Statistical solutions of hyperbolic systems of conservation laws: numerical approximation

U. S. Fjordholm

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

K. Lye

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

S. Mishra

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

F. Weber

Department of Mathematical Sciences, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
franzisw@andrew.cmu.edu

Statistical solutions are time-parameterized probability measures on spaces of integrable functions, which have been proposed recently as a framework for global solutions and uncertainty quantification for multi-dimensional hyperbolic system of conservation laws. By combining high-resolution finite volume methods with a Monte Carlo sampling procedure, we present a numerical algorithm to approximate statistical solutions. Under verifiable assumptions on the finite volume method, we prove that the approximations, generated by the proposed algorithm, converge in an appropriate topology to a statistical solution. Numerical experiments illustrating the convergence theory and revealing interesting properties of statistical solutions, are also presented.

1. Introduction

Systems of conservation laws are a large class of nonlinear partial differential equations of the generic form

$$\partial_t u + \nabla_x \cdot f(u) = 0 \tag{1.1a}$$

$$u(x,0) = \bar{u}(x). \tag{1.1b}$$

Here, the unknown $u = u(x,t) : D \times \mathbb{R}_+ \to U$ is the vector of *conserved variables* and $f = (f^1, \dots, f^d) : \mathbb{R}^N \to \mathbb{R}^{N \times d}$ is the *flux function*. Here, we denote $\mathbb{R}_+ := [0, \infty)$ and $U := \mathbb{R}^N$, and we let the physical domain $D \subset \mathbb{R}^d$ be some open, connected set. $\nabla_x \cdot \sigma$ denotes the divergence of a vector field $\sigma(x) = (\sigma^1(x), \dots, \sigma^d(x))$, i.e.,

$\nabla_x \cdot \sigma = \sum_{i=1}^{d} \partial_{x^i} \sigma^i$, where $x = (x^1, \ldots, x^d)$. The system (1.1a) is *hyperbolic* if the flux Jacobian $\partial_u(f \cdot n)$ has real eigenvalues for all $n \in \mathbb{R}^d$ with $|n| = 1$.

Many important models in physics and engineering are described by hyperbolic systems of conservation laws. Examples include the compressible Euler equations of gas dynamics, the shallow water equations of oceanography, the Magneto-Hydro-Dynamics (MHD) equations of plasma physics, and the equations of nonlinear elastodynamics [10].

1.1. *Entropy Solutions.*

It is well known that even if the initial data $\overline{u}$ in (1.1) is smooth, solutions of (1.1) develop discontinuities, such as *shock waves* and *contact discontinuities*, in finite time. Therefore, solutions to (1.1) are sought in the sense of distributions: A function $u \in L^{\infty}(\mathbb{R}^d \times \mathbb{R}_+, \mathbb{R}^N)$ is an *weak solution* of (1.1) if it satisfies

$$\int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \partial_t \varphi(x,t) u(x,t) + \nabla_x \varphi(x,t) \cdot f(u(x,t)) \ dxdt + \int_{\mathbb{R}^d} \varphi(x,0)\overline{u}(x) \ dx = 0 \quad (1.2)$$

for all test functions $\varphi \in C_c^1(\mathbb{R}^d \times \mathbb{R}_+)$.

As weak solutions are not unique [10], it is necessary to augment them with additional admissibility criteria or *entropy conditions* to recover uniqueness. These entropy conditions are based on the existence of a so-called *entropy pair* — a pair of functions $\eta : \mathbb{R}^N \to \mathbb{R}$, $q : \mathbb{R}^N \to \mathbb{R}^d$, with η convex and q satisfying the compatibility condition $q' = \eta' \cdot f'$ (where f' and q' are the Jacobian matrices of f and q). An *entropy solution* of (1.1) is a weak solution that also satisfies the so-called *entropy inequality*

$$\int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \partial_t \varphi(x,t) \eta(u(x,t)) + \nabla_x \varphi(x,t) \cdot q(u(x,t)) \ dxdt + \int_{\mathbb{R}^d} \varphi(x,0)\eta(\overline{u}(x)) \ dx \geqslant 0$$

$$(1.3)$$

for all nonnegative test functions $\varphi \in C_c^1(\mathbb{R}^d \times \mathbb{R}_+)$. Depending on the availability of entropy pairs (η, q), the entropy condition leads to various *a priori* bounds on u: If, say, $\eta(u) = |u|^p$ (or some perturbation thereof) for some $p \geqslant 1$ then (1.3) leads to

$$\int_{\mathbb{R}^d} |u(x,t)|^p \ dx \leqslant \int_{\mathbb{R}^d} |\overline{u}(x)|^p \ dx \qquad \forall \ t > 0, \tag{1.4}$$

see e.g. [10,26,29].

The global well-posedness of entropy solutions of (1.1) has been addressed both for (multi-dimensional) scalar conservation laws [32] and for systems in one space dimension (see [25,4,7,29] and references therein). However, there are no global existence results for entropy solutions of multi-dimensional systems of conservation laws with generic initial data. On the other hand, it has been established recently in [11,9] that entropy solutions for some systems of conservation laws (such as isentropic Euler

equations in two space dimensions) may not be unique. This is a strong indication that the paradigm of entropy solutions is not the correct framework for the well-posedness of multi-dimensional systems of hyperbolic conservation laws.

1.2. *Numerical schemes*

A wide variety of numerical methods have been developed to approximate entropy solutions of (1.1) in a robust and efficient manner. These include finite volume, (conservative) finite difference, discontinuous Galerkin (DG) finite element and spectral (viscosity) methods, see textbooks [26,28] for further details. Rigorous convergence results of numerical methods to entropy solutions are only available for scalar conservation laws (see e.g. [26] for *monotone schemes* and [14] for high-order schemes) and for some specific numerical methods for one-dimensional systems ([25] for Glimm's scheme and [29] for front tracking).

There are no rigorous convergence results to entropy solutions for any numerical schemes approximating multi-dimensional systems of conservation laws. To the contrary, several numerical experiments, such as those presented recently in [15,19], strongly suggest that there is no convergence of approximations generated by standard numerical schemes for (1.1), as the mesh is refined. This has been attributed to the emergence of turbulence-like structures at smaller and smaller scales upon mesh refinement (see Figure 4 of [15]).

1.3. *Measure-valued and Statistical solutions.*

Given the lack of well-posedness of entropy solutions for multi-dimensional systems of conservation laws and the lack of convergence of numerical approximations to them, it is natural to seek alternative solution paradigms for (1.1). A possible solution framework is that of *entropy measure-valued solutions*, first proposed by DiPerna in [13]. Measure-valued solutions are *Young measures* [47], that is, space-time parameterized probability measures on the phase space $\mathbb{R}^N$ of (1.1). Global existence of entropy measure-valued solutions has been considered in [13,8] and in [15,19] where the authors constructed entropy measure-valued solutions by proving convergence of a Monte Carlo type ensemble-averaging algorithm, based on underlying entropy stable finite difference schemes.

Although entropy measure-valued solutions for multi-dimensional systems of conservation laws exist globally, it is well known that they are not necessarily unique; see [41,?] and references therein. In particular, one can even construct multiple entropy measure-valued solutions for scalar conservation laws for the same measure-valued initial data [19]. Although generic measure-valued solutions might not be unique, numerical experiments presented in [15] indicate that measure-valued solutions of (1.1), computed with the ensemble-averaging algorithm of [15], are stable with respect to initial perturbations and to the choice of underlying numerical method. This suggests imposing additional constraints on entropy measure-valued solutions in order to recover uniqueness.

In [16], the authors implicated the lack of information about (multi-point) statistical correlations in Young measures as a possible cause of the non-uniqueness of entropy measure-valued solutions. Consequently, they introduced a *stronger* solution paradigm termed *statistical solutions* for hyperbolic systems of conservation laws (1.1). Statistical solutions are time-parameterized probability measures on some Lebesgue space $L^p(D; U)$ satisfying (1.1a) in an averaged sense. The choice of the exponent $p \geqslant 1$ depends on the available *a priori* bounds for solution of (1.1), such as (1.4). It was shown in [16] that probability measures on $L^p(D; U)$ can be identified with (and indeed are equivalent to) *correlation measures* — a hierarchy of Young measures defined on tensorized versions of the domain D and the phase space U in (1.1). Statistical solutions have also been introduced in the context of the incompressible Navier–Stokes equations by Foiaş et al.; see [21] and references therein.

In [16], the authors defined statistical solutions of systems of conservation laws (1.1a) by requiring that the moments of the time-parameterized probability measure on $L^p(D)$ (or equivalently, of the underlying correlation measure) satisfy an infinite set of (tensorized) partial differential equations (PDEs), consistent with (1.1a).

The first member of the hierarchy of correlation measures for a statistical solution is a (classical) Young measure and it can be shown to be an entropy measure-valued solution of (1.1), in the sense of DiPerna [13]. The k-th member ($k \geqslant 2$) of the hierarchy represents k-point spatial correlations. Thus, a statistical solution can be thought of as a measure-valued solution, augmented with information about all possible (multi-point) spatial correlations [16]. Consequently, statistical solutions contain much more information than measure-valued solutions.

In [16], the authors constructed a *canonical* statistical solution for scalar conservation laws in terms of the data-to-solution semi-group of Kruzkhov [32] and showed that this statistical solution is unique under a suitable entropy condition. Numerical approximation of statistical solutions of scalar conservation laws was considered in [17], where the authors proposed Monte Carlo and multi-level Monte Carlo (MLMC) algorithms to compute statistical solutions and showed their convergence.

1.4. *Aims and scope of this paper.*

Given this background, our main aim in this paper is to study statistical solutions for multi-dimensional systems of conservation laws. To this end, we obtain the following results:

- We propose a Monte Carlo ensemble averaging based algorithm for computing statistical solutions of systems of conservation laws. This algorithm is a variant of the Monte Carlo algorithms presented in [15] and [17].
- Under reasonable assumptions on the underlying numerical scheme, we prove convergence of the ensemble-averaging algorithm to a statistical solution. It is highly non-trivial to identify an appropriate topology on time parameterized probability measures on $L^p(D)$ in order to prove conver-

gence of the computed statistical solutions. To this end, we find a suitable topology and prescribe novel sufficient conditions that ensure convergence in this topology.

- We present several numerical experiments that illustrate the robustness of our proposed algorithm and also reveal interesting properties of statistical solutions of (1.1a).

As a consequence of our convergence theorem, we establish a *conditional* global existence result for multi-dimensional systems of conservation laws. Moreover, we also propose an entropy condition under which we prove a *weak-strong* uniqueness result for statistical solutions, that is, we prove that if there exists a statistical solution of sufficient regularity (in a sense made precise in Section 3), then all entropy statistical solutions agree with it.

The rest of the paper is organized as follows: In Section 2, we provide the mathematical framework by describing the concepts of correlation measures and statistical solutions. We also provide characterizations of the topology on probability measures on $L^p(D)$, in which our subsequent numerical approximations will converge. The entropy condition and the weak-strong uniqueness of statistical solutions are presented in Section 3 and the Monte Carlo ensemble-averaging algorithm (and its convergence) is presented in Section 4. Numerical experiments are presented in Section 5 and the results of the paper are summarized and discussed in Section 6.

2. Probability measures on $L^p(D; U)$ and Statistical solutions

In the usual, deterministic interpretation of (1.1a), one attempts to find a function $u = u(t) : D \to U$ satisfying (1.1a) in a weak or strong sense. (Here, as in the introduction, we let $D \subset \mathbb{R}^d$ be an open, connected set and we denote $U := \mathbb{R}^N$.) By contrast, a statistical solution of (1.1a) is a probability measure $\mu = \mu_t$ distributed over such functions u and satisfying (1.1a) in an averaged sense. Solutions of (1.1a) are most naturally found in (a subspace of) $L^p(D; U)$, so μ_t is required to be a probability measure on $L^p(D; U)$ at each time t. In order to write down constitutive equations for μ, it is more natural to work with finite-dimensional projections or *marginals* of μ; these are the so-called *correlation measures* [16]. In this section we provide a self-contained description of correlation measures, probability measures over $L^p(D; U)$, and statistical solutions of (1.1a).

In order to link probability measures on L^p to their finite-dimensional marginals, we prove in Section 2.1 that a sequence of such measures converges weakly if and only if it converges with respect to a certain class $\mathcal{C}_p$ of finite-dimensional observables. In Section 2.2 we introduce correlation measures and we show that these are in a one-to-one relationship with probability measures over L^p, and that they are linked precisely through the finite-dimensional observables $\mathcal{C}_p$. We also prove a compactness result for sequences of correlation measures. In Section 2.3 we treat time-parametrized probability and correlation measures, and we prove measurability and compactness results. Finally, in Section 2.4 we provide the definition of

statistical solutions of (1.1a).

For the sake of clarity, many of the proofs in this rather technical section have been moved to Appendices Appendix A, Appendix B and Appendix C.

Notation 2.1. If X is a topological space, then we let $\mathcal{B}(X)$ denote the Borel σ-algebra on X, we let $\mathcal{M}(X)$ denote the set of signed Radon measures on $(X, \mathcal{B}(X))$, and we let $\mathcal{P}(X) \subset \mathcal{M}(X)$ denote the set of all probability measures on $(X, \mathcal{B}(X))$, i.e., all non-negative $\mu \in \mathcal{M}(X)$ with $\mu(X) = 1$ (see e.g. [2,5,31]). For $k \in \mathbb{N}$ and a multiindex $\alpha \in \{0, 1\}^k$ we write $|\alpha| = \alpha_1 + \cdots + \alpha_k$ and $\bar{\alpha} = \mathbb{1} - \alpha = (1 - \alpha_1, \ldots, 1 - \alpha_k)$, and we let x_α be the vector of length $|\alpha|$ consisting of the elements x_i of x for which α_i is non-zero. For a vector $x = (x_1, \ldots, x_k)$ we write $\hat{x}^i = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$. For a vector $\xi = (\xi_1, \ldots, \xi_k)$ we write $|\xi^\alpha| = |\xi_1|^{\alpha_1} \cdots |\xi_k|^{\alpha_k}$ with the convention $\xi_{\alpha_i} = 1$ if $\alpha_i = 0$.

2.1. *Probability measures on $L^p(D)$ and weak convergence*

If X is any topological space and we are given a sequence $\mu_1, \mu_2, \cdots \in \mathcal{P}(X)$ and some $\mu \in \mathcal{P}(X)$, then we say that $\{\mu_n\}_{n \in \mathbb{N}}$ *converges weakly to μ*, written $\mu_n \rightharpoonup \mu$, if

$$\langle \mu_n, F \rangle \to \langle \mu, F \rangle \qquad \text{as } n \to \infty \tag{2.1}$$

for every $F \in C_b(X)$. (Here and elsewhere, $\langle \mu, F \rangle = \int_X F(x) \, d\mu(x)$ denotes the expectation of F with respect to μ.) We will be particularly interested in the case $X = L^p(D; U)$, so to study weak convergence in this space we need to work with the space $C_b(L^p(D; U))$. In this section we will see that it is sufficient to prove (2.1) for a much smaller class of functionals F, namely those which depend only on finite-dimensional projections of $u \in L^p(D; U)$.

If E and V are Euclidean spaces then a measurable function $g : E \times V \to \mathbb{R}$ is called a *Carathéodory function* if $\xi \mapsto g(x, \xi)$ is continuous for a.e. $x \in E$ and $x \mapsto g(x, \xi)$ is measurable for every $\xi \in V$ (see e.g. [1]). For a number $k \in \mathbb{N}$ and a Carathéodory function $g = g(x, \xi) : D^k \times U^k \to \mathbb{R}$ we define the functional $L_g : L^p(D; U) \to \mathbb{R}$ by

$$L_g(u) := \int_{D^k} g(x_1, \ldots, x_k, u(x_1), \ldots, u(x_k)) \, dx. \tag{2.2}$$

(Here, D^k denotes the product space $D^k = D \times \cdots \times D$, and similarly for U^k.) The above integral is clearly not well-defined for every Carathéodory function g, so we restrict our attention to the following class.

Definition 2.1. For every $k \in \mathbb{N}$, we let $\mathcal{H}^{k,p}(D; U)$ denote the space of Carathéodory functions $g : D^k \times U^k \to \mathbb{R}$ satisfying

$$|g(x, \xi)| \leqslant \sum_{\alpha \in \{0,1\}^k} \varphi_{|\bar{\alpha}|}(x_{\bar{\alpha}}) |\xi^\alpha|^p \qquad \forall \, x \in D^k, \, \xi \in U^k \tag{2.3}$$

for nonnegative functions $\varphi_i \in L^1(D^i)$, $i = 0, 1, \ldots, k$ (with the convention that $L^1(D^0) \cong \mathbb{R}$; see also Example 2.1). We let $\mathcal{H}_1^{k,p}(D; U) \subset \mathcal{H}^{k,p}(D; U)$ denote the subspace of functions g which are locally Lipschitz continuous, in the sense that there is some $r > 0$ and some nonnegative $h \in \mathcal{H}^{k-1,p}(D; U)$ such that

$$\left| g(x, \zeta) - g(y, \xi) \right| \leqslant \sum_{i=1}^k |\zeta_i - \xi_i| \max \left(|\xi_i|, |\zeta_i| \right)^{p-1} h(\hat{x}^i, \hat{\xi}^i) \tag{2.4}$$

for every $x \in D^k$, $y \in B_r(x)$ and $\xi, \zeta \in U^k$. Last, we denote

$$\mathcal{C}^p(D; U) := \left\{ L_g \ : \ g \in \mathcal{H}^{k,p}(D; U), \ k \in \mathbb{N} \right\}$$
$$\mathcal{C}_1^p(D; U) := \left\{ L_g \ : \ g \in \mathcal{H}_1^{k,p}(D; U), \ k \in \mathbb{N} \right\}$$

where L_g is defined in (2.2).

Example 2.1. For $k = 1$ the condition (2.3) asserts that

$$|g(x, \xi)| \leqslant \varphi_1(x) + \varphi_0 |\xi|^p$$

for $0 \leqslant \varphi_1 \in L^1(D)$ and $\varphi_0 \in [0, \infty)$, and for $k = 2$ that

$$|g(x_1, x_2, \xi_1, \xi_2)| \leqslant \varphi_2(x_1, x_2) + \varphi_1(x_1)|\xi_2|^p + \varphi_1(x_2)|\xi_1|^p + \varphi_0 |\xi_1|^p |\xi_2|^p$$

for $0 \leqslant \varphi_2 \in L^1(D^2)$, $0 \leqslant \varphi_1 \in L^1(D)$ and $\varphi_0 \in [0, \infty)$.

We will simply denote $\mathcal{H}^{k,p} = \mathcal{H}^{k,p}(D; U)$, etc. when the domain and image D, U are clear from the context.

Lemma 2.1. *Every functional $L_g \in \mathcal{C}^p$ is well-defined and finite on $L^p(D; U)$. Every functional $L_g \in \mathcal{C}_1^p$ is continuous and is Lipschitz continuous on bounded subsets of $L^p(D; U)$.*

Theorem 2.2. *Let $\mu_n, \mu \in \mathcal{P}(L^p(D; U))$ for $n \in \mathbb{N}$ satisfy $\operatorname{supp} \mu, \operatorname{supp} \mu_n \subset B$ for all $n \in \mathbb{N}$, for some bounded set $B \subset L^p(D; U)$. Then $\mu_n \rightharpoonup \mu$ if and only if $\langle \mu_n, F \rangle \to \langle \mu, F \rangle$ for all $F \in \mathcal{C}_1^p(D; U)$.*

The proofs of the above results can be found in Appendix Appendix A. The "only if" part of Theorem 2.2 is trivial, since every $F \in \mathcal{C}_1^p$ belongs to C_b when restricted to a bounded set; the converse relies on an approximation argument found in [2].

2.2. *Correlation measures*

In short, a *correlation measure* prescribes the joint distribution of some uncertain quantity u at any finite collection of spatial points $x_1, \ldots, x_k$. Below we provide the rigorous definition of correlation measures and then state the result from [16] on the equivalence between correlation measures and probability measures on $L^p(D; U)$.

We denote $\mathcal{H}_0^k(D;U) := L^1\big(D^k, C_0(U^k)\big)$. By identifying the expressions $g(x)(\xi)$ and $g(x,\xi)$, we can view $\mathcal{H}_0^k(D;U)$ as a subspace of $\mathcal{H}^{k,p}(D;U)$ for any $p \geqslant 1$ (with the choice $\varphi_0, \ldots, \varphi_{k-1} \equiv 0$ and $\varphi_k(x) = \|g(x)\|_{C_0(U^k)}$ in (2.3)).

Theorem 2.3. *The dual of $\mathcal{H}_0^k(D;U)$ is the space $\mathcal{H}_0^{k*}(D;U) := L_w^\infty\big(D^k, \mathcal{M}(U^k)\big)$, the space of bounded, weak* measurable maps from D^k to $\mathcal{M}(U^k)$, under the duality pairing*

$$\big\langle \nu^k, g \big\rangle_{\mathcal{H}^k} = \int_{D^k} \big\langle \nu_x^k, g(x) \big\rangle \, dx$$

(where $\big\langle \nu_x^k, g(x) \big\rangle = \int_{U^k} g(x,\xi) \, d\nu_x^k(\xi)$ is the usual duality pairing between Radon measures $\mathcal{M}(U^k)$ and continuous functions $C_b(U^k)$).

For more details and references for the above result, see [3].

Definition 2.2 (Fjordholm, Lanthaler, Mishra [16]). A *correlation measure* is a collection $\boldsymbol{\nu} = (\nu^1, \nu^2, \ldots)$ of maps $\nu^k \in \mathcal{H}_0^{k*}(D;U)$ satisfying for all $k = 1, 2, \ldots$

(i) $\nu_x^k \in \mathcal{P}(U^k)$ for a.e. $x \in D^k$, and the map $x \mapsto \big\langle \nu_x^k, f \big\rangle$ is measurable for every $f \in C_b(U^k)$. (In other words, ν^k is a Young measure from D^k to U^k.)

(ii) *Symmetry:* if σ is a permutation of $\{1, \ldots, k\}$ and $f \in C_0(U^k)$ then $\big\langle \nu_{\sigma(x)}^k, f(\sigma(\xi)) \big\rangle = \big\langle \nu_x^k, f(\xi) \big\rangle$ for a.e. $x \in D^k$.

(iii) *Consistency:* If $f \in C_b(U^k)$ is of the form $f(\xi_1, \ldots, \xi_k) = g(\xi_1, \ldots, \xi_{k-1})$ for some $g \in C_0(U^{k-1})$, then $\big\langle \nu_{x_1,\ldots,x_k}^k, f \big\rangle = \big\langle \nu_{x_1,\ldots,x_{k-1}}^{k-1}, g \big\rangle$ Lebesgue-a.e. $(x_1, \ldots, x_k) \in D^k$.

(iv) L^p *integrability:*

$$\int_D \big\langle \nu_x^1, |\xi|^p \big\rangle \, dx < +\infty. \tag{2.5}$$

(v) *Diagonal continuity:* $\lim_{r \to 0} \omega_r^p(\nu^2) = 0$, where

$$\omega_r^p(\nu^2) := \int_D \fint_{B_r(x)} \big\langle \nu_{x,y}^2, |\xi_1 - \xi_2|^p \big\rangle \, dydx. \tag{2.6}$$

Each element ν^k will be called a *correlation marginal*. The functional ω_r^p is called the *modulus of continuity* of $\boldsymbol{\nu}$. We let $\mathcal{L}^p(D;U)$ denote the set of all correlation measures.

The next result shows that there is a duality relation between correlation marginals and the probability measures $\mu \in \mathcal{P}(L^p)$ discussed in the previous section.

Theorem 2.4 (Fjordholm, Lanthaler, Mishra [16]). *For every correlation measure $\boldsymbol{\nu} \in \mathcal{L}^p(D;U)$ there is a unique probability measure $\mu \in \mathcal{P}(L^p(D;U))$ whose p-th moment is finite,*

$$\int_{L^p} \|u\|_{L^p}^p \, d\mu(u) < \infty \tag{2.7}$$

and such that μ is dual to $\boldsymbol{\nu}$: the identity

$$\int_{D^k} \langle \nu^k, g(x) \rangle \, dx = \int_{L^p} \int_{D^k} g(x, u(x)) \, dx d\mu(u) \tag{2.8}$$

holds for every $g \in \mathcal{H}_0^k(D; U)$ and all $k \in \mathbb{N}$. Conversely, for every $\mu \in \mathcal{P}(L^p(D; U))$ satisfying (2.7) there is a unique correlation measure $\boldsymbol{\nu} \in \mathcal{L}^p(D; U)$ that is dual to μ.

Remark 2.1. By using Lebesgue's dominated convergence theorem, it is not hard to show that the identity (2.8) can be extended to all $g \in \mathcal{H}^{k,p}(D; U)$, as long as both integrals are well-defined. In particular, this is true if μ is supported on a bounded subset of $L^p(D; U)$.

Later on, we will be particularly interested in those $\mu \in \mathcal{P}(L^p)$ that have bounded support. The following lemma shows how the property of having bounded support can be expressed in terms of the corresponding correlation measure.

Lemma 2.2. *Let $\boldsymbol{\nu} \in \mathcal{L}^p(D; U)$ and $\mu \in \mathcal{P}(L^p(D; U))$ be dual to one another. Then*

$$\operatorname*{ess\,sup}_{u \in L^p} \|u\|_{L^p} = \limsup_{k \to \infty} \left(\int_{D^k} \langle \nu_x^k, |\xi_1|^p \cdots |\xi_k|^p \rangle \, dx \right)^{1/kp} \tag{2.9}$$

where the "ess sup" is taken with respect to μ.

Proof. From the identity $\|f\|_{L^\infty(X;\mu)} = \lim_{k \to \infty} \|f\|_{L^k(X;\mu)}$, valid for any finite measure μ, we get

$$\operatorname*{ess\,sup}_{u \in L^p} \|u\|_{L^p(D;U)}^p = \lim_{k \to \infty} \left(\int_{L^p(D;U)} \|u\|_{L^p(D;U)}^{pk} \, d\mu(u) \right)^{1/k}$$

$$= \lim_{k \to \infty} \left(\int_{L^p(D;U)} \int_{D^k} |u(x_1)|^p \cdots |u(x_k)|^p \, dx \, d\mu(u) \right)^{1/k}$$

$$= \lim_{k \to \infty} \left(\int_{D^k} \langle \nu_x^k, |\xi_1|^p \cdots |\xi_k|^p \rangle \, dx \right)^{1/k}. \qquad \square$$

Definition 2.3. We let $\mathcal{L}_b^p(D; U)$ denote the subset of correlation measures $\boldsymbol{\nu} \in \mathcal{L}^p(D; U)$ with bounded support, in the sense that there is an $M > 0$ such that

$$\limsup_{k \to \infty} \left(\int_{D^k} \langle \nu_x^k, |\xi_1|^p \cdots |\xi_k|^p \rangle \, dx \right)^{1/kp} \leqslant M. \tag{2.10}$$

Definition 2.4. If $\boldsymbol{\nu}_n, \boldsymbol{\nu} \in \mathcal{L}^p(D; U)$ for $n \in \mathbb{N}$ then we say that $\boldsymbol{\nu}_n$ converges weak* to $\boldsymbol{\nu}$ as $n \to \infty$ (written $\boldsymbol{\nu}_n \overset{*}{\rightharpoonup} \boldsymbol{\nu}$) if $\nu_n^k \overset{*}{\rightharpoonup} \nu^k$ as $n \to \infty$, that is, if $\langle \nu_n^k, g \rangle_{\mathcal{H}^k} \to \langle \nu^k, g \rangle_{\mathcal{H}^k}$ for all $g \in \mathcal{H}_0^k(D; U)$ and all $k \in \mathbb{N}$.

If $\boldsymbol{\nu}_n, \boldsymbol{\nu} \in \mathcal{L}_b^p(D; U)$ for $n \in \mathbb{N}$ then we say that $(\boldsymbol{\nu}_n)_{n \in \mathbb{N}}$ converges weakly to $\boldsymbol{\nu}$ as $n \to \infty$ (written $\boldsymbol{\nu}_n \rightharpoonup \boldsymbol{\nu}$) if $\langle \nu_n^k, g \rangle_{\mathcal{H}^k} \to \langle \nu^k, g \rangle_{\mathcal{H}^k}$ for every $g \in \mathcal{H}_1^{k,p}(D; U)$.

Note that $\boldsymbol{\nu} \in \mathcal{L}_b^p$ implies that $\langle \nu^k, g \rangle_{\mathcal{H}^k}$ is well-defined and finite for any $g \in \mathcal{H}^{k,p}$ (cf. Definition 2.1).

We next show a compactness result which can be thought of as Kolmogorov's compactness theorem (cf. [29]) for correlation measures.

Theorem 2.5. *Let $\boldsymbol{\nu}_n \in \mathcal{L}^p(D; U)$ for $n = 1, 2, \ldots$ be a sequence of correlation measures such that*

$$\sup_{n \in \mathbb{N}} \langle \nu_n^1, |\xi|^p \rangle_{\mathcal{H}^1} \leqslant c^p \tag{2.11}$$

$$\lim_{r \to 0} \limsup_{n \to \infty} \omega_r^p\big(\nu_n^2\big) = 0 \tag{2.12}$$

for some $c > 0$ (where ω_r^p is defined in Definition 2.2(v)). Then there exists a subsequence $(n_j)_{j=1}^\infty$ and some $\boldsymbol{\nu} \in \mathcal{L}^p(D; U)$ such that

(i) $\boldsymbol{\nu}_{n_j} \overset{*}{\rightharpoonup} \boldsymbol{\nu}$ *as $j \to \infty$, that is, $\langle \nu_{n_j}^k, g \rangle_{\mathcal{H}^k} \to \langle \nu^k, g \rangle_{\mathcal{H}^k}$ for every $g \in \mathcal{H}_0^k(D; U)$ and every $k \in \mathbb{N}$*

(ii) $\langle \nu^1, |\xi|^p \rangle_{\mathcal{H}^1} \leqslant c^p$

(iii) $\omega_r^p(\nu^2) \leqslant \liminf_{n \to \infty} \omega_r^p(\nu_n^2)$ *for every $r > 0$*

(iv) *for $k \in \mathbb{N}$, let $\varphi \in L_{loc}^1(D^k)$ and $\kappa \in C(U^k)$ be nonnegative, and let $g(x, \xi) := \varphi(x)\kappa(\xi)$. Then*

$$\langle \nu^k, g \rangle_{\mathcal{H}^k} \leqslant \liminf_{j \to \infty} \langle \nu_{n_j}^k, g \rangle_{\mathcal{H}^k}. \tag{2.13}$$

(v) *Assume moreover that the domain $D \subset \mathbb{R}^d$ is bounded and that $\boldsymbol{\nu}_n$ have uniformly bounded support, in the sense that (2.10) holds for all $\boldsymbol{\nu}_n$ for a fixed $M > 0$, or equivalently,*

$$\|u\|_{L^p} \leqslant M \qquad \text{for } \mu_n\text{-a.e. } u \in L^p(D; U) \text{ for every } n \in \mathbb{N}. \tag{2.14}$$

Then observables converge strongly:

$$\lim_{j \to \infty} \int_{D^k} \big| \langle \nu_{n_j, x}^k, g(x) \rangle - \langle \nu_x^k, g(x) \rangle \big| \, dx = 0 \tag{2.15}$$

for every $g \in \mathcal{H}_1^{k,p}(D; U)$. In particular, $\boldsymbol{\nu}_{n_j} \rightharpoonup \boldsymbol{\nu}$.

The proof is given in Appendix Appendix B.

Remark 2.2. (2.15) implies in particular that $\langle \nu_{n_j}^k, g \rangle_{\mathcal{H}^k} = \langle \mu_{n_j}, L_g \rangle$ converges for any $g \in \mathcal{H}_1^{k,p}$, where $\mu_n \in \mathcal{P}(L^p)$ is dual to $\boldsymbol{\nu}_n$ (see Theorem 2.4). By Theorem 2.2, this is equivalent to saying that μ_{n_j} converges weakly to μ. Since, by hypothesis, the pth moment of μ_n is uniformly bounded, the sequence μ_n converges to μ in the Wasserstein distance; see Definition 3.1 and [44] .

Remark 2.3. Theorem 2.5 can most likely be extended to provide a complete characterization of compact subsets of $\mathcal{L}^p(D; U)$. Since we only require sufficient conditions for compactness, we do not pursue this generalization here.

2.3. *Time-parameterized probability measures on L^p*

Let $T \in (0, \infty]$. To take into account the evolutionary nature of the PDE (1.1a), we will add time-dependence to the probability measures considered in Section 2.1 by considering maps $\mu : [0, T) \to \mathcal{P}(L^p(D; U))$. Note the distinction between time-parametrized maps $\mu : [0, T) \to \mathcal{P}(L^p(D; U))$ and probability measures γ on, say, the space $L^\infty([0, T); L^p(D; U))$. Every such measure γ would correspond to a unique μ, but not *vice versa*; when "projecting" γ onto μ, any information about correlation between function values $u(t_1)$, $u(t_2)$ at different times t_1, t_2 is lost. Given the evolutionary nature of the PDE (1.1a), we have chosen to work with "μ" measures in order to preserve the direction of time in the underlying PDE.

Notation 2.6. We denote the set of Carathéodory functions depending on space and time by $\mathcal{H}_0^k([0, T), D; U) := L^1([0, T) \times D^k; C_0(U^k))$ and its dual space by $\mathcal{H}_0^{k*}([0, T), D; U) := L_w^\infty([0, T) \times D^k; \mathcal{M}(U^k))$.

Analogously to Definition 2.1, we let $\mathcal{H}^{k,p}([0, T), D; U)$ denote the space of Carathéodory functions $g : [0, T) \times D^k \times U^k \to \mathbb{R}$ satisfying

$$|g(t, x, \xi)| \leqslant \sum_{\alpha \in \{0,1\}^k} \varphi_{|\bar{\alpha}|}(t, x_{\bar{\alpha}})|\xi^\alpha|^p \qquad \forall \, x \in D^k, \, \xi \in U^k \tag{2.16}$$

for nonnegative functions $\varphi_i \in L^\infty([0, T); L^1(D^i))$, $i = 0, 1, \dots, k$. We let $\mathcal{H}_1^{k,p}([0, T), D; U) \subset \mathcal{H}^{k,p}([0, T), D; U)$ denote the subspace of functions g satisfying the local Lipschitz condition

$$\left| g(t, x, \zeta) - g(t, y, \xi) \right| \leqslant \psi(t) \sum_{i=1}^k |\zeta_i - \xi_i| \max \left(|\xi_i|, |\zeta_i| \right)^{p-1} h(t, \hat{x}^i, \hat{\xi}^i) \tag{2.17}$$

for every $x \in D^k$, $y \in B_r(x)$ for some $r > 0$, for some nonnegative $h \in \mathcal{H}^{k-1,p}([0, T), D; U)$ and $0 \leqslant \psi(t) \in L^\infty([0, T))$.

The following lemma shows that it is meaningful to "evaluate" an element $\nu^k \in \mathcal{H}_0^{k*}([0, T), D; U)$ at (almost) any time $t \in [0, T)$.

Lemma 2.3. *Let $\nu^k \in \mathcal{H}_0^{k*}([0, T), D; U)$. Then there exists a map $\rho : [0, T) \to \mathcal{H}_0^{k*}(D; U)$, uniquely defined for a.e. $t \in [0, T)$, such that $t \mapsto \left\langle \rho(t), g \right\rangle_{\mathcal{H}_0^k}$ is measurable for all $g \in \mathcal{H}_0^k(D; U)$, and*

$$\left\langle \nu^k, g \right\rangle_{\mathcal{H}^k} = \int_0^T \left\langle \rho(t), g(t, \cdot) \right\rangle_{\mathcal{H}^k} dt \qquad \forall \, g \in \mathcal{H}_0^k([0, T), D; U).$$

The proof of this lemma is given in Appendix C. Henceforth, we will not make distinctions between these two representations of elements of $\mathcal{H}_0^{k*}([0, T), D; U)$, and denote them both by ν^k.

Definition 2.5. A *time-dependent correlation measure* is a collection $\boldsymbol{\nu} = (\nu^1, \nu^2, \dots)$ of maps $\nu^k \in \mathcal{H}_0^{k*}([0, T), D; U)$ such that

(i) $(\nu_t^1, \nu_t^2, \dots) \in \mathcal{L}^p(D; U)$ for a.e. $t \in [0, T)$

(ii) L^p *integrability:*

$$\underset{t \in [0,T)}{\text{ess sup}} \int_D \left\langle \nu_{t,x}^1, |\xi|^p \right\rangle dx \leqslant c^p < +\infty \tag{2.18}$$

(iii) *Diagonal continuity (DC):*

$$\int_0^{T'} \omega_r^p \left(\nu_t^2 \right) dt \to 0 \qquad \text{as } r \to 0 \text{ for all } T' \in (0,T) \tag{2.19}$$

where ω_r^p was defined in (2.6).

We denote the set of all time-dependent correlation measures by $\mathcal{L}^p([0,T), D; U)$.

Remark 2.4. By Lemma 2.3, the objects ν_t^k are well-defined for a.e. $t \in [0,T)$. Assertion (ii) requires that the L^p bound should be uniform in t, and assertion (iii) requires that the modulus of continuity in the diagonal continuity requirement should be integrable in t.

Next, we prove a time-dependent version of the duality result Theorem 2.4.

Theorem 2.7. *For every time-dependent correlation measure $\boldsymbol{\nu} \in \mathcal{L}^p([0,T), D; U)$ there is a unique (up to subsets of $[0,T)$ of Lebesgue measure 0) map $\mu : [0,T) \to \mathcal{P}(L^p(D; U))$ such that*

(i) the map

$$t \mapsto \left\langle \mu_t, L_g \right\rangle = \int_{L^p} \int_{D^k} g(x, u(x)) \, dx d\mu_t(u) \tag{2.20}$$

is measurable for all $g \in \mathcal{H}_0^k(D; U)$,
(ii) μ is L^p-bounded:

$$\underset{t \in [0,T)}{\text{ess sup}} \int_{L^p} \|u\|_{L^p}^p \, d\mu_t(u) \leqslant c^p < \infty \tag{2.21}$$

(iii) μ is dual to $\boldsymbol{\nu}$: the identity

$$\int_{D^k} \left\langle \nu_t^k, g(x) \right\rangle dx = \int_{L^p} \int_{D^k} g(x, u(x)) \, dx d\mu_t(u) \tag{2.22}$$

holds for a.e. $t \in [0,T)$, every $g \in \mathcal{H}_0^k(D; U)$ and all $k \in \mathbb{N}$.

Conversely, for every $\mu : [0,T) \to \mathcal{P}(L^p(D; U))$ satisfying (i) and (ii), there is a unique correlation measure $\boldsymbol{\nu} \in \mathcal{L}^p([0,T), D; U)$ satisfying (iii).

Proof. Let $\boldsymbol{\nu}$ be given. Then for a.e. $t \in [0,T)$ we have $\boldsymbol{\nu}_t := (\nu_t^1, \nu_t^2, \dots) \in \mathcal{L}^p(D; U)$, so by Theorem 2.4 there exists a unique $\mu_t \in \mathcal{P}(L^p(D; U))$ that is dual to $\boldsymbol{\nu}_t$, in the sense that (iii) holds. From the previous remark we know that $t \mapsto \left\langle \nu_t^k, g \right\rangle_{\mathcal{H}^k}$ is measurable for every $g \in \mathcal{H}_0^k(D; U)$, which (using (iii)) is precisely (i). Property (ii) follows by approximating $\xi \mapsto |\xi|^p$ by functions in $C_0(U)$.

Conversely, given μ satisfying (i) and (ii), Theorem 2.4 gives, for a.e. $t \in [0, T)$, the existence and uniqueness of $\boldsymbol{\nu}_t \in \mathcal{L}^p(D; U)$ satisfying (iii) as well as the L^p-bound (2.18). We claim that $(\boldsymbol{\nu}_t)_{t \in [0,T)}$ defines a time-dependent correlation measure $\boldsymbol{\nu} \in \mathcal{L}^p([0, T), D; U)$. Indeed, define the linear functional ν^k by

$$\langle \nu^k, \theta \otimes g \rangle_{\mathcal{H}^k} := \int_0^T \theta(t) \langle \nu_t^k, g \rangle_{\mathcal{H}^k} \, dt \qquad \forall \, \theta \in L^1([0, T)), \ g \in \mathcal{H}_0^k(D; U), \ k \in \mathbb{N}.$$

Then ν^k is well-defined on tensor product test functions $\theta(t)g(x)$, and

$$\left| \langle \nu^k, \theta \otimes g \rangle_{\mathcal{H}^k} \right| \leqslant \|\theta\|_{L^1([0,T))} \left\| \langle \mu_\cdot, L_g \rangle \right\|_{L^\infty([0,T))} \leqslant \|\theta\|_{L^1} \|L_g\|_{C^0(L^p)}$$

$$= \|\theta\|_{L^1} \|g\|_{\mathcal{H}_0^k} = \|\theta \otimes g\|_{L^1([0,T) \times D^k; C_0(U^k))}.$$

Extending ν^k by linearity to all of $L^1([0, T) \times D^k; C_0(U^k))$ produces a unique element $\nu^k \in L^1([0, T) \times D^k; C_0(U^k))^* \cong L_w^\infty([0, T) \times D^k; \mathcal{M}(U^k))$. Defining the collection $\boldsymbol{\nu} = (\nu^1, \nu^2, \dots)$, it only remains to show that ν^2 satisfies the diagonal continuity requirement (2.19). Indeed, since

$$\omega_r^p(\nu_t^2) = \int_{L^p} \omega_r^p(u) \, d\mu_t(u) \to 0 \qquad \text{as } r \to 0$$

for a.e. $t \in [0, T)$, the requirement (2.19) follows from the dominated convergence theorem. $\qquad \square$

We denote the set of all maps $\mu : [0, T) \to \mathcal{P}(L^p(D; U))$ that are dual to some $\boldsymbol{\nu} \in \mathcal{L}^p([0, T), D; U)$ as $\mathcal{P}_T(L^p(D; U))$.

We conclude this section by proving a version of the compactness theorem for time-dependent correlation measures.

Theorem 2.8. *Let $\boldsymbol{\nu}_n \in \mathcal{L}^p([0, T), D; U)$ for $n = 1, 2, \dots$ be a sequence of correlation measures such that*

$$\sup_{n \in \mathbb{N}} \operatorname*{ess\,sup}_{t \in [0,T)} \int_D \langle \nu_{n;t,x}^1, |\xi|^p \rangle \, dx \leqslant c^p < +\infty \tag{2.23}$$

$$\lim_{r \to 0} \limsup_{n \to \infty} \int_0^{T'} \omega_r^p(\nu_{n,t}^2) \, dt = 0 \tag{2.24}$$

for some $c > 0$ and all $T' \in [0, T)$. Then there exists a subsequence $(n_j)_{j=1}^\infty$ and some $\boldsymbol{\nu} \in \mathcal{L}^p([0, T), D; U)$ such that

(i) *$\boldsymbol{\nu}_{n_j} \overset{*}{\rightharpoonup} \boldsymbol{\nu}$ as $j \to \infty$, that is, $\langle \nu_{n_j}^k, g \rangle_{\mathcal{H}^k} \to \langle \nu^k, g \rangle_{\mathcal{H}^k}$ for every $g \in \mathcal{H}_0^k([0, T), D; U)$ and every $k \in \mathbb{N}$*

(ii) *$\langle \nu_t^1, |\xi|^p \rangle_{\mathcal{H}^1} \leqslant c^p$ for a.e. $t \in [0, T)$*

(iii) *$\int_0^{T'} \omega_r^p(\nu_t^2) \, dt \leqslant \liminf_{n \to \infty} \int_0^{T'} \omega_r^p(\nu_{n,t}^2) \, dt$ for every $r > 0$ and $T' \in [0, T)$*

(iv) *for $k \in \mathbb{N}$, let $\varphi \in L_{loc}^1([0, T) \times D^k)$ and $\kappa \in C(U^k)$ be nonnegative, and let $g(t, x, \xi) := \varphi(t, x)\kappa(\xi)$. Then*

$$\langle \nu^k, g \rangle_{\mathcal{H}^k} \leqslant \liminf_{j \to \infty} \langle \nu_{n_j}^k, g \rangle_{\mathcal{H}^k}. \tag{2.25}$$

14   *U. S. Fjordholm, K. Lye, S. Mishra and F. Weber*

(v) *Assume moreover that $D \subset \mathbb{R}^d$ is bounded, $T < \infty$ and that $\boldsymbol{\nu}_n$ have uniformly bounded support, in the sense that*

$$\|u\|_{L^p} \leqslant M \qquad \text{for } \mu_t^n\text{-a.e. } u \in L^p(D; U) \text{ for every } n \in \mathbb{N}, \text{ a.e } t \in (0, T), \tag{2.26}$$

with $\mu_t^n \in \mathcal{P}_T(L^p(D; U))$ being dual to $\boldsymbol{\nu}_n$, then the following observables converge strongly:

$$\lim_{j \to \infty} \int_{D^k} \left| \int_0^T \left\langle \nu_{n_j; t, x}^k - \nu_{t, x}^k, g(t, x) \right\rangle dt \right| dx = 0 \tag{2.27}$$

for every $g \in \mathcal{H}_1^{k, p}([0, T), D; U)$.

We skip the proof of this theorem as is very similar to that of Theorem 2.5.

Remark 2.5. A closer look at the convergence statement (2.27) reveals that we can expect pointwise a.e. convergence in space of the ensemble averages of the observables $g \in \mathcal{H}_1^{k, p}([0, T), D; U)$. On the other hand, time averaging in (2.27) seems essential. In other words, we have convergence of time averages of ensemble averages of the observables.

2.4. *Statistical solutions*

Using correlation measures we can now define statistical solutions of (1.1a). We need the following assumptions on the flux function in (1.1a),

$$|f(u)| \leqslant C(1 + |u|^p) \qquad \forall\, u \in U,$$
$$|f(u) - f(v)| \leqslant C|u - v| \max(|u|, |v|)^{p-1}, \qquad \forall\, u, v \in U. \tag{2.28}$$

for some constant $C > 0$ and $1 \leqslant p < \infty$. The value of p is given by available *a priori* bounds for solutions of (1.1a), for instance, from the entropy condition (1.3) (cf. (1.4)). For example, both the shallow water equations and the isentropic Euler equations are L^2-bounded, at least for solutions away from vacuum [10].

Statistical solutions are correlation measures (or equivalently, probability measures over L^p) satisfying the differential equation (1.1a) in a certain averaged sense. The full derivation can be found in [16], and we only provide the definition here.

Definition 2.6. Let $\bar{\mu} \in \mathcal{P}(L^p(D; U))$ have bounded support,

$$\|u\|_{L^p(D; U)} \leqslant M \qquad \text{for } \bar{\mu}\text{-a.e. } u \in L^p(D; U) \tag{2.29}$$

for some $M > 0$. A *statistical solution* of (1.1a) with initial data $\bar{\mu}$ is a time-dependent map $\mu : [0, T) \mapsto \mathcal{P}(L^p(D; U))$ such that each μ_t has bounded support, and such that the corresponding correlation measures $(\nu_t^k)_{k \in \mathbb{N}}$ satisfy

$$\partial_t \left\langle \nu_{t, x}^k, \xi_1 \otimes \cdots \otimes \xi_k \right\rangle + \sum_{i=1}^k \nabla_{x_i} \cdot \left\langle \nu_{t, x}^k, \xi_1 \otimes \cdots \otimes f(\xi_i) \otimes \cdots \otimes \xi_k \right\rangle = 0 \tag{2.30}$$

in the sense of distributions, i.e.,

$$\int_{\mathbb{R}_+} \int_{D^k} \langle \nu_{t,x}^k, \xi_1 \otimes \cdots \otimes \xi_k \rangle : \partial_t \varphi + \sum_{i=1}^{k} \langle \nu_{t,x}^k, \xi_1 \otimes \cdots \otimes f(\xi_i) \otimes \cdots \otimes \xi_k \rangle : \nabla_{x_i} \varphi \, dx dt$$

$$+ \int_{D^k} \langle \bar{\nu}_x^k, \xi_1 \otimes \cdots \otimes \xi_k \rangle : \varphi\big|_{t=0} \, dx = 0$$

for every $\varphi \in C_c^\infty\big(D^k \times \mathbb{R}_+, \ U^{\otimes k}\big)$ and for every $k \in \mathbb{N}$. (Here, $\bar{\boldsymbol{\nu}}$ denotes the correlation measure associated with the initial probability measure $\bar{\mu}$.)

Remark 2.6. If the initial data $\bar{\mu}$ and a resulting statistical solution μ_t are both atomic, i.e. $\bar{\mu} = \delta_{\bar{u}}$ and $\mu_t = \delta_u$ with $\bar{u} \in L^p(D; U)$ and $u \in L^p((0,T) \times D; U)$, then it is easy to see that a statistical solution in the above sense reduces to a weak solution of (1.1a). Thus, weak solutions are statistical solutions.

Remark 2.7. The evolution equation for the first correlation marginal of the statistical solution, i.e., for $k = 1$ in (2.30), is equivalent to the definition of a measure-valued solution of (1.1a) [13,15]. Thus, a statistical solution can be thought of as a measure-valued solution augmented with information about all possible multi-point correlations. Hence, *a priori*, a statistical solution contains significantly more information than a measure-valued solution.

3. Dissipative Statistical solutions and weak-strong uniqueness

In analogy with weak solutions, it is necessary to impose additional admissibility criteria for statistical solutions in order to ensure uniqueness and stability. In [16], the authors proposed an entropy condition for statistical solutions of scalar conservation laws. This condition was based on a non-trivial generalization of the Kruzkhov entropy condition to the framework of time-parameterized probability measures on $L^1(D)$. It was shown in [16] that these *entropy statistical solutions* were unique and stable in the 1-Wasserstein metric on $\mathcal{P}(L^1(D))$, with respect to perturbations of the initial data.

Although one can extend the entropy condition of [16] to statistical solutions for systems of conservation laws (1.1a), it is not possible to obtain uniqueness and stability of such entropy statistical solutions. Instead, one has to seek alternative notions of stability for systems of conservation laws.

A possible weaker framework for uniqueness (stability) is that of *weak-strong uniqueness*, see [46,10] and references therein. Within this framework, one imposes certain *entropy conditions* and proves that the resulting entropy solutions will coincide with a strong (classical) solution if such a solution exists. Weak-strong uniqueness for systems of conservation laws with strictly convex entropy functions is shown in [10]. In fact, one can even prove weak-strong uniqueness results for the much weaker notion of entropy or dissipative measure-valued solutions of systems of conservation laws, see [12,6,15].

Our aim in this section is to propose a suitable notion of *dissipative statistical solution*s and prove a *weak-strong* uniqueness result for such solutions. Stability of solutions will be measured in the *Wasserstein distance*, whose definition we recall first.

Definition 3.1. Let X be a separable Banach space and let $\mu, \rho \in \mathcal{P}(X)$ have finite pth moments, i.e. $\int_X |x|^p d\mu(x) < \infty$ and $\int_X |x|^p d\rho(x) < \infty$. The p-Wasserstein distance between μ and ρ is defined as

$$W_p(\mu, \rho) = \left(\inf_{\pi \in \Pi(\mu, \rho)} \int_{X^2} |x - y|^p \, d\pi(x, y) \right)^{\frac{1}{p}}; \tag{3.1}$$

where the infimum is taken over the set $\Pi(\mu, \rho) \subset \mathcal{P}(X^2)$ of all transport plans from μ to ρ, i.e. those $\pi \in \mathcal{P}(X^2)$ satisfying

$$\int_{X^2} F(x) + G(y) \, d\pi(x, y) = \int_X F(x) \, d\mu(x) + \int_X G(y) \, d\rho(y) \qquad \forall \, F, G \in C_b(X)$$

(see e.g. [44]).

As in [16], our *entropy condition* for statistical solutions will rely on a comparison with probability measures that are convex combinations of Dirac masses, i.e. $\rho \in \mathcal{P}(L^2(D))$ such that $\rho = \sum_{i=1}^M \alpha_i \delta_{u_i}$ for coefficients $\alpha_i \geqslant 0$, $\sum_i \alpha_i = 1$ and functions $u_1, \ldots, u_M \in L^2(D)$. From [16], we observe that whenever ρ is of this *M-atomic form*, there is a one-to-one correspondence between transport plans $\pi \in \Pi(\mu, \rho)$ and elements of the set

$$\Lambda(\alpha, \mu) := \left\{ (\mu_1, \ldots, \mu_M) \; : \; \mu_1, \ldots, \mu_M \in \mathcal{P}(L^2(D; U)) \text{ and } \sum_{i=1}^M \alpha_i \mu_i = \mu \right\},$$

defined for any $\alpha = (\alpha_1, \ldots, \alpha_M) \in \mathbb{R}^M$ satisfying $\alpha_i \geqslant 0$ and $\sum_{i=1}^M \alpha_i = 1$. The set $\Lambda(\alpha, \mu)$ is never empty since $(\mu, \ldots, \mu) \in \Lambda(\alpha, \mu)$ for any choice of coefficients $\alpha_1, \ldots, \alpha_M$. Note that the set $\Lambda(\alpha, \mu)$ depends on the target measure ρ *only* through the weights $\alpha_1, \ldots, \alpha_M$.

Using this decomposition of transport plans with respect to M-atomic probability measures, we define the notion of dissipative statistical solution as follows.

Definition 3.2. Assume that the system of conservation laws (1.1a) is equipped with an entropy function η. A statistical solution μ of (1.1a) is a *dissipative statistical solution* if

(i) for every choice of coefficients $\alpha_1, \ldots, \alpha_M > 0$ with $\sum_{i=1}^M \alpha_i = 1$ and for every $(\bar{\mu}_1, \ldots, \bar{\mu}_M) \in \Lambda(\alpha, \bar{\mu})$, there exists a function $t \mapsto (\mu_{1,t}, \ldots, \mu_{M,t}) \in \Lambda(\alpha, \mu_t)$, such that each measure $\mu_i \in \mathcal{P}_T(L^p(D; U))$ is a statistical solution of (1.1a) with initial data $\bar{\mu}_i$,

(ii) for all test functions $0 \leqslant \theta(t) \in C_c^\infty(\mathbb{R}_+)$,

$$\int_{\mathbb{R}_+} \int_{L^p(D,U)} \int_D \eta(u(x))\theta'(t) \, dx d\mu_t(u) dt + \int_{L^p(D,U)} \int_D \eta(\bar{u}(x))\theta(0) dx d\bar{\mu}(\bar{u}) \geqslant 0. \tag{3.2}$$

We remark that the first condition in the above definition demands that the decomposition of a statistical solution into the components μ_i is still consistent with the underlying conservation law (1.1a). On the other hand, the second condition (3.2) amounts to requiring that the total entropy of μ decreases in time.

First, we investigate the stability of a dissipative statistical solution of (1.1a) with respect to statistical solutions built from finitely many classical solutions of (1.1).

Lemma 3.1. *Let $T > 0$, set $p = 2$, assume that*

$$\|f''\|_{L^\infty(\mathbb{R}^N)} < \infty \tag{3.3}$$

(where we denoted by f'' the Hessian of f, i.e. $(f''(u))_{ijk} = \partial_{u^j}\partial_{u^k}f^i(u)$, $i, j, k = 1, \ldots, N$), and assume that the conservation law (1.1a) is equipped with an entropy pair (η, q) for which

$$c \leqslant (\eta''(u)v, v) \leqslant C \qquad \forall\, u \in \mathbb{R}^N,\, v \in \mathbb{R}^N \text{ with } |v| = 1 \tag{3.4}$$

(where η'' denotes the Hessian matrix of η) for $c, C > 0$. Let $\mu \in \mathcal{P}_T(L^2(D; U))$ be a dissipative statistical solution of (1.1a), and for $t \in [0, T)$ let $\rho_t = \sum_{i=1}^M \alpha_i \delta_{v_i(t)}$ for coefficients $\alpha_i > 0$, $\sum_{i=1}^M \alpha_i = 1$, and classical solutions $v_1, \ldots, v_M \in W^{1,\infty}(D \times [0, T); U)$ of (1.1a). Then

$$W_2(\mu_t, \rho_t) \leqslant e^{Ct} W_2(\mu_0, \rho_0) \qquad \forall\, t \in [0, T), \tag{3.5}$$

where $C = C(R) \geqslant 0$ is a constant only depending on $R := \max_{i=1,\ldots,M} \|v_i\|_{W^{1,\infty}(D \times \mathbb{R}_+, U)}$.

Proof. It is straightforward to verify that ρ_t as defined above is a statistical solution of (1.1a) with initial data $\bar{\rho} := \sum_{i=1}^M \alpha_i \delta_{\bar{v}_i}$, where $\bar{v}_i := v_i(0)$.

Let $\bar{\mu}^* = (\bar{\mu}_1^*, \ldots, \bar{\mu}_M^*) \in \Lambda(\alpha, \bar{\mu})$ define a transport plan that minimizes the transport cost between $\bar{\mu} := \mu_0$ and $\bar{\rho}$, that is,

$$W_2(\bar{\mu}, \bar{\rho}) = \left(\sum_{i=1}^M \alpha_i \int_{L^2} \|u - \bar{v}_i\|_{L^2}^2 \, d\bar{\mu}_i^*(u) \right)^{\frac{1}{2}}. \tag{3.6}$$

(Here and in the remainder of this proof, we denote $L^2 = L^2(D; U)$.) As μ_t is a dissipative statistical solution, there exists a map $t \mapsto (\mu_{1,t}^*, \ldots, \mu_{M,t}^*) \in \Lambda(\alpha, \mu_t)$ such that

$$\sum_{i=1}^M \alpha_i \left(\int_0^T \int_{L^2} \int_D u(x) \partial_t \varphi_i(x, t) + f(u(x)) \cdot \nabla_x \varphi_i(x, t) \, dx \, d\mu_{i,t}^*(u) \, dt \right.$$

$$\left. + \int_{L^2} \int_D \bar{u} \varphi_i(x, 0) \, dx \, d\bar{\mu}_i^*(\bar{u}) \right) = 0 \quad (3.7)$$

for every $\varphi_1, \ldots, \varphi_M \in C_c^\infty(D \times [0, T))$. For each $1 \leqslant i \leqslant M$, we have that

$$
\int_0^T \!\!\int_{L^2} \!\!\int_D v_i(x,t) \partial_t \varphi_i + f(v_i(t,x)) \cdot \nabla_x \varphi_i dx d\mu_{i,t}^*(u) dt + \int_{L^2} \!\!\int_D \bar{v}_i(x) \varphi_i(x,0) dx d\bar{\mu}_i^*(\bar{u})
$$

$$
= \int_0^T \!\!\int_{L^2} \!\!\int_D \partial_t(v_i \varphi_i) + \nabla_x \cdot (f(v_i)\varphi_i) dx d\mu_{i,t}^*(u) dt + \int_{L^2} \!\!\int_D \bar{v}_i(x) \varphi_i(x,0) dx d\bar{\mu}_i^*(\bar{u})
$$

$$
- \underbrace{\int_0^T \!\!\int_{L^2} \!\!\int_D \varphi_i\big(\partial_t v_i + \nabla_x \cdot f(v_i)\big) dx d\mu_{i,t}^*(u) dt}_{= \,0,\ \text{as } v_i \text{ is a classical solution of (1.1a)}}
$$

$$
= - \int_D v_i(x,0)\varphi_i(x,0) dx + \int_D \bar{v}_i(x)\varphi_i(x,0) dx = 0.
$$

Multiplying the above with α_i and summing over i, we obtain

$$
\sum_{i=1}^M \alpha_i \bigg(\int_0^T \!\!\int_{L^2} \!\!\int_D v_i \partial_t \varphi_i + f(v_i) \cdot \nabla_x \varphi \, dx d\mu_{i,t}^*(u) dt
$$

$$
+ \int_{L^2} \!\!\int_D \bar{v}_i \varphi_i(x,0) dx d\bar{\mu}_i^*(\bar{u}) \bigg) = 0. \tag{3.8}
$$

Subtracting (3.8) from (3.7) and choosing as a test function the vector-valued function $\varphi_i = \eta'(v_i(x,t))\theta(t)$ for some scalar test function $0 \leqslant \theta(t) \in C_c^\infty(\mathbb{R}_+)$ (here, η' denotes the vector-valued derivative of η with respect to u), and using the fact that

$$
\partial_t \varphi_i = \eta'(v_i)\theta'(t) + \theta(t)\eta''(v_i)\partial_t v_i = \eta'(v_i)\theta'(t) - \theta(t)f'(v_i) \cdot \nabla_x \eta'(v_i),
$$
$$
\partial_j \varphi_i = \theta(t)\partial_j \eta'(v_i)
$$

yields

$$
0 = \sum_{i=1}^M \alpha_i \bigg(\int_0^T \!\!\int_{L^2} \!\!\int_D (u - v_i) \cdot \partial_t \varphi_i + (f(u) - f(v_i)) \cdot \nabla_x \varphi_i \, dx d\mu_{i,t}^*(u) dt
$$

$$
+ \int_{L^2} \!\!\int_D (\bar{u} - \bar{v}_i) \cdot \varphi_i(x,0) \, dx d\bar{\mu}_i^*(\bar{u}) \bigg)
$$

$$
= \sum_{i=1}^M \alpha_i \bigg(\int_0^T \!\!\int_{L^2} \!\!\int_D \eta'(v_i) \cdot (u - v_i)\theta'(t) \, dx d\mu_{i,t}^*(u) dt
$$

$$
+ \int_{L^2} \!\!\int_D \eta'(\bar{v}_i) \cdot (\bar{u} - \bar{v}_i)\theta(0) \, dx d\bar{\mu}_i^*(\bar{u})
$$

$$
+ \int_0^T \!\!\int_{L^2} \!\!\int_D \theta(t) \underbrace{\big(f(u) - f(v_i) - f'(v_i)(u - v_i)\big) \cdot \nabla_x \eta'(v_i)}_{=: \, \mathcal{Z}(u|v_i)} \, dx d\mu_{i,t}^*(u) dt \bigg) \tag{3.9}
$$

As v_i is a classical solution of (1.1a) and $\mu_{1,t}^*, \ldots, \mu_{M,t}^*$ are probability measures,

we obtain from the entropy conservation of v_i that

$$\sum_{i=1}^{M} \alpha_i \left(\int_0^T \int_{L^2} \int_D \eta(v_i)\theta'(t)dxd\mu_{i,t}^*(u)dt + \int_{L^2} \int_D \eta(\bar{v}_i)\theta(0)dxd\bar{\mu}_i^*(\bar{u}) \right) = 0,$$

(3.10)

for the same test function θ used in (3.9).

Since μ_t is a dissipative statistical solution of (1.1a), we have

$$0 \leqslant \int_{\mathbb{R}_+} \int_{L^2} \int_D \eta(u(x))\theta'(t)dxd\mu_t dt + \int_{L^2} \int_D \eta(\bar{u}(x))\theta(0)dxd\bar{\mu}(\bar{u})$$

$$= \sum_{i=1}^{M} \alpha_i \left(\int_{\mathbb{R}_+} \int_{L^2} \int_D \eta(u(x))\theta'(t)dxd\mu_{i,t}^*dt + \int_{L^2} \int_D \eta(\bar{u}(x))\theta(0)dxd\bar{\mu}_i^*(\bar{u}) \right).$$

Subtracting (3.10) and (3.9) from the above inequality we obtain

$$\sum_{i=1}^{M} \alpha_i \int_0^T \int_{L^2} \int_D \underbrace{\left(\eta(u) - \eta(v_i) - \eta'(v_i)(u - v_i) \right)}_{=:\,\mathcal{H}(u|v_i)} \theta'(t) \, dxd\mu_{i,t}^*(u)dt$$

$$+ \sum_{i=1}^{M} \alpha_i \int_{L^2} \int_D \left(\eta(\bar{u}) - \eta(\bar{v}_i) - \eta'(\bar{v}_i)(\bar{u} - \bar{v}_i) \right)\theta(0) \, dxd\bar{\mu}_i^*(\bar{u})$$

(3.11)

$$+ \sum_{i=1}^{M} \alpha_i \int_0^T \int_{L^2} \int_D \theta(t)\mathcal{Z}(u|v_i) \, dxd\mu_{i,t}^*(u)dt \geqslant 0$$

As η is strictly convex (the lower bound in (3.4)), we have

$$\mathcal{H}(u|v_i) \geqslant c|u - v_i|^2.$$

Similarly, using (3.3) and the fact that $\|v_i\|_{W^{1,\infty}} \leqslant R$, we obtain that

$$\max \left(\mathcal{Z}(u|v_i), \mathcal{H}(u|v_i) \right) \leqslant C|u - v_i|^2.$$

Using the above estimates in (3.11) and choosing the test function $\theta(s) = \chi_{(0,t]}(s)$ (by approximating with smooth functions) yields

$$\sum_{i=1}^{M} \alpha_i \int_{L^2} \int_D |u(x) - v_i(x,t)|^2 dxd\mu_{i,t}^*(u) \leqslant \sum_{i=1}^{M} \alpha_i \int_{L^2} \int_D |\bar{u}(x) - \bar{v}_i(x)|^2 dxd\bar{\mu}_i^*(\bar{u})$$

$$+ C \sum_{i=1}^{M} \alpha_i \int_0^t \int_{L^2} \int_D |u(x) - v_i(x,s)|^2 dxd\mu_{i,s}^*(u)ds.$$

(3.12)

Applying the integral form of Grönwall's inequality to the above estimate results in

$$\sum_{i=1}^{M} \alpha_i \int_{L^2} \|u - v_i\|_{L^2}^2 dxd\mu_{i,t}^*(u) \leqslant e^{Ct} \sum_{i=1}^{M} \alpha_i \int_{L^2} \|\bar{u} - \bar{v}_i\|_{L^2}^2 dxd\bar{\mu}_i^*(\bar{u})$$

$$= e^{Ct} W_2(\bar{\mu}, \bar{\rho})^2,$$

the last step following from (3.6). As $(\mu_{1,t}^*, \ldots, \mu_{M,t}^*) \in \Lambda(\alpha, \mu_t)$, the above inequality implies (3.5) and concludes the proof. □

The estimate (3.5) implies stability of dissipative statistical solutions with respect to probability measures that are convex combinations of Dirac masses, concentrated on classical solutions of (1.1a). We can extend such a stability result to a more general class of strong solutions:

Definition 3.3. A statistical solution μ is a *strong statistical solution* of (1.1a) if there is some $R > 0$ such that for every $n \in \mathbb{N}$, there exists a $\rho_n \in \mathcal{P}_T(L^2(D, U))$ of the form $\rho_{n,t} = \sum_{i=1}^{N_n} \alpha_i \delta_{v_i(t)}$ such that each v_i is a classical solution of (1.1a) and $v_i \in B_R \subset W^{1,\infty}(D \times (0, T))$, for all $1 \leqslant i \leqslant N_n$ and such that

$$W_2(\mu_t, \rho_{n,t}) \leqslant \frac{1}{n} \qquad \forall\ t \in [0, T] \tag{3.13}$$

We can now prove our main weak-strong uniqueness result:

Theorem 3.1. *Let $\bar{\mu} \in \mathcal{P}(L^2(D; U))$. Then under the same assumptions as in Lemma 3.1, if there exists a strong statistical solution μ of (1.1a), then it is unique in the class of dissipative statistical solutions.*

Proof. For a fixed $n \in \mathbb{N}$, Definition 3.3 implies the existence of a statistical solution $\rho_t = \sum_{i=1}^{N_n} \alpha_i \delta_{v_i(t)}$ such that

$$W_2(\mu_t, \rho_t) \leqslant \frac{1}{n}. \tag{3.14}$$

Moreover, $v_i(x, t) \in B_R$ is a classical solution of (1.1a) with initial data $\bar{v}_i$ for all $i = 1, \ldots, N_n$.

Let γ_t be another dissipative statistical solution of (1.1a) with $\gamma_0 = \bar{\mu}$. By Lemma 3.1 we have

$$W_2(\rho_t, \gamma_t) \leqslant \frac{e^{Ct}}{n}$$

for some $C = C(R)$. Using (3.14) and the triangle inequality yields

$$W_2(\mu_t, \gamma_t) \leqslant \frac{e^{Ct} + 1}{n}.$$

Letting $n \to \infty$ concludes the proof of uniqueness of strong solutions. □

Remark 3.1. If we assume that the initial data $\bar{\mu}$ is such that $\mathrm{supp}(\bar{\mu}) \subset B_{R_0}$, with B_{R_0} being the ball of radius $R_0 > 0$ in $W^{1,\infty}(D; U)$, then by classical results on local well-posedness for (1.1a) with strictly convex entropies [10], there exist $T(R_0), R(R_0) > 0$, such that for every initial data $\bar{v} \in \mathrm{supp}(\bar{\mu})$, there exists a corresponding classical solution $v \in W^{1,\infty}(D \times [0, T(R_0)])$ and $\|v\|_{W^{1,\infty}(D \times [0, T(R_0)])} \leqslant R(R_0)$. Moreover, the data-to-solution map $\mathcal{S}_t : \mathrm{supp}\,\bar{\mu} \to L^2(D; U)$, is well defined for all $0 \leqslant t \leqslant T(R_0)$, and continuous because

$$\|\mathcal{S}_t(\bar{v}) - \mathcal{S}_t(\hat{v})\|_2 \leqslant e^{R(R_0)t} \|\bar{v} - \hat{v}\|_2, \qquad \forall\ t \in [0, T(R_0)].$$

Letting $\mu_t = \mathcal{S}_t \# \bar{\mu}$ for all $0 \leqslant t \leqslant T(R_0)$, one can verify that μ_t is indeed a dissipative statistical solution of (1.1a). Moreover, it is strong in the sense of Definition 3.3. Consequently, we can establish that *as long as the underlying initial data is supported on smooth functions, the resulting statistical solutions are locally well-posed.*

4. Numerical approximation of statistical solutions.

In this section, we will construct statistical solutions for the system of conservation laws (1.1a) by proposing an algorithm to numerically approximate it. We show, under reasonable hypotheses on the underlying numerical schemes, that the approximations constructed by this algorithm converge to a (dissipative) statistical solution of (1.1a). As in [15,19,17], the algorithm will be based on a finite volume spatio-temporal discretization and Monte Carlo sampling of the underlying probability space. The spatial domain D will everywhere be assumed to be bounded.

4.1. *Multidimensional finite volume framework*

In this section we briefly describe numerically approximating conservation laws with finite volume and finite difference methods. For a complete review, one can consult [26,28,29,34].

We discretize the computational spatial domain as a collection of cells

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

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,

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

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

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} u^\Delta_{i^1,\ldots,i^d}(t) + \sum_{m=1}^{d} \frac{1}{\Delta} \bigg(&F^{m,\Delta}\big(u^\Delta_{\mathbf{i}-(q-1)\mathbf{e}_m}(t),\ldots,u^\Delta_{\mathbf{i}+q\mathbf{e}_m}(t)\big) \\
&-F^{m,\Delta}\big(u^\Delta_{\mathbf{i}-q\mathbf{e}_m}(t),\ldots,u^\Delta_{\mathbf{i}+(q-1)\mathbf{e}_m}(t)\big) \bigg) = 0 \\
&u^\Delta_{i^1,\ldots,i^d}(0) = \bar{u}(x_{i^1,\ldots,i^d})
\end{aligned} \tag{4.1}$$

where $\mathbf{e}_1,\ldots,\mathbf{e}_d$ are the canonical unit vectors in $\mathbb{R}^d$, and $F^{m,\Delta}$ is a *numerical flux function* in direction $m = 1,\ldots,d$. We say that the scheme is a *$(2q+1)$-point scheme*, when the numerical flux function $F^{m,\Delta}$ can be written as a function of

$u_{\mathbf{i}+j\mathbf{e}_m}^\Delta(t)$ for $j = -q+1, \ldots, q$. We furthermore assume the numerical flux function is consistent with f and locally Lipschitz continuous, which amounts to requiring that for every bounded set $K \subset \mathbb{R}^N$, there exists a constant $C > 0$ such that for $m = 1, \ldots, d$,

$$\left| F^{m,\Delta}\big(u_{\mathbf{i}-(q-1)\mathbf{e}_m}^\Delta(t), \ldots, u_{\mathbf{i}+q\mathbf{e}_m}^\Delta(t)\big) - f^m(u_{\mathbf{i}}^\Delta)\right| \leqslant C \sum_{j=-q+1}^{q} \left| u_{\mathbf{i}}^\Delta(t) - u_{\mathbf{i}+j\mathbf{e}_m}^\Delta(t)\right|, \tag{4.2}$$

whenever $u_{\mathbf{i}-(q-1)\mathbf{e}_m}^\Delta(t), \ldots, u_{\mathbf{i}+q\mathbf{e}_m}^\Delta(t) \in K$. For the sake of notational simplicity we will write

$$F_{\mathbf{i}+1/2\mathbf{e}_m}^m(u) = F^{m,\Delta}\big(u_{\mathbf{i}-(q-1)\mathbf{e}_m}, \ldots, u_{\mathbf{i}+q\mathbf{e}_m}\big) \qquad \text{for } \mathbf{i} \in \mathbb{Z}^d, \ 1 \leqslant m \leqslant d.$$

We let $\mathcal{S}_t^\Delta : L^p(D) \to L^p(D)$ be the spatially discrete numerical evolution operator defined by (4.1), mapping $\bar{u} \mapsto u^\Delta(t)$. Since $\mathcal{S}_t^\Delta$ is the composition of a projection from L^p onto piecewise constant functions and a continuous evolution under an ordinary differential equation, we see that $\mathcal{S}_t^\Delta$ is measurable.

The current form of (4.1) is continuous in time, and one needs to employ a time stepping method to discretize the ODE system in time, usually through some strong stability preserving Runge–Kutta method [27].

As the operator $\mathcal{S}^\Delta$ is a measurable map, we can define an approximation of a statistical solution of (1.1a) with initial data $\bar{\mu}$ by

$$\mu_t^\Delta = \mathcal{S}_t^\Delta \# \bar{\mu}. \tag{4.3}$$

Henceforth, μ_t^Δ is referred to as an *approximate statistical solution*.

4.2. *Convergence of approximate statistical solutions.*

In this section, we investigate the convergence of the approximate statistical solutions μ_t^Δ as the mesh is refined, i.e., as $\Delta \to 0$.

Theorem 4.1. *Consider the system of conservation laws* (1.1a) *with initial data $\bar{\mu} \in \mathcal{P}(L^p(D; U))$ for some $1 \leqslant p < \infty$, such that $\mathrm{supp}(\bar{\mu}) \subset B_R(0) \subset L^p(D; U)$, with $B_R(0)$ being the ball of radius R and center 0, for some $R > 0$. Assume that the semi-discrete finite volume scheme* (4.1) *satisfies the following conditions:*

(i) L^p bounds:

$$\Delta^d \sum_{\mathbf{i} \in \mathbb{Z}^d} \left| u_{\mathbf{i}}^\Delta(t)\right|^p \leqslant C\Delta^d \sum_{\mathbf{i} \in \mathbb{Z}^d} |\bar{u}_{\mathbf{i}}|^p \qquad \forall \, t \in [0, T), \ \forall \, \bar{u} \in L^p(D; U). \tag{4.4}$$

(ii) Weak BV bounds: There exists $s \geqslant p$ such that

$$\Delta^d \int_0^T \sum_{m=1}^d \sum_{\mathbf{i} \in \mathbb{Z}^d} \left| u_{\mathbf{i}+\mathbf{e}_m}^\Delta(t) - u_{\mathbf{i}}^\Delta(t)\right|^s dt \leqslant C\Delta, \tag{4.5}$$

with the constant $C = C(\|\bar{u}\|_p)$ only depending on the L^p-norm of the initial data $\bar{u}$.

(iii) Approximate scaling*: There exists a constant $C > 0$, possibly depending on the initial data $\bar{\mu}$ but independent of the grid size Δ, such that for every $\ell > 1$*

$$S_{\ell\Delta}^p(\mu^\Delta) \leqslant C\ell^{\frac{1}{s}} S_\Delta^p(\mu^\Delta). \tag{4.6}$$

Here, $S_r^p(\mu)$ is the structure function *associated with the time parameterized probability measure $\mu_t \in \mathcal{P}_T(L^p(D;U))$ (equivalently, time-dependent correlation measure $\boldsymbol{\nu} \in \mathcal{L}^p([0,T),D;U)$), defined as*

$$S_r^p(\mu) := \left(\int_0^T \int_{L^p(D)} \int_D \fint_{B_r(x)} |u(x) - u(y)|^p dy dx d\mu_t(u) dt \right)^{\frac{1}{p}}. \tag{4.7}$$

Then there is a subsequence $\Delta' \to 0$ such that the approximate statistical solutions $\mu^{\Delta'}$ converge strongly to some $\mu \in \mathcal{P}_T(L^p(D,U))$, in the sense of Theorem 2.8(v).

Proof. We will show that that the approximate statistical solutions μ_t^Δ, defined in (4.3), satisfy the conditions of Theorem 2.8 and hence converge (up to a subsequence). To this end, we can readily verify from the uniform L^p bounds (4.4) and the fact that $\operatorname{supp} \bar{\mu} \subset B_R(0)$, that

$$\operatorname{supp}(\mu_t^\Delta) \subset B_{CR}(0) \qquad \forall\, t \in [0,T),$$

with C being the constant in (4.4).

Let $u^\Delta(t) = \mathcal{S}_t^\Delta \bar{u}$ be the solution generated by the scheme (4.1). Denoting

$$V_\Delta(\bar{u}) = \int_0^T \sum_{m=1}^d \sum_{\mathbf{i} \in \mathbb{Z}^d} \left| u_{\mathbf{i}+\mathbf{e}_m}^\Delta(t) - u_{\mathbf{i}}^\Delta(t) \right|^p dt, \tag{4.8}$$

we obtain from the weak BV estimate (4.5) and Hölder's inequality that (recall that D was assumed to be bounded)

$$V_\Delta(\bar{u}) \leqslant C(T,d,R)\Delta^{\frac{p}{s}-d}. \tag{4.9}$$

The above inequality holds for every $\bar{u} \in B_R(0) \subset L^p(D;U)$.

Next, for any $r \leqslant \Delta$, a straightforward but tedious calculation yields

$$
\begin{aligned}
\left(S_r^p(\mu^\Delta) \right)^p &= \int_0^T \int_{L^p} \int_D \fint_{B_r(x)} |u(x) - u(y)|^p dy dx d\mu_t^\Delta(u) dt \\
&= \int_0^T \int_{L^p} \int_D \fint_{B_r(x)} \left| \mathcal{S}_t^\Delta \bar{u}(x) - \mathcal{S}_t^\Delta \bar{u}(y) \right|^p dy dx d\bar{\mu}(\bar{u}) dt \quad \text{(by (4.3))} \\
&= \int_{L^p} \int_0^T \int_D \fint_{B_r(x)} \left| u^\Delta(x,t) - u^\Delta(y,t) \right|^p dy dx dt d\bar{\mu}(\bar{u}) \\
&\leqslant C_d \Delta^{d-1} r \int_{L^p} V_\Delta(\bar{u}) d\bar{\mu}(\bar{u}) \\
&\leqslant C(T,d,R) C_d r^{\frac{p}{s}} \quad \text{(by (4.9) and } r \leqslant \Delta\text{),}
\end{aligned}
$$

where $C_d = 3^{d-1}$ results from successive applications of the triangle inequality. Hence, summarizing the above calculation, we obtain that for any $r \leqslant \Delta$,

$$S_r^p(\mu^\Delta) \leqslant Cr^{\frac{1}{s}}, \tag{4.10}$$

for a constant C that depends on the dimension, the support of the initial probability measure and the final time but is independent of the grid size Δ.

Now, for any $\ell > 1$ and $r = \ell\Delta$, we have

$$\begin{aligned}
S_r^p(\mu^\Delta) = S_{\ell\Delta}^p(\mu^\Delta) &\leqslant C\ell^{\frac{1}{s}} S_\Delta^p(\mu^\Delta) && \text{(by scaling (4.6))} \\
&\leqslant C\ell^{\frac{1}{s}} \Delta^{\frac{1}{s}} && \text{(by (4.10))} \\
&= Cr^{\frac{1}{s}}.
\end{aligned}$$

Here, the constant is independent of Δ. By combining the above estimate with (4.10), we obtain that

$$S_r^p(\mu^\Delta) \leqslant Cr^{\frac{1}{s}} \tag{4.11}$$

for any $r > 0$.

Given the independence of the constant in (4.11) with respect to the grid size Δ, we see from (4.11) that the condition of *uniform diagonal continuity* (2.24) is satisfied. Hence, up to a subsequence still indexed by Δ, μ^Δ converges to some $\mu \in \mathcal{P}_T(L^p(D; U))$. $\qquad\square$

Several remarks on the assumptions in the above theorem follow.

Remark 4.1. There are many examples of finite volume/difference schemes of the form (4.1) which satisfy the uniform L^p bound (4.4) and the weak BV bound (4.5). Assume that the system of conservation laws (1.1a) possesses an entropy function η that satisfies

$$C_1(1 + |u|^p) \leqslant \eta(u) \leqslant C_2(1 + |u|^p) \qquad \forall\, u \in U \tag{4.12}$$

for constants $C_1, C_2 > 0$ and $p \in [1, \infty)$. Then the uniform L^p bound (4.4) follows for any scheme of form (4.1) that satisfies a *discrete entropy inequality*,

$$\frac{\mathrm{d}}{\mathrm{d}t}\eta\big(u_{\mathbf{i}}^\Delta(t)\big) + \sum_{m=1}^d \frac{1}{\Delta}\left(Q_{\mathbf{i}+\frac{1}{2}\mathbf{e}_m}^{m,\Delta}(t) - Q_{\mathbf{i}-\frac{1}{2}\mathbf{e}_m}^{m,\Delta}(t)\right) \leqslant 0, \tag{4.13}$$

with a numerical entropy flux $Q^{m,\Delta}$ that is consistent with the entropy flux q^m in (1.3) for $1 \leqslant m \leqslant d$.

In many cases the weak BV bound (4.5) also follows from the discrete entropy inequality (4.13). Examples of schemes which satisfy the discrete entropy inequality (4.13) and the weak BV bound (4.5) are the so-called entropy stable Lax–Wendroff schemes and the TeCNO schemes of [18].

Remark 4.2. The approximate scaling assumption (4.6) can be thought of as a weaker version of the so-called *self-similarity at small scales* assumption of Kolmogorov in his K41 theory for fully developed turbulence in incompressible fluid

flows, see hypothesis H2, equation (6.3), page 75 of [24]. Kolmogorov based his hypothesis on the fact that smooth solutions of the incompressible Navier–Stokes (Euler) equations scale exactly. Similar considerations also apply to several prototypical examples of systems of conservation laws (1.1a). In particular, for the compressible Euler equations (5.1) (in any space dimension), it can be readily checked that if $u(x, t)$ is a solution, then $\ell^\theta u(\ell x, \ell t)$ is also a solution for any $\theta, \ell > 0$. Hence, it is reasonable to hypothesize scaling, analogous to the Kolmogorov hypothesis, for systems of conservation laws.

It is essential to also point out the differences in our hypothesis (4.6) to the standard Kolmogorov hypothesis for turbulent incompressible flows. First, our hypothesis pertains only to the numerical solution, generated by the finite volume scheme (4.1). Moreover, we require mere inequalities in the scaling law (4.6), in contrast to the standard Kolmogorov hypothesis of equality.

Remark 4.3. Intermittency is widely accepted to be a characteristic of turbulent flows, see [24]. It is believed that intermittency stems from the fact that turbulent solutions do not scale exactly as in the Kolmogorov hypothesis. We automatically incorporate a form of intermittency by only requiring an upper bound in (4.6), instead of an equality. Hence, the scaling exponent in (4.6) can depend explicitly on the underlying length scale, provided that it is bounded below by $1/s$. This encodes a form of intermittency in the approximate solutions.

Remark 4.4. Another approach to incorporating intermittency and relaxing the scaling condition (4.6) is to consider a decomposition of the approximate statistical solution μ_t^Δ into a mean flow and a fluctuation. Defining the mean flow by

$$\widehat{u}^\Delta(x, t) = \langle \nu_{t,x}^{1,\Delta}, \xi \rangle, \tag{4.14}$$

we see that the mean flow is well defined for almost every $(x, t) \in D \times (0, T)$. Similarly, we can define *fluctuations* of $\tilde{\mu}^\Delta \in \mathcal{P}_T(L^p(D; U))$ by its action on all observables $g \in \mathcal{H}^{k,p}([0, T], D; U)$,

$$\langle \tilde{\mu}^\Delta, L_g \rangle = \int_0^T \int_{L^p(D;U)} \int_{D^k} g(x, t, u(x) - \widehat{u}^\Delta(x, t)) \, dx d\tilde{\mu}_t^\Delta(u) dt. \tag{4.15}$$

We can relax the assumption (4.6) by requiring that only the structure function associated with the fluctuation scales approximately, i.e.,

$$S_{\ell\Delta}^p(\tilde{\mu}^\Delta) \leqslant C\ell^{1/s} S_\Delta^p(\tilde{\mu}^\Delta) \qquad \forall\, \ell > 1. \tag{4.16}$$

If we further assume that the mean flow is BV and L^∞, i.e.,

$$\max\left(\left\| \widehat{u}^\Delta \right\|_{L^\infty((0,T)\times D)}, \left\| \widehat{u}^\Delta \right\|_{L^\infty((0,T);BV(D))} \right) \leqslant C, \tag{4.17}$$

for some constant that is independent of the mesh size Δ, then a straightforward

but tedious calculation yields for any $r = \ell\Delta$

$$
\begin{aligned}
S_r^p(\mu^\Delta) &\leqslant \bar{C} \left\| \widehat{u}^\Delta \right\|_{L^\infty((0,T);BV(D))} r^{\frac{1}{p}} + S_{\ell\Delta}^p(\tilde{\mu}^\Delta) \\
&\leqslant \bar{C} r^{\frac{1}{p}} + C \ell^{\frac{1}{s}} S_\Delta^p(\tilde{\mu}^\Delta) && \text{(by (4.16))} \\
&\leqslant \bar{C} r^{\frac{1}{p}} + C \ell^{\frac{1}{s}} \Delta^{\frac{1}{s}} && \text{(by (4.10))} \\
&= \bar{C} r^{\frac{1}{p}} + C r^{\frac{1}{s}}.
\end{aligned}
$$

Thus, the condition (2.24) in Theorem 2.8 is satisfied in this case. A similar argument can be made by imposing some form of (uniform) Hölder continuity on the mean flow.

4.3. *Consistency of the numerical method*

We fix an initial measure $\bar{\mu} \in \mathcal{P}(L^p(D, U))$. For any $u \in L^p(D, U)$ we define the local average of u as

$$
u_{\mathbf{i}} = \frac{1}{|\mathcal{C}_{\mathbf{i}}|} \int_{\mathcal{C}_{\mathbf{i}}} u(x) \, dx \qquad \text{for } \mathbf{i} \in \mathbb{Z}^d,
$$

where $|\mathcal{C}_{\mathbf{i}}|$ denotes the d-dimensional Lebesgue measure of $\mathcal{C}_{\mathbf{i}}$. We now state the "Lax–Wendroff theorem" for our numerical method, that is, consistency of the method with the PDE.

Theorem 4.2 (Lax–Wendroff theorem for statistical solutions). *Let the initial data $\bar{\mu}$ have bounded support, $\operatorname{supp} \bar{\mu} \subset B_K(0) \subset L^p(D, U)$ for some $K > 0$. Let μ^Δ be given by (4.3) for $\Delta > 0$, and assume that for some sequence $\Delta_n \to 0$, the sequence $\{\mu^{\Delta_n}\}_{n\in\mathbb{N}}$ converges strongly to μ in $\mathcal{P}_T(L^p(D, U))$, in the sense of Theorem 2.8(v). Assume moreover that the following weak BV bound is fulfilled:*

$$
\Delta^d \int_0^T \int_{L^p(D)} \sum_{m=1}^d \sum_{\mathbf{i} \in \mathbb{Z}^d} |u_{\mathbf{i}} - u_{\mathbf{i}-\mathbf{e}_m}|^s \, d\mu_t^\Delta(u) \, dt \leqslant C\Delta, \tag{4.18}
$$

for some $0 < s$. Then μ_t is a statistical solution of (1.1).

Given the complicated notation and very technical nature of the proof of the above theorem, we illustrate the main steps of the proof in a very special case, namely $k = 2$ for a one-dimensional scalar conservation law ($d = N = 1$). The proof in the general case is postponed to Appendix Appendix D.

Proof. [Proof for the second moment of a scalar conservation law in one spatial dimension] We consider a scalar conservation law ($N = 1$) in one spatial dimension ($d = 1$). By (4.4), there is some $K_T > 0$ such that

$$
\operatorname{supp} \mu_t^\Delta \subset B_{K_T}(0) \qquad \forall \, t \in [0, T]. \tag{4.19}
$$

Let $\{u_i^\Delta\}_{i\in\mathbb{Z}}$ be computed by (4.1). Denote $F^\Delta = F^{1,\Delta}$ and, for $u \in L^p(D)$, write

$$
F_{i+1/2}^\Delta(u) = F^\Delta\big(u_{i-q+1}, \ldots, u_{i+q}\big), \qquad i \in \mathbb{Z}.
$$

For all pairs of cells i, j, we have, by the product rule,

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(u_i^\Delta(t)u_j^\Delta(t)\big) = -\frac{1}{\Delta}\big(F_{i+1/2}^\Delta(u^\Delta(t)) - F_{i-1/2}^\Delta(u^\Delta(t))\big)u_j^\Delta$$
$$- \frac{1}{\Delta}u_i^\Delta\big(F_{j+1/2}^\Delta(u^\Delta(t)) - F_{j-1/2}^\Delta(u^\Delta(t)(t))\big).$$

Hence, for arbitrary $\varphi \in C_c^\infty(\mathbb{R}^2 \times [0, T))$, we get

$$0 = \int_0^T u_i^\Delta(t)u_j^\Delta(t)\partial_t\varphi(x_i, x_j, t)\, dt + \varphi(x_i, x_j, 0)u_j^\Delta(0)u_i^\Delta(0)$$
$$- \int_0^T \bigg(\frac{1}{\Delta}\big(F_{i+1/2}^\Delta(u^\Delta(t)) - F_{i-1/2}^\Delta(u^\Delta(t))\big)u_j^\Delta$$
$$+ \frac{1}{\Delta}u_i^\Delta\big(F_{j+1/2}^\Delta(u^\Delta(t)) - F_{j-1/2}^\Delta(u^\Delta(t))\big)\bigg)\varphi(x_i, x_j, t)\, dt.$$

Multiply by Δ^2, sum over all $i, j \in \mathbb{Z}$ and perform a summation-by-parts to obtain

$$0 = \Delta^2 \sum_{i,j\in\mathbb{Z}} \bigg(\int_0^T u_i^\Delta(t)u_j^\Delta(t)\partial_t\varphi(x_i, x_j, t)\, dt + u_j^\Delta(0)u_i^\Delta(0)\varphi(x_i, x_j, 0)\bigg)$$
$$+ \Delta^2 \sum_{i,j\in\mathbb{Z}} \int_0^T \bigg(F_{i+1/2}^\Delta(u^\Delta(t))u_j^\Delta(t)\frac{\varphi(x_{i+1}, x_j, t) - \varphi(x_i, x_j, t)}{\Delta}$$
$$+ u_i^\Delta(t)F_{j+1/2}^\Delta(u^\Delta(t))\frac{\varphi(x_i, x_{j+1}, t) - \varphi(x_i, x_j, t)}{\Delta}\bigg)\, dt$$
$$= \Delta^2 \sum_{i,j\in\mathbb{Z}} \bigg(\int_0^T u_i^\Delta(t)u_j^\Delta(t)\partial_t\varphi(x_i, x_j, t)\, dt + u_j^\Delta(0)u_i^\Delta(0)\varphi(x_i, x_j, 0)\bigg)$$
$$+ \Delta^2 \sum_{i,j\in\mathbb{Z}} \int_0^T \bigg(F_{i+1/2}^\Delta(u^\Delta(t))u_j^\Delta(t)\partial_1^\Delta\varphi(x_i, x_j, t) + u_i^\Delta(t)F_{j+1/2}^\Delta(u^\Delta(t))\partial_2^\Delta\varphi(x_i, x_j, t)\bigg)\, dt$$

where we have denoted $\partial_1^\Delta\varphi(x, y, t) = \frac{\varphi(x+\Delta, y, t) - \varphi(x, y, t)}{\Delta}$, and similarly for $\partial_2^\Delta\varphi$. From the special form (4.3) of μ_t^Δ, we therefore have

$$0 = \Delta^2 \sum_{i,j\in\mathbb{Z}} \bigg(\int_0^T \int_{L^p} u_i u_j \partial_t\varphi(x_i, x_j, t)\, d\mu_t^\Delta(u)\, dt + \int_{L^p} u_i u_j \varphi(x_i, x_j, 0)\, d\bar\mu(u)\bigg)$$
$$+ \Delta^2 \sum_{i,j\in\mathbb{Z}} \int_0^T \int_{L^p} \bigg(F_{i+1/2}^\Delta(u)u_j\partial_1^\Delta\varphi(x_i, x_j, t) + u_i F_{j+1/2}^\Delta(u)\partial_2^\Delta\varphi(x_i, x_j, t)\bigg)\, d\mu_t^\Delta(u)\, dt.$$

We write now

$$\Delta^2 \sum_{i,j\in\mathbb{Z}} \int_0^T \int_{L^p} F_{i+1/2}^\Delta(u) u_j \partial_1^\Delta \varphi(x_i, x_j, t) \, d\mu_t^\Delta(u) \, dt$$

$$= \Delta^2 \sum_{i,j\in\mathbb{Z}} \int_0^T \int_{L^p} f(u_i) u_j \partial_1^\Delta \varphi(x_i, x_j, t) \, d\mu_t^\Delta(u) \, dt$$

$$+ \Delta^2 \sum_{i,j\in\mathbb{Z}} \int_0^T \int_{L^p} \left(F_{i+1/2}^\Delta(u) - f(u_i) \right) u_j \partial_1^\Delta \varphi(x_i, x_j, t) \, d\mu_t^\Delta(u) \, dt.$$

The last term vanishes as $\Delta \to 0$, since

$$\left| \Delta^2 \sum_{i,j\in\mathbb{Z}} \int_0^T \int_{L^p} \left(F_{i+1/2}^\Delta(u) - f(u_i) \right) u_j \partial_1^\Delta \varphi(x_i, x_j, t) \, d\mu_t^\Delta(u) \, dt \right|$$

$$\leqslant \int_0^T \int_{L^p} \Delta^2 \sum_{i,j\in\mathbb{Z}} \left| F_{i+1/2}^\Delta(u) - f(u_i) \right| |u_j| |\partial_1^\Delta \varphi(x_i, x_j, t)| \, d\mu_t^\Delta(u) \, dt$$

$$\leqslant \int_0^T \int_{L^p} \left(\Delta \sum_{i\in\mathbb{Z}} \left| F_{i+1/2}^\Delta(u) - f(u_i) \right| \left\| \partial_1^\Delta \varphi(x_i, \cdot, t) \right\|_{L^{p'}(\mathbb{R})} \right) \left(\Delta \sum_{j\in\mathbb{Z}} |u_j|^p \right)^{1/p} d\mu_t^\Delta(u) \, dt$$

$$\leqslant K_T \int_0^T \int_{L^p} \Delta \sum_{i\in\mathbb{Z}} \left| F_{i+1/2}^\Delta(u) - f(u_i) \right| \left\| \partial_1^\Delta \varphi(x_i, \cdot, t) \right\|_{L^{p'}(\mathbb{R})} \, d\mu_t^\Delta(u) \, dt$$

(*by* (4.19))

$$\leqslant K_T \int_0^T \int_{L^p} \Delta \sum_{i\in\mathbb{Z}} \sum_{i'=i-q+1}^{i+q} \left| u_i - u_{i'} \right| \left\| \partial_1^\Delta \varphi(x_i, \cdot, t) \right\|_{L^{p'}(\mathbb{R})} \, d\mu_t^\Delta(u) \, dt$$

(*by the Lipschitz continuity* (4.2))

$$\leqslant C K_T \int_0^T \int_{L^p} \Delta \sum_{i\in\mathbb{Z}} \left| u_i - u_{i-1} \right| \left\| \partial_1^\Delta \varphi(x_i, \cdot, t) \right\|_{L^{p'}(\mathbb{R})} \, d\mu_t^\Delta(u) \, dt$$

$$\leqslant C K_T \underbrace{\left(\int_0^T \int_{L^p} \Delta \sum_{i\in\mathbb{Z}} \left| u_i - u_{i-1} \right|^s \, d\mu_t^\Delta(u) \, dt \right)^{1/s}}_{\to 0 \text{ as } \Delta\to 0, \text{ by } (4.18)} \underbrace{\left(\int_0^T \Delta \sum_{i\in\mathbb{Z}} \left\| \partial_1^\Delta \varphi(x_i, \cdot, t) \right\|_{L^{p'}(\mathbb{R})}^{s'} \, dt \right)^{1/s'}}_{\text{bounded as } \Delta\to 0}$$

$$\to 0.$$

A similar computation holds for the integral involving $F_{j+1/2}^\Delta(u)$. Setting $\Delta = \Delta_n$

then gives

$$0 = \lim_{n\to\infty} \left(\Delta_n^2 \sum_{i,j\in\mathbb{Z}} \left(\int_0^T \int_{L^p} u_i u_j \partial_t \varphi(x_i, x_j, t) \, d\mu_t^{\Delta_n}(u) \, dt + \int_{L^p} u_i u_j \varphi(x_i, x_j, 0) \, d\bar{\mu}(u) \right) \right.$$

$$\left. + \Delta_n^2 \sum_{i,j\in\mathbb{Z}} \int_0^T \int_{L^p} \Big(f(u_i) u_j \partial_1^{\Delta_n} \varphi(x_i, x_j, t) + u_i f(u_j) \partial_2^{\Delta_n} \varphi(x_i, x_j, t) \Big) \, d\mu_t^{\Delta_n}(u) \, dt \right)$$

(as u is piecewise constant $\mu_t^{\Delta_n}$-almost surely)

$$= \lim_{n\to\infty} \left(\int_0^T \int_{L^p} \int_{\mathbb{R}^2} u(x)u(y)\partial_t\varphi(x,y,t)\,dxdy\,d\mu_t^{\Delta_n}(u)\,dt + \int_{L^p} u(x)u(y)\varphi(x,y,0)\,dxdy\,d\mu_0^{\Delta}(u) \right.$$

$$\left. + \int_0^T \int_{L^p} \int_{\mathbb{R}^2} \Big(f(u(x))u(y)\partial_1\varphi(x,y,t) + u(x)f(u(y))\partial_2\varphi(x,y,t) \Big) \, dxdy\,d\mu_t^{\Delta_n}(u)\,dt \right)$$

$$= \int_0^T \int_{L^p} \int_{\mathbb{R}^2} u(x)u(y)\partial_t\varphi(x,y,t)\,dxdy\,d\mu_t(u)\,dt + \int_{L^p} u(x)u(y)\varphi(x,y,0)\,dxdy\,d\bar{\mu}(u)$$

$$+ \int_0^T \int_{L^p} \int_{\mathbb{R}^2} \Big(f(u(x))u(y)\partial_1\varphi(x,y,t) + u(x)f(u(y))\partial_2\varphi(x,y,t) \Big) \, dxdy\,d\mu_t(u)\,dt,$$

which completes the proof. □

Remark 4.5. We can readily show that the limit statistical solution μ_t is a dissipative statistical solution, assuming that the underlying finite volume method satisfies the discrete entropy inequality (4.13). To this end, for every choice of coefficients $\alpha_1, \ldots, \alpha_M > 0$ with $\sum_{i=1}^M \alpha_i = 1$ and every $(\bar{\mu}_1, \ldots, \bar{\mu}_M) \in \Lambda(\alpha, \bar{\mu})$, we construct $\mu_{i,t}^\Delta = \mathcal{S}_t^\Delta \# \bar{\mu}_i$ as the approximate statistical solution generated by the scheme (4.1). By the convergence theorem 4.1, we can show that each $\mu_{i,t}^\Delta$ converges (possibly along a further subsequence), in the topology of Theorem 2.8, to $\mu_{i,t} \in \mathcal{P}_T(L^p(D; U))$ as $\Delta \to 0$. By Theorem 4.2, each μ_i is a statistical solution of (1.1a) with initial data $\bar{\mu}_i$, and the condition (3.2) is a straightforward consequence of the discrete entropy inequality (4.13) and the growth condition (4.12).

4.4. *Monte Carlo algorithm*

While (4.3) provides an abstract definition of μ_t^Δ, it is not amenable to practical computations, since it requires the computation of the trajectory of the numerical solution operator for almost all possible initial data $\bar{u} \in \operatorname{supp} \bar{\mu}$. We will further approximate μ_t^Δ by sampling it for a large ensemble of initial data, drawn from the initial probability measure.

The Monte Carlo algorithm has been shown to be robust in tackling high dimensional problems with low regularity [39,40,?,?], and has later been demonstrated to perform very well for computing measure valued solutions [15].

Algorithm 1 (Monte Carlo Algorithm).

Data: Initial $\bar{\mu} \in \mathcal{P}(L^p(D, U))$, mesh width $\Delta > 0$, numerical evolution operator $\mathcal{S}^\Delta$, number of samples $M \in \mathbb{N}$

For some probability space $(\Omega, \boldsymbol{\Omega}, \mathbb{P})$, let $\bar{u}_1, \ldots, \bar{u}_M : \Omega \to L^p(D; U)$ be independent random variables with distribution $\bar{\mu}$;

for $m = 1, \ldots, M$ **do**

$\quad$ Evolve the sample in time, $u_m^\Delta(t) = \mathcal{S}_t^\Delta(\bar{u}_m)$

end

Estimate statistical solution by the empirical measure

$$\mu_t^{\Delta, M}(\omega) := \frac{1}{M} \sum_{m=1}^{M} \delta_{u_m^\Delta(\omega; \cdot, t)}. \tag{4.20}$$

In the rest of this paper we will refer to the above algorithm simply as "the Monte Carlo Algorithm". Note that for any admissible observable $g \in \mathcal{H}_1^{k,p}([0, T], D; U)$, using (4.3) and (4.20), we obtain that

$$\begin{aligned}
\langle \mu^\Delta, L_g \rangle &= \int_0^T \int_{L^p} \int_{D^k} g(x, t, u) dx d\mu_t^\Delta dt = \int_{L^p} \int_0^T \int_{D^k} g\left(x, t, \mathcal{S}_t^\Delta \bar{u}(x)\right) dx dt d\bar{\mu} \\
&\approx \frac{1}{M} \sum_{m=1}^{M} \left(\int_0^T \int_{D^k} g\left(x, t, \mathcal{S}_t^\Delta \bar{u}_m(x)\right) dx dt \right) = \langle \mu^{\Delta, M}, L_g \rangle.
\end{aligned} \tag{4.21}$$

Remark 4.6. One should note that the probability measure $\mu_T^{\Delta, M}$ is indeed a random probability measure depending on some probability space Ω from which $\bar{u}_1, \ldots \bar{u}_M$ are being drawn.

Using well-known results for weak convergence of the Monte Carlo algorithm [43], we can prove that the Monte Carlo approximation of the statistical solution converges as the number of samples is increased.

Theorem 4.3. *Let $\bar{\mu} \in \mathcal{P}(L^p(D, U))$ have bounded support, let $\mathcal{S}^\Delta$ be some numerical evolution operator, and let $\mu_t^{\Delta, M}$ be defined through the Monte Carlo Algorithm. Let μ_t^Δ be defined by (4.3). Then for every admissible observable $g \in \mathcal{H}_1^{k,p}([0, T], D; U)$, we have*

$$\mathbb{E}\left[\left\langle \mu_T^{\Delta, M} - \mu_T^\Delta, L_g \right\rangle^2 \right] \leqslant \frac{\langle \mu_T^\Delta, L_g^2 \rangle - \langle \mu_T^\Delta, L_g \rangle^2}{M}, \tag{4.22}$$

where

$$\langle \mu_T^\Delta, L_g^2 \rangle := \int_{L^p} \int_{D^2} g(x, u(x)) g(y, u(y)) \, dx dy \, d\mu_T^\Delta(u).$$

The proof of the above theorem follows by standard arguments for proving convergence of Monte Carlo approximations. It is analogous to the proof of Monte Carlo convergence to statistical solutions of scalar conservation laws, see Theorem 2

of [17]. Note that the right hand side in (4.22) is bounded on account of the hypothesis (2.16) on admissible observables $g \in \mathcal{H}_1^{k,p}([0,T], D; U)$.
]

5. Numerical experiments

For all the numerical experiments in this section, we consider the two-dimensional compressible Euler equations,

$$
\frac{\partial}{\partial t}
\begin{pmatrix} \rho \\ \rho w^x \\ \rho w^y \\ E \end{pmatrix}
+ \frac{\partial}{\partial x_1}
\begin{pmatrix} \rho w^x \\ \rho (w^x)^2 + p \\ \rho w^x w^y \\ (E+p)w^x \end{pmatrix}
+ \frac{\partial}{\partial x_2}
\begin{pmatrix} \rho w^y \\ \rho w^x w^y \\ \rho (w^y)^2 + p \\ (E+p)w^y \end{pmatrix}
= 0. \qquad (5.1)
$$

The system is closed with the equation of state

$$
E = \frac{p}{\gamma - 1} + \frac{\rho \left((w^x)^2 + (w^y)^2 \right)}{2}.
$$

We set $\gamma = 1.4$ for all experiments.

5.1. *Kelvin–Helmholtz problem*

We start with this well-known test case for the development of instabilities in fluid flows, which was also extensively studied in [15].

The Kelvin–Helmholtz initial data is a shear flow, separating two states of varying density and pressure,

$$
u_0(\omega; x_1, x_2) =
\begin{cases}
u_L & I_1(\omega; x_1) \leqslant x_2 \leqslant I_2(\omega; x_1) \\
u_R & \text{otherwise}
\end{cases}
\qquad \text{for } (x_1, x_2) \in D = [0,1]^2.
$$

$$(5.2)$$

We assign periodic boundary conditions, and the two states are given as $\rho_L = 2$, $\rho_R = 1$, $w_L^x = -0.5$, $w_R^x = 0.5$, $w_L^y = w_R^y = 0$ and $p_L = p_R = 2.5$. The interfaces between the two states are given as

$$
I_i(x, \omega) = \frac{2(i-1)+1}{4} + \epsilon \sum_{j=1}^{K} a_j^i(\omega) \sin(2\pi(x + b_j^i(\omega))), \qquad (5.3)
$$

where $K = 10$, $\epsilon > 0$, and a_j^i and b_j^i (for $i = 1, 2$, $j = 1, \ldots, K$) are independent uniformly distributed random variables on the interval $[0,1]$. We normalize $(a_j^i)_{j=1}^K$ such that $\sum_j a_j^i = 1$.

The initial measure $\bar{\mu}$ is given by the distribution of the random field u_0. Note that although $\bar{\mu}$ is a probability measure on the infinite-dimensional space $L^p(D; U)$, it is only concentrated on a 40-dimensional subset of this space.

As was already shown in [15], there is no convergence for single realizations (samples) of the problem (5.1), (5.2) as the mesh is refined. We observe this behavior

from Figure 1, where we display the approximate density at time $T = 2$, computed with a second-order high-resolution finite volume scheme using an HLLC approximate Riemann solver and WENO reconstruction, together with a second-order SSP Runge–Kutta time integrator, on a sequence of successively refined meshes. As seen from the figure, structures at finer and finer scales are generated upon mesh refinement, impeding convergence. This lack of convergence is also verified from Figure 2 (A), where the so-called Cauchy rates of the density are shown, i.e., quantities of the form

$$\text{Cauchy}_p(\Psi, \Delta, T) := \|\Psi^\Delta(\cdot, T) - \Psi^{\frac{\Delta}{2}}(\cdot, T)\|_{L^p(D)}, \tag{5.4}$$

with Ψ^Δ being any observable of the solution computed on a mesh with mesh size Δ. On the other hand, the theory developed in Section 4.1 suggests that observables

(a) 128×128

(b) 256×256

(c) 512×512

(d) 1024×1024
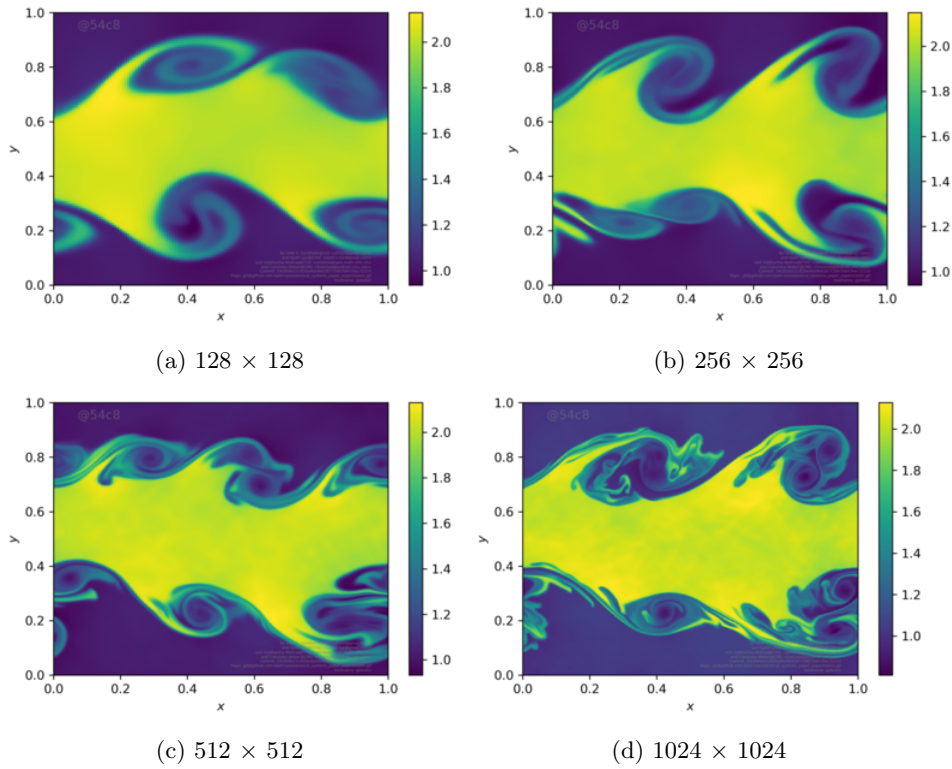
Fig. 1: The approximate density of the Kelvin–Helmholtz instability (5.2) with a fixed $\omega \in \Omega$ for different mesh resolutions with the same initial data. The scheme used is an HLL3 flux with a WENO2 reconstruction algorithm. In this experiment, $\epsilon = 0.01$ and $T = 2$.

$g \in \mathcal{H}_1^{k,p}([0, T], D; U)$ of the ensemble of solutions, for all k, should converge on mesh

refinement. We start by considering observables with respect of the first marginal ν^1 of the approximate statistical solution. In particular, we consider the mean and variance of the density variable ρ, given by

$$M^\Delta(x,t) := \langle \nu_{t,x}^{1,\Delta}, \rho \rangle, \quad V^\Delta(x,t) := \langle \nu_{t,x}^{1,\Delta}, \rho^2 - (M^\Delta(x,t))^2 \rangle. \qquad (5.5)$$

The above quantities are defined a.e. in $D \times [0, T]$ and $\nu^{1,\Delta}$ is the first marginal of the approximate statistical solution $\mu_t^{\Delta,M}$ generated by the Monte Carlo Algorithm. It is straightforward to check that the mean and the variance are admissible observables, in the sense of Convergence Theorem 2.8.

We plot the mean and the variance of the density at time $T = 2$ in Figure 3. As seen from this figure, and in contrast to single samples, the mean and variance clearly converge upon mesh refinement. Moreover, the variance is also concentrated along the so-called *mixing zone*, which spreads out from the initial interface. The convergence of the mean and the variance is further verified from Figure 2(B,C), where the Cauchy rates (5.4) of the mean and variance M^Δ, V^Δ are displayed as a function of mesh resolution.

(a) A single sample, $\Psi^\Delta = \rho^\Delta$

(b) Mean, $\Psi^\Delta = M^\Delta$

(c) Variance, $\Psi^\Delta = V^\Delta$
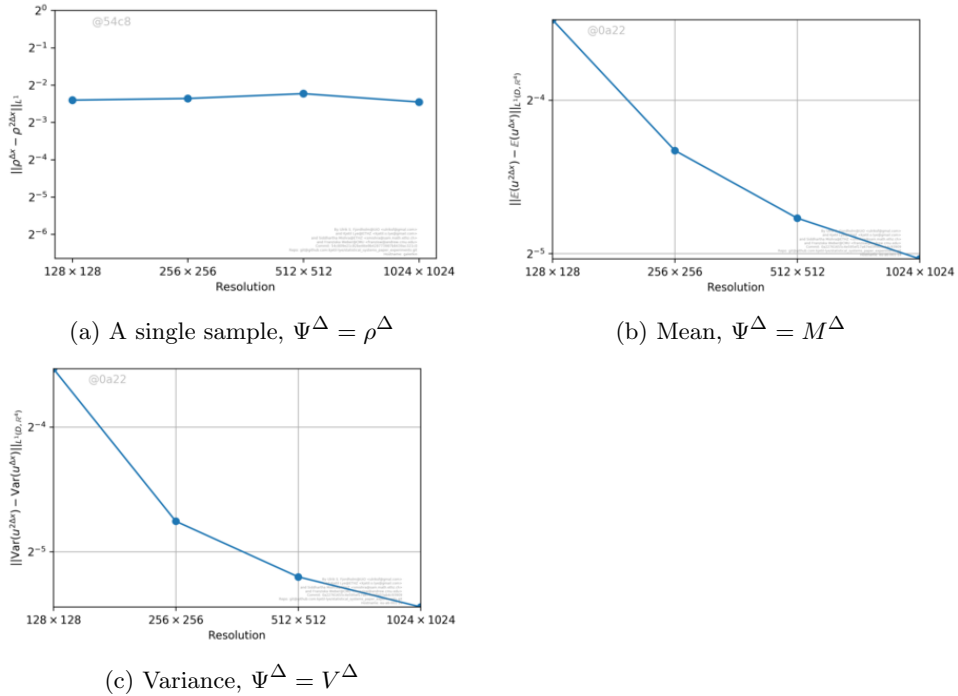
Fig. 2: Cauchy rates (5.4) for the approximate density in the Kelvin–Helmholtz problem (5.2) for different mesh resolutions. The scheme used is an HLL3 flux with a WENO2 reconstruction algorithm. In this experiment, $\epsilon = 0.01$. Here $T = 2$.

(a) Mean at different resolutions.

(b) Variance at different resolutions.
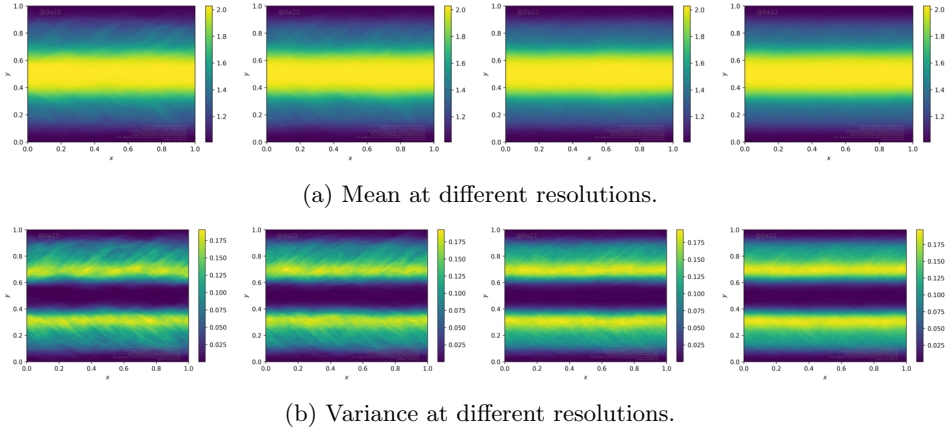
Fig. 3: The approximate mean (top row) and variance (bottom row) of the density of the Kelvin–Helmholtz instability (5.2) for mesh resolutions of (from left to right) 128^2, 256^2, 512^2 and 1024^2 points. The scheme used is an HLL3 flux with a WENO2 reconstruction algorithm. In this experiment, $\epsilon = 0.01$. The number of samples used, M, was set equal to the resolution N, and the final time is $T = 2$.

To quantify the convergence of the distribution of $\mu^{\Delta,M}$ we consider the following Cauchy rates,

$$\mathcal{W}_p^{k,\Delta}(T) := \left(\int_{D^k} W_p\left(\nu_{T,x}^{k,\Delta}, \nu_{T,x}^{k,\frac{\Delta}{2}} \right)^p dx \right)^{\frac{1}{p}}, \tag{5.6}$$

with W_p being the Wasserstein metric defined in (3.1) and $\nu^{k,\Delta}$ is the k-th correlation marginal of the (approximate) statistical solution $\mu_t^{\Delta,M}$ generated by the Monte Carlo Algorithm. One can check that for all $k \in \mathbb{N}$

$$\mathcal{W}_p^{k,\Delta}(t) \leqslant C(k,p) W_p\left(\mu_t^{\Delta}, \mu_t^{\frac{\Delta}{2}} \right) \quad \text{for a.e. } t \in (0,T) \tag{5.7}$$

(see Appendix Appendix E for a proof). As the Wasserstein metric metrizes the weak topology on probability measures, we may conclude from Theorems 2.8 and 4.1 that under the assumptions of some form of time continuity, the right hand side of (5.7) goes to zero as $\Delta \to 0$. This convergence is verified in Figure 4 (A), where we plot the corresponding Cauchy rates for the distance (5.6) with respect to the density (see Appendix Appendix F for details about how the Wasserstein distance was computed numerically).

Next, we consider the computation of observables with respect to the *second correlation marginal* $\nu^{2,\Delta,M}$ of the approximate statistical solution $\mu^{\Delta,M}$. The most interesting observable in this regard is the approximate *structure function* $S_r^p\left(\mu^{\Delta,M}\right)$ (4.7). This is clearly an admissible observable in the sense of Theorem 2.8. For computational purposes, it is easier to compute the *time-sections* of the structure

function, namely

$$\omega_r^p\big(\nu_t^{2,\Delta,M}\big) := \int_D \fint_{B_r(x)} \big\langle \nu_{t,x,y}^{2,\Delta,M}, |\xi_1 - \xi_2|^p \big\rangle \, dy dx \qquad (5.8)$$

for $t \in [0,T]$. In Figure 5 we plot $\omega_r^2\big(\nu_T^{2,\Delta,M}\big)^{1/p}$ for a sequence of mesh sizes Δ. Moreover, we consider three different setups in the figure. In Figure 5 (A) and (B), we set $\epsilon = 0.1$ in (5.3) and $T = 2$ and $T = 4$, respectively, and in Figure 5 (C) we set $\epsilon = 0.01$ and $T = 2$. As seen from the figures, it is clear that the approximate structure functions converge as the mesh is refined. Moreover, the structure functions (approximately) behave as

$$\omega_r^p\big(\nu_T^{2,\Delta,M}\big)^{1/p} \sim C(T) r^{\theta_p(T)}. \qquad (5.9)$$

The computed values of $\theta_p(T)$ are seen in the legend in Figure 5.

The numerical convergence of structure functions is further verified in Figure 4 (B), where we plot the Cauchy rates (5.4) with $g(r,T) = \omega_r^2\big(\nu_T^{2,\Delta,M}\big)$ with $p = 1$. In this figure $T = 2$ and $\epsilon = 0.01$. Clearly, the structure function (as a function of length scale r) converges as the mesh is refined.

Finally, we compute the Wasserstein Cauchy rates $\mathcal{W}_1^{2,\Delta}(T)$ (5.6) with respect to the density, over successively refined mesh sizes and display the result in Figure 4 (C). This figure clearly shows that there is convergence (with respect to mesh resolution and number of Monte Carlo samples) in this metric.

Next, we compute the two-point correlation marginal $\nu_{t,x,y}^{2,\Delta,M}$ for point pairs $(x,y) \in D$. We visualize this Young measure by empirical histograms, plotted in Figure 6. In this figure we show the empirical histogram of the two-point correlation Young measure of the density, on successively refined grids, for two different point pairs $x = (0.7, 0.7)$, $y = (0.4, 0.2)$ (left column) and $x = (0.7, 0.7)$, $y = (0.7, 0.8)$ (right column). The figure indicates that the empirical histograms of the two-point correlation marginals converge on mesh refinement. Moreover, there is a clear difference in the correlation structures at different point pairs.

5.2. *Richtmeyer–Meshkov problem*

Our second test case is the well-studied Richtmeyer–Meshkov problem (see [15] and references therein), which involves a very complicated solution of the compressible Euler equations (5.1), modeling the complex interaction of strong shocks with unstable interfaces. The underlying initial data is given as
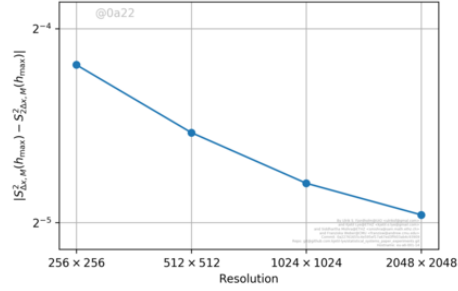
$$p(x) = \begin{cases} 20 & \text{if } |x| < 0.1 \\ 1 & \text{otherwise.} \end{cases} \qquad \rho(x) = \begin{cases} 2 & \text{if } |x| < I(x,\omega) \\ 1 & \text{otherwise} \end{cases} \qquad w^x = w^y = 0 \quad (5.10)$$

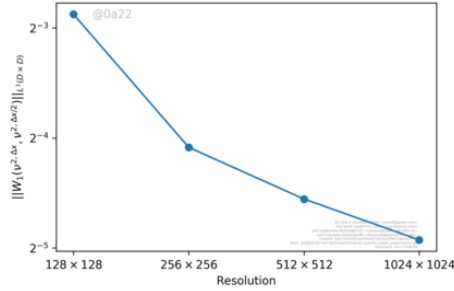We assign periodic boundary conditions on $D = [0,1]^2$. The interface between the two states is given as

$$I(x,\omega) = 0.25 + \epsilon \sum_{j=1}^{K} a_j(\omega) \sin(2\pi(x + b_j(\omega))), \qquad (5.11)$$
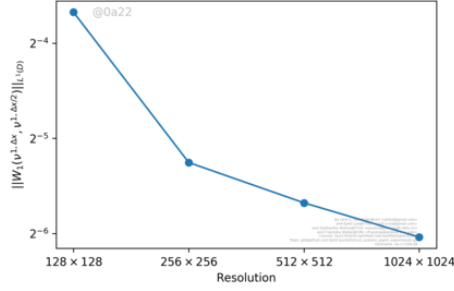
(a) $\mathcal{W}_1^{1,\Delta}(T)$ as in (5.6) (vertical axis) plotted against Δ^{-1} (horizontal axis) with $\epsilon = 0.05$ and $T = 2$.

(b) Cauchy rates (5.4) with $\Psi^\Delta(r, T) = \omega_r^2(\nu_T^{2,\Delta,M})$ with $p = 1$ and $r = 1/32$, $T = 2$ and $\epsilon = 0.01$.

(c) $\mathcal{W}_1^{2,\Delta}(T)$ (5.6) (vertical axis) vs Δ^{-1} (horizontal axis) with $\epsilon = 0.05$ and $T = 2$.

Fig. 4: Cauchy rates for the Kelvin–Helmholtz problem.

where $K = 10$, $\epsilon > 0$, and a_j and b_j (for $j = 1, \ldots, K$) are uniform random variables on the interval $[0, 1]$. We normalize the a_j such that $\sum_j a_j = 1$. The initial probability measure $\bar{\mu}$ is given by the law of the above random field, and lies in $\mathcal{P}(L^p(D; U))$ for every $1 \leqslant p < \infty$.

As in the case of the Kelvin–Helmholtz problem, there is no convergence (on mesh refinement) for single samples (realizations). This non-convergence is demonstrated in Figure 7(A), where the Cauchy rates (5.4) with respect to the density at time $T = 5$ are shown. We visualize the density for different mesh resolutions in Figure 8 (top row). As seen from this figure, the solution at this time is very complicated on account of the interaction between the incoming strong shock (which has been reflected, due to periodic boundary conditions) and the unstable interface, which leads to the generation of turbulent small scale eddies.

On the other hand, and as predicted by the convergence theory developed in Section 4.1, statistical observables such as the mean and the variance (5.5) converge on mesh refinement, as shown in Figure 8 (middle and bottom). Furthermore, this figure shows how the small scale structures are averaged out in the mean, whereas the

(a) $p = 2$, $\epsilon = 0.1$, $T = 2$

(b) $p = 2$, $\epsilon = 0.1$, $T = 4$

(c) $p = 2$, $\epsilon = 0.01$, $T = 2$
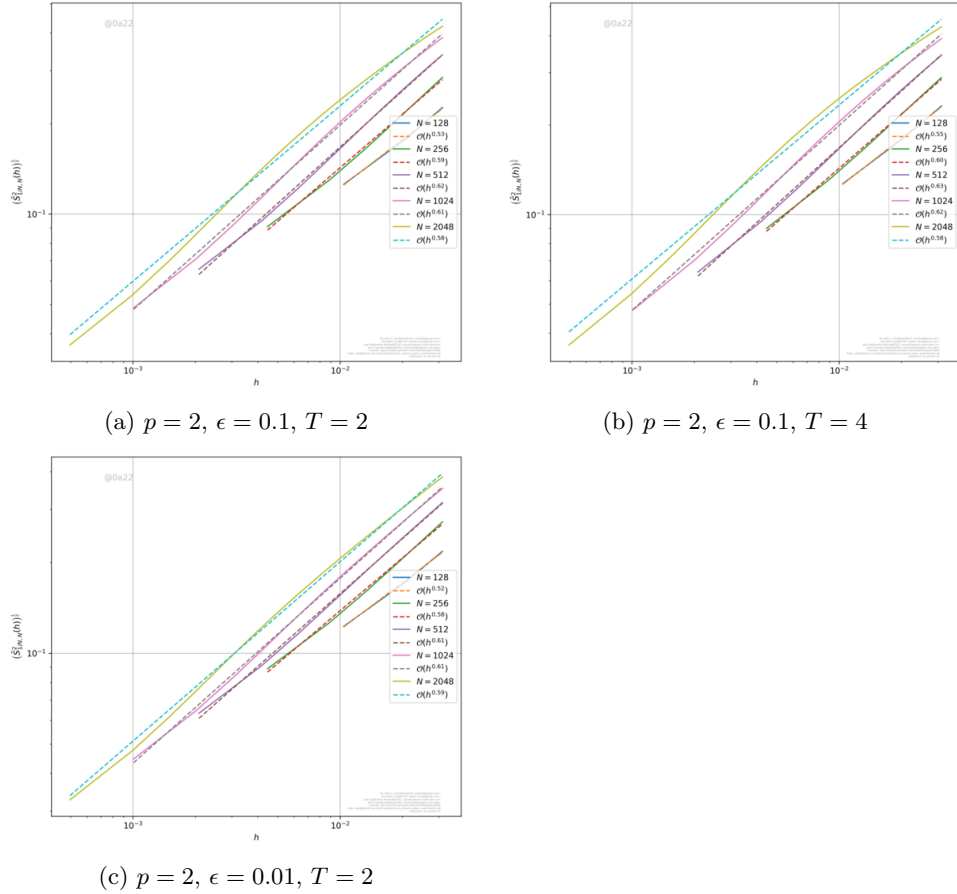
Fig. 5: Structure functions (5.9) for the Kelvin–Helmholtz instability (5.2) for different times T, exponents p and perturbation sizes ϵ. The scheme used is an HLL3 flux with a WENO2 reconstruction algorithm. At each mesh resolution N, $M = N$ samples were used.

small scale information is encoded in the variance, which is concentrated around the mixing zone. We also verify the convergence of the Wasserstein distance $\mathcal{W}_1^{2,\Delta}(T)$ with respect to the density for successively refined meshes in Figure 7 (B).

Next, we compute the time sections of the structure function $\omega_r^2(\nu_t^{2,\Delta,M})$ defined in (5.8). These are shown in Figure 9, where we have used $T = 5$, $r = 2^{-10}, \ldots, 2^{-5}$, and two different values of the perturbation parameter in (5.10). As seen from the figure, the structure function clearly converges on mesh refinement. Moreover, it behaves as in (5.9), with exponents shown in Figure 9. The convergence of the structure function is further verified by plotting the Cauchy rates for the structure function, as a function of the length scale r in Figure 7 (C).
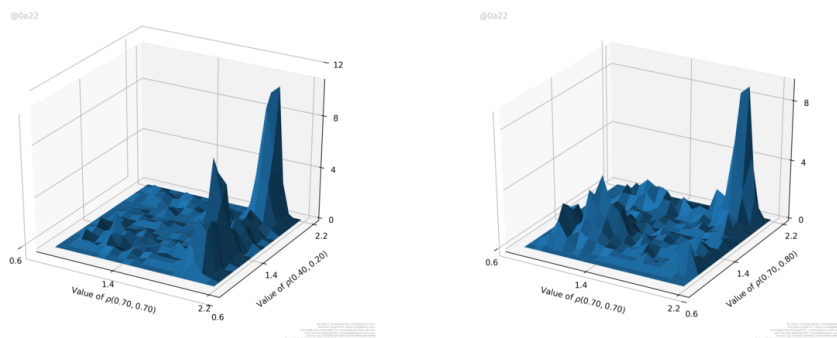
(a) $N = 128$
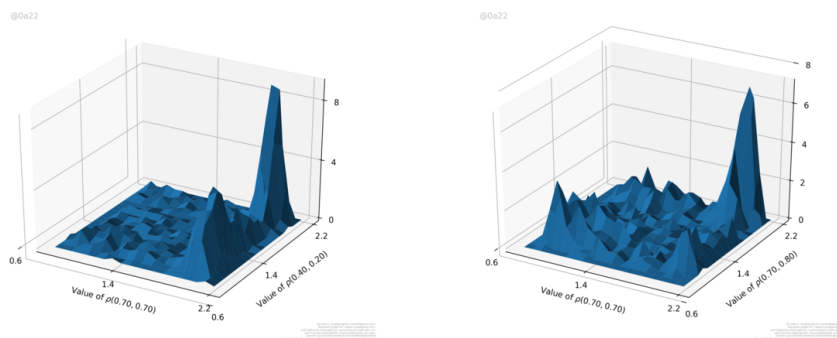
(b) $N = 256$
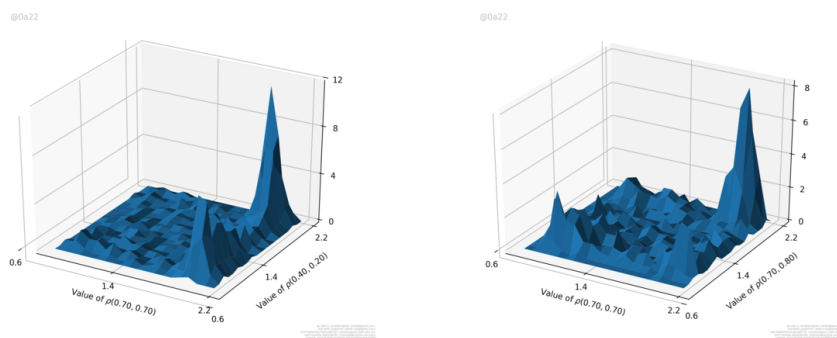
(c) $N = 512$
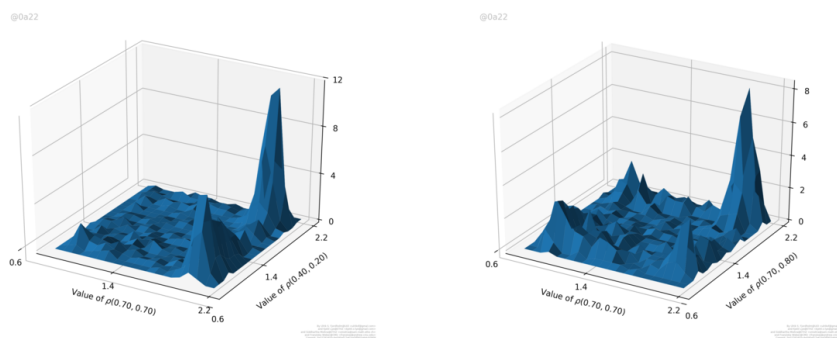
(d) $N = 1024$

Fig. 6: Two-dimensional histograms for the correlation measure at $((0.7, 0.7), (0.4, 0.2))$ (left column) and

(a) A single sample, $\Psi^\Delta = \rho^\Delta$ in (5.4).

(b) $\mathcal{W}_1^{1,\Delta}(T)$, defined in (5.6).

(c) $\Psi^\Delta = \omega_r^2(\nu_T^{2,\Delta,M})$ with $p = 1$ and $r = 1/32$.

(d) $\mathcal{W}_1^{2,\Delta}(T)$.
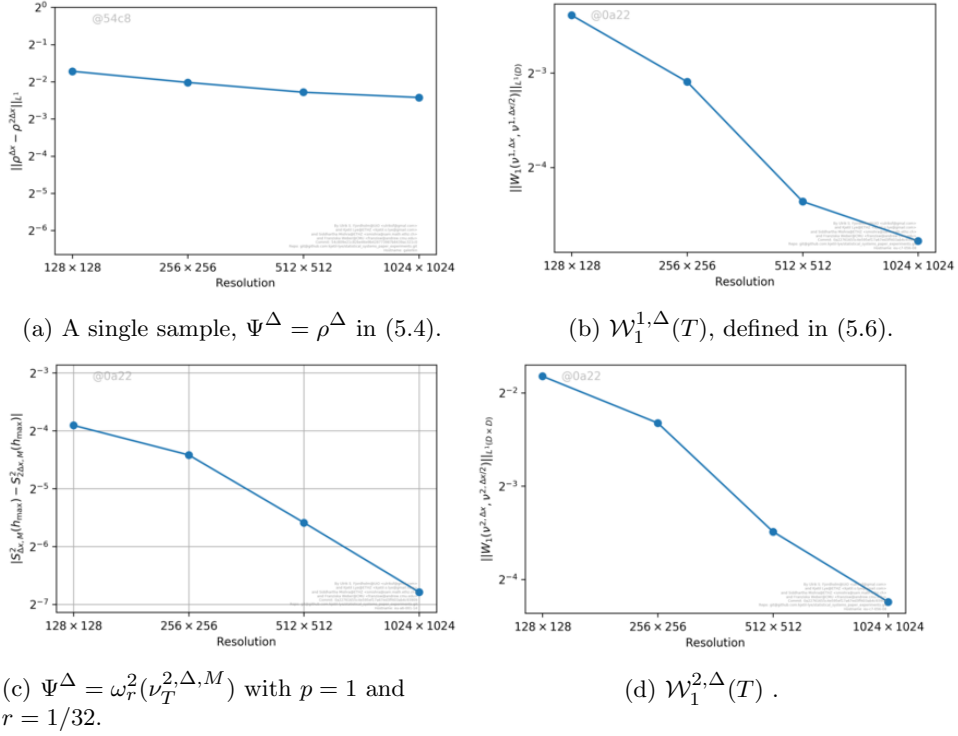
Fig. 7: Cauchy rates (5.4) (vertical axis) versus Δ (horizontal axis) for the Richtmeyer–Meshkov problem using $\epsilon = 0.06$ and $T = 5$.

In Figure 7(D) we plot the Wasserstein distance $\mathcal{W}_1^{2,\Delta}(T)$ for the density and $T = 5$, on a sequence of successively refined meshes. As shown in the figure, this distance converges on mesh refinement.

Finally, in Figure 10, we plot histograms that represent the two-point correlation measure for the density at two different point pairs and at time $T = 5$. These histograms show that the two-point correlation structure for this initial datum is very different from the correlation structure for the Kelvin–Helmholtz problem (Figure 6).

5.3. *Fractional Brownian motion*

The initial probability measure $\bar{\mu} \in \mathcal{P}(L^p(D; U))$ in the previous two numerical experiments was realized as a probability measure on high- but finite-dimensional subsets of L^p. We now consider initial probability measures that are concentrated on genuinely infinite-dimensional subsets of $L^p(D; U)$.

We will assume that the initial probability measure for the two-dimensional compressible Euler equations (5.1) corresponds to a *fractional Brownian motion*.
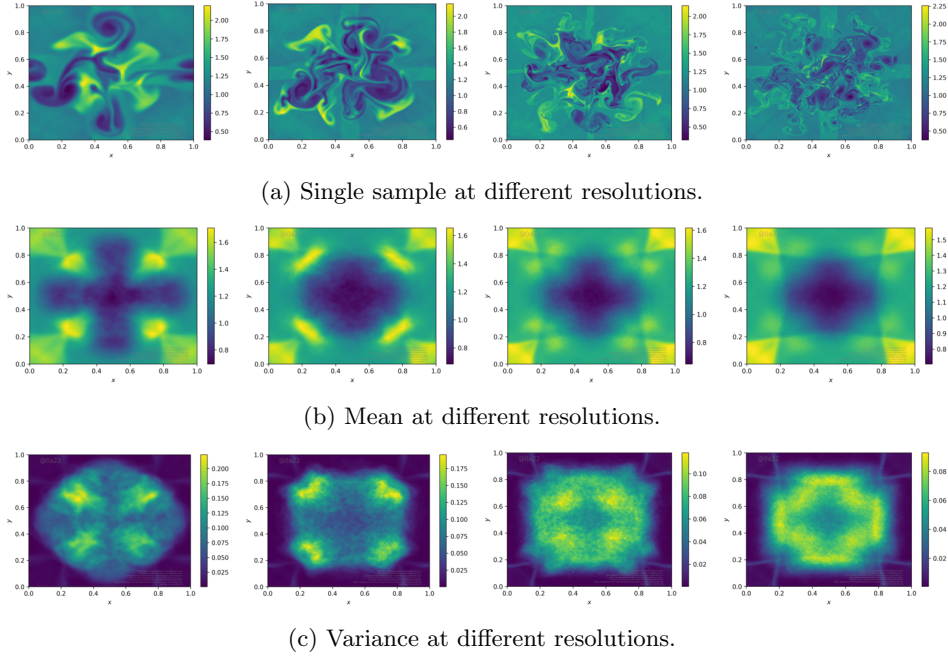
(a) Single sample at different resolutions.

(b) Mean at different resolutions.

(c) Variance at different resolutions.

Fig. 8: Approximate density for the Richtmeyer–Meshkov problem (5.10) using $\epsilon = 0.06$ and at $T = 5$. All results are based on a scheme with the HLLC flux and MC reconstruction, computed at resolutions with (from left to right) 128^2, 256^2, 512^2 and 1024^2 points.

Introduced by Mandelbrot et al. [38], fractional Brownian motion can be seen as a generalization of standard Brownian motion with a scaling exponent different from $1/2$.

We consider the following initial data:

$$
\begin{aligned}
w_0^{x,H}(\omega; x) &:= B_1^H(\omega; x), \qquad \rho_0 = 4, \\
w_0^{y,H}(\omega; x) &:= B_2^H(\omega; x), \qquad p_0 = 2.5,
\end{aligned} \qquad \text{for } \omega \in \Omega, \ x \in [0,1]^2
$$

where B_1^H and B_2^H are two independent two-dimensional fractional Brownian motions with Hurst index $H \in (0,1)$. Standard Brownian motion corresponds to a Hurst index of $H = 1/2$.

To generate fractional Brownian motion, we use the random midpoint displacement method originally introduced by Lévy [35] for Brownian motion, and later adapted for fractional Brownian motion [23,45]. Consider the uniform grid $0 = x_{1/2} < \cdots < x_{N+1/2} = 1$ with $x_{i+1/2} = i\Delta$ and $\Delta = \frac{1}{N}$, where $N = 2^k + 1$ is the number of cells for some $k \in \mathbb{N}$. We first fix the corners

$$
w_{1,N}^{x,H,\Delta x}(\omega; 0) = w_{1,1}^{x,H,\Delta x}(\omega; 0) = w_{N,N}^{x,H,\Delta x}(\omega; 0) = w_{N,1}^{x,H,\Delta x}(\omega; 0) = 0 \qquad \omega \in \Omega,
$$

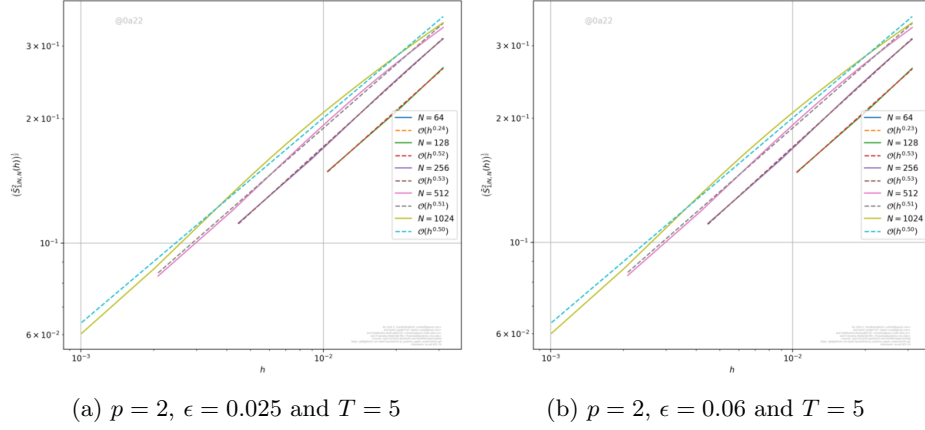(a) $p = 2$, $\epsilon = 0.025$ and $T = 5$ (b) $p = 2$, $\epsilon = 0.06$ and $T = 5$

Fig. 9: Structure functions for the Richtmeyer–Meshkov problem (5.10) for different perturbation sizes ϵ. The scheme used is an HLL3 flux with a MC reconstruction algorithm. At each mesh resolution N, $M = N$ samples were used.

recursively update the values on the edges as

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

and correspondingly for $w^{x,H,\Delta x}_{2^{k-l}j,2^{k-l-1}(2i+1)}(\omega;0)$. For the values in the center of the cells, we use the following expression:

$$w^{x,H,\Delta x}_{2^{k-l-1}(2j+1),2^{k-l-1}(2j+1)}(\omega;0) = \frac{1}{4}\left(w^{x,H,\Delta x}_{2^{k-l}(j+1),2^{k-l}i}(\omega;0) + w^{x,H,\Delta x}_{2^{k-l}j,2^{k-l}i}(\omega;0)\right.$$
$$\left. w^{x,H,\Delta x}_{2^{k-l}(j+1),2^{k-l}(i+1)}(\omega;0) + w^{x,H,\Delta x}_{2^{k-l}(j+1),2^{k-l}(i+1)}(\omega;0)\right)$$
$$+ \sqrt{\frac{1-2^{2H-2}}{2^{2lH}}}X_{2^l+j,2^{k-l}i}(\omega)$$

for $l = 0, \ldots, k$ and for $i, j = 0, \ldots, 2^l$. Here $(X_{k,n})_{(k,n) \in \mathbb{N}^2}$ is a collection of normally distributed random variables with mean 0 and variance 1. That is, we bisect every cell and set the middle value to the average of the neighbouring values plus some Gaussian random variable. The same procedure is repeated for $w^{y,H,\Delta}$. See Figure 11 for a sample of the initial velocity field with standard Brownian motion, i.e., with $H = 0.5$.

The initial probability measure is given by the law of the above random field and the dimension of its support increases with decreasing mesh size. Hence, in the limit $\Delta \to 0$ we are approximating a probability measure supported on an infinite dimensional subspace of $L^2(D; U)$
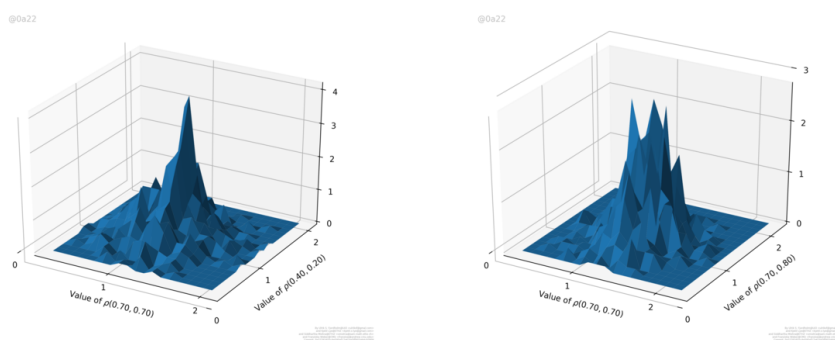
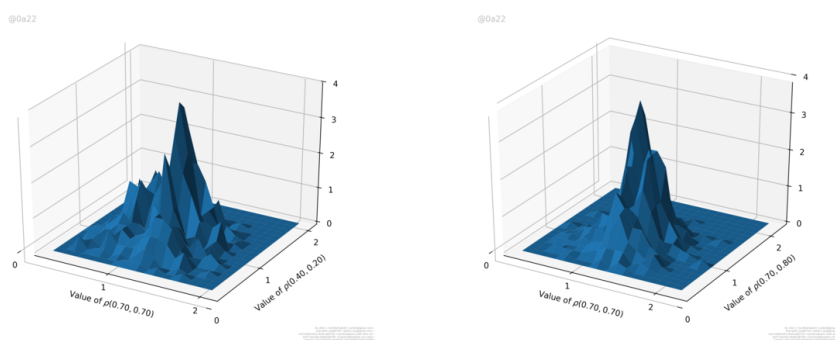42 *U. S. Fjordholm, K. Lye, S. Mishra and F. Weber*
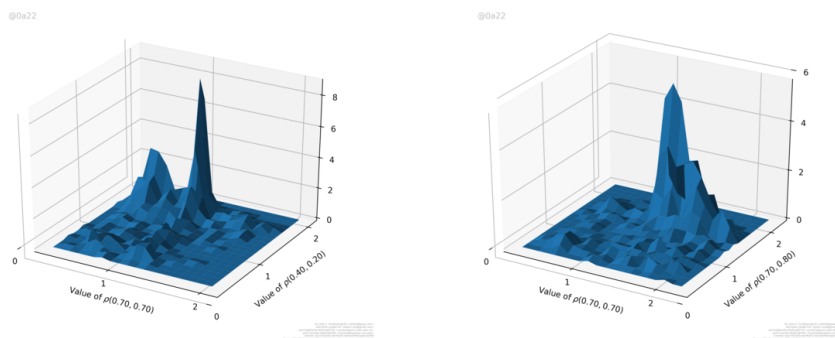
(a) $N = 128$
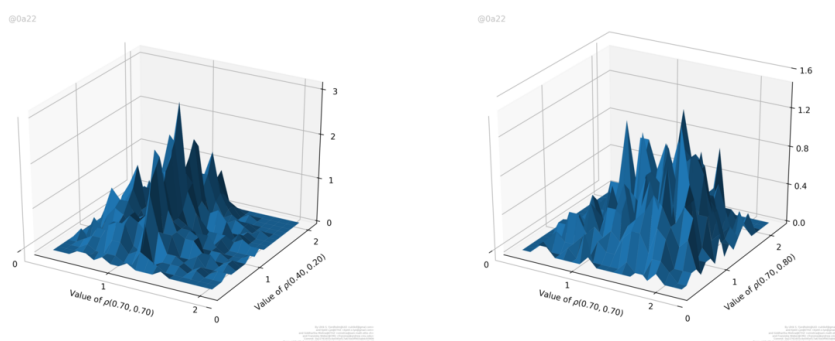
(b) $N = 256$

(c) $N = 512$

(d) $N = 1024$

Fig. 10: Two-dimensional histograms for the correlation measure at $((0.7, 0.7), (0.4, 0.2))$ (left column) and $((0.7, 0.7), (0.4, 0.2))$ (right column) for different resolutions for the density in the
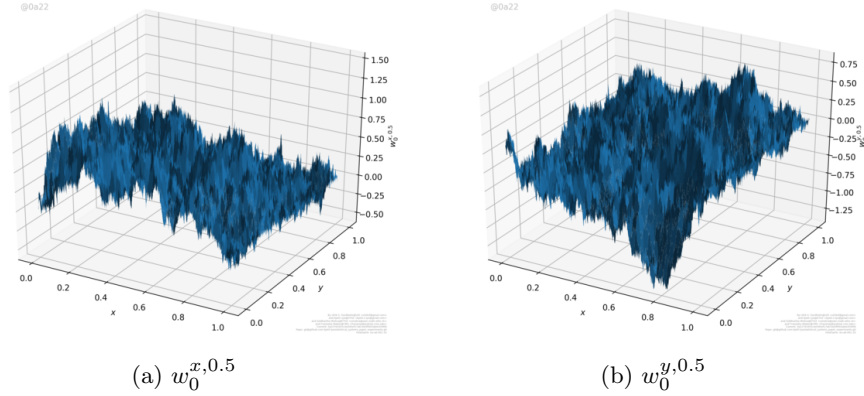
(a) $w_0^{x,0.5}$ (b) $w_0^{y,0.5}$

Fig. 11: Two samples of velocity fields with the Brownian motion initial data.

We compute the statistical solutions with the Monte Carlo Algorithm with the fractional Brownian motion initial data for two different Hurst indices, $H = 0.1$ and $H = 0.5$. Statistical observables corresponding to the one-point correlation marginal, such as the mean and variance, converge on mesh refinement, as shown in Figure 12. We also plot the mean and variance of the density at the highest mesh resolution of 1024^2 grid points and final time $T = 0.25$, for the two different Hurst indices, in Figure 13. As seen from the figure, there is a clear difference in the spatial structure of the mean and the variance as the Hurst index is changed. Moreover, the spatial structure of these statistical quantities is much more complicated than in the case of the Kelvin–Helmholtz and Richtmeyer–Meshkov initial data, with no clear large scale structures such as shocks. On the other hand, the statistical quantities have more small-scale structures. This is more pronounced in the $H = 0.1$ case than for standard Brownian motion.

For $r = 2^{-10}, \ldots, 2^{-5}$ we plot the (time sections of) the structure function $\omega_r^2(\nu_t^{2,\Delta,M})$ at time $T = 0.25$ in Figure 14. The structure functions clearly converge on mesh refinement. This is also verified for both Hurst indices in Figure 15 (A,B), where we plot the Cauchy rates (5.4) for the structure function, with respect to the length scale r. Moreover, the structure functions scale as in (5.9).

In Figure 15 (C,D), we plot the Cauchy rates of the Wasserstein distance $\mathcal{W}_1^{2,\Delta}(0.25)$ with respect to grid resolution, for both Hurst indices. We verify from this figure that these distances also converge on mesh refinement and sample augmentation.

Finally, in Figure 16, we plot histograms representing the two-point correlation marginals of the density, computed on the finest grid resolution of 1024^2 grid points, for two different point pairs. The figure shows that the two-point correlation structure is again very different for different Hurst indices, and from the correlation structures for the previous numerical experiments.

(a) $H = 0.1$, mean

(b) $H = 0.1$, variance

(c) $H = 0.5$, mean

(d) $H = 0.5$, variance

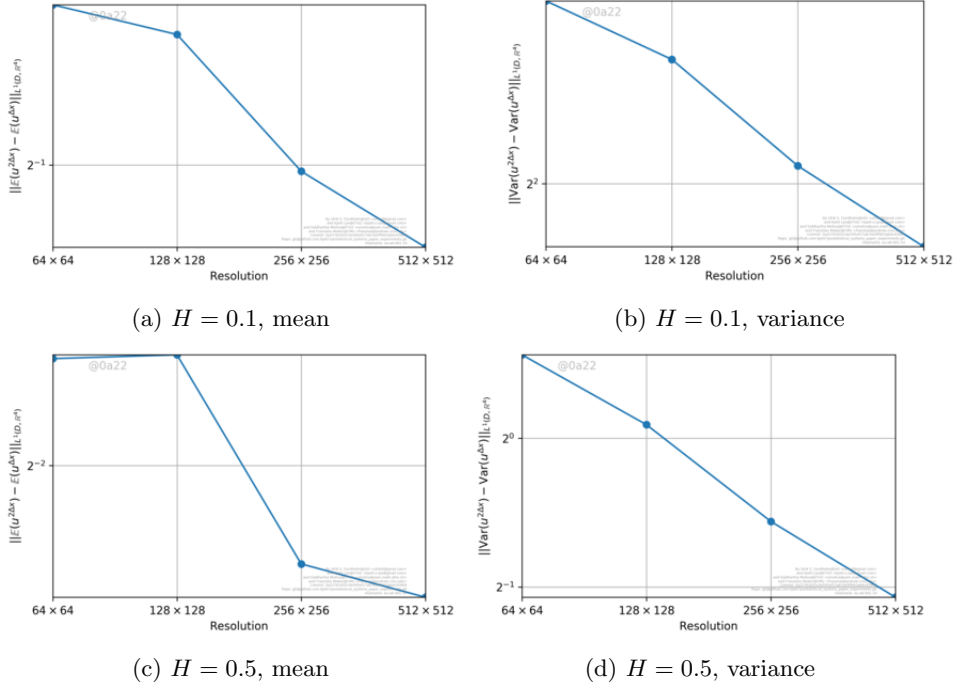Fig. 12: Cauchy rates (5.4) for the mean and the variance of the density ρ for the fractional Brownian motion initial data at time $T = 0.25$. The scheme used is an HLL3 flux with a WENO2 reconstruction algorithm.
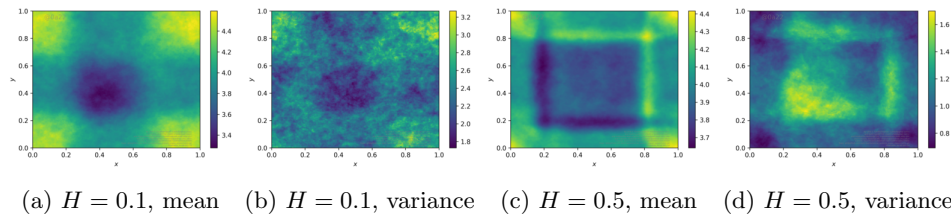

(a) $H = 0.1$, mean   (b) $H = 0.1$, variance   (c) $H = 0.5$, mean   (d) $H = 0.5$, variance

Fig. 13: The mean and the variance (of the density) at the highest resolution of 1024^2 mesh points and 1024 Monte Carlo samples, for fractional Brownian motion initial data with two different Hurst indices and at time $T = 0.25$. The scheme used is an HLL3 flux with a WENO2 reconstruction algorithm.

5.4. *Stability of the computed statistical solution*

A priori, the computed statistical solution depends on the specifics of the underlying initial probability measure $\bar{\mu}$ as well as on the details of the numerical scheme
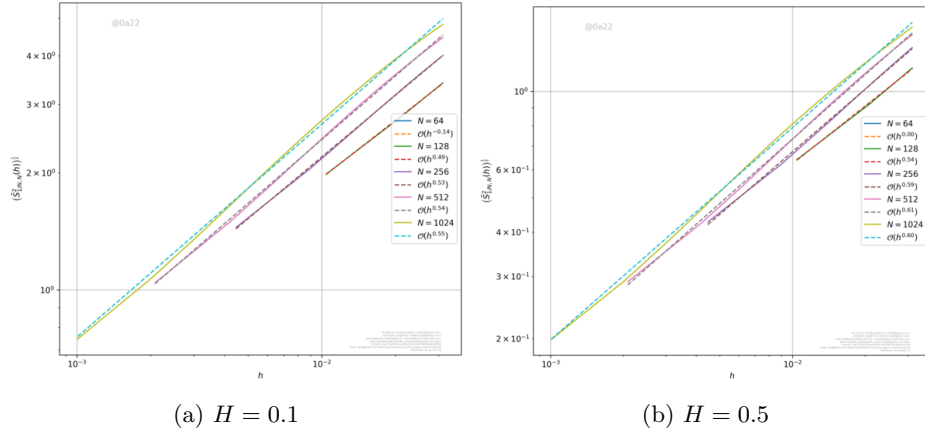
Fig. 14: Structure function (5.8) for $p = 2$ for different grid resolutions at time $T = 0.25$ for two different Hurst indices, corresponding to the fractional Brownian motion initial data.

(4.1), used within the Monte Carlo Algorithm. We investigate the stability of the computed statistical solution with respect to these parameters in the specific case of the Kelvin–Helmholtz problem (5.2).

In [15], the authors had already demonstrated the stability of numerically computed measure-valued solutions with respect to variations of the underlying numerical method, or to the size and type of perturbations to the Kelvin–Helmholtz initial data. As the computed measure-valued solution in [15] is identical to the first correlation marginal of our computed statistical solution, we can deduce that the observables with respect to the first correlation marginal are also stable. Therefore, we investigate the stability of observables with respect to the second correlation marginal $\nu^{2,\Delta,M}$. The results are summarized below.

- *Stability with respect to amplitude of perturbations.* We vary the size of the perturbation parameter ϵ in (5.3) over two orders of magnitude, from $\epsilon = 0.001$ to $\epsilon = 0.1$. The computed (time section of) structure function (5.8) for $p = 2$ and at time $T = 2$, on the finest resolution of 1024^2 points and 1024 Monte Carlo samples, is shown in Figure 17 (A, left). As seen from the figure, the computed structure functions are very close to each other and scale as (5.9) with $\theta \approx 0.61$. This indicates stability of the computed structure function with respect to the amplitude of perturbations in the initial data. This stability is further verified in Figure 17 (A, right) where we plot the Wasserstein distance $\left\| W_1\left(\nu^{2,\epsilon}_{T,x}, \nu^{2,\frac{\epsilon}{2}}_{T,x}\right)\right\|_{L^1(D^2)}$ with respect to the density, at time $T = 2$ for different values of the perturbation parameter. The plot shows (linear) convergence with the decay of the perturbation, indicating stability of the computed statistical
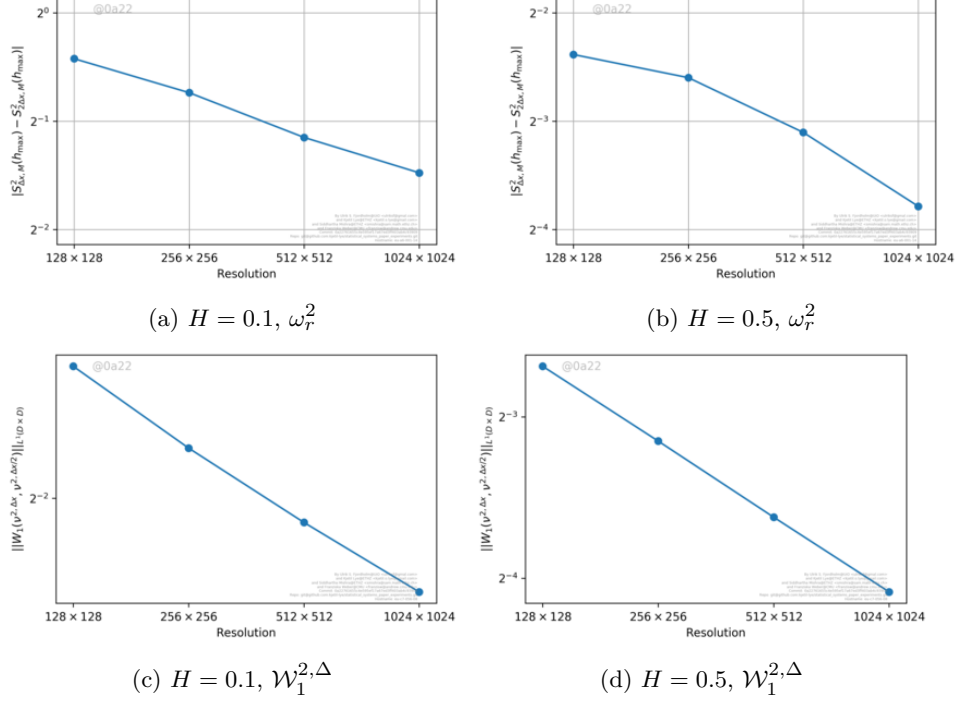
46   *U. S. Fjordholm, K. Lye, S. Mishra and F. Weber*

(a) $H = 0.1$, ω_r^2

(b) $H = 0.5$, ω_r^2

(c) $H = 0.1$, $\mathcal{W}_1^{2,\Delta}$

(d) $H = 0.5$, $\mathcal{W}_1^{2,\Delta}$

Fig. 15: Convergence for different two-point statistical observables for the fractional Brownian motion initial data at time $T = 0.25$. Top row: (Time sections of) structure function (5.8) with $p = 2$. Bottom row: Wasserstein distance (5.6) with $p = 1$, $k = 2$.

solution *vis a vis* perturbation amplitude.

- *Stability with respect to type of perturbations.* In all the numerical experiments for the Kelvin–Helmholtz initial data, we have assumed that the random variables a_j, b_j in (5.3) are chosen from a uniform distribution. Here, we choose these random variables from a standard normal distribution. This amounts to varying the corresponding initial probability measure for (5.2). The consequent change in the structure function (5.8) for two different amplitudes of the perturbation parameter ϵ in (5.3) are shown in Figure 17 (B, left). The figure clearly shows that the computed structure functions are very close to the ones computed with the uniform distribution. This stability with respect to the type of perturbation is further verified in Figure 17 (B, right) where we plot the $\left\| W_1 \left(\nu_{T,x}^{2,\epsilon}, \hat{\nu}_{T,x}^{2,\epsilon} \right) \right\|_{L^1(D^2)}$ at time $T = 2$. Here, $\nu, \hat{\nu}$ refer to the correlation measures, computed with the uniform and standard normal random variables, respectively. The plot shows convergence with the decay of the perturbation, indicating stability of the computed statistical solution, *vis a vis* perturbation

(a) $H = 0.1$
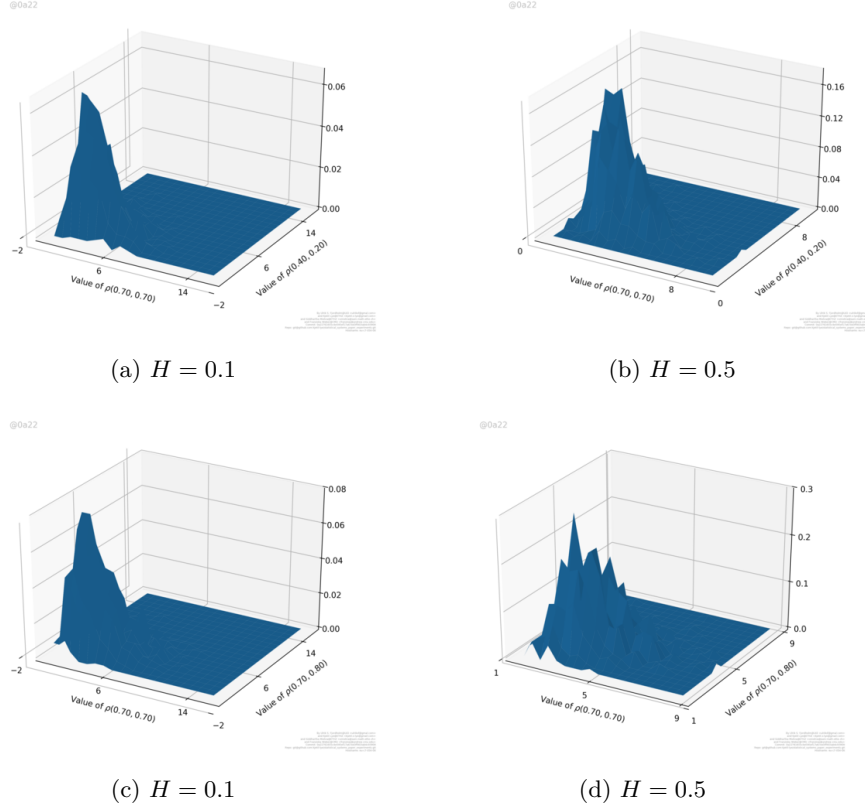
(b) $H = 0.5$

(c) $H = 0.1$

(d) $H = 0.5$

Fig. 16: Histograms representing the two-point correlation marginal for the density at time $T = 0.25$ and two different Hurst indices, at two different point pairs. Top row: At points $(0.4, 0.2)$ and $(0.7, 0.8)$. Bottom row: At points $(0.7, 0.7)$ and $(0.7, 0.8)$. All figures are generated with mesh resolution of 1024^2 points and with 1024 samples.

type.

- *Stability with respect to choice of numerical scheme.* In order to investigate the stability of the computed statistical solutions to the choice of the underlying numerical scheme in (4.1), we vary the reconstruction procedure, i.e., we use a high-resolution finite volume scheme based on the HLLC flux, but with either MC or WENO2 reconstructions (see e.g. [34]). The choice of the reconstruction leads to change in the sub-grid scale numerical viscosity of the overall approximation. We plot the structure function (5.8) in Figure 17 (C, left) and observe a very minor change in the structure function. This issue is investigated further in Figure 17 (C, right) where the Wasserstein distances

$\left\| W_1 \left(\nu^{2,\Delta}_{\text{WENO2},T,x}, \nu^{2,\Delta}_{\text{MC},T,x} \right) \right\|_{L^1(D^2)}$ at time $T = 2$ are plotted. Here $\nu^{2,\Delta}_{\text{WENO2}}$ is the second correlation marginal, computed with the WENO2 reconstruction procedure, and $\nu^{2,\Delta}_{\text{MC}}$ is the second correlation marginal, computed with the MC reconstruction procedure. We observe convergence of this distance with respect to resolution. This allows us to conclude that the statistical solutions are stable with respect to the choice of the underlying numerical method, at least for this Kelvin–Helmholtz problem.

5.5. *Statistical steady state and regularity*

In addition to the four numerical experiments reported in the last section, i.e., Kelvin–Helmholtz, Richtmeyer–Meshkov, and fractional Brownian motion with two different Hurst indices of $H = 0.1$ and $H = 0.5$, we have performed two further numerical experiments. Both of them consider the two-dimensional compressible Euler equations with the following initial data:

- Fractional Brownian motion initial data with Hurst index $H = 0.75$.
- Shock-vortex interaction initial data, see Section 6.3.2 of [18] and references therein.

The Monte Carlo Algorithm is used to compute the approximate statistical solution $\mu^{\Delta,M}_t$ for these additional sets of initial data.

We focus on the (time-sections of) the structure function (5.8) and find that in all six numerical experiments, the structure function behaved as (5.9). The exponent $\theta_p(t)$ as a function of time, for $p = 1, 2, 3$ and for each numerical experiment is shown in Figure 18. We observe the following from this figure.

- First, the exponent $\theta_p(t)$ reaches a steady state rather quickly, when compared to the dynamic behaviour of the solution. In other words, Figure 18 seems to suggest that statistical equilibrium is reached significantly faster than the (deterministic) steady state for individual realizations. Hence, the system evolves dynamically for each sample, while the whole ensemble has already reached statistical equilibrium. The time scale at which this statistical equilibrium is reached depends on the specifics of the initial data.
- For all the experiments except the shock-vortex interaction, there is a very interesting behavior of the structure function (5.8) with respect to time. In particular, the exponent $\theta_p(t)$ for $p = 2, 3$ clearly increases with time, indicating that the non-linear evolution *statistically regularizes* the solution in some manner. The exception is for the shock-vortex interaction where this exponent remains constant with time. This can be explained by the fact that the shock-vortex interaction results in a solution whose total variation (TV) norm is bounded. Hence, one can readily verify that $\theta_p(t) = \frac{1}{p}$, which is approximately realized in the computations. On the other hand and, as shown in Figure 19, the (average) BV-norm blows up for all the remaining test cases. Hence, in these problems,
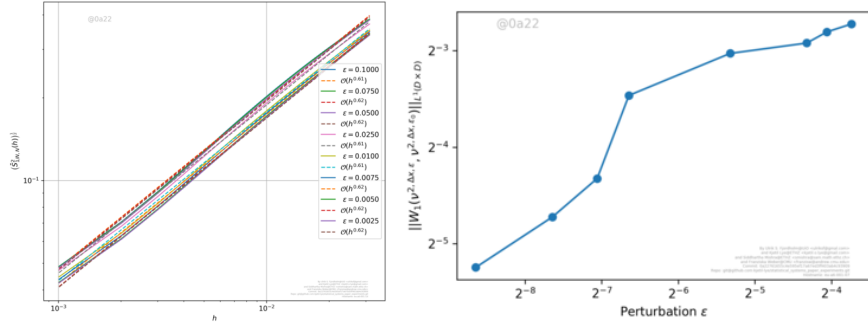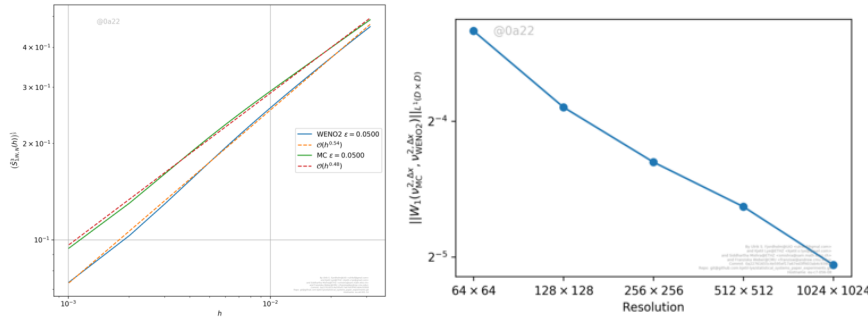
(a) Different amplitudes of perturbation.
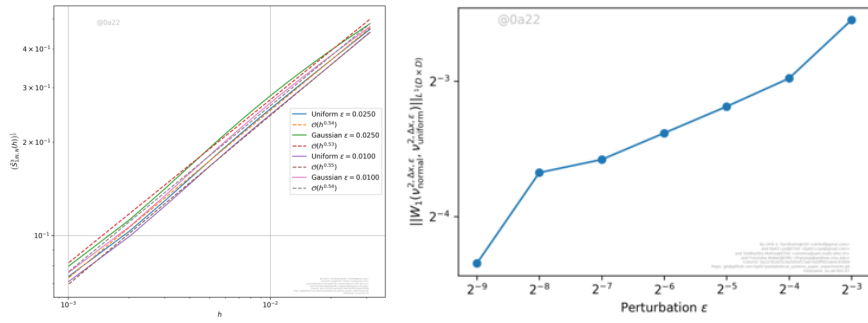
(b) Different distributions of parameters.


(c) Different numerical schemes.

Fig. 17: Stability of the statistical solution with respect to variations of different parameters in the Kelvin–Helmholtz problem (5.2). Left column: The structure function (5.8) for $p = 2$. Right column: Different Wasserstein distances. All computations are at time $T = 2$, computed on a fine grid of 1024^2 points and with 1024 Monte Carlo samples.

there is a complex interaction of structures at different length scales that leads to a subtle statistical regularity.
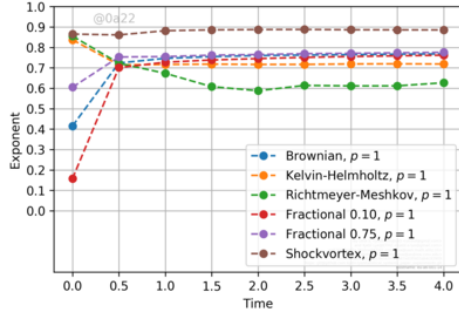
(a) $p = 1$
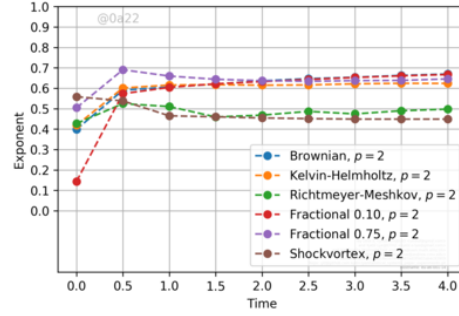
(b) $p = 2$
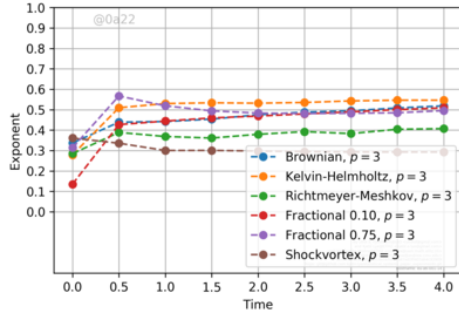
(c) $p = 3$

Fig. 18: The evolution of the approximate scaling exponents of the structure functions as a function of time.

Fig. 19: BV norm as a function of resolution for different initial data.

- Last, but not least, we observe that in all the numerical experiments we have considered, the structure functions scale as (5.9). Hence, the approximate scaling assumption (4.6) in Theorem 4.1 is always observed to be satisfied.

5.6. *Reproducing the numerical experiments*
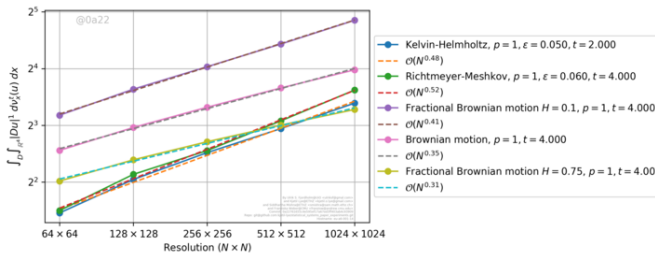
All experiments were carried out using the open source Alsvinn simulator [36]. Please consult the two GitHub repositories

```
     https://github.com/kjetil-lye/systemspaper_experiments
 https://github.com/kjetil-lye/statistical_systems_paper_experiments
```

for a full description on how the experiments were carried out, along with the raw data and post processing scripts.

6. Discussion

We consider hyperbolic systems of conservation laws (1.1). Although both scalar conservation laws and one-dimensional systems are well-posed within the standard solution framework of entropy solutions, it is now clearly established that entropy solutions for multi-dimensional systems are not unique, nor are they amenable to numerical approximation. On the other hand, numerical evidence presented in [15,19] and references therein suggests that a statistical notion of solutions might be more appropriate for (1.1), even if the initial data and other underlying parameters are deterministic.

Entropy measure-valued solutions [13,15] present one possible solution framework. Although global existence and numerical approximation results for entropy measure-valued solutions are available, it is well-known that measure-valued solutions are non-unique, even for scalar conservation laws. As argued in [19], this is largely on account of a lack of information about multi-point spatial correlations.

Inspired by the need to incorporate correlations, the authors of [16] proposed the framework of *statistical solutions* for hyperbolic systems of conservation laws. Statistical solutions are time-parameterized probability measures on the space of p-integrable functions. They were shown in [16] to be equivalent to adding information about all possible multi-point correlations to the measure-valued solution. The time-evolution of these measures is prescribed by a system of nonlinear tensorized moment-transport equations (2.30). Under an additional entropy condition, the well-posedness of statistical solutions for scalar conservation laws was shown in [16] and the numerical approximation of statistical solutions for scalar conservation laws was considered in [17].

Our main aim in this paper was to propose a numerical algorithm to approximate statistical solutions of multi-dimensional hyperbolic systems of conservation laws. To this end, we combined a high-resolution finite volume method (4.1) with a Monte Carlo sampling procedure to obtain our Monte Carlo Algorithm which computes statistical solutions.

Our second aim was to prove a weak-strong uniqueness result, in the sense of e.g. [46]. We proposed an entropy condition and a notion of dissipative statistical solutions (Definition 3.2) and proved a weak-strong uniqueness result for these dissipative statistical solutions. This provides us with a conditional uniqueness result

for statistical solutions, i.e., if they exist, then strong statistical solutions (Definition 3.3) are unique. In particular, we obtain short time existence and uniqueness results for dissipative statistical solutions.

The task of proving convergence of the numerical approximations $\mu_t^{\Delta,M}$, generated by the Monte Carlo Algorithm, was rather intricate. First, we had to completely characterize an appropriate topology on the space of time-parameterized probability measures on p-integrable functions. This topology is based on the topology induced by the underlying correlation measures. We showed that this induced topology is equivalent to the weak topology on the space of probability measures on L^p, for any fixed time, but it also induces appropriate extensions when time is varied. The resulting compactness theorem 2.8 delineates the class of admissible observables that converge in this topology. Essentially, this theorem boils down to the convergence of time averages of (multi-point) statistical quantities of interest such as the mean, variance, multi-point correlation functions and structure functions (4.7). We believe that this topology on time-parameterized probability measures on L^p, and novel sufficient conditions for ensuring convergence in it, might have independent applications in probability theory and stochastic analysis.

Next, we proved in Theorem 4.1 that, under certain assumptions on the underlying finite volume schemes, the approximate statistical solutions converge upon mesh refinement in the aforementioned topology to a time-parameterized probability measure on L^p. A Lax–Wendroff theorem was proved, showing that the limit measure is indeed a statistical solution. Finally, a standard Monte Carlo convergence argument was used to guarantee convergence, under sample augmentation, of the approximations generated by the Monte Carlo Algorithm.

The assumptions in Theorem 4.1 include L^p stability and the weak BV bound (4.5), which are satisfied by many existing high-resolution entropy stable finite volume schemes, such as the TeCNO schemes of [18]. On the other hand, we also required a subtle (approximate) scaling assumption (4.6) that is an analogue of the well-known scaling assumptions in Kolmogorov's theory for homogeneous, isotropic turbulence [24]. Although this assumption was verified in all the numerical experiments, we were unable to prove it here. Hence, we have provided a conditional existence result for statistical solutions of multi-dimensional systems of conservation laws in this paper.

We present extensive numerical experiments for the two-dimensional compressible Euler equations to illustrate our Monte Carlo Algorithm. The results validate the convergence analysis and demonstrate convergence, on mesh refinement and sample augmentation, for statistical quantities of interest such as the mean, variance, structure functions, one-point probability density functions and two-point joint probability density functions. Moreover, we observe convergence in appropriate Wasserstein distances (5.6) for the multi-point correlation measures. Summarizing the results of the numerical experiments, we conclude that one observes *convergence of all interesting statistical observables* in our framework. This should

be contrasted to the state of the art, where deterministic quantities do not converge on mesh refinement [15]. Thus, we provide rigorous justification of the computability of statistical quantities of interest in the context of multi-dimensional systems of conservation laws.

Furthermore, we discover from the numerical experiments that

- The computed statistical solutions are remarkably stable with respect to different variations. In particular, we varied the amplitude of initial perturbations, the type of initial perturbations leading to different probability measures on L^p, and also the underlying numerical method. In all these cases, we observed that the computed statistical solutions were stable with respect to these perturbations. This observed stability augurs very well for identifying further constraints or admissibility criteria in order to obtain uniqueness of statistical solutions.
- The correlation structure of the statistical solutions seem to reach an equilibrium at significantly shorter time scales than the actual flow. This behavior is clearly seen in the variation of structure functions over time (Figure 18). Thus, statistical stationarity might be reached much faster than the actual evolution would suggest.
- There seems to be a subtle gain in regularity, as measured by the decay exponent of the structure function, for the statistical solutions (Figure 18). Qualitatively, it seems that mixing of the underlying structures leads to a gain in regularity. This observed regularity needs to be studied further.

The results in this paper can be extended in different directions. On the theoretical side, a key question is whether the scaling assumption (4.6) can be proved for some numerical approximations or relaxed in an appropriate manner. This would pave the way for a unconditional global existence result for statistical solutions.

As formulated, (dissipative) statistical solutions are not necessarily unique. If we start with a deterministic initial data, i.e., setting $\bar{\mu} = \delta_{\bar{u}}$, then we can apply the construction of [11,9] to obtain infinitely many entropy solutions for the same initial data, by the fact that each deterministic solution defines a statistical solution. Thus, infinitely many statistical solutions are possible for the same (deterministic) initial data. On the other hand, numerical experiments strongly suggest that the computed statistical solutions are stable. Thus, we need to find further admissibility criteria to single out a unique statistical solution. Moreover, the observed gain in regularity might provide additional constraints to obtain uniqueness.

On the computational side, the Monte Carlo Algorithm can be very expensive, even prohibitively expensive in three space dimensions. Hence, it is imperative to consider alternatives to accelerate it. Alternatives such as the multi-level Monte Carlo method and quasi-Monte Carlo method are considered in a forthcoming publication. Another alternative would be to use deep learning, such as in [37], to accelerate the Monte Carlo algorithm.

Appendix A. Proof of Theorem 2.2

In order to characterize weak convergence of probability measures on $L^p(D; U)$, we will use a construction from [2] which we summarize here. Assume that X is Polish (i.e., a complete and separable metric space) and let $X_0 \subset X$ be a dense, countable subspace. For fixed $q_1, q_2, q_3 \in \mathbb{Q}$ with $q_2, q_3 \in (0, 1)$ and every $v \in X_0$, define $H : X \to \mathbb{R}$ by

$$H(u) := \min\big(q_1 + q_2 d(u, v),\ q_3\big). \tag{A.1}$$

The collection $\widehat{\Lambda}_0$ of all such "hat functions" H is clearly countable. Let

$$\Lambda_0 := \big\{q \min(H_1, \ldots, H_m)\ :\ q \in \mathbb{Q},\ H_1, \ldots, H_m \in \widehat{\Lambda}_0 \text{ for any } m \in \mathbb{N}\big\}. \tag{A.2}$$

Then Λ_0 is also countable, and it can be shown that weak convergence of a sequence $(\mu_n)_n$ is equivalent to

$$\langle \mu_n, F \rangle \to \langle \mu, F \rangle \qquad \forall\, F \in \Lambda_0 \tag{A.3}$$

(see the aforementioned reference).

We prove first Lemma 2.1.

Proof. [Proof of Lemma 2.1] If $g \in \mathcal{H}^{k,p}(D; U)$ and $u \in L^p(D; U)$ then

$$
\begin{aligned}
|L_g(u)| &\leqslant \int_{D^k} |g(x, u(x_1), \ldots, u(x_k))|\, dx \\
&\leqslant \sum_{\alpha \in \{0,1\}^k} \int_{D^k} \varphi_{|\hat\alpha|}(x_{\hat\alpha}) |u(x_1)|^{\alpha_1 p} \cdots |u(x_k)|^{\alpha_k p}\, dx \\
&= \sum_{\alpha \in \{0,1\}^k} \|\varphi_{|\hat\alpha|}\|_{L^1(D^{|\hat\alpha|})} \|u\|_{L^p(D)}^{|\alpha|p} \\
&= \sum_{i=0}^{k} \binom{k}{i} \|\varphi_i\|_{L^1(D^i)} \|u\|_{L^p(D)}^{p(k-i)} < \infty.
\end{aligned}
$$

If $g \in \mathcal{H}_1^{k,p}$ and $u, v \in L^p(D; U)$ then (recall $\hat{u}^j(x) := (u(x_1), \ldots, u(x_{j-1}), u(x_{j+1}), \ldots, u(x_k))$)

$$
\begin{aligned}
|L_g(u) - L_g(v)| &\leqslant C \sum_{i=1}^{k} \int_{D^k} |u(x_i) - v(x_i)| \max\big(|u(x_i)|, |v(x_i)|\big)^{p-1} \prod_{j \neq i} h(\hat{x}^j, \hat{u}^j(x))\, dx \\
&\leqslant Ck \|u - v\|_{L^p} \big\| \max(|u|, |v|) \big\|_{L^p}^{(p-1)/p} \big(\|\varphi_1\|_{L^1} + \|u\|_{L^p}^p\big)^{k-1},
\end{aligned}
$$

whence L_g is continuous (and in fact locally Lipschitz). $\qquad\square$

Lemma Appendix A.1. *Let $F \in \left(\mathcal{C}^p\right)^k$ for some $k \in \mathbb{N}$, i.e. $F(u) = (F_1(u), \ldots, F_k(u))$ for $F_1, \ldots, F_k$ lying either in $\mathcal{C}^p$ or in $\mathcal{C}_1^p$. Let $P : \mathbb{R}^k \to \mathbb{R}$ be a polynomial in k variables. Then $P \circ F$ lies in $\mathcal{C}^p$ or in $\mathcal{C}_1^p$, respectively.*

Proof. $\mathcal{C}^p$ is clearly closed under scalar multiplication, as $\alpha L_g = L_{\alpha g}$ for any $\alpha \in \mathbb{R}$. It is therefore enough to show that $\mathcal{C}^p$ ($\mathcal{C}_1^p$, resp.) is closed under addition and multiplication. Let $g_i : D^{k_i} \times U^{k_i} \to \mathbb{R}$ for $i = 1, 2$ be Carathéodory functions satisfying (2.3), (2.4). Then

$$L_{g_1}(u) L_{g_2}(u) = \int_{D^{k_1}} g_1(x, u(x_1), \ldots, u(x_{k_1})) \, dx \int_{D^{k_2}} g_2(x, u(x_1), \ldots, u(x_{k_2})) \, dx$$

$$= \int_{D^k} g(x, u(x_1), \ldots, u(x_k)) \, dx = L_g(u),$$

where $k = k_1 + k_2$ and

$$g(x, \xi) = g_1(x_1, \ldots, x_{k_1}, \xi_1, \ldots, \xi_{k_1}) g_2(x_{k_1+1}, \ldots, x_k, \xi_{k_1+1}, \ldots, \xi_k).$$

The function g is readily seen to satisfy (2.3), and using (2.3) it can be seen that g also satisfies (2.4) whenever g_1, g_2 do so. This shows that $\mathcal{C}^p$ ($\mathcal{C}_1^p$, resp.) is closed under multiplication. To show that $L_{g_1} + L_{g_2} \in \mathcal{C}^p$ (or $\mathcal{C}_1^p$, respectively), assume that, say, $k_1 \leqslant k_2$. If $k_1 = k_2$ then clearly $L_{g_1} + L_{g_2} \in \mathcal{C}^p$ (or $\mathcal{C}_1^p$, respectively), so assume that $k_1 < k_2$. Let $k_0 = k_2 - k_1$ and let $\varphi \in L^1(D^{k_0})$ satisfy $\varphi \geqslant 0$ and $\int_{D^{k_0}} \varphi(x) \, dx = 1$. Define $g : D^{k_2} \times U^{k_2} \to \mathbb{R}$ by

$$g(x, \xi) = g_1(x_1, \ldots, x_{k_1}, \xi_1, \ldots, \xi_{k_1}) \varphi(x_{k_1+1}, \ldots, x_{k_2}) + g_2(x, \xi).$$

It is now straightforward to verify that $L_{g_1} + L_{g_2} = L_g$ and that g satisfies (2.3), (2.4). $\qquad\square$

Proof. [Proof of Theorem 2.2] We may assume that B is closed. By Lemma 2.1, every $L_g \in \mathcal{C}_1^p$ is continuous and bounded on B. For any $L_g \in \mathcal{C}_1^p$ there is, by the Tietze extension theorem[22], an $F \in C_b(L^p(D; U))$ satisfying $F = L_g$ on B, so we see that if $\mu_n \rightharpoonup \mu$ then $\langle \mu_n, L_g \rangle = \langle \mu_n, F \rangle \to \langle \mu, F \rangle = \langle \mu, L_g \rangle$.

Assume conversely that $\langle \mu_n, F \rangle \to \langle \mu, F \rangle$ for all $F \in \mathcal{C}_1^p$. We claim that the "hat functions" $H : L^p(D; U) \to \mathbb{R}$ defined in (A.1) can be approximated uniformly on B by functions in $\mathcal{C}_1^p$. Let $v \in L^p(D; U)$ and $q_1, q_2, q_3 \in \mathbb{Q}$ be as in (A.1), and let $R := \sup_{u \in B} \|u - v\|_{L^p}^p < \infty$. Let $\psi(s) = \min\left(q_1 + q_2|s|^{1/p}, \, q_3\right)$, which is a continuous, bounded function on $\mathbb{R}$. Let $\varepsilon > 0$ and let $P : \mathbb{R} \to \mathbb{R}$ be a polynomial such that $\|\psi - P\|_{C([-R,R])} < \varepsilon$. The function $u \mapsto \|u - v\|_{L^p}^p$ clearly lies in $\mathcal{C}_1^p$, so Lemma Appendix A.1 implies that also $u \mapsto P\left(\|u - v\|_{L^p}^p\right)$ can be written as a function $L_g \in \mathcal{C}_1^p$. For any $u \in B$ we then have

$$|H(u) - L_g(u)| = \left|\psi\left(\|u - v\|_{L^p}^p\right) - P\left(\|u - v\|_{L^p}^p\right)\right| < \varepsilon.$$

This proves the claim.

Next, let $F \in \Lambda_0$, where Λ_0 is given by (A.2). Let $\varepsilon > 0$, let $g_1, \ldots, g_m$ be such that $\|H_i - L_{g_i}\|_{C_b(B)} < \varepsilon$ for $i = 1, \ldots, m$, and let $P : \mathbb{R}^m \to \mathbb{R}$ be a polynomial such that

$$\sup_{r \in [-\varepsilon, 1+\varepsilon]^m} \big| \min(r_1, \ldots, r_m) - P(r) \big| < \varepsilon.$$

By Lemma Appendix A.1 we can write $P \circ \big(L_{g_1}, \ldots, L_{g_m} \big) = L_g$ for some $L_g \in \mathcal{C}_1^p$. We conclude that

$$
\begin{aligned}
\sup_{u \in B} \big| F(u) - L_g(u) \big| &= \sup_{u \in B} \big| \min(H_1(u), \ldots, H_m(u)) - P(L_{g_1}(u), \ldots, L_{g_m}(u)) \big| \\
&\leqslant \sup_{u \in B} \big| \min(H_1(u), \ldots, H_m(u)) - \min(L_{g_1}(u), \ldots, L_{g_m}(u)) \big| \\
&\quad + \sup_{u \in B} \big| \min(L_{g_1}(u), \ldots, L_{g_m}(u)) - P(L_{g_1}(u), \ldots, L_{g_m}(u)) \big| \\
&\leqslant 2\varepsilon,
\end{aligned}
$$

by the 1-Lipschitz continuity of the min-function and the approximation properties of $g_1, \ldots, g_m$ and P. We can conclude that (A.3) holds, and hence $\mu_n \rightharpoonup \mu$ weakly. $\square$

Appendix B. The compactness theorem

Throughout this appendix we will use the cutoff functions

$$
\theta(s) := \begin{cases} 1 & |s| \leqslant 1 \\ 2 - |s| & 1 < |s| < 2 \\ 0 & |s| \geqslant 2, \end{cases} \qquad \theta_R(v) := \theta(|v|/R), \qquad \zeta_R(v) := v\theta_R(v)
$$

(for some $R \geqslant 1$), defined for $s \in \mathbb{R}$ and $v \in U$, where $U \subset \mathbb{R}^N$ as before. Note that $\zeta_R(v) = v$ for $|v| < R$ and $\zeta_R(v) = 0$ for $|v| > 2R$, and that $\|\zeta_R\|_{\mathrm{Lip}} \leqslant 2$.

For a function $u \in L^p(D)$ we define the modulus of continuity

$$\omega_r^p(u) := \int_D \fint_{B_r(0)} |u(x + z) - u(x)|^p \, dz dx.$$

Lemma Appendix B.1.   *If $u \in L^q(D, U)$ for some $q \in [1, \infty)$ then*

$$\big\| \zeta_R \circ u - u \big\|_{L^q(D)} \leqslant 3\omega_r^q(u)^{1/q}$$

whenever $R \geqslant \frac{1}{|B_r|^{1/q}} \|u\|_{L^q(D)}$.

Proof.   Denote $u_r(x) := \fint_{B_r(0)} u(x + z) \, dz$. Then $|u_r(x)| \leqslant \frac{1}{|B_r|^{1/q}} \|u\|_{L^q(D)}$ for every $x \in D$, so that if $R \geqslant \frac{1}{|B_r|^{1/q}} \|u\|_{L^q(D)}$ then $u_r = \zeta_R \circ u_r$. It follows that

$$
\begin{aligned}
\big\| \zeta_R \circ u - u \big\|_{L^q(D)} &\leqslant \big\| \zeta_R \circ u - \zeta_R \circ u_r \big\|_{L^q(D)} + \|u_r - u\|_{L^q(D)} \\
&\leqslant \big(\|\zeta_R\|_{\mathrm{Lip}} + 1 \big) \|u_r - u\|_{L^q(D)} \\
&\leqslant 3\omega_r^q(u)^{1/q},
\end{aligned}
$$

where we have used the fact that $\|u_r - u\|_{L^q} \leqslant \omega_r^q(u)^{1/q}$. $\hspace{2cm}$ □

Proof. [Proof of Theorem 2.5] For every $k \in \mathbb{N}$, the sequence $(\nu_n^k)_{n \in \mathbb{N}} \subset \mathcal{H}_0^{k*}(D; U)$ is bounded with norm $\|\nu_n^k\|_{\mathcal{H}^{k*}} \equiv 1 < \infty$, and so has a weak*-convergent subsequence. Thus, we can extract a diagonal subsequence $(n_j)_{j \in \mathbb{N}}$ such that $\nu_{n_j}^k \overset{*}{\rightharpoonup} \nu^k \in \mathcal{H}_0^{k*}(D; U)$ for every $k \in \mathbb{N}$. For the sake of notational simplicity we denote $n = n_j$ for the remainder of this proof.

We start by showing iv. Let κ and φ be nonnegative functions, as prescribed. We may assume that $\liminf_{n \to \infty} \langle \nu_n^k, g \rangle_{\mathcal{H}^k} < \infty$, for otherwise there is nothing to prove. For $R > 0$, let $\kappa_R \in C_c(U^k)$ and $\varphi_R \in L^1(D^k)$ be the functions $\kappa_R(\xi) = \kappa(\xi)\theta_R(\xi)$ and $\varphi_R(x) = \varphi(x)\theta_R(x)$. Then $\kappa_R \to \kappa$ and $\varphi_R \to \varphi$ pointwise almost everywhere as $R \to \infty$. If $g_R(x, \xi) := \varphi_R(x)\kappa_R(\xi)$ then clearly $g_R \to g$ as $R \to \infty$ almost everywhere, and $g_R \in \mathcal{H}_0^k(D; U)$ for every R. We then find that

$$
\begin{aligned}
\langle \nu^k, g \rangle_{\mathcal{H}^k} &= \lim_{R \to \infty} \langle \nu^k, g_R \rangle_{\mathcal{H}^k} && \text{(Dominated Convergence Theorem)} \\
&= \lim_{R \to \infty} \liminf_{n \to \infty} \langle \nu_n^k, g_R \rangle_{\mathcal{H}^k} && \text{(since } \nu_n^k \overset{*}{\rightharpoonup} \nu^k\text{)} \\
&\leqslant \lim_{R \to \infty} \liminf_{n \to \infty} \langle \nu_n^k, g \rangle_{\mathcal{H}^k} && \text{(since } g_R \leqslant g \text{ for all } R\text{)} \\
&= \liminf_{n \to \infty} \langle \nu_n^k, g \rangle_{\mathcal{H}^k},
\end{aligned}
$$

which is (2.13).

Since $\nu_{n;x}^k \geqslant 0$ for all k, n and a.e. $x \in D^k$, we have $\nu_x^k \geqslant 0$ for all k and a.e. $x \in D^k$. To see that $\nu_x^k(U^k) = 1$ for a.e. $x \in D^k$, let $A_1, \ldots, A_k \subset D$ be bounded Borel sets of positive Lebesgue measure and define $A = A_1 \times \cdots \times A_k$. Then since $\nu_{n;x}^k$ are probability measures and converge weak* as we have just shown,

$$
1 \geqslant \fint_A \langle \nu_{n;x}^k, \theta_R \rangle \, dx \to \fint_A \langle \nu_x^k, \theta_R \rangle \, dx \qquad \text{as } n \to \infty.
$$

Conversely,

$$
\begin{aligned}
\fint_A 1 - \langle \nu_x^k, \theta_R \rangle \, dx &= \lim_{n \to \infty} \fint_A 1 - \langle \nu_{n;x}^k, \theta_R \rangle \, dx = \lim_{n \to \infty} \fint_A \langle \nu_{n;x}^k, 1 - \theta_R \rangle \, dx \\
&\leqslant \limsup_{n \to \infty} \frac{1}{R^p} \fint_A \langle \nu_{n;x}^k, |\xi|^p \rangle \, dx \\
&\leqslant \limsup_{n \to \infty} \frac{1}{R^p} \sum_{i=1}^{k} \fint_A \langle \nu_{n;x}^k, |\xi_i|^p \rangle \, dx \\
&\leqslant \frac{1}{R^p} \sum_{i=1}^{k} \frac{c^p}{|A_i|},
\end{aligned}
$$

where we have first used Chebyshev's inequality and then the uniform L^p bound (2.11). Passing to $R \to \infty$ in the two estimates above shows that

$$
\fint_A \nu_x^k(U^k) \, dx = 1. \tag{B.1}
$$

Since $A_1, \ldots, A_k$ were arbitrary, it follows that $\nu_x^k(U^k) = 1$ for a.e. $x \in D^k$.

The limit $\boldsymbol{\nu}$ clearly satisfies the symmetry condition. For consistency, we need to show that for any $f \in C_0(U^{k-1})$ and $\varphi \in C_c(D^k)$, we have

$$\int_{D^k} \varphi(x)\langle \nu_x^k, f(\xi_1, \ldots, \xi_{k-1})\rangle \, dx = \int_{D^k} \varphi(x)\langle \nu_{x_1,\ldots,x_{k-1}}^{k-1}, f(\xi_1, \ldots, \xi_{k-1})\rangle \, dx \tag{B.2}$$

Define $\bar{f}_R \in C_0(U^k)$ by $\bar{f}_R(\xi_1, \ldots, \xi_k) = f(\xi_1, \ldots, \xi_{k-1})\theta_R(\xi_k)$. We estimate the difference between the left- and right-hand sides of (B.2) by $E_1 + E_2 + E_3 + E_4$, where

$$E_1 = \left| \int_{D^k} \varphi(x)\langle \nu_x^k, f(\xi_1, \ldots, \xi_{k-1}) - \bar{f}_R(\xi)\rangle \, dx \right|$$

$$E_2 = \left| \int_{D^k} \varphi(x)\left(\langle \nu_x^k, \bar{f}_R \rangle - \langle \nu_{n;x}^k, \bar{f}_R \rangle \right) \, dx \right|$$

$$E_3 = \left| \int_{D^k} \varphi(x)\left(\langle \nu_{n;x}^k, \bar{f}_R \rangle - \langle \nu_{n;\hat{x}_k}^{k-1}, f \rangle \right) \, dx \right|$$

$$E_4 = \left| \int_{D^k} \varphi(x)\left(\langle \nu_{n;\hat{x}_k}^{k-1}, f \rangle - \langle \nu_{\hat{x}_k}^{k-1}, f \rangle \right) \, dx \right|$$

where $\hat{x}_k = (x_1, \ldots, x_{k-1})$. The fact that $E_1 \to 0$ as $R \to \infty$ follows from the Dominated Convergence Theorem. For every $R > 0$ we have $E_2, E_4 \to 0$ as $n \to \infty$ by the weak* convergence of ν_n^k and ν_n^{k-1}, respectively. For E_3 we can write

$$E_3 = \left| \int_{D^k} \varphi(x)\langle \nu_{n;x}^k, f(\xi_1, \ldots, \xi_{k-1})(\theta_R(\xi_k) - 1)\rangle \, dx \right|$$

$$\leqslant \|f\|_{C_0(U^{k-1})}\|\varphi\|_{C_0(D^k)} \int_{\mathrm{supp}(\varphi)} \langle \nu_{n;x_k}^1, 1 - \theta_R(\xi)\rangle \, dx$$

$$\leqslant \|f\|_{C_0(U^{k-1})}\|\varphi\|_{C_0(D^k)} \frac{C}{R^p} \to 0$$

as $R \to \infty$ (as in the proof of (B.1)). Thus, for a given $\varepsilon > 0$ we can first select R such that $E_1, E_3 < \varepsilon$, and then select n such that $E_2, E_4 < \varepsilon$. This proves (B.2).

The L^p bound ii follows from (2.13) with $k = 1$, $\varphi \equiv 1$ and $\kappa(\xi) = |\xi|^p$, since $\langle \nu^1, |\xi|^p\rangle_{\mathcal{H}^1} \leqslant \liminf_n \langle \nu_n^1, |\xi|^p\rangle_{\mathcal{H}^1} \leqslant c^p$, by (2.11). To prove the diagonal continuity property iii of $\boldsymbol{\nu}$ we use (2.13) with $k = 2$, $\varphi(x, y) = \frac{\mathbf{1}_{B_r(x)}(y)}{|B_r(x)|}$ and $\kappa(\xi_1, \xi_2) = |\xi_1 - \xi_2|^p$, and note that $\omega_r^p(\nu^2) = \langle \nu^2, g\rangle_{\mathcal{H}^2}$. The diagonal continuity of ν^2 then follows from $\omega_r^p(\nu^2) \leqslant \liminf_{n\to\infty} \omega_r^p(\nu_n^2) \to 0$ as $r \to 0$, by (2.12).

We prove v for $k = 1$ and give a sketch of the general case. Let $g \in \mathcal{H}_1^{1,p}(D; U)$, that is, g is a Carathéodory function such that $|g(x, \xi)| \leqslant \varphi_1(x) + \varphi_0|\xi|^p$ for some $\varphi_0 > 0$ and nonnegative $\varphi_1 \in L^1(D)$, as well as $|g(x, \xi) - g(y, \zeta)| \leqslant C\max(|\xi|, |\zeta|)^{p-1}|\xi - \zeta|$ for some $C > 0$. Define

$$u_n(x) := \langle \nu_{n;x}^1, g(x, \xi)\rangle, \qquad u_n^R(x) := \langle \nu_{n;x}^1, \theta_R(|\xi|^p)g(x, \xi)\rangle,$$

$$u(x) := \langle \nu_x^1, g(x, \xi)\rangle, \qquad u^R(x) := \langle \nu_x^1, \theta_R(|\xi|^p)g(x, \xi)\rangle.$$

The sequence $\{u_n\}_{n\in\mathbb{N}}$ satisfies

$$\int_D \fint_{B_r(0)} |u_n(x+z) - u_n(x)|\, dzdx$$

$$\leqslant \int_D \fint_{B_r(0)} \left|\langle \nu_{n,x,x+z}^2, g(x+z,\xi_2) - g(x,\xi_1)\rangle\right| dzdx$$

$$\leqslant C \int_D \fint_{B_r(0)} \langle \nu_{n,x,x+z}^2, \max(|\xi_1|,|\xi_2|)^{p-1}|\xi_2 - \xi_1|\rangle\, dzdx$$

$$\leqslant C \left(\int_D \fint_{B_r(0)} \langle \nu_{n,x,x+z}^2, \max(|\xi_1|,|\xi_2|)^p\rangle\, dzdx\right)^{(p-1)/p} \tag{B.3}$$

$$\left(\int_D \fint_{B_r(0)} \langle \nu_{n,x,x+z}^2, |\xi_2 - \xi_1|^p\rangle\, dzdx\right)^{1/p}$$

$$\leqslant 2Cc^{p-1}\omega_r^p\left(\nu_n^2\right)^{1/p},$$

so together with the compactness of D and the uniform bound $\|u_n\|_{L^1(D)} \leqslant \|\varphi_1\|_{L^1(D)} + \varphi_0 c^p$ we can apply Kolmogorov's Compactness Theorem [29], [42] to conclude that $\{u_n\}_{n\in\mathbb{N}}$ is precompact in $L^1(D)$. Hence, there exists a subsequence $\{u_{n_l}\}_{l\in\mathbb{N}}$ and some $\bar{u} \in L^1(D)$ such that $u_{n_l} \to \bar{u}$ as $l \to \infty$ in $L^1(D)$. From the weak* convergence of ν_n^1 we know that $u_n^R \rightharpoonup u^R$ as $n \to \infty$ weakly in $L^1(D)$ for every $R > 0$. Lebesgue's dominated convergence theorem implies that $u^R \to u$ as $R \to \infty$ in $L^1(D)$.

We claim that $u_n^R \to u_n$ as $R \to \infty$ in $L^1(D)$, *uniformly* in n. Indeed, choosing $R > 0$ such that $R \geqslant \frac{1}{|B_r|^{1/p}}M$ (where M is the constant in (2.14)), we get

$$\int_D |u_n(x) - u_n^R(x)|\, dx = \int_D \left|\langle \nu_{n;x}^1, (1-\theta_R(|\xi|^p))g(x,\xi)\rangle\right| dx$$

$$\leqslant \int_D \langle \nu_{n;x}^1, (1-\theta_R(|\xi|^p))\varphi_1(x)\rangle\, dx + \varphi_0 \int_D \langle \nu_{n;x}^1, (1-\theta_R(|\xi|^p))|\xi|^p\rangle\, dx$$

$$= \underbrace{\int_{L^p} \int_D (1-\theta_R(|u(x)|^p))\varphi_1(x)\, dxd\mu_n(u)}_{=:F_1}$$

$$+ \varphi_0 \underbrace{\int_{L^p} \int_D (1-\theta_R(|u(x)|^p))|u(x)|^p\, dxd\mu_n(u)}_{=:F_2}.$$

The first term can be bounded by

$$F_1 \leqslant \int_{L^p} \int_{\{x\in D\,:\,|u(x)|^p>R\}} \varphi_1(x)\, dxd\mu_n(u)$$

$$\leqslant \int_{L^p} \sup\left\{\int_{D'} \varphi_1(x)\, dx\,:\, D' \subset D,\, |D'| \leqslant \|u\|_{L^p}^p/R\right\} d\mu_n(u)$$

$$\leqslant \sup\left\{\int_{D'} \varphi_1(x)\, dx\,:\, D' \subset D\,:\, |D'| \leqslant c^p/R\right\}$$

where we in the second inequality used Chebyshev's inequality and in the third inequality the uniform L^p bound (2.11). The above converges to 0 as $R \to \infty$, uniformly in n. For the second term we have

$$
\begin{aligned}
F_2 &= \int_{L^p} \int_D \big| |u(x)|^p - \zeta_R(|u(x)|^p) \big| \, dx d\mu_n(u) = \int_{L^p} \big\| |u|^p - \zeta_R \circ |u|^p \big\|_{L^1(D)} \, d\mu_n(u) \\
&\leqslant 3 \int_{L^p} \omega_r^1(|u|^p) \, d\mu_n(u) \leqslant 6pc^{p-1} \int_{L^p} \omega_r^p(u)^{1/p} \, d\mu_n(u) \\
&\leqslant 6pc^{p-1} \omega_r^p(\nu_n^2)^{1/p}
\end{aligned}
$$

where the first inequality follows from Lemma Appendix B.1 with $q = 1$, the second inequality follows an estimate similar to (B.3), and the third inequality is Hölder's inequality. The final term above vanishes as $R \to \infty$ uniformly in n, by the uniform diagonal continuity assumption (2.12). It follows that for any $\psi \in L^\infty(D)$,

$$
\begin{aligned}
\Big| \int_D \psi(u - \bar{u}) \, dx \Big| &\leqslant \|\psi\|_{L^\infty} \Big(\|u - u^R\|_{L^1} + \|u_{n_l}^R - u_{n_l}\|_{L^1} + \|u_{n_l} - \bar{u}\|_{L^1} \Big) \\
&\quad + \Big| \int_D \psi\big(u^R - u_{n_l}^R\big) \, dx \Big|,
\end{aligned}
$$

all of which vanish as $R \to \infty$ and $l \to \infty$. We conclude that $\bar{u} = u$, whence $u_{n_l} \to u$ as $l \to \infty$. By the uniqueness of the limit u, we get convergence of the whole sequence: $u_n \to u$ in $L^1(D)$ as $n \to \infty$.

For general $k \in \mathbb{N}$ we prove only that $u_n(x) := \big\langle \nu_{n;x}^k, g(x, \xi) \big\rangle$ satisfies a bound of the form (B.3), and leave the rest to the reader. We write first

$$
\begin{aligned}
&\int_{D^k} \fint_{B_r(0)^k} |u_n(x + z) - u_n(x)| \, dz dx \\
&\leqslant \int_{D^k} \fint_{B_r(0)^k} \big\langle \nu_{n;x,x+z}^{2k}, \big| g\big(x + z, \xi_{k+1}, \dots, \xi_{2k}\big) - g\big(x, \xi_1, \dots, \xi_k\big) \big| \big\rangle \, dz dx \\
&\leqslant C \sum_{l=1}^k \int_{D^k} \fint_{B_r(0)^k} \big\langle \nu_{n;x,x+z}^{2k}, |\xi_{l+k} - \xi_l| \max(|\xi_l|, |\xi_{l+k}|)^{p-1} h(\hat{x}^l, \hat{\xi}^l) \big\rangle \, dz dx
\end{aligned}
$$

(cf. Definition 2.1 with $\hat{\xi}^l = (\xi_1, \dots, \xi_{l-1}, \xi_{l+1}, \dots, \xi_k)$). Consider, say, the last sum-

mand above:

$$\int_{D^k} \fint_{B_r(0)^k} \left\langle \nu_{n;x,x+z}^{2k}, |\xi_{2k} - \xi_k| \max(|\xi_k|, |\xi_{2k}|)^{p-1} h(\hat{x}^k, \hat{\xi}^k) \right\rangle dzdx$$

$$= \int_{D^k} \fint_{B_r(0)} \left\langle \nu_{n;x,x_k+z_k}^{k+1}, |\xi_{k+1} - \xi_k| \max(|\xi_k|, |\xi_{k+1}|)^{p-1} h(\hat{x}^k, \hat{\xi}^k) \right\rangle dz_k dx$$

$$= \int_{L^p(D)} \int_{D^k} \fint_{B_r(0)} \left|u(x_k + z_k) - u(x_k)\right| \max(|u(x_k)|, |u(x_k + z_k)|)^{p-1}$$
$$h(x_1, \ldots, x_{k-1}, u(x_1), \ldots, u(x_{k-1})) \, dz_k dx d\mu_n(u)$$

$$= \int_{L^p(D)} \left(\int_D \fint_{B_r(0)} |u(x_k + z_k) - u(x_k)| \max(|u(x_k)|, |u(x_k + z_k)|)^{p-1} \, dz_k dx_k \right)$$
$$\left(\int_{D^{k-1}} h(x_1, \ldots, x_{k-1}, u(x_1), \ldots, u(x_{k-1})) \, d\hat{x}^k \right) d\mu_n(u)$$

$$\leqslant 2 \int_{L^p(D)} \left(\int_D \fint_{B_r(0)} |u(x_k + z_k) - u(x_k)|^p \, dz_k dx_k \right)^{1/p} \left(\int_D |u(x_k)|^p \, dx_k \right)^{(p-1)/p}$$
$$C\left(1 + \|u\|_{L^p}^{p(k-1)}\right) d\mu_n(u)$$

$$\leqslant 2C\omega_r^p \left(\nu_n^2\right)^{1/p} M^{p-1}(1 + M^{p(k-1)})$$

where the second-last inequality follows from Hölder's inequality and the boundedness of L_h (cf. the proof of Lemma 2.1), and the last inequality follows from (2.14). By the uniform diagonal continuity of $\boldsymbol{\nu}_n$, the above vanishes uniformly as $r \to 0$. The rest of the proof follows as in the proof for ν^1. $\qquad\square$

Appendix C. Time-dependent correlation measures

Proof. [Proof of Lemma 2.3] By assumption, $(t, x) \mapsto \left\langle \nu_{t,x}^k, g \right\rangle$ is measurable for all $g \in C_0(U^k)$, so by separability of $\mathcal{H}_0^k(D; U)$, the map $t \mapsto \int_{D^k} \left\langle \nu_{t,x}^k, g(x) \right\rangle dx$ is also measurable for any $g \in \mathcal{H}_0^k(D; U)$. Let $E_0 \subset \mathcal{H}_0^k(D; U)$ be a countable, dense subset of the unit sphere in $\mathcal{H}_0^k(D; U)$, and let

$$\mathcal{T} := \bigcap_{g \in E_0} \left\{ \text{Lebesgue points for } t \mapsto \int_{D^k} \left\langle \nu_{t,x}^k, g(x) \right\rangle dx \right\},$$

a set whose complement $[0, T) \setminus \mathcal{T}$ has zero Lebesgue measure. The set $E := \operatorname{span} E_0$ is dense in $\mathcal{H}_0^k(D; U)$, and every $t \in \mathcal{T}$ is still a Lebesgue point for $t \mapsto \int_{D^k} \left\langle \nu_{t,x}^k, g(x) \right\rangle dx$ whenever $g \in E$. For every $t \in \mathcal{T}$ we define the functional $\rho(t) : E \to \mathbb{R}$ by

$$\rho(t)(g) := \int_{D^k} \left\langle \nu_{t,x}^k, g(x) \right\rangle dx.$$

Then $\rho(t)$ is linear: For any $g, h \in E$ and $\alpha \in \mathbb{R}$ we have

$$
\begin{aligned}
\rho(t)(\alpha g + h) &= \lim_{h \to 0^+} \fint_{t-h}^{t+h} \int_{D^k} \langle \nu_{s,x}^k, \alpha g(x) + h(x) \rangle \, dx ds \\
&= \lim_{h \to 0^+} \left(\alpha \fint_{t-h}^{t+h} \int_{D^k} \langle \nu_{s,x}^k, g(x) \rangle \, dx ds + \fint_{t-h}^{t+h} \int_{D^k} \langle \nu_{s,x}^k, h(x) \rangle \, dx ds \right) \\
&= \alpha \rho(t)(g) + \rho(t)(h).
\end{aligned}
$$

Moreover, $\rho(t)$ is continuous, as

$$
\begin{aligned}
|\rho(t)(g)| &\leqslant \lim_{h \to 0^+} \fint_{t-h}^{t+h} \int_{D^k} |\langle \nu_{t,x}^k, g(x) \rangle| \, dx dt \leqslant \lim_{h \to 0^+} \fint_{t-h}^{t+h} \int_{D^k} \|g(x)\|_{C_0} \, dx dt \\
&= \|g\|_{\mathcal{H}_0^k(D;U)}.
\end{aligned}
$$

It follows that for every $t \in \mathcal{T}$ the functional $\rho(t)$ can be extended uniquely to a continuous linear functional on $\mathcal{H}_0^k(D;U)$. The remaining claims in the lemma are now readily checked. $\qquad \square$

Appendix D. Proof of Theorem 4.2

We write

$$
u_{\mathbf{i}}^{t,\Delta} = (\mathcal{S}_t^\Delta \bar{u})_{\mathbf{i}}
$$

where $\mathcal{S}^\Delta$ is the numerical evolution operator; see Section 4.1.

Proof. Let μ_t^Δ be the approximate statistical solution, defined in (4.3). Let $k \in \mathbb{N}$, and let $\varphi \in C_c^\infty(D^k)$ be a test function. Fix a multiindex $\mathbf{i} = (\mathbf{i}_1, \ldots, \mathbf{i}_k) \in (\mathbb{Z}^d)^k$, By changing the order of integration, we have

$$
\begin{aligned}
\int_{\mathbb{R}_+} \int_{L^p(D,U)} u_{\mathbf{i}_1} \cdots u_{\mathbf{i}_k} \partial_t \varphi(\mathbf{x_i}, t) \, d\mu_t^\Delta(u) dt &= \int_{\mathbb{R}_+} \int_{L^p(D,U)} u_{\mathbf{i}_1}^{t,\Delta} \cdots u_{\mathbf{i}_k}^{t,\Delta} \partial_t \varphi(\mathbf{x_i}, t) \, d\bar{\mu}(\bar{u}) \, dt \\
&= \int_{L^p(D,U)} \int_{\mathbb{R}_+} u_{\mathbf{i}_1}^{t,\Delta} \cdots u_{\mathbf{i}_k}^{t,\Delta} \partial_t \varphi(\mathbf{x_i}, t) \, dt \, d\bar{\mu}(\bar{u}),
\end{aligned}
$$

where

$$
\mathbf{x_i} := \begin{pmatrix} x_{\mathbf{i}_1} & \cdots & x_{\mathbf{i}_k} \end{pmatrix}.
$$

Since $\{u_{\mathbf{j}}^{\cdot,\Delta}\}_{\mathbf{j} \in \mathbb{Z}^d}$ solves (4.1) with initial data $\{\bar{u}_{\mathbf{j}}\}_{\mathbf{j} \in \mathbb{Z}^d}$ for every $\bar{u} \in L^p(D,U)$, we get that

$$
\begin{aligned}
&\int_{L^p(D,U)} \int_{\mathbb{R}_+} u_{\mathbf{i}_1}^{t,\Delta} \cdots u_{\mathbf{i}_k}^{t,\Delta} \partial_t \varphi(\mathbf{x_i}, t) \\
&- \frac{1}{\Delta} \sum_{n=1}^k \sum_{m=1}^d u_{\mathbf{i}_1}^{t,\Delta} \cdots \left(F_{\mathbf{i}_n + 1/2\mathbf{e}_m}^m (\mathcal{S}^\Delta(t)(\bar{u})) - F_{\mathbf{i}_n - 1/2\mathbf{e}_m}^m (\mathcal{S}^\Delta(t)(\bar{u})) \right) \cdots u_{\mathbf{i}_k}^{t,\Delta} \varphi(\mathbf{x_i}, t) \, dt \\
&\hspace{6cm} + \varphi(\mathbf{x_i}, 0) u_{\mathbf{i}_1}^{0,\Delta} \cdots u_{\mathbf{i}_k}^{0,\Delta} \, d\bar{\mu}(\bar{u}) = 0.
\end{aligned}
$$

Multiplying by Δ^{kd} and summing over $(\mathbb{Z}^d)^k$, we get

$$
\sum_{\mathbf{i}\in(\mathbb{Z}^d)^k} \Delta^{kd} \int_{L^p(D,U)} \Bigg(\int_{\mathbb{R}_+} \Bigg(u_{\mathbf{i}_1}^{t,\Delta}\cdots u_{\mathbf{i}_k}^{t,\Delta}\partial_t\varphi(\mathbf{x_i},t)
$$
$$
-\frac{1}{\Delta}\sum_{n=1}^{k}\sum_{m=1}^{d} u_{\mathbf{i}_1}^{t,\Delta}\cdots\Bigg(F_{\mathbf{i}_n+1/2\mathbf{e}_m}^m(\mathcal{S}^\Delta(t)(\bar u))-F_{\mathbf{i}_n-1/2\mathbf{e}_m}^m(\mathcal{S}^\Delta(t)(\bar u))\Bigg)\cdots u_{\mathbf{i}_k}^{t,\Delta}\varphi(\mathbf{x_i},t)\Bigg)\,dt
$$
$$
+\varphi(\mathbf{x_i},0)u_{\mathbf{i}_1}^{0,\Delta}\cdots u_{\mathbf{i}_k}^{0,\Delta}\Bigg)\,d\bar\mu(u)=0.
$$

For $1\leqslant n\leqslant k$ and $1\leqslant m\leqslant d$, summation by parts gives

$$
\sum_{\mathbf{i}\in(\mathbb{Z}^d)^k}\Delta^{dk}\int_{\mathbb{R}_+}\int_{L^p(D,U)} u_{\mathbf{i}_1}\cdots\Bigg(F_{\mathbf{i}_n+1/2\mathbf{e}_m}^m(u))-F_{\mathbf{i}_n-1/2\mathbf{e}_m}^m(u))\Bigg)\cdots u_{\mathbf{i}_k}\varphi(\mathbf{x_i},t)\,dt\,d\mu_t^\Delta(u)
$$
$$
=\sum_{\mathbf{i}\in(\mathbb{Z}^d)^k}\Delta^{dk}\int_{\mathbb{R}_+}\int_{L^p(D,U)} u_{\mathbf{i}_1}\cdots F_{\mathbf{i}_n+1/2\mathbf{e}_m}^m(u)\cdots u_{\mathbf{i}_k}\Bigg(\varphi(\mathbf{x_i},t)-\varphi(\mathbf{x}_{\mathbf{i}+\mathbf{e}_{n,m}},t)\Bigg)\,dt\,d\mu_t^\Delta(u)
$$

We furthermore have that

$$
\sum_{\mathbf{i}\in(\mathbb{Z}^d)^k}\Delta^{dk}\int_{\mathbb{R}_+}\int_{L^p(D,U)} u_{\mathbf{i}_1}\cdots F_{\mathbf{i}_n+1/2\mathbf{e}_m}^m(u)\cdots u_{\mathbf{i}_k}\Bigg(\frac{\varphi(\mathbf{x_i},t)-\varphi(\mathbf{x}_{\mathbf{i}+\mathbf{e}_{n,m}},t)}{\Delta}\Bigg)\,d\mu_t^\Delta(u)\,dt
$$
$$
=\sum_{\mathbf{i}\in(\mathbb{Z}^d)^k}\Delta^{dk}\int_{\mathbb{R}_+}\int_{L^p(D,U)} u_{\mathbf{i}_1}\cdots f^m(u_{\mathbf{i}_n})\cdots u_{\mathbf{i}_k}\Bigg(\frac{\varphi(\mathbf{x_i},t)-\varphi(\mathbf{x}_{\mathbf{i}+\mathbf{e}_{n,m}},t)}{\Delta}\Bigg)\,d\mu_t^\Delta(u)\,dt
$$
$$
-\sum_{\mathbf{i}\in(\mathbb{Z}^d)^k}\Delta^{dk}\int_{\mathbb{R}_+}\int_{L^p(D,U)} u_{\mathbf{i}_1}\cdots\Bigg(f^m(u_{\mathbf{i}_n})-F_{\mathbf{i}_n+1/2\mathbf{e}_m}^m(u)\Bigg)\cdots u_{\mathbf{i}_k}\Bigg(\frac{\varphi(\mathbf{x_i},t)-\varphi(\mathbf{x}_{\mathbf{i}+\mathbf{e}_{n,m}},t)}{\Delta}\Bigg)\,d\mu_t^\Delta(u)\,dt
$$

and by the Lipschitz continuity (4.2), we have

$$
\left|\sum_{\mathbf{i}\in(\mathbb{Z}^d)^k}\Delta^{dk}\int_{\mathbb{R}_+}\int_{L^p(D,U)} u_{\mathbf{i}_1}\cdots\Bigg(f^m(u_{\mathbf{i}_n})-F_{\mathbf{i}_n+1/2\mathbf{e}_m}^m(u)\Bigg)\cdots u_{\mathbf{i}_k}\Bigg(\frac{\varphi(\mathbf{x_i},t)-\varphi(\mathbf{x}_{\mathbf{i}+\mathbf{e}_{n,m}},t)}{\Delta}\Bigg)\,d\mu_t^\Delta(u)\,dt\right|
$$
$$
\leqslant\sum_{\mathbf{i}\in(\mathbb{Z}^d)^k}\Delta^{dk}\int_{\mathbb{R}_+}\int_{L^p(D,U)}\left| u_{\mathbf{i}_1}\cdots\Bigg(f^m(u_{\mathbf{i}_n})-F_{\mathbf{i}_n+1/2\mathbf{e}_m}^m(u)\Bigg)\cdots u_{\mathbf{i}_k}\Bigg(\frac{\varphi(\mathbf{x_i},t)-\varphi(\mathbf{x}_{\mathbf{i}+\mathbf{e}_{n,m}},t)}{\Delta}\Bigg)\right|\,d\mu_t^\Delta(u)\,dt
$$
$$
\leqslant C\sum_{\mathbf{i}\in(\mathbb{Z}^d)^k}\Delta^{dk}\sum_{q=-p+1}^{p}\int_{\mathbb{R}_+}\int_{L^p(D,U)}\left| u_{\mathbf{i}_1}\right|\cdots\left| u_{\mathbf{i}_n+q\mathbf{e}_m}-u_{\mathbf{i}_n}\right|\cdots\left| u_{\mathbf{i}_k}\right|\left|\frac{\varphi(\mathbf{x_i},t)-\varphi(\mathbf{x}_{\mathbf{i}+\mathbf{e}_{n,m}},t)}{\Delta}\right|\,d\mu_t^\Delta(u)\,dt
$$
$$
\leqslant C\sum_{\mathbf{i}\in(\mathbb{Z}^d)^k}\Delta^{dk}\int_{\mathbb{R}_+}\int_{L^p(D,U)}\left| u_{\mathbf{i}_1}\right|\cdots\left| u_{\mathbf{i}_n}-u_{\mathbf{i}_n-\mathbf{e}_m}\right|\cdots\left| u_{\mathbf{i}_k}\right|\left|\frac{\varphi(\mathbf{x_i},t)-\varphi(\mathbf{x}_{\mathbf{i}-\mathbf{e}_{n,m}},t)}{\Delta}\right|\,d\mu_t^\Delta(u)\,dt
$$

By Hölder's inequality we get

$$
\sum_{\mathbf{i}\in(\mathbb{Z}^d)^k} \Delta^{dk} \int_{\mathbb{R}_+} \int_{L^p(D,U)} \left|u_{\mathbf{i}_1}\right| \cdots \left|u_{\mathbf{i}_n} - u_{\mathbf{i}_n-\mathbf{e}_m}\right| \cdots \left|u_{\mathbf{i}_k}\right| \left|\frac{\varphi(\mathbf{x}_{\mathbf{i}},t) - \varphi(\mathbf{x}_{\mathbf{i}-\mathbf{e}_{n,m}},t)}{\Delta}\right| \, d\mu_t^\Delta(u) \, dt
$$

$$
\leqslant C \left(\Delta^{kd} \int_{\mathbb{R}_+} \int_{L^p(D,U)} \sum_{\mathbf{i}\in(\mathbb{Z}^d)^k} \left|u_{\mathbf{i}_n} - u_{\mathbf{i}_n-\mathbf{e}_m}\right|^s \, d\mu_t^\Delta(u) \, dt \right)^{1/s}
$$

$$
\cdot \left(\Delta^{kd} \sum_{\mathbf{i}\in(\mathbb{Z}^d)^k} \int_{\mathbb{R}_+} \int_{L^p(D,U)} \left|u_{\mathbf{i}_1}\right|^{\frac{s}{s-1}} \cdots \left|u_{\mathbf{i}_k}\right|^{\frac{s}{s-1}} \left|\frac{\varphi(\mathbf{x}_{\mathbf{i}},t) - \varphi(\mathbf{x}_{\mathbf{i}-\mathbf{e}_{n,m}},t)}{\Delta}\right|^{\frac{s}{s-1}} \, d\mu_t^\Delta(u) \, dt \right)^{\frac{s-1}{s}}.
$$

By the weak BV assumption (4.18), we get

$$
\lim_{\Delta\to 0} \Delta^{kd} \int_{\mathbb{R}_+} \int_{L^p(D,U)} \sum_{\mathbf{i}\in(\mathbb{Z}^d)^k} \left|u_{\mathbf{i}_n} - u_{\mathbf{i}_n-\mathbf{e}_m}\right|^s \, d\mu_t^\Delta(u) \, dt
$$

$$
= \lim_{\Delta\to 0} \Delta^d \int_{\mathbb{R}_+} \int_{L^p(D,U)} \sum_{\mathbf{i}\in\mathbb{Z}^d} \left|u_{\mathbf{i}} - u_{\mathbf{i}-\mathbf{e}_m}\right|^s \, d\mu_t^\Delta(u) \, dt = 0.
$$

We furthermore note that since $\varphi \in C_c^\infty$, we have

$$
\Delta^{kd} \sum_{\mathbf{i}\in(\mathbb{Z}^d)^k} \int_{\mathbb{R}_+} \int_{L^p(D,U)} \left|u_{\mathbf{i}_1}\right|^{\frac{s}{s-1}} \cdots \left|u_{\mathbf{i}_k}\right|^{\frac{s}{s-1}} \left|\frac{\varphi(\mathbf{x}_{\mathbf{i}},t) - \varphi(\mathbf{x}_{\mathbf{i}-\mathbf{e}_{n,m}},t)}{\Delta}\right|^{\frac{s}{s-1}} \, d\mu_t^\Delta(u) \, dt \leqslant C.
$$

Hence we get

$$
0 = \lim_{\Delta\to 0} \sum_{\mathbf{i}\in(\mathbb{Z}^d)^k} \Delta^{kd} \int_{L^p(D,U)} \left(\int_{\mathbb{R}_+} \left(u_{\mathbf{i}_1}^{t,\Delta} \cdots u_{\mathbf{i}_k}^{t,\Delta} \partial_t \varphi(\mathbf{x}_{\mathbf{i}},t) \right. \right.
$$

$$
- \frac{1}{\Delta} \sum_{n=1}^{k} \sum_{m=1}^{d} u_{\mathbf{i}_1}^{t,\Delta} \cdots \left(F_{\mathbf{i}_n+1/2\mathbf{e}_m}^m(\mathcal{S}^\Delta(t)(\bar{u})) - F_{\mathbf{i}_n-1/2\mathbf{e}_m}^m(\mathcal{S}^\Delta(t)(\bar{u})) \right) \cdots u_{\mathbf{i}_k}^{t,\Delta} \varphi(\mathbf{x}_{\mathbf{i}},t) \right) \, dt
$$

$$
\left. + \varphi(\mathbf{x}_{\mathbf{i}},0) u_{\mathbf{i}_1}^{0,\Delta} \cdots u_{\mathbf{i}_k}^{0,\Delta} \right) \, d\bar{\mu}(\bar{u})
$$

$$
= \int_{\mathbb{R}_+} \int_{L^p(D,U)} \int_{(\mathbb{R}^d)^k} \left(u(x_1) \cdots u(x_k) \partial_t \varphi(\mathbf{x},t) \right.
$$

$$
\left. + \sum_{n=1}^{k} \sum_{m=1}^{d} u(x_1) \cdots f^m(u(x_n)) \cdots u(x_k) \partial_{nm}\varphi(\mathbf{x},t) \right) \, d\mathbf{x} \, d\mu^t(u) \, dt
$$

$$
+ \int_{L^p(D,U)} \int_{(\mathbb{R}^d)^k} \bar{u}(x_1) \cdots \bar{u}(x_k) \varphi(\mathbf{x},0) \, d\mathbf{x} \, d\bar{\mu}(\bar{u}),
$$

which completes the proof. $\qquad\square$

Appendix E. Proof of (5.7)

Proposition Appendix E.1. *Let $D \subset \mathbb{R}^d$ be bounded and let $\mu, \tilde{\mu} \in \mathcal{P}(L^p(D;U))$ both have bounded p-th moment. Then*

$$\left(\int_{D^k} W_p\big(\nu_x^k, \tilde{\nu}_x^k\big)^p \, dx \right)^{1/p} \leqslant C W_p(\mu, \tilde{\mu}) \tag{E.1}$$

for some $C = C(p, k, |D|)$, where $(\nu^k)_{k\in\mathbb{N}}, (\tilde{\nu}^k)_{k\in\mathbb{N}}$ are the correlation measures corresponding to $\mu, \tilde{\mu}$.

Proof. Let $\pi \in \mathcal{P}(L^p(D;U)^2)$ be an optimal transport plan for $W_p(\mu, \tilde{\mu})$, and let $(\sigma^k)_{k\in\mathbb{N}} \in \mathcal{L}^p(D;U^2)$ be its corresponding correlation measure. We claim that the marginals of σ_x^k are ν_x^k and $\tilde{\nu}_x^k$, respectively, for a.e. $x \in D^k$. Indeed, if $g \in \mathcal{H}_0^k(D;U)$ then, denoting $\bar{g}(x, \xi, \zeta) := g(x, \xi)$,

$$\big\langle \sigma^k, \bar{g} \big\rangle = \big\langle \pi, L_{\bar{g}} \big\rangle = \int_{L^p(D;U)^2} L_{\bar{g}}(u, v) \, d\pi(u, v) = \int_{L^p(D;U)^2} L_g(u) \, d\pi(u, v)$$

$$= \int_{L^p(D;U)} L_g(u) \, d\mu(u) = \big\langle \nu^k, g \big\rangle.$$

A similar computation holds for the second marginal. Since g is arbitrary, the claim follows. We can therefore estimate

$$\int_{D^k} W_p\big(\nu_x^k, \tilde{\nu}_x^k\big)^p \, dx \leqslant \int_{D^k} \int_{U^{2k}} |\xi - \zeta|^p \, d\sigma_x^k(\xi, \zeta) \, dx$$

$$= \sum_{l=1}^{k} \int_{D^k} \int_{U^{2k}} |\xi_l - \zeta_l|^p \, d\sigma_x^k(\xi, \zeta) \, dx$$

$$= k \int_{D^k} \int_{U^2} |\xi_1 - \zeta_1|^p \, d\sigma_{x_1}^1(\xi, \zeta) \, dx$$

$$= k|D|^{k-1} \int_D \int_{U^2} |\xi - \zeta|^p \, d\sigma_x^1(\xi, \zeta) \, dx$$

$$= k|D|^{k-1} \int_{L^p(D;U)^2} \int_D |u(x) - v(x)|^p \, dx \, d\pi(u, v) = W_p(\mu, \tilde{\mu})^p. \qquad \square$$

Appendix F. Computing the Wasserstein distance for sums of Diracs

In Section 5, we computed the Wasserstein distance between different numerical solutions. The numerical solutions are always sums of Diracs, which simplifies the computations greatly.

For two k-th correlation marginals ν^k and $\tilde{\nu}^k$, we are interested in the "$L^p W_p$" distance

$$\big\| W_1\big(\nu_{T,\cdot}^k, \tilde{\nu}_{T,\cdot}^k\big) \big\|_{L^1(D^k)}. \tag{F.1}$$

66   *U. S. Fjordholm, K. Lye, S. Mishra and F. Weber*

F.1. *Computing the Wasserstein distance between the first correlation marginals*

To compute the first correlation marginals, we use the function `scipy.stats.wasserstein_distance` in the `scipy` module for Python [30] to compute the Wasserstein distance. The function computes the Wasserstein distance by going through the CDF, consult the `scipy` documentation for more information.

We approximate the spatial integral as a sum over all the volume averages.

F.2. *Computing the Wasserstein distance between the second correlation marginals*

For the second correlation marginals, we use the function `ot.emd` in the `POT` module for Python [20]. This function uses the Hungarian algorithm [33] to compute the Wasserstein distance between sums of Diracs.

The spatial integral in (F.1) is approximated by taking 10 spatial points in each direction, for a total of 10 000 evaluations of the Wasserstein distance. More concretely, we use the following approximation:

$$\left\| W_1\left(\nu^2_{T,\cdot}, \tilde{\nu}^2_{T,\cdot}\right)\right\|_{L^1(D^2)} \approx \sum_{i,j,\tilde{i},\tilde{j}=1}^{10} \frac{1}{10^4} W_1\left(\nu^2_{T,x_{i,j,\tilde{i},\tilde{j}}}, \tilde{\nu}^2_{T,x_{i,j,\tilde{i},\tilde{j}}}\right),$$

where

$$x_{i,j,\tilde{i},\tilde{j}} = \left(\frac{i}{10}, \frac{j}{10}, \frac{\tilde{i}}{10}, \frac{\tilde{j}}{10}\right) \qquad \text{for } i,j,\tilde{i},\tilde{j} = 1,\ldots,10.$$

References

1. C. D. Aliprantis and K. Border. *Infinite Dimensional Analysis*. Springer-Verlag Berlin Heidelberg, 3rd edition, 2006.
2. L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
3. J. M. Ball. A version of the fundamental theorem for Young measures. In *PDEs and continuum models of phase transitions (Nice, 1988)*, volume 344 of *Lecture Notes in Phys.*, pages 207–215. Springer, Berlin, 1989.
4. S. Bianchini and A. Bressan. Vanishing viscosity solutions of nonlinear hyperbolic systems. *Ann. of Math*, 161(1):223–342, 2005.
5. P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, Inc., 2008.
6. Y. Brenier, C. De Lellis, and L. Székelyhidi, Jr. Weak-strong uniqueness for measure-valued solutions. *Comm. Math. Phys.*, 305(2):351–361, 2011.
7. A. Bressan. Hyperbolic systems of conservation laws: The one dimensional Cauchy problem. *Oxford University Press*, 200, 2000.
8. J. Brezina and E. Feireisl. Measure-valued solutions to the complete Euler system. *J. Math. Soc. Japan,*, 70(4):1227–1245, 2018.
9. E. Chiodaroli, C. De Lellis, and O. Kreml. Global ill-posedness of the isentropic system of gas dynamics. *Comm. Pure Appl. Math.*, 68(7):1157–1190, 2015.

10. C. M. Dafermos. *Hyperbolic conservation laws in continuum physics*, volume 325 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, fourth edition, 2016.

11. C. De Lellis and L. Székelyhidi, Jr. The Euler equations as a differential inclusion. *Ann. of Math. (2)*, 170(3):1417–1436, 2009.

12. S. Demoulini, D. M. A. Stuart, and A. E. Tzavaras. Weak-strong uniqueness of dissipative measure-valued solutions for polyconvex elastodynamics. *Arch. Rat. Mech. Anal.*, 205(3):927–961, 2012.

13. R. J. DiPerna. Measure-valued solutions to conservation laws. *Arch. Rational Mech. Anal.*, 88(3):223–270, 1985.

14. U. S. Fjordholm. *High-order accurate entropy stable numerical schemes for hyperbolic conservation laws.* PhD thesis, ETH Zürich, 2013. Dis. no 21025.

15. 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(3):763–827, 2017.

16. U. S. Fjordholm, S. Lanthaler, and S. Mishra. Statistical solutions of hyperbolic conservation laws: foundations. *Arch. Ration. Mech. Anal.*, 226(2):809–849, 2017.

17. U. S. Fjordholm, K. O. Lye, and S. Mishra. Numerical approximation of statistical solutions of scalar conservation laws. *SIAM J. Numer. Anal.*, 56(5):2989–3009, 2018.

18. U. S. Fjordholm, S. Mishra, and E. Tadmor. Arbitrarily high-order accurate entropy stable essentially nonoscillatory schemes for systems of conservation laws. *SIAM J. Numer. Anal.*, 50(2):544–573, 2012.

19. U. S. Fjordholm, S. Mishra, and E. Tadmor. On the computation of measure-valued solutions. *Acta Numerica*, 25:567–679, 2016.

20. R. Flamary and N. Courty. POT Python Optimal Transport library, 2017.

21. C. Foiaş, O. Manley, R. Rosa, and R. Temam. *Navier–Stokes Equations and Turbulence.* Cambridge University Press, 2001.

22. G. B. Folland. *Real analysis.* John Wiley & Sons Inc., 1999.

23. A. Fournier, D. Fussell, and L. Carpenter. Computer rendering of stochastic models. *Commun. ACM*, 25(6):371–384, June 1982.

24. U. Frisch. *Turbulence.* Cambridge University Press, 1995.

25. J. Glimm. Solutions in the large for nonlinear hyperbolic systems of equations. *Comm. Pure Appl. Math.*, 18(4):697–715, 1965.

26. E. Godlewski and P. Raviart. *Hyperbolic systems of conservation laws.* Mathématiques & applications. Ellipses, 1991.

27. S. Gottlieb, C.-W. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Rev.*, 43(1):89–112, 2001.

28. J. Hesthaven. *Numerical methods for conservation laws: From analysis to algorithms.* SIAM, 2018.

29. H. Holden and N. H. Risebro. *Front tracking for hyperbolic conservation laws*, volume 152 of *Applied Mathematical Sciences*. Springer, Heidelberg, second edition, 2015.

30. E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed 1.04.2019].

31. A. Klenke. *Probability Theory. A Comprehensive Course.* Springer London, 2nd edition, 2014.

32. S. N. Kruzkhov. First order quasilinear equations in several independent variables. *USSR Math. Sbornik.*, 10(2):217–243, 1970.

33. H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, mar 1955.

34. R. J. LeVeque. *Numerical methods for conservation laws.* Lectures in Mathematics

68   *U. S. Fjordholm, K. Lye, S. Mishra and F. Weber*

ETH Zürich. Birkhäuser Verlag, Basel, second edition, 1992.

35. 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.

36. K. O. Lye. Alsvinn: The fast finite volume simulator with UQ support, 2018 –. [Online; accessed 1.04.2019].

37. K. O. Lye, S. Mishra, and D. Ray. Deep learning observables in computational fluid dynamics. Preprint, available as arXiv:1903.03040v1.

38. B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437, 1968.

39. 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(280):1979–2018, 2012.

40. 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(8):3365–3388, 2012.

41. S. Schochet. Examples of measure-valued solutions. *Comm. Par. Diff. Eqns.*, 14(5):545–575, 1989.

42. V. N. Sudakov. Criteria of compactness in function spaces. *Uspehi Mat. Nauk (N.S.)*, 12(3(75)):221–224, 1957.

43. A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes.* Springer New York, 1996.

44. C. Villani. *Topics in Optimal Transportation.* Graduate Studies in Mathematics, Vol. 58. American Mathematical Society, 2003.

45. R. F. Voss. Random fractal forgeries. In R. A. Earnshaw, editor, *Fundamental Algorithms for Computer Graphics: NATO Advanced Study Institute directed by J.E. Bresenham, R.A. Earnshaw, M.L.V. Pitteway*, pages 805–835. Springer Berlin Heidelberg, Berlin, Heidelberg, 1991.

46. E. Wiedemann. Weak-strong uniqueness in fluid dynamics. In *Partial differential equations in fluid dynamics, London Math. Soc. Lecture notes 452*, pages 289–326. Cambridge university press., 2018.

47. L. Young. *Lectures on the Calculus of Variations and Optimal Control Theory.* W. B . Saunders, Philadelphia, PA, 1969.