Error control for statistical solutions of hyperbolic systems of conservation laws

Jan Giesselmann[1] · Fabian Meyer[2] · Christian Rohde[2]

Abstract

Statistical solutions have recently been introduced as an alternative solution framework for hyperbolic systems of conservation laws. In this work, we derive a novel a posteriori error estimate in the Wasserstein distance between dissipative statistical solutions and numerical approximations obtained from the Runge-Kutta Discontinuous Galerkin method in one spatial dimension, which rely on so-called regularized empirical measures. The error estimator can be split into deterministic parts which correspond to spatio-temporal approximation errors and a stochastic part which reflects the stochastic error. We provide numerical experiments which examine the scaling properties of the residuals and verify their splitting.

Keywords Hyperbolic conservation laws · Statistical solutions · A posteriori error estimates · Discontinuous Galerkin method

Mathematics Subject Classification Primary: 35L65, 65M15 · Secondary: 65M60, 65M700

1 Introduction

Analysis and numerics of hyperbolic conservation laws have seen a significant shift of paradigms in the last decade. The investigation and approximation of entropy weak solutions was state of the art for a long time. This has changed due to two

✉ Jan Giesselmann
  giesselmann@mathematik.tu-darmstadt.de

1 Department of Mathematics, Technical University of Darmstadt, Dolivostraße 15, 64293 Darmstadt, Germany

2 Institute of Applied Analysis and Numerical Simulation, University of Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany

reasons. Firstly, analytical insights [5, 9] revealed that weak entropic solutions to the Euler equations in several space dimensions are not unique. Secondly, numerical experiments [14, 23] have shown that in simulations of e.g. the Kelvin-Helmholtz instability, numerical solutions do not converge under mesh refinement. In contrast, when families of simulations with slightly varying initial data are considered, averaged quantities like mean, variance and also higher moments are observed to converge under mesh refinement. This has led to several weaker (more statistics inspired) notions of solutions being proposed. We would like to mention dissipative measure valued solutions [13] and statistical solutions [17]. Considering measure valued solutions has a long history in hyperbolic conservation laws and can be traced back to the works of DiPerna [12] considering Young measures. Statistical solutions are time-parametrized probability measures on spaces of integrable functions and have been introduced recently for scalar problems in [15] and for systems in [17]. We would like to mention that in the context of the incompressible Navier-Stokes equations and turbulence, statistical solutions have a long history going back to the seminal work of Foias et al. see [19] and references therein.

The precise definition of statistical solutions is based on an equivalence theorem [17, Theorem 2.8] that relates probability measures on spaces of integrable functions and *correlation measures* on the state space. Correlation measures are measures that determine joint probability distributions of some unknown quantity at any finite collection of spatial points. In this sense, statistical solutions contain more information than e.g. dissipative measure valued solutions and, indeed, for scalar problems uniqueness of entropy dissipative statistical solutions can be proven. This proof is similar to the classical proof of uniqueness of entropy weak solutions for scalar problems. In contrast, for systems in multiple space dimensions the non-uniqueness of entropy weak solutions immediately implies non-uniqueness of dissipative measure valued solutions and statistical solutions. Still, all these concepts satisfy weak-strong uniqueness principles, i.e., as long as a Lipschitz continuous solution exists in any of these classes it is the unique solution in any of these classes. The (technical) basis for obtaining weak-strong uniqueness results is the relative entropy framework of Dafermos and DiPerna [8], which can be extended to dissipative statistical solutions as in [17].

Convergence of numerical schemes for nonlinear systems of hyperbolic conservation laws is widely open (except when the solution is smooth). Exceptions are the one dimensional situation, where convergence of the Glimm scheme is well known [22] and is, indeed, used for constructing the standard Riemann semigroup [4]. For multi-dimensional problems, there is recent progress showing convergence of numerical schemes towards dissipative measure valued solutions [13] with more information on the convergence in case the limit is an entropy weak solution. It seems impossible to say anything about convergence rates in this setting due to the multiplicity of entropy weak solutions.

The work at hand tries to complement the a priori analysis from [17] with a reliable a posteriori error estimator, i.e., we propose a computable upper bound for the numerical approximation error of statistical solutions. This extends results for entropy weak solutions of deterministic and random systems of hyperbolic conservation laws [10, 20, 21, 30] towards statistical solutions. One appealing feature of

our error estimator is that it can be decomposed into different parts corresponding to space-time and stochastic errors. Our analysis relies on the relative entropy framework of Dafermos and DiPerna [8] extended to statistical solutions and, as such, it requires one of the (approximate) solutions that are to be compared to be Lipschitz continuous. Thus, we need to introduce a suitable reconstruction of the numerical solution, see [29] for the general idea, and our estimator is not convergent under mesh refinement once shocks have formed in the exact solution. We would like to mention that there are also other frameworks for providing a posteriori error analysis for hyperbolic systems, namely superconvergence, e.g., [1], dual weighted residuals, e.g., [24] and Bressan's L^1 stability framework, e.g., [27, 28] for one-dimensional systems.

The structure of this work is as follows: Sect. 2 reviews the notion of (dissipative) statistical solutions of hyperbolic conservation laws from [17]. Section 3 is concerned with the numerical approximation of dissipative statistical solutions using empirical measures. Moreover, we recall a reconstruction procedure which allows us to define the so-called regularized empirical measure. In Sect. 4, we present our main a posteriori error estimate between a dissipative statistical solution and its numerical approximation using the regularized empirical measure. We explain in Sect. 5 that our a posterirori error analysis directly extends to statistical solutions to general systems that take values in some compact subset of the state space. Finally, Sect. 6 provides some numerical experiments examining and verifying the convergence and splitting of the error estimators.

2 Preliminaries and notations

We consider the following one-dimensional system of $m \in \mathbb{N}$ nonlinear conservation laws:

$$\begin{cases} \partial_t u(t,x) + \partial_x f(u(t,x)) = 0, & (t,x) \in (0,T) \times D, \\ u(0,x) = \bar{u}(x), & x \in D. \end{cases} \tag{1}$$

Here, $u(t,x) \in \mathcal{U} \subset \mathbb{R}^m$ is the vector of conserved quantities, $\mathcal{U}$ is an open and convex set that is called *state space*, $f \in C^2(\mathcal{U};\mathbb{R}^m)$ is the flux function, $D \subset \mathbb{R}$ is the spatial domain and $T \in \mathbb{R}_+$. We restrict ourselves to the case where $D = (0,1)$ with periodic boundary conditions. The system (1) is called hyperbolic if for any $u \in \mathcal{U}$ the flux Jacobian $Df(u)$ has m real eigenvalues and admits a basis of eigenvectors. We assume that (1) is equipped with an entropy/entropy flux pair (η, q), where the strictly convex function $\eta \in C^2(\mathcal{U};\mathbb{R})$ and $q \in C^2(\mathcal{U};\mathbb{R})$ satisfy $Dq = D\eta Df$.

Most literature on numerical schemes for hyperbolic conservation laws focuses on computing numerical approximations of entropy weak solutions of (1). In contrast, we are interested in computing statistical solutions. We recall the definition of statistical solutions in Sect. 2.1. It is worthwhile to note that statistical solutions were only defined for systems for which $\mathcal{U} = \mathbb{R}^m$ in [17]. We restrict our analysis to this setting in Sects. 2.1 and 4. In Sect. 5, we discuss some issues concerning how to define statistical solutions to general systems.

2.1 Statistical solutions

In this section, we provide a brief overview of the notions needed to define statistical solutions for (1), following the exposition in [17], referring to [17, Sect. 2] for more background, details and proofs.

Let us introduce some notation: For any topological space X let $\mathcal{B}(X)$ denote the Borel σ-algebra on X and $\mathcal{M}(X)$ denotes the set of signed Radon measures on $(X, \mathcal{B}(X))$. In addition, $\mathcal{P}(X)$ denotes the set of probability measures on $(X, \mathcal{B}(X))$, i.e., all non-negative $\mu \in \mathcal{M}(X)$ satisfying $\mu(X) = 1$. We consider $\mathcal{U} = \mathbb{R}^m$ and choose $p \in [1, \infty)$ minimal, such that

$$|f(u)|, |\eta(u)|, |q(u)| \leqslant C(1 + |u|^p), \quad \forall u \in \mathcal{U}$$

holds for some constant $C > 0$. The following classical theorem states the duality between $L^1(D^k; C_0(\mathcal{U}^k))$ and $L^\infty(D^k; \mathcal{M}(\mathcal{U}^k))$.

Theorem 2.1 ([2], p. 211) *For any $k \in \mathbb{N}$ the dual space of $\mathcal{H}_0^k(D, \mathcal{U}) := L^1(D^k; C_0(\mathcal{U}^k))$ is $\mathcal{H}_0^{k*}(D, \mathcal{U}) := L^\infty(D^k; \mathcal{M}(\mathcal{U}^k))$, i.e., the space of bounded, weak*-measurable maps from D^k to $\mathcal{M}(\mathcal{U}^k)$ under the duality pairing*

$$\langle v^k, g \rangle_{\mathcal{H}^k} := \int_{D^k} \langle v_x^k, g(x) \rangle \, dx := \int_{D^k} \int_{\mathcal{U}^k} g(x)(\xi) \, dv_x^k(\xi) \, dx.$$

Definition 2.2 (Correlation measures [15, Def 2.5]) A p-summable correlation measure is a family $\mathbf{v} = (v^1, v^2, \dots)$ with $v^k \in \mathcal{H}_0^{k*}(D; \mathcal{U})$ satisfying for all $k \in \mathbb{N}$ the following properties:

a. v^k is a Young measure from D^k to $\mathcal{U}^k$.
b. *Symmetry*: If σ is a permutation of $\{1, \dots, k\}$, i.e., for $x = (x_1, \dots, x_k)$, we have $\sigma(x) := (x_{\sigma(1)}, \dots, x_{\sigma(k)})$ and if $f \in C_0(\mathcal{U}^k)$ then $\langle v_{\sigma(x)}^k, f(\sigma(\xi)) \rangle = \langle v_x^k, f(\xi) \rangle$ for a.e. $x \in D^k$.
c. *Consistency*: If $f \in C_b(\mathcal{U}^k)$ is of the form $f(\xi_1, \dots, \xi_k) = g(\xi_1, \dots, \xi_{k-1})$ for some $g \in C_0(\mathcal{U}^{k-1})$ then $\langle v_{x_1, \dots, x_k}^k, f \rangle = \langle v_{x_1, \dots, x_{k-1}}^{k-1}, g \rangle$ for a.e. $x = (x_1, \dots, x_k) \in D^k$.
d. *L^p-integrability*: $\int_{D^k} \langle v_x^1, |\xi|^p \rangle \, dx < \infty$.
e. *Diagonal continuity*: $\lim_{r \searrow 0} \int_D \frac{1}{|B_r(x_1)|} \int_{B_r(x_1)} \langle v_{x_1, x_2}^2, |\xi_1 - \xi_2|^p \rangle \, dx_2 dx_1 = 0$.

Let $\mathcal{L}^p(D; \mathcal{U})$ denote the set of all p-summable correlation measures.

Theorem 2.3 (*Main theorem on correlation measures* [15, Thm. 2.7]) *For every correlation measure $\mathbf{v} \in \mathcal{L}^p(D; \mathcal{U})$ there exists a unique probability measure $\mu \in \mathcal{P}(L^p(D; \mathcal{U}))$ whose p-th moment is finite, i.e., $\int_{L^p} \|u\|_{L^p(D; \mathcal{U})}^p d\mu(u) < \infty$ and such that μ is dual to $\mathbf{v}$, i.e.,*

$$\int_{D^k} \langle v_x^k, g(x) \rangle \, dx = \int_{L^p} \int_{D^k} g(x, u(x)) dx d\mu(u) \quad \forall g \in \mathcal{H}_0^k(D, \mathcal{U}), \, \forall k \in \mathbb{N}.$$

Conversely, for every $\mu \in \mathcal{P}(L^p(D; \mathcal{U}))$ with finite p-th moment there is a $\mathbf{v} \in \mathcal{L}^p(D; \mathcal{U})$ that is dual to μ.

To take into account the time-dependence in (1) the authors of [17] suggest to consider time parametrized probability measures. For $T \in (0, \infty]$ consider time parametrized measures $\mu : [0, T) \to \mathcal{P}(L^p(D; \mathcal{U}))$. Note that such a μ does not contain any information about correlation between function values at different times. We denote μ evaluated at time t by μ_t. Let us define $\mathcal{H}_0^{k*}([0, T), D; \mathcal{U}) := L^\infty([0, T) \times D^k; \mathcal{M}(\mathcal{U}^k))$ and notice that it was shown in [17] that it is meaningful to evaluate an element $v^k \in \mathcal{H}_0^{k*}([0, T), D; \mathcal{U})$ at almost every $t \in [0, T)$.

Definition 2.4 A time-dependent correlation measure $\boldsymbol{v}$ is a collection $\boldsymbol{v} = (v^1, v^2, \dots)$ of maps $v^k \in \mathcal{H}_0^{k*}([0, T), D; \mathcal{U})$ such that

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

$$\operatorname*{ess\,sup}_{t \in [0,T)} \int_D \langle v_{t,x}^1, |\xi|^p \rangle \, dx < \infty.$$

c. *Diagonal continuity*

$$\lim_{r \searrow 0} \int_0^{T'} \int_D \frac{1}{|B_r(x_1)|} \int_{B_r(x_1)} \langle v_{t,x_1,x_2}^2, |\xi_1 - \xi_2|^p \rangle \, dx_2 dx_1 \, dt = 0 \quad \forall T' \in (0, T).$$

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

A time-dependent analogue of Theorem 2.3 holds true:

Theorem 2.5 *For every time-dependent correlation measure $\boldsymbol{v} \in \mathcal{L}^p([0, T), D; \mathcal{U})$ there is a unique (up to subsets of $[0, T)$ of Lebesgue measure zero) map $\mu : [0, T) \to \mathcal{P}(L^p(D; \mathcal{U}))$ such that*

a. *the mapping*

$$t \mapsto \int_{L^p} \int_{D^k} g(x, u(x)) \, dx d\mu_t(u)$$

is measurable for all $g \in \mathcal{H}_0^k(D; \mathcal{U})$.
b. *μ is L^p-bounded, i.e.,*

$$\operatorname{ess\,sup}_{t \in [0,T)} \int_{L^p} \|u\|_{L^p(D;\mathcal{U})}^p \, d\mu_t(u) < \infty.$$

c. *Duality*, *i.e.* μ *is dual to* $\mathbf{v}$,

$$\int_{D^k} \langle v_{t,x}^k, g(x) \rangle \, \mathrm{d}x = \int_{L^p} \int_{D^k} g(x, u(x)) \mathrm{d}x d\mu_t(u) \text{ for a.e. } t \in [0, T),$$

$$\forall g \in \mathcal{H}_0^k(D, \mathcal{U}), \ \forall k \in \mathbb{N}.$$

Conversely, for every $\mu : [0, T) \to \mathcal{P}(L^p(D; \mathcal{U}))$ *satisfying (a) and (b) there is a unique correlation measure* $\mathbf{v} \in \mathcal{L}^p([0, T), D; \mathcal{U})$ *such that (c) holds.*

Definition 2.6 (Bounded support) We say some $\bar{\mu} \in \mathcal{P}(L^p(D; \mathcal{U}))$ has bounded support, provided there exists $C > 0$ so that

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

Definition 2.7 (Statistical solution) Let $\bar{\mu} \in \mathcal{P}(L^p(D; \mathcal{U}))$ have bounded support. A statistical solution of (1) with initial data $\bar{\mu}$ is a time-dependent map $\mu : [0, T) \to \mathcal{P}(L^p(D; \mathcal{U}))$ such that each μ_t has bounded support and such that the corresponding correlation measures v_t^k satisfy

$$\partial_t \langle v_{t,x}^k, \xi_1 \otimes \cdots \otimes \xi_k \rangle + \sum_{i=1}^k \partial_{x_i} \langle v_{t,x_1,\dots,x_k}^k, \xi_1 \otimes \cdots \otimes f(\xi_i) \otimes \cdots \otimes \xi_k \rangle = 0 \quad (2)$$

in the sense of distributions, for every $k \in \mathbb{N}$.

Lemma 2.8 *Let* $\bar{\mu} \in \mathcal{P}(L^p(D; \mathcal{U}))$ *have bounded support. Then, every statistical solution* $\mu : [0, T) \to \mathcal{P}(L^p(D; \mathcal{U}))$ *to (1) with initial data* $\bar{\mu}$ *satisfies*

$$\int_0^T \int_{L^p(D; \mathcal{U})} \int_D u(x) \partial_t \cdot \phi(t, x) + f(u(x)) \cdot \partial_x \phi(t, x) \mathrm{d}x d\mu_t(u) dt$$

$$+ \int_{L^p(D; \mathcal{U})} \int_D \bar{u}(x) \cdot \phi(0, x) \mathrm{d}x d\bar{\mu}(\bar{u}) = 0, \quad (3)$$

for any $\phi \in C^\infty([0, T) \times D; \mathbb{R}^m)$, *where* μ_t *denotes* μ *at time t.*

Proof This is a special case of [17, Equation 3.7] for $M = 1$. $\qquad\qquad\square$

In order to show uniqueness (for scalar problems) and weak-strong uniqueness the authors of [15] and [17] require a comparison principle that compares statistical solutions to convex combinations of Dirac measures. Therefore, the *entropy condition* for statistical solutions has two parts. The first imposes stability under convex decompositions and the second is reminiscent of the standard entropy condition for deterministic problems. To state the first condition, we need the following notation: For $\mu \in \mathcal{P}(L^p(D; \mathcal{U}))$, $K \in \mathbb{N}$, $\alpha \in \mathbb{R}^K$ with $\alpha_i \geqslant 0$ and $\sum_{i=1}^K \alpha_i = 1$ we set

$$\Lambda(\alpha, \mu) := \left\{ (\mu_1, \dots, \mu_K) \in (\mathcal{P}(L^p(D; \mathcal{U})))^K \ : \ \sum_{i=1}^K \alpha_i \mu_i = \mu \right\}. \quad (4)$$

The elements of $\Lambda(\alpha, \mu)$ are strongly connected to *transport plans* that play a major role in defining the Wasserstein distance, see Remark 3.5 for details. Now, we are in position to state the selection criterion for statistical solutions.

Definition 2.9 (Dissipative statistical solution) A statistical solution of (1) is called a dissipative statistical solution if

1. for every choice of coefficients $\alpha_i \geqslant 0$, satisfying $\sum_{i=1}^{K} \alpha_i = 1$ and for every $(\bar{\mu}_1, \ldots, \bar{\mu}_K) \in \Lambda(\alpha, \bar{\mu})$, there exists a function $t \mapsto (\mu_{1,t}, \ldots, \mu_{K,t}) \in \Lambda(\alpha, \mu_t)$, such that each $\mu_i : [0, T) \to \mathcal{P}(L^p(D;\mathcal{U}))$ is a statistical solution of (1) with initial measure $\bar{\mu}_i$.
2. it satisfies

$$
\int_0^T \int_{L^p(D;\mathcal{U})} \int_D \eta(u(x))\partial_t\phi(t)\,\mathrm{d}x\mathrm{d}\mu_t(u)\mathrm{d}t + \int_{L^p(D;\mathcal{U})} \int_D \eta(\bar{u}(x))\phi(0)\,\mathrm{d}x\mathrm{d}\bar{\mu}(\bar{u}) \geqslant 0,
$$

for any $\phi \in C_c^\infty([0, T);\mathbb{R}_+)$.

This selection criterion implies weak-(dissipative)-strong uniqueness, i.e., as long as some initial value problem admits a statistical solution supported on finitely many classical solutions, then this is the only dissipative statistical solution (on that time interval), [17, Lemma 3.3]. That result is a major ingredient in the proof of our main result Theorem 4.2 and it is indeed the special case of Theorem 4.2 for $R_k^{st} \equiv 0$. Both results are extensions of the classical relative entropy stability framework going back to seminal works of Dafermos and DiPerna. The attentive reader will note that on the technical level there are some differences between [17, Lemma 3.3] and Theorem 4.2. This is due to the following consideration: If L^2 stability results are to be inferred from the relative entropy framework, upper and lower bounds on the Hessian of the entropy and an upper bound on the Hessian of the flux F are needed. To this end, Fjordholm et.al. restrict their attention to systems for which such bounds exist globally, while we impose conditions (8), (9) and discuss situations in which they are satisfied (including the setting of [17, Lemma 3.3]), see Remark 4.3.

3 Numerical approximation of statistical solutions

This section is concerned with the description of the numerical approximation of statistical solutions. Following [16, 17] the stochastic discretization relies on a Monte-Carlo, resp. collocation approach. Once samples are picked the problem at each sample is deterministic and we approximate it using the Runge–Kutta Discontinuous Galerkin method which we outline briefly. Moreover, we introduce a Lipschitz continuous reconstruction of the numerical solutions which is needed for our a posteriori error estimate in Theorem 4.2. Let us start with the deterministic space and time discretization.

3.1 Space and time discretization: Runge–Kutta discontinuous Galerkin method

We briefly describe the space and time discretization of (1), using the Runge–Kutta Discontinuous Galerkin (RKDG) method from [6]. Let $\mathcal{T} := \{I_l\}_{l=0}^{N_s-1}$, $I_l := (x_l, x_{l+1})$ be a quasi-uniform triangulation of D. We set $h_l := (x_{l+1} - x_l)$, $h_{\max} := \max_l h_l$, $h_{\min} := \min_l h_l$ for the spatial mesh and the (spatial) piecewise polynomial DG spaces for $p \in \mathbb{N}_0$ are defined as

$$V_h^p := \{v : D \to \mathbb{R}^m \mid v\mid_I \in \mathbb{P}_p(I, \mathbb{R}^m), \ \text{for all } I \in \mathcal{T}\}.$$

Here $\mathbb{P}_p$ denotes the space of polynomials of degree at most p and $\mathcal{L}_{V_h^p}$ denotes the L^2-projection into the DG space V_h^p. After spatial discretization of (1) we obtain the following semi-discrete scheme for the discrete solution $u_h \in C^1([0, T); V_h^p)$:

$$\begin{cases} \sum\limits_{l=0}^{N_s-1} \int\limits_{x_l}^{x_{l+1}} \partial_t u_h \cdot \psi_h \, dx \ = \ \sum\limits_{l=0}^{N_s-1} \int\limits_{x_l}^{x_{l+1}} L_h(u_h) \cdot \psi_h \, dx, \quad \forall \psi_h \in V_h^p, \\ u_h(t=0) \qquad\qquad = \mathcal{L}_{V_h^p} \bar{u}, \end{cases} \tag{DG}$$

where $L_h : V_h^p \to V_h^p$ is defined by

$$\sum_{l=0}^{N_s-1} \int_{x_l}^{x_{l+1}} L_h(v_h) \cdot \psi_h \, dx = \sum_{l=0}^{N_s-1} \int_{x_l}^{x_{l+1}} f(v_h) \cdot \partial_x \psi_h \, dx$$

$$- \sum_{l=0}^{N_s-1} F(v_h(x_l^-), v_h(x_l^+)) \cdot [\![\psi_h]\!]_l, \qquad \forall v_h, \psi_h \in V_h^p.$$

Here, $F : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ denotes a consistent, conservative and locally Lipschitz-continuous numerical flux, $\psi_h(x^\pm) := \lim_{s \searrow 0} \psi_h(x \pm s)$ are spatial traces and $[\![\psi_h]\!]_l := (\psi_h(x_l^-) - \psi_h(x_l^+))$ are jumps.

The initial-value problem (DG) is advanced in time by a R-th order strong stability preserving Runge–Kutta (SSP-RK) method [26, 32]. To this end, we let $0 = t_0 < t_1 < \ldots < t_{N_t} = T$ be a (non-equidistant) temporal decomposition of [0, T]. We define $\Delta t_n := (t_{n+1} - t_n)$, $\Delta t := \max_n \Delta t_n$. To ensure stability, the explicit time-stepping scheme has to obey the CFL-type condition

$$\Delta t \leqslant C \frac{h_{\min}}{\lambda_{\max}(2p + 1)},$$

where $\lambda_{\max}$ is an upper bound for absolute values of eigenvalues of the flux Jacobian Df and $C \in (0, 1]$. Furthermore, we let $\Pi_h : \mathbb{R}^m \to \mathbb{R}^m$ be the TVBM minmod slope limiter from [7]. The complete S-stage time-marching algorithm for given n-th time-iterate $u_h^n := u_h(t_n, \cdot) \in V_h^p$ can then be written as follows.

Algorithm 1 TVBM Runge–Kutta time-step

1: Set $u_h^{(0)} = u_h^n = u_h(t_n)$.
2: **for** $s = 1, \ldots, S$ **do**
3:     Compute: $u_h^{(s)} = \Pi_h \left(\sum_{l=0}^{s-1} \delta_{sl} w_h^{sl} \right), \quad w_h^{sl} = u_h^{(l)} + \frac{\beta_{sl}}{\delta_{sl}} \Delta t_n L_h(u_h^{(l)})$.
4: **end for**
5: Set $u_h^{n+1} = u_h^{(S)}$.

Note that the initial condition $u_h(t = 0)$ is also limited by Π_h. The parameters δ_{sl} satisfy the conditions $\delta_{sl} \geqslant 0$, $\sum_{l=0}^{s-1} \delta_{sl} = 1$, and if $\beta_{sl} \neq 0$, then $\delta_{sl} \neq 0$ for all $s = 1, \ldots, S, l = 0, \ldots, s$.

3.2 Reconstruction of numerical solution

Our main a posteriori error estimate Theorem 4.2 is based on the relative entropy method of Dafermos and DiPerna [8] and its extension to statistical solutions as described in the recent work [17]. The a posteriori error estimate requires that the approximating solutions are at least Lipschitz-continuous in space and time. To ensure this property, we reconstruct the numerical solution $\{u_h^n\}_{n=0}^{N_t}$ to a Lipschitz-continuous function in space and time. We do not elaborate upon this reconstruction procedure, to keep notation short and simple, but refer to [10, 21, 30]. The reconstruction process provides a computable space-time reconstruction $u^{st} \in W_\infty^1((0,T);V_h^{p+1} \cap C^0(D))$, which allows us to define the following residual.

Definition 3.1 (Space-time residual) We call the function $\mathcal{R}^{st} \in L^2((0,T) \times D;\mathbb{R}^m)$, defined by

$$\mathcal{R}^{st} := \partial_t u^{st} + \partial_x f(u^{st}), \tag{5}$$

the space-time residual for u^{st}.

We would like to stress that the mentioned reconstruction procedure does not only render u^{st} Lipschitz-continuous, but it is also specifically designed to ensure that the residual, defined in (5), has the same decay properties as the error of the RKDG numerical scheme as h tends to zero, cf. [10].

3.3 Computing the empirical measure

Following [16, 17], we approximate the statistical solution of (1) using empirical measures. In this work, we allow for arbitrary sample points and weights. The sample points can either be obtained by randomly sampling the initial measure (Monte-Carlo), or by using roots of corresponding orthogonal polynomials. In this article we focus on the Monte-Carlo sampling. Let us assume that the sampling points are indexed by the set $\mathcal{K} = \{1, \ldots, K\}$, $K \in \mathbb{N}$ and let us denote

the corresponding weights by $\{w_k\}_{k\in\mathcal{K}}$, i.e., for Monte-Carlo sampling we have $w_k = \frac{1}{K}$, for all $k \in \mathcal{K}$. We use the following Monte-Carlo type algorithm from [17] to compute the empirical measure.

Algorithm 2 Algorithm to compute the empirical measure

1: Given: initial measure $\bar{\mu} \in \mathcal{P}(L^p(D))$
2: For some probability space $(\Omega, \sigma, \mathbb{P})$ let $\bar{u} : \Omega \to L^p(D;\mathcal{U})$ be a random field such that the law of $\bar{u}$ with respect to $\mathbb{P}$ coincides with $\bar{\mu}$
3: Draw K samples from $\bar{u}$ and denote the samples by $\{\bar{u}_k\}_{k\in\mathcal{K}}$
4: For every $k \in \mathcal{K}$ compute a numerical approximation $\{(u_{h,k})^n\}_{n=0}^{N_t}$ of (1) with initial data $\bar{u}_k$ using Algorithm 1
5: Set $\mu_T^{K,h} := \sum_{k\in\mathcal{K}} w_k \delta_{u_{h,k}(T)}$

Definition 3.2 (Empirical measure) Let $\{u_{h,k}(t)\}_{k\in\mathcal{K}}$ be a sequence of approximate solutions of (1) with initial data $\bar{u}_k$, at time $t \in (0, T)$. We define the empirical measure at time $t \in (0, T)$ via

$$\mu_t^{K,h} := \sum_{k\in\mathcal{K}} w_k \delta_{u_{h,k}(t)}. \tag{6}$$

For the a posteriori error estimate in Theorem 4.2, we need to consider a regularized measure, which we define using the reconstructed numerical solution u^{st} obtained from the reconstruction procedure described in Sect. 3.2.

Definition 3.3 [Regularized empirical measure] Let $\{u_k^{st}(t)\}_{k\in\mathcal{K}}$ be a sequence of reconstructed numerical approximations at time $t \in (0, T)$. We define the regularized empirical measure as follows

$$\hat{\mu}_t^K = \sum_{k\in\mathcal{K}} w_k \delta_{u_k^{st}(t)}. \tag{7}$$

Common metrics for computing distances between probability measures on Banach spaces are the Wasserstein-distances. In Theorem 4.2, we bound the error between the dissipative statistical solution of (1) and the discrete regularized measure in the 2-Wasserstein distance.

Definition 3.4 (r-Wasserstein distance) Let $r \in [0, \infty)$, X be a separable Banach space and let $\mu, \rho \in \mathcal{P}(X)$ with finite rth moment, i.e., $\int_X \|u\|_X^r \, d\mu(u) < \infty$, $\int_X \|u\|_X^r \, d\rho(u) < \infty$. The r-Wasserstein distance between μ and ρ is defined as

$$W_r(\mu, \rho) := \left(\inf_{\pi\in\Pi(\mu,\rho)} \int_{X^2} \|u - v\|_X^r \, d\pi(u, v) \right)^{\frac{1}{r}},$$

where $\Pi(\mu, \rho) \subset \mathcal{P}(X^2)$ is the set of all transport plans from μ to ρ, i.e., the set of measures π on X^2 with marginals μ and ρ, i.e.,

$$\pi(A \times X) = \mu(A), \qquad \pi(X \times A) = \rho(A),$$

for all measurable subsets A of X.

Remark 3.5 It is important to recall from [15, Lemma 4.2] that if $\rho \in \mathcal{P}(X)$ is K-atomic, i.e., $\rho = \sum_{i=1}^{K} \alpha_i \delta_{u_i}$ for some $u_1, \dots, u_K \in X$ and some $\alpha_1, \dots, \alpha_K \geqslant 0$ with $\sum_{i=1}^{K} \alpha_i = 1$, then there is a one-to-one correspondence between transport plans in $\Pi(\mu, \rho)$ and elements of $\Lambda(\alpha, \mu)$, which was defined in (4).

4 The a posteriori error estimate

In this section, we present the main a posteriori error estimate between dissipative statistical solutions and regularized numerical approximations. A main ingredient for the proof of Theorem 4.2 is the notion of relative entropy. Since the relative entropy framework is essentially an L^2-framework and due to the quadratic bounds in (8), (9) it is natural to consider the case $p = 2$. Therefore, for the remaining part of this paper we denote $\mathcal{F} := L^2(D;\mathcal{U})$.

Definition 4.1 (Relative entropy and relative flux) Let (η, q) be a strictly convex entropy/entropy flux pair of (1). We define the relative entropy $\eta(\cdot|\cdot) : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$, and relative flux $f(\cdot|\cdot) : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^m$ via

$$\eta(u|v) := \eta(u) - \eta(v) - D\,\eta(v)(u - v),$$
$$f(u|v) := f(u) - f(v) - Df(v)(u - v),$$

for all $u, v \in \mathcal{U}$.

We can now state the main result of this work which is an a posteriori error estimate between dissipative statistical solutions and their numerical approximations using the regularized empirical measure.

Theorem 4.2 Let μ_t be a dissipative statistical solution of (1) as in Definition 2.9 and let $\hat{\mu}_t^K$ be the regularized empirical measure from Definition 3.3. Assume that there exist constants $A, B > 0$ (independent of h, K) such that for all $t \in [0, T)$

$$|u(x) - v(x)|^2 + |f(u(x)|v(x))| \leqslant A\eta(u(x)|v(x)) \leqslant A^2|u(x) - v(x)|^2, \qquad (8)$$

$$|(u(x) - v(x))^{\top} H_v \eta(v(x))(u(x) - v(x))| \leqslant B|u(x) - v(x)|^2 \qquad (9)$$

for a.e. $x \in D$, for $\mu_t - a.e. u \in \mathcal{F}$, for $\hat{\mu}_t^K$ – a.e. $v \in \mathcal{F}$. Here H_v denotes the Hessian matrix w.r.t. v.

Then the distance between μ_t and $\hat{\mu}_t^K$ satisfies

$$W_2(\mu_s, \hat{\mu}_s^K)^2 \leqslant A \left(\sum_{k \in \mathcal{K}} w_k \left[\int\limits_0^s \int\limits_D |\mathcal{R}_k^{st}|^2 \, \mathrm{d}x\mathrm{d}t \right] + A W_2(\bar{\mu}, \hat{\mu}^K)^2 \right) \exp \left(s A^3 B (L+1) \right),$$

for   a.e. $s \in (0, T)$ and $L := \max\limits_{k \in \mathcal{K}} \|\partial_x u_k^{st}\|_{L^\infty((0,s)\times D)}$.   Here   we   denoted
$\hat{\mu}^K := \sum_{k \in \mathcal{K}} w_k \delta_{\bar{u}_k^{st}}.$

Remark 4.3 There are several settings that guarantee validity of (8). Two of them are

1. The global bounds of the Hessians of f and η assumed in [17, Lemma 3.3].
2. Let $\mathcal{C} \Subset \mathcal{U}$ be a compact and convex set. Let us assume that μ_t and $\hat{\mu}_t^K$ are (for all t) supported on solutions having values in $\mathcal{C}$ only. Due to the regularity of f, η and the compactness of $\mathcal{C}$, there exist constants $0 < C_{\bar{f},\mathcal{C}} < \infty$ and $0 < C_{\underline{\eta},\mathcal{C}} < C_{\bar{\eta},\mathcal{C}} < \infty$, such that

$$|u^\top H_v f(v) u| \leqslant C_{\bar{f},\mathcal{C}} |u|^2, \quad C_{\underline{\eta},\mathcal{C}} |u|^2 \leqslant |u^\top H_v \eta(v) u| \leqslant C_{\bar{\eta},\mathcal{C}} |u|^2, \quad \forall u \in \mathbb{R}^m, v \in \mathcal{C}. \tag{10}$$

In this case, A, B from (8), (9) can be chosen as

$$A = \max\{(1 + C_{\bar{f},\mathcal{C}}) C_{\underline{\eta},\mathcal{C}}^{-1}, C_{\bar{\eta},\mathcal{C}}\}, \quad B = C_{\bar{\eta},\mathcal{C}}. \tag{11}$$

Remark 4.4 Let us briefly outline a fundamental difference between Theorem 4.2 and [17, Thm 4.9]. The latter result is an *a priori* estimate that treats the empirical measure (created via Monte Carlo sampling) as a random variable and infers convergence in the mean via the law of large numbers. Theorem 4.2 provides an *a posteriori* error estimate that can be evaluated once the numerical scheme has produced an approximate solution. This process involves sampling the initial data measure $\bar{\mu}$ and the "quality" of these samples enters the error estimate via the term $W_2(\bar{\mu}, \hat{\mu}^K)$.

Remark 4.5 Note that the error bound in Theorem 4.2 depends exponentially on the spatial Lipschitz constant of the regularized empirical measure. Thus, if the exact solution μ_t assigns positive measure to some set of discontinuous functions then the error estimator from Theorem 4.2 is expected to blow up under mesh refinement.

Proof We recall that the weights $\{w_k\}_{k \in \mathcal{K}}$ satisfy $\sum_{k \in \mathcal{K}} w_k = 1$. We further denote the vector of weights by $\vec{w} := (w_1, \ldots, w_K)$, and let $\bar{\mu}^* = (\bar{\mu}_1^*, \ldots, \bar{\mu}_K^*) \in \Lambda(\vec{w}, \bar{\mu})$ correspond to an optimal transport plan between $\bar{\mu}$ and $\hat{\mu}^K$. Because μ_t is a dissipative statistical solution, there exists $(\mu_{1,t}^*, \ldots, \mu_{K,t}^*) \in \Lambda(\vec{w}, \mu_t)$, such that

$$\sum_{k \in \mathcal{K}} w_k \left[\int_0^T \int_{\mathcal{F}} \int_D \left(u \cdot \partial_t \phi_k + f(u) \cdot \partial_x \phi_k \right) \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \right.$$
$$\left. + \int_{\mathcal{F}} \int_D \bar{u} \cdot \phi_k(0, \cdot) \mathrm{d}x \mathrm{d}\bar{\mu}_k^*(\bar{u}) \right] = 0, \tag{12}$$

for every $\phi_k \in C_c^\infty([0,T] \times D; \mathbb{R}^m)$, $k \in \mathcal{K}$. Recalling that each function u_k^{st} is a Lipschitz-continuous solution of the perturbed conservation law (5), considering its weak formulation yields

$$\int_0^T \int_D \left(u_k^{st} \cdot \partial_t \phi_k + f(u_k^{st}) \cdot \partial_x \phi_k \right) \mathrm{d}x \mathrm{d}t + \int_D \bar{u}_k^{st} \cdot \phi_k(0, \cdot) \mathrm{d}x + \int_0^T \int_D \mathcal{R}_k^{st} \cdot \phi_k \, \mathrm{d}x \mathrm{d}t = 0, \tag{13}$$

for every $\phi_k \in C_c^\infty([0,T] \times D; \mathbb{R}^m)$. As $(\mu_{k,t}^*)_{k \in \mathcal{K}}$ are probability measures on $\mathcal{F}$, we obtain (after changing order of integration)

$$0 = \sum_{k \in \mathcal{K}} w_k \left[\int_0^T \int_{\mathcal{F}} \int_D \left(u_k^{st} \cdot \partial_t \phi_k + f(u_k^{st}) \cdot \partial_x \phi_k \right) \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \right.$$
$$\left. + \int_{\mathcal{F}} \int_D \bar{u}_k^{st} \cdot \phi_k(0, \cdot) \mathrm{d}x \mathrm{d}\bar{\mu}_k^*(\bar{u}) + \int_0^T \int_{\mathcal{F}} \int_D \mathcal{R}_k^{st} \cdot \phi_k \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \right]. \tag{14}$$

Subtracting (14) from (12) and using the Lipschitz-continuous test function $\phi_k(t,x) := \mathrm{D}\,\eta(u_k^{st}(t,x))\phi(t)$, where $\phi \in C_c^\infty([0,T]; \mathbb{R}_+)$ yields

$$0 = \sum_{k \in \mathcal{K}} w_k \left[\int_0^T \int_{\mathcal{F}} \int_D \left((u - u_k^{st}) \cdot \partial_t (\mathrm{D}\,\eta(u_k^{st})\phi) \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \right. \right.$$
$$+ \int_0^T \int_{\mathcal{F}} \int_D (f(u) - f(u_k^{st})) \cdot \partial_x (\mathrm{D}\,\eta(u_k^{st})\phi) \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t$$
$$\left. + \int_{\mathcal{F}} \int_D (\bar{u} - \bar{u}_k^{st}) \cdot \mathrm{D}\,\eta(\bar{u}_k^{st})\phi(0) \, \mathrm{d}x \mathrm{d}\bar{\mu}_k^*(\bar{u}) - \int_0^T \int_{\mathcal{F}} \int_D \mathcal{R}_k^{st} \cdot \mathrm{D}\,\eta(u_k^{st})\phi \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \right]. \tag{15}$$

We compute the partial derivatives of $\mathrm{D}\,\eta(u_k^{st}(t,x))\phi(t)$ using product and chain rule

$$\partial_t (\mathrm{D}\,\eta(u_k^{st})\phi) = \partial_t u_k^{st} \cdot H\eta(u_k^{st})\phi + \partial_t \phi \, \mathrm{D}\,\eta(u_k^{st}), \tag{16}$$

$$\partial_x (\mathrm{D}\,\eta(u_k^{st})\phi) = \partial_x u_k^{st} \cdot H\eta(u_k^{st})\phi. \tag{17}$$

Next, we multiply (5) by $D\eta(u_k^{st})$. Upon using the chain rule for Lipschitz-continuous functions and the relationship $Dq(u_k^{st}) = D\eta(u_k^{st})Df(u_k^{st})$ we derive the relation

$$D\eta(u_k^{st}) \cdot \mathcal{R}_k^{st} = \partial_t \eta(u_k^{st}) + \partial_x q(u_k^{st}). \tag{18}$$

Let us consider the weak form of (18) and integrate w.r.t. x, t and $d\mu_{k,t}^*$ for $k \in \mathcal{K}$. Upon changing the order of integration we have

$$
0 = \sum_{k \in \mathcal{K}} w_k \Bigg[\int_0^T \int_{\mathcal{F}} \int_D \eta(u_k^{st}) \partial_t \phi \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \\
+ \int_0^T \int_{\mathcal{F}} \int_D \mathcal{R}_k^{st} \cdot D\eta(u_k^{st}) \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t + \int_{\mathcal{F}} \int_D \eta(\bar{u}_k^{st}) \phi(0) \, \mathrm{d}x \mathrm{d}\bar{\mu}_k^*(\bar{u}) \Bigg] \tag{19}
$$

for any $\phi \in C_c^\infty([0,T];\mathbb{R}_+)$. Since $\mu_{k,t}^*$ is a dissipative statistical solution it satisfies

$$
0 \leqslant \sum_{k \in \mathcal{K}} w_k \Bigg[\int_0^T \int_{\mathcal{F}} \int_D \eta(u) \partial_t \phi \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t + \int_{\mathcal{F}} \int_D \eta(\bar{u}) \phi(0) \, \mathrm{d}x \mathrm{d}\bar{\mu}_k^*(\bar{u}) \Bigg]. \tag{20}
$$

Hence, subtracting (19) from (20) and using the definition of the relative entropy from Definition 4.1 yields

$$
0 \leqslant \sum_{k \in \mathcal{K}} w_k \Bigg[\int_0^T \int_{\mathcal{F}} \int_D \eta(u|u_k^{st}) \partial_t \phi \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \\
+ \int_0^T \int_{\mathcal{F}} \int_D \Big((u - u_k^{st}) \cdot D\eta(u_k^{st}) \partial_t \phi \Big) \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \\
- \int_0^T \int_{\mathcal{F}} \int_D \mathcal{R}_k^{st} \cdot D\eta(u_k^{st}) \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t + \int_{\mathcal{F}} \int_D \Big(\eta(\bar{u}) - \eta(\bar{u}_k^{st}) \Big) \phi(0) \, \mathrm{d}x \mathrm{d}\bar{\mu}_k^*(\bar{u}) \Bigg]. \tag{21}
$$

After subtracting (15) from (21) and using (16), (17) we are led to

$$
\begin{aligned}
0 \leqslant \sum_{k \in \mathcal{K}} w_k \Bigg[& \int_0^T \int_{\mathcal{F}} \int_D \eta(u|u_k^{st}) \partial_t \phi \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \\
& - \int_0^T \int_{\mathcal{F}} \int_D (u - u_k^{st}) \cdot H\eta(u_k^{st}) \partial_t u_k^{st} \phi \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \\
& - \int_0^T \int_{\mathcal{F}} \int_D \left(f(u) - f(u_k^{st}) \right) \cdot H\eta(u_k^{st}) \partial_x u_k^{st} \phi \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \\
& + \int_{\mathcal{F}} \int_D \eta(\bar{u}|\bar{u}_k^{st}) \phi(0) \, \mathrm{d}x \mathrm{d}\bar{\mu}_k^*(\bar{u}) \Bigg].
\end{aligned}
\tag{22}
$$

Rearranging (5) yields

$$
\partial_t u_k^{st} = -\mathrm{D}f(u_k^{st}) \partial_x u_k^{st} + \mathcal{R}_k^{st}.
\tag{23}
$$

Plugging (23) into (22) and after rearranging we have

$$
\begin{aligned}
0 \leqslant \sum_{k \in \mathcal{K}} w_k \Bigg[& \int_0^T \int_{\mathcal{F}} \int_D \eta(u|u_k^{st}) \partial_t \phi \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \\
& - \int_0^T \int_{\mathcal{F}} \int_D \partial_x u_k^{st} \cdot H\eta(u_k^{st}) f(u|u_k^{st}) \phi \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \\
& - \int_0^T \int_{\mathcal{F}} \int_D (u - u_k^{st}) \cdot H\eta(u_k^{st}) \mathcal{R}_k^{st} \phi \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \\
& + \int_{\mathcal{F}} \int_D \eta(\bar{u}|\bar{u}_k^{st}) \phi(0) \, \mathrm{d}x \mathrm{d}\bar{\mu}_k^*(\bar{u}) \Bigg],
\end{aligned}
\tag{24}
$$

where we have used $\mathrm{D}f \cdot H\eta = H\eta \, \mathrm{D}f$. Up to now, the choice of $\phi(t)$ was arbitrary. We fix $s > 0$ and $\epsilon > 0$ and define ϕ as follows

$$
\phi(\sigma) := \begin{cases}
1 & : \sigma < s, \\
1 - \frac{\sigma - s}{\epsilon} & : s < \sigma < s + \epsilon, \\
0 & : \sigma > s + \epsilon.
\end{cases}
$$

According to Theorem 2.5 (a) we have that the mapping

$$
t \mapsto \int_{\mathcal{F}} \int_D \eta(u|u_k^{st}(t, \cdot)) \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u)
\tag{25}
$$

is measurable for all $k \in \mathcal{K}$. Moreover, due to the quadratic bound on the relative entropy, cf. (8), Lebesgue's differentiation theorem states that a.e. $t \in (0, T)$ is a Lebesgue point of (25). Thus, letting $\epsilon \to 0$ we obtain

$$
\begin{aligned}
\sum_{k \in \mathcal{K}} & w_k \int_{\mathcal{F}} \int_D \eta(u|u_k^{st}(s, \cdot)) \, \mathrm{d}x \mathrm{d}\mu_{k,s}^*(u) \\
& \leqslant \sum_{k \in \mathcal{K}} w_k \Bigg[- \int_0^s \int_{\mathcal{F}} \int_D \partial_x u_k^{st} \cdot H\eta(u_k^{st}) f(u|u_k^{st}) \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \\
& \quad - \int_0^s \int_{\mathcal{F}} \int_D (u - u_k^{st}) \cdot H\eta(u_k^{st}) \mathcal{R}_k^{st} \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \\
& \quad + \int_{\mathcal{F}} \int_D \eta(\bar{u}|\bar{u}_k^{st}) \, \mathrm{d}x \mathrm{d}\bar{\mu}_k^*(\bar{u}) \Bigg].
\end{aligned}
\tag{26}
$$

The left hand side of (26) is bounded from below using (8). The first term on the right hand is estimated using the $L^\infty(D)$-norm of the spatial derivative. We estimate the second term on the right hand side by Young's inequality. Finally, we apply (8) and then (9) to both terms. The last term on the right hand side is estimated using (8). We, thus, end up with

$$
\begin{aligned}
\sum_{k \in \mathcal{K}} & w_k \Bigg[A^{-1} \int_{\mathcal{F}} \int_D |u - u_k^{st}(s, \cdot)|^2 \, \mathrm{d}x \mathrm{d}\mu_{k,s}^*(u) \Bigg] \\
& \leqslant \sum_{k \in \mathcal{K}} w_k \Bigg[\int_0^s \left(A^2 B \|\partial_x u_k^{st}(t, \cdot)\|_{L^\infty(D)} + A^2 B \right) \int_{\mathcal{F}} \int_D |u - u_k^{st}(t, \cdot)|^2 \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t \\
& \quad + \int_0^s \int_{\mathcal{F}} \int_D |\mathcal{R}_k^{st}|^2 \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t + A \int_{\mathcal{F}} \int_D |\bar{u} - \bar{u}_k^{st}|^2 \, \mathrm{d}x \mathrm{d}\bar{\mu}_k^*(\bar{u}) \Bigg].
\end{aligned}
$$

Upon using Grönwall's inequality we obtain

$$
\begin{aligned}
\sum_{k \in \mathcal{K}} & w_k \Bigg[\int_{\mathcal{F}} \int_D |u - u_k^{st}(s, \cdot)|^2 \, \mathrm{d}x \mathrm{d}\mu_{k,s}^*(u) \Bigg] \\
& \leqslant \sum_{k \in \mathcal{K}} w_k \Bigg[A \Big(\int_0^s \int_{\mathcal{F}} \int_D |\mathcal{R}_k^{st}|^2 \, \mathrm{d}x \mathrm{d}\mu_{k,t}^*(u) \mathrm{d}t + A \int_{\mathcal{F}} \int_D |\bar{u} - \bar{u}_k^{st}|^2 \, \mathrm{d}x \mathrm{d}\bar{\mu}_k^*(\bar{u}) \Big) \\
& \quad \times \exp \Big(\int_0^s A \Big(A^2 B \|\partial_x u_k^{st}(t, \cdot)\|_{L^\infty(D)} + A^2 B \Big) \mathrm{d}t \Bigg].
\end{aligned}
$$

Using $\max\limits_{k\in\mathcal{K}}\|\partial_x u_k^{st}\|_{L^\infty((0,s)\times D)} =: L$ and recalling that $(\overline{\mu}_k^*)_{k\in\mathcal{K}}$ corresponds to an optimal transport plan and that $(\mu_{k,s}^*)_{k\in\mathcal{K}}$ corresponds to an admissible transport plan, we finally obtain

$$W_2(\mu_s, \hat{\mu}_s^K)^2 \leqslant A\left(\sum_{k\in\mathcal{K}} w_k\left[\int\limits_0^s \int\limits_D |\mathcal{R}_k^{st}|^2 \, dx dt\right] + AW_2(\overline{\mu}, \hat{\mu}^K)^2\right)$$

$$\times \exp\left(\int\limits_0^s \left(A^3 BL + A^3 B\right) dt\right).$$

$\square$

To obtain an error estimate with separated bounds, i.e., bounds that quantify spatio-temporal and stochastic errors, respectively, we split the 2-Wasserstein error in initial data into a spatial and a stochastic part. Using the triangle inequality we obtain

$$W_2(\overline{\mu}, \hat{\mu}^K) \leqslant W_2\left(\overline{\mu}, \sum_{k\in\mathcal{K}} w_k \delta_{\overline{u}_k}\right) + W_2\left(\sum_{k\in\mathcal{K}} w_k \delta_{\overline{u}_k}, \hat{\mu}^K\right). \tag{27}$$

The first term in (27) is a stochastic error, which is inferred from approximating the initial data by an empirical measure. On the other hand, the second term is essentially a spatial approximation error. This can be seen from the following lemma.

Lemma 4.6 *With the same notation as in Theorem* 4.2, *the following inequality holds.*

$$W_2\left(\sum_{k\in\mathcal{K}} w_k \delta_{\overline{u}_k}, \hat{\mu}^K\right)^2 \leqslant \sum_{k\in\mathcal{K}} w_k\|\overline{u}_k - \overline{u}_k^{st}\|_{\mathcal{F}}^2. \tag{28}$$

Equality in (28) *holds provided spatial discretization errors are smaller than distances between samples. A sufficient condition is:*

$$\|\overline{u}_k - \overline{u}_k^{st}\|_{\mathcal{F}}^2 \leqslant \frac{1}{2}\min\limits_{\ell \neq k} \|\overline{u}_k - \overline{u}_\ell\|_{\mathcal{F}}^2 \quad \forall k \in \mathcal{K} \tag{29}$$

Proof Recalling the definition of $\hat{\mu}^K = \sum_{k\in\mathcal{K}} w_k \overline{u}_k^{st}$ and defining the transport plan $\pi^* := \sum_{k\in\mathcal{K}} w_k(\delta_{\overline{u}_k} \otimes \delta_{\overline{u}_k^{st}})$ yields the assertion, because

$$W_2\left(\sum_{k\in\mathcal{K}} w_k \delta_{\overline{u}_k}, \hat{\mu}^K\right)^2 \leqslant \int_{\mathcal{F}^2} \|x - y\|_{L^2(D)}^2 \, d\pi^*(x, y) = \sum_{k\in\mathcal{K}} w_k\|\overline{u}_k - \overline{u}_k^{st}\|_{\mathcal{F}}^2.$$

If (29) holds, π^* is an optimal transport plan. $\square$

Remark 4.7 In contrast to random conservation laws as considered in [30], the stochastic part of the error estimator of Theorem 4.2 is solely given by the discretization

error of the initial data, i.e., given by the second term in (27), which may be amplified due to nonlinear effects. However, there is no stochastic error source during the evolution of the numerical approximation.

5 Extension to general systems (with $\mathcal{U} \neq \mathbb{R}^m$)

Up to this point, we have stuck with the setting from [17] to stress that a posteriori error estimates can be obtained as long as the relative entropy framework from [17] is applicable. Having said this, it is fairly clear that this setting does not cover certain systems of practical interest, e.g. the Euler equations of fluid dynamics.

This raises the question how statistical solutions can be defined for more general systems. It seems to be sufficient to require μ_t to be a measure on some set of functions $\mathcal{F}$ so that $u \in \mathcal{F}$ implies that $u, f(u)$ and $\eta(u)$ are integrable. In particular, such a definition ensures (fairly generically) that $\int_{\mathcal{F}} \int_D f(u) \mathrm{d}x d\mu(u)$ is well-defined, i.e. that $\qquad E : \mathcal{F} \to \mathbb{R}^m, \; u \mapsto E(u) := \int_D f(u) \mathrm{d}x \qquad$ is $\qquad$ measurable. $\qquad$ For $\mathcal{F} = L_f^r := \{u \in L^r : \int_D |f(u(x))| \, \mathrm{d}x < \infty\}$ for some $r \in [1, \infty]$ and $\mathcal{F}$ being equipped with the L^r topology this can be seen as follows: For any $M > 0$ the functional $u \mapsto E_M(u) := \int_D \max\{-M, \min\{M, f(u)\}\} \mathrm{d}x$ is continuous and E is measurable as pointwise limit of continuous functions. The same argument works for $\int_{\mathcal{F}} \int_D \eta(u) \mathrm{d}x d\mu(u)$.

However, it is not clear to us, whether the a posteriori error estimate in Sect. 4 can be extended to all such settings in which statistical solutions can be defined since it requires (8), (9). . One interesting, albeit somewhat restrictive, setting in which (8), (9) hold is the case that all measures under consideration are supported on some set of functions taking values in some compact subset of the state space. We will provide a definition of statistical solutions and an a posteriori error estimate in this setting.

Definition 5.1 ($\mathcal{C}$-valued statistical solution) Let $\mathcal{C} \Subset \mathcal{U}$ be compact and convex. Let $\bar{\mu} \in \mathcal{P}(L^p(D; \mathcal{U}))$ be supported on $L^p(D; \mathcal{C})$. A $\mathcal{C}$-valued statistical solution of (1) with initial data $\bar{\mu}$ is a time-dependent map $\mu : [0, T] \to \mathcal{P}(L^p(D; \mathcal{U}))$ such that each μ_t is supported on $L^p(D; \mathcal{C})$ for all $t \in [0, T)$ and such that the corresponding correlation measures ν_t^k satisfy

$$\partial_t \langle \nu_{t,x}^k, \xi_1 \otimes \cdots \otimes \xi_k \rangle + \sum_{i=1}^k \partial_{x_i} \langle \nu_{t,x_1,\dots,x_k}^k, \xi_1 \otimes \cdots \otimes f(\xi_i) \otimes \cdots \otimes \xi_k \rangle = 0 \quad (30)$$

in the sense of distributions, for every $k \in \mathbb{N}$

Remark 5.2 Note that duality implies that if μ is supported on $L^p(D; \mathcal{C})$ then the corresponding ν^k is supported on $\mathcal{C}^k$ (for all $k \in \mathbb{N}$).

Definition 5.3 (Dissipative $\mathcal{C}$-valued statistical solution) A $\mathcal{C}$-valued statistical solution of (1) is called a dissipative statistical solution provided the conditions from

Definition 2.9 hold with "statistical solution" being replaced by "$\mathcal{C}$-valued statistical solution".

We immediately have the following theorem whose proof follows *mutatis mutandis* from the proof of Theorem 4.2

Theorem 5.4 *Let μ_t be a $\mathcal{C}$-valued dissipative statistical solution of* (1) *with initial data $\bar{\mu}$ as in Definition* 2.9 *and let $\hat{\mu}_t^K$ be the regularized empirical measure as in Definition* 3.3 *and supported on functions with values in $\mathcal{C}$. Then the difference between μ_t and $\hat{\mu}_t^K$ satisfies*

$$W_2(\mu_s, \hat{\mu}_s^K)^2 \leqslant A \left(\sum_{k \in \mathcal{K}} w_k \left[\int_0^s \int_D |\mathcal{R}_k^{st}|^2 \, dx dt \right] + A W_2(\bar{\mu}, \hat{\bar{\mu}}^K)^2 \right) \exp\left(s A^3 B (L+1) \right),$$

for a.e. $s \in (0, T)$ and $L := \max_{k \in \mathcal{K}} \|\partial_x u_k^{st}\|_{L^\infty((0,s) \times D)}$.

6 Numerical experiments

In this section we illustrate how to approximate the 2-Wasserstein distances that occur in Theorem 4.2. Moreover, we examine the scaling properties of the estimators in Theorem 4.2 by means of a smooth and a non-smooth solution of the one-dimensional compressible Euler equations.

6.1 Numerical approximation of Wasserstein distances

We illustrate how to approximate the 2-Wasserstein distance on the example of $W_2(\bar{\mu}, \sum_{k \in \mathcal{K}} w_k \delta_{\bar{u}_k})$ from (27). For the given initial measure $\bar{\mu} \in \mathcal{P}(L^2(D))$ we choose a probability space $(\Omega, \sigma, \mathbb{P})$ and a random field $\bar{u} \in L^2(\Omega; L^2(D))$, such that the law of $\bar{u}$ with respect to $\mathbb{P}$ coincides with $\bar{\mu}$. We approximate the initial measure $\bar{\mu}$ using some empirical measure $\sum_{m \in \mathcal{M}} w_m \delta_{\bar{u}_m}$, for a second sample set $\mathcal{M} := \{1, \ldots, M\}$, where the number of samples $M \gg K$ is significantly larger than the number of samples of the numerical approximation. To distinguish between the two different sample sets $\mathcal{K}$ and $\mathcal{M}$, we write $\{\bar{u}_k^{\mathcal{K}}\}_{k \in \mathcal{K}}$, $\{\bar{u}_m^{\mathcal{M}}\}_{m \in \mathcal{M}}$ and $\{w_k^{\mathcal{K}}\}_{k \in \mathcal{K}}$, $\{w_m^{\mathcal{M}}\}_{m \in \mathcal{M}}$ respectively. Finally, we collect the weights $\{w_k^{\mathcal{K}}\}_{k \in \mathcal{K}}$ in the vector $\vec{w}^{\mathcal{K}}$ and $\{w_m^{\mathcal{M}}\}_{m \in \mathcal{M}}$ in the vector $\vec{w}^{\mathcal{M}}$.

Computing the optimal transport π^* (and thus the 2-Wasserstein distance) between the two atomic measures can be formulated as the following linear program, cf. [31, (2.11)]

$$\pi^* = \underset{\pi \in \Pi\left(\sum_{k \in \mathcal{K}} w_k^{\mathcal{K}} \delta_{\bar{u}_k^{\mathcal{K}}}, \sum_{m \in \mathcal{M}} w_m^{\mathcal{M}} \delta_{\bar{u}_m^{\mathcal{M}}} \right)}{\arg \min} \sum_{k=1}^{K} \sum_{m=1}^{M} \pi_{k,m} \mathbb{C}_{k,m}, \tag{31}$$

where $\quad \Pi\Big(\sum_{k\in\mathcal{K}} w_k^{\mathcal{K}} \delta_{\bar{u}_k^{\mathcal{K}}}, \sum_{m\in\mathcal{M}} w_m^{\mathcal{M}} \delta_{\bar{u}_m^{\mathcal{M}}} \Big) := \{\pi \in \mathbb{R}_+^{K\times M} \mid \pi \mathbb{1}_M = \vec{w}^{\mathcal{M}} \text{ and } \pi^\top \mathbb{1}_K = \vec{w}^{\mathcal{K}} \}$
denotes the set of transport plans and $\mathbb{1}_n := (1, \dots, 1)^\top \in \mathbb{R}^n$. Moreover, the entries of the cost matrix $\mathbb{C} \in \mathbb{R}_+^{K\times M}$ are computed as

$$\mathbb{C}_{k,m} = \|\bar{u}_k^{\mathcal{K}} - \bar{u}_m^{\mathcal{M}}\|_{L^2(D)}^2, \quad \text{for all } k = 1, \dots, K, \ m = 1, \dots, M.$$

The linear program is solved using the network simplex algorithm [3, 31], implemented in the optimal transport library `ot.emd` [18] in python. The 2-Wasserstein distance is finally approximated by

$$W_2\Big(\sum_{k\in\mathcal{K}} w_k \delta_{\bar{u}_k}, \bar{\mu} \Big)^2 \approx W_2\Big(\sum_{k\in\mathcal{K}} w_k^{\mathcal{K}} \delta_{\bar{u}_k^{\mathcal{K}}}, \sum_{m\in\mathcal{M}} w_m^{\mathcal{M}} \delta_{\bar{u}_m^{\mathcal{M}}} \Big)^2. \tag{32}$$

where (32) can be computed with `ot.emd2` $(\vec{w}^{\mathcal{K}}, \vec{w}^{\mathcal{M}}, \mathbb{C})$.

6.2 Numerical experiment for a smooth solution

This section is devoted to the numerical study of the scaling properties of the individual terms in the a posteriori error estimate Theorem 4.2. In the following experiment we consider as instance of (1) the one-dimensional compressible Euler equations for the flow of an ideal gas, which are given by

$$\partial_t \rho + \partial_x m = 0,$$
$$\partial_t m + \partial_x \left(\frac{m^2}{\rho} + p \right) = 0, \tag{33}$$
$$\partial_t E + \partial_x \left((E + p) \frac{m}{\rho} \right) = 0,$$

where ρ describes the mass density, m the momentum and E the energy of the gas. The constitutive law for pressure p reads

$$p = (\gamma - 1)\left(E - \frac{1}{2}\frac{m^2}{\rho} \right),$$

with the adiabatic constant $\gamma = 1.4$. We construct a smooth exact solution of (33) by introducing an additional source term. The exact solution reads as follows

$$\begin{pmatrix} \rho(t,x,\omega) \\ m(t,x,\omega) \\ E(t,x,\omega) \end{pmatrix} = \begin{pmatrix} 2 + 0.2 \cdot \xi(\omega) \cdot \cos(6\pi(x-t)) \\ \big(2 + 0.2 \cdot \xi(\omega) \cdot \cos(6\pi(x-t)) \big)\big(1 + 0.2 \cdot \xi(\omega) \cdot \sin(6\pi(x-t)) \big) \\ \big(2 + 0.2 \cdot \xi(\omega) \cdot \cos(6\pi(x-t)) \big)^2 \end{pmatrix},$$
$$\tag{34}$$

where $\xi \sim \mathcal{U}(0,1)$ is a uniformly distributed random variable. Moreover, the spatial domain is $[0,1]_{\text{per}}$ and we compute the solution up to $T = 0.2$. We sample the initial measure $\bar{\mu}$ and the exact measure μ_T with 10000 samples.

For the remaining part of this section we introduce the notations

$$\mathcal{E}^{\mathrm{det}}(s) := \sum_{k \in \mathcal{K}} w_k \left[\int_0^s \int_D |\mathcal{R}_k^{st}|^2 \, \mathrm{d}x \mathrm{d}t \right], \tag{35}$$

$$\mathcal{E}_0^{\mathrm{det}} := \sum_{k \in \mathcal{K}} w_k \|\bar{u}_k - \bar{u}_k^{st}\|_{\mathcal{F}}^2, \tag{36}$$

$$\mathcal{E}_0^{\mathrm{stoch}} := W_2(\bar{\mu}, \sum_{k \in \mathcal{K}} w_k \delta_{\bar{u}_k})^2. \tag{37}$$

Moreover, when we refer to error, we mean the Wasserstein distance between exact and numerically computed density ρ at time $t = T$.

As numerical solver we use the RKDG Code Flexi [25]. The DG polynomial degrees will always be two and for the time-stepping we use a low storage SSP RK-method of order three as in [26]. The time-reconstruction is also of order three. As numerical fluxes we choose the Lax-Wendroff numerical flux

$$F(u, v) := f(w(u, v)), \quad w(u, v) := \frac{1}{2}\left((u + v) + \frac{\Delta t}{h}(f(v) - f(u))\right), \tag{38}$$

Computing $\mathcal{E}^{\mathrm{det}}, \mathcal{E}_0^{\mathrm{det}}, \mathcal{E}_0^{\mathrm{stoch}}$ requires computing integrals, we approximate them via Gauß-Legendre quadrature where we use seven points in each time-cell and ten points in every spatial cell.
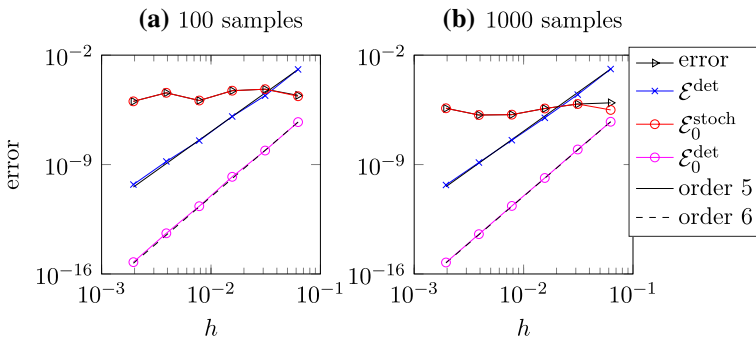
6.2.1 Spatial refinement

In this section we examine the scaling properties of $\mathcal{E}^{\mathrm{det}}(T)$ and $\mathcal{E}_0^{\mathrm{det}}$ when we gradually increase the number of spatial cells. We start initially on a coarse mesh with 16 elements, i.e., $h = 1/16$ and upon each refinement we subdivide each spatial cell uniformly into two cells. To examine a possible influence of the stochastic resolution we consider a coarse and fine sampling with 100 and 1000 samples, respectively. Table 1 and Fig. 1a display the error estimator parts, $\mathcal{E}^{\mathrm{det}}, \mathcal{E}_0^{\mathrm{stoch}}$ and $\mathcal{E}_0^{\mathrm{det}}$ from (35)-(37) when we compute the numerical approximation with 100 samples and Table 2 and Fig. 1b display the same quantities for 1000 samples. We observe that the convergence of $\mathcal{E}^{\mathrm{det}}$ and $\mathcal{E}_0^{\mathrm{det}}$ does not depend on the stochastic resolution. Indeed, both quantities decay when we increase the number of spatial elements and they are uniform with respect to changes in the sample size. Moreover, $\mathcal{E}_0^{\mathrm{det}}$ exhibits the expected order of convergence which is six for a DG polynomial degree of two (note that we compute squared quantities). It can also be observed that $\mathcal{E}^{\mathrm{det}}$ converges with order five, which is due to a suboptimal rate of convergence on the first time-step. This issue has also been discussed in detail in [30, Remark 4.1]. Furthermore, we see that the numerical error remains almost constant upon mesh refinement, since it is dominated by the stochastic resolution error, which is described by $\mathcal{E}_0^{\mathrm{stoch}}$. Since $\mathcal{E}_0^{\mathrm{stoch}}$ reflects the stochastic error it also remains constant upon spatial mesh refinement.

Table 1 Spatial refinement for 100 samples

h-refinement, 100 samples				
h	Error	$\mathcal{E}^{\text{det}}$	$\mathcal{E}_0^{\text{stoch}}$	$\mathcal{E}_0^{\text{det}}$
1/16	2.6610e-05	0.0012	2.3481e-05	5.2125e-07
1/32	6.6226e-05	2.5710e-05	6.6362e-05	7.8654e-09
1/64	5.3067e-05	1.1870e-06	5.3054e-05	1.6428e-10
1/128	1.2506e-05	3.4142e-08	1.2506e-05	2.1617e-12
1/256	3.8927e-05	1.5755e-09	3.8927e-05	4.0005e-14
1/512	1.1071e-05	5.2537e-11	1.1071e-05	5.6958e-16

Numerical example from Sect. 6.2

Fig. 1 Spatial refinement for 100 and 1000 samples. Numerical example from Sect. 6.2

Table 2 Spatial refinement for 1000 samples

| *h*-refinement, 1000 samples				
h	Error	$\mathcal{E}^{\text{det}}$	$\mathcal{E}_0^{\text{stoch}}$	$\mathcal{E}_0^{\text{det}}$
1/16	8.8879e-06	0.0013	3.1289e-06	5.4848e-07
1/32	7.4574e-06	3.0015e-05	7.4498e-06	9.1929e-09
1/64	3.8515e-06	9.7543e-07	3.8531e-06	1.3842e-10
1/128	1.5582e-06	3.6073e-08	1.5582e-06	2.2541e-12
1/256	1.4771e-06	1.3478e-09	1.4771e-06	3.5150e-14
1/512	3.9935e-06	4.9174e-11	3.9935e-06	5.4300e-16

Numerical example from Sect. 6.2

6.2.2 Stochastic refinement

In this section, we consider stochastic refinement, i.e., we increase the number of samples and keep the spatial resolution fixed. Similarly to the numerical example in the previous section we consider a very coarse spatial discretization with 8 elements (Table 3, Fig. 2a) and a fine spatial discretization with 256 elements (Table 4, Fig. 2b). We observe that $\mathcal{E}^{\text{det}}$ and $\mathcal{E}_0^{\text{det}}$ remain constant when we increase the

Table 3 Stochastic refinement for 8 spatial cells

Stochastic refinement, $h = 1/8$				
Samples	Error	$\mathcal{E}^{\text{det}}$	$\mathcal{E}_0^{\text{stoch}}$	$\mathcal{E}_0^{\text{det}}$
2	0.0017	0.3096	0.0014	2.3085e-05
4	0.0011	0.3464	0.0004	2.9522e-05
8	0.0014	0.3784	0.0002	3.5952e-05
16	0.0004	0.3131	0.0002	2.4098e-05
32	0.0014	0.3876	0.0001	3.7421e-05
64	0.0005	0.3345	5.3368e-05	2.7750e-05
128	0.0007	0.3509	2.2416e-05	3.0542e-05
256	0.0008	0.3577	5.2963e-06	3.1865e-05
512	0.0009	0.3692	1.1292e-05	3.3930e-05
1024	0.0007	0.3562	4.2294e-06	3.1578e-05
2048	0.0008	0.3616	1.1969e-06	3.2545e-05

Numerical example from Sect. 6.2
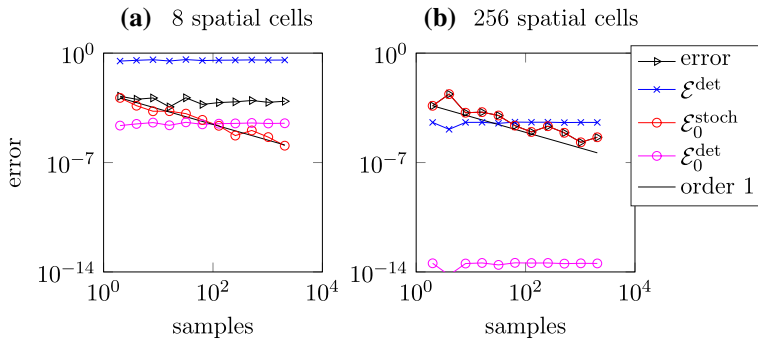
Fig. 2 Stochastic refinement for 8 and 256 spatial cells. Numerical example from Sect. 6.2

Table 4 Stochastic refinement for 256 spatial cells

Stochastic refinement, $h = 1/256$				
Samples	Error	$\mathcal{E}^{\text{det}}$	$\mathcal{E}_0^{\text{stoch}}$	$\mathcal{E}_0^{\text{det}}$
2	0.0004	3.6763e-05	0.0004	3.5764e-14
4	0.0024	1.3302e-05	0.0024	6.2086e-15
8	0.0002	3.7038e-05	0.0002	3.4537e-14
16	0.0002	3.7339e-05	0.0002	3.7582e-14
32	0.0001	3.2223e-05	0.0001	2.8239e-14
64	2.2881e-05	3.8634e-05	2.2881e-05	3.8232e-14
128	9.0158e-06	3.7754e-05	9.0157e-06	3.6971e-14
256	2.1062e-05	3.8056e-05	2.1062e-05	3.7567e-14
512	7.9671e-06	3.6037e-05	7.9671e-06	3.4011e-14
1024	1.9663e-06	3.7004e-05	1.9662e-06	3.5579e-14
2048	4.1378e-06	3.6733e-05	4.1378e-06	3.5074e-14

Numerical example from Sect. 6.2

number of samples. This is the correct behavior, since both residuals reflect spatio-temporal errors. For the coarse spatial discretization we observe that the numerical error does not decrease when increasing the number of samples since the spatial discretization error, reflected by $\mathcal{E}^{\text{det}}$, dominates the numerical error. For the fine spatial discretization the numerical error is dominated by the stochastic resolution error and thus the error converges with the same order. We observe an experimental rate of convergence of one (resp. one half after taking the square root). Finally, we see that $\mathcal{E}_0^{\text{stoch}}$ is independent of the spatial discretization because its convergence is not altered by the spatio-temporal resolution (Tables 5, 6, 7, 8).

6.3 Numerical experiment for a non-smooth solution

In this numerical experiment we consider a spatially non-smooth exact solution given by

$$
\rho(t, x, \omega) = 2 + \sum_{k=1}^{600} \frac{\xi(\omega) \cdot \cos(2\pi k(x - t))}{5k^2}, \quad m(t, x, \omega) = E(t, x, \omega) = \rho(t, x, \omega),
$$

(39)

Table 5 Spatial refinement for 100 samples

h-refinement, 100 samples				
h	Error	$\mathcal{E}^{\text{det}}$	$\mathcal{E}_0^{\text{stoch}}$	$\mathcal{E}_0^{\text{det}}$
1/16	1.0046e-03	5.3974e-01	1.0776e-04	3.0333e-04
1/32	3.1775e-04	4.6198e-01	8.1488e-06	5.1291e-05
1/64	7.7589e-05	2.7323e-01	3.6387e-05	6.2962e-06
1/128	1.5776e-05	1.6943e-01	7.7810e-06	8.8946e-07
1/256	1.5549e-05	9.0180e-02	1.4144e-05	1.1215e-07
1/512	4.5574e-05	4.5733e-02	4.5356e-05	1.3946e-08

Numerical example from Sect. 6.3

Table 6 Spatial refinement for 1000 samples

h-refinement, 100 samples				
h	Error	$\mathcal{E}^{\text{det}}$	$\mathcal{E}_0^{\text{stoch}}$	$\mathcal{E}_0^{\text{det}}$
1/16	1.0338e-03	6.3285e-01	4.4275e-06	3.5692e-04
1/32	3.3777e-04	4.9430e-01	3.9742e-06	5.4882e-05
1/64	4.0502e-03	3.0849e-01	1.8892e-06	7.1184e-06
1/128	9.5235e-06	1.7166e-01	1.3788e-06	9.0221e-07
1/256	2.2359e-06	9.1058e-02	8.0784e-07	1.1350e-07
1/512	4.3406e-06	4.5935e-02	4.1240e-06	1.3992e-08

Numerical example from Sect. 6.3

Table 7 Stochastic refinement for 8 spatial cells

Stochastic refinement, $h = 1/8$				
Samples	Error	$\mathcal{E}^{\text{det}}$	$\mathcal{E}_0^{\text{stoch}}$	$\mathcal{E}_0^{\text{det}}$
2	2.4296e-03	2.1913e-01	9.6970e-04	9.6957e-04
4	1.9059e-03	3.6195e-01	2.5221e-04	2.0300e-03
8	2.0140e-03	5.3951e-01	4.5084e-04	3.4629e-03
16	1.6421e-03	3.4298e-01	1.0539e-04	1.9946e-03
32	1.7357e-03	3.3839e-01	1.2722e-04	1.9963e-03
64	1.5387e-03	4.2101e-01	3.7605e-05	2.5636e-03
128	1.5659e-03	4.0292e-01	4.1701e-05	2.3854e-03
256	1.5662e-03	4.0802e-01	8.0689e-06	2.4552e-03
512	1.5389e-03	4.0178e-01	5.2234e-06	2.4118e-03
1024	1.5331e-03	3.9535e-01	1.2003e-06	2.3499e-03
2048	1.5394e-03	3.8443e-01	1.3741e-06	2.2819e-03

Numerical example from Sect. 6.3

Table 8 Stochastic refinement for 256 spatial cells

Stochastic refinement, $h = 1/256$				
Samples	Error	$\mathcal{E}^{\text{det}}$	$\mathcal{E}_0^{\text{stoch}}$	$\mathcal{E}_0^{\text{det}}$
2	5.5747e-04	9.7655e-02	5.5579e-04	1.2184e-07
4	1.1150e-03	4.0349e-02	1.1150e-03	4.9749e-08
8	1.1873e-03	1.4541e-01	1.1837e-03	1.8119e-07
16	8.9368e-05	1.0356e-01	8.7486e-05	1.2932e-07
32	8.1722e-05	1.0194e-01	7.9884e-05	1.2709e-07
64	1.9237e-05	9.3937e-02	1.7726e-05	1.1724e-07
128	9.8315e-06	9.5595e-02	8.2528e-06	1.1903e-07
256	1.5538e-05	9.5881e-02	1.3931e-05	1.1960e-07
512	2.0009e-05	9.9828e-02	1.8289e-05	1.2455e-07
1024	1.5185e-05	9.4760e-02	1.3624e-05	1.1810e-07
2048	2.3780e-06	9.2924e-02	8.8819e-07	1.1580e-07

Numerical example from Sect. 6.3

where $\xi \sim \mathcal{U}(0, 1)$ is a uniformly distributed random variable. The spatial domain is again $[0, 1]_{\text{per}}$. We compute the solution up to $T = 0.2$ and sample the initial measure $\bar{\mu}$ and the exact measure μ_T with 10000 samples.

6.3.1 Spatial refinement

Again we first examine the scaling properties of $\mathcal{E}^{\text{det}}(T)$ and $\mathcal{E}_0^{\text{det}}$ when we gradually increase the number of spatial cells. We start initially on a coarse mesh with 16 elements, i.e., $h = 1/16$. Strictly speaking the samples $\bar{\rho}_k$ are smooth but the meshes

under consideration are sufficiently coarse so that the numerics sees the k^{-2} decay of Fourier coefficients in $\bar{\rho}_k$ but the cut-off at $k = 600$ is not resolved. Thus, for the meshes under consideration, the decay of error is as if each sample $\{\bar{\rho}_k\}_{k \in \mathcal{K}}$ satisfies $\bar{\rho}_k \in H^s(D)$ for any $s < \frac{3}{2}$ but is not in $H^{3/2}(D)$. We expect the convergence of $\mathcal{E}_0^{\text{det}}$ in terms of the mesh width h to be of order 3, resp. $\frac{3}{2}$ after taking the square root. This behavior can be indeed observed in Fig. 3. Similarly, the numerical error converges with the order $\frac{3}{2}$, up to the point where it is dominated by the stochastic error. The deterministic residual $\mathcal{E}^{\text{det}}$ exhibits a rate of convergence of 1, resp. $\frac{1}{2}$ after taking the square root, which has also been observed in [11].

6.3.2 Stochastic refinement

Finally, we consider stochastic refinement for two fixed spatial meshes. The setup of the meshes and number of stochastic samples is exactly the same as in the numerical experiment from Sect. 6.2.2. In Fig. 3 we observe a similar behavior of the residuals and the numerical error as in Sect. 6.2.2, that is, the stochastic residual $\mathcal{E}_0^{\text{stoch}}$ converges independently of the spatial resolution and on the coarse grid the numerical error does not converge due to the low spatial resolution. On the fine grid, however, it converges with the same order as the stochastic resolution error, which is one, resp. one-half after taking the square-root (Fig. 4).

7 Conclusions

This work provides a first rigorous a posteriori error estimate for numerical approximations of dissipative statistical solutions in one spatial dimension. Our numerical approximations rely on so-called regularized empirical measures, which enable us to use the relative entropy method of Dafermos and DiPerna [8] within the framework of dissipative statistical solutions introduced by the authors of [17]. We derived a splitting of the error estimator into a stochastic and a spatio-temporal part. In addition, we provided numerical examples verifying this splitting. Moreover, our
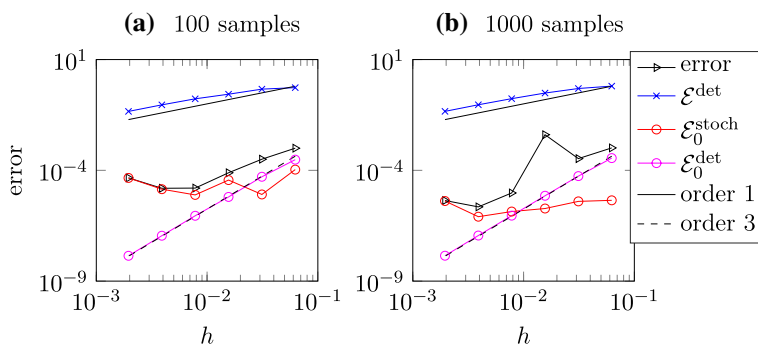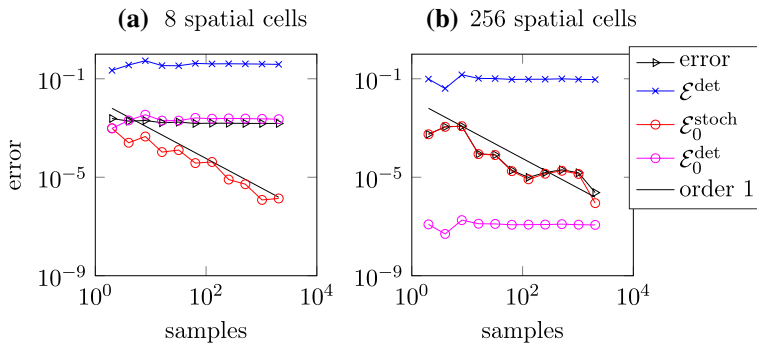

Fig. 3 Spatial refinement for 100 and 1000 samples. Numerical example from Sect. 6.3

Fig. 4 Stochastic refinement for 8 and 256 spatial cells. Numerical example from Sect. 6.3

numerical results confirm that the quantities that occur in our a posteriori error estimate decay with the expected order of convergence.

Further work should focus on a posteriori error control for multi-dimensional systems of hyperbolic conservation laws, especially the correct extension of the space-time reconstruction to two and three spatial dimensions. Moreover, based on the a posteriori error estimate it is possible to design reliable space-time-stochastic numerical schemes for the approximation of dissipative statistical solutions. For random conservation laws and classical $L^2(\Omega; L^2(D))$-estimates, the stochastic part of the error estimator is given by the discretization error in the initial data and an additional stochastic residual which occurs during the evolution, see for example [21, 30]. For dissipative statistical solutions the stochastic part of the error estimator in the 2-Wasserstein distance is directly related to the stochastic discretization error of the initial data which may be amplified in time due to nonlinear effects. This result shows that stochastic adaptivity for dissipative statistical solutions becomes significantly easier compared to random conservation laws since only stochastic discretization errors of the initial data (and their proliferation) need to be controlled. The design of space-stochastic adaptive numerical schemes based on this observation will be the subject of further research.

References

1. Adjerid, S., Devine, K.D., Flaherty, J.E., Krivodonova, L.: A posteriori error estimation for discontinuous Galerkin solutions of hyperbolic problems. Comput. Methods Appl. Mech. Eng. **191**, 1097–1112 (2002)
2. J. M. Ball, A version of the fundamental theorem for Young measures, in PDEs and continuum models of phase transitions (Nice: vol. 344 of Lecture Notes in Phys. Springer, Berlin **1989**, 207–215 (1988)
3. Bonneel, N., Van De Panne, M., Paris, S., Heidrich, W.: Displacement interpolation using lagrangian mass transport. In ACM Transactions on Graphics (TOG), vol. 30, ACM, p. 158 (2011)
4. Bressan, A.: The unique limit of the Glimm scheme. Arch. Rational Mech. Anal. **130**, 205–230 (1995)
5. Chiodaroli, E.: A counterexample to well-posedness of entropy solutions to the compressible Euler system. J. Hyperbolic Differ. Equ. **11**, 493–519 (2014)
6. Cockburn, B., Shu, C.-W.: The Runge-Kutta discontinuous Galerkin method for conservation laws. V. Multidimensional systems. J. Comput. Phys. **141**, 199–224 (1998)
7. Cockburn, B., Shu, C.-W.: Runge–Kutta discontinuous Galerkin methods for convection-dominated problems. J. Sci. Comput. **16**, 173–261 (2001)
8. Dafermos, C.M.: Hyperbolic conservation laws in continuum physics. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 325, 4th edn. Springer-Verlag, Berlin (2016)
9. De Lellis, C., Székelyhidi, L., Jr.: On admissibility criteria for weak solutions of the Euler equations. Arch. Ration. Mech. Anal. **195**, 225–260 (2010)
10. Dedner, A., Giesselmann, J.: A posteriori analysis of fully discrete method of lines discontinuous Galerkin schemes for systems of conservation laws. SIAM J. Numer. Anal. **54**, 3523–3549 (2016)
11. Dedner, A., Giesselmann, J.: Residual error indicators for discontinuous Galerkin schemes for discontinuous solutions to systems of conservation laws. In: Theory numerics and applications of hyperbolic problems I, Aachen, Germany, August 2016, pp. 459–471. Springer, Cham (2018)
12. DiPerna, R.J.: Measure-valued solutions to conservation laws. Arch. Rational Mech. Anal. **88**, 223–270 (1985)
13. Feireisl, E., Lukáčová-Medvidová, M., Mizerová, H.: *K*-convergence as a new tool in numerical analysis, IMA J. Num. Anal. (2019)
14. Fjordholm, U.S., Käppeli, R., Mishra, S., Tadmor, E.: Construction of approximate entropy measure-valued solutions for hyperbolic systems of conservation laws. Found. Comput. Math. **17**, 763–827 (2017)
15. Fjordholm, U.S., Lanthaler, S., Mishra, S.: Statistical solutions of hyperbolic conservation laws: foundations. Arch. Ration. Mech. Anal. **226**, 809–849 (2017)
16. Fjordholm, U.S., Lye, K., Mishra, S.: Numerical approximation of statistical solutions of scalar conservation laws. SIAM J. Numer. Anal. **56**, 2989–3009 (2018)
17. Fjordholm, U.S., Lye, K., Mishra, S., Weber, F.: Statistical solutions of hyperbolic systems of conservation laws: numerical approximation. Math. Models Methods Appl. Sci. **30**, 539–609 (2020)
18. Flamary, R., Courty, N.: POT Python Optimal Transport library, GitHub: https://github.com/rflamary/POT, (2017)
19. Foias, C., Manley, O., Rosa, R., Temam, R.: Navier–Stokes Equations and Turbulence. Encyclopedia of Mathematics and its Applications, vol. 83. Cambridge University Press, Cambridge (2001)
20. Giesselmann, J., Makridakis, C., Pryer, T.: A posteriori analysis of discontinuous Galerkin schemes for systems of hyperbolic conservation laws. SIAM J. Numer. Anal. **53**, 1280–1303 (2015)
21. Giesselmann, J., Meyer, F., Rohde, C.: A posteriori error analysis and adaptive non-intrusive numerical schemes for systems of random conservation laws. BIT **60**, 619–649 (2020)
22. Glimm, J.: Solutions in the large for nonlinear hyperbolic systems of equations. Comm. Pure Appl. Math. **18**, 697–715 (1965)
23. Glimm, J., Sharp, D.H., Lim, H., Kaufman, R., Hu, W.: Euler equation existence, non-uniqueness and mesh converged statistics. Philos. Trans. Roy. Soc. A **373**, 20140282 (2015)

24. Hartmann, R., Houston, P.: Adaptive discontinuous Galerkin finite element methods for nonlinear hyperbolic conservation laws. SIAM J. Sci. Comput. **24**, 979–1004 (2002)
25. Hindenlang, F., Gassner, G.J., Altmann, C., Beck, A., Staudenmaier, M., Munz, C.-D.: Explicit discontinuous Galerkin methods for unsteady problems. Comput. Fluids **61**, 86–93 (2012)
26. Ketcheson, D.I.: Highly efficient strong stability-preserving Runge-Kutta methods with low-storage implementations. SIAM J. Sci. Comput. **30**, 2113–2136 (2008)
27. Laforest, M.: A posteriori error estimate for front-tracking: systems of conservation laws. SIAM J. Math. Anal. **35**, 1347–1370 (2004)
28. Laforest, M.: An a posteriori error estimate for Glimm's scheme, in Hyperbolic problems. Theory, numerics and applications. Proceedings of the 11th international conference on hyperbolic problems, Ecole Normale Supérieure, Lyon, France, July 17–21, (2006), pp. 643–651 Berlin: Springer (2008)
29. Makridakis, C.: Space and time reconstructions in a posteriori analysis of evolution problems. ESAIM Proc. **21**, 31–44 (2007)
30. Meyer, F., Rohde, C., Giesselmann, J.: A posteriori error analysis for random scalar conservation laws using the stochastic Galerkin method. IMA J. Numer. Anal. **40**, 1094–1121 (2020)
31. Peyré, G., Cuturi, M.: Computational optimal transport, foundations and trends in machine. Learning **11**, 355–607 (2019)
32. Shu, C.-W., Osher, S.: Efficient implementation of essentially nonoscillatory shock-capturing schemes. J. Comput. Phys. **77**, 439–471 (1988)