|\mathcal{L}^\Delta(u) - \mathcal{L}(u)\right\|_{L^2(\mu)} e^{|y|^2 + 2M^2},
\end{aligned}$$

for a constant $C = C(\Gamma)$. Using this estimate, we can now bound

$$\int_X \Phi(u)\left[d\mu^{\Delta,y}(u) - d\mu^y(u)\right] \leq \overline{C}\left\|\mathcal{L}^\Delta(u) - \mathcal{L}(u)\right\|_{L^2(\mu)},$$

where

$$\overline{C} = Ce^{|y|^2 + 2M^2}\left(\int_X \|u\|_X^2\, d\mu(u)\right)^{1/2}.$$

Taking the supremum over all $\Phi(u) \in \mathrm{Lip}$, $\|\Phi\|_{\mathrm{Lip}} \leq 1$ on the left, we obtain the claimed estimate. □

4. Data assimilation

4.1. **Problem setting.** In the context of time-dependent PDEs, one is often not only interested in obtaining an estimate for the (initial) state given individual measurements y, but to track the temporal evolution of a system, given measurements y_1, y_2, ... acquired over time. The data assimilation problem seeks to provide a best estimate for the state u of the system at time t, expressed in terms of a posterior probability measure $\nu_t^y(u)$, given the available measurements $y_1, y_2, \dots$. There are at least two types of data assimilation problems: Following standard terminology, we call *filtering*, the problem of determining the posterior $\nu_t^y(u)$ at time $t \in [0, T]$ from the measurements available up to time t, i.e. from measurements in the time-interval $[0, t)$. The filtering problem thus provides the best prediction given a set of past measurements. On the other hand, if the posterior $\nu_t^y(u)$ at $t \in [0, T]$ is obtained "after the fact", i.e. given a set of measurements acquired during the whole time-interval $[0, T]$, then we speak of the *smoothing* problem. The generic data assimilation problem is schematically illustrated in Figure 1.

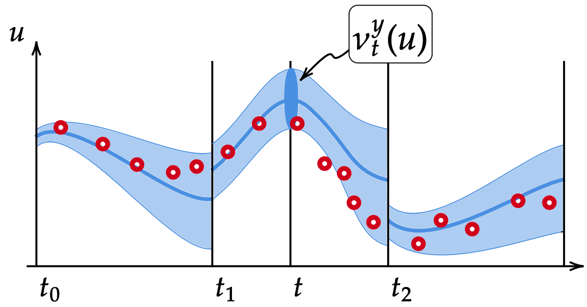

FIGURE 1. Schematic illustration of the data assimilation problem: Measurements (red circles) are used at times $t = t_0$, t_1, ..., to periodically update the posterior measure ν_t^y (indicated by its confidence interval in blue), combining all available information from the deterministic evolution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

In the following, we will focus on the filtering problem, for which we provide a precise formulation below; however, most of the results should apply mutatis mutandis also to the smoothing problem. Due to the weak temporal and spatial regularity properties of the fluid dynamics applications of interest in the present work, simple pointwise measurements of the form $\mathcal{L}(u) = u(x_k, t_k)$ are not well-defined. Thus, we will first discuss an appropriate notion of observables. We make the following definition

***Definition* 4.1** (Eulerian Observables)**.** A mapping $\mathcal{G} : L^1(0,T;L^2_x) \to \mathbb{R}^d$, $u(x,t) \mapsto \mathcal{G}(u) = (\mathcal{G}^1(u), \dots, \mathcal{G}^d(u))$, with $\mathcal{G}^k(u)$ of the form

$$\mathcal{G}^k(u) = \int_0^T \int_D \phi^{(k)}(x,t) g^{(k)}(u(x,t))\, dx\, dt, \tag{4.1}$$

for $u(x,t) \in L^1(0,T;L^2_x)$, is called an **Eulerian observable** (or simply observable), provided that, for all $k = 1, \dots, d$, we have $\phi^{(k)}(x,t) \in L^\infty(D \times [0,T])$ and $g^{(k)}(u)$ is Lipschitz continuous with

$$|g^{(k)}(u) - g^{(k)}(u')| \le C|u - u'|. \tag{4.2}$$

To simplify notation in the following, instead of (4.1) we shall simply write

$$\mathcal{G}(u) = \int_0^T \int_D \phi(x,t) g(u(x,t))\, dx\, dt, \tag{4.3}$$

where $\phi(x,t) := (\phi^{(1)}(x,t), \dots, \phi^{(d)}(x,t))$, $g(u) = (g^{(1)}(u), \dots, g^{(d)}(u))$, and it is understood that the multiplication in (4.3) is carried out componentwise.

It is then straightforward to prove the following result.

***Proposition* 4.2.** An Eulerian observable $\mathcal{G}(u)$ is Lipschitz continuous on $L^1_t([0,T];L^2_x)$, i.e., there exists a constant $C > 0$, such that

$$|\mathcal{G}(u) - \mathcal{G}(u')| \le C \int_0^T \|u - u'\|_{L^2_x}\, dt, \quad \forall u, u' \in L^1([0,T];L^2_x).$$

Proof. This follows immediately from the definition (4.3) of $\mathcal{G}(u)$ and the assumed bound (4.2). $\square$

***Assumption* 4.3** (standing assumption)**.** In the present section, we will make the standing assumption that the approximate solution operators $S_t^\Delta : L^2_x \to L^2_x$ (as well as a possible limit $S_t : L^2_x \to L^2_x$, if it exists) satisfy uniform bounds of the following form:

- Energy admissibility: For any $u \in L^2_x$, we have
$$\|S_t^\Delta(u)\|_{L^2_x} \le C\|u\|_{L^2_x}, \quad \forall\, u \in L^2_x,$$
- Weak time-regularity: There exist constants $L, C > 0$, such that
$$\|S_t^\Delta(u) - S_{t'}^\Delta(u)\|_{H_x^{-L}} \le C|t-t'|, \quad \forall\, u \in L^2_x,\ t,t' \in [0,T],$$
i.e. $t \mapsto S_t^\Delta(u)$ is Lipschitz continuous with values in some negative Sobolev space.

Given a sequence of measurement times $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$ for $N \in \mathbb{N}$, we denote $\delta t_j = t_j - t_{j-1}$. Given observables of the form

$$\mathcal{G}_j : L^1_t([0,\delta t_j); L^2_x) \to \mathbb{R}, \quad \mathcal{G}_j(u) = \int_0^{\delta t_j} \int_D \phi_j(x,t) g_j\left(u(x,t)\right)\, dx\, dt \tag{4.4}$$

the filtering problem at grid scale $\Delta > 0$ is described as follows: The temporal evolution of the system state $u(x,t)$ is modeled by the approximate solution operator S_t^Δ, i.e. $u(x,t) = S_t^\Delta(\overline{u})$, where $\overline{u} = u(x,0)$. We fix a prior $\mu_{\text{prior}} \in \mathcal{P}(L^2_x)$ at the initial time $t = t_0$, representing our best estimate of the state of the system in the absence of measurements. For a sequence of measurements $y_1, \dots, y_N$, we denote $Y_j = (y_1, \dots, y_j)$ the vector of partial measurements up to time t_j. We wish to find a sequence of probability measures $\nu_{t_1}^{\Delta,Y_1}, \nu_{t_2}^{\Delta,Y_2}, \dots, \nu_{t_N}^{\Delta,Y_N}$, where $\nu_{t_j}^{\Delta,Y_j}$ provides a best (probabilistic) estimate of the state of the system at times t_j, given the measurements $Y_j = (y_1, \dots, y_j)$ available up to that time. The measurements are modeled as

$$y_j = \mathcal{L}_j^\Delta(\bar{u}) + \eta_j, \qquad \eta_j \sim \rho_j(y)\, dy, \tag{4.5}$$

where for each j, the noise distribution ρ_j is required to satisfy the assumption 3.1 with a matrix $\Gamma_j \in \mathbb{R}^{d\times d}$ and observable $\mathcal{L}_j^\Delta(\bar{u}) = \mathcal{G}_j(S^\Delta_{t_{j-1}+t}(\bar{u}))$, i.e.,

$$\mathcal{L}_j^\Delta(\bar{u}) = \int_0^{\delta t_j} \int_D \phi_j(x,t) g_j\left(u^\Delta(x, t_{j-1}+t)\right)\, dx\, dt, \tag{4.6}$$

where $u^\Delta(x,t) = S_t^\Delta(\bar{u})$ is the approximate solution corresponding to S_t^Δ, with initial data $\bar{u} = u(x,0)$.

***Remark* 4.4.** More generally, given all measurements $Y_j = (y_1, \dots, y_j)$ obtained in the time interval $[0, t_j]$, we might be interested in ν_t^{Δ,Y_j}, the best probabilistic Bayesian estimate of the state u *at arbitrary time* $t \in [0,T]$, i.e. we can formally consider the conditional probabilities

$$\nu_t^{\Delta,Y_j}(du) = \mathbb{P}[u(\,\cdot\,,t) \in du \,|\, Y_j] = \mathbb{P}[u(\,\cdot\,,t) \in du \,|\, y_1, \dots, y_j],$$

for $t \in [0,T]$. The filtering problem thus considers the case for which all available information at time $t = t_j$ is incorporated in $\nu_{t_j}^{\Delta,Y_j}$, providing the best prediction of the state u at time t_j, given all measurements made during the time-interval $[0, t_j]$.

We note that, under assumption 4.3, Proposition 4.2 implies in particular that

$$\begin{aligned} &\|\mathcal{L}_j^\Delta(u)\|_{L^2(\mu_{\text{prior}})} \le C\left(1 + \|u\|_{L^2(\mu_{\text{prior}})}\right), \\ &\|\mathcal{L}_j^\Delta(u) - \mathcal{L}_j^{\Delta'}(u)\|_{L^2(\mu_{\text{prior}})} \le C \int_{t_{j-1}}^{t_j} \|S_t^\Delta(u) - S_t^{\Delta'}(u)\|_{L^2(\mu_{\text{prior}})}\, dt, \end{aligned} \tag{4.7}$$

where $C = C(\mathcal{G}_j, T) > 0$.

We will denote the log-likelihood function corresponding to the observable $\mathcal{G}_j(u)$ on the j-th time interval $[t_{j-1}, t_j]$ by

$$\Phi_j^{\Delta, y_j}(u) = -\log \rho_j(y_j - \mathcal{G}_j(S_t^\Delta(u))), \quad \forall\, u \in L_x^2. \tag{4.8}$$

$$\begin{tikzcd}
\cdots \arrow[r] & \nu_{t_{j-1}}^{\Delta, Y_{j-1}} \arrow[r] \arrow[d, "y_j"] & \nu_{t_j}^{\Delta, Y_j} \arrow[r] \arrow[d, "y_{j+1}"] & \nu_{t_{j+1}}^{\Delta, Y_{j+1}} \arrow[r] & \cdots \\
& \nu_{t_{j-1}}^{\Delta, Y_j} \arrow[ur] & \nu_{t_j}^{\Delta, Y_{j+1}} \arrow[ur] & &
\end{tikzcd}$$

FIGURE 2. Schematic for the filtering problem: (orange) the correction step incorporates the measurement $y_j = \mathcal{L}_j^\Delta(\bar{u}) + \eta_j$ to update the current best estimate, (blue) the updated estimate is used to predict the next state.

We formalize the filtering problem as follows:

***Definition* 4.5** (Filtering)**.** At the initial time $t = 0$, we fix a prior measure μ_{prior}, and define

$$\nu_{t_0}^{\Delta, Y_0} := S_{0,\#}^\Delta \mu_{\text{prior}}. \tag{4.9}$$

We note that $S_0^\Delta \approx \mathrm{Id}$ is an approximation to the identity. Given times $0 = t_0 < t_1 < \cdots < t_N = T$ and measurements $y_1, \dots, y_N$, the **filtering problem** involves the following two recursive steps.

(1) **Correction step:** Given $\nu_{t_{j-1}}^{\Delta, Y_{j-1}}$ as a prior at time t_{j-1}, solve the Bayesian inverse problem with new measurement $y_j = \mathcal{G}_j(S_t^\Delta(u)) + \eta_j$, for $t \in [0, \delta t_j]$, to obtain a corrected Bayesian estimate

$$d\nu_{t_{j-1}}^{\Delta, Y_j}(u) = \frac{1}{Z_j^\Delta(y_j)} \exp\left(-\Phi_j^{\Delta, y_j}(u)\right) d\nu_{t_{j-1}}^{\Delta, Y_{j-1}}(u). \tag{4.10}$$

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

$$\nu_{t_j}^{\Delta, Y_j} = S_{\delta t_j, \#}^\Delta \nu_{t_{j-1}}^{\Delta, Y_j}, \tag{4.11}$$

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

***Remark* 4.6.** Informally, we can write the correction step (4.10) of the filtering problem as follows:

$$\begin{aligned}\mathbb{P}\left[u(t_{j-1}) \in du \,|\, Y_j\right] = \mathbb{P}&\left[y_j = \mathcal{G}_j(S_t^\Delta(u(t_{j-1}))) \,|\, u(t_{j-1})\right] \\ &\times \mathbb{P}\left[u(t_{j-1}) \in du \,|\, Y_{j-1}\right],\end{aligned}$$

The prediction step (4.11) can be expressed intuitively as

$$\begin{aligned}\mathbb{P}\left[u(t_j) \in du \,|\, Y_j\right] &= \mathbb{P}\left[S_{\delta t_j}^\Delta u(t_{j-1}) \in du \,|\, Y_j\right] \\ &= S_{\delta t_j, \#}^\Delta \mathbb{P}\left[u(t_{j-1}) \in du \,|\, Y_j\right].\end{aligned}$$

The filtering problem is thus defined by recursion, and provides a sequence of best-estimates $\nu_{t_j}^{\Delta, Y_j}$ given the time sequence $0 = t_0, t_1, \dots, t_N$ and measurements $y_1, \dots, y_N$, and based on a fixed prior μ_{prior} at the initial time $t = 0$.

Although the filtering problem is most naturally expressed in terms of the above recursive prediction/correction scheme, it turns out to be beneficial for the analysis of this problem to discuss an equivalent alternative formulation. To this end, we consider $\mu^{\Delta,Y_j} \in \mathcal{P}(L^2_x)$ for $j = 0, \dots, N$, informally given by

$$\mu^{\Delta,Y_j}(du) = \mathbb{P}\left[u(\,\cdot\,,0) \in du \,|\, Y_j\right], \tag{4.12}$$

i.e. the probability of the initial state $u(\,\cdot\,,0) \in du$, given the measurements $Y_j = (y_1, \dots, y_j)$. More precisely, we define $\mu^{\Delta,Y_j}(du)$ as the solution of the BIP with prior μ_{prior} and given the measurement

$$Y_j = (\mathcal{L}^\Delta_1(u), \mathcal{L}^\Delta_2(u), \dots, \mathcal{L}^\Delta_j(u)) + (\eta_1, \eta_2, \dots, \eta_j),$$

and $(\eta_1, \eta_2, \dots, \eta_j)$ the measurement noise. For simplicity, we will assume that the random variables $\eta_1, \dots, \eta_j$ at different time-steps are independent. In this case, the law of $(\eta_1, \dots, \eta_j)$ is a simple product,

$$(\eta_1, \dots, \eta_j) \sim \rho_1(y_1)\, dy_1 \otimes \dots \otimes \rho_j(y_j)\, dy_j,$$

and the solution of the above BIP with prior μ_{prior} is given by

$$d\mu^{\Delta,Y_j}(u) = \frac{1}{\mathcal{Z}^\Delta_j(Y_j)} \exp\left(-\sum_{k=1}^{j} \Phi^{\Delta,y_k}_k \circ S^\Delta_{t_{k-1}}(u) \right) d\mu_{\text{prior}}(u), \tag{4.13}$$

where we note that, by (4.8) and the definition of $\mathcal{L}^\Delta_k(u) = \mathcal{G}_j(S^\Delta_{t_{k-1}+t}(u))$, we have that

$$\begin{aligned} \Phi^{\Delta,y_k}_k \circ S^\Delta_{t_{k-1}}(u) &= -\log \rho_k \left(y_k - \mathcal{G}_k(S^\Delta_{t+t_{k-1}}(u)) \right) \\ &= -\log \rho_k \left(y_k - \mathcal{L}^\Delta_k(u) \right), \quad \forall\, u \in L^2_x, \end{aligned} \tag{4.14}$$

i.e. $\Phi^{\Delta,y_k}_k \circ S^\Delta_{t_{k-1}}$ is the log-likelihood function corresponding to the measurement $y_k = \mathcal{L}_k(u) + \eta_k$, and starting from the initial data $u \in L^2_x$ at time $t = 0$. In (4.13), $\mathcal{Z}^\Delta_j(Y_j)$ is a suitable normalization constant, defined by

$$\mathcal{Z}^\Delta_j(Y_j) = \int_{L^2_x} \exp\left(-\sum_{k=1}^{j} \Phi^{\Delta,y_k}_k \circ S^\Delta_{t_{k-1}}(u) \right) d\mu_{\text{prior}}(u),$$

for $Y_j = (y_1, \dots, y_j)$. We note that $\mathcal{L}^\Delta_k(u) = \mathcal{G}_j(S^\Delta_{t_{k-1}+t}(u))$ (cp. equation (4.6)) can be written as

$$\mathcal{L}^\Delta_k\,(u) = \int_{t_{k-1}}^{t_k} \int_D \phi_k(x, t - t_{k-1}) g_k\left(S^\Delta_t(u)\right)\, dx\, dt,$$

i.e. $\mathcal{L}_k$ provides a measurement of the solution $S^\Delta_t(u)$ with initial data u (at $t = 0$) over the time interval $[t_{k-1}, t_k]$. Consistent with the above identity for μ^{Δ,Y_j} (which is valid for $j \geq 1$), we define

$$\mu^{\Delta,Y_0} := \mu_{\text{prior}}, \tag{4.15}$$

corresponding to the empty sum in (4.13).

We can now state the following proposition, providing an alternative formulation of the filtering problem:

***Proposition* 4.7.** Let $\nu^{\Delta,Y_j}_{t_j}$ denote the recursively computed sequence of probability measures in the filtering problem (cp. Definition 4.5). Let μ^{Δ,Y_j} be given by (4.13). Then, we have the identity

$$\nu^{\Delta,Y_j}_{t_j} = S^\Delta_{t_j,\#} \mu^{\Delta,Y_j}, \tag{4.16}$$

i.e. $\nu^{\Delta,Y_j}_{t_j}$ is given by the push-forward of μ^{Δ,Y_j} to time $t = t_j$.

***Remark* 4.8.** The content of Proposition 4.7 is intuitively clear: The measure $\mu^{\Delta,Y_j}(u)$ provides the best Bayesian estimate for the initial state $u(\,\cdot\,,t)$ at $t=0$ given the measurements $Y_j=(y_1,\dots,y_j)$ acquired over the interval $[0,t_j]$. Proposition 4.7 expresses the fact that the best Bayesian estimate for the state $u(x,t_j)$ at time t_j should simply be given by evolving the best initial estimate μ^{Δ,Y_j} (given Y_j), forward in time to $t=t_j$, via the solution operator $S^{\Delta}_{t_j}$.

***Remark* 4.9.** Proposition 4.7 also indicates a consistent definition of ν_t^{Δ,Y_j} for *any* $t\in[0,T]$. Indeed, the best Bayesian estimate for $u(\,\cdot\,,t)$ given the measurements Y_j is simply given by

$$\nu_t^{\Delta,Y_j} = S^{\Delta}_{t,\#}\mu^{\Delta,Y_j}, \tag{4.17}$$

We now come to the proof of Proposition 4.7.

Proof of Proposition 4.7. We proceed by induction on j. The case $j=0$ is trivial, since

$$\nu_{t_0}^{\Delta,Y_0} \overset{(4.9)}{\overset{\downarrow}{=}} S^{\Delta}_{t_0,\#}\mu_{\text{prior}} \overset{(4.15)}{\overset{\downarrow}{=}} S^{\Delta}_{t_0,\#}\mu^{\Delta,Y_0}.$$

For $j\geq 1$, we integrate against an arbitrary, integrable (cylindrical) test function $\Phi(u)$ to find, from the prediction step (4.11) of the filtering problem:

$$\int_{L^2_x}\Phi(u)\,d\nu_{t_j}^{\Delta,Y_j} = \int_{L^2_x}\Phi(u)\,d\left[S^{\Delta}_{\delta t_j,\#}\nu_{t_{j-1}}^{\Delta,Y_j}\right] = \int_{L^2_x}\Phi\left(S^{\Delta}_{\delta t_j}(u)\right)d\nu_{t_{j-1}}^{\Delta,Y_j}(u).$$

Substitution of the correction step (4.10), yields

$$\int_{L^2_x}\Phi(u)\,d\nu_{t_{j-1}}^{\Delta,Y_j} = \int_{L^2_x}\Phi\left(S^{\Delta}_{\delta t_j}(u)\right)q_j^{\Delta}(u)d\nu_{t_{j-1}}^{\Delta,Y_{j-1}}(u),$$

where

$$q_j^{\Delta}(u) = \frac{1}{Z_j^{\Delta}(y_j)}\exp\left(-\Phi_j^{\Delta,y_j}(u)\right).$$

By the induction hypothesis, the measure $\nu_{t_{j-1}}^{\Delta,Y_{j-1}}$ can be written as a push-forward (4.16):

$$\nu_{t_{j-1}}^{\Delta,Y_{j-1}} = S^{\Delta}_{t_{j-1},\#}\mu^{\Delta,Y_{j-1}}.$$

Thus, substituting above, we find

$$\begin{aligned}\int_{L^2_x}\Phi(u)\,d\nu_{t_j}^{\Delta,Y_j} &= \int_{L^2_x}\Phi(S^{\Delta}_{\delta t_j}(u))\,q_j^{\Delta}(u)\,d\left[S^{\Delta}_{t_{j-1},\#}\mu^{\Delta,Y_{j-1}}(u)\right]\\ &= \int_{L^2_x}\Phi(S^{\Delta}_{t_j}(u))q_j^{\Delta}(S^{\Delta}_{t_{j-1}}(u))\,d\mu^{\Delta,Y_{j-1}}(u),\end{aligned}$$

where we have used that $S^{\Delta}_{t_{j-1}}\circ S^{\Delta}_{\delta t_j} = S^{\Delta}_{t_j}$ to simplify the argument of Φ in the last step. We now note that, by our definition of q_j^{Δ} and $\mu^{\Delta,Y_{j-1}}$, we have

$$\begin{aligned}q_j^{\Delta}(S^{\Delta}_{t_{j-1}}(u))\,d\mu^{\Delta,Y_{j-1}}(u) &\propto \exp\left(-\Phi_j^{\Delta,y_j}\circ S^{\Delta}_{t_{j-1}}(u)\right)\\ &\qquad\times\exp\left(-\sum_{k=1}^{j-1}\Phi_k^{\Delta,y_k}\circ S^{\Delta}_{t_{k-1}}(u)\right)d\mu_{\text{prior}}(u)\\ &= \exp\left(-\sum_{k=1}^{j}\Phi_k^{\Delta,y_k}\circ S^{\Delta}_{t_{k-1}}(u)\right)d\mu_{\text{prior}}(u),\end{aligned}$$

with a proportionality constant that can be determined by normalization. From our definition (4.13), the last expression is equal to μ^{Δ,Y_j}, and hence

$$\int_{L^2_x} \Phi(u)\, d\nu^{\Delta,Y_j}_{t_j}(u) = \int_{L^2_x} \Phi(S^\Delta_{t_j}(u))\, d\mu^{\Delta,Y_j}(u) = \int_{L^2_x} \Phi(u)\, d\left[S^\Delta_{t_j,\#}\mu^{\Delta,Y_j}\right](u).$$

Since Φ was an arbitrary (cylindrical) test function, the claimed identity follows. $\square$

***Remark* 4.10.** We recall that by Remark 3.10, for any $\phi \in L^1(\mu_{\text{prior}})$, the mapping

$$Y_j \mapsto \int_{L^2_x} \phi(u)\, d\mu^{\Delta,Y_j}(u),$$

is analytic in Y_j. As a consequence of Proposition 4.7, it follows that also

$$Y_j \mapsto \int_{L^2_x} \phi(u)\, d\nu^{\Delta,Y_j}_t(u) = \int_{L^2_x} \phi(S^\Delta_t(u))\, d\mu^{\Delta,Y_j}(u),$$

is analytic in Y_j, *independently of the smoothness of the solution operator* S^Δ_t.

4.2. **Stability with respect to measurements.** 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 representation (4.13) of the previous section and the stability results for the BIP in section 3. Due to the low a priori time-regularity of the time-dependent mapping $t \mapsto \nu^{\Delta,Y_j}_t$, we will formulate the stability in the space $L^1_t(\mathcal{P}) = L^1([0,T]; \mathcal{P}(L^2_x))$ [19], defined as the set of all weak-$*$ measurable mappings $[0,T] \to \mathcal{P}(L^2_x)$, $t \mapsto \nu_t$, such that

$$\int_0^T \|u\|_{L^2_x}\, d\nu_t(u)\, dt < \infty,$$

with metric

$$d_T(\nu_t, \nu'_t) := \int_0^T W_1(\nu_t, \nu'_t)\, dt, \qquad \forall\, \nu_t, \nu'_t \in L^1([0,T]; \mathcal{P}(L^2_x)).$$

This space has been introduced in [19, Definition 2.3]; it is not difficult to prove that $(L^1_t(\mathcal{P}), d_T)$ is a complete metric space (cp. [19, Proposition 2.4]).

We can now state the following lemma

***Lemma* 4.11.** Let $T > 0$. Let $\mu_{\text{prior}} \in \mathcal{P}_1(L^2_x)$ be a prior such that $\|u\|_{L^1(\mu_{\text{prior}})} < \infty$. Let ν^{Δ,Y_j}_t be given by (4.17) for $t \in [0,T]$, so that, formally, $\nu^{\Delta,Y_j}_t(du) = \mathbb{P}\left[u(\cdot,t) \in du \,|\, Y_j\right]$. Then for any $R > 0$, there exists $C = C(R) > 0$, such that for any $t, \delta t \geq 0$, we have

$$\int_t^{t+\delta t} W_1\left(\nu^{\Delta,Y_j}_\tau, \nu^{\Delta,Y_j{}'}_\tau\right) d\tau \leq C\delta t \left(\sum_{k=1}^j |y_k - y'_k|^2_{\Gamma_k}\right)^{1/2}, \tag{4.18}$$

for all $Y_j = (y_1, \dots, y_j)$, $Y'_j = (y'_1, \dots, y'_j)$ such that $\sqrt{\sum_{k=1}^j |y_k|^2_{\Gamma_k}} \leq R$, $\sqrt{\sum_{k=1}^j |y'_k|^2_{\Gamma_k}} \leq R$.

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

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

By (4.17), we have $\nu_t^{\Delta,Y_j} = S_{t,\#}^{\Delta}\mu^{\Delta,Y_j}$, where μ^{Δ,Y_j} solves a BIP and is given by (4.13). Since μ^{Δ,Y_j} is the solution of a standard BIP with noise $\eta = (\eta_1, \dots, \eta_j)$ satisfying assumption 3.1, then by Lemma 3.7, we obtain

$$\left\| \frac{d\mu^{\Delta,Y_j}}{d\mu_{\text{prior}}} - \frac{d\mu^{\Delta,Y_j'}}{d\mu_{\text{prior}}} \right\|_{L^\infty(\mu_{\text{prior}})} \le C|Y_j - Y_j'|_\Gamma. \tag{4.19}$$

Let $\Phi(u) \in \mathrm{Lip}(L^2_x)$ be a function with Lipschitz constant ≤ 1. Then there exists $g(u)$ such that

$$\Phi(u) - \Phi(0) = g(u)\|u\|_{L^2_x}, \qquad |g(u)| \le 1.$$

Now note that

$$\begin{aligned}
\int_{L^2_x} \Phi(u)\left[d\nu_t^{\Delta,Y_j} - d\nu_t^{\Delta,Y_j'}\right] &= \int_{L^2_x} [\Phi(u) - \Phi(0)]\left[d\nu_t^{\Delta,Y_j} - d\nu_t^{\Delta,Y_j'}\right] \\
&= \int_{L^2_x} g(u)\|u\|_{L^2_x} S_{t,\#}^{\Delta}\left[d\mu^{\Delta,Y_j} - d\mu^{\Delta,Y_j'}\right] \\
&= \int_{L^2_x} g(S_t^{\Delta}(u))\|S_t^{\Delta}(u)\|_{L^2_x}\left[\frac{d\mu^{\Delta,Y_j}}{d\mu_{\text{prior}}} - \frac{d\mu^{\Delta,Y_j'}}{d\mu_{\text{prior}}}\right] d\mu_{\text{prior}}(u) \\
&\le \int_{L^2_x} \left|g(S_t^{\Delta}(u))\right| \|S_t^{\Delta}(u)\|_{L^2_x}\left|\frac{d\mu^{\Delta,Y_j}}{d\mu_{\text{prior}}} - \frac{d\mu^{\Delta,Y_j'}}{d\mu_{\text{prior}}}\right| d\mu_{\text{prior}}(u) \\
&\overset{\substack{|g(u)|\le 1,\\ \|S_t^{\Delta}(u)\|_{L^2_x} \le C\|u\|_{L^2_x}}}{\le} C\int_{L^2_x} \|u\|_{L^2_x}\left|\frac{d\mu^{\Delta,Y_j}}{d\mu_{\text{prior}}} - \frac{d\mu^{\Delta,Y_j'}}{d\mu_{\text{prior}}}\right| d\mu_{\text{prior}}(u) \\
&\le C\left(\int_{L^2_x} \|u\|_{L^2_x}\, d\mu_{\text{prior}}(u)\right)\left\|\frac{d\mu^{\Delta,Y_j}}{d\mu_{\text{prior}}} - \frac{d\mu^{\Delta,Y_j'}}{d\mu_{\text{prior}}}\right\|_{L^\infty(\mu_{\text{prior}})}.
\end{aligned}$$

Taking the supremum over all $\Phi(u)$ such that $\|\Phi\|_{\mathrm{Lip}} \le 1$ on the left, and noting the upper bound (4.19) on the last term, we find

$$W_1\left(\nu_t^{\Delta,Y_j}, \nu_t^{\Delta,Y_j{}'}\right) \le C|Y_j - Y_j'|_\Gamma,$$

where the constant $C > 0$ is independent of Y_j, Y_j'. Integrating in time, we obtain the claimed inequality

$$\int_t^{t+\delta t} W_1\left(\nu_t^{\Delta,Y_j}, \nu_t^{\Delta,Y_j{}'}\right) dt \le C\delta t|Y_j - Y_j'|_\Gamma.$$

□

We will finally state a general stability theorem for the solution of the filtering problem. To this end, we introduce the following notation

***Definition* 4.12.** Given times $0 = t_0 < t_1 < \dots < t_N = T$, and measurements $y_1, \dots, y_N$, we denote by $\nu^{\Delta,\boldsymbol{y}}$, with $\boldsymbol{y} = (y_1, \dots, y_N)$ the solution of the associated filtering problem, i.e.,

$$\nu_t^{\Delta,\boldsymbol{y}} := \begin{cases} \nu_t^{\Delta,Y_0}, & t \in [0, t_1), \\ \nu_t^{\Delta,Y_1}, & t \in [t_1, t_2), \\ \vdots & \\ \nu_t^{\Delta,Y_{N-1}}, & t \in [t_{N-1}, t_N), \\ \nu_t^{\Delta,Y_N}, & t \ge t_N, \end{cases} \tag{4.20}$$

***Theorem* 4.13.** Let $\nu_t^{\Delta,\boldsymbol{y}}$ denote the solution of the filtering problem with prior $\mu_{\mathrm{prior}} \in \mathcal{P}_1(L^2_x)$, and measurements $\boldsymbol{y} = (y_1, \dots, y_N)$. Then for any $R > 0$, there exists $C = C(R,T)$, such that

$$\int_0^T W_1\left(\nu_t^{\Delta,\boldsymbol{y}}, \nu_t^{\Delta,\boldsymbol{y}'}\right) dt \le C|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma, \tag{4.21}$$

for all $\boldsymbol{y}, \boldsymbol{y}'$ such that $|\boldsymbol{y}|_\Gamma$, $|\boldsymbol{y}'|_\Gamma \le R$. Here, we use the norm

$$|\boldsymbol{y}|_\Gamma := \left(\sum_{k=1}^N |y_k|^2_{\Gamma_k}\right)^{1/2}.$$

Proof. The claimed stability estimate follows readily from Lemma 4.11: Indeed, $\nu_t^{\Delta,\boldsymbol{y}}$ is defined piece-wise in time, for $t \in [0,T) = [t_0, t_N)$, as

$$\nu_t^{\Delta,\boldsymbol{y}} = \sum_{k=1}^N 1_{[t_{k-1},t_k)}(t)\, \nu_t^{\Delta,Y_{k-1}}.$$

$$\begin{aligned}\int_0^T W_1\left(\nu_t^{\Delta,\boldsymbol{y}}, \nu_t^{\Delta,\boldsymbol{y}'}\right) dt &= \sum_{k=1}^N \int_{t_{k-1}}^{t_k} W_1\left(\nu_t^{\Delta,Y_{k-1}}, \nu_t^{\Delta,Y'_{k-1}}\right) dt \\ &\le C \sum_{k=1}^N \delta t_k |Y_{k-1} - Y'_{k-1}|_\Gamma \\ &\le CT|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma.\end{aligned}$$

□

4.3. **Compactness properties.** Our second main result for the filtering problem is a conditional compactness result, motivated by the study of statistical solutions of the compressible and incompressible Euler equations in [9, 19, 20]. In [19], the authors study the forward problem for statistical initial data μ a probability measure on $L^2_x := L^2(\mathbb{T}^d)$. They prove that under Assumption 4.3, the sequence of discretized approximate solutions $\mu_t^\Delta := (S_t^\Delta)_\# \mu$ (push-forward by the discretized solution operator) is compact in $\mathcal{P}_1(L^2_x)$, provided that the following measure of average two-point correlations

$$\mathscr{S}_2^T(\mu_t^\Delta; r) := \left(\int_0^T \int_{L^2_x} \mathscr{S}_2(u;r)^2 \, d\mu_t^\Delta(u)\, dt\right)^{1/2}, \tag{4.22}$$

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

$$\mathscr{S}_2(u;r) := \left(\int_D \fint_{B_r(0)} |u(x+h) - u(x)|^2 \, dh\, dx\right)^{1/2}, \tag{4.23}$$

measures the average of two-point correlations of u: More precisely, if μ_t^Δ is of the form $\mu_t^\Delta = S^\Delta_{t,\#}\mu_0$, $\mu_0 \in \mathcal{P}_2(L^2_x)$, with $S_t^\Delta : L^2_x \to L^2_x$ satisfying assumption 4.3, and if we have $\mathscr{S}_2^T(\mu_t^\Delta; r) \le \phi(r)$, for some modulus of continuity $\phi(r)$ (i.e., $\phi(r) > 0$, $\forall r$ and $\lim_{r\to 0} \phi(r) = 0$), uniformly in Δ, then μ_t^Δ is compact in $L^1_t(\mathcal{P})$. 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 $D = \mathbb{T}^d$. Numerical evidence for the uniform boundedness of these structure functions for the statistical forward problem has been presented for a variety of initial probability measures μ supported on rough initial data of the two-dimensional incompressible Euler equations in [19, 20], and in the context of hyperbolic conservation laws in [9].

We formulate this observation motivated by the numerical experiments in [9, 19, 20] abstractly as the following assumption:

***Assumption* 4.14.** The prior μ_{prior} has finite second moments,

$$\int_{L^2_x} \|u\|^2_{L^2_x} \, d\mu_{\mathrm{prior}}(u) < \infty,$$

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

$$\mathscr{S}_2^T(S^\Delta_{t,\#}\mu_{\mathrm{prior}}; r) \le \phi(r), \quad \forall r > 0, \ t \in [0,T], \tag{4.24}$$

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

***Remark* 4.15.** Let $\mu = \mu_{\mathrm{prior}} \in \mathcal{P}(L^2_x)$ be a probability measure with finite second moments. We note that under our standing Assumption 4.3 on the uniform boundedness of the S^Δ_t, and if S^Δ_t converges to S_t in $L^1([0,T]; L^1(\mu))$, then Assumption 4.14 is automatically satisfied. Indeed, for any $\Phi \in \mathrm{Lip}(L^2_x)$ with $\|\Phi\|_{\mathrm{Lip}} \le 1$, we have

$$\begin{aligned}\int_{L^2_x} \Phi(u) \left[d\left(S^\Delta_{t,\#}\mu\right) - d\left(S_{t,\#}\mu\right)\right] &= \int_{L^2_x} \left[\Phi(S^\Delta_t(u)) - \Phi(S_t(u))\right] d\mu(u) \\ &\le \int_{L^2_x} \|S^\Delta_t(u) - S_t(u)\|_{L^2_x} \, d\mu(u).\end{aligned}$$

Taking the supremum over all such Φ and integrating over $[0,T]$, we obtain

$$\int_0^T W_1\left(S^\Delta_{t,\#}\mu, S_{t,\#}\mu\right) \, dt \le \int_{L^2_x} \int_0^T \|S^\Delta_t(u) - S_t(u)\|_{L^2_x} \, dt \, d\mu(u).$$

Thus, the assumption that $S^\Delta_t(u) \to S_t(u)$ in $L^1([0,T]; L^1(\mu))$ implies that

$$\int_0^T W_1(S^\Delta_{t,\#}\mu, S_{t,\#}\mu) \, dt \to 0, \quad (\Delta \to 0),$$

i.e., that $S^\Delta_{t,\#}\mu \to S_{t,\#}\mu$ in $L^1_t(\mathcal{P}) = L^1([0,T]; \mathcal{P}(L^2_x))$. In particular, $S^\Delta_{t,\#}\mu$ is compact in $L^1_t(\mathcal{P})$, from which it follows (cp. Proposition A.2) that there exists a modulus of continuity $\phi(r)$, such that $\mathscr{S}_2^T(S^\Delta_{t,\#}\mu; r) \le \phi(r)$.

We also note that if there exists a set $A \subset L^2_x$, such that $\mu_{\mathrm{prior}}(A) = 1$, and $S^\Delta_t(u) \to S_t(u)$ point-wise for all $u \in A$, and almost all $t \in [0,T]$, then $S^\Delta_t(u) \to S_t(u)$ in $L^1([0,T]; L^1(\mu))$. Indeed, this follows from the point-wise bound

$$\|S^\Delta_t(u) - S_t(u)\|_{L^2_x} \le \|S^\Delta_t(u)\|_{L^2_x} + \|S_t(u)\|_{L^2_x} \le 2\|u\|_{L^2_x},$$

the fact that $\int \|u\|_{L^2_x} \, d\mu(u) < \infty$, and the dominated convergence theorem.

Conditional on Assumption 4.14, we can now prove a compactness result for the filtering problem:

***Lemma* 4.16.** Let $\nu^{\Delta,\boldsymbol{y}}_t$ be the solution of the filtering problem with prior $\mu_{\mathrm{prior}} \in \mathcal{P}_2(L^2_x)$, such that $\|u\|_{L^2(\mu_{\mathrm{prior}})} < \infty$, and measurements $\boldsymbol{y} = (y_1, \dots, y_N)$. If assumption 4.14 holds, then $\nu^{\Delta,\boldsymbol{y}}_t$ is a compact sequence in $L^1_t(\mathcal{P})$, as $\Delta \to 0$.

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

$$\begin{aligned}
\mathscr{S}_2^T\left(\nu_t^{\Delta,\boldsymbol{y}};r\right)^2 &= \int_0^T \int_{L_x^2} \mathscr{S}_2(u;r)^2\, d\nu_t^{\Delta,\boldsymbol{y}}(u)\, dt \\
&= \sum_{k=1}^N \int_{t_{k-1}}^{t_k} \int_{L_x^2} \mathscr{S}_2(u;r)^2\, d\nu_t^{\Delta,\boldsymbol{y}}(u)\, dt \\
&= \sum_{k=1}^N \int_{t_{k-1}}^{t_k} \int_{L_x^2} \mathscr{S}_2(u;r)^2\, d\left[S_{t,\#}^{\Delta}\mu^{\Delta,Y_{k-1}}\right](u)\, dt \\
&= \sum_{k=1}^N \int_{t_{k-1}}^{t_k} \int_{L_x^2} \mathscr{S}_2(S_t^\Delta(u);r)^2\, d\mu^{\Delta,Y_{k-1}}(u)\, dt.
\end{aligned}$$

We recall that on the last line, $\mu^{\Delta,Y_{k-1}}$ is the solution of the BIP for the initial data (4.13). Using the uniform boundedness result for such BIP, Lemma 3.6, we conclude that

$$\begin{aligned}
\frac{d\mu^{\Delta,Y_{k-1}}}{d\mu_{\text{prior}}} &\le \exp\left(\sum_{j=1}^{k-1}|y_j|_{\Gamma_j}^2 + \sum_{j=1}^{k-1}\|\mathcal{L}_j^\Delta\|_{L^2(\mu_{\text{prior}})}^2\right) \\
&\le \exp\left(|\boldsymbol{y}|_\Gamma^2 + \sum_{j=1}^{N}\|\mathcal{L}_j^\Delta\|_{L^2(\mu_{\text{prior}})}^2\right).
\end{aligned}$$

By our standing boundedness assumption 4.3, and the upper bound on observables (cp. (4.7)), it follows that there exists a constant $C>0$, such that

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

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

$$\frac{d\mu^{\Delta,Y_{k-1}}}{d\mu_{\text{prior}}} \le C,$$

and hence

$$\begin{aligned}
\mathscr{S}_2^T\left(\nu_t^{\Delta,\boldsymbol{y}};r\right)^2 &= \sum_{k=1}^N \int_{t_{k-1}}^{t_k} \int_{L_x^2} \mathscr{S}_2(S_t^\Delta(u);r)^2\, d\mu^{\Delta,Y_{k-1}}(u)\, dt \\
&\le \sum_{k=1}^N \int_{t_{k-1}}^{t_k} \int_{L_x^2} \mathscr{S}_2(S_t^\Delta(u);r)^2\, C d\mu_{\text{prior}}(u)\, dt \\
&= C\sum_{k=1}^N \int_{t_{k-1}}^{t_k} \int_{L_x^2} \mathscr{S}_2(u;r)^2\, d\left[S_{t,\#}^\Delta\mu_{\text{prior}}\right](u)\, dt \\
&= C\int_0^T \int_{L_x^2} \mathscr{S}_2(u;r)^2\, d\left[S_{t,\#}^\Delta\mu_{\text{prior}}\right](u)\, dt \\
&= C\mathscr{S}_2^T\left(S_{t,\#}^\Delta\mu_{\text{prior}};r\right)^2 \\
&\le C\phi(r)^2.
\end{aligned}$$

The last estimate follows from assumption 4.14. Thus, $\mathscr{S}_2^T(\nu_t^{\Delta,\boldsymbol{y}};r) \le \sqrt{C}\phi(r)$ is uniformly bounded by a modulus of continuity, implying compactness in $L_t^1(\mathcal{P})$. □

Combining the uniform stability result, Theorem 4.13 with the point-wise compactness result, Lemma 4.16, we can formulate the following theorem:

***Theorem* 4.17.** Fix a prior $\mu_{\mathrm{prior}} \in \mathcal{P}_1(L^2_x)$, such that $\|u\|_{L^2(\mu_{\mathrm{prior}})} < \infty$. Let $0 = t_0 < t_1 < \dots < t_N = T$ be a strictly increasing sequence of times for fixed $N \in \mathbb{N}$. Let $\boldsymbol{y} = (y_1, y_2, \dots, y_N) \in \mathbb{R}^{d\times N}$ be a sequence of measurements. Let $\nu_t^{\Delta,\boldsymbol{y}}$, $j = 0, \dots, N$, be the solution of the associated filtering problem. If assumption 4.14 holds, then the sequence $\nu_t^{\Delta,\boldsymbol{y}}$ is pre-compact in $C_{\mathrm{loc}}(\mathbb{R}^{d\times N}; L^1_t(\mathcal{P}))$, as $\Delta \to 0$. In fact, there exists a subsequence $\Delta_k \to 0$, and $\mu_t^{*,\boldsymbol{y}}$ with

$$\boldsymbol{y} \mapsto \nu_t^{*,\boldsymbol{y}} \in \mathrm{Lip}_{\mathrm{loc}}\left(\mathbb{R}^{N\times d}; L^1_t(\mathcal{P})\right),$$

such that

$$d_T\left(\nu_t^{\Delta_k,\boldsymbol{y}}, \nu_t^{*,\boldsymbol{y}}\right) = \int_0^T W_1\left(\nu_t^{\Delta_k,\boldsymbol{y}}, \nu_t^{*,\boldsymbol{y}}\right)\, dt \to 0,$$

converges locally uniformly in $\boldsymbol{y}$.

Proof. By Theorem 4.13, the mapping

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

is uniformly bounded on any compact subset $K \subset \mathbb{R}^{d\times N}$ and uniformly equicontinuous on K. By Lemma 4.16, the sets

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

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. $\square$

***Remark* 4.18.** In practice, a very popular choice of priors are Gaussian priors $\mu_{\mathrm{prior}} \sim \mathcal{N}(m,\Gamma)$ on function spaces, i.e. priors μ_{prior} such that each finite-dimensional projection is Gaussian. We point out in passing that Theorems 3.9, 3.12, 4.13 and 4.17 on the stability and compactness properties of approximate posteriors apply in particular, when the prior is Gaussian.

4.4. **Consistency with the canonical solution.** We finally discuss the consistency of the above convergence result for the approximate filtering problems based on the discretized solution operator S_t^Δ, and the limiting filtering problem with solution operator S_t. More precisely, we show that if $S_t^\Delta(u) \to S_t(u)$ converges in a suitable sense, then $\nu_t^{\Delta,\boldsymbol{y}} \to \nu_t^{\boldsymbol{y}}$ in $L^1_t(\mathcal{P})$, where $\nu_t^{\boldsymbol{y}}$ denotes the solution of the limiting filtering problem.

***Theorem* 4.19.** Assume that $\mu_{\mathrm{prior}} \in \mathcal{P}_1(L^2_x)$ is such that $\|u\|_{L^2(\mu_{\mathrm{prior}})} < \infty$. Then there exists a constant $C > 0$, independent of Δ, such that

$$\int_0^T W_1\left(\nu_t^{\Delta,\boldsymbol{y}}, \nu_t^{\boldsymbol{y}}\right)\, dt \le C \int_0^T \|S_t^\Delta(u) - S_t(u)\|_{L^2(\mu_{\mathrm{prior}})}\, dt.$$

In particular, if $S_t^\Delta(u) \to S_t(u)$ in $L^1([0,T]; L^2(\mu_{\mathrm{prior}}))$, then $\nu_t^{\Delta,\boldsymbol{y}} \to \nu_t^{\boldsymbol{y}}$ in $L^1_t(\mathcal{P})$.

Proof. Fix $t \in [0,T]$ and $j \in \{0.\dots, N-1\}$, such that $t \in [t_j, t_{j+1}]$. Then, by the definition of $\nu_t^{\Delta,\boldsymbol{y}}$, we have

$$\nu_t^{\Delta,\boldsymbol{y}} = \nu_t^{\Delta,Y_j} = S^\Delta_{t,\#}\mu^{\Delta,Y_j},$$

where the last equality follows form (4.17). Given $\Phi \in \mathrm{Lip}$, with $\|\Phi\|_{\mathrm{Lip}} \le 1$ and $\Phi(0) = 0$, we find

$$\begin{aligned}
\int_{L^2_x} \Phi(u) \left[d\nu_t^{\Delta,\boldsymbol{y}}(u) - d\nu_t^{\boldsymbol{y}}(u) \right] &= \int_{L^2_x} \Phi(u) \left[S^\Delta_{t,\#} d\mu^{\Delta,Y_j}(u) - S_{t,\#} d\mu^{Y_j}(u) \right] \\
&= \int_{L^2_x} \Phi(u) \left[S^\Delta_{t,\#} d\mu^{\Delta,Y_j}(u) - S^\Delta_{t,\#} d\mu^{Y_j}(u) \right] \\
&\quad + \int_{L^2_x} \Phi(u) \left[S^\Delta_{t,\#} d\mu^{Y_j}(u) - S_{t,\#} d\mu^{Y_j}(u) \right] \\
&=: (I) + (II).
\end{aligned}$$

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

$$\begin{aligned}
(I) &= \int_{L^2_x} \Phi(u) \left[S^\Delta_{t,\#} d\mu^{\Delta,Y_j}(u) - S^\Delta_{t,\#} d\mu^{Y_j}(u) \right] \\
&= \int_{L^2_x} \Phi(S^\Delta_t(u)) \left[\frac{d\mu^{\Delta,Y_j}}{d\mu_{\mathrm{prior}}} - \frac{d\mu^{Y_j}}{d\mu_{\mathrm{prior}}} \right] d\mu_{\mathrm{prior}}(u) \\
&\le C \int_{L^2_x} \|u\|_{L^2_x} \left| \frac{d\mu^{\Delta,Y_j}}{d\mu_{\mathrm{prior}}} - \frac{d\mu^{Y_j}}{d\mu_{\mathrm{prior}}} \right| d\mu_{\mathrm{prior}}(u) \\
&\le C \|u\|_{L^2(\mu_{\mathrm{prior}})} \left\| \frac{d\mu^{\Delta,Y_j}}{d\mu_{\mathrm{prior}}} - \frac{d\mu^{Y_j}}{d\mu_{\mathrm{prior}}} \right\|_{L^2(\mu_{\mathrm{prior}})}.
\end{aligned}$$

The last term can be estimated using Lemma 3.8, recalling that μ^{Δ,Y_j} is defined as the posterior with prior μ_{prior} and given the measurements $(\mathcal{L}^\Delta_1, \dots, \mathcal{L}^\Delta_j)$ of the form (4.6). Lemma 3.8 therefore yields

$$\left\| \frac{d\mu^{\Delta,Y_j}}{d\mu_{\mathrm{prior}}} - \frac{d\mu^{Y_j}}{d\mu_{\mathrm{prior}}} \right\|_{L^2(\mu_{\mathrm{prior}})} \le C \left(\sum_{\ell=1}^{j} \|\mathcal{L}^\Delta_\ell(u) - \mathcal{L}_\ell(u)\|^2_{L^2(\mu_{\mathrm{prior}})} \right)^{1/2},$$

for some constant $C > 0$ depending only on the prior μ_{prior}; here, we have used the fact that Y_j is fixed, and that $\|\mathcal{L}^\Delta_\ell(u)\|_{L^2(\mu_{\mathrm{prior}})}, \|\mathcal{L}_\ell(u)\|_{L^2(\mu_{\mathrm{prior}})} \le C(1 + \|u\|_{L^2(\mu_{\mathrm{prior}})}) < \infty$ are bounded independently of $\Delta > 0$, which allows us to bound the additional exponential factor in Lemma 3.8 uniformly in Δ. Continuing, we note that the observables are Lipschitz continuous by assumption; Indeed, by (4.7), we have

$$\|\mathcal{L}^\Delta_\ell(u) - \mathcal{L}_\ell(u)\|_{L^2(\mu_{\mathrm{prior}})} \le C \int_{t_{\ell-1}}^{t_\ell} \|S^\Delta_t(u) - S_t(u)\|_{L^2(\mu_{\mathrm{prior}})} \, dt.$$

It follows that

$$(I) \le C \left(\sum_{\ell=1}^{j} \left[\int_{t_{\ell-1}}^{t_\ell} \left\| S^\Delta_t(u) - S_t(u) \right\|_{L^2(\mu_{\mathrm{prior}})} \, dt \right]^2 \right)^{1/2}.$$

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

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

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

$$
\begin{aligned}
(I) &\le C \int_0^T \left(\sum_{\ell=1}^j |F(t,\ell)|^2\right)^{1/2} dt = C \sum_{\ell=1}^j \int_{t_{\ell-1}}^{t_\ell} |F(t,\ell)|\, dt \\
&\le C \int_0^T \|S_t^\Delta(u) - S_t(u)\|_{L^2(\mu_{\mathrm{prior}})}\, dt.
\end{aligned}
$$

To estimate the second term, we note that

$$
\begin{aligned}
\int_{L^2_x} \Phi(u) \left[S^\Delta_{t,\#} d\mu^{Y_j}(u) - S_{t,\#} d\mu^{Y_j}(u)\right] &= \int_{L^2_x} \left[\Phi(S_t^\Delta(u)) - \Phi(S_t(u))\right] \, d\mu^{Y_j}(u) \\
&\le \int_{L^2_x} \left\|S_t^\Delta(u) - S_t(u)\right\|_{L^2_x} \, d\mu^{Y_j}(u) \\
&\le C \int_{L^2_x} \left\|S_t^\Delta(u) - S_t(u)\right\|_{L^2_x} \, d\mu_{\mathrm{prior}}(u) \\
&\le C \left\|S_t^\Delta(u) - S_t(u)\right\|_{L^2(\mu_{\mathrm{prior}})}.
\end{aligned}
$$

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

$$
\begin{aligned}
\int_{L^2_x} \Phi(u) \left[d\nu_t^{\Delta,\boldsymbol{y}}(u) - d\nu_t^{\boldsymbol{y}}(u)\right] \le C \left\|S_t^\Delta(u) - S_t(u)\right\|_{L^2(\mu_{\mathrm{prior}})} \\
+ C \int_0^T \left\|S_t^\Delta(u) - S_t(u)\right\|_{L^2(\mu_{\mathrm{prior}})} \, dt.
\end{aligned}
$$

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

$$
\int_0^T W_1\left(\nu_t^{\Delta,\boldsymbol{y}}, d\nu_t^{\boldsymbol{y}}\right) \, dt \le C \int_0^T \left\|S_t^\Delta(u) - S_t(u)\right\|_{L^2(\mu_{\mathrm{prior}})} \, dt,
$$

where $C > 0$ is independent of Δ. □

5. Applications

In the present section, we discuss several concrete applications of the abstract results obtained in the previous sections.

5.1. Incompressible Euler.

The incompressible Euler equations model the motion of an ideal inviscid fluid, and are given by the following system of PDEs for the fluid velocity field $u = u(x,t)$:

$$
\begin{cases}
\partial_t u + \operatorname{div}(u \otimes u) + \nabla p = 0, \\
\operatorname{div}(u) = 0, \\
u(\,\cdot\,,0) = \overline{u}.
\end{cases}
\tag{5.1}
$$

Here, $p = p(x,t)$ is the scalar pressure, which can be determined from $u(x,t)$ via solution of the elliptic equation, $-\Delta p = \operatorname{div}(\operatorname{div}(u \otimes u))$.

In the following, we will focus on the periodic case with domain $D = \mathbb{T}^d$, and dimension $d \in \{2,3\}$. Physically meaningful solutions of (5.1) are required to satisfy an energy admissibility constraint of the form $\|u(t)\|_{L^2_x} \le \|\overline{u}\|_{L^2}$ for all $t \in [0,T]$, so that $u(t) \in L^2_x(\mathbb{T}^d;\mathbb{R}^d)$ is uniformly bounded in time. In particular, we consider solutions in the space $u \in L^\infty_t([0,T];L^2_x)$.

5.1.1. *Spectral hyper-viscosity scheme.* Popular numerical discretizations of the forward problem for the incompressible Euler equations on periodic domains are spectral methods [3, 2, 16, 11]. In this section, we will review the spectral (hyper-)viscosity methods, originally proposed by Tadmor [30] in the context of scalar conservation laws, and further investigated in [17, 18] in the context of the incompressible Euler equations.

We write $u^\Delta(x,t) = \sum_{|k|_\infty \le N} \widehat{u}_k^\Delta(t) e^{ik\cdot x}$, where now and in the following we shall consistently denote $\Delta = 1/N$, and we denote $|k|_\infty := \max_{i=1,\dots,d} |k_i|$. We consider the following spectral viscosity approximation [30, 31] of the incompressible Euler equations,

$$\begin{cases} \partial_t u^\Delta + \mathcal{P}_N(u^\Delta \cdot \nabla u^\Delta) + \nabla p^\Delta = -\epsilon_N |\nabla|^{2\sigma}(Q_N * u^\Delta), \\ \qquad\qquad\qquad\qquad \mathrm{div}(u^\Delta) = 0, \\ \qquad\qquad\qquad\qquad u^\Delta|_{t=0} = \mathcal{P}_N \overline{u}. \end{cases} \tag{5.2}$$

Here $\mathcal{P}_N$ is the spatial Fourier projection operator, mapping an arbitrary function $f(x,t)$ onto the first N Fourier modes: $\mathcal{P}_N f(x,t) = \sum_{|k|_\infty \le N} \widehat{f}_k(t) e^{ik\cdot x}$. Q_N is a Fourier multiplier of the form

$$Q_N(x) = \sum_{m_N < |k| \le N} \widehat{Q}_k e^{ik\cdot x}, \tag{5.3}$$

and we assume $0 \le \widehat{Q}_k \le 1$.

The underlying idea of the SV method is that the numerical dissipation which is required to stabilize the method, is only applied on the upper part of the spectrum, i.e. for $|k| > m_N$, thus preserving the formal spectral accuracy of the method. The hyperviscosity parameter $\sigma \ge 1$ can be chosen larger to enforce more numerical dissipation on the high Fourier modes, thus enabling a larger part of the Fourier spectrum to remain free of numerical diffusion.

The Fourier multiplier Q_N is defined via its Fourier coefficients $\widehat{Q}_k$, which, for an additional parameter $\theta > 0$, are subject to the constraints:

$$\widehat{Q}_k = 0, \quad \text{for } |k| \le m_N, \quad 1 - \left(\frac{m_N}{|k|}\right)^{(2\sigma-1)/\theta} \le \widehat{Q}_k \le 1. \tag{5.4}$$

The parameters m_N, ϵ_N, θ in [31] are chosen such that

$$m_N \sim N^\theta, \quad \epsilon_N \sim \frac{1}{N^{2\sigma-1}}, \quad 0 \le \theta < \frac{2\sigma-1}{2\sigma}. \tag{5.5}$$

5.1.2. *A priori estimates and consistency for the SV scheme.* Multiplying the evolution equation (5.2) by u^Δ and integrating by parts, we obtain the following energy balance,

$$\|u^\Delta(t)\|_{L^2_x}^2 + 2\epsilon_N \sum_{|k|_\infty \le N} \int_0^t \widehat{Q}_k |k|^{2\sigma} |\widehat{u}_k^\Delta(\tau)|^2 \, d\tau \le \|\overline{u}\|_{L^2_x}^2.$$

In particular, for any admissible choice of the parameters of the SV scheme, we obtain the a priori energy bound

$$\|u^\Delta(t)\|_{L^2_x} \le \|\overline{u}\|_{L^2_x}, \quad \forall\, t \in [0,T]. \tag{5.6}$$

We also recall [17, Lemma 3.2] that the SV scheme is consistent with the incompressible Euler equations, in the sense that for any initial data $\overline{u} \in L^2_x$, the sequence u^Δ converges (up to a subsequence) in the sense of Young measures to an energy admissible measure-valued solution [17], as $\Delta \to 0$. In fact, we have the following simple Lemma:

***Lemma* 5.1.** The approximate solution operator $S_t^\Delta : L_x^2 \to L_x^2$ obtained from the SV scheme (5.2) at grid scale $\Delta = 1/N$ satisfies assumption 4.3.

Proof. Energy admissibility has already been derived preceding (5.6). The simple argument to show temporal Lipschitz continuity with values in a sufficiently negative Sobolev space H_x^{-L} has e.g. been provided in [17, Remark 3.3]. □

It has been shown [1] that if there exists a strong solution $u \in C([0,T]; L_x^2)$ for given initial data $\overline{u}$, such that

$$\int_0^T \|\nabla u(t) + \nabla u(t)^T\|_{L_x^\infty} \, dt < \infty, \tag{5.7}$$

then this strong solution u is unique in the class of energy admissible measure-valued solutions. As a consequence of this weak-strong uniqueness result, we conclude that in fact $u^\Delta \to u$ converges e.g. in $L_t^2([0,T]; L_x^2)$ (in fact, $L_t^p([0,T]; L_x^2)$ for all $p < \infty$). We collect this observation in the following proposition.

***Proposition* 5.2.** Let $\overline{u} \in L_x^2$ be given initial data for the incompressible Euler equations. If there exists a unique strong solution $u = S_t(\overline{u})$ of (5.1) with initial data $\overline{u}$ and such that (5.7) holds, then the approximate solution $u^\Delta = S_t^\Delta(\overline{u})$ computed by the SV scheme converges to $S_t(\overline{u})$. More precisely, we have

$$\int_0^T \|S_t^\Delta(\overline{u}) - S_t(\overline{u})\|_{L_x^2}^2 \, dt \to 0, \qquad \text{as } \Delta \to 0.$$

***Remark* 5.3.** Paired with classical short-term existence and smoothness results for the incompressible Euler equations [26], the previous proposition provides in particular a general (short-time) convergence result for the SV scheme for smooth initial data.

In the two-dimensional case, $d = 2$, the vorticity is known to be advected by the flow, implying that, at least formally, L^p-norms of $\omega = \mathrm{curl}(u)$ can be controlled. The SV scheme ensures L^p-control on the vorticity $\omega^\Delta = \mathrm{curl}(u^\Delta)$ for $p = 2$: In the *two-dimensional case*, we have the following enstrophy bound (see e.g. [18, Proposition 4.2])

$$\|\omega^\Delta(t)\|_{L_x^2} \le \|\overline{\omega}\|_{L_x^2}, \quad \forall\, t \in [0,T], \tag{5.8}$$

where $\overline{\omega} = \mathrm{curl}(\overline{u})$ is the vorticity of the initial data. If the initial vorticity $\overline{\omega} \in L_x^\infty$ is bounded, it has been shown by Yudovich [34], that there exists a solution $u = S_t(\overline{u})$ of the incompressible Euler equations with uniformly bounded vorticity $\|\mathrm{curl}(u)\|_{L_x^\infty} \le \|\overline{\omega}\|_{L_x^\infty}$. Furthermore, this solution $S_t(u)$ is unique in the class of solution with bounded vorticity [34]. Later, it has been pointed out by Liu and Xin [25], that the proof of uniqueness in [34, 35] actually extends to provide a weak-strong uniqueness result in a wider class: If v is another weak solution of the incompressible Euler equations with vorticity bound $\|\mathrm{curl}(v(t))\|_{L_x^p} \le C$, for any $p > 4/3$, then $v \equiv u$ is the unique Yudovich solution.[2] As a consequence of this weak-strong uniqueness result and the enstrophy bound (5.8), we obtain

[2] In fact, the Yudovich-class weak-strong uniqueness result of [25] can be slightly extended to prove that Yudovich solutions are unique in the class of weak solutions with a L_x^p vorticity bound for *any* $p > 1$. Since this extension is not necessary in the present case, we do not provide a detailed proof here.

***Proposition* 5.4.** If $\overline{u}$ is initial data for the two-dimensional incompressible Euler equations with bounded vorticity, $\|\overline{\omega}\|_{L^\infty_x} < \infty$, then the approximate solutions $u^\Delta = S^\Delta_t(\overline{u})$ converge strongly in $L^2_t([0,T];L^2_x)$ to the unique Yudovich solution $S_t(\overline{u})$, i.e.

$$\int_0^T \|S^\Delta_t(\overline{u}) - S_t(\overline{u})\|_{L^2_x}\,dt \to 0, \quad \text{as } \Delta \to 0.$$

A second consequence of the enstrophy bound (5.8) is a uniform estimate on the structure function:

***Proposition* 5.5.** If $\overline{u} \in L^2_x$ is initial data for the two-dimensional incompressible Euler equations with bounded enstrophy, $\|\overline{\omega}\|_{L^2_x} < \infty$ with $\overline{\omega} = \mathrm{curl}(\overline{u})$, then there exists a constant $C > 0$, such that for any $\Delta > 0$, the structure function obeys the bound

$$\mathscr{S}_2(S^\Delta_t(\overline{u}); r) \le Cr\|\overline{\omega}\|_{L^2_x}, \qquad \forall\, t \in [0,T],\ r \ge 0.$$

Proof. By definition, we have for any $u \in H^1_x$:

$$\mathscr{S}_2(u;r)^2 = \fint_{B_r(0)} \int_D |u(x+h) - u(x)|^2\,dx\,dh \quad = \fint_{B_r(0)} \|u(\,\cdot\,+h) - u(\,\cdot\,)\|^2_{L^2_x}\,dh.$$

The estimate $\|u(\,\cdot\,+h) - u(\,\cdot\,)\|_{L^2_x} \le C\|\nabla u\|_{L^2_x}|h|$ is classical. Furthermore, it follows from the incompressibility of u that $\|\nabla u\|_{L^2_x} = \|\mathrm{curl}(u)\|_{L^2_x}$. Hence,

$$\begin{aligned}
\mathscr{S}_2(u;r)^2 &= \fint_{B_r(0)} \|u(\,\cdot\,+h) - u(\,\cdot\,)\|^2_{L^2_x}\,dh \\
&\le \fint_{B_r(0)} C\|\mathrm{curl}(u)\|^2_{L^2_x}|h|^2\,dh \\
&\le C\|\mathrm{curl}(u)\|^2_{L^2_x} r^2.
\end{aligned}$$

Setting $u = S^\Delta_t(\overline{u})$, we thus find

$$\mathscr{S}_2(S^\Delta_t(\overline{u}); r) \le Cr\|\mathrm{curl}(S^\Delta_t(u))\|_{L^2_x} \le Cr\|\overline{\omega}\|_{L^2_x},$$

where the last inequality follows from (5.8). □

5.1.3. *The well-posed case.* Combining the general results for the Bayesian inverse and filtering problems in sections 3 and 4, and the above convergence results for the spectral viscosity scheme, we can now prove:

***Theorem* 5.6.** If $\mu_{\mathrm{prior}} \in \mathcal{P}(L^2_x)$ is a prior, and if there exists $M > 0$, $s > d/2 + 2$, such that $\mu_{\mathrm{prior}}(B^s_M) = 1$, where

$$B^s_M := \left\{\overline{u} \in L^2_x \cap H^s_x \,\middle|\, \|\overline{u}\|_{H^s_x} \le M\right\} \subset L^2_x,$$

then there exists a time interval $[0,T]$ with $T = T(M,s) > 0$, such that the BIP and filtering problems for the incompressible Euler equations are well-posed on $[0,T]$: Given measurements in the time-interval $[0,T]$, there exists a unique solution μ^y for the BIP and $\nu^{\boldsymbol{y}}_t$ for the filtering problem. The posteriors μ^y and $\nu^{\boldsymbol{y}}_t$ are W_1-stable with respect to measurements, in the sense of (3.23) and (4.21), respectively. Furthermore, the approximations $\mu^{\Delta,y}$ and $\nu^{\Delta,\boldsymbol{y}}_t$ obtained by the numerical discretization with the SV scheme converge to this solution as $\Delta \to 0$, in the 1-Wasserstein norm W_1.

Proof. We first observe that there exists a $T > 0$, such that the initial value problem for the incompressible Euler equations is well-posed on $[0, T]$, for all initial data $u \in B^s_M$. In fact, by Sobolev embedding, there exists $T > 0$ such that the quantity (5.7) is finite. In particular, by Proposition 5.2, $S^\Delta_t(u) \to S_t(u)$ converges to the unique solution for all initial data $u \in B^s_M$ and $t \in [0, T]$. From this point-wise convergence and the following uniform bound on the measurements

$$|\mathcal{L}^\Delta(u)|_\Gamma = |\mathcal{G}(S^\Delta_t(u))|_\Gamma \le C\|S^\Delta_t(u)\|^2_{L^2_x} \le C\|u\|^2_{L^2_x} \le CM^2,$$

for all $u \in B^s_M$, it now follows from dominated convergence that

$$\|\mathcal{L}^\Delta(u) - \mathcal{L}(u)\|_{L^2(\mu_{\text{prior}})} \to 0, \quad (\Delta \to 0).$$

In particular, by the consistency Theorem 3.14 for the BIP, it follows that the approximate posterior of the BIP $\mu^{\Delta,y} \to \mu^y$ converges wrt. to the 1-Wasserstein metric to the unique solution in the limit $\Delta \to 0$. Furthermore, by Theorem 3.9, the posteriors $\mu^{\Delta,y}$ are uniformly stable with respect to the measurements y (cp. equation (3.23)).

We next discuss the filtering problem. By Lemma 5.1, the SV scheme satisfies Assumption 4.3. Theorem 4.13 implies that the posteriors $\nu^{\boldsymbol{y}}_t$ are uniformly stable with respect to the measurements $\boldsymbol{y}$. Due to the pointwise convergence $S^\Delta_t(u) \to S_t(u)$ for all $u \in B^s_M$ and the uniform bound

$$\|S^\Delta_t(u) - S_t(u)\|_{L^2_x} \le 2M,$$

Lebesgue's dominated convergence theorem implies that

$$\lim_{\Delta \to 0} \int_0^T \|S^\Delta_t(u) - S_t(u)\|_{L^2(\mu_{\text{prior}})}\, dt = 0.$$

The consistency theorem 4.19 therefore shows that $\nu^{\Delta,\boldsymbol{y}}_t \to \nu^{\boldsymbol{y}}_t$ in $L^1(\mathcal{P})$. □

In the two-dimensional case, the above result can be improved:

***Theorem* 5.7.** If $\mu_{\text{prior}} \in \mathcal{P}(L^2_x)$ is a prior for the *two-dimensional* incompressible Euler equations, such that

$$\int \|\mathrm{curl}(u)\|^2_{L^\infty_x}\, d\mu_{\text{prior}}(u) < \infty,$$

then the BIP and filtering problems for the incompressible Euler equations are well-posed and the numerical solutions converge as in the conclusion of Theorem 5.6 on $[0, T]$, for any $T > 0$.

Proof. The condition

$$\int \|\mathrm{curl}(u)\|^2_{L^\infty_x}\, d\mu_{\text{prior}}(u) < \infty,$$

implies that μ_{prior} is concentrated on Yudovich initial data. The strong convergence $S^\Delta_t(u) \to S_t(u)$ to the unique Yudovich solution for such initial data u has been shown in Proposition 5.4. The remainder of the proof follows verbatim as in the proof of Theorem 5.6. □

5.1.4. *The ill-posed case.* Beyond the short-time existence, uniqueness and stability results for the incompressible Euler equations with smooth initial data there are currently no general a priori well-posedness results for the forward problem in the three-dimensional case. In the two-dimensional case, existence results are known for initial data with vorticity $\overline{\omega} \in L^p$, $p \ge 1$, as well as for less regular initial data with a essential sign restriction, of the form $\overline{\omega} = \overline{\omega}_0 + \overline{\omega}_1$, such that $\overline{\omega}_0 \in \mathcal{BM}_+$, $\overline{\omega}_0 \ge 0$ a bounded Radon measure and $\overline{\omega}_1 \in L^1$ [6, 33]. Uniqueness remains unknown for such rough flows beyond the class considered by Yudovich, even if $\overline{\omega} \in L^p$, for $p < \infty$.

Thus, the forward problem may be ill-posed for general initial data $u \in L^2_x$ for the incompressible Euler equations, in both two and three dimensions. Despite this possible lack of stability and compactness for the forward problem, the general results of Section 3 imply that the Bayesian inverse problem is stable with respect to measurements and compact in the 1-Wasserstein norm for approximations obtained from the SV scheme.

***Theorem* 5.8.** If $\mu_{\mathrm{prior}} \in \mathcal{P}_1(L^2_x)$ is any prior for the incompressible Euler equations in either two or three dimensions, then the posteriors $\mu^{\Delta,y}$ of the BIP (3.3) for the incompressible Euler equations are uniformly stable in y, in the sense of (3.23), for any $\Delta > 0$. Furthermore, the posteriors $\mu^{\Delta,y}$ form a compact sequence in $\mathcal{P}_1$.

For the filtering problem, we have the following result:

***Theorem* 5.9.** If $\mu_{\mathrm{prior}} \in \mathcal{P}_1(L^2_x)$ is a prior for the incompressible Euler equations for $d = 2$ or $d = 3$, then the approximate solutions $\nu_t^{\Delta,\boldsymbol{y}}$ of the filtering problem computed by the SV scheme are uniformly stable with respect to the measurements $\boldsymbol{y}$, in the sense of (4.21), for any $\Delta > 0$. In addition, if either

(a) there exists a modulus of continuity such that

$$\mathscr{S}_2^T(S^{\Delta}_{t,\#}\mu_{\mathrm{prior}}; r) \leq \phi(r), \quad \forall\, \Delta > 0,\ r \geq 0,$$

or

(b) $d = 2$ and μ_{prior} satisfies

$$\int \|\mathrm{curl}(u)\|^2_{L^2_x} \, d\mu_{\mathrm{prior}}(u) < \infty,$$

then the posteriors $\nu_t^{\Delta,\boldsymbol{y}}$ form a compact sequence in $L^1_t(\mathcal{P})$.

***Remark* 5.10.** Numerical evidence that assumption (a) of Theorem 5.9 is verified for a large range of priors supported on rough initial data, at least in the two-dimensional case, has been presented in [19, 20].

***Remark* 5.11.** We emphasize that the proof of the uniform local Lipschitz-stability

$$d_T\left(\nu_t^{\Delta,\boldsymbol{y}}, \nu_t^{\Delta,\boldsymbol{y}'}\right) \leq C|\boldsymbol{y} - \boldsymbol{y}'|_\Gamma,$$

has been rigorously established from a priori estimates, and is not conditional on any assumptions on the structure functions. We believe this stability result to be of particular importance to practitioners in data assimilation.

5.2. **Incompressible Navier-Stokes.** We consider the incompressible Navier-Stokes equations:

$$\begin{cases} \partial_t u + \mathrm{div}\,(u \otimes u) + \nabla p = \nu \Delta u, \\ \mathrm{div}(u) = 0, \\ u(\,\cdot\,, 0) = \overline{u}, \end{cases} \tag{5.9}$$

with viscosity $\nu > 0$. For simplicity we shall again focus on the case of periodic boundary conditions. It is well-known that *in the two-dimensional case*, the Navier-Stokes are well-posed on L^2_x, for any fixed value of the viscosity $\nu > 0$ [24, Theorem 3.1]. In the three-dimensional case, it has been shown in the celebrated work of Leray [22] that energy admissible solutions exist, but their uniqueness remains an open question. Again, we consider the numerical approximation by spectral methods, analogous to (5.2), leading now to the discretized system

$$\begin{cases} \partial_t u^\Delta + \mathcal{P}_N(u^\Delta \cdot \nabla u^\Delta) + \nabla p^\Delta = \nu \Delta u^\Delta, \\ \operatorname{div}(u^\Delta) = 0, \\ u^\Delta|_{t=0} = \mathcal{P}_N \overline{u}. \end{cases} \tag{5.10}$$

Multiplying the first equation of (5.10) by u^Δ and integrating over space and the time interval $[0,t]$, we find the a priori energy estimate

$$\frac{1}{2}\|u^\Delta(t)\|_{L^2_x}^2 + \nu \int_0^t \|\nabla u^\Delta\|_{L^2_x}^2 \, dt = \frac{1}{2}\|u^\Delta(0)\|_{L^2_x}^2 \le \frac{1}{2}\|\overline{u}\|_{L^2_x}^2. \tag{5.11}$$

Furthermore, from (5.10), we have

$$\partial_t u^\Delta = -\mathcal{P}_N \operatorname{div}\left(u^\Delta \otimes u^\Delta\right) - \nabla p^\Delta + \nu \Delta u^\Delta.$$

Due to the uniform L^2-bound $\|u^\Delta\|_{L^2_x} \le \|\overline{u}\|_{L^2_x}$, it is not hard to see that the terms on the right hand side are uniformly bounded in H^{-L}_x for sufficiently large $L > 0$, with an upper bound depending only on $\|\overline{u}\|_{L^2_x}$. Thus, it follows that $u^\Delta(t) = S^\Delta_t(\overline{u}) \in \operatorname{Lip}([0,T]; H^{-L}_x)$ for some $L > 0$. In particular, we conclude that assumption 4.3 is satisfied for the spectral numerical approximants of the Navier-Stokes equations. Owing to the energy estimate (5.11), we also find

***Lemma* 5.12.** Let $\mu_{\mathrm{prior}} \in \mathcal{P}(L^2_x)$ be a prior for the incompressible Navier-Stokes equations (5.9), such that $\int_{L^2_x} \|u\|_{L^2_x}^2 \, d\mu_{\mathrm{prior}}(u) < \infty$. Let $S^\Delta_t : L^2_x \to L^2_x$ denote the approximate solution operator obtained from the spectral scheme (5.10). Then we have the following structure function estimate:

$$\mathscr{S}_2^T(S^\Delta_{t,\#}\mu_{\mathrm{prior}}; r) \le \frac{r}{\sqrt{2\nu}} \left(\int_{L^2_x} \|u\|_{L^2_x}^2 \, d\mu_{\mathrm{prior}}(u) \right)^{1/2}.$$

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

Proof. By definition, we have

$$\mathscr{S}_2^T(S^\Delta_{t,\#}\mu_{\mathrm{prior}}; r)^2 = \int_{L^2_x} \int_0^T \mathscr{S}_2(S^\Delta_t(u); r)^2 \, dt \, d\mu_{\mathrm{prior}}(u),$$

where, setting $u^\Delta = S^\Delta_t(u)$,

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

By (5.11), with $u^\Delta(t) = S^\Delta_t(u)$, we have

$$\int_0^T \|\nabla u^\Delta\|_{L^2_x}^2 \, dt \le \frac{1}{2\nu}\|u\|_{L^2_x}^2.$$

Thus, from the above estimates, we can now conclude that

$$\int_0^T \mathscr{S}_2(S^\Delta_t(u); r)^2 \, dt = \int_0^T \mathscr{S}_2(u^\Delta; r)^2 \, dt \le r^2 \int_0^T \|\nabla u^\Delta\|_{L^2_x}^2 \, dt \le \frac{r^2}{2\nu}\|u\|_{L^2_x}^2.$$

Integration against $d\mu_{\mathrm{prior}}$ yields

$$\mathscr{S}_2^T(S^{\Delta}_{t,\#}\mu_{\mathrm{prior}}; r)^2 \le \frac{r^2}{2\nu}\int_{L^2_x} \|u\|^2_{L^2_x}\, d\mu_{\mathrm{prior}}(u),$$

as claimed. □

As a result of these a priori estimates for the incompressible Navier-Stokes equations and the general compactness results for Bayesian inverse problems derived in the present work, we can now state:

***Theorem* 5.13.** If $\mu_{\mathrm{prior}} \in \mathcal{P}_1(L^2_x)$ is any prior for the incompressible Navier-Stokes equations with viscosity $\nu > 0$, then the posteriors $\mu^{\Delta,y}$ of the BIP (3.3) are uniformly stable in y, in the sense of (3.23), for any $\Delta > 0$. Furthermore, the posteriors $\mu^{\Delta,y}$ form a compact sequence in $\mathcal{P}_1$ and any limit point $\mu^{*,y}$ is absolutely continuous with respect to the prior μ_{prior}.

For the filtering problem, we obtain the following result:

***Theorem* 5.14.** If $\mu_{\mathrm{prior}} \in \mathcal{P}_1(L^2_x)$ is a prior for the incompressible Navier-Stokes equations (5.9) with fixed viscosity $\nu > 0$ (for $d = 2$ or $d = 3$), and if μ_{prior} has finite second moment

$$\int \|u\|^2_{L^2_x}\, d\mu_{\mathrm{prior}}(u) < \infty,$$

then the approximate solutions $\nu_t^{\Delta,\boldsymbol{y}}$ of the filtering problem for the Navier-Stokes equations computed by the spectral scheme (5.10) are uniformly stable with respect to the measurements $\boldsymbol{y}$, in the sense of (4.21) and the posteriors $\nu_t^{\Delta,\boldsymbol{y}}$ form a compact sequence in $L^1_t(\mathcal{P}_1)$.

5.3. **Hyperbolic systems of conservation laws.** Finally, we apply the results of this work to the numerical approximation of Bayesian inverse problems for hyperbolic systems of conservation laws. Again, we take as our domain $D = [0, 2\pi]^n$ with periodic boundary conditions. We recall that a system of conservation laws is a PDE of the form

$$\partial_t u^i + \sum_{j=1}^n \partial_j \left(F^{ij}(u)\right) = 0, \tag{5.12}$$

describing the temporal evolution of m conserved quantities $u^1, \dots, u^m : D \times [0, T] \to \mathbb{R}$, and $F^{ij} : \mathbb{R}^m \to \mathbb{R}$ are the fluxes. It is convenient to write the system (5.12) in the succinct form

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

where $u = (u^1, \dots, u^m) : D \times [0, T] \to \mathbb{R}^m$ and $F = (F^{ij}) : \mathbb{R}^m \to \mathbb{R}^{m\times n}$. The system of conservation laws (5.12) 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, include the compressible Euler equations of gas dynamics, the shallow water equations of oceanography, the Magneto-Hydro-Dynamics (MHD) equations of plasma physics, and the equations of nonlinear elastodynamics [4].

As is well-known, even in the special case of a scalar conservation law ($m = 1$), weak solutions to (5.12) are not necessarily unique. It is therefore necessary to augment hyperbolic conservation laws (5.12) 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 (5.12) is called an *entropy* weak solution, provided that, in addition to (5.12), also

$$\partial_t \eta(u) + \operatorname{div}(q(u)) \le 0, \tag{5.13}$$

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

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

Note that in this case, the entropy admissibility condition (5.13) implies, upon integration over D, an a priori bound of the form

$$\|u(t)\|_{L^2} \le C\|\overline{u}\|_{L^2},$$

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

5.3.1. *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 [9]. We briefly review the form of these numerical schemes, following [9, Section 4.1]. For a more complete review, we refer to e.g. [13, 23].

The computational spatial domain is discretized by a collection of cells

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

with corresponding cell midpoints

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

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

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

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

$$\frac{d}{dt} u^\Delta_{i^1,\dots,i^n}(t) + \sum_{k=1}^n \frac{1}{\Delta} \Big(F^{k,\Delta}(u^\Delta_{\mathbf{i}-(q-1)\mathbf{e}_k}(t), \dots, u^\Delta_{\mathbf{i}+q\mathbf{e}_k}(t)) \tag{5.14}$$

$$-F^{k,\Delta}(u^\Delta_{\mathbf{i}-q\mathbf{e}_k}(t), \dots, u^\Delta_{\mathbf{i}+(q-1)\mathbf{e}_k}(t)) \Big) = 0, \tag{5.15}$$

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

As in [9], we make the following assumptions on the discretization (5.14):

***Assumption* 5.15.** We assume that the finite volume scheme (5.14) is consistent in the sense that there exists a constant $C > 0$ such that for $k = 1, \dots, d$,

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

and the discretized solutions satisfy

(1) L^2 bound: There exists $C > 0$ such that

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

(2) weak BV bound: There exists $s \ge 2$, such that

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

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

***Remark* 5.16.** It is not difficult to see that, under Assumption 5.15, 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 standing boundedness assumption 4.3 of Section 4. Furthermore, as pointed out in [9, Remark 4.2], many examples of finite volume/difference schemes can be shown to satisfy Assumption 5.15. Examples include the so-called entropy stable Lax-Wendroff schemes and the TeCNO schemes of [10].

Based on the results of sections 3 and 4, we immediately obtain the following result:

***Theorem* 5.17.** Assume that the FV scheme (5.14) satisfies Assumption 5.15. If $\mu_{\text{prior}} \in \mathcal{P}_1(L^2_x)$ is any prior, then the posteriors $\mu^{\Delta,y}$ of the BIP (3.3) for the hyperbolic conservation law are uniformly stable in y, in the sense of (3.23), for any $\Delta > 0$. Furthermore, the posteriors $\mu^{\Delta,y}$ form a compact sequence in $\mathcal{P}_1$.

For the filtering problem, we have the following result:

***Theorem* 5.18.** Assume that the FV scheme (5.14) satisfies Assumption 5.15. 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 (4.21), for any $\Delta > 0$. In addition, if the prior $\mu_{\text{prior}} \in \mathcal{P}_1(L^2_x)$ satisfies Assumption 4.14, then the posteriors $\nu^{\Delta,\boldsymbol{y}}_t$ form a compact sequence in $L^1_t(\mathcal{P})$.

***Remark* 5.19.** The validity of Assumption 4.14 has been investigated for a diverse set of initial priors in [9]. The numerical presented in [9] strongly suggest that it is fulfilled for the cases considered there, and we conjecture that the structure functions (4.24) are uniformly bounded for a wide range of priors of practical relevance.

***Remark* 5.20.** 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, and the compactness results of [19] 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 $p \neq 2$.

6. Discussion

The Bayesian framework has been well-established as a suitable formulation of inverse problems arising in the context of PDEs (1.1). The well-posedness of the Bayesian inverse problem has been demonstrated, as long as the *forward problem* for the underlying PDE is well-posed i.e., the solution operator (1.2) exists, is unique and depends continuously and stably on the data.

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 not 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 [8, 9, 12].

Our main aim in this paper was to investigate Bayesian inverse problems for such PDEs with an *ill-posed forward problem.* In section 3, we considered a sequence of approximate posteriors $\mu^{\Delta,y}$, with Δ corresponding to a numerical approximation parameter (mesh size) and y being finite-dimensional (noisy) measurements. We were able to prove,

- (stability) uniform in $\Delta > 0$ stability of $\mu^{\Delta,y}$ with respect to the measurements y in the 1-Wasserstein norm,
- (compactness) compactness of the approximate solution sequence $\{\mu^{\Delta,y}\}_{\Delta>0}$ in the space of probability measures $\mathcal{P}_1(X)$ with respect to the Wasserstein metric,
- (consistency) convergence in $\mathcal{P}_1(X)$ to the canonical posterior $\mu^{*,y}$, provided that the observables converge in an average L^2-sense.

All of these results are obtained under only mild boundedness assumptions on the approximate observables and on the measurement noise (e.g. satisfied by Gaussian noise). The general compactness properties allow us to define a set of candidate solutions to the BIP, generated by the numerical scheme. 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 (cf. Remark 3.13).

Building upon these general considerations for the abstract BIP, a derivation of similar stabilty, compactness and consistency properties for the filtering problem has been given in Section 4. In this case, the approximate posterior measures $t \mapsto \mu_t^{\Delta,y}$ are time-dependent, and are updated at discrete times to incorporate information obtained from measurements. In contrast to the abstract BIP, the filtering problem as formulated in Section 4 involves a recursive process, alternating between evolving the current posterior to the next discrete time step, where it serves as a prior for the new measurements, and using the new measurements to obtain the next posterior. In a suitable space of time-parametrized probability measures, we show that a similar uniform stability result with respect to the measurements as for the abstract BIP also holds for this formulation of the filtering problem. Perhaps astonishingly, even though perturbations to the measurement $\boldsymbol{y}$ perturb $\mu_{t_i}^{\Delta,\boldsymbol{y}}$ at each time-step and the filtering problem involves a successive application of a push-forward $S^{\Delta}_{t,\#}\mu^{\Delta,\boldsymbol{y}}_{t_i}$ by the discretized solution operator S^{Δ}_t, our stability result holds under a mere boundedness assumption on S^{Δ}_t, and *does not require* any uniform continuity of the mapping $\mu \mapsto S^{\Delta}_{t,\#}\mu$. In practice, the boundedness assumption usually corresponds to a discrete energy or entropy inequality, which is satisfied by suitably designed numerical schemes. In addition to this general stability result, we prove compactness of the approximate solution sequence $\mu^{\Delta,\boldsymbol{y}}_t$ for the filtering problem, under the assumption of a uniform bound on the second-order structure function. The structure function measures two point-correlations in the flow, and is a very natural quantity in the study of turbulence. If the solution of the forward problem possesses unique solutions almost surely with respect to the prior, then we prove that the numerically obtained solutions of the filtering problem (obtained by a consistent numerical scheme) converge to the expected canonical solution of the filtering problem.

The applicability of the abstract results of sections 3 and 4 to the numerical approximation of Bayesian inverse problems encountered in practice is discussed in Section 5. We consider three representative model problems: the incompressible Euler equations (in 2d and 3d), the incompressible Navier-Stokes equations (in 3d) and a class of hyperbolic systems of conservation laws. For the incompressible Euler equations, we consider the numerical approximation by spectral schemes and verify the sufficient conditions for stability, compactness and consistency by a priori analysis for a class of priors in 2d. In 3d, the general stability and consistency properties continue to hold by the same a priori considerations; the compactness property holds under the additional assumption of a physically motivated bound on the structure functions. For the incompressible Navier-Stokes equations (in 3d), we prove the conditions for stability and compactness by a priori analysis, for numerical solutions obtained by spectral schemes. Finally, in the context of systems of hyperbolic conservation laws, we consider the entropy stable discretization by finite-volume schemes. Again, we prove stability for the BIP and filtering problems without any additional assumptions; compactness is obtained provided an average bound on the structure function holds. We point out that numerical evidence that the required bound on the structure function holds, has been demonstrated by numerical experiments for a number of initial data priors in [9, 19, 20], and is further motivated by physical considerations.

The well-posedness results in the context of Bayesian inversion 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 mild conditions on the numerical scheme, the approximate solutions of the BIP and data assimilation problems are stable with respect to perturbations of the measurements, *independently of the numerical resolution.* The general compactness results presented in this work will be of importance in determining suitable selection criteria to single out a "canonical" posterior amongst the set of candidate solutions. We proposed to do so in a forthcoming paper.

Appendix A. Mathematical complements

We recall the Arzela-Aszoli 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.

The following theorem relates the compactness of a family of measures $\mu_t^\Delta \in L^1_t(\mathcal{P})$, where $L^1_t(\mathcal{P}) = L^1_t([0,T];\mathcal{P}(L^2_x(\mathbb{T}^d)))$ is metrized by the distance $d_T(\mu_t,\nu_t)$,

$$d_T(\mu_t,\nu_t) := \int_0^T W_1(\mu_t,\nu_t)\,dt, \quad (W_1 = \text{1-Wasserstein distance}),$$

to the uniform decay of their structure functions

$$S_2^T(\mu_t^\Delta; r) := \left(\int_0^T \int_{L^2_x} \int_{\mathbb{T}^d} \fint_{B_r(0)} |u(x+h) - u(x)|^2 \,dh\,dx\,d\mu_t(u)\,dt\right)^{1/2}.$$

***Proposition* A.2.** Let $S^\Delta : [0,T] \times L^2_x \to L^2_x$, be a family of operators, such that $(t,u) \mapsto S^\Delta_t(u)$ is continuous for each $\Delta > 0$. Assume that there exists a constant $C > 0$, such that $\|S^\Delta_t(u)\|_{L^2_x} \le C\|u\|_{L^2_x}$ for all $t \in [0,T]$ and $\Delta > 0$. Let $\mu_0 \in \mathcal{P}(L^2_x)$ be a probability measure on $L^2_x(\mathbb{T}^d)$ with finite second moments, i.e. such that

$$\int_{L^2_x} \|u\|^2_{L^2_x} \,d\mu_0(u) < \infty,$$

and define a family of probability measures $\{\mu^\Delta_t\}_{\Delta>0}$ as the push-forward of μ_0 under S^Δ_t, i.e. $\mu^\Delta_t := S^\Delta_{t,\#}\mu_0 \in L^1_t(\mathcal{P})$. If $\{\mu^\Delta_t\}_{\Delta>0}$ is relatively compact in $L^1_t(\mathcal{P})$, then there exists a modulus of continuity $\phi : [0,1] \to [0,\infty)$, $r \mapsto \phi(r)$, such that

$$S_2^T(\mu^\Delta_t; r) \le \phi(r), \quad \forall\, \Delta > 0,\ \forall\, r \in [0,1].$$

Proof. To prove the claim, it suffices to show that $\limsup_{r\to 0} \sup_{\Delta>0} S_2^T(\mu^\Delta_t; r) = 0$, i.e. that

$$\limsup_{r\to 0} \sup_{\Delta>0} \left(\int_0^T \int_{L^2_x} \int_{\mathbb{T}^d} \fint_{B_r(0)} |u(x+h) - u(x)|^2 \,dh\,dx\,d\mu^\Delta_t(u)\,dt\right)^{1/2} = 0.$$

We first note that the quantity on the left is finite for any $\Delta > 0$ and $r \in [0,1]$, since we trivially have

$$S_2^T(\mu^\Delta_t; r) \le 2\left(\int_0^T \int_{L^2_x} \|u\|^2_{L^2_x} \,d\mu^\Delta_t(u)\,dt\right)^{1/2},$$

and by assumption on the second moment of μ_0 and the uniform boundedness of the operators $S^\Delta_t : L^2_x \to L^2_x$, there exists a constant $C > 0$, such that

$$\begin{aligned}\int_{L^2_x} \|u\|^2_{L^2_x} \,d\mu^\Delta_t(u) &= \int_{L^2_x} \|u\|^2_{L^2_x} \,d\left(S^\Delta_{t,\#}\mu_0\right)(u) = \int_{L^2_x} \|S^\Delta_t(u)\|^2_{L^2_x} \,d\mu_0(u) \\ &\le \int_{L^2_x} C^2 \|u\|^2_{L^2_x} \,d\mu_0(u) < \infty,\end{aligned}$$

is uniformly bounded for any $t \in [0,T]$.

To show that $S_2^T(\mu^\Delta_t; r)$ converges to 0, uniformly as $r \to 0$, we will use mollification. In the following, we denote by u_ϵ the ϵ-mollification of $u \in L^2_x$, $u_\epsilon(x) = (\rho_\epsilon * u)(x)$, where $\rho_\epsilon(x) := \epsilon^{-d}\rho(x/\epsilon)$, and $\rho \ge 0$ is a smooth function supported in a ball of radius 1, such that $\int_{B_1(0)} \rho(x)\,dx = 1$. We now note that

$$S_2^T(\mu_t^\Delta; r) \le \left(\int_0^T \int_{L_x^2} \int_{\mathbb{T}^d} \fint_{B_r(0)} |u_\epsilon(x+h) - u_\epsilon(x)|^2 \,dh\,dx\,d\mu_t(u)\,dt\right)^{1/2} + \sup_{\Delta>0} 2\left(\int_0^T \int_{L_x^2} \|u - u_\epsilon\|_{L_x^2}^2 \,d\mu_t^\Delta(u)\,dt\right)^{1/2}.$$

For any $u \in L_x^2$ and $\epsilon > 0$, we have

$$\int_{\mathbb{T}^d} \fint_{B_r(0)} |u_\epsilon(x+h) - u_\epsilon(x)|^2 \,dh\,dx \le \|\nabla u_\epsilon\|_{L_x^2}^2 r^2 \le C\|u\|_{L_x^2}^2 \left(\frac{r}{\epsilon}\right)^2.$$

In particular, this implies that

$$\begin{aligned}
\limsup_{r\to 0} \sup_{\Delta>0} &\int_0^T \int_{L_x^2} \int_{\mathbb{T}^d} \fint_{B_r(0)} |u_\epsilon(x+h) - u_\epsilon(x)|^2 \,dh\,dx\,d\mu_t^\Delta(u)\,dt \\
&\le \limsup_{r\to 0} CT\left(\int_{L_x^2} \|u\|_{L_x^2}^2 \,d\mu_0(u)\right)\left(\frac{r}{\epsilon}\right)^2 \\
&= 0.
\end{aligned}$$

Hence, we have

$$\limsup_{r\to 0} \sup_{\Delta>0} S_2^T(\mu_t^\Delta; r) \le 2\left(\sup_{\Delta>0} \int_0^T \int_{L_x^2} \|u - u_\epsilon\|_{L_x^2}^2 \,d\mu_t^\Delta(u)\,dt\right)^{1/2}, \tag{A.1}$$

for any $\epsilon > 0$. The statement of this proposition will thus follows from the following claim: If the family $\{\mu_t^\Delta\}_{\Delta>0}$ is compact in $L_t^1(\mathcal{P})$, then we have

$$\limsup_{\epsilon\to 0} \sup_{\Delta>0} \int_0^T \int_{L_x^2} \|u - u_\epsilon\|_{L_x^2}^2 \,d\mu_t^\Delta(u)\,dt = 0.$$

To see why, fix $M > 0$ and denote $B_M = \left\{u \in L_x^2 \,\middle|\, \|u\|_{L_x^2} \le M\right\}$. Then

$$\begin{aligned}
\int_0^T \int_{L_x^2} \|u - u_\epsilon\|_{L_x^2}^2 \,d\mu_t^\Delta(u)\,dt &= \int_0^T \int_{L_x^2 \cap B_M^c} \|u - u_\epsilon\|_{L_x^2}^2 \,d\mu_t^\Delta(u)\,dt \\
&\quad + \int_0^T \int_{L_x^2 \cap B_M} \|u - u_\epsilon\|_{L_x^2}^2 \,d\mu_t^\Delta(u)\,dt \\
&=: (I^\Delta) + (II^\Delta).
\end{aligned} \tag{A.2}$$

We can estimate

$$\begin{aligned}
(I^\Delta) &\le \int_0^T \int_{L_x^2 \cap B_M^c} \left\{2\|u\|_{L_x^2}^2 + 2\|u_\epsilon\|_{L_x^2}^2\right\} d\mu_t^\Delta(u)\,dt \\
&\le 4\int_0^T \int_{L_x^2 \cap B_M^c} \|u\|_{L_x^2}^2 \,d\mu_t^\Delta(u)\,dt \\
&= 4\int_0^T \int_{L_x^2 \cap [S_t^\Delta]^{-1}(B_M^c)} \|S_t^\Delta(u)\|_{L_x^2}^2 \,d\mu_0(u)\,dt
\end{aligned}$$

By our assumption on the boundedness of S_t^Δ, there exists a constant $C > 0$, such that $\|S_t^\Delta(u)\|_{L_x^2} \le C\|u\|_{L_x^2}$, for all $u \in L_x^2$, and uniformly for $\Delta > 0$. This implies that

$$\sup_{\Delta>0} \int_{L_x^2 \cap [S_t^\Delta]^{-1}(B_M^c)} \|S_t^\Delta(u)\|_{L_x^2}^2 \,d\mu_0(u) \le \int_{L_x^2 \cap B_{M/C}^c} C^2\|u\|_{L_x^2}^2 \,d\mu_0(u),$$

and hence

$$\sup_{\Delta>0} (I^\Delta) = \sup_{\Delta>0} \int_0^T \int_{L^2_x \cap B^c_M} \|u - u_\epsilon\|^2_{L^2_x} \, d\mu^\Delta_t(u)\, dt$$
$$\leq 4TC^2 \int_{L^2_x \cap B^c_{M/C}} \|u\|^2_{L^2_x} \, d\mu_0(u).$$

Let $\delta > 0$ be arbitrary. Since $\int_{L^2_x} \|u\|^2_{L^2_x} \, d\mu_0(u) < \infty$, we can fix $M > 0$ sufficiently large, such that

$$\sup_{\Delta>0} (I^\Delta) < \delta. \tag{A.3}$$

To estimate the second term, we note that by the assumed relative compactness of $\{\mu^\Delta_t\} \subset L^1_t(\mathcal{P})$ and for the same $\delta, M > 0$ as above, there exists a finite collection $\{\mu^{\Delta_k}_t\}_{k=1,\dots,N}$, such that for any $\Delta > 0$, there exists $k \in \{1, \dots, N\}$, such that $d_T(\mu^\Delta_t, \mu^{\Delta_k}_t) < \delta/4M$. It then follows that for any index $\Delta > 0$, we have

$$\begin{aligned}
(II^\Delta) &= \int_0^T \int_{L^2_x \cap B_M} \|u - u_\epsilon\|^2_{L^2_x} \, d\mu^\Delta_t(u)\, dt \\
&\leq \int_0^T \int_{L^2_x \cap B_M} 2M \|u - u_\epsilon\|_{L^2_x} \, d\mu^\Delta_t(u)\, dt \\
&= 2M \int_0^T \int_{L^2_x} \|u - u_\epsilon\|_{L^2_x} \left[d\mu^\Delta_t(u) - d\mu^{\Delta_k}_t(u) \right] dt \\
&\qquad + 2M \int_0^T \int_{L^2_x} \|u - u_\epsilon\|_{L^2_x} \, d\mu^{\Delta_k}_t(u)\, dt \\
&\leq 4M \int_0^T W_1(\mu^\Delta_t, \mu^{\Delta_k}_t)\, dt + 2M \int_0^T \int_{L^2_x} \|u - u_\epsilon\|_{L^2_x} \, d\mu^{\Delta_k}_t(u)\, dt \\
&< \delta + 2M \int_0^T \int_{L^2_x} \|u - u_\epsilon\|_{L^2_x} \, d\mu^{\Delta_k}_t(u)\, dt,
\end{aligned}$$

for any $\epsilon > 0$. In the second to last step, we have used the 2-Lipschitz continuity of $u \mapsto \|u - u_\epsilon\|_{L^2_x}$ and Kantorovich duality for W_1. In the last step, we used the fact that $\{\mu^{\Delta_k}_t\}_{k=1,\dots,N}$ is a $\delta/4M$-net for $\{\mu^\Delta_t\} \subset L^1_t(\mathcal{P})$. We further note that the integrand $u \mapsto \|u - u_\epsilon\|$ converges pointwise to 0 as $\epsilon \to 0$, and is uniformly bounded by the (integrable) function $(u \mapsto 2\|u\|_{L^2_x}) \in L^1(\mu^{\Delta_k}_t \otimes dt)$ for any $k = 1, \dots, N$. By the Lebesgue dominated convergence theorem, it thus follows that

$$\lim_{\epsilon \to 0} \int_0^T \int_{L^2_x} \|u - u_\epsilon\|_{L^2_x} \, d\mu^{\Delta_k}_t(u)\, dt = 0,$$

for any $k = 1, \dots, N$. Since the set $\{\mu^{\Delta_k}_t\}_{k=1,\dots,N}$ is finite, we conclude that

$$\sup_{\Delta>0} (II^\Delta) < \delta + \lim_{\epsilon \to 0} \max_{k=1,\dots,N} 2M \int_0^T \int_{L^2_x} \|u - u_\epsilon\|_{L^2_x} \, d\mu^{\Delta_k}_t(u)\, dt = \delta.$$

Together with the estimate (A.3), the definition of (I^Δ), (II^Δ) (A.2), and (A.1), we finally conclude that

$$\limsup_{r \to 0} \sup_{\Delta>0} S^T_2(\mu^\Delta_t; r) < \delta.$$

But $\delta > 0$ was arbitrary, so we must in fact have

$$\limsup_{r \to 0} \sup_{\Delta>0} S^T_2(\mu^\Delta_t; r) = 0,$$

as claimed. □

References

[1] Y. Brenier, C. De Lellis, and L. Székelyhidi, *Weak-strong uniqueness for measure-valued solutions*, Communications in mathematical physics, 305 (2011), p. 351.

[2] A. Chorin., *Numerical solution of the Navier-Stokes equations*, Math. Comput., 22 (1968), pp. 745–762.

[3] Y. M. H. D. Gottlieb and S. Orszag, *Theory and application of spectral methods*, in Spectral methods for PDEs, SIAM, 1984, pp. 1–54.

[4] C. M. Dafermos, C. M. Dafermos, C. M. Dafermos, G. Mathématicien, C. M. Dafermos, and G. Mathematician, *Hyperbolic conservation laws in continuum physics*, vol. 3, Springer, 2005.

[5] C. De Lellis and L. Székelyhidi, Jr., *The Euler equations as a differential inclusion*, Ann. of Math. (2), 170 (2009), pp. 1417–1436.

[6] J.-M. Delort, *Existence de Nappes de Toubillon en Dimension Deux*, J. Am. Math. Soc., 4 (1991), pp. 553–586.

[7] P. Dupuis and R. S. Ellis, *A weak convergence approach to the theory of large deviations*, vol. 902, John Wiley & Sons, 2011.

[8] U. S. Fjordholm, R. Käppeli, S. Mishra, and E. Tadmor, *Construction of approximate entropy measure-valued solutions for hyperbolic systems of conservation laws*, Foundations of Computational Mathematics, 17 (2017), pp. 763–827.

[9] U. S. Fjordholm, K. Lye, S. Mishra, and F. Weber, *Statistical solutions of hyperbolic systems of conservation laws: Numerical approximation*, Mathematical Models and Methods in Applied Sciences, 30 (2020), pp. 539–609.

[10] U. S. Fjordholm, S. Mishra, and E. Tadmor, *Arbitrarily high-order accurate entropy stable essentially nonoscillatory schemes for systems of conservation laws*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 544–573.

[11] S. Ghoshal., *An analysis of numerical errors in large eddy simulations of turbulence*, J. Comput. Phys., 125 (1996), pp. 187–206.

[12] J. Glimm, J. Grove, X. Li, W. Oh, and D. Sharp, *A critical analysis of rayleigh–taylor growth rates*, Journal of Computational Physics, 169 (2001), pp. 652 – 677.

[13] E. Godlewski and P.-A. Raviart, *Numerical approximation of hyperbolic systems of conservation laws*, vol. 118, Springer Science & Business Media, 2013.

[14] L. Herrmann, C. Schwab, and J. Zech, *Deep neural network expression of posterior expectations in bayesian PDE inversion*, Inverse Problems, 36 (2020), p. 125011.

[15] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, vol. 160, Springer Science & Business Media, 2006.

[16] G. S. Karamanos and G. E. Karniadakis, *A spectral vanishing viscosity method for large-eddy simulations*, J. Comput. Phys., 163 (2000), pp. 22–50.

[17] S. Lanthaler and S. Mishra, *Computation of measure-valued solutions for the incompressible Euler equations*, Mathematical Models and Methods in Applied Sciences, 25 (2015), pp. 2043–2088.

[18] ———, *On the convergence of the spectral viscosity method for the two-dimensional incompressible euler equations with rough initial data*, Foundations of Computational Mathematics, (2019), pp. 1–54.

[19] S. Lanthaler, S. Mishra, and C. Parés-Pulido, *Statistical solutions of the incompressible euler equations*, 2019.

[20] S. Lanthaler, S. Mishra, and C. Parés-Pulido, *On the conservation of energy in two-dimensional incompressible flows*, 2020.

[21] J. Latz, *On the Well-posedness of Bayesian Inverse Problems*, SIAM/ASA Journal on Uncertainty Quantification, 8 (2020), pp. 451–482.

[22] J. Leray et al., *Sur le mouvement d'un liquide visqueux emplissant l'espace*, Acta mathematica, 63 (1934), pp. 193–248.

[23] R. J. LeVeque and R. J. Leveque, *Numerical methods for conservation laws*, vol. 3, Springer, 1992.

[24] P.-L. Lions, *Mathematical topics in fluid mechanics. Vol. 1*, vol. 3 of Oxford Lecture Series in Mathematics and its Applications, The Clarendon Press, Oxford University Press, New York, 1996. Incompressible models, Oxford Science Publications.

[25] J.-G. Liu and Z. Xin, *Convergence of vortex methods for weak solutions to the 2-d euler equations with vortex sheet data*, Communications on Pure and Applied Mathematics, 48 (1995), pp. 611–628.

[26] A. J. Majda and A. L. Bertozzi, *Vorticity and Incompressible Flow*, Cambridge Texts in Applied Mathematics, Cambridge University Press, 2001.

[27] T. Schneider, S. Lan, A. Stuart, and J. Teixeira, *Earth systemmodeling* 2.0*: A blueprint for models that learn from observations and targeted high-resolution simulations*, Geophysical Research Letters, 44 (2017), pp. 12396–12417.

[28] B. Sprungk, *On the local lipschitz stability of bayesian inverse problems*, Inverse Problems, 36 (2020), p. 055015.

[29] A. M. Stuart, *Inverse problems: a bayesian perspective*, Acta numerica, 19 (2010), pp. 451–559.

[30] E. Tadmor, *Convergence of spectral methods for nonlinear conservation laws*, SIAM J. Numer. Anal., 26 (1989).

[31] ———, *Burgers' equation with vanishing hyper-viscosity*, Communications in Mathematical Sciences, 2 (2004), pp. 317–324.

[32] A. Tarantola, *Inverse problem theory and methods for model parameter estimation*, SIAM, 2005.

[33] I. Vecchi and S. Wu, *On l1-vorticity for 2-d incompressible flow*, manuscripta mathematica, 78 (1993), pp. 403–412.

[34] V. I. Yudovich, *Non-stationary flow of an ideal incompressible liquid*, USSR Computational Mathematics and Mathematical Physics, 3 (1963), pp. 1407–1456.

[35] V. I. Yudovich, *Uniqueness theorem for the basic nonstationary problem in the dynamics of an ideal incompressible fluid*, Mathematical Research Letters, 2 (1995), pp. 27–38.