Well-posedness of Bayesian inverse problems for hyperbolic conservation laws

S. Mishra and D. Ochsner and A. M. Ruf and F. Weber

Research Report No. 2021-24
July 2021

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

Funding ERC: 770880 COMANFLO

Well-posedness of Bayesian Inverse Problems for Hyperbolic Conservation Laws

Siddhartha Mishra*, David Ochsner†, Adrian M. Ruf*, Franziska Weber‡

July 20, 2021

Abstract

We study the well-posedness of the Bayesian inverse problem for scalar hyperbolic conservation laws where the statistical information about inputs such as the initial datum and (possibly discontinuous) flux function are inferred from noisy measurements. In particular, the Lipschitz continuity of the measurement to posterior map as well as the stability of the posterior to approximations, are established with respect to the Wasserstein distance. Numerical experiments are presented to illustrate the derived estimates.

Key words. Inverse problem, Bayesian, Wasserstein distance, conservation laws

AMS subject classification. 65M32, 65C50, 35L64

1 Introduction

Hyperbolic systems of conservation laws are a large class of nonlinear PDEs which model a wide variety of phenomena in the sciences and engineering. The generic form of these PDEs is given by [8],

$$\begin{aligned} w_t + \nabla_x \cdot f(w) &= 0, \quad (x,t) \in \mathbb{R}^d \times (0,T), \\ w(x,0) &= \bar{w}(x), \quad x \in \mathbb{R}^d. \end{aligned} \tag{1.1}$$

Here, the solution field $w\colon \mathbb{R}^d \times [0,T] \to \mathbb{R}^m$ is the vector of conserved variables, $f\colon \mathbb{R}^m \to \mathbb{R}^m$ is the so-called *flux function* and $\bar{w}$ is the initial datum.

Prototypical examples of systems of conservation laws include the compressible Euler equations of fluid dynamics, the shallow-water equations of oceanography, the MHD equations of plasma physics and the equations of nonlinear elasticity. The simplest examples are the so-called *scalar conservation laws* i.e., (1.1) with $m = 1$, with the well-known Burgers' equation being a prototype.

It is well-known that solutions of even scalar conservation laws develop discontinuities, such as *shock waves*, for smooth initial data. Thus, the solutions of hyperbolic conservation laws are sought in the sense of distributions. However, these *weak solutions* are supplemented with additional admissibility criteria or *entropy conditions* to recover uniqueness [8].

The most studied aspect of PDEs such as hyperbolic conservation laws is the so-called *forward problem* i.e., given the inputs (initial datum and flux function) $u = (\bar{w}, f)$, find the *entropy solution*

*Seminar for Applied Mathematics (SAM), ETH Zürich, Rämistrasse 101, Zürich, Switzerland (`siddhartha.mishra@sam.math.ethz.ch` and `adrian.ruf@sam.math.ethz.ch`)

†Department of Mathematics, ETH Zürich, Rämistrasse 101, Zürich, Switzerland (`david.ochsner@math.ethz.ch`)

‡Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, USA (`franzisw@andrew.cmu.edu`)

The research of Siddhartha Mishra and Adrian M. Ruf is partially supported by the European Research Council Consolidator grant ERC-COG 770880 COMANFLO. Franziska Weber is partially supported by NSF DMS 1912854 and NSF OIA-DMR 2021019.

w of (1.1). Often, one is not necessarily interested in the whole solution field w of (1.1), but rather in *observables* or *quantities of interest* of the solution. Hence, the forward problem reduces to an evaluation of the mapping $\mathcal{G}$,

$$\mathcal{G}\colon X \to Y,\, u \mapsto y = \mathcal{G}(u),$$

which maps inputs $u \in X$ into observables $\mathcal{G}(u) \in Y$ of the solution, with X, Y being suitable Banach spaces.

However in practice, the inputs u (which correspond to the initial datum and flux function in the context of hyperbolic conservation laws (1.1)) may not be known exactly. Rather, one has to *infer* them from *measurements* of the observables. Hence, one is often interested in the so-called *Inverse problem*, which amounts to finding information about the inputs u, given *noisy measurements* of the form;

$$y = \mathcal{G}(u) + \eta,$$

with η being a random variable encoding measurement noise.

It is well-known that the *deterministic* version of the inverse problem may be *ill-posed* [22]. Although *regularization* procedures have been widely developed in the last few decades to address this ill-posedness of the deterministic inverse problem, it is a widely held view that *statistical approaches* might be better suited in this context. A very popular statistical approach [22] models the prior knowledge about the inputs u in terms of a *prior probability measure* $\mu_0 \in \mathrm{Prob}(X)$. Then the famous *Bayes' theorem* can be used to update our knowledge of the inputs u (and consequently the solution w) in terms of a *posterior measure* $\mu^y \in \mathrm{Prob}(X)$, conditioned on the noisy measurements $y \in Y$. The posterior measure is given by the following expression of its Radon-Nikodym derivative,

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z(y)} \exp(-\Phi(u; y)), \quad Z(y) = \int_X \exp(-\Phi(u; y))\, \mathrm{d}\mu_0(u) \tag{1.2}$$

Here, Φ is the *log-likelihood* with respect to the measurements y.

It is to be noted that the Bayesian formulation encodes a regularized version of the underlying deterministic problem as the latter is a *maximum a posteriori* (MAP) estimator of the former, with a suitable choice of the prior [22].

The well-posedness of the Bayesian inverse problem refers to the rigorous demonstration of *existence and uniqueness* of the posterior measure μ^y, its continuous dependence and stability with respect to perturbations of the measurements y. Moreover, in practice, one *approximates* the posterior computationally, for instance, by sampling from it with a Metropolis–Hastings-type Markov chain Monte Carlo (MCMC) algorithm. This in turn requires one to evaluate the likelihood in terms of *numerical approximations* $\mathcal{G}^\Delta$ of the forward operator $\mathcal{G}$. Here, Δ is a numerical regularization parameter such as the mesh size or the time step. The accuracy of the resulting *approximate posterior* $\mu^{y,\Delta}$ is also of great interest.

The well-posedness of the Bayesian inverse problems has been studied extensively in recent years and is nicely summarized in [22]. It has been clearly established that the Bayesian inverse problem is well-posed as long as the forward map $\mathcal{G}$ is *Lipschitz continuous*, with respect to suitable topologies. Even weaker assumptions on the forward map have been investigated recently in [14, 21].

Furthermore, these abstract assumptions on well-posedness have been verified and illustrated for a variety of elliptic, parabolic and linear hyperbolic PDEs, see [22] and references therein. The application of this theory to nonlinear hyperbolic PDEs, such as hyperbolic conservation laws (1.1) is currently not available, except in [10] where the authors study an example of a scalar conservation law with uncertain flux.

Given this context, our main goal in this paper is to study and establish well-posedness of the Bayesian inverse problem for hyperbolic conservation laws. We will focus on the scalar case ($m = 1$ in (1.1)) as no rigorous well-posedness results are available for the forward problem for systems of conservation laws, particularly in several space dimensions.

To this end, we will also study the Lipschitz continuity of the posterior measure with respect to measurements in the *Wasserstein distance* on probability measures. We observe that the standard framework for Bayesian inverse problems [22] uses the *Hellinger distance* to investigate stability with respect to perturbations. However, the Wasserstein distance offers some advantages over the Hellinger distance. To illustrate this, consider two measures which are absolutely continuous with respect to a Gaussian reference measure (e.g., the prior measure μ_0) and which are a distance ε apart in the Wasserstein distance. Then the difference between the means of the two measures is bounded by ε (cf. Remark 2.6). In contrast, if the two measures are a distance ε apart in the Hellinger metric then the difference in the means is only bounded by $C\varepsilon$ (see [22, Lem. 6.37]) where the constant C depends on the second moments and, in particular, can be arbitrarily large. Thus by bounding the Wasserstein distance, we can more effectively control the change in the posterior, caused either by perturbations of the measurement or by replacing the underlying forward map with a (numerical) approximation.

We apply these abstract stability results to establish the well-posedness of the Bayesian inverse problem, for inferring initial data as well as flux functions of scalar conservation laws, from measurements. Moreover, we extend the results to cover the Bayesian inverse problem for a conservation law, corresponding to a flux function that can vary discontinuously in the space variable. Thus, we establish the first rigorous well-posedness results for Bayesian inverse problems for these nonlinear hyperbolic PDEs.

The remainder of this paper is organized as follows: In Section 2 we provide the general well-posedness theory in the spirit of [22], but employing the Wasserstein distance instead of the Hellinger distance. Section 3 contains general approximation results for posterior measures given by (1.2). We study inverse problems for multi-dimensional scalar conservation laws and one-dimensional scalar conservation laws with discontinuous flux in Section 4. Lastly, in Section 5 we present a series of numerical experiments illustrating, in particular, the convergence of the approximated posterior distribution under refinement of the finite-dimensional approximation.

2 Well-posedness of general Bayesian inverse problems in the Wasserstein distance

The probability measure of interest is defined through a density with respect to a prior reference measure μ_0 which, by shift of origin, we take to have mean zero. Further, we assume that this prior measure is Gaussian with covariance operator $\mathcal{C}$. We write $\mu_0 = \mathcal{N}(0, \mathcal{C})$.

Assumption 2.1. *For some separable Banach space X with $\mu_0(X) = 1$, the function $\Phi\colon X \times Y \to \mathbb{R}$ satisfies the following:*

(i) *for every $\varepsilon > 0$ and $r > 0$ there is $M = M(\varepsilon, r) \in \mathbb{R}$ such that for all $u \in X$ and $y \in Y$ with $\|y\|_Y < r$*

$$\Phi(u; y) \geq M - \varepsilon \|u\|_X^2 ;$$

(ii) *for every $r > 0$ there is a $K = K(r) > 0$ such that for all $u \in X$ and $y \in Y$ with $\|u\|_X, \|y\|_Y < r$*

$$\Phi(u; y) \leq K;$$

(iii) *for every $r > 0$ there exists $L = L(r) > 0$ such that for all $u, u' \in X$ and $y \in Y$ with $\|u\|_X, \|u'\|_X, \|y\|_Y < r$*

$$|\Phi(u; y) - \Phi(u'; y)| \leq L \|u - u'\|_X .$$

(iv) *for all $\varepsilon > 0$ and $r > 0$ there is $C = C(\varepsilon, r) \in \mathbb{R}$ such that for all $y, y' \in Y$ with $\|y\|, \|y'\| < r$ and for all $u \in X$*

$$|\Phi(u; y) - \Phi(u; y')| \leq \exp\left(\varepsilon \|u\|_X^2 + C\right) \|y - y'\|_Y .$$

Note that Assumption 2.1 (1) and (2) will lead to bounds on the normalization constant Z from above and below. Assumption 2.1 (3) and (4) are Lipschitz conditions in u and y respectively.

For Bayesian inverse problems in which a finite number of observations are made and the observation error η is mean zero Gaussian with covariance matrix Γ, the potential Φ has the form

$$\Phi(u;y) = \frac{1}{2}\left|y - \mathcal{G}(u)\right|_\Gamma^2, \tag{2.1}$$

where $y \in \mathbb{R}^m$ is the data, $\mathcal{G}\colon X \to \mathbb{R}^m$ is the observation operator, and $|\cdot|_\Gamma$ is a covariance weighted norm on $\mathbb{R}^m$. In this case, we can translate Assumption 2.1 in terms of $\mathcal{G}$.

Assumption 2.2. *For some separable Banach space X with $\mu_0(X) = 1$, the function $\mathcal{G}\colon X \to \mathbb{R}^m$ satisfies the following:*

(i) *for every $\varepsilon > 0$ there is $M = M(\varepsilon) \in \mathbb{R}$ such that for all $u \in X$*

$$\left|\mathcal{G}(u)\right|_\Gamma \leq \exp\left(\varepsilon \left\|u\right\|_X^2 + M\right);$$

(ii) *for every $r > 0$ there is a $K = K(r) > 0$ such that for all $u, u' \in X$ with $\left\|u\right\|_X, \left\|u'\right\|_X <$ r*

$$\left|\mathcal{G}(u) - \mathcal{G}(u')\right|_\Gamma \leq K \left\|u - u'\right\|_X.$$

Lemma 2.3 ([22, Lem. 2.8])**.** *Assume that $\mathcal{G}\colon X \to \mathbb{R}^m$ satisfies Assumption 2.2 and that μ_0 is a Gaussian measure with $\mu_0(X) = 1$. Then $\Phi\colon X \times \mathbb{R}^m \to \mathbb{R}$ given by* (2.1) *satisfies Assumption 2.1 with $(y, \|\cdot\|_Y) = (\mathbb{R}^m, |\cdot|_\Gamma)$. In particular, if $\mathcal{G}$ satisfies Assumption 2.2* (1) *then Φ given by* (2.1) *satisfies Assumption 2.1* (1),(2), *and* (4)*.*

Proof. Assumption 2.1 (1) is trivially satisfied since Φ is nonnegative. Let now $r > 0$, $u, u' \in X$ and $y, y' \in \mathbb{R}^m$ all with norm less than r. Using the exponential bound on $\mathcal{G}$ (with $\varepsilon = 1$) we find

$$\begin{aligned}\Phi(u;y) &\leq |y|_\Gamma^2 + |\mathcal{G}(u)|_\Gamma^2 \\ &\leq r^2 + \exp\left(\left\|u\right\|_X^2 + M\right) \\ &\leq r^2 + \exp(r^2 + M)\end{aligned}$$

which gives Assumption 2.1 (2). Assumption 2.1 (3) follows from Assumption 2.2 (2) because

$$\begin{aligned}\left|\Phi(u;y) - \Phi(u',y)\right| &\leq \frac{1}{2}\left|2y - \mathcal{G}(u) - \mathcal{G}(u')\right|_\Gamma \left|\mathcal{G}(u) - \mathcal{G}(u')\right|_\Gamma \\ &\leq C\left(|y| + N\left\|u - u'\right\|_X\right) K \left\|u - u'\right\|_X \\ &\leq C\left(|y| + \left\|u\right\|_X + \left\|u'\right\|_X\right)\left\|u - u'\right\|_X \\ &\leq Cr\left\|u - u'\right\|_X.\end{aligned}$$

Lastly, for $\varepsilon > 0$, using the exponential bound on $\mathcal{G}$ we get

$$\begin{aligned}\left|\Phi(u;y) - \Phi(u,y')\right| &\leq \frac{1}{2}\left|y + y' - 2\mathcal{G}(u)\right|_\Gamma \left|y - y'\right|_\Gamma \\ &\leq C\left(|y| + |y'| + \exp\left(\varepsilon\left\|u\right\|_X^2 + M\right)\right)|y - y'| \\ &\leq \exp\left(\varepsilon\left\|u\right\|_X^2 + C(\varepsilon, r)\right)|y - y'|.\end{aligned}$$

$\square$

The following theorem is due to Stuart and shows that μ^y given by (1.2) is a well-defined probability measure provided Φ satisfies a Lipschitz condition in u.

Theorem 2.4 ([22, Thm. 4.1])**.** *Let Φ satisfy Assumption 2.1* (1), (2), *and* (3) *and assume that μ_0 is a Gaussian measure satisfying $\mu_0(X) = 1$. Then μ^y given by* (1.2) *is a well-defined probability measure on X.*

We have the following immediate corollary for Bayesian inverse problems with a finite number of observations and Φ of the form (2.1).

Corollary 2.5. *Assume that $\Phi\colon X \times Y \to \mathbb{R}$ is given by* (2.1) *and let $\mathcal{G}$ satisfy Assumption 2.2. Let further μ_0 be a Gaussian measure satisfying $\mu_0(X) = 1$. Then μ^y given by* (1.2) *is a well-defined probability measure on X.*

2.1 Well-posedness in the Wasserstein distance

The 1-Wasserstein distance between two probability measures μ and μ' with finite first moments

$$\int_X \|u\|_X \,\mathrm{d}\mu(u), \int_X \|u\|_X \,\mathrm{d}\mu'(u) < \infty$$

is defined as

$$W_1(\mu, \mu') = \sup_{\substack{\psi \in \mathcal{C}_b(X) \\ \|\psi\|_{\mathrm{Lip}} \leq 1}} \int_X \psi(u) \,\mathrm{d}(\mu - \mu')(u),$$

see [24]. Note that by the Fernique Theorem all moments of u in X are finite under a Gaussian measure (cf. Theorem A.1).

Remark 2.6. *The difference between the first moments of two probability measures μ and μ' is bounded by the Wasserstein distance between those measures:*

$$\left\| \int_X u \,\mathrm{d}\mu(u) - \int_X u \,\mathrm{d}\mu'(u) \right\|_X = \left\| \int_X u \,\mathrm{d}(\mu - \mu')(u) \right\|_X \leq \int_X \|u\|_X \,\mathrm{d}(\mu - \mu')(u) \leq W_1(\mu, \mu').$$

We show that the posterior measure μ^y is Lipschitz continuous with respect to the data y in the 1-Wasserstein distance. This constitutes a well-posedness result for the posterior measure. The result, and proof, is similar to that in [22] concerning well-posedness in the Hellinger distance.

Theorem 2.7 (Well-posedness in W_1)**.** *Let Φ satisfy Assumption 2.1* (1), (2), *and* (4)*. Assume also that μ_0 is a Gaussian measure satisfying $\mu_0(X) = 1$ and that for all $y \in Y$ the measure μ^y is absolutely continuous with respect to μ_0, $\mu^y \ll \mu_0$, with Randon–Nikodým derivative given by* (1.2)*. Then $y \mapsto \mu^y$ is Lipschitz continuous with respect to the 1-Wasserstein distance: if μ^y and $\mu^{y'}$ are two measures corresponding to data y and y' then for all $r > 0$ there exists $C = C(r) > 0$ such that, if $\|y\|_Y, \|y'\|_Y < r$, then*

$$W_1(\mu^y, \mu^{y'}) \leq C \|y - y'\|_Y .$$

Proof. In the following, we will write Z and Z' for $Z(y)$ and $Z(y')$ respectively (where Z is defined in (1.2)). From Assumption 2.1 (2) we get for any $r > 0$ and $\|y\|_Y < r$

$$|Z| \geq \int_{\{\|u\|_X < r\}} \exp(-L) \,\mathrm{d}\mu_0(u) \geq \exp(-L)\mu_0(\{\|u\|_X < r\}).$$

This lower bound is positive since μ_0 has full measure on X and is Gaussian so that all balls in X have positive probability. We have an analogous lower bound for $|Z'|$.

Using the estimate

$$|\exp(a) - \exp(b)| \leq (\exp(a) \vee \exp(b))|a - b|, \tag{2.2}$$

Assumption 2.1 (1), (4) and the fact that μ_0 is a Gaussian measure so that the Fernique Theorem A.1 applies, we find for $\|y\|_Y, \|y'\|_Y < r$

$$|Z - Z'| \leq \int_X (\exp(-\Phi(u; y)) \vee \exp(-\Phi(u; y'))) \,|\Phi(u; y) - \Phi(u; y')| \,\mathrm{d}\mu_0(u)$$

$$
\begin{aligned}
&\leq C \int_X \exp\left(\varepsilon \|u\|_X^2 - M\right) \exp\left(\varepsilon \|u\|_X^2 + C\right) \|y - y'\|_Y \,\mathrm{d}\mu_0(u) \\
&= C \int_X \exp\left(2\varepsilon \|u\|_X^2\right) \mathrm{d}\mu_0(u) \, \|y - y'\|_Y \\
&= C \|y - y'\|_Y .
\end{aligned}
$$

Now, let $\psi \in \mathcal{C}_b(X)$ with $\|\psi\|_{\mathrm{Lip}} \leq 1$. Since μ^y and $\mu^{y'}$ are probability measures, we have

$$
\begin{aligned}
\int_X \psi(u) \,\mathrm{d}(\mu^y - \mu^{y'})(u) &= \int_X (\psi(u) - \psi(0)) \,\mathrm{d}(\mu^y - \mu^{y'})(u) + \int_X \psi(0) \,\mathrm{d}(\mu^y - \mu^{y'})(u) \\
&= \int_X (\psi(u) - \psi(0)) \,\mathrm{d}(\mu^y - \mu^{y'})(u) \\
&= \int_X (\psi(u) - \psi(0)) \left(\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) - \frac{\mathrm{d}\mu^{y'}}{\mathrm{d}\mu_0}(u) \right) \mathrm{d}\mu_0(u) \\
&= \int_X (\psi(u) - \psi(0)) \left(Z^{-1} \exp(-\Phi(u;y)) - (Z')^{-1} \exp(-\Phi(u;y')) \right) \mathrm{d}\mu_0(u) \\
&= I_1 + I_2
\end{aligned}
$$

where

$$
\begin{aligned}
I_1 &= \int_X (\psi(u) - \psi(0)) Z^{-1} \left(\exp(-\Phi(u;y)) - \exp(-\Phi(u;y')) \right) \mathrm{d}\mu_0(u), \\
I_2 &= \int_X (\psi(u) - \psi(0)) \left(Z^{-1} - (Z')^{-1} \right) \exp(-\Phi(u;y')) \,\mathrm{d}\mu_0(u)
\end{aligned}
$$

Using again the estimate (2.2), the fact that $\|\psi\|_{\mathrm{Lip}} \leq 1$, Assumption 2.1 (1), and (4) we obtain

$$
\begin{aligned}
Z I_1 &\leq \int_X |\psi(u) - \psi(0)| \left| \exp(-\Phi(u;y)) - \exp(-\Phi(u;y')) \right| \mathrm{d}\mu_0(u) \\
&\leq \int_X \|u\|_X \left(\exp(-\Phi(u;y)) \vee \exp(-\Phi(u;y')) \right) |\Phi(u;y)) - \Phi(u;y')| \,\mathrm{d}\mu_0(u) \\
&\leq C \left(\int_X \|u\|_X \exp\left(2\varepsilon \|u\|_X^2\right) \mathrm{d}\mu_0(u) \right) \|y - y'\|_Y .
\end{aligned}
$$

Since all moments of u in X are finite under the Gaussian measure μ_0 by the Fernique Theorem, the integral in the last line can be bounded by using the Cauchy–Schwarz inequality and again the Fernique Theorem. Since Z is bounded from below by a positive constant, this gives a bound on I_1.

Using the fact that $\|\psi\|_{\mathrm{Lip}} \leq 1$, Assumption 2.1 (1), the above bound on $|Z - Z'|$ and again the fact that Z and Z' are bounded from below by a positive constant, we get

$$
\begin{aligned}
I_2 &\leq \int_X \|u\|_X \left| Z^{-1} - (Z')^{-1} \right| \exp(-\Phi(u;y')) \,\mathrm{d}\mu_0(u) \\
&\leq C \left| Z^{-1} - (Z')^{-1} \right| \int_X \|u\|_X \exp(\varepsilon \|u\|_X^2) \,\mathrm{d}\mu_0(u) \\
&\leq C \left(Z^{-2} \vee (Z')^{-2} \right) |Z - Z'| \\
&\leq C \|y - y'\|_Y .
\end{aligned}
$$

Here we used the same arguments as before to bound the integral $\int_X \|u\|_X \exp\left(\varepsilon \|u\|_X^2\right) \mathrm{d}\mu_0(u)$. Combining the bounds for I_1 and I_2 gives the desired continuity result in the Wasserstein distance. □

Remark 2.8. *In the proof of Theorem 2.7 we only use the assumption that μ_0 is Gaussian to deduce that there exists $\alpha > 0$ such that $\int_X \exp(\alpha \|u\|_X^2)\, \mathrm{d}\mu_0(u) < \infty$. Therefore, the statement of Theorem 2.7 readily extends to any prior measure μ_0 with this property.*

For Bayesian inverse problems with finite data the potential has the form (2.1) where $y \in \mathbb{R}^m$ is the data $\mathcal{G}\colon X \to \mathbb{R}^m$ is the observation operator, and $|\cdot|_\Gamma$ is a covariance weighted norm on $\mathbb{R}^m$. By Lemma 2.3 we know that Assumption 2.2 implies Assumption 2.1 for Φ given by (2.1). Thus, we have the following corollary of Theorem 2.7.

Corollary 2.9. *Assume that $\Phi\colon X \times Y \to \mathbb{R}$ is given by (2.1) and let $\mathcal{G}$ satisfy Assumption 2.2 (1). Assume further that μ_0 is a Gaussian measure satisfying $\mu_0(X) = 1$ and that for all $y \in Y$ the measure μ^y is absolutely continuous with respect to μ_0, $\mu^y \ll \mu_0$, with Randon–Nikodým derivative given by (1.2). Then $y \mapsto \mu^y$ is Lipschitz continuous with respect to the 1-Wasserstein distance: if μ^y and $\mu^{y'}$ are two measures corresponding to data y and y' then for all $r > 0$ there exists $C = C(r) > 0$ such that, if $\|y\|_Y, \|y'\|_Y < r$, then*

$$W_1(\mu^y, \mu^{y'}) \leq C \|y - y'\|_Y$$

3 Approximation of posterior measures in the Wasserstein distance

In order to implement algorithms designed to sample the posterior measure μ^y, we need to make finite-dimensional approximations. Since the dependence on y is not relevant in this section, we suppress it notationally and study measures μ given by

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(u) = \frac{1}{Z}\exp(-\Phi(u)) \tag{3.1}$$

where the normalization constant Z is given by

$$Z = \int_X \exp(-\Phi(u))\, \mathrm{d}\mu_0(u).$$

We approximate μ by approximating Φ over some N-dimensional subspace of X. Specifically, we define μ^N by

$$\frac{\mathrm{d}\mu^N}{\mathrm{d}\mu_0}(u) = \frac{1}{Z^N}\exp\left(-\Phi^N(u)\right) \tag{3.2}$$

where

$$Z^N = \int_X \exp\left(-\Phi^N(u)\right) \mathrm{d}\mu_0(u).$$

The following theorem bounds the 1-Wasserstein distance between μ and μ^N in terms of the error in approximating Φ. Note that this effectively translates approximation results for Φ—which are determined by the forward problem—into approximation results for the posterior μ.

Theorem 3.1. *Assume that the measures μ and μ^N are both absolutely continuous with respect to μ_0, satisfying $\mu_0(X) = 1$, with Randon–Nikodým derivative given by (3.1) and (3.2) and that Φ and Φ^N satisfy Assumption 2.1 (1) and (2) with constants uniform in N. Assume also that for any $\varepsilon > 0$ there is $K = K(\varepsilon) > 0$ such that*

$$\left|\Phi(u) - \Phi^N(u)\right| \leq K \exp\left(\varepsilon \|u\|_X^2\right) \Psi(N), \tag{3.3}$$

where $\Psi(N) \to 0$ as $N \to \infty$. Then the measures μ and μ^N are close with respect to the 1-Wasserstein distance: there is a constant C, independent of N, such that

$$W_1(\mu, \mu^N) \leq C\Psi(N).$$

Proof. The normalization constants Z and Z^N satisfy lower bounds independent of N which are identical to that proved for Z in the course of establishing Theorem 2.7.

Using the estimate (2.2), Assumption 2.1 (1), (3.3), and the fact that μ_0 is a Gaussian measure so that the Fernique Theorem A.1 applies, we find

$$\begin{aligned}
\left|Z - Z^N\right| &\leq \int_X \left(\exp(-\Phi(u)) \vee \exp(-\Phi^N(u))\right) \left|\Phi(u) - \Phi^N(u)\right| \mathrm{d}\mu_0(u) \\
&\leq C \int_X \exp\left(\varepsilon \|u\|_X^2 - M\right) \exp\left(\varepsilon \|u\|_X^2\right) K\Psi(N) \,\mathrm{d}\mu_0(u) \\
&= C \int_X \exp\left(2\varepsilon \|u\|_X^2\right) \mathrm{d}\mu_0(u)\Psi(N) \\
&= C\Psi(N).
\end{aligned}$$

Now, let $\psi \in \mathcal{C}_b(X)$ with $\|\psi\|_{\mathrm{Lip}} \leq 1$. Since μ and μ^N are probability measures, we have

$$\begin{aligned}
\int_X \psi(u) \,\mathrm{d}\left(\mu - \mu^N\right)(u) &= \int_X (\psi(u) - \psi(0)) \,\mathrm{d}\left(\mu - \mu^N\right)(u) + \int_X \psi(0) \,\mathrm{d}\left(\mu - \mu^N\right)(u) \\
&= \int_X (\psi(u) - \psi(0)) \,\mathrm{d}\left(\mu - \mu^N\right)(u) \\
&= \int_X (\psi(u) - \psi(0)) \left(\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(u) - \frac{\mathrm{d}\mu^N}{\mathrm{d}\mu_0}(u)\right) \mathrm{d}\mu_0(u) \\
&= \int_X (\psi(u) - \psi(0)) \left(Z^{-1}\exp(-\Phi(u)) - \left(Z^N\right)^{-1}\exp(-\Phi^N(u))\right) \mathrm{d}\mu_0(u) \\
&= I_1 + I_2
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \int_X (\psi(u) - \psi(0)) Z^{-1}\left(\exp(-\Phi(u)) - \exp(-\Phi^N(u))\right) \mathrm{d}\mu_0(u), \\
I_2 &= \int_X (\psi(u) - \psi(0)) \left(Z^{-1} - \left(Z^N\right)^{-1}\right) \exp(-\Phi^N(u)) \,\mathrm{d}\mu_0(u).
\end{aligned}$$

Using the estimate (2.2) again as well as the fact that $\|\psi\|_{\mathrm{Lip}} \leq 1$, Assumption 2.1 (1), and (3.3) we obtain

$$\begin{aligned}
ZI_1 &\leq \int_X |\psi(u) - \psi(0)| \left|\exp(-\Phi(u)) - \exp(-\Phi^N(y))\right| \mathrm{d}\mu_0(u) \\
&\leq \int_X \|u\|_X \left(\exp(-\Phi(u)) \vee \exp(-\Phi^N(u))\right) \left|\Phi(u)) - \Phi^N(u)\right| \mathrm{d}\mu_0(u) \\
&\leq C\left(\int_X \|u\| \exp\left(2\varepsilon \|u\|_X^2\right) \mathrm{d}\mu_0(u)\right) \Psi(N).
\end{aligned}$$

Since all moments of u in X are finite under the Gaussian measure μ_0 by the Fernique Theorem, the integral in the last line can be bounded by using the Cauchy–Schwarz inequality and again the Fernique Theorem. Since Z is bounded from below by a positive constant, this gives a bound on I_1.

Using the fact that $\|\psi\|_{\mathrm{Lip}} \leq 1$, Assumption 2.1 (1), the above bound on $\left|Z - Z^N\right|$ and again the fact that Z and Z^N are bounded from below by a positive constant independent of N, we get

$$\begin{aligned}
I_2 &\leq \int_X \|u\|_X \left|Z^{-1} - (Z^N)^{-1}\right| \exp(-\Phi^N(u)) \,\mathrm{d}\mu_0(u) \\
&\leq C\left|Z^{-1} - (Z^N)^{-1}\right| \int_X \|u\|_X \exp(\varepsilon \|u\|_X^2) \,\mathrm{d}\mu_0(u)
\end{aligned}$$

$$
\begin{aligned}
&\le C\left(Z^{-2} \vee (Z^N)^{-2}\right)\left|Z - Z^N\right| \\
&\le C\Psi(N).
\end{aligned}
$$

Here we used the same arguments as before to bound the integral $\int_X \|u\|_X \exp\left(\varepsilon \|u\|_X^2\right) \mathrm{d}\mu_0(u)$. Combining the bounds for I_1 and I_2 gives the desired continuity result in the Wasserstein distance. □

Again, if the data is finite, the potential has the form (2.1), where $y \in \mathbb{R}^m$ is the data, $\mathcal{G}\colon X \to \mathbb{R}^m$ is the observation operator, and $|\cdot|_\Gamma$ is a covariance weighted norm on $\mathbb{R}^m$. If $\mathcal{G}^N$ is an approximation to $\mathcal{G}$ and we define

$$
\Phi^N(u;y) := \left|y - \mathcal{G}^N(u)\right|_\Gamma \tag{3.4}
$$

then we can define an approximation μ^N to μ as in (3.2) and we have the following corollary.

Corollary 3.2. *Assume that the measures μ and μ^N are both absolutely continuous with respect to μ_0, satisfying $\mu_0(X) = 1$, with Randon–Nikodým derivative given by* (3.1), (2.1) *and* (3.2), (3.4) *respectively. Assume also that $\mathcal{G}$ is approximated by a function $\mathcal{G}^N$ with the property that for any $\varepsilon > 0$ there is $K' = K'(\varepsilon) > 0$ such that*

$$
\left|\mathcal{G}(u) - \mathcal{G}^N(u)\right| \le K' \exp\left(\varepsilon \|u\|_X^2\right) \Psi(N), \tag{3.5}
$$

where $\Psi(N) \to 0$ as $N \to \infty$. If $\mathcal{G}$ and $\mathcal{G}^N$ satisfy Assumption 2.2 (1) *uniformly in N, then the measures μ and μ^N are close with respect to the* 1*-Wasserstein distance: there is a constant C, independent of N, such that*

$$
W_1(\mu, \mu^N) \le C\Psi(N).
$$

Proof. Using Assumption 2.2 (1) we get for all $\varepsilon > 0$ and $y \in \mathbb{R}^m$

$$
\begin{aligned}
\left|\Phi(u) - \Phi^N(u)\right| &\le \frac{1}{2}\left|2y - \mathcal{G}(u) - \mathcal{G}^N(u)\right|_\Gamma \left|\mathcal{G}(u) - \mathcal{G}^N(u)\right|_\Gamma \\
&\le C\left(|y| + \exp\left(\varepsilon \|u\|_X^2 + M\right)\right) \exp\left(\varepsilon \|u\|_X^2\right) \Psi(N) \\
&\le C(2\varepsilon, y) \exp\left(2\varepsilon \|u\|_X^2\right) \Psi(N)
\end{aligned}
$$

such that (3.3) holds and, in view of Lemma 2.3, we can apply Theorem 3.1. □

In Theorem 3.1 it is necessary that the constant in the error bound (3.3) for approximating the function Φ by Φ^N is integrable by use of the Fernique Theorem A.1. In case such integrability is not at hand, we can still derive the convergence result, albeit at possibly weaker rates.

Theorem 3.3. *Assume that the measures μ and μ^N are both absolutely continuous with respect to μ_0, satisfying $\mu_0(X) = 1$, with Randon–Nikodým derivative given by* (3.1) *and* (3.2) *and that Φ and Φ^N satisfy Assumption 2.1* (1) *and* (2) *with constants uniform in N. Assume also that for any $R > 0$ there is $K = K(R) > 0$ such that for all $u \in X$ with $\|u\|_X \le R$*

$$
\left|\Phi(u) - \Phi^N(u)\right| \le K\Psi(N), \tag{3.6}
$$

where $\Psi(N) \to 0$ as $N \to \infty$. Then

$$
W_1(\mu, \mu^N) \to 0
$$

as $N \to \infty$.

Proof. The normalization constants Z and Z^N satisfy lower bounds independent of N which are identical to that proved for Z in the course of establishing Theorem 2.7.

Using the estimate (2.2), Assumption 2.1 (1), and (3.6), we find

$$
\begin{aligned}
\left|Z - Z^N\right| &\le \int_X \left|\exp(-\Phi(u)) - \exp(-\Phi^N(u))\right| \mathrm{d}\mu_0(u) \\
&\le \int_{\{\|u\|_X \le R\}} \exp(\varepsilon \|u\|_X^2 - M) \left|\Phi(u) - \Phi^N(u)\right| \mathrm{d}\mu_0(u) \\
&\quad + \int_{\{\|u\|_X > R\}} 2\exp(\varepsilon \|u\|_X^2 - M) \,\mathrm{d}\mu_0(u) \\
&\le \exp(\varepsilon R^2 - M) K(R) \Psi(N) + J_R \\
&=: K_1(R)\Psi(N) + J_R
\end{aligned}
$$

where

$$
J_R = \int_{\{\|u\|_X > R\}} 2\exp(\varepsilon \|u\|_X^2 - M) \,\mathrm{d}\mu_0(u).
$$

Because of the Fernique Theorem A.1, $J_R \to 0$ as $R \to \infty$. Therefore, for any $\delta > 0$ we can choose R sufficiently large such that $J_R < \delta$. By choosing N large enough that $K_1(R)\Psi(N) < \delta$, we get $\left|Z - Z^N\right| < 2\delta$. Therefore, we have $Z^N \to Z$ as $N \to \infty$.

Now, let $\psi \in \mathcal{C}_b(X)$ with $\|\psi\|_{\mathrm{Lip}} \le 1$. Since μ and μ^N are probability measures, we have

$$
\begin{aligned}
\int_X \psi(u) \,\mathrm{d}\left(\mu - \mu^N\right)(u) &= \int_X (\psi(u) - \psi(0)) \,\mathrm{d}\left(\mu - \mu^N\right)(u) + \int_X \psi(0) \,\mathrm{d}\left(\mu - \mu^N\right)(u) \\
&= \int_X (\psi(u) - \psi(0)) \,\mathrm{d}\left(\mu - \mu^N\right)(u) \\
&= \int_X (\psi(u) - \psi(0)) \left(\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(u) - \frac{\mathrm{d}\mu^N}{\mathrm{d}\mu_0}(u)\right) \mathrm{d}\mu_0(u) \\
&= \int_X (\psi(u) - \psi(0)) \left(Z^{-1}\exp(-\Phi(u)) - \left(Z^N\right)^{-1} \exp(-\Phi^N(u))\right) \mathrm{d}\mu_0(u) \\
&= I_1 + I_2
\end{aligned}
$$

where

$$
\begin{aligned}
I_1 &= \int_X (\psi(u) - \psi(0)) Z^{-1} \left(\exp(-\Phi(u)) - \exp(-\Phi^N(u))\right) \mathrm{d}\mu_0(u), \\
I_2 &= \int_X (\psi(u) - \psi(0)) \left(Z^{-1} - \left(Z^N\right)^{-1}\right) \exp(-\Phi^N(u)) \,\mathrm{d}\mu_0(u).
\end{aligned}
$$

Using the estimate (2.2), the fact that $\|\psi\|_{\mathrm{Lip}} \le 1$, Assumption 2.1 (1), and (3.6) we obtain

$$
\begin{aligned}
I_1 &\le Z^{-1} \int_X |\psi(u) - \psi(0)| \left|\exp(-\Phi(u)) - \exp(-\Phi^N(y))\right| \mathrm{d}\mu_0(u) \\
&\le Z^{-1} \int_{\{\|u\|_X \le R\}} \|u\|_X \exp(\varepsilon \|u\|_X^2 - M) K(R) \Psi(N) \,\mathrm{d}\mu_0(u) \\
&\quad + Z^{-1} \int_{\{\|u\|_X > R\}} 2\|u\|_X \exp(\varepsilon \|u\|_X^2 - M) \,\mathrm{d}\mu_0(u) \\
&\le Z^{-1} R \exp(\varepsilon R^2 - M) K(R) \Psi(N) + Z^{-1} \int_{\{\|u\|_X > R\}} 2\|u\|_X \exp(\varepsilon \|u\|_X^2 - M) \,\mathrm{d}\mu_0(u).
\end{aligned}
$$

Since Z is bounded from below and all momemts of u in X are finite under the Gaussian measure μ_0, an argument similar to the one above for $\left|Z - Z^N\right|$ shows that $I_1 \to 0$ as $N \to \infty$.

Using the fact that $\|\psi\|_{\mathrm{Lip}} \le 1$, Assumption 2.1 (1), the bound on $\left|Z - Z^N\right|$ and again the fact that Z and Z^N are bounded from below by a positive constant independent of N, we get as before

$$
\begin{aligned}
I_2 &\leq \int_X \|u\|_X \left|Z^{-1} - (Z^N)^{-1}\right| \exp(-\Phi^N(u))\,\mathrm{d}\mu_0(u) \\
&\leq C \left|Z^{-1} - (Z^N)^{-1}\right| \int_X \|u\|_X \exp(\varepsilon \|u\|_X^2)\,\mathrm{d}\mu_0(u) \\
&\leq C \left(Z^{-2} \vee (Z^N)^{-2}\right) |Z - Z^N|.
\end{aligned}
$$

Thus, we have $I_2 \to 0$ as $N \to \infty$. Combining gives the claimed continuity result in the Wasserstein distance. ☐

If the data is finite, we can derive the following Corollary in analogy to Corollary 3.2.

Corollary 3.4. *Assume that the measures μ and μ^N are both absolutely continuous with respect to μ_0, satisfying $\mu_0(X) = 1$, with Randon–Nikodým derivative given by* (3.1), (2.1) *and* (3.2), (3.4) *respectively. Assume also that $\mathcal{G}$ is approximated by a function $\mathcal{G}^N$ with the property that for any $R > 0$ there is $K' = K'(R) > 0$ such that for all $u \in X$ with $\|u\|_X \leq R$*

$$
\left|\mathcal{G}(u) - \mathcal{G}^N(u)\right| \leq K'\Psi(N),
$$

where $\Psi(N) \to 0$ as $N \to \infty$. If $\mathcal{G}$ and $\mathcal{G}^N$ satisfy Assumption 2.2 (1) *uniformly in N, then*

$$
W_1(\mu, \mu^N) \to 0
$$

as $N \to \infty$.

Proof. Using Assumption 2.2 (1) (with $\varepsilon = 1$) we get for all $R > 0$, $u \in X$ with $\|u\|_X \leq R$, and $y \in \mathbb{R}^m$

$$
\begin{aligned}
\left|\Phi(u) - \Phi^N(u)\right| &\leq \frac{1}{2} \left|2y - \mathcal{G}(u) - \mathcal{G}^N(u)\right|_\Gamma \left|\mathcal{G}(u) - \mathcal{G}^N(u)\right|_\Gamma \\
&\leq C \left(|y| + \exp\left(\|u\|_X^2 + M(1)\right)\right) K(R)\Psi(N) \\
&\leq C \left(|y| + \exp(R^2 + M(1))\right) K(R)\Psi(N)
\end{aligned}
$$

such that (3.6) holds and, in view of Lemma 2.3, we can apply Theorem 3.3. ☐

4 Bayesian inverse problems for conservation laws

In this section we use stability and convergence rate estimates for scalar conservation laws to establish that the associated inverse problems may be placed in the general framework for Bayesian inverse problems in the Wasserstein distance. To this end, we consider scalar conservation laws for which the available theory is very mature as well as scalar conservation laws with discontinuous flux where stability in the model parameters and convergence rates were established only very recently. We start by recalling the necessary well-posedness results for entropy solutions of scalar conservation laws.

4.1 Scalar conservation laws in several space dimensions

We consider the Cauchy problem for scalar conservation laws of the form

$$
\begin{aligned}
w_t + \nabla_x \cdot f(w) = 0, &\quad (x,t) \in \mathbb{R}^d \times (0,T), \\
w(x,0) = \bar{w}(x), &\quad x \in \mathbb{R}^d.
\end{aligned}
\tag{4.1}
$$

Here, the unknown is $w\colon \mathbb{R}^d \times [0,T] \to \mathbb{R}$ and $f = (f_1, \ldots, f_d) \in \mathcal{C}^{0,1}(\mathbb{R};\mathbb{R}^d)$ is the flux function.

4.1.1 Entropy solutions

Since weak solutions of (4.1) are not unique we consider entropy solutions in the following sense.

Definition 4.1. *We call a function $w \in \mathrm{L}^\infty(\mathbb{R}^d \times (0,T)) \cap \mathcal{C}([0,T]; \mathrm{L}^1(\mathbb{R}^d))$ an entropy solution of* (4.1) *if for all $c \in \mathbb{R}$*

$$\int_0^T \int_{\mathbb{R}^d} \left(|w-c|\, \varphi_t + \operatorname{sgn}(w-c) \sum_{j=1}^d (f_j(u) - f_j(c)) \varphi_{x_j} \right) \mathrm{d}x\, \mathrm{d}t + \int_{\mathbb{R}^d} |\bar{w}(x) - c|\, \varphi(x,0)\, \mathrm{d}x \geq 0$$

for all nonnegative $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^d \times [0,T))$.

It is well-known that the Cauchy problem (4.1) admits, for each $\bar{w} \in \left(\mathrm{L}^1 \cap \mathrm{BV}\right)(\mathbb{R}^d)$, a unique entropy solution and we summarize the classical results on existence and uniqueness of entropy solutions in the following theorem (see, e.g., [9]).

Theorem 4.2.

1. *For every $\bar{w} \in \mathrm{L}^\infty(\mathbb{R}^d)$,* (4.1) *admits a unique entropy solution $w \in \mathrm{L}^\infty(\mathbb{R}^d \times (0,T))$.*
2. *For every $t > 0$, the solution operator S_t given by*

$$S_t \bar{w} = w(\cdot, t)$$

satisfies

(i) *$S_t \colon \mathrm{L}^1(\mathbb{R}^d) \to \mathrm{L}^1(\mathbb{R}^d)$ is a contraction, i.e.,*

$$\|S_t \bar{w} - S_t \hat{w}\|_{\mathrm{L}^1(\mathbb{R}^d)} \leq \|\bar{w} - \hat{w}\|_{\mathrm{L}^1(\mathbb{R}^d)}$$

for all $\bar{w}, \hat{w} \in \mathrm{L}^1(\mathbb{R}^d)$.

(ii) *S_t maps $(\mathrm{L}^1 \cap \mathrm{BV})(\mathbb{R}^d)$ into itself and*

$$\mathrm{TV}(S_t \bar{w}) \leq \mathrm{TV}(\bar{w})$$

for all $\bar{w} \in (\mathrm{L}^1 \cap \mathrm{BV})(\mathbb{R}^d)$.

(iii) *For every $\bar{w} \in (\mathrm{L}^1 \cap \mathrm{L}^\infty)(\mathbb{R}^d)$*

$$\|S_t \bar{w}\|_{\mathrm{L}^1(\mathbb{R}^d)} \leq \|\bar{w}\|_{\mathrm{L}^1(\mathbb{R}^d)}\,, \tag{4.2}$$

$$\|S_t \bar{w}\|_{\mathrm{L}^\infty(\mathbb{R}^d)} \leq \|\bar{w}\|_{\mathrm{L}^\infty(\mathbb{R}^d)}\,. \tag{4.3}$$

(iv) *$t \mapsto S_t$ is a uniformly continuous mapping from $\mathrm{L}^1(\mathbb{R}^d)$ into $\mathcal{C}_b([0,\infty); \mathrm{L}^1(\mathbb{R}^d))$ and*

$$\|t \mapsto S_t \bar{w}\|_{\mathcal{C}([0,T]; \mathrm{L}^1(\mathbb{R}^d))} \leq \|\bar{w}\|_{\mathrm{L}^1(\mathbb{R}^d)}$$

for all $\bar{w} \in \mathrm{L}^1(\mathbb{R}^d)$.

In the following, our notation for the solution operator will not only carry the dependence on the initial datum, but also on the flux. We will write

$$S_t(\bar{w}, f) = w(\cdot, t)$$

and understand S_t as a map from $\left(\mathrm{L}^1 \cap \mathrm{L}^\infty \cap \mathrm{BV}\right)(\mathbb{R}^d) \times \mathcal{C}^{0,1}(\mathbb{R}; \mathbb{R}^d)$ to $\mathrm{L}^1(\mathbb{R}^d)$ with the properties listed above. The following theorem shows that this map is locally Lipschitz continuous.

Theorem 4.3 ([11, Thm. 4.3])**.** *Assume $\bar{w}, \hat{w} \in \left(\mathrm{L}^1 \cap \mathrm{L}^\infty \cap \mathrm{BV}\right)(\mathbb{R}^d)$ and $f, g \in \mathcal{C}^{0,1}(\mathbb{R}; \mathbb{R}^d)$. Then the solution operator satisfies*

$$\|S_t(\bar{w}, f) - S_t(\hat{w}, g)\|_{\mathrm{L}^1(\mathbb{R}^d)} \leq \|\bar{w} - \hat{w}\|_{\mathrm{L}^1(\mathbb{R}^d)} + t \min\left(\mathrm{TV}(\bar{w}), \mathrm{TV}(\hat{w})\right) \|f - g\|_{\mathrm{Lip}} \tag{4.4}$$

for every $0 \leq t \leq T$.

4.1.2 Finite volume methods

We briefly describe the conventional approach of numerically approximating solutions of scalar conservation laws through finite volume methods (cf. [15, 6, 12]).

We discretize the spatial computational domain with cells

$$\mathcal{C}_{i_1,\dots,i_d} := (x^1_{i_1-1/2}, x^1_{i_1+1/2}) \times \ldots \times (x^d_{i_d-1/2}, x^d_{i_d+1/2}) \subset \mathbb{R}^d$$

with corresponding cell midpoints

$$x_{i_1,\dots,i_d} := \left(\frac{x^1_{i_1+1/2} + x^1_{i_1-1/2}}{2}, \dots, \frac{x^d_{i_d+1/2} + x^d_{i_d-1/2}}{2} \right).$$

For simplicity, we assume that the mesh is equidistant, meaning

$$x^k_{i_k+1/2} - x^k_{i_k-1/2} = \Delta x, \qquad \text{for all } k = 1, \dots, d \text{ and } i_k \in \mathbb{Z},$$

for some $\Delta x > 0$. We consider a uniform discretization in time with time step $\Delta t > 0$ such that the time interval $[0, T]$ is partitioned into intervals $[t^n, t^{n+1})$ where $t^n = n\Delta t$ and that $\lambda := \frac{\Delta t}{\Delta x}$ is constant and satisfies a standard CFL condition based on the maximum wave speed (see e.g. [9]).

We consider the following numerical scheme:

$$\begin{aligned} w^{n+1}_{i_1,\dots,i_d} &= w^n_{i_1,\dots,i_d} - \lambda \sum_{k=1}^{d} \left(F^{k,n}_{i_1,\dots,i_k+1/2,\dots,i_d} - F^{k,n}_{i_1,\dots,i_k-1/2,\dots,i_d} \right), \\ w^0_{i_1,\dots,i_d} &= \frac{1}{\Delta x^d} \int_{\mathcal{C}_{i_1,\dots,i_d}} \bar{w}(x)\, \mathrm{d}x, \end{aligned} \tag{4.5}$$

where $F^{k,n}$ is a numerical flux function in direction k. In a $(2p+1)$-point scheme, the numerical flux function $F^{k,n}_{i_1,\dots,i_k+1/2,\dots,i_d}$ can be written as a function of the $2p$ values $\left(w^n_{i_1,\dots,i_k+j,\dots,i_D}\right)^p_{j=-p+1}$. Furthermore, we assume that the numerical flux function is consistent with f and locally Lipschitz continuous, i.e., for every bounded set $K \subset \mathbb{R}$, there exists a constant $C > 0$ such that for $k = 1, \dots, d$,

$$\left| F^{k,n}_{i_1,\dots,i_k+1/2,\dots,i_d} - f_k\left(w^n_{i_1,\dots,i_d}\right) \right| \le C \sum_{j=-p+1}^{p} \left| w^n_{i_1,\dots,i_k+j,\dots,i_d} - w^n_{i_1,\dots,i_d} \right|$$

whenever $w^n_{i_1,\dots,i_k-p+1,\dots,i_d}, \dots, w^n_{i+1,\dots,i_k+p,\dots,i_d} \in K$. Finally, we consider monotone finite volume methods where the right-hand side of (4.5) is nondecreasing in each argument.

We define the numerical solution operator

$$S^{\Delta x}_t \colon \left(\mathrm{L}^1 \cap \mathrm{L}^\infty \cap \mathrm{BV}\right)(\mathbb{R}^d) \times \mathcal{C}^{0,1}(\mathbb{R}; \mathbb{R}^d) \to \left(\mathrm{L}^1 \cap \mathrm{L}^\infty \cap \mathrm{BV}\right)(\mathbb{R}^d)$$

by

$$(S^{\Delta x}_t(\bar{w}, f))(x) = w^n_{i_1,\dots,i_d}, \qquad (x, t) \in \mathcal{C}_{i_1,\dots,i_d} \times [t^n, t^{n+1}).$$

The following convergence rate estimate is due to Kutznetsov.

Theorem 4.4 ([12, Thm. 4]). *Let $\bar{w} \in \left(\mathrm{L}^1 \cap \mathrm{L}^\infty \cap \mathrm{BV}\right)(\mathbb{R}^d)$, $f \in \mathcal{C}^{0,1}(\mathbb{R}; \mathbb{R}^d)$, $S_t(\bar{w}, f)$ the corresponding entropy solution of* (4.1) *and $S^{\Delta x}_t(\bar{w}, f)$ the numerical approximation given by* (4.5). *Then we have the following convergence rate estimate:*

$$\left\| S_t(\bar{w}, f) - S^{\Delta x}_t(\bar{w}, f) \right\|_{\mathrm{L}^1(\mathbb{R}^d)} \le C \left(\mathrm{TV}(\bar{w}) + \|f\|_{\mathrm{Lip}} \, \mathrm{TV}(\bar{w}) \right) \Delta x^{1/2} \tag{4.6}$$

for all $0 \le t \le T$ where C is independent of $\Delta t, \Delta x, \bar{w}$, and f.

Note that the convergence rate estimate (4.6) is optimal in the sense that the exponent $1/2$ cannot be improved without further assumptions on the initial datum [3, 19] (see [18] for an overview of the literature regarding optimal convergence rates).

4.1.3 Bayesian inverse problems for scalar conservation laws

We will now use the above well-posedness and approximation results to show that the abstract framework of Sections 2 and 3 can be applied to Bayesian inverse problems for scalar conservation laws where the inputs $u = (\bar{w}, f)$ are inferred from measurements of the observables. To that end, we define $X = \left(\mathrm{L}^1 \cap \mathrm{L}^\infty \cap \mathrm{BV}\right)(\mathbb{R}^d) \times V$ equipped with the norm

$$\|(\bar{w}, f)\|_X = \|\bar{w}\|_{\mathrm{L}^1(\mathbb{R}^d)} + \mathrm{TV}(\bar{w}) + \|\bar{w}\|_{\mathrm{L}^\infty(\mathbb{R}^d)} + \|f\|_V$$

where V is some separable Banach space embedded in $\mathcal{C}^{0,1}(\mathbb{R};\mathbb{R}^d)$. Specifically, in light of the Sobolev Embedding Theorem we can take $V = \mathrm{W}^{2,p}(\mathbb{R};\mathbb{R}^d)$ for any $1 < p < \infty$, for example $V = \mathrm{H}^2(\mathbb{R};\mathbb{R}^d)$. We then consider observation operators of the form $\mathcal{G} \colon X \to \mathbb{R}^m$ given by

$$(\mathcal{G}(\bar{w}, f))_j = \int_0^T \int_{\mathbb{R}^d} \psi_j(x,t) g_j((S_t(\bar{w}, f))(x))\,\mathrm{d}x\,\mathrm{d}t, \qquad j = 1, \ldots, m, \tag{4.7}$$

for $\psi_j \in \mathrm{L}^1(\mathbb{R}^d \times (0,T)) \cap \mathrm{L}^1(0,T;\mathrm{L}^\infty(\mathbb{R}^d))$ and $g_j \in \mathcal{C}^1(\mathbb{R};\mathbb{R}^d)$ with $\|g_j\|_{\mathcal{C}^1(\mathbb{R};\mathbb{R}^d)} < \infty$.

The following lemma shows that the Bayesian inverse problem of determining the initial datum $\bar{w}$ and the flux function f given observations of the form (4.7) is well-posed.

Lemma 4.5. *The observation operator $\mathcal{G}$ defined by* (4.7) *satisfies Assumption 2.2. Therefore, by Corollary 2.9, the Bayesian inverse problem associated with the observation operator $\mathcal{G}$ is well-posed.*

Proof. It suffices to consider the case $m = 1$. Using the L^∞ bound (4.3) of the solution operator we find

$$\begin{aligned}
|\mathcal{G}(\bar{w}, f)| &\leq \int_0^T \int_{\mathbb{R}^d} |\psi(x,t)|\,|g((S_t(\bar{w}, f))(x))|\,\mathrm{d}x\,\mathrm{d}t \\
&\leq \int_0^T \int_{\mathbb{R}^d} |\psi(x,t)|\,|g((S_t(\bar{w}, f))(x)) - g(0)|\,\mathrm{d}x\,\mathrm{d}t + \int_0^T \int_{\mathbb{R}^d} |\psi(x,t)|\,|g(0)|\,\mathrm{d}x\,\mathrm{d}t \\
&\leq \|g'\|_\infty \|\psi\|_{\mathrm{L}^1(\mathbb{R}^d \times (0,T))} \max_{0 \leq t \leq T} \|S_t(\bar{w}, f)\|_{\mathrm{L}^\infty(\mathbb{R}^d)} + \|g\|_\infty \|\psi\|_{\mathrm{L}^1(\mathbb{R}^d \times (0,T))} \\
&\leq \|g'\|_\infty \|\psi\|_{\mathrm{L}^1(\mathbb{R}^d \times (0,T))} \|\bar{w}\|_{\mathrm{L}^\infty(\mathbb{R}^d)} + \|g\|_\infty \|\psi\|_{\mathrm{L}^1(\mathbb{R}^d \times (0,T))} \\
&=: C_1 \|\bar{w}\|_{\mathrm{L}^\infty(\mathbb{R}^d)} + C_2.
\end{aligned}$$

Let now $\varepsilon > 0$. Using the estimates $\ln(x) \leq x$ and $\exp(ax) \leq \exp\left(\varepsilon x^2 + \frac{a^2}{\varepsilon}\right)$ we get

$$\begin{aligned}
C_1 \|\bar{w}\|_{\mathrm{L}^\infty(\mathbb{R}^d)} + C_2 &\leq \exp(C_1 \|\bar{w}\|_{\mathrm{L}^\infty(\mathbb{R}^d)}) + C_2 \\
&\leq (1 + C_2) \exp(C_1 \|\bar{w}\|_{\mathrm{L}^\infty(\mathbb{R}^d)}) \\
&\leq (1 + C_2) \exp\left(\varepsilon \|\bar{w}\|^2_{\mathrm{L}^\infty(\mathbb{R}^d)} + \frac{C_1^2}{\varepsilon}\right) \\
&\leq \exp\left(\varepsilon \|(\bar{w}, f)\|_X^2 + \frac{C_1^2}{\varepsilon} + 1 + C_2\right)
\end{aligned}$$

which shows that Assumption 2.2 (1) is satisfied. On the other hand, for $r > 0$ and $(\bar{w}, f), (\hat{w}, g) \in X$ with $\|(\bar{w}, f)\|_X, \|(\hat{w}, g)\|_X < r$, because of (4.4) we have

$$\begin{aligned}
&|\mathcal{G}(\bar{w}, f) - \mathcal{G}(\hat{w}, g)| \\
&\leq \int_0^T \int_{\mathbb{R}^d} |\psi(x,t)|\,|g((S_t(\bar{w}, f))(x)) - g((S_t(\hat{w}, g))(x))|\,\mathrm{d}x\,\mathrm{d}t \\
&\leq \|g'\|_\infty \int_0^T \|\psi(\cdot, t)\|_{\mathrm{L}^\infty(\mathbb{R}^d)} \|S_t(\bar{w}, f) - S_t(\hat{w}, g)\|_{\mathrm{L}^1(\mathbb{R}^d)}\,\mathrm{d}t
\end{aligned}$$

$$\le \|\psi\|_{\mathrm{L}^1(0,T;\mathrm{L}^\infty(\mathbb{R}^d))} \|g'\|_\infty \Big(\|\bar{w}-\hat{w}\|_{\mathrm{L}^1(\mathbb{R}^d)} + T\min(\mathrm{TV}(\bar{w}),\mathrm{TV}(\hat{w}))\,\|f-g\|_{\mathrm{Lip}}\Big)$$
$$\le \|\psi\|_{\mathrm{L}^1(0,T;\mathrm{L}^\infty(\mathbb{R}^d))} \|g'\|_\infty \max(1, CTr)\,\|(\bar{w},f)-(\hat{w},g)\|_X$$

such that Assumption 2.2 (2) is satisfied. □

Using the finite volume method (4.5), we can define an approximation to $\mathcal{G}$ by replacing the solution operator S in (4.7) by the numerical solution operator $S^{\Delta x}$,

$$\big(\mathcal{G}^{\Delta x}(\bar{w},f)\big)_j = \int_0^T \int_{\mathbb{R}^d} \psi_j(x,t) g_j\big((S_t^{\Delta x}(\bar{w},f))(x)\big)\,\mathrm{d}x\,\mathrm{d}t, \qquad j=1,\dots,m. \tag{4.8}$$

Lemma 4.6. *The approximation $\mathcal{G}^{\Delta x}$ defined in (4.8) of the observation operator $\mathcal{G}$ defined in (4.7) satisfies (3.5) in Corollary 3.2 with $\Psi(\Delta x^{-1}) = \sqrt{\Delta x}$.*

Proof. Let $\varepsilon > 0$. Using the convergence rate estimate (4.6) as well as the estimates $\ln(x) \le x$ and $\exp(ax) \le \exp\left(\varepsilon x^2 + \frac{a^2}{\varepsilon}\right)$ we find

$$\begin{aligned}
\big|\mathcal{G}(\bar{w},f) - \mathcal{G}^{\Delta x}(\bar{w},f)\big| &\le \|g'\|_\infty \int_0^T \|\psi(\cdot,t)\|_{\mathrm{L}^\infty(\mathbb{R}^d)} \big\|S_t(\bar{w},f) - S_t^{\Delta x}(\bar{w},f)\big\|_{\mathrm{L}^1(\mathbb{R}^d)}\,\mathrm{d}t \\
&\le \|g'\|_\infty \|\psi\|_{\mathrm{L}^1(0,T;\mathrm{L}^\infty(\mathbb{R}^d))} C\left(\mathrm{TV}(\bar{w}) + \|f\|_{\mathrm{Lip}}\,\mathrm{TV}(\bar{w})\right)\sqrt{\Delta x} \\
&\le C\left(\|(\bar{w},f)\|_X + \|(\bar{w},f)\|_X^2\right)\sqrt{\Delta x} \\
&\le C\exp\left(\ln(\|(\bar{w},f)\|_X) + \ln(\|(\bar{w},f)\|_X + 1)\right)\sqrt{\Delta x} \\
&\le C\exp\left(2\,\|(\bar{w},f)\|_X + 1\right)\sqrt{\Delta x} \\
&\le C\exp\left(\varepsilon\,\|(\bar{w},f)\|_X^2 + \frac{4}{\varepsilon} + 1\right)\sqrt{\Delta x} \\
&= C\exp\left(\frac{4}{\varepsilon} + 1\right)\exp\left(\varepsilon\,\|(\bar{w},f)\|_X^2\right)\sqrt{\Delta x}.
\end{aligned}$$

□

Remark 4.7. *Note that the assertion of Lemma 4.6 also holds for $\psi_j = \delta_T$, $j=1,\dots,m$, where δ_T is the Dirac delta function. In that case $\mathcal{G}^{\Delta x}$ takes the form*

$$\big(\mathcal{G}^{\Delta x}(\bar{w},f)\big)_j = \int_{\mathbb{R}^d} g_j\big((S_T^{\Delta x}(\bar{w},f))(x)\big)\,\mathrm{d}x, \qquad j=1,\dots,m,$$

which we will use in Section 5 for our numerical experiments.

4.2 Scalar conservation laws with discontinuous flux in one dimension

As a second application, we consider the Cauchy problem for scalar conservation laws with discontinuous flux of the form

$$\begin{aligned}
w_t + f(k(x),w)_x &= 0, \quad (x,t) \in \mathbb{R}\times(0,T), \\
w(x,0) &= \bar{w}(x), \quad x \in \mathbb{R}
\end{aligned} \tag{4.9}$$

where the flux is strictly increasing in w and has a possibly discontinuous spatial dependency through the coefficient k.

Note that if the spatial dependency coefficient k is piecewise constant with finitely many discontinuities we effectively consider standard conservation laws where the flux function changes across finitely many points in space. In particular, this includes the important so-called two-flux case

$$w_t + \big(H(x)f(w) + (1-H(x))g(w)\big)_x = 0$$

where H is the Heaviside function.

4.2.1 Adapted entropy solutions

We assume that the flux is strictly increasing in w and consider solutions in the sense of adapted entropy solutions (see [4, 1]). To that end, we define for $p \in \mathbb{R}$ the function $c_p \colon \mathbb{R} \to \mathbb{R}$ through the equation

$$f(k(x), c_p(x)) = p \qquad \text{for all } x \in \mathbb{R}.$$

This equation has a unique solution for each $x \in \mathbb{R}$ since the flux is strictly increasing in w.

Definition 4.8 ([4, 1])**.** *We call a function $w \in \mathrm{L}^\infty(\mathbb{R} \times (0,T)) \cap \mathcal{C}([0,T]; \mathrm{L}^1(\mathbb{R}))$ an adapted entropy solution of (4.9) if for all $p \in \mathbb{R}$*

$$\int_0^T \int_{\mathbb{R}} \left(|w - c_p(x)|\, \varphi_t + \operatorname{sgn}(w - c_p(x))(f(k(x), w) - f(k(x), c_p(x)))\varphi_x \right) \mathrm{d}x\, \mathrm{d}t \\ + \int_{\mathbb{R}} |\bar{w}(x) - c_p(x)|\, \varphi(x, 0)\, \mathrm{d}x \geq 0$$

for all nonnegative $\varphi \in \mathcal{C}_c^\infty(\mathbb{R} \times [0,T))$.

Since stability results for (4.9) with respect to the modeling parameters $\bar{w}$, k, and f are only available under the assumption that k is piecewise constant with finitely many discontinuities, we will restrict the exposition to that case from this point on. However, we want to remark that more general results regarding existence and uniqueness of adapted entropy solutions are available in the literature and we refer the reader to [23, 16, 1].

Theorem 4.9. *Let $f \in \mathcal{C}^2(\mathbb{R}^2; \mathbb{R})$ be strictly increasing in w in the sense that $f_w \geq \alpha > 0$, and assume that $f(k^*, 0) = 0$ for all $k^* \in \mathbb{R}$. Let further k be piecewise constant with finitely many discontinuities and $\bar{w} \in (\mathrm{L}^\infty \cap \mathrm{BV})(\mathbb{R})$. Then there exists a unique entropy solution w of (4.9) and the solution operator S_t given by*

$$S_t \bar{w} = w(\cdot, t)$$

satisfies

(i) *For all $0 \leq t \leq T$*

$$\|S_t \bar{w}\|_{\mathrm{L}^1(\mathbb{R})} \leq \|\bar{w}\|_{\mathrm{L}^1(\mathbb{R})},$$
$$\|S_t \bar{w}\|_{\mathrm{L}^\infty(\mathbb{R})} \leq \frac{C_f}{\alpha} \|\bar{w}\|_{\mathrm{L}^\infty(\mathbb{R})},$$

and

$$\mathrm{TV}(S_t \bar{w}) \leq C(\mathrm{TV}(\bar{w}) + \mathrm{TV}(k))$$

where C_f denotes the maximal Lipschitz constant of f.

(ii) *For all $x \in \mathbb{R}$*

$$\mathrm{TV}_{[0,T]}(t \mapsto (S_t \bar{w})(x)) \leq C\,\mathrm{TV}(\bar{w}).$$

Proof. The existence and uniqueness statement follows from the theory developed by Baiti and Jenssen [4]. The L^1, L^∞, and TV bounds follow from [17, Thm. 4.1], [23, Thm. 1.4], and [3, Lem. 4.6] respectively. □

Similarly to before, we will denote the solution operator by $S_t(\bar{w}, k, f)$ to highlight the dependence on k and f as well. We have the following Lipschitz continuity result.

Theorem 4.10 ([17, Thm. 4.1])**.** *Let f and g be flux functions satisfying the assumptions of Theorem 4.9, k and l be piecewise constant functions with finitely many discontinuities and $\bar{w}, \hat{w} \in (\mathrm{L}^\infty \cap \mathrm{BV})(\mathbb{R})$. Then the solution operator satisfies*

$$\|S_t(\bar{w}, k, f) - S_t(\hat{w}, l, g)\|_{\mathrm{L}^1(\mathbb{R})} \leq \|\bar{w} - \hat{w}\|_{\mathrm{L}^1(\mathbb{R})} + C\left(\|k - l\|_{\mathrm{L}^\infty(\mathbb{R})} + \|f_w - g_w\|_{\mathrm{L}^\infty(\mathbb{R}^2;\mathbb{R})} \right). \tag{4.10}$$

for every $0 \leq t \leq T$.

Note that the constant C in (4.10) depends linearly on (products of) the L^∞ and TV norms of $\bar{w}$ and $\hat{w}$, the Lipschitz constants of f and g and the maximum number of discontinuities in k and l.

4.2.2 Finite volume methods

We will now present a class of finite volume methods for (4.9) introduced in [3]. As before, we discretize the domain $\mathbb{R} \times [0,T]$ using the spatial and temporal grid discretization parameters Δx and Δt. The resulting grid cells we denote by $\mathcal{C}_j = (x_{j+1/2}, x_{j-1/2})$ in space and $[t^n, t^{n+1})$ in time for points $x_{j+1/2}$, such that $x_{j+1/2} - x_{j-1/2} = \Delta x$, $j \in \mathbb{Z}$, and $t^n = n\Delta t$ for $n = 0, \dots, M+1$.

For a given coefficient k we denote by ξ_i, $i = 1, \dots, N$, its discontinuities and by $D_i = (\xi_i, \xi_{i+1})$, $i = 0, \dots, N$, the subdomains where k is constant. Here we have used the notation $\xi_0 = -\infty$ and $\xi_{N+1} = +\infty$. Furthermore, we will write

$$f^{(i)} = f(k(x), \cdot), \qquad \text{for } x \in D_i,\ i = 0, \dots, N.$$

In the following, we will assume that the grid is aligned in such a way that all discontinuities of k lie on cell interfaces, i.e., $\xi_i = x_{P_i - 1/2}$ for some integers P_i, $i = 1, \dots, N$. In general, this can be achieved by considering a globally nonuniform grid that is uniform on each D_i and taking $\Delta x = \max_{i=0,\dots,N} \Delta x_i$ where Δx_i is the grid discretization parameter in D_i.

The finite volume method we consider is the following [3]:

$$\begin{aligned} w_j^{n+1} &= w_j^n - \lambda \left(f^{(i)}(w_j^n) - f^{(i)}(w_{j-1}^n) \right), \quad n \geq 0,\ P_i < j < P_{i+1},\ 0 \leq i \leq N, \\ w_{P_i}^{n+1} &= \left(f^{(i)} \right)^{-1} \left(f^{(i-1)} \left(w_{P_i - 1}^{n+1} \right) \right), \quad n \geq 0,\ 0 < i \leq N, \\ w_j^0 &= \frac{1}{\Delta x} \int_{\mathcal{C}_j} \bar{w}(x) \,\mathrm{d}x, \quad j \in \mathbb{Z}, \end{aligned} \tag{4.11}$$

where $P_0 = -\infty$, $P_{N+1} = +\infty$, and $\lambda = \Delta t / \Delta x$. We assume that the grid discretization parameters satisfy the following CFL condition:

$$\lambda \max_i \max_u \left(f^{(i)} \right)' (w) \leq 1. \tag{4.12}$$

Note that the definition of $w_{P_i}^{n+1}$ in (4.11) represents a discrete version of the Rankine–Hugoniot condition which in the setting of conservation laws with discontinuous flux holds across discontinuities of k. Here, we use the ghost cells $\mathcal{C}_{P_i}$, $i = 1, \dots, N$ to explicitly enforce the Rankine–Hugoniot condition on the discrete level.

We define the numerical solution operator $S_t^{\Delta x}$ by

$$(S_t^{\Delta x}(\bar{w}, k, f))(x) = w_j^n, \qquad (x,t) \in \mathcal{C}_j \times [t^n, t^{n+1}).$$

The following lemma shows that the finite volume method is stable in L^∞ and L^1.

Lemma 4.11 ([2, Lem. 5.1])**.** *Let f, k, and $\bar{w}$ satisfy the assumptions of Theorem 4.9. If the numerical scheme* (4.11) *satisfies the CFL condition* (4.12) *we have the following stability estimates:*

$$\left\| S_t^{\Delta x}(\bar{w}, k, f) \right\|_{\mathrm{L}^\infty(\mathbb{R})} \leq \frac{C_f}{\alpha} \left\| \bar{w} \right\|_{\mathrm{L}^\infty(\mathbb{R})}$$

and

$$\left\| S_t^{\Delta x}(\bar{w}, k, f) \right\|_{\mathrm{L}^1(\mathbb{R})} \leq \left\| \bar{w} \right\|_{\mathrm{L}^1(\mathbb{R})} + C \mathrm{TV}(\bar{w}) \Delta x.$$

Theorem 4.12 ([3, Thm. 5.1])**.** *Let f, k, and $\bar{w}$ satisfy the assumptions of Theorem 4.9. Let $S_t(\bar{w}, k, f)$ denote the corresponding adapted entropy solution of* (4.9) *and $S_t^{\Delta x}(\bar{w}, k, f)$ the numerical approximation given by* (4.11)*. Then we have the following convergence rate estimate*

$$\left\| S_t(\bar{w}, k, f) - S_t^{\Delta x}(\bar{w}, k, f) \right\|_{\mathrm{L}^1(\mathbb{R})} \leq C \Delta x^{1/2} \tag{4.13}$$

for all $0 \leq t \leq T$. Like in (4.6)*, the constant C depends polynomially on $\mathrm{TV}(\bar{w}), \|f\|_{\mathrm{Lip}}$ and in this case the number of discontinuities of k.*

4.2.3 Bayesian inverse problems for scalar conservation laws with discontinuous flux

We consider a given, fixed set of points $\xi_1 < \xi_2 < \ldots < \xi_{N-1}$ for $N \gg 1$ representing the possible points of discontinuity of the coefficient k. We identify the space

$$V \coloneqq \{k \in \mathrm{L}^\infty(\mathbb{R}) \mid k \text{ is piecewise constant with discontinuities among the points } \xi_1, \ldots, \xi_{N-1}\}$$

(as a subspace of $\mathrm{L}^\infty(\mathbb{R})$) with $(\mathbb{R}^N, \|\cdot\|_\infty)$ by associating $k \in V$ with the vector $(k_i)_{i=1}^N \in \mathbb{R}^N$ representing the values of k between neighboring points ξ_i and ξ_{i+1}. We then consider the Bayesian inverse problem with $X = (\mathrm{L}^\infty \cap \mathrm{BV})(\mathbb{R}) \times \mathbb{R}^N$ and define the observation operator $\mathcal{G}\colon X \to \mathbb{R}^m$ by

$$(\mathcal{G}(\bar{w}, k))_j = \int_0^T \int_{\mathbb{R}} \psi_j(x,t) g_j((S_t(\bar{w}, k))(x)) \,\mathrm{d}x \,\mathrm{d}t, \qquad j = 1, \ldots, m, \tag{4.14}$$

for $\psi_j \in \mathrm{L}^1(\mathbb{R}^d \times (0,T)) \cap \mathrm{L}^1(0,T;\mathrm{L}^\infty(\mathbb{R}^d))$ and $g_j \in \mathcal{C}^1(\mathbb{R};\mathbb{R}^d)$ with $\|g_j\|_{\mathcal{C}^1(\mathbb{R};\mathbb{R}^d)} < \infty$. Note that here we keep the flux f fixed since the assumption $f_w \geq \alpha > 0$ is incompatible with a Banach space setting. As before, we use the finite volume method (4.11) to define an approximation to $\mathcal{G}$ in the following way:

$$\big(\mathcal{G}^{\Delta x}(\bar{w}, k)\big)_j = \int_0^T \int_{\mathbb{R}} \psi_j(x,t) g_j\big((S_t^{\Delta x}(\bar{w}, k))(x)\big) \,\mathrm{d}x \,\mathrm{d}t, \qquad j = 1, \ldots, m. \tag{4.15}$$

Lemma 4.13. *The observation operator defined by* (4.14) *satisfies Assumption 2.2 and the approximation $\mathcal{G}^{\Delta x}$ satisfies* (3.5) *with $\Psi(\Delta x^{-1}) = \sqrt{\Delta x}$.*

Proof. In light of the stability estimate (4.10) and the convergence rate (4.13), the proof can be carried out in the same way, *mutatis mutandis*, as the proofs of Lemmas 4.5 and 4.6. ☐

5 Numerical experiments

In this section, we illustrate our theoretical results by presenting a series of numerical experiments. We employ a Metropolis–Hastings method to generate a Markov chain which samples from the posterior μ^y. Such methods require a proposal kernel and here we choose the following standard random walk (see [22, 5]):

- Set $n = 0$ and pick $u^{(0)}$.
- Propose $v^{(n)} = u^{(n)} + \beta\xi^{(n)}$ where $\xi^{(n)} \sim \mathcal{N}(0, \mathcal{C})$.
- Set $u^{(n+1)} = v^{(n)}$ with probability $a(u^{(n)}, v^{(n)})$.
- Set $u^{(n+1)} = u^{(n)}$ otherwise.
- $n \to n+1$.

The underlying acceptance probability is defined as

$$a(u,v) = \min(1, \exp(I(u) - I(v)))$$

where

$$I(u) = \Phi(u) + \frac{1}{2}\left\|\mathcal{C}^{-1/2}u\right\|_X^2.$$

If we generate $\xi^{(n)}$ and the uniform random variable used in the accept-reject step independently of each other for each n and independently of their values for different n then this construction gives rise to a Markov chain $(u^{(n)})_{n=0}^\infty$ which is distributed according to μ^y given by (1.2) [22].

The algorithm has three scalar hyperparameters which need to be specified. First, the stepsize β which controls the size of the move, second the burn-in b, i.e., the number of samples which are

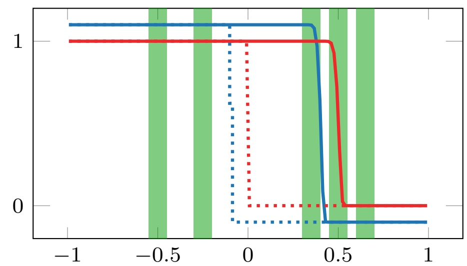


Figure 1: *Experiment 1:* Initial data (dotted lines) and numerical solutions (solid lines) for $(\delta_1^p, \delta_2^p, \sigma_0^p) = (0.1, -0.1, -0.1)$ (blue) corresponding to the prior mean and $(\delta_1^*, \delta_2^*, \sigma_0^*) = (0, 0, 0)$ (red) corresponding to the ground truth. The numerical solutions are calculated using the Rusanov scheme and the grid discretization parameter $\Delta x = 2/128$.

discarded in order to minimize the contribution of the initial value $u^{(0)}$, and the sample interval τ which is the number of states which are discarded between two observations.

The best choices of hyperparameters, corresponding to short burn-in and smaller step-size in the steady-state can be achieved by letting β vary with the step-count, i.e., $\beta = \beta(k)$. We chose a piecewise linear function for β, where in the beginning the steps are large and decrease linearly until a certain number of steps, after which it stays constant. Ultimately, the step-size is problem-dependent and has to be adjusted for each problem, for instance by a grid search.

5.1 Inverse problem for the shock location and amplitude in a Riemann problem for Burgers' equation

In our first numerical experiment we consider Burgers' equation

$$\begin{aligned} w_t + \left(\frac{w^2}{2}\right)_x &= 0, \quad (x,t) \in (-1,1) \times (0,T), \\ w(x,0) &= \bar{w}(x), \quad x \in (-1,1), \end{aligned}$$

with outflow boundary conditions. Given numerical solutions at a specified time we want to infer the initial datum which we assume is of the form

$$\bar{w}^{(\delta_1,\delta_2,\sigma_0)}(x) = \begin{cases} 1 + \delta_1, & x < \sigma_0, \\ \delta_2, & x > \sigma_0, \end{cases}$$

parameterized by $u = (\delta_1, \delta_2, \sigma_0) \in \mathbb{R}^3$. In order to infer the parameters $(\delta_1, \delta_2, \sigma_0)$ by observing (an approximation of) the solution w at time $T = 1$ we define the observation operator

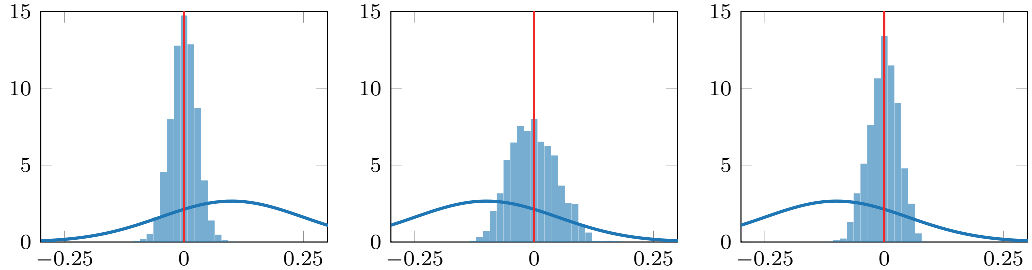
$$(\mathcal{G}(\delta_1, \delta_2, \sigma_0)))_j = 10 \int_{x_j - 0.05}^{x_j + 0.05} S_T^{\Delta x} \left(\bar{w}^{(\delta_1,\delta_2,\sigma_0)}\right) \mathrm{d}x, \qquad j = 1, \ldots, 5,$$

where $S_T^{\Delta x}$ denotes the numerical solution operator and the measurement points are $(x_j)_{j=1}^5 = (-0.5, -0.25, 0.35, 0.5, 0.65)$.

We consider observational noise $\eta \sim \mathcal{N}(0, \gamma^2 \mathcal{I}_5)$ with $\gamma = 0.05$ and prior $\mu_0 \sim \mathcal{N}(u^p, \mathcal{C})$ with mean $u^p = (\delta_1^p, \delta_2^p, \sigma_0^p) = (0.1, -0.1, -0.1)$ and covariance matrix $\mathcal{C} = \varphi^2 \mathcal{I}_3$, $\varphi = 0.15$. The ground truth we want to recover is $u^* = (\delta_1^*, \delta_2^*, \sigma_0^*) = (0, 0, 0)$. Figure 1 shows the initial data corresponding to the prior mean and the ground truth as well as corresponding numerical solutions computed at time $T = 1$. The measurement intervals used in the observation operator are highlighted in green.

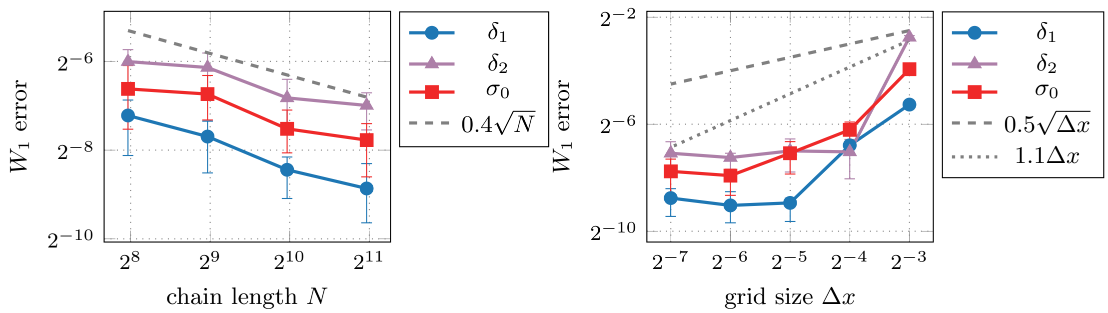
As for the step size used in the Metropolis–Hastings method, we chose

$$\beta(k) = \begin{cases} \beta_0 - \frac{\beta_0 - \beta_1}{k_b} k, & k \le k_b, \\ \beta_1, & k > k_b, \end{cases} \tag{5.1}$$



(a) δ_1 (b) δ_2 (c) σ_0

Figure 2: *Experiment 1:* Histograms corresponding to the Metropolis-Hastings approximation of the posterior using a chain length of 2500 and $\Delta x = 2/128$. The ground truth and the prior are shown in red and blue respectively.



(a) W_1 error against chain length. (b) W_1 error against grid size.

Figure 3: *Experiment 1:* W_1 error as a function of the chain length with a fixed grid discretization parameter $\Delta x = 2/128$ (left) and as a function of the grid discretization parameter Δx for a fixed chain length ($N = 2500$) (right)

where $(\beta_0, \beta_1, k_b) = (0.05, 0.001, 250)$. This allowed us to use $b = 500$ and $\tau = 20$, i.e., after discarding the first 500 states use every 20th state to approximate the posterior.

Figure 2 shows the histograms of the approximated posterior computed by the Metropolis–Hastings method with a chain length of 2500 and using $\Delta x = 2/128$ for the underlying finite volume method for the forward problem. The resulting posteriors all peak at the ground truth parameter values. The posteriors indicate the uncertainties inherent in estimating these parameters. The posterior of δ_2 has the largest spread indicating comparatively slightly larger uncertainty in this parameter. This appears to be a consequence of the placement of the measurement intervals since only the rightmost measurement interval around the point $x_5 = 0.75$ contributes towards inferring the parameter δ_2. The mean of the approximated posterior is $u_{\text{mean}} = (\delta_1, \delta_2, \sigma_0) \approx (-0.0004, -0.0010, -0.0012)$ and the *maximum a posteriori* (MAP) estimator is $u_{\text{MAP}} = (\delta_1, \delta_2, \sigma_0) \approx (0.0136 - 0.0195, -0.0037)$ both very close to the ground truth $u^* = (0, 0, 0)$.

In Figure 3 we investigate the convergence of the approximated posterior measured in the 1-Wasserstein distance with respect to the length of the chain as well as with respect to the grid discretization parameter Δx used in the finite volume method of the forward problem. Specifically, in Figure 3 (a) we consider chain lengths $250, 500, 1000$, and 2000 while keeping the grid discretization parameter $\Delta x = 2/128$ constant. On the other hand, in Figure 3 (b) we use $16, 32, 64, 128$, and 256 cells in the domain $(-1, 1)$ while keeping the chain length $N = 2500$ constant. We compute each Wasserstein error shown in Figure 3 as

$$\frac{1}{K}\sum_{k=1}^{K} W_1\left(U_k^{N,\Delta x}, U_{\mathrm{Ref}}^{N^*,\Delta x^*}\right)$$

where $\left(U_k^{N,\Delta x}\right)_{k=1}^{K}$ is an ensemble of K Markov chains all of length N and using the same grid discretization parameter Δx and $U_{\mathrm{Ref}}^{N^*,\Delta x^*}$ is a reference solution. In the case of convergence with respect to the chain length we computed the reference solution $U_{\mathrm{Ref}}^{N^*,\Delta x^*}$ as an average of an ensemble of K Markov chains using $N^* = 4000$ and $\Delta x^* = 2/128$. For the convergence with respect to Δx we computed $U_{\mathrm{Ref}}^{N^*,\Delta x^*}$ again as an average of K Markov chains using $N^* = 2500$ and $\Delta x^* = 2/512$. In both experiments we used an ensembles of size $K = 30$.

Figure 3 shows that both errors decrease at approximately the expected rate (for the grid size the expected rate is $\sqrt{\Delta x}$, cf. Lemma 4.6 and Corollary 3.2). It is clear from Figure 3 (right) that there is a saturation of convergence with respect to some finer grid sizes. This can be explained by the fact that the sampling error with respect to the chain length (see Figure 3 (left)) has already been reached and dominates the discretization error due to the numerical method.

5.2 Inverse problem for the transport speed and jump amplitude for a Riemann problem with flux discontinuity

In our second experiment we consider the conservation law with discontinuous flux

$$w_t + (k(x)f(w) + (1 - k(x))g(w))_x = 0 \tag{5.2}$$

where k is the Heaviside function and g and f are the Transport respectively Burgers flux, i.e., $g^{(a)}(w) = aw$, and $f(w) = \frac{w^2}{2}$. Equation (5.2) corresponds to switching from the Transport equation to Burgers equation across the flux interface at $x = 0$. We use the initial datum

$$\bar{w}^{(\delta)}(x) = \begin{cases} 0.5 + \delta, & x < -0.5, \\ 2, & x > -0.5, \end{cases}$$

on the domain $(-1, 1)$ with outflow boundary conditions and our aim is to infer the left state of the Riemann initial datum, i.e., δ, as well as the transport speed a by observing the (numerical) solution at time $T = 1$. Specifically, we consider the observation operator

$$\big(\mathcal{G}(\delta, a)\big)_j = 10\int_{x_j-0.075}^{x_j+0.075} S_T^{\Delta x}\big(\bar{w}^{(\delta)}, g^{(a)}\big)\,\mathrm{d}x, \qquad j = 1, \dots, 6,$$

where $S_T^{\Delta x}$ is the numerical solution operator defined in (4.11) and the measurement points are $(x_j)_{j=1}^{6} = (-0.5, 0.1, 0.3, 0.5, 0.7, 0.9)$. We consider observational noise $\eta \sim \mathcal{N}(0, \gamma^2\mathcal{I}_6)$ with $\gamma = 0.05$ and prior $\mu_0 \sim (u^p, \mathcal{C})$ with mean $u^p = (\delta^p, a^p) = (0.1, 0.9)$ and covariance matrix $\mathcal{C} = \varphi^2\mathcal{I}_2$, $\varphi = 0.15$. The ground truth we want to recover is $(\delta^*, a^*) = (0, 1)$. Figure 4 illustrates the initial data and numerical solutions corresponding to the prior mean and ground truth parameters.

Figure 5 shows the histograms of the approximated posterior. Here, we used a chain length of 2500 and $\Delta x = 128$ and $\lambda = 0.4$ in the finite volume approximation (4.11). The mean of the approximated posterior is $u_{\mathrm{mean}} = (\delta, a) \approx (-0.00732, 1.00119)$ and the MAP estimator us $u_{\mathrm{MAP}} = (\delta, a) \approx (-0.00414, 1.00076)$.

Figure 6 again illustrates the convergence of the approximated posterior measured in the 1-Wasserstein distance with respect to the length of the chain and with respect to the grid discretization parameter Δx. We see that the observed order of convergence with respect to Δx in this experiment is strictly higher than the order $\sqrt{\Delta x}$ which our theory guarantees. This observation is in line with the fact that the experimental order of convergence for finite volume methods is typically closer to one.

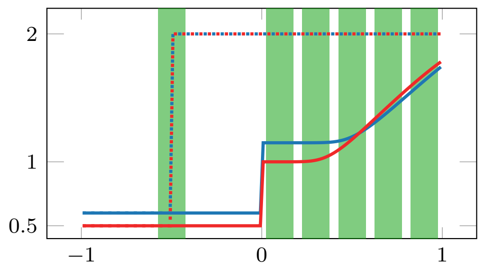
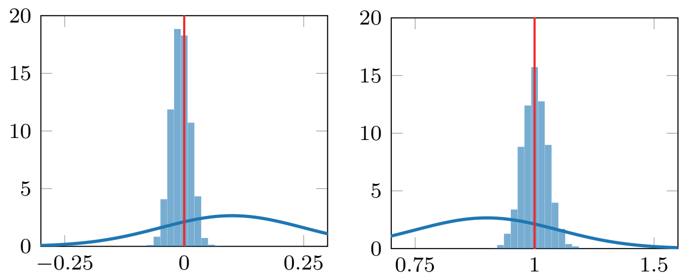


Figure 4: *Experiment 2:* Initial data (dotted lines) and numerical solutions (solid lines) for $\delta = 0.1$ and $a = 0.9$ (blue) corresponding to the prior mean and $\delta = 0$ and $a = 1.$ (red) corresponding to the ground truth. The numerical solutions are calculated using the scheme (4.11) with $\lambda = 0.4$ and grid discretization parameter $\Delta x = 2/128$.



(a) δ (b) a

Figure 5: *Experiment 2:* Histograms corresponding to the Metropolis-Hastings approximation of the posterior using a chain length of 2500 and $\Delta x = 2/128$. The ground truth and the prior are shown in red and blue respectively.

5.3 An inverse problem for systems of conservation laws.

While our theory does not cover systems of conservation laws, even in one space dimension, due to a lack of rigorous stability results in the literature, we demonstrate with the following numerical experiment that Bayesian inverse problems for systems of conservation laws, at least in one space dimension, might still be well-approximated with the MCMC type sampling algorithms presented here.

We consider the one-dimensional Euler equations

$$w_t + f(w)_x = 0$$

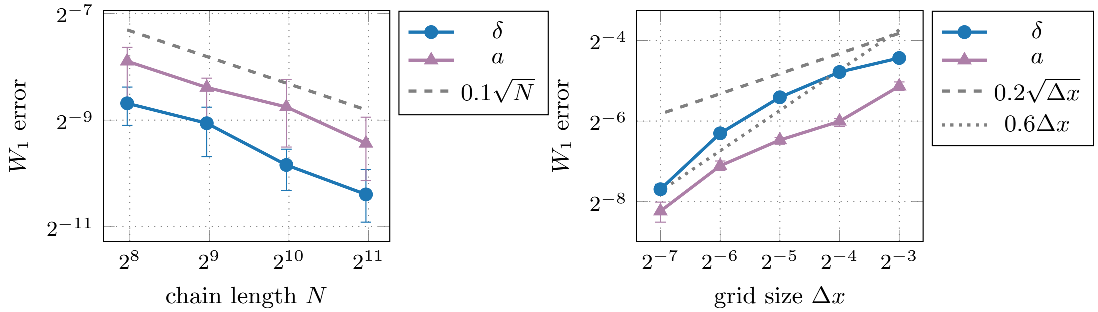
$$w = \begin{pmatrix} \rho \\ \rho v \\ E \end{pmatrix}, \qquad f(w) = \begin{pmatrix} \rho v \\ \rho v^2 + p \\ (E+p)v \end{pmatrix},$$

where the density ρ, velocity v and energy E are unknown and the pressure p and the energy are related by the following equation of state:

$$E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho v^2, \qquad \text{for } \gamma = 1.4.$$

We consider Sod's shock tube problem [20] on the domain $(0,1)$ with outflow boundary conditions and initial discontinuity at $x = 0.5$. We want to infer the initial datum $\bar{w}^{(\delta_L,\gamma_L,\beta_L,\delta_R,\gamma_R,\beta_R)}$ which we assume takes the left and right states

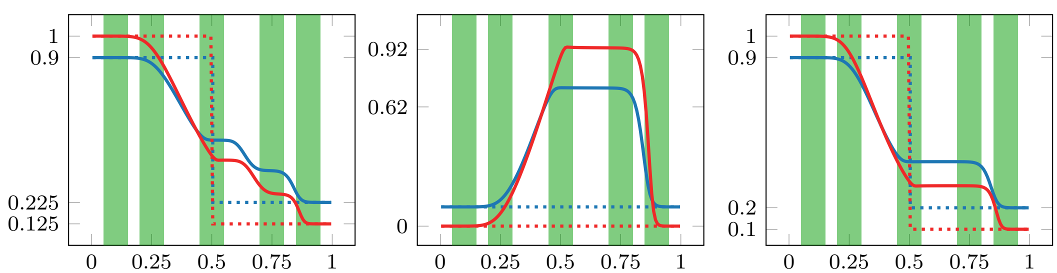
$$\begin{pmatrix} \rho_L \\ v_L \\ p_L \end{pmatrix} = \begin{pmatrix} 1+\delta_L \\ \gamma_L \\ 1+\beta_L \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} \rho_R \\ v_R \\ p_R \end{pmatrix} = \begin{pmatrix} 0.125+\delta_R \\ \gamma_R \\ 0.1+\beta_R \end{pmatrix}$$



(a) W_1 error against chain length.

(b) W_1 error against grid size.

Figure 6: *Experiment 2:* W_1 error as a function of the chain length with a fixed grid discretization parameter $\Delta x = 2/128$ (left) and as a function of the grid discretization parameter Δx for a fixed chain length ($N = 2500$) (right)



(a) ρ

(b) v

(c) p

Figure 7: *Experiment 3:* Initial data (dotted lines) and numerical solutions (solid lines) for $u^p = (\delta_L, \gamma_L, \beta_L, \delta_R, \gamma_R, \beta_R) = (-0.1, 0.1, -0.1, 0.1, 0, 1, 0.1)$ (blue) corresponding to the prior mean and $u^* = (\delta_L^*, \gamma_L^*, \beta_L^*, \delta_R^*, \gamma_R^*, \beta_R^*) = (0, 0, 0, 0, 0, 0)$ (red) corresponding to the ground truth. The numerical solutions are calculated using the HLLC scheme and grid discretization parameter $\Delta x = 1/128$.

to the left respectively to the right of the initial discontinuity $x = 0.5$. To that end we consider the observation operator

$$(\mathcal{G}(\delta_L, \gamma_L, \beta_L, \delta_R, \gamma_R, \beta_R))_j = 10 \int_{x_j - 0.05}^{x_j + 0.05} S_T^{\Delta x} \left(\bar{w}^{(\delta_L, \gamma_L, \beta_L, \delta_R, \gamma_R, \beta_R)} \right) \mathrm{d}x, \qquad j = 1, \ldots, 5,$$

at time $T = 0.2$ and for the measurement points $(x_j)_{j=1}^5 = (0.1, 0.25, 0.5, 0.75, 0.9)$. Here $S_T^{\Delta x}$ is a numerical solution operator and in the subsequent experiment we will employ the HLLC method.

We consider observational noise $\eta \sim \mathcal{N}(0, \gamma^2 \mathcal{I}_{15})$ with $\gamma = 0.05$ and prior $\mu_0 \sim \mathcal{N}(u^p, \mathcal{C})$ with mean $u^p = (\delta_L^p, \gamma_L^p, \beta_L^p, \delta_R^p, \gamma_R^p, \beta_R^p) = (-0.1, 0.1, -0.1, 0.1, 0, 1, 0.1)$ and covariance matrix $\mathcal{C} = \varphi^2 \mathcal{I}_6$ with $\varphi = 0.15$. The ground truth we want to recover is $u^* = (\delta_L^*, \gamma_L^*, \beta_L^*, \delta_R^*, \gamma_R^*, \beta_R^*) = (0, 0, 0, 0, 0, 0)$. Figure 7 shows the initial data corresponding to the prior mean and the ground truth as well as corresponding numerical solutions. We chose the constant step size $\beta = 0.0005$ and burn-in $b = 500$ and sample interval $\tau = 10$. The histograms of the approximated posteriors are shown in Figure 8 and we observe overall good approximation with the only possible exception of the right state of the velocity. Here, we used a chain length of 1500 and $\Delta x = 1/128$. The means of the approximated posterior are $u_{\text{mean}} \approx (0.0078, -0.0064, 0.0084, 0.02, 0.046, 0.012)$ and the MAP estimators are $u_{\text{MAP}} \approx (0.0092, 0.012, 0.012, 0.014, 0.013, 0.0011)$ both very close to zero.

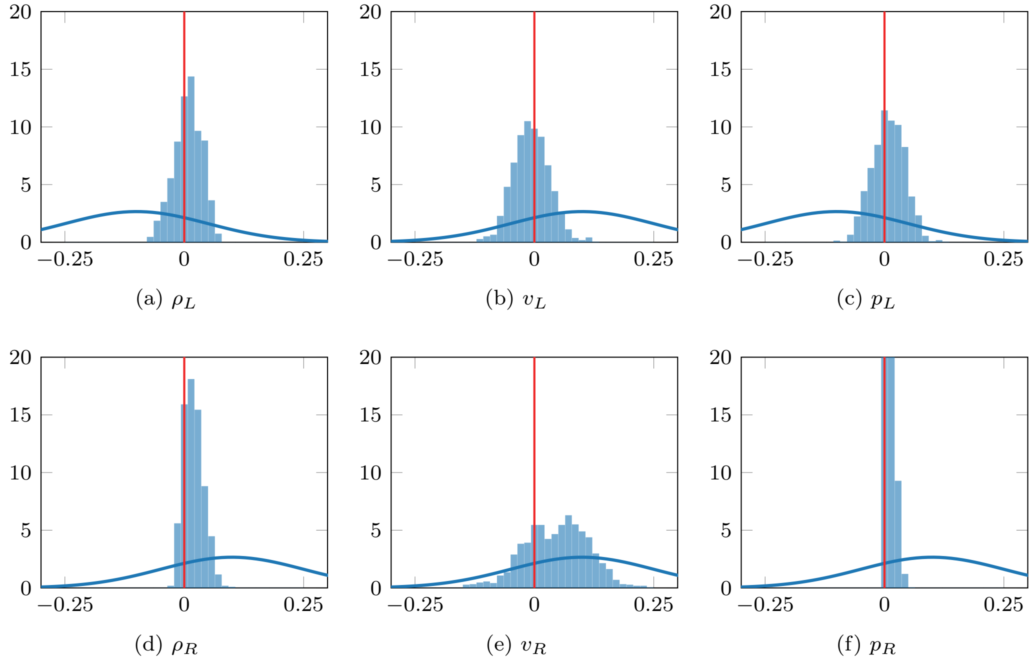


Figure 8: *Experiment 3:* Histograms corresponding to the Metropolis-Hastings approximation of the posterior using a chain length of 1500 and $\Delta x = 1/128$. The ground truth and the prior are shown in red and blue respectively.

6 Summary

We studied the well-posedness of the Bayesian inverse problem for scalar hyperbolic conservation laws in this paper. To this end, we started with the abstract framework for well-posedness that was formalized in [22]. In contrast to [22], we investigated Lipschitz continuity of the measurement to posterior map with respect to the Wasserstein metric. This allows us to more effectively control important statistical moments, such as the means of the posteriors. Moreover, the Lipschitz continuity of *approximate posteriors*, with respect to variations in the approximation parameter, was also derived, allowing us to establish convergence rates with respect to spatio-temporal numerical approximations of the underlying forward map.

These abstract results were verified for scalar conservation laws, in the context of a Bayesian inverse problem corresponding to inferring the initial datum and flux functions, from noisy measurements of the observables of entropy solutions. Moreover, we also demonstrated the well-posedness of the Bayesian inverse problems for conservation laws with a flux function, that is possibly discontinuous in the space variable. In both cases, explicit stability estimates were obtained for the variation of the posterior in the Wasserstein distance, with respect to measurement perturbations or approximations.

Finally, we illustrated the theoretical results with numerical experiments, where we verified the convergence rates for the posterior with respect to the spatio-temporal discretization. Our theory and experiments illustrated the fact that the Bayesian inverse problem is both well-posed and can be approximated quite well numerically, even for these nonlinear hyperbolic PDEs with discontinuous solutions.

Our focus in this paper was on scalar conservation laws as the forward map, in this case, is well-posed and is Lipschitz continuous with respect to the data and to approximations. Extending these results to hyperbolic systems of conservation laws is very challenging. In one space dimension, it is well known that entropy solutions exist and are unique, at least for initial data with small total variation. However, the lack of stability results, particularly with respect to fluxes, inhibits the direct application of our theory in this case. Nevertheless, we presented a numerical experiment to show that the Bayesian inverse problem is computable. However, for systems of conservation laws in several space dimensions, the forward map might not even be globally defined. The well-posedness of the Bayesian inverse problem for such *ill-posed* PDEs is discussed in the recent paper [13].

A Appendix

Theorem A.1 (Fernique Theorem [7, Thm. 2.7])**.** *If $\mu = \mathcal{N}(0, \mathcal{C})$ is a Gaussian measure on some Banach space X, so that $\mu(X) = 1$, then there exists $\alpha > 0$ such that*

$$\int_X \exp\left(\alpha \|x\|_X^2\right) \mu(\mathrm{d}x) < \infty.$$

The Fernique Theorem implies in particular that all moments of u under Gaussian measures μ are finite as can be seen in the following way: Since $\|u\|_X^p = \exp(p \ln \|u\|_X) \leq \exp(p \|u\|_X)$ we find

$$\begin{aligned}
\int_X \|u\|_X^p \,\mathrm{d}\mu(u) &\leq \int_X \exp(p \|u\|_X) \,\mathrm{d}\mu(u) \\
&= \int_X \left(\exp(p \|u\|_X) \chi_{\left\{\|u\|_X \geq \frac{p}{\alpha}\right\}} + \exp(p \|u\|_X) \chi_{\left\{\|u\|_X < \frac{p}{\alpha}\right\}}\right) \mathrm{d}\mu(u) \\
&\leq \int_X \exp\left(\alpha \|u\|_X^2\right) \mathrm{d}\mu(u) + \exp\left(\frac{p^2}{\alpha}\right) \mu\left(\left\{\|u\|_X < \frac{p}{\alpha}\right\}\right) \\
&< \infty.
\end{aligned}$$

References

[1] E. Audusse and B. Perthame, *Uniqueness for scalar conservation laws with discontinuous flux via adapted entropies*, Proc. Roy. Soc. Edinburgh Sect. A, 135 (2005), pp. 253–265.

[2] J. Badwaik, C. Klingenberg, N. H. Risebro, and A. M. Ruf, *Multilevel Monte Carlo finite volume methods for random conservation laws with discontinuous flux*, M2AN Math. Model. Numer. Anal., 55 (2021), pp. 1039–1065.

[3] J. Badwaik and A. M. Ruf, *Convergence rates of monotone schemes for conservation laws with discontinuous flux*, SIAM J. Numer. Anal., 58 (2020), pp. 607–629.

[4] P. Baiti and H. K. Jenssen, *Well-posedness for a class of 2×2 conservation laws with L^∞ data*, J. Differential Equations, 140 (1997), pp. 161 – 185.

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

[6] M. G. Crandall and A. Majda, *Monotone difference approximations for scalar conservation laws*, Math. Comp., 34 (1980), pp. 1–21.

[7] G. Da Prato and J. Zabczyk, *Stochastic equations in infinite dimensions*, Cambridge university press, 2014.

[8] C. M. Dafermos, *Hyperbolic conservation laws in continuum physics*, vol. 325, Springer, 2010.

[9] E. Godlewski and P.-A. Raviart, *Hyperbolic systems of conservation laws*, Ellipses, 1991.

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

[11] H. Holden and N. H. Risebro, *Front tracking for hyperbolic conservation laws*, vol. 152, Springer, 2015.

[12] N. Kuznetsov, *Accuracy of some approximate methods for computing the weak solutions of a first-order quasi-linear equation*, USSR Computational Mathematics and Mathematical Physics, 16 (1976), pp. 105–119.

[13] S. Lanthaler, S. Mishra, and F. Weber, *On the well-posedness of Bayesian inversion for PDEs with ill-posed forward problems*, arXiv preprint arXiv:2107.07593, (2021).

[14] J. Latz, *On the well-posedness of Bayesian inverse problems*, SIAM/ASA J. Uncertainty Quantification, 8 (2020), pp. 451–482.

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

[16] B. Piccoli and M. Tournus, *A general BV existence result for conservation laws with spatial heterogeneities*, SIAM J. Math. Anal., 50 (2018), pp. 2901–2927.

[17] A. M. Ruf, *Flux-stability for conservation laws with discontinuous flux and convergence rates of the front tracking method*, IMA J. Numer. Anal., (2021).

[18] A. M. Ruf, E. Sande, and S. Solem, *The optimal convergence rate of monotone schemes for conservation laws in the Wasserstein distance*, J. Sci. Comput., 80 (2019), pp. 1764–1776.

[19] F. Sabac, *The optimal convergence rate of monotone finite difference methods for hyperbolic conservation laws*, SIAM J. Numer. Anal., 34 (1997), pp. 2306–2318.

[20] G. A. Sod, *A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws*, J. Comput. Phys., 27 (1978), pp. 1–31.

[21] B. Sprungk, *On the local Lipschitz stability of Bayesian inverse problems*, Inverse Problems, 36 (2020), p. 055015.

[22] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.

[23] J. D. Towers, *An existence result for conservation laws having BV spatial flux heterogeneities - without concavity*, J. Differential Equations, 269 (2020), pp. 5754 – 5764.

[24] C. Villani, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.