NUMERICAL APPROXIMATION OF STATISTICAL SOLUTIONS OF SCALAR CONSERVATION LAWS

U. S. FJORDHOLM*, K. LYE†, AND S. MISHRA‡

Abstract. We propose efficient numerical algorithms for approximating statistical solutions of scalar conservation laws. The proposed algorithms combine finite volume spatio-temporal approximations with Monte Carlo and multi-level Monte Carlo discretizations of the probability space. Both sets of methods are proved to converge to the entropy statistical solution. We also prove that there is a considerable gain in efficiency resulting from the multi-level Monte Carlo method over the standard Monte Carlo method. Numerical experiments illustrating the ability of both methods to accurately compute multi-point statistical quantities of interest are also presented.

Key words. scalar conservation laws, burgulence, statistical solutions, finite volume methods, Monte Carlo, Wasserstein

AMS subject classifications. 35L65, 65M08, 65C05, 65C30

1. Introduction. Hyperbolic systems of conservation laws are a large class of nonlinear partial differential equations that arise in a wide variety of models in physics and engineering [10]. Prominent examples include the shallow water equations of oceanography, compressible Euler equations of aerodynamics and the magnetohydrodynamics (MHD) equations of plasma physics.

The simplest example for this class of PDEs is provided by the scalar conservation law

$$(1.1) \qquad \begin{aligned} u_t + \nabla_x \cdot f(u) &= 0 \\ u(x,0) &= \bar{u}(x) \end{aligned}$$

where $u : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}$ is the conserved variable and $f : \mathbb{R} \to \mathbb{R}^d$ is the flux function. A prototypical example of a scalar conservation law is the Burgers equation, i.e (1.1) with $d = 1$ and flux function $f(u) = \frac{1}{2}u^2$.

It is well known that discontinuities, such as shock waves, can arise in the solution of (1.1) even when the initial data $\bar{u}$ is smooth. Hence, one seeks solutions of (1.1) in the sense of distributions. These *weak* solutions are not necessarily unique and need to be augmented with additional admissibility criteria called *entropy conditions* to ensure uniqueness. *Entropy solutions* of the multi-dimensional scalar conservation law (1.1) exist as long as $\bar{u} \in L^1 \cap L^\infty(\mathbb{R}^d)$, and are unique and stable with respect to the initial data:

$$\left\| \mathcal{S}(t)\bar{u} - \mathcal{S}(t)\bar{v} \right\|_{L^1(\mathbb{R}^d)} \leqslant \|\bar{u} - \bar{v}\|_{L^1(\mathbb{R}^d)} \qquad \forall \, \bar{u}, \bar{v} \in L^1 \cap L^\infty(\mathbb{R}^d).$$

Here, $\mathcal{S}(t) : \bar{u} \mapsto u(t)$ denotes the data to solution operator of (1.1).

On the other hand, similar well-posedness results for systems of conservation laws are only available in one space dimension [6]. There are no global existence results for generic multi-dimensional systems of conservation laws. Moreover, it is now known

---

*Department of Mathematics, University of Oslo, Postboks 1053 Blindern, 0316 Oslo, Norway (ulriksf@math.uio.no).

†Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland (kjetil.lye@sam.math.ethz.ch).

‡Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland (siddhartha.mishra@sam.math.ethz.ch).

that entropy solutions for (some examples) of multi-dimensional systems may not be unique [11, 9].

Given this paucity of well-posedness results for systems of conservation laws, it is natural to seek alternative solution frameworks. A possible solution paradigm is that of entropy measure valued solutions [12], where the sought for solution is not necessarily an integrable function but rather a Young measure, i.e. a space-time parametrized probability measure. Although entropy measure-valued solutions exist globally (even for multi-dimensional systems) and can be approximated with Monte Carlo type ensemble averaging numerical algorithms [13], these solutions are not unique, even for scalar conservation laws (1.1), in particular in the case when the initial Young measure is not concentrated on a single integrable function. The reason behind this non-uniqueness is the lack of information on spatial correlations (conditional probabilities) that is intrinsic to the notion of Young measures.

In a recent paper [14], the authors proposed another solution concept, namely *statistical solutions*, to supplement measure-valued solutions with information about multi-point spatial correlations. Statistical solutions are time-parametrized probability measures on $L^p(D)$, where $D \subset \mathbb{R}^d$ is the spatial domain and $1 \leqslant p < \infty$. It was proved in [14] that these probability measures on spaces of p-integrable functions are identified with a hierarchy of Young measures, termed as *correlation measures*. Consequently, the time-evolution of statistical solutions is completely determined in terms of an infinite family of nonlinear PDEs for a class of moments of the corresponding correlation measures. It was shown in [14] that *entropy statistical solutions* of the scalar conservation law (1.1) always exist, are unique and stable with respect to initial data. This is in contrast to earlier notions of statistical solutions for Burgers' equation that are defined as probability measures on distributions [8, 7, 3]. See also [29] for a different notion of statistical solutions for the incompressible Navier–Stokes equations.

Statistical solutions can also be considered as a framework for uncertainty quantification (UQ), i.e. the modeling, analysis and efficient numerical approximation of uncertainty in the solutions of PDEs, given uncertainties in their inputs such as the initial and boundary data, fluxes, coefficients etc. [4]. For instance, the initial datum $\bar{u}$ in (1.1) is a measured (observed) quantity, and these measurements are not exact. Often, one needs to model the underlying input uncertainty and the resulting solution uncertainty statistically. It is common practice to model both input and solution uncertainty in terms of random fields (often characterized by a large but finite number of parameters). A popular notion of uncertainty modeling for conservation laws is that of *random entropy solutions* introduced in [27, 28]. However, random entropy solutions rely on a particular parametrization of the underlying uncertainty. On the other hand, the framework of statistical solutions offers a parametrization-independent modeling of uncertainty where only the law of random fields, i.e. a probability measure on the underlying function space, is described and evolved, cf. [14, 1].

The study of scalar conservation laws with random initial conditions also plays an important role in turbulence research [2]. The results on *burgulence* can also be reinterpreted within the framework of statistical solutions of (1.1).

Given the above discussion, we focus on the efficient numerical approximation of statistical solutions for scalar conservation laws (1.1) in this article. We will approximate the spatio-temporal domain with standard finite volume/finite difference numerical schemes. The probability space is discretized with a Monte Carlo method. We prove that that this finite-volume Monte Carlo method converges, in an appropriate topology, to the entropy statistical solution of (1.1) as the mesh is refined and

the number of Monte Carlo samples is increased. A complexity analysis demonstrates that this procedure can be rather expensive on account of the slow convergence (in terms of samples) of the Monte Carlo method.

In order to improve computational efficiency, we propose and analyze a multi-level Monte Carlo (MLMC) version of our numerical method. MLMC methods were first proposed in [20, 17] and were applied to UQ for conservation laws in [27]. We adapt the MLMC procedure to our setup and prove that the MLMC method also converges to the entropy statistical solution. Moreover, we show that the MLMC method is significantly more efficient than its Monte Carlo counterpart. We illustrate the performance of both set of methods on a suite of numerical test problems.

The rest of the article is organized as follows. We define statistical solutions of (1.1) in Section 2. Finite volume schemes for approximating the underlying deterministic problem are described in Section 3. Monte Carlo and multi-level Monte Carlo methods are presented in Sections 4 and 5, respectively. We present a set of numerical experiments illustrating our methods in Section 6.

2. Statistical solutions of conservation laws. In this section we briefly describe the framework of statistical solutions introduced in [14]. As mentioned above, we are interested in the situation where instead of an initial data $\bar{u} \in L^1 \cap L^\infty(\mathbb{R}^d)$ for (1.1), we are given some $\bar{\mu} \in \mathcal{P}(L^1(\mathbb{R}^d))$ with $\mathrm{supp}\,\bar{\mu} \subset L^\infty(\mathbb{R}^d)$, that is, a probability distribution over different initial data $\bar{u} \in L^1 \cap L^\infty(\mathbb{R}^d)$. For ease of notation we define $\mathcal{P}_0(\mathbb{R}^d)$ to be the set of admissible initial data:

$$(2.1) \qquad \mathcal{P}_0(\mathbb{R}^d) := \big\{ \bar{\mu} \in \mathcal{P}(L^1(\mathbb{R}^d)) \mid \mathrm{supp}\,\mu \text{ is a bounded subset of } L^\infty(\mathbb{R}^d) \big\}.$$

A statistical solution of this initial value problem is a map $t \mapsto \mu_t \in \mathcal{P}_0(\mathbb{R}^d)$ which satisfies the PDE (1.1) in a certain sense. In [14], the authors showed that any probability measure $\mu \in \mathcal{P}(L^1(\mathbb{R}^d))$ can be described equivalently as a *correlation measure*—a hierarchy $\boldsymbol{\nu} = (\nu^1, \nu^2, \dots)$ in which each element ν^k provides the joint probability distribution $\nu^k_{x_1,\dots,x_k}$ of the solution values $u(x_1), \dots, u(x_k)$ at any choice of spatial points $x_1, \dots, x_k \in \mathbb{R}^d$. The evolution equation for μ_t is most naturally described in terms of its corresponding correlation measure, yielding an infinite hierarchy of evolution equations. In particular, the equation for the one-point distribution ν^1_x coincides with that of *measure-valued solutions*, cf. DiPerna [12]. Most statistical observables of a fluid flow can be directly expressed in terms of correlation marginals. For instance, all one-point statistics such as the mean flow can be expressed in terms of ν^1, while the *structure functions* of turbulence theory are easily expressed in terms of ν^2 (see Section 6).

DEFINITION 2.1. *Let $d, N \in \mathbb{N}$, let $q \in [1, \infty)$, let $D \subset \mathbb{R}^d$ be an open set (the "spatial domain") and (for notational convenience) denote $\mathcal{U} = \mathbb{R}^N$ ("phase space"). A correlation measure from D to $\mathcal{U}$ is a collection $\boldsymbol{\nu} = (\nu^1, \nu^2, \dots)$ of maps satisfying for every $k \in \mathbb{N}$:*
(i) ν^k is a Young measure from D^k to $\mathcal{U}^k$.
(ii) Symmetry: if σ is a permutation of $\{1, \dots, k\}$ and $f \in C_0(\mathcal{U}^k)$ then $\langle \nu^k_{\sigma(x)}, f(\sigma(\xi)) \rangle = \langle \nu^k_x, f(\xi) \rangle$ for a.e. $x \in D^k$.
(iii) Consistency: If $f \in C_b(\mathcal{U}^k)$ is of the form $f(\xi_1, \dots, \xi_k) = g(\xi_1, \dots, \xi_{k-1})$ for some $g \in C_0(\mathcal{U}^{k-1})$, then $\langle \nu^k_{x_1,\dots,x_k}, f \rangle = \langle \nu^{k-1}_{x_1,\dots,x_{k-1}}, g \rangle$ for almost every $(x_1, \dots, x_k) \in D^k$.

(iv) L^q integrability:

$$\int_D \langle \nu_x^1, |\xi|^q \rangle \, dx < \infty. \tag{2.2}$$

(v) Diagonal continuity (DC): $\lim_{\varepsilon \to 0} d_\varepsilon^q(\nu^2) = 0$, *where*

$$d_\varepsilon^q(\nu^2) := \left(\int_D \fint_{B_\varepsilon(x)} \langle \nu_{x,y}^2, |\xi_1 - \xi_2|^q \rangle \, dy dx \right)^{1/q}. \tag{2.3}$$

(Here, $\fint_B = \frac{1}{|B|} \int_B$ is the average over B.)

The sense in which correlation measures and probability measures on L^p are equivalent is made more precise in the following definition:

DEFINITION 2.2. *A probability measure $\mu \in \mathcal{P}(L^q(D; \mathcal{U}))$ is said to be* dual *to a correlation measure $\boldsymbol{\nu}$ from D to $\mathcal{U}$ provided*

$$\int_{D^k} \langle \nu_x^k, g(x, \cdot) \rangle \, dx = \int_{L^2} \int_{D^k} g(x, v(x_1), \ldots, v(x_k)) \, dx d\mu(v) \tag{2.4}$$

for every $k \in \mathbb{N}$ and for every Caratheodory function $g : D^k \to C_b(\mathcal{U}^k)$.

It was shown in [14] that every probability measure $\mu \in \mathcal{P}(L^q(D; \mathcal{U}))$ is dual to a unique correlation measure $\boldsymbol{\nu}$, and *vice versa*. Using this duality, we can now state the definition of statistical solutions.

DEFINITION 2.3. *We say that a weak*-measurable map $t \mapsto \mu_t \in \mathcal{P}_0(\mathbb{R}^d)$ with corresponding spatial correlation measures $\boldsymbol{\nu}_t = (\nu_t^k)_{k \in \mathbb{N}}$, is a statistical solution of* (1.1) *with initial data $\bar{\mu} \in \mathcal{P}_0(\mathbb{R}^d)$ if*

$$\begin{aligned}
\int_{\mathbb{R}_+} \int_{(\mathbb{R}^d)^k} \langle \nu_{t,x}^k, \xi^1 \cdots \xi^k \rangle \partial_t \phi + \sum_{i=1}^k \langle \nu_{t,x}^k, \xi^1 \cdots f(\xi^i) \cdots \xi^k \rangle \nabla_x \phi \, dx \, dt \\
+ \int_{(\mathbb{R}^d)^k} \langle \bar{\nu}_x^k, \xi^1 \cdots \xi^k \rangle \phi|_{t=0} \, dx = 0,
\end{aligned} \tag{2.5}$$

for all $\phi \in C_c^\infty\left(\left(\mathbb{R}^d\right)^k \times \mathbb{R}_+ \right)$ and all $k \in \mathbb{N}$. (See (2.1) *for the definition of $\mathcal{P}_0(\mathbb{R}^d)$.)*

We define the *canonical statistical solution* as $\mu_t = \mathcal{S}(t) \# \bar{\mu}$, that is, the measure satisfying

$$\int_{L^1} G(u) \, d\mu_t(u) = \int_{L^1} G(\mathcal{S}(t)\bar{u}) \, d\bar{\mu}(\bar{u})$$

for every $G \in C_b(L^1(\mathbb{R}^d))$. It is straightforward to show that this indeed constitutes a statistical solution of (1.1) with initial data $\bar{\mu}$. Moreover, in [14] it was shown that under an additional entropy condition on the statistical solution, any two *entropy statistical solutions* (see Section 4.2 of [14] for a precise definition) μ_t, ρ_t satisfy

$$W_1(\mu_t, \rho_t) \leqslant W_1(\bar{\mu}, \bar{\rho}),$$

where W_1 is the 1-Wasserstein distance on the set of probability measures $\mathcal{P}(L^1(\mathbb{R}^d))$. In particular, the entropy statistical solution is unique and coincides with the canonical statistical solution. We refer to [14] for further details.

3. Numerical approximation of statistical solutions. In this section we introduce numerical approximations of statistical solutions, which are based on standard finite volume methods (FVM) for (1.1). In Section 3.1 we give a short description of FVM for scalar conservation laws, and in Section 3.2 we prove that FVM applied to statistical ensembles converge as the mesh is refined. We postpone the discretization of the probability space to Section 4.

3.1. Finite volume methods for conservation laws. This section briefly describes the conventional approach of numerically approximating conservation laws through finite volume and finite difference methods. For a complete review, one can consult [24].

We discretize the computational spatial domain as a collection of cells

$$\left\{ (x^1_{i_1-1/2}, x^1_{i_1+1/2}) \times \cdots \times (x^d_{i^d-1/2}, x^d_{i^d+1/2}) \right\}_{(i^1,\ldots,i^d)} \subset \mathbb{R}^d,$$

with corresponding cell midpoints

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

For simplicity, we assume that our mesh is equidistant, that is,

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

for some $\Delta > 0$. We will describe the semi-discrete case. For each cell, marked by $(i^1, \ldots, i^d)$, we let $u^\Delta_{i^1,\ldots,i^d}(t)$ denote the averaged value in the cell at time $t \geqslant 0$.

We use the following semi-discrete formulation:

(3.1)
$$\frac{\mathrm{d}}{\mathrm{d}t} u^\Delta_{i^1,\ldots,i^d}(t) + \sum_{k=1}^{d} \frac{1}{\Delta} \left(F^{k,\Delta}_{i^1,\ldots,i^k+1/2,\ldots,i^d}(t) - F^{k,\Delta}_{i^1,\ldots,i^k-1/2,\ldots,i^d}(t) \right) = 0$$

$$u^\Delta_{i^1,\ldots,i^d}(0) = \fint_{(x^1_{i_1-1/2}, x^1_{i_1+1/2}) \times \cdots \times (x^d_{i^d-1/2}, x^d_{i^d+1/2})} u_0(x)\, dx$$

where $F^{k,\Delta}$ is a *numerical flux function* in direction k. In a $(2p+1)$-*point scheme,* the numerical flux function $F^{k,\Delta}_{i^1,\ldots,i^k+1/2,\ldots,i^d}(t)$ can be written as a function of $u^\Delta_{i^1,\ldots,j^k,\ldots,i^k}(t)$ for $j^k = i^k - p + 1, \ldots, i^k + p$. We furthermore assume the numerical flux function is consistent with f and locally Lipschitz continuous, which amounts to requiring that for every bounded set $K \subset \mathbb{R}$, there exists a constant $C > 0$ such that for $k = 1, \ldots, d$,

$$\left| F^{k,\Delta}_{i^1,\ldots,j^k+1/2,\ldots,i^d} - f(u^\Delta_{i^1,\ldots,i^d}) \right| \leqslant C \sum_{j^k=i^1-p+1}^{i^k+p} \left| u^\Delta_{i^1,\ldots,i^d} - u^\Delta_{i^1,\ldots,j^k,\ldots,i^d} \right|,$$

whenever $u^\Delta_{i^1,\ldots,i^k-p+1,\ldots,i^d}, \ldots, u^\Delta_{i^1,\ldots,i^k+p,\ldots,i^d} \in K$.

We let $\mathcal{S}^\Delta(t) : L^1 \cap L^\infty(\mathbb{R}^d) \to L^1 \cap L^\infty(\mathbb{R}^d)$ be the spatially discrete numerical evolution operator corresponding to (3.1). Since $\mathcal{S}^\Delta$ is the composition of a projection from L^∞ onto piecewise constant functions and a continuous evolution under an ordinary differential equation, we see that $\mathcal{S}^\Delta$ is measurable.

The current form of (3.1) is continuous in time, and one needs to employ a time stepping method to discretize the ODE in time, usually through some strong stability preserving Runge–Kutta method [19].

3.2. Spatial convergence of statistical solutions for scalar equations.
We are interested in measuring convergence to the entropy statistical solution of (1.1)
in the 1-Wasserstein distance on $\mathcal{P}(L^1(\mathbb{R}^d))$, defined here as

$$(3.2) \qquad W_1(\mu_1, \mu_2) = \inf_{\pi \in \Pi(\mu_1, \mu_2)} \left(\int_{L^1(\mathbb{R}^d)^2} \|u_1 - u_2\|_{L^1(\mathbb{R}^d)} \, d\pi(u_1, u_2) \right)$$

where $\Pi(\mu_1, \mu_2)$ is the set of probability measures with first and second marginals μ_1
and μ_2, respectively (see e.g. [32] for definitions). Letting $\mathcal{S}^\Delta$ be a numerical evolution
operator (as described in the previous section), we set

$$\mu_t^\Delta := \mathcal{S}^\Delta(t) \# \bar{\mu}.$$

THEOREM 3.1. *Let $\bar{\mu} \in \mathcal{P}_0(\mathbb{R}^d)$ be the initial data. For every $t \in \mathbb{R}_+$, define*
$\mu_t^\Delta := \mathcal{S}^\Delta(t) \# \bar{\mu}$ *and* $\mu_t := \mathcal{S}(t) \# \bar{\mu}$. *Assume the numerical scheme has a convergence
rate of s for the given initial data, i.e.,*

$$(3.3) \qquad \int_{L^1(\mathbb{R}^d)} \left\| \mathcal{S}(t)(\bar{u}) - \mathcal{S}^\Delta(t)(\bar{u}) \right\|_{L^1(\mathbb{R}^d)} d\bar{\mu}(\bar{u}) \leqslant C\Delta^s,$$

for some $s > 0$. Then

$$(3.4) \qquad W_1(\mu_t^\Delta, \mu_t) \leqslant C\Delta^s.$$

Proof. Define the measure $\gamma \in \mathcal{P}\left(L^1(\mathbb{R}^d) \times L^1(\mathbb{R}^d)\right)$ by

$$\gamma = \left(\mathcal{S}(t), \mathcal{S}^\Delta(t) \right) \# \bar{\mu}.$$

One can trivially check that $\gamma \in \Pi(\mu_t, \mu_t^\Delta)$. We obtain

$$W_1(\mu_t, \mu_t^\Delta) \leqslant \int_{L^1(\mathbb{R}^d)^2} \|u_1 - u_2\|_{L^1(\mathbb{R}^d)} \, d\gamma(u_1, u_2)$$

$$= \int_{L^1(\mathbb{R}^d)} \left\| \mathcal{S}(t)(\bar{u}) - \mathcal{S}^\Delta(t)(\bar{u}) \right\|_{L^1(\mathbb{R}^d)} d\bar{\mu}(\bar{u}).$$

Inserting the assumptions of (3.3) yields the claim. $\qquad \square$

The above theorem is used below to obtain convergence rates for monotone schemes
approximating scalar conservation laws (see [18, 21] for definitions).

COROLLARY 3.2. *Assume the numerical scheme $\mathcal{S}^\Delta$ is monotone. Let $\bar{\mu} \in \mathcal{P}_0(\mathbb{R}^d)$
be such that*

$$\mathrm{TV}(\bar{\mu}) := \int_{L^1(\mathbb{R}^d)} \mathrm{TV}(\bar{u}) \, d\bar{\mu}(\bar{u}) < \infty.$$

For every $t \in \mathbb{R}_+$, define $\mu_t := \mathcal{S}(t) \# \bar{\mu}$ and $\mu_t^\Delta := \mathcal{S}^\Delta(t) \# \bar{\mu}$. Then

$$(3.5) \qquad W_1(\mu_t^\Delta, \mu_t) \leqslant C\Delta^{1/2} \, \mathrm{TV}(\bar{\mu}).$$

Proof. The convergence estimate of Kuznetsov [22] for monotone schemes approx-
imating scalar conservation laws yields

$$\left\| \mathcal{S}(t)(\bar{u}) - \mathcal{S}^\Delta(t)(\bar{u}) \right\|_{L^1(\mathbb{R}^d)} \leqslant C\sqrt{\Delta} \, \mathrm{TV}(\bar{u})$$

for every $\bar{u} \in L^1$ with $\mathrm{TV}(\bar{u}) < \infty$. The estimate (3.5) now follows from (3.4). $\qquad \square$

4. Monte Carlo approximations of statistical solutions. Let $\bar{\mu} \in \mathcal{P}(L^1(\mathbb{R}^d))$ be the given initial measure. We choose a probability space $(\Omega, \mathcal{X}, \mathbb{P})$ and a random field $\bar{u} \in L^2(\Omega; L^1(\mathbb{R}^d))$ such that the law of $\bar{u}$ with respect to $\mathbb{P}$ is $\bar{\mu}$. In most real-world applications, the uncertainty in initial data is usually described in terms of such a random field $\bar{u}$, for instance, one given as a parametric function $\bar{u} : [0,1]^Q \times \mathbb{R}^d \to \mathbb{R}$, with possibly $Q \gg 1$. Draw M independent and identically distributed (i.i.d.) samples $\bar{u}_1, \ldots, \bar{u}_M$ of $\bar{u}$ and for $1 \leqslant k \leqslant M$, set

$$u_k(\omega, \cdot, t) := \mathcal{S}^\Delta(t)(\bar{u}_k).$$

We define the Monte Carlo $\mu_t^{\Delta, M}$ approximation as

$$\mu_t^{\Delta, M} := \frac{1}{M} \sum_{k=1}^M \delta_{u_k(\omega; \cdot, t)},$$

or equivalently, through its action on an integrable function,

$$\left\langle \mu_t^{\Delta, M}, G \right\rangle = \frac{1}{M} \sum_{k=1}^M G(u_k(\omega; \cdot, t)) \qquad G \in C_b(L^1(\mathbb{R}^d)).$$

4.1. Convergence of Monte Carlo. The relevant notion of convergence on $\mathcal{P}(L^1(\mathbb{R}^d))$ is that of weak convergence. We say that a sequence $\{\rho_n\}_{n \in \mathbb{N}} \subset \mathcal{P}(L^1(\mathbb{R}^d))$ converges weakly to $\rho \in \mathcal{P}(L^1(\mathbb{R}^d))$ if

$$\left\langle \rho_n, G \right\rangle \to \left\langle \rho, G \right\rangle \qquad \text{for all } G \in C_b(L^1(\mathbb{R}^d)).$$

It can be shown that weak convergence preserves the probability measure structure, that is, $\rho \geqslant 0$ and $\rho(X) = 1$.

Using standard Monte Carlo techniques we obtain the following convergence theorem for our Monte Carlo method.

THEOREM 4.1. *Let $\bar{\mu} \in \mathcal{P}_0(\mathbb{R}^d)$ be the initial data. For $t \in \mathbb{R}_+$, let $\mu_t^\Delta := \mathcal{S}^\Delta(t) \# \bar{\mu}$ and let $\mu_t^{\Delta, M}$ be the Monte Carlo approximation. Then*

$$(4.1) \qquad \left\| \left\langle \mu_t^{\Delta, M} - \mu_t^\Delta, G \right\rangle \right\|_{L^2(\Omega)} = \frac{\sqrt{\mathrm{Var}\left[G(u^{\Delta x}(\cdot, \cdot, t)) \right]}}{M^{1/2}}$$

for all $G \in C_b(L^1(\mathbb{R}^d))$. Here,

$$(4.2) \qquad \mathrm{Var}\left[G(u^{\Delta x}(\cdot, \cdot, t)) \right] := \int_{L^1(\mathbb{R}^d)} G(u)^2 \, d\mu_t^\Delta(u) - \left(\int_{L^1(\mathbb{R}^d)} G(u) \, d\mu_t^\Delta(u) \right)^2.$$

Moreover, if the assumptions of Theorem 3.1 are fulfilled then observables converge as $M \to \infty$, $\Delta \to 0$ in $L^2(\Omega)$:

$$(4.3) \qquad \left\| \left\langle \mu_t^{\Delta, M} - \mu_t, G \right\rangle \right\|_{L^2(\Omega)} \leqslant H^\Delta(G) + C_2 M^{-1/2},$$

where $H^\Delta(G) := \left| \left\langle \mu_t - \mu_t^\Delta, G \right\rangle \right| \to 0$ as $\Delta \to 0$ whenever $G \in C_b(L^1(\mathbb{R}^d))$. If G is Lipschitz continuous then

$$\left\| \left\langle \mu_t^{\Delta, M} - \mu_t, G \right\rangle \right\|_{L^2(\Omega)} = \mathcal{O}(\Delta^s + M^{-1/2}).$$

Proof. We have

$$(4.4) \qquad \langle \mu_t^\Delta, G \rangle = \int_\Omega G(u^\Delta(\omega', \cdot, t)) \, d\mathbb{P}(\omega').$$

Hence,

$$\left\| \langle \mu_t^{\Delta,M} - \mu_t^\Delta, G \rangle \right\|_{L^2(\Omega)}^2 = \int_\Omega \left[\int_\Omega G(u^\Delta(\omega', \cdot, t)) \, d\mathbb{P}(\omega') - \frac{1}{M} \sum_{k=1}^M G(u_k(\omega; \cdot, t)) \right]^2 d\mathbb{P}(\omega)$$
$$= \frac{\mathrm{Var}[G(u^{\Delta x}(\cdot, \cdot, t))]}{M},$$

where the key observation is that $G(u^\Delta(\omega, \cdot, t))$ is a real-valued random variable, hence we can directly apply the standard Monte Carlo estimate found in [23]. This proves (4.1).

To prove (4.3) we apply the triangle inequality to get

$$\left\| \langle \mu_t^{\Delta,M} - \mu_t, G \rangle \right\|_{L^2(\Omega)} \leqslant \left\| \langle \mu_t - \mu_t^\Delta, G \rangle \right\|_{L^2(\Omega)} + \left\| \langle \mu_t^\Delta - \mu_t^{\Delta,M}, G \rangle \right\|_{L^2(\Omega)}$$
$$\leqslant H^\Delta(G) + \frac{C_2}{M^{1/2}},$$

the second inequality following from (4.1) and the fact that $\langle \mu_t - \mu_t^\Delta, G \rangle$ does not depend on the stochastic parameter ω. Theorem 3.1 implies that $\mu_t^\Delta \to \mu_t$ for all t, which is equivalent to $H^\Delta(G) \to 0$ for all $G \in C_b$. The last assertion follows from the estimate $H^\Delta(G) \leqslant \|G\|_{\mathrm{Lip}} W_1(\mu_t, \mu_t^\Delta) = \mathcal{O}(\Delta^s)$. ☐

Note that if $G \in C_b(L^1(\mathbb{R}^d))$ then $\mathrm{Var}[G(u^{\Delta x}(\cdot, \cdot, t))] \leqslant \|G\|_{C_b(L^1)}^2 < \infty$, and hence the right-hand side of (4.1) is finite. Examples of such test functions G include finite-dimensional observables $G(u) = g(\langle \phi_1, u \rangle, \ldots, \langle \phi_n, u \rangle)$ for functionals $\phi_1, \ldots, \phi_n \in L^1(\mathbb{R}^d)^* = L^\infty(\mathbb{R}^d)$ and $g \in C_b(\mathbb{R}^n)$.

REMARK 1. *It can be shown that the convergence rate of the Monte Carlo method, measured in the Wasserstein metric for probability measures on $\mathbb{R}^d$, deteriorates as d grows [16], and hence one can not expect to obtain a convergence rate for the Monte Carlo method in the Wasserstein metric for probability measures on L^1. However, as the weak topology on probability measures is metrized by the Wasserstein metric, Theorem 4.1 enables us to conclude that the Monte Carlo method does converge with respect to the Wasserstein metric as $M \to \infty$.*

4.2. Work analysis for Monte Carlo. The *work* of a numerical method is the number of floating point operations it consumes. The classical explicit finite volume method has a work estimate of

$$(4.5) \qquad \mathrm{Work}_{\mathrm{FVM}}(\Delta, \Delta t) = C\Delta^{-d}\Delta t^{-1},$$

for some constant $C > 0$. Applying the CFL condition $\Delta t = C_{\mathrm{CFL}}\Delta$ gives

$$\mathrm{Work}_{\mathrm{FVM}}(\Delta, \Delta t) = \mathrm{Work}_{\mathrm{FVM}}^\Delta = C_{\mathrm{FVM}}\Delta^{-d-1},$$

for some constant $C_{\mathrm{FVM}} > 0$. Thus, the work required to run the Monte Carlo algorithm scales as

$$\mathrm{Work}_{\mathrm{MC}}^{\Delta,M} = M\mathrm{Work}_{\mathrm{FVM}}^\Delta = C_{\mathrm{FVM}}M\Delta^{-d-1}.$$

We assume that the spatial error scales as

$$W_1(\mu_t, \mu_t^\Delta) = \mathcal{O}(\Delta^s),$$

and we choose the number of samples such that the Monte Carlo error is asymptotically the same as the spatial error. That is, we choose

$$M = \Delta^{-2s}.$$

This gives the optimal work estimate

(4.6) $$\text{Work}_{\text{MC}}^{\Delta,M} = C_{\text{FVM}}\Delta^{-d-1-2s},$$

for some constant C_{FVM} depending on the finite volume scheme and its implementation.

5. Multilevel Monte Carlo approximation. We define the multi-level Monte Carlo approximation similar to that found in [17] and [27]. For *fixed* $L \in \mathbb{N}$, we assume we have a family of nested meshes with mesh lengths $\Delta_0, \ldots, \Delta_L$, where

(5.1) $$\Delta_l = 2^{-l}\Delta_0 \qquad \text{for } l = 1, \ldots, L,$$

for some given base resolution $\Delta_0 > 0$. In the convergence theory of multi-level Monte Carlo we will quantify the convergence as $\Delta_0 \to 0$. Following the notation in the previous section, we define the multi-level Monte Carlo approximation of μ_t to be

(5.2) $$\mu_t^{\mathbf{\Delta},\mathbf{M}} := \mu_t^{\Delta_0,M_0} + \sum_{l=1}^{L}\left(\mu_t^{\Delta_l,M_l} - \mu_t^{\Delta_{l-1},M_l}\right),$$

where $\mathbf{M} = (M_0, \ldots, M_L) \in \mathbb{N}^{L+1}$ and $\mathbf{\Delta} = (\Delta_0, \ldots, \Delta_L)$. In other words, we set

$$\mu_t^{\mathbf{\Delta},\mathbf{M}}(\omega) := \frac{1}{M_0}\sum_{k=1}^{M_0}\delta_{u_k^{\Delta_0}(\omega;\cdot,t)} + \sum_{l=1}^{L}\frac{1}{M_l}\sum_{k=1}^{M_l}\left(\delta_{u_k^{\Delta_l}(\omega;\cdot,t)} - \delta_{u_k^{\Delta_{l-1}}(\omega;\cdot,t)}\right).$$

Note that $\mu_t^{\mathbf{\Delta},\mathbf{M}}(\omega)$ will in general be a signed measure, rather than a probability measure, on $L^1(\mathbb{R}^d)$.

THEOREM 5.1. *Let $L \in \mathbb{N}$ be fixed. Let $\bar{\mu} \in \mathcal{P}_0(\mathbb{R}^d)$ be the initial data, and for any $t \in \mathbb{R}_+$ and $\Delta > 0$, let $\mu_t^\Delta := \mathcal{S}^\Delta(t)\#\bar{\mu}$ and let $\mu_t^{\mathbf{\Delta},\mathbf{M}}$ be the multi-level Monte Carlo approximation* (5.2). *Then*

$$\left\|\langle \mu_t^{\mathbf{\Delta},\mathbf{M}} - \mu_t^{\Delta_L}, G\rangle\right\|_{L^2(\Omega)} \leqslant \frac{\text{Var}[G(u^{\Delta_0}(\cdot,t))]^{1/2}}{M_0^{1/2}}$$

$$+ \sum_{l=1}^{L}\frac{\text{Var}\left[G(u^{\Delta_l}(\cdot,t)) - G(u^{\Delta_{l-1}}(\cdot,t))\right]^{1/2}}{M_l^{1/2}}$$

for every $G \in C_b(L^1(\mathbb{R}^d))$. Moreover, for every $l = 1, \ldots, L$,

$$\text{Var}\left[G(u^{\Delta_l}(\cdot,t)) - G(u^{\Delta_{l-1}}(\cdot,t))\right] \to 0 \qquad \text{as } \Delta_0 \to 0.$$

Proof. We observe that

$$\mu^{\Delta_L} = \mu^{\Delta_0} + \sum_{l=1}^{L} \left(\mu^{\Delta_l} - \mu^{\Delta_{l-1}} \right).$$

Therefore,

$$\left\| \langle \mu_t^{\mathbf{\Delta},\mathbf{M}} - \mu_t^{\Delta_L}, G \rangle \right\|_{L^2(\Omega)} \leqslant \left\| \langle \mu_t^{\Delta_0,M_0} - \mu_t^{\Delta_0}, G \rangle \right\|_{L^2(\Omega)}$$
$$+ \sum_{l=1}^{L} \left\| \langle \mu_t^{\Delta_l,M_l} - \mu_t^{\Delta_{l-1},M_l}, G \rangle - \langle \mu_t^{\Delta_l} - \mu_t^{\Delta_{l-1}}, G \rangle \right\|_{L^2(\Omega)}.$$

Using (4.1) from Theorem 4.1 we can bound the first term as

$$\left\| \langle \mu_t^{\Delta,M_0} - \mu_t^{\Delta_0}, G \rangle \right\|_{L^2(\Omega)} \leqslant C_G M_0^{-1/2}.$$

For the second term, we observe that

$$\left\| \langle \mu_t^{\Delta_l,M_l} - \mu_t^{\Delta_{l-1},M_l}, G \rangle - \langle \mu_t^{\Delta_l} - \mu_t^{\Delta_{l-1}}, G \rangle \right\|_{L^2(\Omega)}$$
$$= \left\| \frac{1}{M_l} \sum_{k=1}^{M_l} \left(G(u_k^{\Delta_l}(\cdot,t)) - G(u_k^{\Delta_{l-1}}(\cdot,t)) \right) - \int_\Omega \left(G(u^{\Delta_l}(\omega';\cdot,t)) - G(u^{\Delta_{l-1}}(\omega';\cdot,t)) \right) d\mathbb{P}(\omega') \right\|_{L^2(\Omega)}.$$

Using a standard Monte Carlo estimate gives

$$\left\| \langle \mu_t^{\Delta_l,M_l} - \mu_t^{\Delta_{l-1},M_l}, G \rangle - \langle \mu_t^{\Delta_l} - \mu_t^{\Delta_{l-1}}, G \rangle \right\|_{L^2(\Omega)} \leqslant \frac{\text{Var}\left[G(u^{\Delta_l}(\cdot,t)) - G(u^{\Delta_{l-1}}(\cdot,t)) \right]}{M_l^{1/2}}.$$

Summing over l and adding the first estimate yields the first statement.

For the last statement of the theorem, we observe that since G is continuous and $(u^\Delta)_{\Delta>0}$ forms a convergent sequence as $\Delta \to 0$, we have that $G(u^\Delta)$ forms a pointwise Cauchy sequence. Appealing to the dominated convergence theorem and the boundedness of G yields the statement. □

REMARK 2. *The above theorem guarantees that $\mu^{\mathbf{\Delta},\mathbf{M}}$ converges as long as*

$$\frac{\text{Var}\left[G(u^{\Delta_0}(\cdot,t)) \right]^{1/2}}{M_0^{1/2}} + \sum_{l=1}^{L} \frac{\text{Var}\left[G(u^{\Delta_l}(\cdot,t)) - G(u^{\Delta_{l-1}}(\cdot,t)) \right]^{1/2}}{M_l^{1/2}} \to 0.$$

The interesting case, considered in Subsection 5.1 and section 6, occurs when MLMC is faster than Monte Carlo. This happens when

$$\text{Var}\left[G(u^{\Delta_l}(t\cdot,t)) - G(u^{\Delta_{l-1}}(\cdot,t)) \right]^{1/2} \ll \text{Var}\left[G(u^{\Delta_0}(\cdot,t)) \right].$$

REMARK 3. *For general $G \in C_b(L^1(\mathbb{R}^d))$, we can only guarantee that*

(5.3) $$\text{Var}\left[G(u^{\Delta_l}(\cdot,t)) - G(u^{\Delta_{l-1}}(\cdot,t)) \right] \to 0.$$

However, for practical computations, G is known, and we can get a much better estimate on the decay of (5.3).

5.1. Work analysis of MLMC. In this section we extend the analysis of Subsection 4.2 to the MLMC algorithm. In the computation of the MLMC approximation $\mu^{\boldsymbol{\Delta},\mathbf{M}}$ we compute M_l finite volume simulations with resolution Δ_l for each $l = 0, \ldots, L$, and M_l finite volume simulations with resolution Δ_{l-1} for $l = 1, \ldots, L$. If $M_l \leqslant M_{l-1}$ then the latter can be neglected, and we find that the work performed by the MLMC algorithm is

$$
\begin{aligned}
\text{Work}_{\text{MLMC}}^{\boldsymbol{\Delta},\mathbf{M}} &= \sum_{l=0}^{L} M_l \text{Work}_{\text{FVM}}^{\Delta_l} \\
&= \sum_{l=0}^{L} C_{\text{FVM}} M_l \Delta_l^{-d-1} \\
&= C_{\text{FVM}} \sum_{l=0}^{L} M_l 2^{l(d+1)} \Delta_0^{-d-1}.
\end{aligned}
\tag{5.4}
$$

The number of samples per level, M_l, has so far been unspecified. It is common to optimize the number of samples for a given convergence rate. We handle the general case, and optimize with respect to the number of samples, where the variance across the levels is abstractly given as

$$
\text{Var}\left[G(u^{\Delta_l}) - G(u^{\Delta_{l-1}}) \right] = V_l.
\tag{5.5}
$$

THEOREM 5.2. *Assume that*

$$
V_l = \mathcal{O}(\Delta_l^r)
$$

for some $r > 0$. Choose the number of samples per level so that the error per level is equidistributed, that is, choose

$$
M_0 = \Delta_L^{-2s}, \qquad M_l = 2^{r(L-l)} \Delta_L^{r-2s}, \qquad l = 1, \ldots, L.
$$

Then

$$
\text{Work}_{MLMC}^{\boldsymbol{\Delta},\mathbf{M}} \leqslant C_{FVM} \Delta_L^{-d-1-2s} \left(2^{-L(d+1)} + \Delta_L^r L \right).
\tag{5.6}
$$

REMARK 4. *The estimate* (5.6) *shows that the MLMC method is more efficient than Monte Carlo, by a factor depending on L. To see this, let $\boldsymbol{\Delta}$ and $\mathbf{M}$ be as prescribed. Comparing the error estimate of the Monte Carlo method (Theorem 4.1) with the same for MLMC (Theorem 5.1), one sees that the Monte Carlo method with mesh size Δ_L has the same error as MLMC when $M = \Delta_L^{-2s}/L^2 = M_L \Delta_L^{-r}/L^2$. Comparing the corresponding work estimates yields*

$$
\text{Work}_{MLMC}^{\boldsymbol{\Delta},\mathbf{M}} \leqslant L^2 2^{-L(d+1)+1} \, \text{Work}_{MC}^{\Delta_L, \Delta_L^{-2s}/L^2},
$$

(assuming that Δ_L is sufficiently small, $\Delta_L \leqslant \left(\frac{2^{-L(d+1)}}{L} \right)^{\frac{1}{r}}$). Thus, the gain in efficiency with the MLMC method is a constant which increases exponentially with L.

Proof of Theorem 5.2. We compute

$$
\begin{aligned}
\frac{\mathrm{Work}^{\mathbf{\Delta},\mathbf{M}}_{\mathrm{MLMC}}}{C_{\mathrm{FVM}}} &= \sum_{l=0}^{L} \Delta_l^{-d-1} M_l \\
&= \sum_{l=0}^{L} (2^{L-l}\Delta_L)^{-d-1} M_l \\
&= (2^L\Delta_L)^{-d-1} M_0 + \sum_{l=1}^{L} (2^{L-l}\Delta_L)^{-d-1} M_l \\
&= 2^{-L(d+1)}\Delta_L^{-d-1-2s} + \sum_{l=1}^{L} (2^{-d-1})^{L-l}\Delta_L^{-d-1} 2^{r(L-l)}\Delta_L^{r-2s} \\
&= 2^{-L(d+1)}\Delta_L^{-d-1-2s} + (2^{-d-1+r})^L \Delta_L^{-d-1+r-2s} \sum_{l=1}^{L} (2^{d+1-r})^l \\
&\leqslant 2^{-L(d+1)}\Delta_L^{-d-1-2s} + (2^{-d-1+r})^L \Delta_L^{-d-1+r-2s} L (2^{d+1-r})^L \\
&= 2^{-L(d+1)}\Delta_L^{-d-1-2s} + \Delta_L^{-d-1+r-2s} L,
\end{aligned}
$$

which is what we wanted. □

REMARK 5. *It is possible to obtain a similar work estimate by minimizing the work for a given error and using a Lagrange multiplier technique [17].*

6. Numerical Examples. In this section, we will test the Monte Carlo and the multi-level Monte Carlo on a suite of numerical experiments. Our model equation for scalar conservation laws is the one-dimensional Burgers equation

$$
\begin{aligned}
(6.1) \qquad u_t + \left(\frac{u^2}{2}\right)_x &= 0 &\text{for } x \in D := [0,1], \ t \leqslant T \\
u(x,0) &= \bar{u}(x) &\text{for } x \in D
\end{aligned}
$$

with suitable boundary conditions.

We will compute statistical quantities of interest for the approximated solution. These include one-point statistics such as the mean and the variance. We will also compute a two-point *local structure function* of the solution, which is given by

$$
(6.2) \qquad S_p(x,t;h) := \int_{\mathbb{R}^2} |u_1 - u_2|^p \, d\nu^2_{t,x,x+h}(u_1,u_2) \qquad x \in D, h \geqslant 0
$$

for some $p \geqslant 1$. We also define the integrated structure functions as

$$
(6.3) \qquad \bar{S}_p(t;h) = \int_D S_p(x,t;h) \, dx = \int_{L^1(D)} \int_D |u(x+h) - u(x)|^p \, dx \, d\mu_t(u).
$$

We assume that the initial measure has integrable L^p-norms,

$$
\int_{L^1(D)} \|\bar{u}\|_{L^p(\mathbb{R}^d)}^p \, d\bar{\mu}(\bar{u}) < \infty.
$$

By Minkowski's inequality, this assumption implies that the structure function $\bar{S}_p$ is finite, and we may therefore approximate it using the Monte Carlo algorithm.

We let $S_p^{\Delta,M}(x,t;h)$, $\bar{S}_p^{\Delta,M}(t;h)$ and $S_p^{\boldsymbol{\Delta},\mathbf{M}}(x,t;h)$, $\bar{S}_p^{\boldsymbol{\Delta},\mathbf{M}}(t;h)$ denote the local and integrated structure functions for the single- and multi-level Monte Carlo methods, respectively.

We define a perturbed version of the three-point moment by setting

$$(6.4) \qquad M_p(x,t;h_1,h_2) := \int_{\mathbb{R}^3} (u_1 - u_2)(u_1 - u_3)^2 \, d\nu_{t,x,x+h_1,x+h_2}^3(u_1,u_2,u_3).$$

This three-point moment is also bounded assuming, e.g.,

$$\int_{L^1(D)} \|\bar{u}\|_{L^3(D)}^3 \, d\bar{\mu}(\bar{u}) < \infty.$$

6.1. Uncertain shock location. We consider Burgers' equation with the initial data

$$(6.5) \qquad \bar{u}(\omega,x) = \begin{cases} 1 & x < X(\omega) + \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \qquad x \in [0,1]$$

with Dirichlet boundary conditions. Here, X is a a uniformly distributed random variable on $[-\frac{1}{10}, \frac{1}{10}]$. We can explicitly compute the mean and the variance in this case. Given the one-dimensional stochastic space for this scalar problem, the structure function (6.2) can also be explicitly computed,

$$S_p(x,t;h) = \begin{cases} 0 & x + \frac{1}{10} + h < \frac{1}{2}(1+t) \\ 0 & x - \frac{1}{10} + h > \frac{1}{2}(1+t) \\ 5\left(\min(\frac{1}{10}, x + h - \frac{1}{2}(1+t)) - \max(-\frac{1}{10}, x - \frac{1}{2}(1+t)))\right) & \text{otherwise.} \end{cases}$$

Since the function $\omega \mapsto u(\omega,\cdot,\cdot)$ is uniformly bounded in $L^1([0,1])$, one can verify that the functions S_p can be viewed as a function in $C_b(L^1(\mathbb{R}^d))$.

We use the Monte Carlo and multi-level Monte Carlo algorithm to approximate different statistical quantities of interest. By Theorem 5.2, we need to estimate the decay of V_l in (5.5). From numerical experiments we estimate that $V_l = \mathcal{O}(\Delta)$ for $G = S_p$. For the Monte Carlo simulations we set the number of samples equal to the number of finite volume cells, while for the MLMC simulations we set

$$M_l = \begin{cases} 16 \cdot 2^{L-l} & \text{if } l = 1,\ldots,L \\ 1/\Delta^L & \text{if } l = 0. \end{cases}$$

For the spatial discretization we use a finite volume solver, based on the Godunov flux with piecewise linear WENO reconstruction and an SSP Runge–Kutta method of second order. We set the CFL constant to be 0.475.

We plot the point-wise mean and variance in Figure 1, and the corresponding convergence histories in Figure 2. The plots show that both the MC and MLMC methods approximate the mean and variance quite accurately. The convergence plots also show that the MLMC method is (an order of magnitude) more efficient when compared to the MC method.

For a comparison between the numerical and analytically computed structure functions, see Figure 3. In addition, we perform a convergence study with respect to the analytical solution in Figure 4. As is clear from Figure 4, the work required

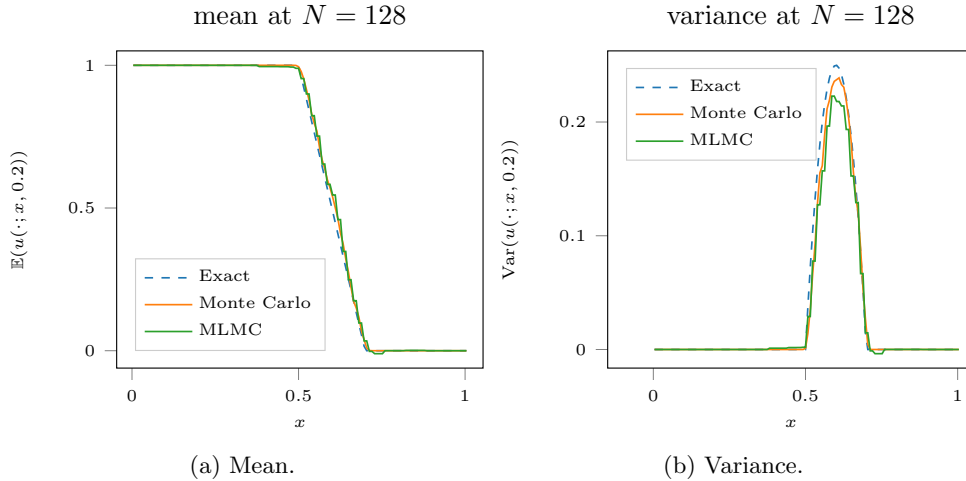(a) Mean.   (b) Variance.

Fig. 1: Mean and variance for Burgers' equation and initial data given in (6.5).

to achieve a given error in the Monte Carlo simulation is significantly greater than
the computational work required with MLMC method. In fact, the gain in efficiency
is a couple of orders of magnitude. Moreover, in Figure 4 (right), we plot the error
vs. work for the Monte Carlo and MLMC algorithms in computing the third moment
(6.4). Again the gain in efficiency with the MLMC method is significant.

To confirm the assertion of Theorem 3.1, we measure the Wasserstein distance
between the generated sequence and the exact solution in Figure 5a. Since the Monte
Carlo ensemble is a stochastic quantity, the measured error will also be stochastic.
We therefore run 10 experiments for each mesh resolution and measure the average
error. As is clear from the figure, the convergence rate of the mean of the error is
close to $1/2$, which agrees well with the Monte Carlo error in (4.1). However, as we
can see, the variance of the error is rather large relative to the error.

Because the initial data (6.5) only depends on a one-dimensional parameter, we
can replace the stochastic Monte Carlo integration rule with a deterministic midpoint
integration rule. Figure 5b clearly shows a convergence rate of 1 for the deterministic
integration rule.

6.2. Fractional Brownian motion. As is standard in the *burgulence* literature
[2, 30, 5], we perform experiments for Burgers' equation (6.1) with the initial data set
to be *fractional Brownian motion*. Introduced by Mandelbrot et al. [26], fractional
Brownian motion can be seen as a generalization of standard Brownian motion with
a scaling exponent different than $1/2$. We set

$$u_0^H(\omega; x) := B^H(\omega; x) \qquad \omega \in \Omega, \ x \in [0, 1],$$

where B^H is fractional Brownian motion with Hurst exponent $H \in (0, 1)$. Brownian
motion corresponds to a Hurst exponent of $H = 1/2$.

To generate fractional Brownian motion, we use the random midpoint displace-
ment method originally introduced by Lévy [25] for Brownian motion, and later
adapted for fractional Brownian motion [15, 33]. Consider a uniform partition $0 =
x_{1/2} < \cdots < x_{N+1/2} = 1$ with $x_{i+1/2} - x_{i-1/2} \equiv \Delta$, where $N = 2^k + 1$ is the number of
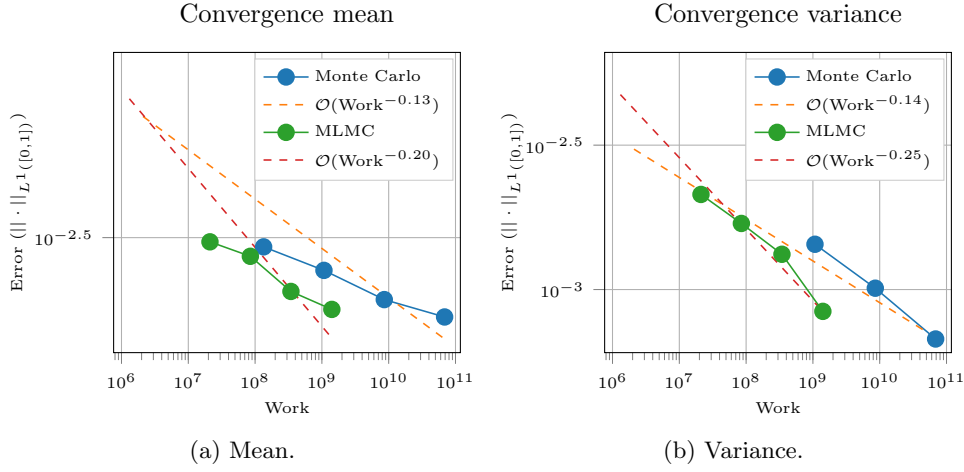
Convergence mean

Convergence variance

(a) Mean.

(b) Variance.

Fig. 2: Convergence for mean and variance for Burgers' equation and initial data given in (6.5).
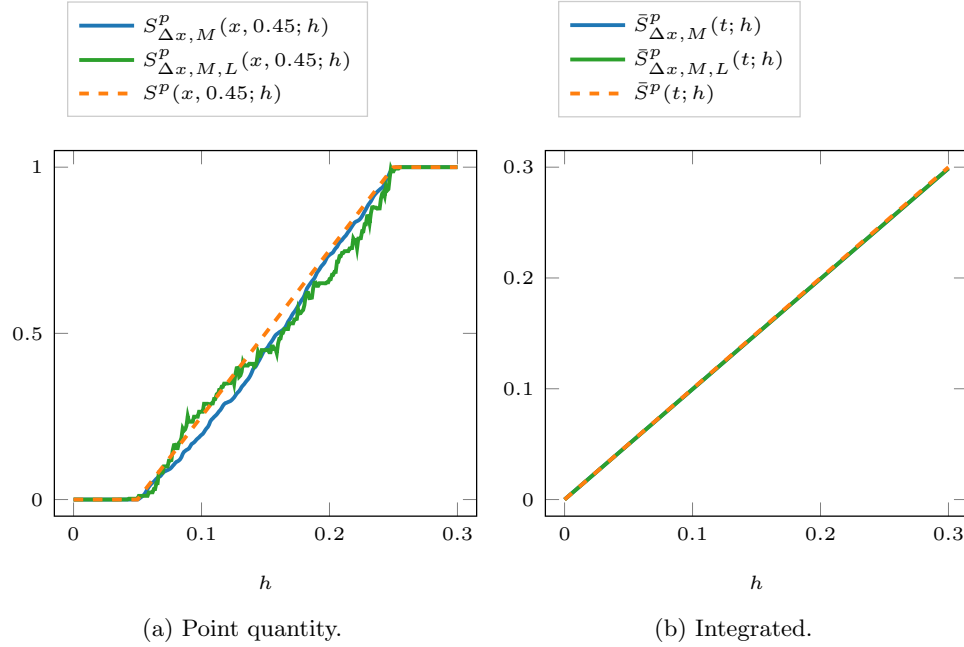
(a) Point quantity.

(b) Integrated.

Fig. 3: Comparison of two point structure function numerically computed and analytic solution for Burgers' equation and initial data given in (6.5). Here $\Delta = 1/1024$ and $M = 1024$.
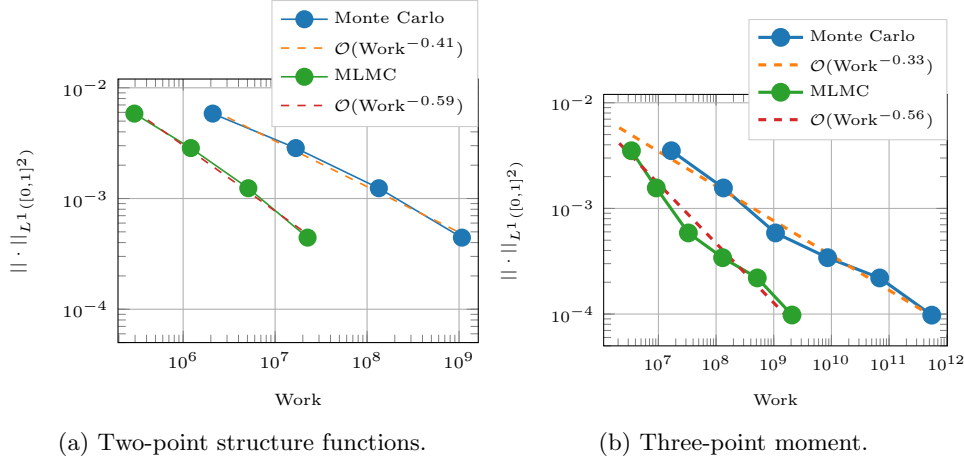
(a) Two-point structure functions.   (b) Three-point moment.

Fig. 4: Convergence errors of the two-point structure function (6.3) and three-point moment (6.4) with initial data given as the uncertain shock location (6.5).

(a) Monte Carlo algorithm   (b) Midpoint rule
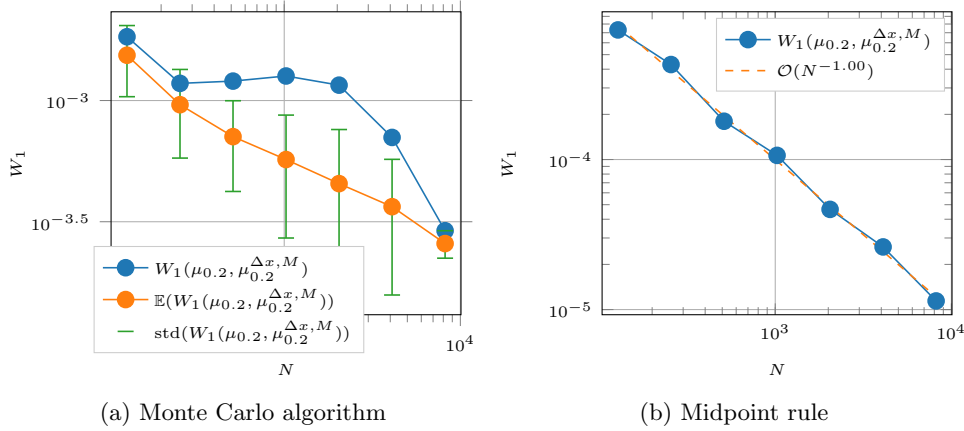
Fig. 5: Wasserstein convergence for the initial data given in (6.5) at $T = 0.2$.

cells for some $k \in \mathbb{N}$. We first fix the endpoints

$$u_1^{H,\Delta x}(\omega; 0) = 0 \qquad u_N^{H,\Delta x}(\omega; 0) = X_0(\omega),$$

where $(X_k)_{k \in \mathbb{N}}$ is a collection of normally distributed random variables with mean 0 and variance 1. Recursively, we set

$$u_{2^{k-l-1}(2j+1)}^{H,\Delta x}(\omega; 0) = \frac{1}{2}\left(u_{2^{k-l}(j+1)}^{H,\Delta x}(\omega; 0) + u_{2^{k-l}j}^{H,\Delta x}(\omega; 0)\right) + \sqrt{\frac{1 - 2^{2H-2}}{2^{2lH}}}X_{2^l+j}(\omega)$$

for $l = 0, \dots, k$ and for $j = 0, \dots, 2^l$. That is, we bisect every interval and set the middle value to the average of the neighbouring values plus some Gaussian random variable. See Figure 6 (left) for an example with $H = 0.01$.

We run experiments for the standard Brownian motion ($H = 0.5$) and an even rougher (pathwise) initial datum, corresponding to $H = 0.01$. In both cases, one can easily check that the initial total variation $\mathrm{TV}(\bar{u}^H(\omega, \cdot))$ is in fact infinite for almost all ω, and hence the Kuznetszov error estimate [22] does not apply. Furthermore, since the initial does not have bounded support, we can not view S_p as a function in $C_b(L^1(\mathbb{R}^d))$, but numerical experiments indicate that the variance is finite.

By Theorem 5.2, we need to estimate the decay of V_l in (5.5). From numerical experiments, we estimate that $V_l = \mathcal{O}(\Delta)$ for $G = S_p$. For the Monte Carlo simulations we set the number of samples equal to the number of finite volume cells, while for the MLMC simulations we set

$$M_l = \begin{cases} 16 \cdot 2^{L-l} & \text{if } l = 1, \ldots, L \\ 1/\Delta^L & \text{if } l = 0 \end{cases}$$

For the Finite Volume Method, we use the Godunov flux with WENO2 reconstruction and an SSP Runge–Kutta method of second order. We set the CFL constant to be 0.475.

We measure the decay of V_l defined in (5.5) numerically for $H = 0.01$ and $H = 0.5$ and see that V_l as a function of the mesh width h behaves approximately as $\mathcal{O}(h)$ (results not displayed here). This happens in spite of the fact that the Kuznetszov error estimate does not apply.

Our aim is to numerically approximate structure functions. Following [30], one intuitively argues that if the (pathwise) solution consists of a disjoint set of shocks, well-separated by rarefaction waves for almost all times, then the exponent of the structure function (6.3) will be dominated by the behavior at shocks and the structure functions will scale as $\mathcal{O}(h)$ for any $1 \leqslant p < \infty$. Given the rigorous results of Sinai [31], one expects this scaling to hold in the case of standard Brownian motion initial data.

In Figure 7 we approximate the scaling exponents of $\bar{S}_p$ for different values of H. For $H = 0.5$, the structure functions scale close to $\mathcal{O}(h)$ for $p = 1, 2, 3$, which agrees with the results of [2, 31]. Similarly for $H = 0.01$, even though the initial data is much rougher than standard Brownian motion, we observe from Figure 6, that the initally highly oscillatory pathwise solution very quickly evolves into a set of well-separated shocks separated by rarefaction waves. As argued before, one expects a scaling exponent of approximately $\mathcal{O}(h)$ for all p. This is indeed verified in the approximate structure functions for $p = 1, 2, 3$, shown in Figure 7 (right).

We also measure the numerical convergence rate of the Monte Carlo and Multilevel Monte Carlo algorithm against a reference solution computed with $\Delta x = 1/4096$ and $M = 4096$. The results are shown in Figure 8b. As is expected, the MLMC algorithm outperforms the single-level Monte Carlo algorithm. Indeed, the error versus work asymptotic of a MLMC simulation is equal to that of the runtime asymptotics for a single sample.

This example clearly illustrates the ability of both the MC and MLMC methods to compute statistical quantities in a realistic and well-studied problems and demonstrates the considerable gain in efficiency for the MLMC method over the MC method.

6.3. Cubic conservation law with Brownian initial data. We redo the experiments in the previous section, but with the cubic conservation law,

$$(6.6) \qquad \begin{aligned} u_t + \left(\frac{u^3}{3}\right)_x &= 0 && \text{for } x \in D := [0, 1], \ t \leqslant T \\ u(x, 0) &= \bar{u}(x) && \text{for } x \in D. \end{aligned}$$
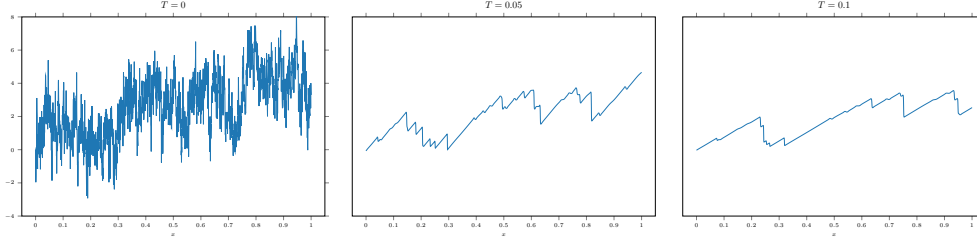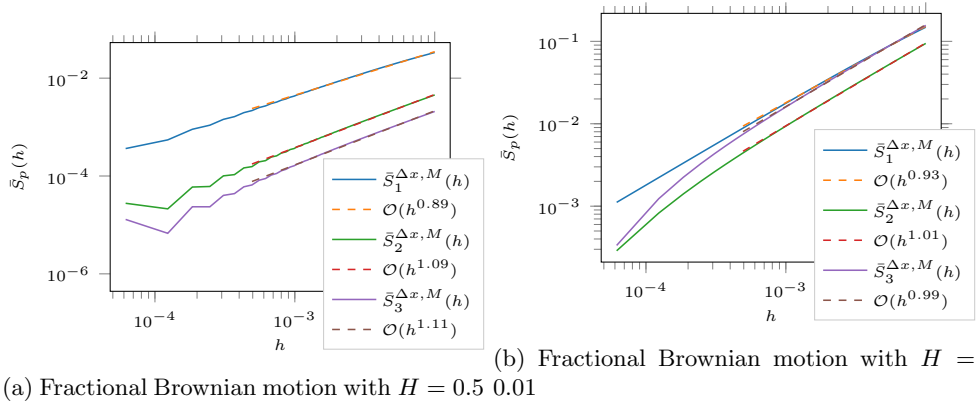
Fig. 6: Single sample simulation results for the Burgers' equation with initial data given as fractional Brownian motion with $H = 0.01$. In this example, $\Delta = 1/1024$.

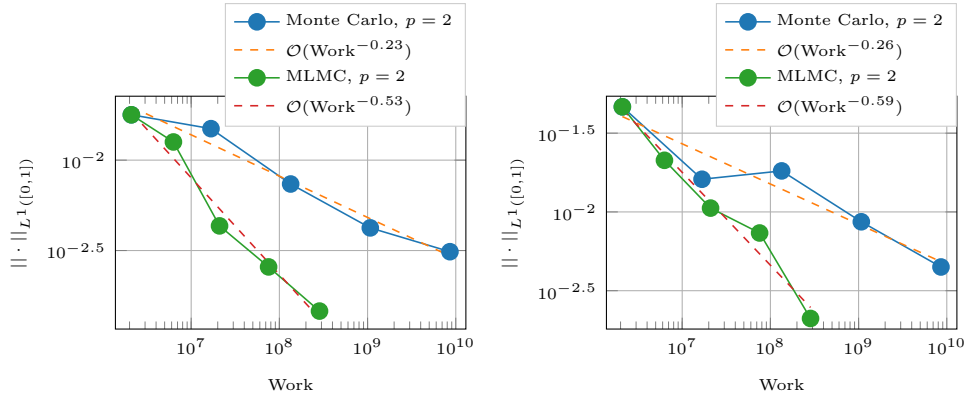

(a) Fractional Brownian motion with $H = 0.5$

(b) Fractional Brownian motion with $H = 0.01$

Fig. 7: Structure functions for Burgers' equation with initial data given as fractional Brownian motion. In this example, $\Delta = 1/16384$ and $M = 16384$.

In this case, it is not possible to derive explicit expressions or asymptotices for the structure functions as in the case of Burgers' equation [31], as no explicit Hopf–Lax-type formulas are available for a non-convex flux function. Therefore, numerical simulations are the main tool in calculating statistical quantities of interest. We approximate the structure functions (6.6) with initial data given as Brownian motion, in other words, we set $u_0(x, \omega) = B_x^{1/2}(\omega)$. The results are shown in Figure 9. As is clear from the figure, we get the same linear scaling of the structure functions for any $1 \leqslant p < \infty$, indicating that the initially highly oscillatory solution breaks down into shocks, well separated by rarefactions, as in the case of Burgers' equation.

7. Conclusion. We consider statistical solutions for scalar conservation laws. Statistical solutions are time-parametrized probability measures on $L^1(\mathbb{R}^d)$, whose time evolution is specified in terms of an infinite family of PDEs for the corresponding correlation measures. Statistical solutions provide a framework for *parametrization free* uncertainty quantification in conservation laws with random initial data and their well-posedness was established in [14].

We design an efficient algorithm for the numerical approximation of statistical solutions of scalar conservation laws. This algorithm is based on standard finite volume (difference) methods for the spatio-temporal discretization and (multi-level)

(a) Numerical error for the two point structure function against a reference solution for Burgers' equation with Brownian motion as initial data.

(b) Numerical error for the two point structure function against a reference solution for Burgers' equation with fractional Brownian motion with $H = 0.01$ as initial data.

Fig. 8: Error vs. work for the Monte Carlo and MLMC methods for Burgers' equation with fractional Brownian motion initial data.
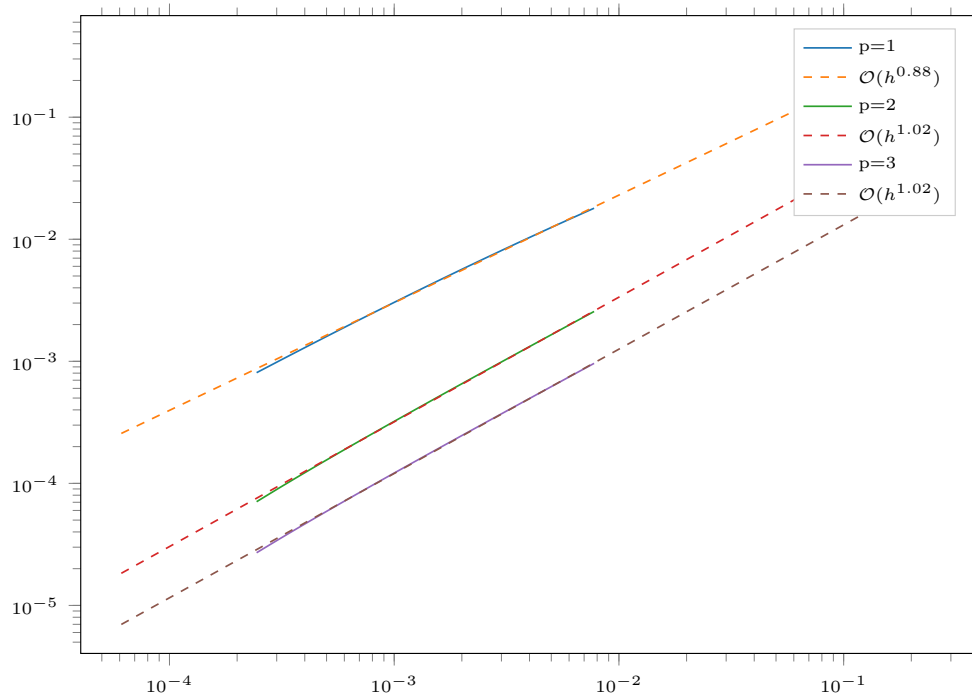

Fig. 9: Structure function scaling for a cubic conservation law ($f(u) = 1/3u^3$).

Monte Carlo methods for discretizing the probability space.

We prove that both the Monte Carlo and the multi-level Monte Carlo based algorithms converge to the *entropy statistical solutions* of scalar conservation laws, in the Wasserstein metric on probability measures on integrable functions, as the mesh is refined and the number of samples increased. We also present a complexity analysis and prove that there is a considerable gain in computational efficiency of the multi-level Monte Carlo method over the Monte Carlo method.

We present a set of numerical experiments for Burgers' equation to illustrate the ability of both the Monte Carlo and the multi-level Monte Carlo algorithms to approximate the statistical solution accurately and to demonstrate the gain in efficiency resulting from the MLMC method. These experiments involve rough random initial conditions such as fractional Brownian motion (in space), and the computed results are consistent with those published in the burgulence literature [2, 30]. We also compute statistical quantities of interest for the cubic conservation law. In this case, it is not possible to obtain analytical formulas for structure functions. However, numerical results show that the structure functions scale as in the case of Burgers' equation with rough random initial conditions.

The convergence of our numerical methods in the scalar case was underpinned by the presence of (pathwise) convergent numerical methods for the underlying deterministic problem. Such convergence results are not available for systems of conservation laws. Consequently, the design of convergent numerical approximations for statistical solutions of systems of conservation laws is very challenging and is addressed in a forthcoming paper.

REFERENCES

[1] R. ABGRALL AND S. MISHRA, *Uncertainty quantification for hyperbolic systems of conservation laws*, in Handbook of Numerical Methods for Hyperbolic Problems: Basic and Fundamental Issues, vol. 18 of Handbook of Numerical Analysis, Elsevier, 2016, pp. 507–544.

[2] J. BEC AND K. KHANIN, *Burgers turbulence*, Phys. Rep., 447 (2007), pp. 1–66.

[3] J. BERTOIN, *The inviscid Burgers equation with Brownian initial velocity*, Comm. Math. Phys., 193 (1998), pp. 397–406.

[4] *Uncertainty Quantification in Computational Fluid Dynamics*, vol. 92 of Lecture Notes in Computational Science and Engineering, Springer International Publishing, 2014.

[5] A. BORITCHEV, *Decaying turbulence in the generalised Burgers equation*, Arch. Ration. Mech. Anal., 214 (2014), pp. 331–357, https://doi.org/10.1007/s00205-014-0766-5, https://doi.org/10.1007/s00205-014-0766-5.

[6] A. BRESSAN., *Hyperbolic systems of conservation laws: The one dimensional Cauchy problem*, Oxford university press, 200.

[7] L. CARRARO AND J. DUCHON., *Burgers equation with initial conditions with homogeneous and independent increments.*, Ann. Inst. H. Poincaré Anal. Non Lineare, 199, pp. 431–458.

[8] L. CARRARO AND J. DUCHON., *Intrinsic statistical solutions of the burgers equation and levy processes.*, C.R. Math. Acad. Sci. Paris, 319 (1994), pp. 855–858.

[9] E. CHIODAROLI, C. DE LELLIS, AND O. KREML, *Global ill-posedness of the isentropic system of gas dynamics*, Comm. Pure Appl. Math., 68 (2015), pp. 1157–1190, https://doi.org/10.1002/cpa.21537, http://dx.doi.org/10.1002/cpa.21537.

[10] C. M. DAFERMOS, *Hyperbolic conservation laws in continuum physics*, vol. 325 of Grundlehren

der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, fourth ed., 2016, https://doi.org/10.1007/978-3-662-49451-6, http://dx.doi.org/10.1007/978-3-662-49451-6.

[11] C. De Lellis and L. Székelyhidi, Jr., *The Euler equations as a differential inclusion*, Ann. of Math. (2), 170 (2009), pp. 1417–1436, https://doi.org/10.4007/annals.2009.170.1417, http://dx.doi.org/10.4007/annals.2009.170.1417.

[12] R. J. DiPerna, *Measure-valued solutions to conservation laws*, Arch. Rational Mech. Anal., 88 (1985), pp. 223–270, https://doi.org/10.1007/BF00752112, http://dx.doi.org/10.1007/BF00752112.

[13] U. S. Fjordholm, R. Käppeli, S. Mishra, and E. Tadmor, *Construction of approximate entropy measure-valued solutions for hyperbolic systems of conservation laws*, Found. Comput. Math., 17 (2017), pp. 763–827, https://doi.org/10.1007/s10208-015-9299-z, http://dx.doi.org/10.1007/s10208-015-9299-z.

[14] U. S. Fjordholm, S. Lanthaler, and S. Mishra, *Statistical solutions of hyperbolic conservation laws: foundations*, Arch. Ration. Mech. Anal., 226 (2017), pp. 809–849, https://doi.org/10.1007/s00205-017-1145-9, http://dx.doi.org/10.1007/s00205-017-1145-9.

[15] A. Fournier, D. Fussell, and L. Carpenter, *Computer rendering of stochastic models*, Commun. ACM, 25 (1982), pp. 371–384, https://doi.org/10.1145/358523.358553, http://doi.acm.org/10.1145/358523.358553.

[16] N. Fournier and A. Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probability Theory and Related Fields, 162 (2015), pp. 707–738, https://doi.org/10.1007/s00440-014-0583-7, http://dx.doi.org/10.1007/s00440-014-0583-7, https://arxiv.org/abs/1312.2128.

[17] M. B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617, https://doi.org/10.1287/opre.1070.0496, http://dx.doi.org/10.1287/opre.1070.0496.

[18] E. Godlewski and P. Raviart, *Hyperbolic systems of conservation laws*, Mathématiques & applications, Ellipses, 1991, https://books.google.no/books?id=hzvvAAAAMAAJ.

[19] S. Gottlieb, C.-W. Shu, and E. Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM Rev., 43 (2001), pp. 89–112, https://doi.org/10.1137/S003614450036757X, http://dx.doi.org/10.1137/S003614450036757X.

[20] S. Heinrich, *Multilevel monte carlo methods*, Lecture Notes in Computer Science, (2001), pp. 58–67, https://doi.org/10.1007/3-540-45346-6_5.

[21] H. Holden and N. H. Risebro, *Front tracking for hyperbolic conservation laws*, vol. 152 of Applied Mathematical Sciences, Springer, Heidelberg, second ed., 2015, https://doi.org/10.1007/978-3-662-47507-2, http://dx.doi.org/10.1007/978-3-662-47507-2.

[22] 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, https://doi.org/10.1016/0041-5553(76)90046-X, http://www.sciencedirect.com/science/article/pii/004155537690046X.

[23] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Classics in Mathematics, Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.

[24] R. J. LeVeque, *Numerical methods for conservation laws*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, second ed., 1992, https://doi.org/10.1007/978-3-0348-8629-1, http://dx.doi.org/10.1007/978-3-0348-8629-1.

[25] P. Lévy, *Processus stochastiques et mouvement brownien*, Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics], Éditions Jacques Gabay, Sceaux, 1992. Followed by a note by M. Loève, Reprint of the second (1965) edition.

[26] B. B. Mandelbrot and J. W. Van Ness, *Fractional Brownian motions, fractional noises and applications*, SIAM Rev., 10 (1968), pp. 422–437, https://doi.org/10.1137/1010093, http://dx.doi.org/10.1137/1010093.

[27] S. Mishra and C. Schwab, *Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data*, Math. Comp., 81 (2012), pp. 1979–2018, https://doi.org/10.1090/S0025-5718-2012-02574-9, http://dx.doi.org/10.1090/S0025-5718-2012-02574-9.

[28] S. Mishra, C. Schwab, and J. Šukys, *Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions*, J. Comput. Phys., 231 (2012), pp. 3365–3388, https://doi.org/10.1016/j.jcp.2012.01.011, http://dx.doi.org/10.1016/j.jcp.2012.01.011.

[29] C. F. s, O. Manley, R. Rosa, and R. Temam., *Navier–Stokes Equations and Turbulence*, Cambridge University Press, 2001.

[30] Z.-S. She, E. Aurell, and U. Frisch., *The inviscid Burgers equation with initial data of Brownian type*, Comm. Math. Phys., 148 (1992), pp. 623–641.

[31] Y. G. SINAI, *Statistics of shocks in solutions of inviscid burgers equation*, Communications in Mathematical Physics, 148 (1992), pp. 601–621, https://doi.org/10.1007/BF02096550, https://doi.org/10.1007/BF02096550.

[32] C. VILLANI, *Topics in Optimal Transportation*, Graduate Studies in Mathematics, Vol. 58, American Mathematical Society, 2003.

[33] R. F. VOSS, *Random Fractal Forgeries*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1991, pp. 805–835, https://doi.org/10.1007/978-3-642-84574-1_34, http://dx.doi.org/10.1007/978-3-642-84574-1_34.