Convergence order of upwind type schemes for transport equations with discontinuous coefficients

François Delarue[*], Frédéric Lagoutière[†], Nicolas Vauchelet[‡]

January 19, 2017

Abstract

An analysis of the error of the upwind scheme for transport equation with discontinuous coefficients is provided. We consider here a velocity field that is bounded and one-sided Lipschitz continuous. In this framework, solutions are defined in the sense of measures along the lines of Poupaud and Rascle's work. We study the convergence order of the upwind scheme in the Wasserstein distances. More precisely, we prove that in this setting the convergence order is 1/2. We also show the optimality of this result. In the appendix, we show that this result also applies to other "diffusive" "first order" schemes and to a forward semi-Lagrangian scheme.

Keywords: upwind finite volume scheme, forward semi-Lagrangian scheme, convergence order, conservative transport equation, continuity equation, measure-valued solution.

2010 AMS subject classifications: 35D30, 35L65, 65M12, 65M15.

1 Introduction

This paper is devoted to the numerical analysis of an upwind scheme for the linear transport equation in conservative form (continuity equation) with discontinuous coefficients. In space dimension d, this equation reads

$$\partial_t \rho + \operatorname{div}\big(a\rho\big) = 0, \qquad t > 0, \quad x \in \mathbb{R}^d, \tag{1.1}$$

and is complemented with the initial condition $\rho(0,\cdot) = \rho^{ini}$.

We consider a rather weak regularity of the velocity, bounded and one-sided (right) Lipschitz continuous (OSL for short):

$$a \in L^\infty([0,+\infty); L^\infty(\mathbb{R}^d))^d$$

and there exists $\alpha \in L^1_{loc}([0,+\infty))$ such that

$$\langle a(t,x) - a(t,y), x - y\rangle \le \alpha(t)|x-y|^2, \tag{1.2}$$

where $\langle\cdot,\cdot\rangle$ stands for the Euclidean scalar product in $\mathbb{R}^d$. Note that, as a is assumed to be bounded, the right Lipschitz continuity coefficient $\alpha(t)$ is non-negative, for any $t > 0$.

[*]Laboratoire J.-A. Dieudonné, UMR CNRS 7351, Univ. Nice, Parc Valrose, 06108 Nice Cedex 02, France. Email: delarue@unice.fr

[†]Univ Lyon, UniversitÃ© Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 blvd. du 11 novembre 1918, F-69622 Villeurbanne cedex, France. Email: lagoutiere@math.univ-lyon1.fr

[‡]LAGA - UMR 7539 Institut Galilée Université Paris 13, 99, avenue Jean-Baptiste Clément 93430 Villetaneuse - France. Email: vauchelet@math.univ-paris13.fr

Since the velocity a is not assumed Lipschitz continuous, the definition of solutions with the characteristic curves is not straightforward. In [17], Filippov proposed a notion of solution which extends the classical one. Using this so-called Filippov flow, Poupaud and Rascle proposed in [26] a notion of solutions to the conservative linear transport equation (1.1). They are defined as $Z_{\#}\rho^{ini}$ where Z is the Filippov flow corresponding to the velocity a (note that a stability result of the flow has been proved recently in [4]). In dimension 1, these solutions are equivalent to the duality solutions defined in [8, 9]. Duality solutions have also been defined in higher dimensions in [10] but the theory is still not complete for the conservative transport equation.

In the present setting, solutions to the continuity equation can form Dirac masses and then are defined in the sense of measures. The numerical approximation of measure valued solutions to (1.1) requires a particular care. In dimension 1, Gosse & James, [18], proposed a class of finite difference numerical schemes that includes the one dimensional upwind scheme. Using the setting of duality solutions, the convergence of these schemes has been obtained in the sense of measures. However no error estimates are provided. More recently, Bouchut, Eymard & Prignet, in [7], have proposed a different strategy in any dimension with a finite volume scheme defined by the characteristics (the flow is assumed to be given). The convergence is proved, on general admissible meshes, in the sense of measures, but no error estimates are provided.

We here present an error analysis of an upwind scheme for the continuity equation (1.1) when the coefficient a is one-sided Lipschitz continuous. More precisely, we prove that the order of convergence of the scheme is $1/2$ in Wasserstein distances W_p.

The convergence order of the upwind scheme for transport equations has received a lot of attention. When the velocity field is Lipschitz continuous, this scheme is known to be first order convergent in the L^∞ norm for any smooth initial data in $\mathcal{C}^2(\mathbb{R}^d)$ *and for well-suited meshes*, provided a stability (Courant-Friedrichs-Lewy) condition holds: see [5]. However, for non-smooth initial data or on more general meshes, this order of convergence falls down to $1/2$, in L^p norms. This result has been first proved in the Cartesian framework by Kuznetsov in [20] (this analysis is actually done for the entropy solutions of scalar (nonlinear) hyperbolic equations). On quite general meshes, for $L^1(\mathbb{R}^d) \cap BV(\mathbb{R}^d)$ initial data, the result was tackled in [24] and [12], in quite similar settings but with very different methods, in the L^∞ in time and L^1 in space norm (we here will use the formalism developed in [12]). At last, let us mention that for initial data that are Lipschitz-continuous, the convergence to the $1/2 - \varepsilon$ order in the L^∞ norm (for any $\varepsilon > 0$) is proved in [23] and again in [12]. We also can mention [15] and [31] for related results. We emphasize that the techniques used in [23, 24] and [12] are totally different. In the former, the technique is based on entropy estimates, whereas in the latter, the proof relies on the construction of a stochastic characteristic defined as a Markov chain.

Up to our knowledge, there are no error estimates for the upwind scheme or, more generally, finite volume schemes, when the velocity field is less smooth. When the velocity field is given in $L^1((0,T);(W^{1,1}(\Omega))^d)$ for $\Omega \subset \mathbb{R}^d$, the *convergence* of numerical solutions, obtained thanks to an upwind scheme, towards renormalized solutions of the transport equation is studied in [11].

In this work we perform a numerical analysis of the upwind scheme in the weak framework where the velocity field is one-sided Lipschitz continuous. More precisely, our main result is the following.

Result (see Theorem 4.1 for a precise statement). Let ρ^{ini} be a probability measure on $\mathbb{R}^d$ such that $\int_{\mathbb{R}^d} |x|^p \rho^{ini}(dx) < \infty$ for some $p \geq 1$. We assume that the velocity field a belongs to $L^\infty([0,+\infty); L^\infty(\mathbb{R}^d))^d$ and satisfies the OSL condition (1.2). Let ρ be the solution of the transport equation (1.1) with initial datum ρ^{ini} (whose existence and uniqueness is recalled in Theorem 2.4 below). Let $\rho_{\Delta x}$ be the numerical approximation computed thanks to the upwind scheme on a Cartesian grid mesh with space step Δx. Then, under a (usual) Courant-Friedrichs-Lewy condition linking the time step Δt and the cell size Δx, there exists a constant $C \geq 0$, depending only on a,

p and ρ^{ini}, such that we have

$$\forall\, t \geq 0, \qquad W_p\big(\rho(t), \rho_{\Delta x}(t)\big) \leq C\left(\sqrt{\Delta x\left(t + \int_0^t \alpha(s)ds\right)} + \Delta x\right) \exp\left((1+\Delta t)\int_0^t \alpha(s)ds\right). \quad (1.3)$$

In this result, W_p denotes the Wasserstein distance of order $p \geq 1$, whose definition will be recalled below.

Remark 1.1 *A natural question is then: equipped with this result, and now assuming that the solution is, let us say, with bounded variation in space, do we recover a convergence order in a strong norm such as L^1 in space? The answer is yes, thanks to an interpolation estimate by Santambrogio ([28]), which proof is reported in the Appendix: let f, g be two non-negative functions in $L^1(\mathbb{R}^d)$ with mass equal to 1. There exists a constant $C \in \mathbb{R}$ such that*

$$||f-g||_{L^1} \leq C|f-g|_{BV}^{1/2} W_1(f,g)^{1/2}$$

where $|\cdot|_{BV}$ denotes the total variation semi-norm on $\mathbb{R}^d$. Thus, we recover a $1/4$ convergence order in L^1. This is not optimal: it is known that the convergence order is $1/2$, but, to reach $1/2$, one should use the additional smoothness of the solution in the W_1 estimate to obtain the convergence to the first order.

The main idea of the proof of the theorem is, in the spirit of [12], to show that, similar to the exact solution, the numerical solution can be interpreted as the pushforward of the initial condition by a (numerical) flow. However, this flow is stochastic whilst that associated with the original equation is obviously deterministic. The numerical (deterministic) solution is then represented as the *expectation* of the pushforward of the initial datum by the stochastic flow.

Finally, we emphasize that although our result is fully established for an upwind scheme, the approach developed in this work can be easily extended to other schemes, as it is explained in the appendix. Meanwhile, it is worth mentioning that, although our strategy is shown to work on any general mesh for the semi-Lagrangian scheme discussed in the appendix, it works on a Cartesian grid only for the upwind scheme under study. This is a major difference with [12], in which the analysis of the upwind scheme is performed on a general mesh. The rationale for this difference is as follows: In [12], the strategy for handling the upwind scheme on non-Cartesian grids relies on a time reversal argument and, somehow, on the analysis of the characteristics associated with the velocity field $-a$. Whilst there is no difficulty for doing so in the Lipschitz setting, this is of course much more challenging under the weaker OSL condition (1.2) since the ordinary differential equation driven by $-a$ is no more well-posed. We hope to address this question in future works.

The outline of the paper is the following. Section 2 is devoted to general definitions and notations that will be used throughout the paper (in particular, we recall the notion of measure solutions to the transport equation (1.1) as defined in [26]). In Section 3, we define the upwind scheme on a Cartesian mesh and provide some basic properties for this scheme. Section 4 is devoted to the statement and the proof of our main result: the convergence with order $1/2$ of the upwind scheme on a Cartesian grid. Finally, in order to illustrate the optimality of this order of convergence, we provide in Section 5 first an explicit computation of the error in a simple case, and then some numerical experiments in dimension 1. An appendix provides an extension to other numerical (similar) schemes and the proof of the lemma used in the preceding remark.

2 Measure solutions to the continuity equation

All along the paper, we will make use of the following notations. We denote by $\mathcal{M}_b(\mathbb{R}^d)$ the space of finite signed measures on $\mathbb{R}^d$ equipped with the Borel σ-field $\mathcal{B}(\mathbb{R}^d)$. For $\rho \in \mathcal{M}_b(\mathbb{R}^d)$, we denote by $|\rho|(\mathbb{R}^d)$ its total variation, or total mass. The space of measures $\mathcal{M}_b(\mathbb{R}^d)$ is endowed with the weak topology $\sigma(\mathcal{M}_b(\mathbb{R}^d), \mathcal{C}_0(\mathbb{R}^d))$, where $\mathcal{C}_0(\mathbb{R}^d)$ is the set of continuous functions on $\mathbb{R}^d$ that tend to 0 at ∞. We then define $\mathcal{S}_\mathcal{M} := \mathcal{C}([0, +\infty); \mathcal{M}_b(\mathbb{R}^d) - \sigma(\mathcal{M}_b, \mathcal{C}_0(\mathbb{R}^d)))$ and we equip it with the topology of uniform convergence on finite intervals of the form $[0, T]$, with $T > 0$.

For ρ a measure in $\mathcal{M}_b(\mathbb{R}^d)$ and Z a measurable map (throughout the paper, *measurability* is understood as measurability with respect to the Borel σ-fields when these latter are not specified), we denote by $Z_\#\rho$ the pushforward measure of ρ by Z; by definition, it satisfies

$$\int_{\mathbb{R}^d} \phi(x)\, Z_\#\rho(dx) = \int_{\mathbb{R}^d} \phi(Z(x))\, \rho(dx) \quad \text{for any } \phi \in \mathcal{C}_0(\mathbb{R}^d),$$

or, equivalently,

$$Z_\#\rho(A) = \rho\big(Z^{-1}(A)\big) \quad \text{for any } A \in \mathcal{B}(\mathbb{R}^d).$$

Moreover, we denote by $\mathcal{P}(\mathbb{R}^d)$ the subspace of $\mathcal{M}_b(\mathbb{R}^d)$ made of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Also, we let $\mathcal{P}_p(\mathbb{R}^d)$ the space of probability measures with finite p-th order moment, $p \geq 1$:

$$\mathcal{P}_p(\mathbb{R}^d) = \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^p \mu(dx) < \infty \right\}.$$

Finally, for any probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and any integrable random variable X from $(\Omega, \mathcal{A})$ into $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we denote by $\mathbb{E}(X)$ the expectation of X.

2.1 Wasserstein distance

The space $\mathcal{P}_p(\mathbb{R}^d)$ is endowed with the Wasserstein distance W_p defined by (see e.g. [2, 32, 33])

$$W_p(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x|^p \, \gamma(dx, dy) \right\}^{1/p} \tag{2.4}$$

where $\Gamma(\mu, \nu)$ is the set of measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν, i.e.

$$\Gamma(\mu, \nu) = \Big\{ \gamma \in \mathcal{P}_p(\mathbb{R}^d \times \mathbb{R}^d);\ \forall\, \xi \in \mathcal{C}_0(\mathbb{R}^d),\ \int_{\mathbb{R}^d \times \mathbb{R}^d} \xi(y_1) \gamma(dy_1, dy_2) = \int_{\mathbb{R}^d} \xi(y_1) \mu(dy_1),$$
$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \xi(y_2) \gamma(dy_1, dy_2) = \int_{\mathbb{R}^d} \xi(y_2) \nu(dy_2) \Big\}.$$

It is known that in the definition of W_p the infimum is actually a minimum (see [32, 29]). A measure that fulfils the minimum in the definition (2.4) of W_p is called an *optimal plan*. The set of optimal plans is denoted by $\Gamma_0(\mu, \nu)$. Thus for all $\gamma_0 \in \Gamma_0(\mu, \nu)$, we have

$$W_p(\mu, \nu)^p = \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x|^p \, \gamma_0(dx, dy).$$

We will make use of the following properties of the Wasserstein distance. Given two measurable maps $X, Y : \mathbb{R}^d \to \mathbb{R}^d$, we have the inequality

$$W_p(X_\#\mu, Y_\#\mu) \leq \|X - Y\|_{L^p(\mu)}. \tag{2.5}$$

Indeed, $\pi = (X, Y)_\#\mu \in \Gamma(X_\#\mu, Y_\#\mu)$ and $\int_{\mathbb{R}^d} |x - y|^p \, \pi(dx, dy) = \|X - Y\|^p_{L^p(\mu)}$.

More generally, for any probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and any two random variables $X, Y : \Omega \to \mathbb{R}^d$ such that $\mathbb{E}[|X|^p]$ and $\mathbb{E}[|Y|^p]$ are finite, we have

$$\big(W_p(X_\#\mathbb{P}, Y_\#\mathbb{P})\big)^p \leq \mathbb{E}\big(|X - Y|^p\big), \tag{2.6}$$

where $X_\#\mathbb{P}$ and $Y_\#\mathbb{P}$ denote the respective distributions of X and Y under $\mathbb{P}$. The proof follows from the same argument as above: $\pi = (X, Y)_\#\mathbb{P} \in \Gamma(X_\#\mathbb{P}, Y_\#\mathbb{P})$ and $\int_{\mathbb{R}^d} |x - y|^p \, \pi(dx, dy) = \mathbb{E}[|X - Y|^p]$.

2.2 Weak measure solutions for linear conservation laws

We recall in this section some useful results on weak measure solutions to the conservative transport equation (1.1), when driven by an initial datum $\rho(0,\cdot) = \rho^{ini} \in \mathcal{M}_b(\mathbb{R}^d)$ and a vector field a that satisfies the OSL condition.

We start by the following definition of characteristics [17]:

Definition 2.1 *Let us assume that $a : [0,+\infty) \times \mathbb{R}^d \ni (t,x) \mapsto a(t,x) \in \mathbb{R}^d$ is a (measurable) vector field. A Filippov characteristic $Z(\cdot; s, x)$ stemmed from $x \in \mathbb{R}^d$ at time $s \geq 0$ is a continuous function $[s,+\infty) \ni t \mapsto Z(t;s,x) \in \mathbb{R}^d$ such that $Z(s;s,x) = x$, $\frac{\partial}{\partial t} Z(t;s,x)$ exists for a.e. $t \geq s$ and*

$$\frac{\partial}{\partial t} Z(t;s,x) \in \big\{\mathrm{Convess}(a)(t,\cdot)\big\}\big(Z(t;s,x)\big) \quad \textit{for a.e. } t \geq s.$$

From now on, we will use the notation $Z(t,x) = Z(t;0,x)$.

In this definition, $\mathrm{Convess}(E)$ denotes the essential convex hull of the set E: let us remind briefly the definition for the sake of completeness (see [17, 3] for more details). We denote by $\mathrm{Conv}(E)$ the classical convex hull of E, i.e., the smallest closed convex set containing E. Given the vector field $a(t,\cdot) : \mathbb{R}^d \to \mathbb{R}^d$, its essential convex hull at point x is defined as

$$\big\{\mathrm{Convess}(a)(t,\cdot)\big\}(x) = \bigcap_{r>0} \bigcap_{N \in \mathcal{N}_0} \mathrm{Conv}\big[a\big(t, B(x,r) \setminus N\big)\big],$$

where $\mathcal{N}_0$ is the set of zero Lebesgue measure sets. Then, we have the following existence and uniqueness result of Filippov characteristics under the assumption that the vector field a is one-sided Lipschitz continuous.

Theorem 2.2 ([17]) *Let $a \in L^1_{loc}([0,+\infty); L^\infty(\mathbb{R}^d))^d$ be a vector field satisfying the OSL condition (1.2). Then there exists a unique Filippov flow Z associated with this vector field, meaning that there exists a unique characteristic for any initial condition $(s,x) \in [0,+\infty) \times \mathbb{R}^d$. This flow does not depend on the choice of the representative (up to a $dt \otimes dx$ null set) of the velocity field a as long as this version satisfies the OSL condition pointwise. Moreover, we have the semi-group property: For any $t, \tau, s \in [0,+\infty)$ such that $t \geq \tau \geq s$ and $x \in \mathbb{R}^d$,*

$$Z(t;s,x) = Z(\tau;s,x) + \int_\tau^t a(\sigma, Z(\sigma;s,x))\, d\sigma.$$

Importantly, we have the following Lipschitz continuous estimate on the Filippov characteristic:

Lemma 2.3 ([17]) *Let $a \in L^1([0,+\infty), L^\infty(\mathbb{R}^d))^d$ satisfy the OSL condition (1.2) and Z be the associated flow. Then, for all $t \geq s$ in $[0,+\infty)$, we have*

$$L_Z(t;s) := \sup_{x,y \in \mathbb{R}^d, x \neq y} \frac{|Z(t;s,x) - Z(t;s,y)|}{|x-y|} \leq e^{\int_s^t \alpha(\sigma)\, d\sigma} \tag{2.7}$$

($L_Z(t;s)$ is the Lipschitz constant of the flow Z).

Proof. For $x, y \in \mathbb{R}^d$, we compute

$$\frac{d}{dt}|Z(t;s,x) - Z(t;s,y)|^2 = 2\big\langle a(t, Z(t;s,x)) - a(t, Z(t;s,y)), Z(t;s,x) - Z(t;s,y)\big\rangle.$$

Using the OSL estimate (1.2), we deduce

$$\frac{d}{dt}|Z(t;s,x) - Z(t;s,y)|^2 \leq 2\alpha(s)|Z(t;s,x) - Z(t;s,y)|^2.$$

Thanks to a Grönwall lemma, we get

$$|Z(t;s,x) - Z(t;s,y)|^2 \leq e^{\int_s^t 2\alpha(\sigma)\,d\sigma}|x-y|^2,$$

which completes the proof. □

An important consequence of this result is the existence and uniqueness of weak measure solutions for the conservative linear transport equation. This has been obtained by Poupaud & Rascle in [26].

Theorem 2.4 ([26]) *Let $a \in L^1_{loc}([0,+\infty), L^\infty(\mathbb{R}^d))^d$ be a vector field satisfying the OSL condition* (1.2)*. Then, for any $\rho^{ini} \in \mathcal{M}_b(\mathbb{R}^d)$, there exists a unique measure solution ρ in $\mathcal{S}_\mathcal{M}$ to the conservative transport equation* (1.1) *with initial datum $\rho(0,\cdot) = \rho^{ini}$ such that $\rho(t) = Z(t)_\#\rho^{ini}$, where Z is the unique Filippov flow, i.e. for any $\phi \in \mathcal{C}_0(\mathbb{R}^d)$, we have*

$$\int_{\mathbb{R}^d} \phi(x)\rho(t,dx) = \int_{\mathbb{R}^d} \phi(Z(t,x))\rho^{ini}(dx), \qquad \textit{for} \quad t \geq 0.$$

Note that actually, in this result, the expression $\rho(t) = Z(t)_\#\rho^{ini}$ is somehow understood as a *definition* of solution to the Cauchy problem. From now on, we will interpret the solutions to (1.1) in this sense.

To conclude this section, we recall the stability estimate of the flow due to Bianchini and Gloyer [4, Theorem 1.1]. This estimate reads as a bound for the difference between the flows Z_1 and Z_2 associated with two velocity fields a_1 and a_2 in $L^1([0,\infty), L^\infty(\mathbb{R}^d))^d$ satisfying the OSL condition (1.2). For any $r > 0$ any $x \in B(0,r)$, it holds that

$$|Z_1(t;s,x) - Z_2(t;s,x)|^2 \leq C\int_s^t \|a_1(\sigma,.) - a_2(\sigma,.)\|^{1/d}_{L^1(B(0,2R))}\,d\sigma, \tag{2.8}$$

where $R = r + a_\infty T$, and $a_\infty = \max\{\|a_1\|_\infty, \|a_2\|_\infty\}$ and C is a constant that only depends on the dimension.

Remark that this estimate, which is also proved in the same paper to be optimal, is a bad hint to obtain a result as the one we will prove here, because the stability of the characteristics with respect to perturbations of the velocity field decreases as the space dimension increases. Based on this estimate, one could imagine that a similar phenomenon should occur when estimating the error of a numerical scheme for (1.1). Indeed (as it will be the case in the next section), the analysis of the scheme should consist in regarding the numerical solution as the solution of an equation of the same type as (1.1) but driven by an approximating velocity field. Then, it would be tempting to compare both solutions by means of (2.8). However, our result shows that this strategy is non-optimal, at least for the scheme studied in the paper. Our analysis exploits the fact that, in our case, the structure of the approximating velocity is actually very close to that of the original velocity field.

3 Definition of the scheme and basic properties

3.1 Numerical discretization

From now on, we consider a velocity field $a \in L^\infty([0,+\infty), L^\infty(\mathbb{R}^d))^d$ and we choose a representative $\widehat{a}$ in the equivalence class of a in $L^\infty([0,+\infty), L^\infty(\mathbb{R}^d))^d$: $\widehat{a}$ is defined everywhere and is jointly measurable in time and space; it is $dt \otimes dx$ a.e. equal to a; and, it satisfies the condition (1.2) everywhere. In order to simplify the presentation, we will keep the notation a instead of $\widehat{a}$ and we write $a = (a_1, \ldots, a_d)$.

We denote by $\Delta t > 0$ the time step and consider a Cartesian grid with step $\Delta x_i > 0$ in the ith direction, $i = 1, \ldots, d$, and $\Delta x = \max_i \Delta x_i$. For $i = 1, \ldots, d$, we note e_i the ith vector of the canonical basis of $\mathbb{R}^d$. We define the multi-indices $J = (J_1, \ldots, J_d) \in \mathbb{Z}^d$, the space cells $C_J = [(J_1 - \frac{1}{2})\Delta x_1, (J_1 + \frac{1}{2})\Delta x_1) \times \ldots [(J_d - \frac{1}{2})\Delta x_d, (J_d + \frac{1}{2})\Delta x_d)$ and their center $x_J = (J_1 \Delta x_1, \ldots, J_d \Delta x_d)$. Finally, we set $t^n = n\Delta t$.

For a given non-negative measure $\rho^{ini} \in \mathcal{P}(\mathbb{R}^d)$, we define for $J \in \mathbb{Z}^d$,

$$\rho_J^0 = \int_{C_J} \rho^{ini}(dx) \geq 0, \tag{3.9}$$

which actually is to be understood as $\rho_J^0 = \rho^{ini}(C_J)$. Since ρ^{ini} is a probability measure, the total mass of the system is $\sum_{J \in \mathbb{Z}^d} \rho_J^0 = 1$. We denote by ρ_J^n an approximation of the value $\rho(t^n)(C_J)$, for $J \in \mathbb{Z}^d$, and we propose to compute this approximation by using an upwind-typed scheme (see for example [21, 22] for general considerations on schemes for transport equations): more precisely, we let by induction

$$\rho_J^{n+1} = \rho_J^n - \sum_{i=1}^{d} \frac{\Delta t}{\Delta x_i} \big(({a_i}_J^n)^+ \rho_J^n - ({a_i}_{J+e_i}^n)^- \rho_{J+e_i}^n - ({a_i}_{J-e_i}^n)^+ \rho_{J-e_i}^n + ({a_i}_J^n)^- \rho_J^n \big), \qquad n \in \mathbb{N}, J \in \mathbb{Z}^d. \tag{3.10}$$

The notation $(a)^+ = \max\{0, a\}$ stands for the positive part of the real number a and $(a)^- = \max\{0, -a\}$ for the negative part. Remark that this scheme can also be viewed as an Engquist-Osher-typed scheme (although this scheme was developped in an homogeneous in space non-linear frame, in [16]). The numerical velocity is defined, for $i = 1, \ldots, d$, by

$$ {a_i}_J^n = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} a_i(s, x_J)\, ds. \tag{3.11}$$

Remark 3.1 *There is at least one other traditional upwind scheme, for which the velocity is to be computed at the interface: $a^n_{i+1/2}$. The difficulty with this one is explained at the end of Section 1 in the appendix. The question of the convergence of this other upwind scheme remains an open problem when $d > 1$).*

Remark 3.2 *The discretization of the velocity requires the computation of the mean value in time of the velocity field in formula* (3.11)*. However, if one assumes the velocity field to be Lipschitz continuous in time, uniformly in space, then ${a_i}_J^n$ can be replaced by $a_i(t^n, x_J)$. We refer to Remark 4.7 below for a short account on the new form of the main estimate* (1.3)*.*

Remark 3.3 *In dimension 1, Scheme* (3.10) *reads*

$$\rho_j^{n+1} = \rho_j^n - \frac{\Delta t}{\Delta x} \Big((a_j^n)^+ \rho_j^n - (a_{j+1}^n)^- \rho_{j+1}^n - (a_{j-1}^n)^+ \rho_{j-1}^n + (a_j^n)^- \rho_j^n \Big).$$

We will make use of the following interpretation of this scheme. Defining $\rho_{\Delta x}^n = \sum_{j \in \mathbb{Z}} \rho_j^n \delta_{x_j}$, we construct the approximation at time t^{n+1} with the following two steps:

- *The Dirac mass ρ_j^n, located at position x_j, moves with velocity a_j^n to the position $x_j + a_j^n\Delta t$. Assuming a Courant-Friedrichs-Lewy condition $||a||_\infty \Delta t \leq \Delta x$, the point $x_j + a_j^n \Delta t$ belongs to the interval $[x_j, x_{j+1}]$ if $a_j^n \geq 0$, or to the interval $[x_{j-1}, x_j]$ if $a_j^n \leq 0$.*
- *Then we split the mass ρ_j^n between x_j and x_{j+1} if $a_j^n \geq 0$ or between x_{j-1} and x_j if $a_j^n \leq 0$. We use a linear splitting rule: say whenever $a_j^n \geq 0$, the mass $\rho_j^n \times a_j^n \Delta t/\Delta x$ is sent to grid point x_{j+1} whereas $\rho_j^n \times (1 - a_j^n\Delta t/\Delta x)$ is sent to grid point x_j. We let the reader verify that this gives the scheme defined above.*

3.2 Properties of the scheme

Throughout the analysis, a_∞ stands for $||a||_{L^\infty(\mathbb{R}\times\mathbb{R}^d)}$ whenever $a \in L^\infty([0,+\infty), L^\infty(\mathbb{R}^d))^d$. We will assume further that $\Delta x \leq 1$.

The following lemma states a Courant-Friedrichs-Lewy-like (CFL) condition ensuring that the scheme preserves nonnegativity:

Lemma 3.4 *Let $a \in L^\infty([0,+\infty), L^\infty(\mathbb{R}^d))^d$ and let $(\rho_J^0)_{J\in\mathbb{Z}^d}$ be defined by* (3.9) *for some $\rho^{ini} \in \mathcal{P}_p(\mathbb{R}^d)$, $p \geq 1$. Assume further that the following CFL condition holds:*

$$a_\infty \sum_{i=1}^d \frac{\Delta t}{\Delta x_i} \leq 1. \tag{3.12}$$

Then the sequence $(\rho_J^n)_{n,J}$ computed thanks to the scheme defined in (3.10)–(3.11) *is non-negative: for all $J \in \mathbb{Z}^d$ and $n \in \mathbb{N}$, $\rho_J^n \geq 0$.*

Proof.
We can rewrite equation (3.10) as

$$\rho_J^{n+1} = \rho_J^n \left[1 - \sum_{i=1}^d \frac{\Delta t}{\Delta x_i}|a_{iJ}^{\,n}|\right] + \sum_{i=1}^d \rho_{J+e_i}^n \frac{\Delta t}{\Delta x_i}(a_{iJ+e_i}^{\,n})^- + \sum_{i=1}^d \rho_{J-e_i}^n \frac{\Delta t}{\Delta x_i}(a_{iJ-e_i}^{\,n})^+. \tag{3.13}$$

From definition (3.11), we have $|a_{iJ}^{\,n}| \leq a_\infty$ for $i = 1, \ldots, d$. Thus assuming Condition (3.12), we deduce that in (3.13) all the coefficients in front of ρ_J^n, $\rho_{J-e_i}^n$ and $\rho_{J+e_i}^n$, $i = 1, \ldots, d$, are non-negative. By a straightforward induction argument, as $\rho_J^0 \geq 0$ for all $J \in \mathbb{Z}^d$, $\rho_J^{n+1} \geq 0$ for all $J \in \mathbb{Z}^d$. □

In the next lemma, we collect some useful properties of the scheme, among which mass conservation and finiteness of the pth order moment:

Lemma 3.5 *Let $a \in L^\infty([0,+\infty), L^\infty(\mathbb{R}^d))^d$ and let $(\rho_J^0)_{J\in\mathbb{Z}^d}$ be defined by* (3.9) *for some $\rho^{ini} \in \mathcal{P}_p(\mathbb{R}^d)$, $p \geq 1$. Let us assume that the CFL condition* (3.12) *is satisfied. Then the sequence $(\rho_J^n)_{n\in\mathbb{N},J\in\mathbb{Z}^d}$ given by the numerical scheme* (3.10)–(3.11) *satisfies:*
(i) Conservation of the mass: for all $n \in \mathbb{N}^*$, we have

$$\sum_{J\in\mathbb{Z}^d} \rho_J^n = \sum_{J\in\mathbb{Z}^d} \rho_J^0 = 1.$$

(ii) Bound on the pth moment: there exists a constant $C_p > 0$, only depending on a_∞, the dimension d and the exponent p, such that, for all $n \in \mathbb{N}$, we have

$$M_1^n := \sum_{J\in\mathbb{Z}^d} |x_J|\rho_J^n \leq M_1^0 + C_1 t^n, \qquad \textit{when } p = 1,$$
$$M_p^n := \sum_{J\in\mathbb{Z}^d} |x_J|^p\rho_J^n \leq e^{C_p t^n}\big(M_p^0 + C_p\big), \qquad \textit{when } p > 1,$$

where we recall that $t^n = n\Delta t$.

(iii) Support: let us define $\lambda_i = \Delta t/\Delta x_i$. If ρ^{ini} has a bounded support then the numerical approximation at finite time T has a bounded support too. More precisely, assuming that $\rho^0_{\Delta x}$ is compactly supported in $B(0,R)$, then for any $T \geq 0$, for any integer $n \leq T/\Delta t$, we have

$$\rho^n_J = 0 \quad \textit{for any } J \in \mathbb{Z}^d \textit{ such that } x_J \notin B\Big(0, R + \frac{T}{\min_{i=1,\dots,d}\{\lambda_i\}}\Big).$$

Proof. According to Lemma 3.4, the weights $((\rho^n_J)_{J\in\mathbb{Z}^d})_{n\in\mathbb{N}}$ are non-negative.

(i) The mass conservation is directly obtained by summing equation (3.10) over J.

(ii) Let $p \geq 1$. By a discrete integration by parts on (3.10), we get

$$\begin{aligned}\sum_{J\in\mathbb{Z}^d} |x_J|^p \rho^{n+1}_J = \quad & \sum_{J\in\mathbb{Z}^d} |x_J|^p \rho^n_J - \sum_{i=1}^d \frac{\Delta t}{\Delta x_i} \sum_{J\in\mathbb{Z}^d} (a_{iJ}^n)^+ \big(|x_J|^p - |x_{J+e_i}|^p\big)\rho^n_J \\ & + \sum_{i=1}^d \frac{\Delta t}{\Delta x_i} \sum_{J\in\mathbb{Z}^d} (a_{iJ}^n)^- \big(|x_J|^p - |x_{J-e_i}|^p\big)\rho^n_J.\end{aligned} \tag{3.14}$$

Letting $c_p = p2^{p-1}$, one has

$$\Big| |x|^p - |x \pm \Delta x_i e_i|^p \Big| \leq c_p \Big(\Delta x_i |x|^{p-1} + \Delta x_i^p \Big). \tag{3.15}$$

Indeed, thanks to the convexity of $x \mapsto |x|^p$, we have

$$\begin{cases} |x|^p \geq |x \pm \Delta x_i e_i|^p \mp p|x \pm \Delta x_i e_i|^{p-2} \langle x \pm \Delta x_i e_i, \Delta x_i e_i\rangle, \\ |x \pm \Delta x_i e_i|^p \geq |x|^p \pm p|x|^{p-2}\langle x, \Delta x_i e_i\rangle. \end{cases} \tag{3.16}$$

Above, $|x \pm \Delta x_i e_i|^{p-2}\langle x \pm \Delta x_i e_i, \Delta x_i e_i\rangle$ is understood as 0 when $x \pm \Delta x_i e_i = 0$, and similarly for $|x|^{p-2}\langle x, \Delta x_i e_i\rangle$.

Now, the first line in (3.16), yields

$$|x \pm \Delta x_i e_i|^p - |x|^p \leq p\Delta x_i |x \pm \Delta x_i e_i|^{p-1} \leq p2^{p-1} \left(\Delta x_i |x|^{p-1} + \Delta x_i^p\right)$$

(actually true with 2^{p-2} instead of 2^{p-1}), whilst the second line gives

$$|x \pm \Delta x_i e_i|^p - |x|^p \geq -p|x|^{p-1}\Delta x_i \geq -p2^{p-1}\left(\Delta x_i |x|^{p-1} + \Delta x_i^p\right).$$

We easily get (3.15).

Then, using inequality (3.15) together with the mass conservation and the fact that $\Delta x \leq 1$, we deduce from (3.14):

$$\begin{aligned}\sum_{J\in\mathbb{Z}^d} |x_J|^p \rho^{n+1}_J \quad & \leq \sum_{J\in\mathbb{Z}^d} |x_J|^p \rho^n_J + c_p \Delta t \sum_{i=1}^d \sum_{J\in\mathbb{Z}^d} \Big(|x_J|^{p-1} + \Delta x_i^{p-1}\Big) \rho^n_J \, |a_{iJ}^n| \\ & \leq \sum_{J\in\mathbb{Z}^d} |x_J|^p \rho^n_J + c_p d \Delta t a_\infty \Big(\sum_{J\in\mathbb{Z}^d} |x_J|^{p-1} \rho^n_J + 1\Big),\end{aligned}$$

which may be rewritten as

$$M^{n+1}_p \leq M^n_p + c_p d\Delta t a_\infty \big(M^n_{p-1} + 1\big),$$

where, by mass conservation, $M_0^n = 1$. For $p = 1$, we get the result with $C_1 = 2c_p da_\infty = 2da_\infty$ by a straightforward induction. For $p > 1$, thanks to Hölder's inequality, $M_{p-1}^n \leq (M_p^n)^{(p-1)/p} \leq (M_p^n + 1)^{(p-1)/p} \leq M_p^n + 1$ and, we get

$$\begin{aligned} M_p^{n+1} &\leq M_p^n + c_p d \Delta t a_\infty \big(M_p^n + 2 \big) \\ &\leq (1 + C_p \Delta t) M_p^n + C_p \Delta t \end{aligned}$$

with $C_p = 2c_p da_\infty$. We conclude the proof using a discrete Grönwall lemma.

(iii) By definition of the numerical scheme (3.10), we clearly have that the support increases of only one cell in each direction at each time step. Therefore, after n iterations, the support has increased of less than

$$n \times \max_{i=1,\dots,d} \{\Delta x_i\} = \frac{n\, \Delta t}{\min_{i=1\dots,d}\{\lambda_i\}} \leq \frac{T}{\min_{i=1,\dots,d}\{\lambda_i\}},$$

provided that $T \geq n\Delta t$. □

3.3 Probabilistic interpretation

Following the idea in [12], we associate random characteristics with the here above upwind scheme 3.10. The construction of these characteristics is based upon the trajectories of a Markov chain admitting $\mathbb{Z}^d$ as state space. Here the random characteristics will be *forward* characteristics whilst they are *backward* in [12]. The rationale for considering forward characteristics lies in the fact that the equation is conservative; subsequently, the expression of the solution provided by Theorem 2.4 is based upon a forward flow. On the opposite, the equation considered in [12] is of the non-conservative form $\partial_t \rho + a \cdot \nabla \rho = 0$ and the expression of the solution involves backward characteristics.

Throughout the analysis, we will denote by $\Omega = (\mathbb{Z}^d)^{\mathbb{N}}$ the canonical space for the Markov chain. The canonical process is denoted by $(K^n)_{n\in\mathbb{N}}$ (namely K^n maps $\omega = (\omega^n)_{n\in\mathbb{N}} \in \Omega$ onto the n^{th} coordinate ω^n of ω): K^n must be understood as the n^{th} site occupied by a random process taking values in $\mathbb{Z}^d$. Notice that we here adopt a non-standard notation for the time index as we put it in superscript instead of subscript; although it does not fit the common habit, we feel it more consistent with the notation used above for defining the numerical scheme.

We equip Ω with the standard Kolmogorov σ-field $\mathcal{A}$ generated by sets of the form $\prod_{n\in\mathbb{N}} A^n$, with $A^n \subset \mathbb{Z}^d$ for all $n \in \mathbb{N}$ and, for some integer $n_0 \geq 0$, $A^n = \mathbb{Z}^d$ for $n \geq n_0$. In other words, $\mathcal{A}$ is the smallest σ-field that renders each K^n, $n \in \mathbb{N}$, measurable. Indeed, for any integer $n_0 \geq 0$ and any subsets $A^0, \dots, A^{n_0} \subset \mathbb{Z}^d$, the pre-image $(K^0, \dots, K^{n_0})^{-1}(A^0 \times \cdots \times A^{n_0})$ is precisely the cylinder $A^0 \times \cdots \times A^{n_0} \times \prod_{n>n_0} \mathbb{Z}^d$. The canonical filtration generated by $(K^n)_{n\in\mathbb{N}}$ is denoted by $\mathbb{F} = (\mathcal{F}^n = \sigma(K^0, \cdots, K^n))_{n\in\mathbb{N}}$. For each $n \geq 0$, $\mathcal{F}^n$ is the sub-σ-field of $\mathcal{A}$ containing events of the form $A^{(n)} \times \prod_{k>n} \mathbb{Z}^d$, with $A^{(n)} \subset (\mathbb{Z}^d)^{n+1}$. Informally, $\mathcal{F}^n$ stands for the information that an observer would collect by observing the random characteristic up until time n (or equivalently the realizations of $K^0, \dots, K^n$).

We then endow the pair $(\Omega, \mathcal{A})$ with a collection of probability measures $(\mathbb{P}_\mu)_{\mu\in\mathcal{P}(\mathbb{Z}^d)}$, $\mathcal{P}(\mathbb{Z}^d)$ denoting the set of probability measures on $\mathbb{Z}^d$, such that, for all $\mu \in \mathcal{P}(\mathbb{Z}^d)$, $(K^n)_{n\in\mathbb{N}}$ is a time-inhomogeneous Markov chain under $\mathbb{P}_\mu$ with initial law μ, namely $K^0{}_\#\mathbb{P}_\mu = \mu$ (*i.e.* $\mathbb{P}_\mu(K_0 = J) = \mu(J)$, which means that the initial starting cell is picked at random according to the law μ;

sometimes, we will also write $K^0 \sim \mu$), and with transition matrix at time $n \geq 0$:

$$P^n_{J,L} = \begin{cases} 1 - \sum_{i=1}^{d} \frac{\Delta t}{\Delta x_i} |a_i{}^n_J| & \text{when } L = J, \\ \frac{\Delta t}{\Delta x_i} (a_i{}^n_J)^+ & \text{when } L = J + e_i, \quad \text{for } i = 1, \dots, d, \\ \frac{\Delta t}{\Delta x_i} (a_i{}^n_J)^- & \text{when } L = J - e_i, \quad \text{for } i = 1, \dots, d, \\ 0 & \text{otherwise,} \end{cases} \tag{3.17}$$

where we used (3.11) under the assumption that $a \in L^\infty([0, +\infty), L^\infty(\mathbb{R}^d))^d$ satisfies the CFL condition (3.12) (we assume it to be in force throughout the section).

For any $\mu \in \mathcal{P}(\mathbb{Z}^d)$, we write $\mathbb{E}_\mu$ for the expectation under $\mathbb{P}_\mu$. Also, for any $\mu \in \mathcal{P}(\mathbb{Z}^d)$ and any $n \in \mathbb{N}$, the conditional probability $\mathbb{P}_\mu(\cdot|\mathcal{F}^n)$ is just denoted by $\mathbb{P}^n_\mu(\cdot)$; similarly, the conditional expectation $\mathbb{E}_\mu(\cdot|\mathcal{F}^n)$ is denoted by $\mathbb{E}^n_\mu(\cdot)$. Moreover, in statements that are true independently of the initial distribution $\mu \in \mathcal{P}(\mathbb{Z}^d)$, we often drop the index μ in the symbols $\mathbb{P}_\mu$ and/or $\mathbb{E}_\mu$. For instance, we may write:

$$\forall n \in \mathbb{N}, \ \forall L \in \mathbb{Z}^d, \quad \mathbb{P}^n\big(K^{n+1} = L\big) = P^n_{K^n, L}.$$

Whenever μ is the Dirac mass at some $J \in \mathbb{Z}^d$, namely $\mu = \delta_J$, we write $\mathbb{P}_J$ instead of $\mathbb{P}_\mu$ and similarly for $\mathbb{E}$. Notice that, for any $\mu \in \mathcal{P}(\mathbb{Z}^d)$, $\mathbb{P}_\mu$ is entirely determined by μ and the collection $(\mathbb{P}_J)_{J \in \mathbb{Z}^d}$:

$$\mathbb{P}_\mu(\cdot) = \sum_{J \in \mathbb{Z}^d} \mu_J \mathbb{P}_J(\cdot).$$

In most cases, it thus suffices to restrict the analysis of the Markov chain to the cases when $\mu = \delta_J$, for $J \in \mathbb{Z}^d$.

The following lemma gives the connection between the sequence of weights $((\rho^n)_{J \in \mathbb{Z}^d})_{n \in \mathbb{N}}$ introduced in the previous section (defined by the upwind scheme) and the Markov chain with transition matrix P:

Lemma 3.6 *Given an initial distribution $\rho^0 = (\rho^0_J)_{J \in \mathbb{Z}^d} \in \mathcal{P}(\mathbb{Z}^d)$, define, for any $n \in \mathbb{N}$, $\rho^n = (\rho^n_J)_{J \in \mathbb{Z}^d}$ through the scheme* (3.10), *namely*

$$\begin{aligned} \rho^{n+1}_J &= \rho^n_J - \sum_{i=1}^{d} \frac{\Delta t}{\Delta x_i} \Big((a_i{}^n_J)^+ \rho^n_J - (a_i{}^n_{J+e_i})^- \rho^n_{J+e_i} - (a_i{}^n_{J-e_i})^+ \rho^n_{J-e_i} + (a_i{}^n_J)^- \rho^n_J \Big) \\ &= \rho^n_J P^n_{J,J} + \sum_{i=1}^{d} \big(\rho^n_{J-e_i} P^n_{J-e_i,J} + \rho^n_{J+e_i} P^n_{J+e_i,J} \big). \end{aligned}$$

Then, for any $n \in \mathbb{N}$, one has $\rho^n = K^n{}_\# \mathbb{P}_{\rho^0}$ (equivalently ρ^n is the law of K^n when the chain is initialized with ρ^0).

Proof. For any $\mu \in \mathcal{P}(\mathbb{Z}^d)$, we have

$$K^{n+1}{}_\# \mathbb{P}_\mu = \sum_{L \in \mathbb{Z}^d} \delta_L \mathbb{P}_\mu\big(K^{n+1} = L\big) = \sum_{L \in \mathbb{Z}^d} \sum_{J \in \mathbb{Z}^d} \delta_L \mathbb{P}_\mu\big(K^n = J\big) P^n_{J,L}.$$

Therefore,

$$K^{n+1}{}_{\#}\mathbb{P}_\mu = \sum_{L\in\mathbb{Z}^d} \delta_L \Big(\mathbb{P}_\mu\big(K^n = L\big) P^n_{L,L} \\ + \sum_{i=1}^d \big(\mathbb{P}_\mu\big(K^n = L - e_i\big) P^n_{L-e_i,L} + \mathbb{P}_\mu\big(K^n = L + e_i\big) P^n_{L+e_i,L} \big) \Big).$$

Choosing $\mu = \rho^0$, the result follows from a straightforward induction. $\square$

Now that we have associated a Markov chain with the weights involved in the definition of the upwind scheme, we can define, as announced, the corresponding random characteristics. A random characteristic consists of a sequence of random variables $(X^n)_{n\in\mathbb{N}}$ from $(\Omega, \mathcal{A})$ into $\mathbb{R}^d$:

$$\forall n \in \mathbb{N},\ \forall \omega \in \Omega, \quad X^n(\omega) = x_{K^n(\omega)}, \tag{3.18}$$

where we recall that $x_J = (J_1 \Delta x_1, \ldots, J_d \Delta x_d)$ whenever $J = (J_1, \ldots, J_d) \in \mathbb{Z}^d$.

Proposition 3.7 *Let X^n be the random variable defined by* (3.18) *through the Markov chain admitting P in* (3.17) *as transition matrix.*

(i) For all $J \in \mathbb{Z}^d$, we have, with probability one under $\mathbb{P}_J$,

$$\mathbb{E}^n_J(X^{n+1} - X^n) = \int_{t^n}^{t^{n+1}} a(s, X^n)\, ds. \tag{3.19}$$

(ii) Defining $\rho^n_{\Delta x} = \sum_{J\in\mathbb{Z}^d} \rho^n_J \delta_{x_J}$, we have $\rho^n_{\Delta x} = X^n{}_{\#}\mathbb{P}_{\rho^0_{\Delta x}}$.

Proof. (i) For $J \in \mathbb{Z}^d$, we compute the conditional expectation given the trajectory of the Markov chain up until time n:

$$\mathbb{E}^n_J(X^{n+1} - X^n) = \sum_{i=1}^d \left(\Delta x_i e_i (a^n_{iK^n})^+ \frac{\Delta t}{\Delta x_i} - \Delta x_i e_i (a^n_{iK^n})^- \frac{\Delta t}{\Delta x_i} \right).$$

We deduce (3.19) by using definition (3.11).

(ii) For any $\mu \in \mathcal{P}(\mathbb{Z}^d)$, we have, for all $n \in \mathbb{N}$,

$$X^n{}_{\#}\mathbb{P}_\mu = \sum_{L\in\mathbb{Z}^d} \delta_{x_L} \mathbb{P}_\mu\big(K^n = L\big) = \sum_{L\in\mathbb{Z}^d} \delta_{x_L} \big(K^n{}_{\#}\mathbb{P}_\mu\big)_L.$$

The claim follows from Lemma 3.6. $\square$

4 Order of convergence

This section is devoted to the proof of the main result of our paper, that is the 1/2 order of convergence of the numerical approximation constructed by the upwind scheme (3.10). The precise statement of the result is:

Theorem 4.1 *Let $\rho^{ini} \in \mathcal{P}_p(\mathbb{R}^d)$, for some $p \geq 1$. Let us assume that $a \in L^\infty([0,\infty), L^\infty(\mathbb{R}^d))^d$ and satisfies the one-sided Lipschitz continuity condition* (1.2)*. Let $\rho = Z_\# \rho^{ini}$ be the unique measure solution in the sense of Poupaud and Rascle to the conservative transport equation* (1.1) *with initial datum ρ^{ini} given by Theorem 2.4. Let us define*

$$\rho^n_{\Delta x} = \sum_{J\in\mathbb{Z}^d} \rho^n_J \delta_{x_J},$$

where the approximation sequence $((\rho^n_J)_{J\in\mathbb{Z}^d})_{n\in\mathbb{N}}$ is computed thanks to the scheme (3.9)*–*(3.10)*–*(3.11)*. We assume that the CFL condition* (3.12) *holds. Then, there exists a non-negative constant C, depending upon p, ρ^{ini} and a_∞ only, such that, for all $n \in \mathbb{N}^*$,*

$$W_p\big(\rho(t^n), \rho^n_{\Delta x}\big) \leq C\, e^{(1+\Delta t)\int_0^{t^n} \alpha(s)\, ds} \left[\sqrt{\Delta x\left(t^n + \int_0^{t^n} \alpha(s) ds\right)} + \Delta x\right].$$

The proof of this theorem is postponed to Section 4.3. We need first to establish some useful estimates on the distance between the Filippov characteristics generated by a one sided Lipschitz continuous velocity field and the approximated characteristics.

4.1 Approximation of the flow

In the following lemma, we provide an estimate for the distance between the exact Filippov flow and an approximating flow computed through an explicit Euler discretization:

Lemma 4.2 *Let $a \in L^\infty([0,\infty), L^\infty(\mathbb{R}^d))^d$ satisfying Condition* (1.2)*. Let us consider Z the Filippov flow associated to the velocity field a and define by induction*

$$Y^{n+1} = Y^n + \int_{t^n}^{t^{n+1}} a(s, Y^n)\, ds,$$

with the initial condition $Y^0 = Z(0)$. There exists a universal constant C such that, for all $n \in \mathbb{N}$,

$$|Y^n - Z(t^n)| \leq C a_\infty e^{(1+\Delta t)\int_0^{t^n} \alpha(s)\, ds} \sqrt{\Delta t\left(t^n + (1+\Delta t)\int_0^{t^n} \alpha(s)\, ds\right)}.$$

Remark 4.3 *The order $1/2$ in this estimate may not be optimal, but it is sufficient for our purpose.*

Proof. For a given value of $n \in \mathbb{N}$, let us define, for $t \in [t^n, t^{n+1}]$, $Y(t) = Y^n + \int_{t^n}^t a(s, Y^n)\, ds$. By definition of the characteristics, we have, for any $t \in [t^n, t^{n+1}]$,

$$|Y(t) - Z(t)|^2 = \left| Y^n - Z(t^n) + \int_{t^n}^t \big(a(s, Y^n) - a(s, Z(s))\big)\, ds \right|^2.$$

Expanding the right hand side, we get

$$\begin{aligned} |Y(t) - Z(t)|^2 \leq |Y^n - Z(t^n)|^2 &+ 2\int_{t^n}^t \big\langle Y^n - Z(s), a(s, Y^n) - a(s, Z(s))\big\rangle\, ds \\ &+ 2\int_{t^n}^t \big\langle Z(s) - Z(t^n), a(s, Y^n) - a(s, Z(s))\big\rangle\, ds + (2a_\infty \Delta t)^2. \end{aligned} \tag{4.20}$$

Using condition (1.2), we deduce

$$|Y(t)-Z(t)|^2 \leq \ |Y^n - Z(t^n)|^2 + 2\int_{t^n}^t \alpha(s)|Y^n - Z(s)|^2\, ds \\ +2\int_{t^n}^t \big\langle Z(s)-Z(t^n), a(s,Y^n)-a(s,Z(s))\big\rangle\, ds + (2a_\infty \Delta t)^2.$$

Moreover, since the field a is bounded, we have $|Z(s)-Z(t^n)| \leq a_\infty |s-t^n|$ and so $\int_{t^n}^t |Z(s)-Z(t^n)|ds \leq a_\infty (t-t^n)^2/2$. Thus,

$$|Y(t)-Z(t)|^2 \leq |Y^n - Z(t^n)|^2 + 2\int_{t^n}^t \alpha(s)|Y^n - Z(s)|^2\, ds + 6a_\infty^2 \Delta t^2,$$

and, as we also have $|Y(s)-Y^n| \leq a_\infty|s-t^n|$, we get, for any $\varepsilon > 0$,

$$\begin{aligned}|Y(t)-Z(t)|^2 &\leq |Y^n - Z(t^n)|^2 \\ &+ 2(1+\varepsilon)\int_{t^n}^t \alpha(s)|Y(s)-Z(s)|^2\, ds + 6a_\infty^2\Delta t^2 + 2(1+\frac{1}{\varepsilon})a_\infty^2 \Delta t^2 \int_{t^n}^t \alpha(s)\, ds,\end{aligned} \tag{4.21}$$

where we used the standard Young inequality $(|Y(s)-Z(s)|+|Y^n - Y(s)|)^2 \leq (1+\varepsilon)|Y(s)-Z(s)|^2 + (1+1/\varepsilon)|Y^n - Y(s)|^2$.

Thanks to a continuous Grönwall lemma, the two characteristics thus satisfy

$$|Y^{n+1} - Z(t^{n+1})|^2 \leq \left[|Y^n - Z(t^n)|^2 + 2a_\infty^2\Delta t^2\left(3 + (1+\frac{1}{\varepsilon})\int_{t^n}^{t^{n+1}} \alpha(s)\, ds\right)\right] e^{2(1+\varepsilon)\int_{t^n}^{t^{n+1}} \alpha(s)\, ds}.$$

As $Y^0 = Z(0)$, a discrete Grönwall lemma leads to

$$|Y^n - Z(t^n)|^2 \leq 2a_\infty^2 \Delta t^2 \left(3n + (1+\frac{1}{\varepsilon})\int_0^{t^n} \alpha(s)\, ds\right) e^{2(1+\varepsilon)\int_0^{t^n}\alpha(s)\, ds}.$$

By taking $\varepsilon = \Delta t$, we deduce,

$$|Y^n - Z(t^n)|^2 \leq 2a_\infty^2 \Delta t \left(3t^n + (\Delta t + 1)\int_0^{t^n} \alpha(s)\, ds\right) e^{2(1+\Delta t)\int_0^{t^n}\alpha(s)\, ds}.$$

This completes the proof. □

Remark 4.4 *Whenever a is L-Lipschitz in time (uniformly in space) and $a(s,Y^n)$ is replaced by $a(t^n,Y^n)$ in the recursive definition of the sequence $(Y^n)_{n\in\mathbb{N}}$ in the statement of Lemma 4.2, there is an additional term in (4.20), coming from the time discretization of the velocity field. This term has the form:*

$$2\int_{t^n}^t \Big\langle Y^n - Z(s), a(t^n,Y^n) - a(s,Y^n)\Big\rangle ds, \tag{4.22}$$

which is less than $2L\Delta t \int_{t^n}^t |Y^n - Z(s)|ds$. By Young's inequality, we get

$$2\int_{t^n}^t \Big\langle Y^n - Z(s), a(s,Y^n) - a(t^n,Y^n)\Big\rangle ds \leq L\Delta t \int_{t^n}^t |Y^n - Z(s)|^2 ds + L\Delta t^2.$$

This gives a similar inequality to (4.21) but with α replaced by $\alpha + L\Delta t$ and a_∞ replaced by $a_\infty + L$. The corresponding version of Lemma 4.2 is easily derived.

Alternatively, we may perform all the above computations with respect to the OSL constant of $a(t^n,\cdot)$ instead of $a(s,\cdot)$. Instead of (4.22)*, we then focus on*

$$2\int_{t^n}^{t}\Big\langle Y^n - Z(s), a(t^n, Z(s)) - a(s, Z(s))\Big\rangle ds.$$

Then we obtain the same conclusion, with α replaced by $\alpha + L\Delta t$ and a_∞ replaced by $a_\infty + L$, but also the integral $\int_0^{t_n}\alpha(s)ds$ in the statement has to be replaced by the Riemann sum $\Delta t\sum_{k=0}^{n-1}\alpha(t^k)$.

4.2 Distance between the Euler scheme and the random characteristics

Lemma 4.5 *Under the CFL condition* (3.12)*, consider the random characteristics $(X^n)_{n\in\mathbb{N}}$ defined in* (3.18)*. Then, for any initial condition $\mu \in \mathcal{P}(\mathbb{Z}^d)$, it holds that, with probability 1 under $\mathbb{P}_\mu$, for all $n\in\mathbb{N}$,*

$$X^{n+1} = X^n + \int_{t^n}^{t^{n+1}} a(s, X^n)\, ds + h^n, \tag{4.23}$$

where h^n is an $\mathcal{F}^{n+1}$-measurable $\mathbb{R}^d$-valued random variable that satisfies

$$\mathbb{E}^n_\mu(h^n) = 0\,;\quad |h^n| \le 2\Delta x\,;\quad \forall p\ge 1,\quad \mathbb{E}^n_\mu\big(|h^n|^p\big) \le 2^p d a_\infty \Delta t \Delta x^{p-1}. \tag{4.24}$$

In particular, if we define iteratively the following sequence of random variables $(\hat{Y}^n)_{n\in\mathbb{N}}$ (constructed on the space $(\Omega,\mathcal{A})$ that supports the random characteristics):

$$\hat{Y}^{n+1} = \hat{Y}^n + \int_{t^n}^{t^{n+1}} a(s, \hat{Y}^n)\, ds,$$

with the (random) initial datum $\hat{Y}^0 = X^0 = x_{K^0}$, then, provided that $\Delta x \le 1$, there exists, for any $p\ge 1$, a non-negative constant C_p, only depending on p, d and a_∞, such that

$$\forall n\in\mathbb{N},\quad \mathbb{E}_\mu\big(|X^n - \hat{Y}^n|^p\big)^{1/p} \le C_p\, e^{\int_0^{t^n}\alpha(s)\,ds}\big(\sqrt{t^n\Delta x} + \Delta x\big). \tag{4.25}$$

Proof. The expansion (4.23), with h^n satisfying $\mathbb{E}^n_\mu(h^n) = 0$ for each $n\in\mathbb{N}$, is a direct consequence of *(i)* in Proposition 3.7. By construction, we can write $h^n = X^{n+1} - X^n - \mathbb{E}^n_\mu(X^{n+1} - X^n)$. Since $|X^{n+1} - X^n| \le \Delta x$, we deduce that $|h^n| \le 2\Delta x$. Moreover, for any $p \ge 1$,

$$\mathbb{E}^n_\mu\big(|h^n|^p\big) \le 2^{p-1}\Big[\mathbb{E}^n_\mu\big(|X^{n+1} - X^n|^p\big) + \mathbb{E}^n_\mu\big(|\mathbb{E}^n_\mu(X^{n+1} - X^n)|^p\big)\Big] \le 2^p\mathbb{E}^n_\mu\big(|X^{n+1} - X^n|^p\big).$$

Then, the bound for $\mathbb{E}^n_\mu(|h^n|^p)$ follows from the fact that:

$$\begin{aligned}\mathbb{E}^n_\mu\big(|X^{n+1} - X^n|^p\big) &= \mathbb{E}^n_\mu\big(|X^{n+1} - X^n|^p\mathbf{1}_{\{K^{n+1}\neq K^n\}}\big)\\ &\le \sum_{i=1}^d \Delta x_i^p \mathbb{P}^n_\mu\big(K_i^{n+1}\neq K_i^n\big) \le \sum_{i=1}^d \Delta x_i^p\big(P^n_{K^n,K^n+e_i} + P^n_{K^n,K^n-e_i}\big) \le d a_\infty \Delta x^{p-1}\Delta t.\end{aligned}$$

We split the proof of the second claim (4.25), into two steps. In the first step, we will estimate $\mathbb{E}(|X^n - \hat{Y}^n|^p)^{1/p}$, for $p\in[1,2]$. The second step is devoted to the analysis of $\mathbb{E}(|X^n - \hat{Y}^n|^p)^{1/p}$ when $p>2$. This step is rather more technical. Indeed, the case $p=2$ is more natural because it corresponds to the use of the scalar product, involved in the one-sided Lipschitz continuity.

First step. From definition (4.23), we obtain, after an obvious expansion,

$$\begin{aligned}|X^{n+1}-\hat{Y}^{n+1}|^2 \le |X^n-\hat{Y}^n|^2 + 2\int_{t^n}^{t^{n+1}} \Big\langle X^n-\hat{Y}^n, a(s,X^n)-a(s,\hat{Y}^n)\Big\rangle\, ds \\ + 2\langle X^n-\hat{Y}^n, h^n\rangle + |h^n|^2 + 4a_\infty\big(a_\infty\Delta t^2+\Delta t\Delta x\big),\end{aligned} \tag{4.26}$$

using the fact that a is bounded and that $|h^n| \le 2\Delta x$, see (4.24).

Using the CFL condition (3.12) in order to bound $a_\infty\Delta t^2$ by $\Delta t\Delta x/d$, we get

$$\begin{aligned}|X^{n+1}-\hat{Y}^{n+1}|^2 \le |X^n-\hat{Y}^n|^2 + 2\int_{t^n}^{t^{n+1}} \Big\langle X^n-\hat{Y}^n, a(s,X^n)-a(s,\hat{Y}^n)\Big\rangle\, ds \\ + 2\langle X^n-\hat{Y}^n, h^n\rangle + |h^n|^2 + Ca_\infty\Delta t\Delta x,\end{aligned}$$

for a constant C that only depends on d and whose value is allowed to increase from line to line. Since a satisfies the one-sided Lipschitz continuity condition (1.2), we get that, with probability 1 under $\mathbb{P}_\mu$, for all $n \in \mathbb{N}$,

$$\begin{aligned}&|X^{n+1}-\hat{Y}^{n+1}|^2 \\ &\le \left(1+2\int_{t^n}^{t^{n+1}} \alpha(s)\, ds\right)|X^n-\hat{Y}^n|^2 + 2\langle X^n-\hat{Y}^n, h^n\rangle + |h^n|^2 + Ca_\infty\Delta t\Delta x.\end{aligned} \tag{4.27}$$

Recalling that $\mathbb{E}^n_\mu(h^n)=0$ and noticing that $X^n-\hat{Y}^n$ is $\mathcal{F}^n$ is measurable, we have

$$\mathbb{E}^n_\mu\big(\langle X^n-\hat{Y}^n, h^n\rangle\big) = \big\langle X^n-\hat{Y}^n, \mathbb{E}^n_\mu(h^n)\big\rangle = 0. \tag{4.28}$$

Now taking the conditional expectation $\mathbb{E}^n_\mu$ in (4.27) and recalling from the preliminary step of the proof that $\mathbb{E}^n_\mu(|h^n|^2) \le Ca_\infty\Delta t\Delta x$ (with $C=4d$), we obtain

$$\mathbb{E}^n_\mu\big(|X^{n+1}-\hat{Y}^{n+1}|^2\big) \le \left(1+2\int_{t^n}^{t^{n+1}} \alpha(s)\, ds\right)|X^n-\hat{Y}^n|^2 + Ca_\infty\Delta t\Delta x,$$

for a new value of C.

Taking the expectation $\mathbb{E}_\mu$ (using the projective property of the conditional expectation $\mathbb{E}_\mu[\cdot] = \mathbb{E}_\mu[\mathbb{E}^n_\mu(\cdot)]$), applying a discrete version of Grönwall's lemma, and using also the fact that the initial datum verify $X^0 = \hat{Y}^0$, we deduce

$$\forall n\in\mathbb{N}, \quad \mathbb{E}_\mu\big(|X^n-\hat{Y}^n|^2\big) \le Ca_\infty e^{2\int_0^{t^n}\alpha(s)\,ds} n\Delta t\Delta x. \tag{4.29}$$

Therefore, for $p\in[1,2]$, we have, thanks to Hölder's inequality,

$$\Big(\mathbb{E}_\mu\big(|X^n-\hat{Y}^n|^p\big)\Big)^{1/p} \le \Big(\mathbb{E}_\mu\big(|X^n-\hat{Y}^n|^2\big)\Big)^{1/2} \le e^{\int_0^{t^n}\alpha(s)\,ds}\sqrt{Ca_\infty t^n\Delta x}, \tag{4.30}$$

which concludes the proof when $p\in[1,2]$.

Second step. In order to handle the case $p\ge 2$, we use an induction. We assume that, for some $p\in\mathbb{N}\setminus\{0,1\}$, there exists a constant c, only depending on p, d and a_∞, such that, for all $1\le m\le 2(p-1)$, for all $n\in\mathbb{N}$,

$$\mathbb{E}_\mu(|X^n-\hat{Y}^n|^m)^{1/m} \le ce^{\int_0^{t^n}\alpha(s)\,ds}\Big(\sqrt{t^n\Delta x}+\Delta x\Big), \tag{4.31}$$

which is obviously true when $p=2$ thanks to (4.30). From (4.27), we get

$$|X^{n+1}-\hat{Y}^{n+1}|^{2p} \leq \Big(e^{2\int_{t^n}^{t^{n+1}} \alpha(s)\,ds} |X^n-\hat{Y}^n|^2 + 2\langle X^n-\hat{Y}^n, h^n\rangle + |h^n|^2 + Ca_\infty \Delta t \Delta x \Big)^p .$$

Expanding the right hand side, we obtain

$$\begin{aligned} |X^{n+1}-\hat{Y}^{n+1}|^{2p} &\leq e^{2p\int_{t^n}^{t^{n+1}} \alpha(s)\,ds} |X^n-\hat{Y}^n|^{2p} \\ &\quad + p e^{2(p-1)\int_{t^n}^{t^{n+1}} \alpha(s)\,ds} |X^n-\hat{Y}^n|^{2(p-1)} \Big(2\langle X^n-\hat{Y}^n, h^n\rangle + |h^n|^2 + Ca_\infty \Delta t \Delta x \Big) \\ &\quad + \sum_{k=2}^{p} \binom{p}{k} e^{2(p-k)\int_{t^n}^{t^{n+1}} \alpha(s)\,ds} |X^n-\hat{Y}^n|^{2(p-k)} \Big(2\langle X^n-\hat{Y}^n, h^n\rangle + |h^n|^2 + Ca_\infty \Delta t \Delta x \Big)^k . \end{aligned} \tag{4.32}$$

By the same token as in (4.28), notice that (the constant C being allowed to increase from line to line as long as it only depends on d)

$$\begin{aligned} &\mathbb{E}_\mu^n \Big[|X^n-\hat{Y}^n|^{2(p-1)} \Big(2\langle X^n-\hat{Y}^n, h^n\rangle + |h^n|^2 + Ca_\infty \Delta t \Delta x \Big) \Big] \\ &\quad = \mathbb{E}_\mu^n \Big[|X^n-\hat{Y}^n|^{2(p-1)} \Big(Ca_\infty \Delta t \Delta x + |h^n|^2 \Big) \Big] \leq Ca_\infty \Delta t \Delta x \, |X^n-\hat{Y}^n|^{2(p-1)}, \end{aligned}$$

where we used the last estimate of (4.24) for the last inequality.

We proceed in a similar way with the last term in (4.32). Allowing the constant C to depend upon p, we have, for all $k \in \{2,\dots,p\}$,

$$\begin{aligned} &\mathbb{E}_\mu^n \Big[|X^n-\hat{Y}^n|^{2(p-k)} \Big(2\langle X^n-\hat{Y}^n, h^n\rangle + |h^n|^2 + Ca_\infty \Delta t \Delta x \Big)^k \Big] \\ &\quad \leq C\mathbb{E}_\mu^n \Big[|X^n-\hat{Y}^n|^{2p-k} |h^n|^k + |X^n-\hat{Y}^n|^{2(p-k)} |h^n|^{2k} \Big] + Ca_\infty^k \Delta t^k \Delta x^k |X^n-\hat{Y}^n|^{2(p-k)} \\ &\quad \leq C\Delta t \Delta x^{k-1} |X^n-\hat{Y}^n|^{2p-k} + C\Delta t \Delta x^{2k-1} |X^n-\hat{Y}^n|^{2(p-k)}, \end{aligned}$$

where, once again, we used (4.24) together with the CFL condition to pass from the second to the third line. In the last line, we allowed C to depend on d and p, but also on a_∞.

Returning to (4.32) and taking the expectation therein (using the fact that $\mathbb{E}_\mu[\cdot] = \mathbb{E}_\mu[\mathbb{E}_\mu^n(\cdot)]$), we finally get that:

$$\begin{aligned} \mathbb{E}_\mu\big[|X^{n+1}-\hat{Y}^{n+1}|^{2p}\big] &\leq e^{2p\int_{t^n}^{t^{n+1}} \alpha(s)\,ds} \mathbb{E}_\mu\big[|X^n-\hat{Y}^n|^{2p}\big] \\ &\quad + Ce^{2(p-1)\int_{t^n}^{t^{n+1}} \alpha(s)\,ds} \sum_{k=2}^{2p} \Big(\mathbb{E}_\mu\big[|X^n-\hat{Y}^n|^{2p-k}\big] \Delta t \Delta x^{k-1} \Big). \end{aligned}$$

Pay attention that sum above runs from 2 to $2p$ instead of 2 to p in the original inequality (4.32). Plugging the induction property (4.31), we get, for all $n \in \mathbb{N}$,

$$
\begin{aligned}
&\mathbb{E}_\mu\big[|X^{n+1}-\hat{Y}^{n+1}|^{2p}\big]\\
&\quad\le e^{2p\int_{t^n}^{t^{n+1}}\alpha(s)\,ds}\mathbb{E}_\mu\big[|X^n-\hat{Y}^n|^{2p}\big]+Ce^{2p\int_0^{t^{n+1}}\alpha(s)\,ds}\sum_{k=2}^{2p}\Delta t\Delta x^{k-1}\Big(\sqrt{t^n\Delta x}+\Delta x\Big)^{2p-k}\\
&\quad\le e^{2p\int_{t^n}^{t^{n+1}}\alpha(s)\,ds}\mathbb{E}_\mu\big[|X^n-\hat{Y}^n|^{2p}\big]+Ce^{2p\int_0^{t^{n+1}}\alpha(s)\,ds}\Delta t\bigg(\Delta x^{2p-1}+\sum_{k=2}^{2p}(t^n)^{p-k/2}\Delta x^{p+k/2-1}\bigg),
\end{aligned}
\tag{4.33}
$$

where we used the bound $(\sqrt{t^n\Delta x}+\Delta x)^{2p-k}\le C[(t^n)^{p-k/2}\Delta x^{p-k/2}+\Delta x^{2p-k}]$ in order to pass from the second to the third line. Notice now that

$$
\begin{aligned}
\sum_{k=2}^{2p}(t^n)^{p-k/2}\Delta x^{p+k/2-1}&=\Delta x^p\sum_{k=2}^{2p}\sqrt{t^n}^{2p-k}\sqrt{\Delta x}^{k-2}\\
&\le\Delta x^p\Big(\sqrt{t^n}+\sqrt{\Delta x}\Big)^{2p-2}\le C\Delta x^p\big((t^n)^{p-1}+\Delta x^{p-1}\big).
\end{aligned}
$$

Plugging into (4.33), we get

$$
\begin{aligned}
&\mathbb{E}_\mu\big[|X^{n+1}-\hat{Y}^{n+1}|^{2p}\big]\\
&\quad\le e^{2p\int_{t^n}^{t^{n+1}}\alpha(s)\,ds}\mathbb{E}_\mu\big[|X^n-\hat{Y}^n|^{2p}\big]+Ce^{2p\int_0^{t^{n+1}}\alpha(s)\,ds}\Delta t\Delta x^p\big((t^n)^{p-1}+\Delta x^{p-1}\big).
\end{aligned}
$$

Iterating over n and recalling that $X^0=\hat{Y}^0$ and $n\Delta t=t^n$, we deduce that, for all $n\in\mathbb{N}$,

$$
\mathbb{E}_\mu\big[|X^n-\hat{Y}^n|^{2p}\big]\le Ce^{2p\int_0^{t^n}\alpha(s)\,ds}t^n\Delta x^p\big((t^n)^{p-1}+\Delta x^{p-1}\big).
$$

We finally obtain, for all $n\in\mathbb{N}$,

$$
\mathbb{E}_\mu\big[|X^n-\hat{Y}^n|^{2p}\big]\le Ce^{2p\int_0^{t^n}\alpha(s)\,ds}\Delta x^p\big(t^n+\Delta x\big)^p,
$$

where we used Young's inequality to bound $t^n\Delta x^{2p-1}$ by $C((t^n)^p\Delta x^p+\Delta x^{2p})$. Then,

$$
\mathbb{E}_\mu\big[|X^n-\hat{Y}^n|^{2p}\big]^{1/2p}\le Ce^{\int_0^{t^n}\alpha(s)\,ds}\Big(\sqrt{t^n\Delta x}+\Delta x\Big),
$$

By Hölder's inequality, we conclude that (4.31) holds for $2(p-1)\le m\le 2p$. By induction, (4.31) is satisfied for all $p\in\mathbb{N}^*$.

Conclusion. Finally, from (4.30) and (4.31), we conclude the proof. $\square$

Remark 4.6 *In full analogy with Remark 4.4, we may discuss the case when a is L-Lipschitz in time (uniformly in space) and $a(s,\cdot)$ is replaced by $a(t^n,\cdot)$ in the recursive definitions of the sequences $(X^n)_{n\in\mathbb{N}}$ and $(\hat{Y}^n)_{n\in\mathbb{N}}$ in the statement of Lemma 4.2. Then, the final result is the same provided that the integral $\int_0^{t_n}\alpha(s)ds$ is replaced by $\Delta t\sum_{k=0}^{n-1}\alpha(t^k)$.*

4.3 Proof of Theorem 4.1

Let $\rho_{\Delta x}^n = \sum_{J\in\mathbb{Z}^d} \rho_J^n \delta_{x_J}$ be the measure associated to the numerical solution given by the scheme (3.9)–(3.10)–(3.11) at time t^n. Since $\rho^{ini} \in \mathcal{P}_p(\mathbb{R}^d)$, we first notice from Lemma 3.5 that $\rho_{\Delta x}^n \in \mathcal{P}_p(\mathbb{R}^d)$. By Proposition 3.7, we have $\rho_{\Delta x}^n = X^n{}_\#\mathbb{P}_{\rho_{\Delta x}^0}$ where X^n is defined in (3.18) and $\rho_{\Delta x}^0 = \sum_{J\in\mathbb{Z}^d} \rho_J^0 \delta_{x_J}$, with ρ^0 defined in (3.9). Let ρ be the exact solution of Theorem 2.4, $\rho(t) = Z(t)_\#\rho^{ini}$ where Z is the Filippov flow associated to the one-sided Lipschitz continuous velocity field a.

Consider now the two sequences $(Y^n)_{n\in\mathbb{N}}$ and $(\hat{Y}^n)_{n\in\mathbb{N}}$ respectively defined in Lemmas 4.2 and 4.5. Each Y^n is regarded as a mapping from $\mathbb{R}^d$ into itself: the initial condition Y^0 (also equal to $Z(0)$) in the statement of Lemma 4.2 is given by the identity mapping $\mathbb{R}^d \ni x \mapsto x \in \mathbb{R}^d$, that is $Y^0(x) = x$ for all $x \in \mathbb{R}^d$. We then call $Y^n(x)$ the value of Y^n in the statement of Lemma 4.2 when $Y^0(x) = x$. When $\mathbb{R}^d$ is equipped with the distribution $\rho_{\Delta x}^0$, the distribution of Y^n writes $Y^n{}_\#\rho_{\Delta x}^0$. In comparison with, each $\hat{Y}^n$ is a random variable from Ω to $\mathbb{R}^d$: when $\hat{Y}^0$ (also equal to X^0) has the distribution $\rho_{\Delta x}^0$, the distribution of $\hat{Y}^n$ writes $\hat{Y}^n{}_\#\mathbb{P}_{\rho^0}$. It is then crucial to observe that $\hat{Y}^n(\omega)$ may be regarded as $Y^n(X^0(\omega))$. In particular, if both Y^0 and $\hat{Y}^0$ have $\rho_{\Delta x}^0$ as common law (although the mappings are constructed on different spaces), then Y^n and $\hat{Y}^n$ also have the same distribution, namely $Y^n{}_\#\rho_{\Delta x}^0 = \hat{Y}^n{}_\#\mathbb{P}_{\rho^0}$.

As a consequence of the above discussion, we deduce from the triangle inequality:

$$\begin{aligned} W_p\big(\rho_{\Delta x}^n, \rho(t^n)\big) \leq\ & W_p\big(X^n{}_\#\mathbb{P}_{\rho^0}, \hat{Y}^n{}_\#\mathbb{P}_{\rho^0}\big) + W_p\big(Y^n{}_\#\rho_{\Delta x}^0, Z(t^n)_\#\rho_{\Delta x}^0\big) \\ & + W_p\big(Z(t^n)_\#\rho_{\Delta x}^0, Z(t^n)_\#\rho^{ini}\big). \end{aligned} \tag{4.34}$$

We will bound each term of the right hand side separately.

Initial datum. Let us first consider the last term in the right hand side of (4.34). We have

$$W_p(Z(t^n)_\#\rho_{\Delta x}^0, Z(t^n)_\#\rho^{ini}) \leq L_Z(t^n;0) W_p(\rho_{\Delta x}^0, \rho^{ini}).$$

Indeed, let π be an optimal map in $\Gamma_0(\rho_{\Delta x}^0, \rho^{ini})$, i.e.

$$W_p(\rho_{\Delta x}^0, \rho^{ini})^p = \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \pi(dx,dy).$$

Then $\gamma = (Z(t^n), Z(t^n))_\#\pi$ is a map with marginals $Z(t^n)_\#\rho_{\Delta x}^0$ and $Z(t^n)_\#\rho^{ini}$. It implies

$$W_p\big(Z(t^n)_\#\rho_{\Delta x}^0, Z(t^n)_\#\rho^{ini}\big)^p \leq \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \gamma(dx,dy) = \int_{\mathbb{R}^d\times\mathbb{R}^d} |Z(t^n,x) - Z(t^n,y)|^p \pi(dx,dy).$$

From Lemma 2.3, we know that the flow Z is Lipschitz continuous with Lipschitz constant $L_Z(t^n;0)$. Thus

$$W_p\big(Z(t^n)_\#\rho_{\Delta x}^0, Z(t^n)_\#\rho^{ini}\big) \leq L_Z(t^n;0) \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \pi(dx,dy)\right)^{1/p}.$$

Precisely, using inequality (2.7) in Lemma 2.3, we deduce

$$W_p\big(Z(t^n)_\#\rho_{\Delta x}^0, Z(t^n)_\#\rho^{ini}\big) \leq e^{\int_0^t \alpha(s)ds} W_p(\rho_{\Delta x}^0, \rho^{ini}). \tag{4.35}$$

Now, for $\rho^{ini} \in \mathcal{P}_p(\mathbb{R}^d)$, we recall the definition

$$\rho_{\Delta x}^0 = \sum_{J\in\mathbb{Z}^d} \rho_J^0 \delta_{x_J}, \qquad \text{with} \qquad \rho_J^0 = \int_{C_J} \rho^{ini}(dx) = \rho^{ini}(C_J). \tag{4.36}$$

Let us define $\tau : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ by $\tau(\sigma, x) = \sigma x_J + (1-\sigma)x$, for $x \in C_J$. We have that $\tau(0,\cdot) = \mathrm{id}$ and $\tau(1,\cdot)_\# \rho^0 = \rho^0_{\Delta x}$. Thus

$$\begin{aligned} W_p(\rho^0_{\Delta x}, \rho^{ini})^p &\leq \int_{\mathbb{R}^d} |x-y|^p \, (\mathrm{id} \times \tau(1,\cdot))_\# \rho^{ini}(dx, dy) \\ &\leq \sum_{J \in \mathbb{Z}^d} \int_{C_J} |x - x_J|^p \, \rho^{ini}(dx), \end{aligned} \tag{4.37}$$

where we use (4.36) for the last inequality. We deduce $W_p(\rho^0_{\Delta x}, \rho^{ini}) \leq \Delta x$. Injecting this latter inequality into (4.35), we obtain

$$W_p\big(Z(t^n)_\# \rho^0_{\Delta x}, Z(t^n)_\# \rho^{ini}\big) \leq e^{\int_0^{t^n} \alpha(s) ds} \Delta x. \tag{4.38}$$

Second term. For the second term of the right hand side of (4.34), by the standard property (2.5) of the Wasserstein distance, one has

$$W_p\big(Y^n{}_\# \rho^0_{\Delta x}, Z(t^n)_\# \rho^0_{\Delta x}\big) \leq \|Y^n - Z(t^n)\|_{L^p(\rho^0_{\Delta x})} \leq \sup_{J \in \mathbb{Z}^d} |Y^n(x_J) - Z(t^n, x_J)|.$$

Then applying Lemma 4.2, we deduce that there exists a non-negative constant C such that

$$\begin{aligned} W_p(Y^n{}_\# \rho^0_{\Delta x}, Z(t^n)_\# \rho^0_{\Delta x}) &\leq C a_\infty e^{(1+\Delta t) \int_0^{t^n} \alpha(s) ds} \sqrt{\Delta t \left(t^n + (1+\Delta t) \int_0^{t^n} \alpha(s)\, ds \right)} \\ &\leq C e^{(1+\Delta t) \int_0^{t^n} \alpha(s) ds} \sqrt{\Delta x \left(t^n + \int_0^{t^n} \alpha(s)\, ds \right)}, \end{aligned} \tag{4.39}$$

where we used again the CFL condition (3.12) and where C only depends on d and a_∞.

First term. We consider finally the first term in the right hand side of (4.34). By (2.6),

$$W_p\big(X^n{}_\# \mathbb{P}_{\rho^0}, \hat{Y}^n{}_\# \mathbb{P}_{\rho^0}\big) \leq \mathbb{E}_{\rho^0}\big[|X^n - \hat{Y}^n|^p\big]^{1/p}.$$

From Lemma 4.5, we deduce that there exists a constant C_p, only depending on p, d and a_∞, such that

$$W_p\big(X^n{}_\# \mathbb{P}_{\rho^0}, \hat{Y}^n{}_\# \mathbb{P}_{\rho^0}\big) \leq C_p \, e^{\int_0^{t^n} \alpha(s)\, ds} \left(\sqrt{t^n \Delta x} + \Delta x \right). \tag{4.40}$$

Conclusion. Injecting inequalities (4.38), (4.39) and (4.40) into (4.34) we deduce, for all $n \in \mathbb{N}^*$,

$$W_p\big(\rho^n_{\Delta x}, \rho(t^n)\big) \leq C_p \, e^{(1+\Delta t) \int_0^{t^n} \alpha(s) ds} \left(\sqrt{\Delta x \left(t^n + \int_0^{t^n} \alpha(s)\, ds \right)} + \Delta x \right),$$

for a new value of C_p.

Remark 4.7 *When a is L-Lipschitz in time (uniformly in space) and $a(s,\cdot)$ is replaced by $a(t^n,\cdot)$ in the definition of the upwind scheme* (3.11)*, Remarks 4.4 and 4.6 say that the final result holds true but with α replaced by $\alpha + L\Delta t$ and with $\int_0^{t_n} \alpha(s) ds$ replaced by the Riemann sum $\Delta t \sum_{k=0}^{n-1} \alpha(t^k)$.*

5 One-dimensional examples

In the aim to show the optimality of the present result, we perform both an exact computation of the error in a very simple case and provide some numerical simulations, in dimension 1. We first recall that, in the one-dimensional case, the expression of the Wasserstein distance simplifies. This simplification may be used for a numerical investigation, as it has been proposed in the pioneering work [19]. Indeed, any probability measure μ on the real line $\mathbb{R}$ can be described thanks to its cumulative distribution function $F(x) = \mu((-\infty, x])$, which is a right-continuous and non-decreasing function with $F(-\infty) = 0$ and $F(+\infty) = 1$. Then we can define the generalized inverse Q_μ of F (or monotone rearrangement of μ) by $Q_\mu(z) = F^{-1}(z) := \inf\{x \in \mathbb{R}/F(x) > z\}$; it is a right-continuous and non-decreasing function, defined on $[0, 1]$. For every non-negative Borel-measurable map $\xi : \mathbb{R} \to \mathbb{R}$, we have

$$\int_{\mathbb{R}} \xi(x)\mu(dx) = \int_0^1 \xi(Q_\mu(z))\, dz.$$

In particular, $\mu \in \mathcal{P}_p(\mathbb{R})$ if and only if $Q_\mu \in L^p((0, 1))$. Moreover, in the one-dimensional setting, there exists a unique optimal transport plan realizing the minimum in (2.4). More precisely, if μ and ν belong to $\mathcal{P}_p(\mathbb{R})$, with monotone rearrangements Q_μ and Q_ν, then $\Gamma_0(\mu, \nu) = \{(Q_\mu, Q_\nu)_\# \mathbb{L}_{(0,1)}\}$ where $\mathbb{L}_{(0,1)}$ is the restriction to $(0, 1)$ of the Lebesgue measure. Then we have the explicit expression of the Wasserstein distance (see [27, 33])

$$W_p(\mu, \nu) = \left(\int_0^1 |Q_\mu(z) - Q_\nu(z)|^p\, dz \right)^{1/p}, \tag{5.41}$$

and the map $\mu \mapsto Q_\mu$ is an isometry between $\mathcal{P}_p(\mathbb{R})$ and the convex subset of (essentially) non-decreasing functions of $L^p((0, 1))$.

We will take advantage of this expression (5.41) of the Wasserstein distance in dimension 1 in our numerical simulations to estimate the numerical error of the upwind scheme (3.10). This scheme in dimension 1 on a Cartesian mesh reads, with time step Δt and cell size Δx:

$$\rho_j^{n+1} = \rho_j^n - \frac{\Delta t}{\Delta x} \left((a_j^n)^+ \rho_j^n - (a_{j+1}^n)^- \rho_{j+1}^n - (a_{j-1}^n)^+ \rho_{j-1}^n + (a_j^n)^- \rho_j^n \right).$$

With this scheme, we define the probability measure $\rho_{\Delta x}^n = \sum_{j\in\mathbb{Z}} \rho_j^n \delta_{x_j}$. Then the generalized inverse of $\rho_{\Delta x}^n$, denoted by $Q_{\Delta x}$, is given by

$$Q_{\Delta x}(z) = x_{j+1}, \qquad \text{for } z \in \Big[\sum_{k\le j} \rho_k^n, \sum_{k\le j+1} \rho_k^n \Big).$$

5.1 Optimality of the convergence order when $a \equiv 1$ and $2\Delta t = \Delta x$

We here consider the very simple case where $a = 1$ and $\rho^{ini} = \delta_0$, thus $\rho_j^0 = \delta_{0j}$. In this case, the solution of the transport equation with velocity a and initial data ρ^{ini} is given by $\rho(t) = \delta_t$. For the sake of simplicity, we choose Δt and Δx such that $\Delta t/\Delta x = 1/2$. The numerical scheme thus simplifies to $\rho_j^{n+1} = \rho_j^n - 1/2(\rho_j^n - \rho_{j-1}^n)$, and it is a simple exercise to show that then the numerical solution is

$$\rho_j^n = \begin{cases} 0 \text{ if } j < 0, \\ \binom{n}{j}(1/2)^n \text{ if } 0 \le j \le n, \\ 0 \text{ if } j > n. \end{cases}$$

For any discrete time t^n, $n \in \mathbb{N}$, $W_1(\rho(t), \rho^n_{\Delta x})$ is the sum over j of the distance $|j\Delta x - n\Delta t|$ of the cell number j to the position of the Dirac mass of the exact solution, multiplied by the mass associated with this cell, ρ^n_j (formula (5.41)):

$$W_1(\rho(t^n), \rho^n_{\Delta x}) = \sum_{j=0}^{n} \rho^n_j |j\Delta x - n\Delta t| = \sum_{j=0}^{n} \binom{n}{j} (1/2)^n |j\Delta x - n\Delta t|$$

The right-hand side may be written as $\Delta x \times \mathbb{E}[|S_n - \mathbb{E}(S_n)|]$, where S_n is a binomial random variable with n as number of trials and $1/2$ as parameter of success. Recalling that the variance of S_n is $n/4$, we know from the central limit theorem that

$$\lim_{n\to\infty} \frac{2}{\sqrt{n}} \mathbb{E}[|S_n - \mathbb{E}(S_n)|] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |x| \exp\left(-\frac{x^2}{2}\right) dx = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{+\infty} x \exp\left(-\frac{x^2}{2}\right) dx = \frac{\sqrt{2}}{\sqrt{\pi}}. \tag{5.42}$$

Therefore,

$$W_1(\rho(t^n), \rho^n_{\Delta x}) \sim_{n\to\infty} \frac{1}{\sqrt{2\pi}} \sqrt{n\Delta x^2} = \frac{1}{\sqrt{2\pi}} \sqrt{2n\Delta t\Delta x} = \frac{1}{\sqrt{\pi}} \sqrt{t^n \Delta x}, \tag{5.43}$$

which proves the optimality of the one half order of convergence. Remark furthermore that, due to the linearity of both the equation and the scheme, this provides a direct proof of the convergence to the one half order of the scheme with any initial probability measure datum (when the velocity a is constant, and at least when $a\Delta t/\Delta x = 1/2$). Observe also that the distance between the left and right-hand sides in (5.43) can be explicitly bounded in terms of n by means of Berry-Esseen's theorem, which is a standard result in probability theory for estimating the rate of convergence in the central limit theorem for sums of random variables with a finite moment of order 3 (which is obviously the case here). In our framework, Berry-Esseen's theorem asserts that there exists a constant $C \geq 0$, independent of Δx, such that, for all $n \geq 1$ and $x \in \mathbb{R}$:

$$\left| F_n(x) - \Phi(x) \right| \leq \frac{C}{\sqrt{n}(1 + |x|^3)}, \tag{5.44}$$

see Petrov [25] or Shiryaev [30], where we let:

$$F_n(x) = \mathbb{P}\left[2\frac{S_n - \mathbb{E}(S_n)}{\sqrt{n}} \leq x \right], \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{z^2}{2}\right) dz, \quad x \in \mathbb{R}.$$

Returning to (5.42), we get:

$$\begin{aligned} \frac{2}{\sqrt{n}} \mathbb{E}\left[|S_n - \mathbb{E}(S_n)|\right] &= \int_0^{+\infty} \mathbb{P}\left[\frac{2|S_n - \mathbb{E}(S_n)|}{\sqrt{n}} \geq x \right] dx \\ &= \int_0^{+\infty} \Big(F_n(-x) + 1 - F_n(x-) \Big) dx, \end{aligned} \tag{5.45}$$

where we used the notation $F_n(x-) = \lim_{y \nearrow x} F_n(y)$. Since $F_n(x-) = F_n(x)$ for almost every $x \in \mathbb{R}$, we can easily replace $F_n(x-)$ by $F_n(x)$ in the last term right above. Of course, a similar identity holds true for the L^1-norm of a standard Gaussian random variable Z:

$$\mathbb{E}\left[|Z|\right] = \int_0^{+\infty} \Big(\Phi(-x) + 1 - \Phi(x) \Big) dx. \tag{5.46}$$

Recall from (5.42) that $\mathbb{E}[|Z|] = \sqrt{2}/\sqrt{\pi}$. Making the difference between (5.45) and (5.46) and invoking (5.44), we deduce that:

$$\left| \frac{2}{\sqrt{n}} \mathbb{E}\left[|S_n - \mathbb{E}(S_n)|\right] - \frac{\sqrt{2}}{\sqrt{\pi}} \right| \leq \frac{2C}{\sqrt{n}} \int_0^{+\infty} \frac{1}{1 + x^3} dx.$$

Allowing the constant C to increase from line to line and recalling that $W_1(\rho(t^n),\rho^n_{\Delta x}) = \Delta x \times \mathbb{E}[|S_n - \mathbb{E}(S_n)|]$, we end up with:

$$\forall n \geq 1, \quad \left| W_1(\rho(t^n),\rho^n_{\Delta x}) - \frac{1}{\sqrt{\pi}}\sqrt{t^n\Delta x}\right| \leq \frac{C}{n}.$$

Of course, one may bypass the use of the central limit theorem and perform the computations explicitly. Choose for instance n of the form $n = 2k$, $k \in \mathbb{N}$. Then, thanks to the parity of the binomial coefficients and to the fact that, for $j = k$, $2k\Delta t = k\Delta x = j\Delta x$, we have

$$\begin{aligned}
&W_1(\rho(t^n),\rho^n_{\Delta x})\\
&= 2\sum_{j=0}^{k-1}\binom{2k}{j}(1/2)^{2k}(2k\Delta t - j\Delta x) = 2\sum_{j=0}^{k-1}\binom{2k}{j}(1/2)^{2k}(k\Delta x - j\Delta x)\\
&= 2\Delta x\sum_{j=0}^{k-1}\binom{2k}{j}(1/2)^{2k}(k-j) = 2\Delta x\left(k\sum_{j=0}^{k-1}\binom{2k}{j}(1/2)^{2k} - k\sum_{j=1}^{k-1}\binom{2k-1}{j-1}(1/2)^{2k-1}\right)\\
&= 2k\Delta x\left(\sum_{j=0}^{k-1}\binom{2k}{j}(1/2)^{2k} - \sum_{j=0}^{k-2}\binom{2k-1}{j}(1/2)^{2k-1}\right).
\end{aligned}$$

Using the two identities

$$\begin{aligned}
2\sum_{j=0}^{k-1}\binom{2k}{j}(1/2)^{2k} &= 1 - \binom{2k}{k}(1/2)^{2k},\\
2\sum_{j=0}^{k-2}\binom{2k-1}{j}(1/2)^{2k-1} &= 1 - 2\binom{2k-1}{k-1}(1/2)^{2k-1},
\end{aligned}$$

this rewrites

$$\begin{aligned}
W_1(\rho(t^n),\rho^n_{\Delta x}) &= k\Delta x\left(1 - \binom{2k}{k}(1/2)^{2k} - 1 + \binom{2k-1}{k-1}(1/2)^{2k-2}\right)\\
&= k\Delta x\left(4\binom{2k-1}{k-1} - \binom{2k}{k}\right)(1/2)^{2k} = k\Delta x\binom{2k}{k}(1/2)^{2k}.
\end{aligned}$$

From Stirling's formula, we thus recover that

$$W_1(\rho(t^n),\rho^n_{\Delta x}) \sim_{n\to\infty} k\Delta x\frac{4^k}{\sqrt{k\pi}}(1/2)^{2k} = \frac{1}{\sqrt{\pi}}\sqrt{t^n\Delta x}.$$

5.2 Numerical illustration

We present in the following several numerical examples for which we compute the numerical error in the Wasserstein distance W_1 using formula (5.41). For these computations, we choose the final time $T = 2$ and the computational domain is $[-2.5, 2.5]$. We compute the error in the Wasserstein distance W_1 for different space and time step to estimate the convergence order.

Example 1. We consider a velocity field given by $a(t,x) = 1$ for $x < 0$ and $a(t,x) = \frac{1}{2}$ for $x \geq 0$. Since a is non-increasing, a satisfies the OSL condition (1.2). For this example, we choose the initial datum $\rho^{ini} = \delta_{x_0}$ with $x_0 = -0.5$. Then the solution to the transport equation (1.1) is given by

$$\rho(t,x) = \delta_{t+x_0}(x) \quad \text{for } t < -x_0; \quad \rho(t,x) = \delta_{\frac{1}{2}(t+x_0)}(x) \quad \text{for } t \geq -x_0.$$

Then the generalized inverse is given for $z \in [0,1)$ by $Q_\rho(t,z) = t + x_0$ if $t < -x_0$, $Q_\rho(t,z) = \frac{1}{2}(t+x_0)$ if $t \geq -x_0$. Therefore, denoting $u_j^n = \sum_{k\leq j} \rho_k^n$, we can compute easily the error at time $t^n = n\Delta t$,

$$e^n := W_1(\rho(t^n), \rho_{\Delta x}^n) = \sum_{k\in\mathbb{Z}} \int_{u_{k-1}^n}^{u_k^n} |x_k - Q_\rho(t^n, z)| dz.$$

Then we define the numerical error as $e = \max_{n\leq T/\Delta t} e^n$. We display in Figure 1 the numerical error with respect to the number of nodes in logarithmic scale computed with this procedure for $\Delta t = 0.8\Delta x$. We observe that the computed numerical error is of order $1/2$. This suggests the optimality of the result in Theorem 4.1.

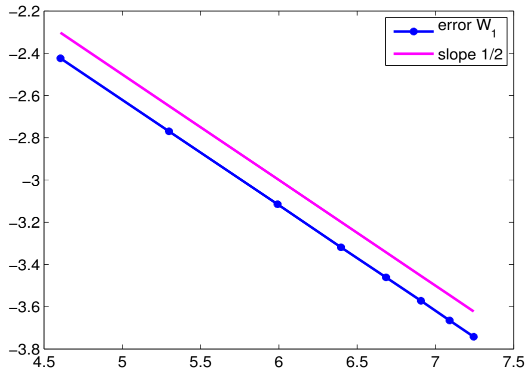


Figure 1: Numerical error with respect to the number of nodes in logarithmic scale for the upwind scheme in Wasserstein distance W_1 in the case of example 1: initial datum given by a Dirac mass and velocity field $a(t,x) = 1$ for $x < 0$ and $a(t,x) = 1/2$ for $x \geq 0$.

Example 2. We consider the same velocity field as above, given by $a(t,x) = 1$ for $x < 0$ and $a(t,x) = \frac{1}{2}$ for $x \geq 0$. However, we choose for initial datum the piecewise constant function $\rho^{ini} = \mathbf{1}_{[-1,1]}$. Then the solution to the transport equation (1.1) is given by

$$\rho(t,x) = \begin{cases} \mathbf{1}_{[-1+t,0)} + 2\,\mathbf{1}_{[0,t/2)} + \mathbf{1}_{[t/2,1+t/2)}, & \text{for } t \leq 1, \\ 2\,\mathbf{1}_{[\frac{1}{2}(t-1),\frac{t}{2})} + \mathbf{1}_{[t/2,1+t/2)}, & \text{for } t > 1. \end{cases}$$

We perform the numerical computation as in the first example. Figure 2 displays a comparison between the numerical solution $\rho_{\Delta x}$ and the exact solution ρ at time $T = 2$, with $\Delta x = 2/1000$ and $\Delta t = 0.8\Delta x$. As expected we observe numerical diffusion. Figure 3-left reports the numerical error in Wasserstein distance W_1 with respect to the number of nodes in logarithmic scale, with $\Delta t = 0.8\Delta x$ is given. We observe that the numerical error seems to be of order 1 in this case. However, since the solution stays in L^1, we can estimate the numerical error in L^1, which is provided in Figure 3-right. We observe that this numerical error is of order $1/2$.

Example 3. We consider the velocity field $a(t,x) = 2$ for $x < \min(t,1)$ and $a(t,x) = 1$ for $x \geq \min(t,1)$. Since a is non-increasing with respect to x, it satisfies the one-sided Lipschitz continuity condition. The initial datum is given by: $\rho^{ini} = \mathbf{1}_{[-1,0]}$. In this case, the solution to the transport equation (1.1) is given by

$$\rho(t,x) = \begin{cases} \mathbf{1}_{[-1+2t,t)} + t\,\delta_t, & \text{for } t < 1, \\ \delta_t, & \text{for } t \geq 1. \end{cases}$$

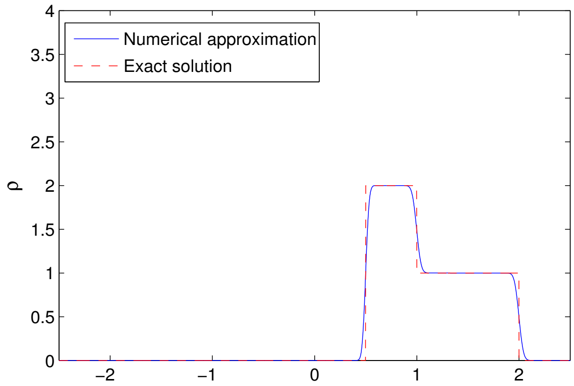


Figure 2: Comparison between the numerical approximation obtained with the upwind scheme and the exact solution at time $T = 2$ for the second example.

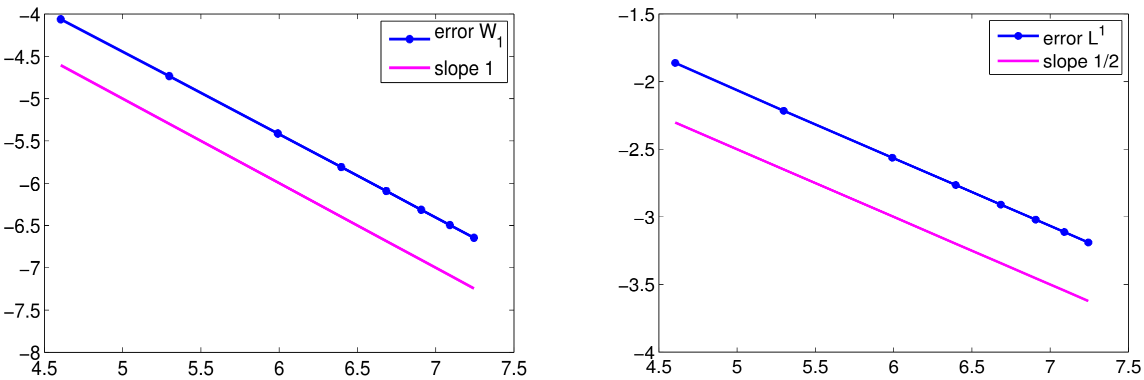


Figure 3: Left: Numerical error with respect to the number of nodes for the upwind scheme in Wasserstein distance W_1 in logarithmic scale for the second example. Right: Numerical error for the upwind scheme in L^1 norm in logarithmic scale for the second example.

We deduce the expression of the generalized inverse,

$$Q_\rho(z) = \begin{cases} (z-1+t)\mathbf{1}_{[0,1-t)} + \mathbf{1}_{[1-t,1)}, & \text{for } t < 1, \\ t, & \text{for } t \geq 1. \end{cases}$$

Performing the numerical computation, we obtain the numerical error displayed in Figure 4, in the same way for the preceding two examples, with $\Delta t = 0.4\Delta x$. We observe that in this case the order of the convergence is $1/2$. Compared to example 2, although the initial datum is regular (piecewise constant), we have the formation of a Dirac delta in finite time. Then the solution is defined as a measure and the observed numerical order of convergence falls down to $1/2$.

As a conclusion, these numerical results seem to indicate that as long as the numerical solution belongs to $L^1 \cap BV(\mathbb{R}^d)$, the convergence of the upwind scheme is of order 1 in Wasserstein distance. However, when the velocity field is only bounded and one-sided Lipschitz continuous, the solution might no longer be a function; for instance in example 3 above, Dirac deltas are created for $t > 0$ althought the initial datum is piecewise constant. When such singularities appear, examples 1 and 3 indicate that convergence order falls down to $1/2$ which shows the optimality of Theorem 4.1.

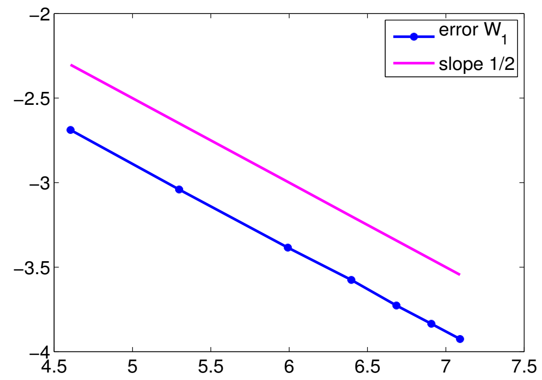


Figure 4: Numerical error with respect to the number of nodes in logarithmic scale for the upwind scheme in Wasserstein distance W_1 in the case of example 3 for which a Dirac delta is created from the initial datum $\rho^{ini} = \mathbf{1}_{[-1,0]}$.

Appendix

1 Generalization to other finite volume schemes

For simplicity of the notations, we have presented our analysis for an upwind scheme. In this appendix, we generalize our approach to other schemes on Cartesian grids. With the notation above, for $J \in \mathbb{Z}^d$, we consider the scheme:

$$\rho_J^{n+1} = \rho_J^n - \sum_{i=1}^d \frac{\Delta t}{\Delta x_i}\Big(g^n_{J+\frac12 e_i}(\rho_J^n, \rho_{J+e_i}^n) - g^n_{J-\frac12 e_i}(\rho^n_{J-e_i}, \rho^n_J)\Big), \tag{A.1}$$

where we take the general form for the flux

$$g^n_{J+\frac12 e_i}(u,v) = \zeta^n_{J+\frac12 e_i} u - \beta^n_{J+\frac12 e_i} v.$$

We make the following assumptions on the coefficients:

$$0 \le \zeta^n_{J+\frac12 e_i} \le \zeta_\infty, \quad 0 \le \beta^n_{J+\frac12 e_i} \le \beta_\infty. \tag{A.2}$$

Then, equation (A.1) can be rewritten as

$$\rho_J^{n+1} = \rho_J^n\Big(1 - \sum_{i=1}^d \frac{\Delta t}{\Delta x_i}\big(\zeta^n_{J+\frac12 e_i} + \beta^n_{J-\frac12 e_i}\big)\Big) + \sum_{i=1}^d \frac{\Delta t}{\Delta x_i}\Big(\beta^n_{J+\frac12 e_i}\rho^n_{J+e_i} + \zeta^n_{J-\frac12 e_i}\rho^n_{J-e_i}\Big).$$

Assuming that the following CFL condition holds

$$(\beta_\infty + \zeta_\infty)\sum_{i=1}^d \frac{\Delta t}{\Delta x_i} \le 1, \tag{A.3}$$

the scheme is clearly non-negative. We define then the random characteristics as in Section 3.3 by (3.18) where the transition matrix at time n is now given by

$$P^n_{J,L} = \begin{cases} 1 - \displaystyle\sum_{i=1}^{d} \frac{\Delta t}{\Delta x_i} (\zeta^n_{J+\frac12 e_i} + \beta^n_{J-\frac12 e_i}) & \text{when } L = J, \\ \dfrac{\Delta t}{\Delta x_i} \zeta^n_{J+\frac12 e_i} & \text{when } L = J + e_i, \quad \text{for } i = 1, \ldots, d, \\ \dfrac{\Delta t}{\Delta x_i} \beta^n_{J-\frac12 e_i} & \text{when } L = J - e_i, \quad \text{for } i = 1, \ldots, d, \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that Lemma 3.6 and Proposition 3.7 (ii) hold true with this random characteristics. We compute

$$\mathbb{E}^n_J\big(X^{n+1} - X^n\big) = \sum_{i=1}^{d} \left(\Delta x_i \, \zeta^n_{J+\frac12 e_i} \frac{\Delta t}{\Delta x_i} - \Delta x_i \, \beta^n_{J-\frac12 e_i} \frac{\Delta t}{\Delta x_i} \right) e_i.$$

Thus

$$\mathbb{E}^n_J\big(X^{n+1} - X^n\big) = \Delta t \sum_{i=1}^{d} (\zeta^n_{J+\frac12 e_i} - \beta^n_{J-\frac12 e_i}) e_i.$$

We deduce the following result:

Proposition A.1 *Under the assumptions of Theorem 4.1 on ρ^{ini} and a, assume further that the bounds* (A.2) *and the CFL condition* (A.3) *hold true.*

If moreover the weights $(((\zeta^n_{J+e_i/2})_{i=1,\ldots,d})_{J\in\mathbb{Z}^d})_{n\in\mathbb{N}}$ and $(((\beta^n_{J+e_i/2})_{i=1,\ldots,d})_{J\in\mathbb{Z}^d})_{n\in\mathbb{N}}$ satisfy

$$\zeta^n_{J+\frac12 e_i} - \beta^n_{J-\frac12 e_i} = {a_i}^n_J, \tag{A.4}$$

where ${a_i}^n_J$ is given in (3.11)*, then, the result of Theorem 4.1 still holds true for the scheme* (A.1)*.*

Indeed, thanks to (A.4), Proposition 3.7 (i) holds true. Thus, we can redo the proof of Theorem 4.1 in this framework.

Example A.2

- *We first observe that, for $\zeta^n_{J+\frac12 e_i} = ({a_i}^n_J)^+$ and $\beta^n_{J-\frac12 e_i} = -({a_i}^n_J)^-$,* (A.4) *is satisfied. This choice corresponds to the upwind scheme* (3.10) *considered in this paper.*
- *If we now consider the Rusanov scheme (see [22, 6]), we then have $\zeta^n_{J+\frac12 e_i} = \frac12({a_i}^n_J + a_\infty)$ and $\beta^n_{J-\frac12 e_i} = \frac12(-{a_i}^n_J + a_\infty)$. We easily check that* (A.2) *and* (A.4) *are satisfied. Thus our result also shows that the Rusanov scheme, when applied to a conservative transport equation with a velocity field that is only L^∞ and OSL, has an order $1/2$ in distance W_p, $p \geq 1$.*

Finally, as pointed out in Remark 3.1, we observe that for another traditional upwind scheme given by:

$$\rho^{n+1}_J = \rho^n_J - \sum_{i=1}^{d} \frac{\Delta t}{\Delta x_i} \Big(({a_i}^n_{J+\frac12 e_i})^+ \rho^n_J - ({a_i}^n_{J+\frac12 e_i})^- \rho^n_{J+e_i} - ({a_i}^n_{J-\frac12 e_i})^+ \rho^n_{J-e_i} + ({a_i}^n_{J-\frac12 e_i})^- \rho^n_J \Big),$$

where ${a_i}^n_{J+\frac{1}{2}e_i} = \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} a_i(s, x_{J+\frac{1}{2}e_i})\, ds$, we have

$$\mathbb{E}_J^n(X^{n+1} - X^n) = \int_{t^n}^{t^{n+1}} \sum_{i=1}^{d} \Big(a_i(s, X^n + \frac{\Delta x_i}{2} e_i)^+ - a_i(s, X^n - \frac{\Delta x_i}{2} e_i)^- \Big) e_i\, ds.$$

Then the statements of Proposition 3.7 (i) does not hold. Consequently, we cannot use the techniques developed in this paper.

2 Application of the technique to a scheme on unstructured meshes

In this section, we explain shortly how to obtain the error estimate for a forward semi-Lagrangian scheme defined on an unstructured mesh. For the sake of simplicity, we present the case of a triangular mesh in dimension 2, but this approach can be easily extended to any mesh made of simplices, in any dimension. Basic references on forward semi-Lagrangian schemes are [13] and [14] (although they concern schemes on structured quadrilateral meshes).

2.1 Numerical algorithm

Let us consider a triangular mesh $\mathcal{T} = (T_k)_{k\in\mathbb{Z}}$ with nodes $(x_i)_{i\in\mathbb{Z}}$. We assume this mesh to be conformal: A summit cannot belong to an open edge of the grid. The triangles $(T_k)_{k\in\mathbb{Z}}$ are assumed to satisfy $\bigcup_{k\in\mathbb{Z}} T_k = \mathbb{R}^2$ and $T_k \cap T_l = \emptyset$ if $k \neq l$ (in particular, the cells are here not assumed to be closed nor open). For any triangle T with summits x, y, z, we will use also the notation $(x, y, z) = T$. We denote by $\mathcal{V}(T) = \mathcal{V}(x, y, z)$ the area of this triangle, and $h(T)$ its height (defined as the minimum of the three heights of the triangle T). We make the assumption that the mesh satisfies $\hbar := \inf_{k\in\mathbb{Z}} h(T_k) > 0$.

For any node x_i, $i \in \mathbb{Z}$, we denote by $K(i)$ the set of indices indexing triangles that have x_i as a summit, and we denote by $\mathcal{T}_i$ the set of all triangles of $\mathcal{T}$ that have x_i as a summit: thus $\mathcal{T}_i = \{T_k; k \in K(i)\}$.

For any triangle T_k, $k \in \mathbb{Z}$, we denote by

$$I(k) = \{I_1(k), I_2(k), I_3(k)\} \tag{A.5}$$

the set of indices indexing the summits of T_k (for some arbitrary order, whose choice has no importance for the sequel).

Here is the derivation of the forward semi-Lagrangian scheme, whose rigorous definition is given next, in (A.9). Let us emphasize that this is not a finite volume scheme.

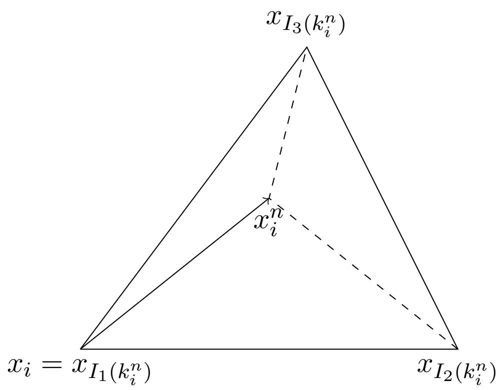
- For an initial distribution ρ^{ini} of the PDE (1.1), define the probability weights $(\rho_i^0)_{i\in\mathbb{Z}}$ through the following procedure: Consider the one-to-one mapping $\iota : \mathbb{Z} \ni k \mapsto \iota(k) \in \mathbb{Z}$ such that, for each $k \in \mathbb{Z}$, $x_{\iota(k)}$ is a node of the triangle T_k; ι is thus a way to associate a node with a cell; then, for all $i \in \mathbb{Z}$, let $\rho_i^0 = \sum_{k;\iota(k)=i} \rho^{ini}(T_k)$. Observe from (4.37) that $\rho_{\Delta x}^0 = \sum_{j\in\mathbb{Z}} \rho_j^0 \delta_{x_j}$ is an approximation of ρ^{ini}.

- Assume that, for a given $n \in \mathbb{N}$, we already have probability weights $(\rho_i^n)_{i\in\mathbb{Z}}$ such that $\rho_{\Delta x}^n = \sum_{j\in\mathbb{Z}} \rho_j^n \delta_{x_j}$ is an approximation of $\rho(t^n, \cdot)$, where ρ is the solution to (1.1) with ρ^{ini} as initial condition. Similar to (3.11), let us denote $a_i^n = \Delta t^{-1} \int_{t^n}^{t^{n+1}} a(s, x_i)\, ds$, and $x_i^n = x_i + a_i^n \Delta t$, for $i \in \mathbb{Z}$. Under the CFL-like condition

$$a_\infty \Delta t \leq \hbar, \tag{A.6}$$

x_i^n belongs to one (and only one) of the elements of $\mathcal{T}_i$. We denote by k_i^n the index of this triangle: $x_i^n \in T_{k_i^n}$.

- The basic idea now is to use a linear splitting rule between the summits of the triangle $T_{k_i^n}$: the mass ρ_i^n is sent to these three points $x_{I_1(k_i^n)}$, $x_{I_2(k_i^n)}$, $x_{I_3(k_i^n)}$ according to the *barycentric coordinates* of x_i^n in the triangle. In some sense, this scheme is a natural extension of the one-dimensional upwind scheme to greater dimensions (see the interpretation of the one-dimensional upwind scheme provided in Remark 3.3).



Let $T = (x, y, z) \in \mathcal{T}$, and $\xi \in T$. We define the barycentric coordinates of ξ with respect to x, y and z, λ_x^T, λ_y^T and λ_z^T:

$$\lambda_x^T(\xi) = \frac{\mathcal{V}(\xi, y, z)}{\mathcal{V}(T)}, \quad \lambda_y^T(\xi) = \frac{\mathcal{V}(\xi, x, z)}{\mathcal{V}(T)}, \quad \lambda_z^T(\xi) = \frac{\mathcal{V}(\xi, x, y)}{\mathcal{V}(T)}, \tag{A.7}$$

and then have $\xi = \lambda_x^T(\xi)x + \lambda_y^T(\xi)y + \lambda_z^T(\xi)z$. Note also that $\lambda_x^T(\xi) + \lambda_y^T(\xi) + \lambda_z^T(\xi) = 1$. Therefore, we have the following fundamental property, which will be used in the sequel:

$$\lambda_x^T(\xi)(x - \zeta) + \lambda_y^T(\xi)(y - \zeta) + \lambda_z^T(\xi)(z - \zeta) = \xi - \zeta, \tag{A.8}$$

for any $\zeta \in \mathbb{R}^2$.

Considering $x_i^n \in T_{k_i^n}$, we will use the barycentric coordinates of x_i^n with respect to the summits $(x_j)_{j \in I(k_i^n)}$ of $T_{k_i^n}$. For notational convenience, let us denote

$$\lambda_{i,j}^n = \lambda_{x_j}^T(x_i^n) \qquad \text{when} \quad T = T_{k_i^n}.$$

The numerical scheme reads:

$$\rho_j^{n+1} = \sum_{i \in \gamma(j)} \rho_i^n \lambda_{i,j}^n, \qquad j \in \mathbb{Z}, \ n \in \mathbb{N}, \tag{A.9}$$

where, for a given $j \in \mathbb{Z}$, we denote by $\gamma(j)$ the set of all indices $i \in \mathbb{Z}$ indexing nodes x_i such that $x_i + a_i^n \Delta t$ belongs to a triangle that has x_j as a summit :

$$\gamma(j) = \{i \in \mathbb{Z} \,/\, \text{there exists } k \in K(j) \text{ such that } x_i + a_i^n \Delta t \in T_k\}.$$

2.2 Probabilistic interpretation

As in Section 3.3, we define a random characteristic associated to the scheme (A.9). Letting $\Omega = \mathbb{Z}^{\mathbb{N}}$ and defining the canonical process $(I^n)_{n\in\mathbb{N}}$ as we defined $(K^n)_{n\in\mathbb{N}}$ above (the definition is the same but we prefer to use the letter I instead of K; we make this clear right below), we equip Ω with the Kolmogorov σ-field $\mathcal{A}$ and with a collection of probability measures $(\mathbb{P}_\mu)_{\mu\in\mathcal{P}(\mathbb{Z})}$, such that, for each $\mu \in \mathcal{P}(\mathbb{Z})$, $(I^n)_{n\in\mathbb{N}}$ is a time-inhomogeneous Markov chain under $\mathbb{P}_\mu$, with μ as initial distribution and with transition matrix:

$$P^n_{i,j} = \begin{cases} \lambda^n_{i,j} & \text{when } j \in I(k^n_i), \\ 0 & \text{otherwise,} \end{cases} \tag{A.10}$$

where we used the notation $I(k^n_i)$ introduced in (A.5), that is to say, more precisely,

$$P^n_{i,j} = \begin{cases} \lambda^{T_{k^n_i}}_{x_j}(x^n_i) & \text{when } j \in I(k^n_i), \\ 0 & \text{otherwise,} \end{cases}$$

with the notation in (A.7). Pay attention that the chain $(I^n)_{n\in\mathbb{N}}$ here takes values in the set of indices indexing the nodes of the grid whilst the chain $(K^n)_{n\in\mathbb{N}}$ used in the analysis of the upwind scheme (see Section 3.3) takes values in the set of indices indexing the cells of the grid. This is the rationale for using different letters.

Then, we let the random characteristics be the sequence of random variables $(X^n)_{n\in\mathbb{N}}$ from $(\Omega, \mathcal{A})$ into $\mathbb{R}^2$ defined by

$$\forall\, n \in \mathbb{N},\ \forall \omega \in \Omega, \quad X^n(\omega) = x_{I^n(\omega)}. \tag{A.11}$$

We now check that Proposition 3.7 still holds true with this definition of the random characteristics:

Proposition A.1 *Let $(X^n)_{n\in\mathbb{N}}$ be the random characteristics defined by* (A.10)–(A.11).
(i) Defining $\rho^n_{\Delta x} = \sum_{j\in\mathbb{Z}} \rho^n_j \delta_{x_j}$, we have $\rho^n_{\Delta x} = X^n{}_{\#}\mathbb{P}_{\rho^0_{\Delta x}}$.
(ii) For all $j \in \mathbb{Z}$, we have, with probability one under $\mathbb{P}_j$, $\mathbb{E}^n_j(X^{n+1} - X^n) = a^n_{I^n}\Delta t$.

Proof. (i) This result follows straightforwardly from the proof of (ii) in Proposition 3.7, but with the transition matrix defined in (A.10).

(ii) From a direct computation, we have

$$\mathbb{E}^n_j(X^{n+1} - X^n) = \sum_{\ell \in I(k^n_{I^n})} \lambda^n_{I^n,\ell}(x_\ell - x_{I^n}).$$

Thanks to Property (A.8),

$$\mathbb{E}^n_j(X^{n+1} - X^n) = x^n_{I^n} - x_{I^n} = a^n_{I^n}\Delta t,$$

which completes the proof. □

2.3 Convergence order

By the same token as in Section 4, we can use Proposition A.1 and Lemmas 4.2 and 4.5 to prove that the numerical scheme (A.9) is of order 1/2:

Theorem A.2 *Let $\rho^{ini} \in \mathcal{P}_p(\mathbb{R}^d)$ for $p \geq 1$. Let us assume that $a \in L^\infty([0,\infty); L^\infty(\mathbb{R}^2))^2$ and satisfies the OSL condition* (1.2). *Let $\rho = Z_\# \rho^{ini}$ be the unique measure solution to the aggregation equation with initial datum ρ^{ini} in the sense of Theorem 2.4. Let us consider a triangular conformal mesh $(T_k)_{k\in\mathbb{Z}}$ with nodes $(x_j)_{j\in\mathbb{Z}}$ such that $\hbar = \inf_{k\in\mathbb{Z}} h(T_k) > 0$. We denote by Δx the longest edge in the mesh. We define*

$$\rho^n_{\Delta x} = \sum_{j\in\mathbb{Z}} \rho^n_j \delta_{x_j},$$

where the approximation sequence (ρ^n_j) is computed thanks to the scheme (A.9). *We assume that the CFL condition* (A.6) *holds. Then, there exists a non-negative constant C, such that for all $n \in \mathbb{N}^*$,*

$$W_p(\rho(t^n), \rho^n_{\Delta x}) \leq C e^{2\int_0^{t^n} \alpha(s)ds} \big(\sqrt{t^n \Delta x} + \Delta x\big).$$

Proof. In order to repeat the arguments developed in Section 4, the main point is to check the analogue of the second inequality in (4.24).

To do so, we need to bound $\lambda^n_{i,j}$ for $j \neq i$. Clearly,

$$\sum_{j\neq i} \lambda^n_{i,j} = \frac{\mathcal{V}(x^n_i, x_i, z) + \mathcal{V}(x^n_i, x_i, y)}{\mathcal{V}(T_{k^n_i})},$$

where y and z are the two summits of $T_{k^n_i}$ that are different from x_i, namely $\{y, z\} = I(k^n_i) \setminus \{x_i\}$. Since $|x^n_i - x_i| \leq a_\infty \Delta t$, we have

$$\mathcal{V}(x^n_i, x_i, y) \leq \frac{a_\infty \Delta t \Delta x}{2}, \quad \mathcal{V}(x^n_i, x_i, z) \leq \frac{a_\infty \Delta t \Delta x}{2},$$

from which we get

$$\sum_{j\neq i} \lambda^n_{i,j} \leq \frac{a_\infty \Delta t \Delta x}{\mathcal{V}(T_{k^n_i})} \leq 2a_\infty \frac{\Delta t}{\hbar}.$$

This permits to implement the strategy proposed in Lemma 4.5, as long as the constants therein are allowed to depend upon $1/\hbar$, which shows the need for requiring $\hbar > 0$. □

3 An interpolation result

Proposition A.1 ([28]) *There exists a constant C such that, for any f, g, non-negative functions in $BV(\mathbb{R}^d)$ such that $\int_{\mathbb{R}^d} f = \int_{\mathbb{R}^d} g = 1$, it holds that*

$$||f - g||_1 \leq C|f - g|^{1/2}_{BV} W_1(f, g)^{1/2},$$

where $|\cdot|_{BV}$ denotes the BV semi-norm.

Proof. Let $h \in L^\infty(\mathbb{R}^d)$ such that $||h||_{L^\infty} = 1$. Let ρ be a smoothing kernel and $\rho_\varepsilon(x) = \rho(x/\varepsilon)/\varepsilon^d$, for any $\varepsilon > 0$. Let us denotes $h_\varepsilon = h \star \rho_\varepsilon$, $f_\varepsilon = f \star \rho_\varepsilon$, $g_\varepsilon = g \star \rho_\varepsilon$, where $\star$ stands for the convolution product. One has

$$\int_{\mathbb{R}^d} h(f - g) = \int_{\mathbb{R}^d} h\big(f - g - (f_\varepsilon - g_\varepsilon)\big) + \int_{\mathbb{R}^d} h(f_\varepsilon - g_\varepsilon).$$

Let us now estimate both integrals in the right hand term above. On the one hand, we have

$$\int_{\mathbb{R}^d} h\big(f - g - (f_\varepsilon - g_\varepsilon)\big) \leq ||h||_{L^\infty} ||f - g - (f_\varepsilon - g_\varepsilon)||_{L^1} \leq C\varepsilon |f - g|_{BV},$$

for $C = \int_{\mathbb{R}^d} |x|\rho(x)\,dx$; indeed,

$$\begin{aligned}\int_{\mathbb{R}^d} |f - f_\varepsilon| &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \frac{(f(x) - f(x-y))\rho(y/\varepsilon)}{\varepsilon^d}\,dy \right| dx \\ &\le \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x) - f(x - \varepsilon y)|\rho(y)\,dy\,dx \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x) - f(x-\varepsilon y)|\,dx \right) \rho(y)\,dy \le \varepsilon |f|_{BV} \int_{\mathbb{R}^d} |y|\rho(y)\,dy,\end{aligned}$$

the last inequality being due to the fact that $\int_{\mathbb{R}^d} |f(x) - f(x-\varepsilon y)|\,dx \le \varepsilon |f|_{BV}|y|$ for any y: see for example Remark 3.25 in [1]. On the other hand,

$$\int_{\mathbb{R}^d} h(f_\varepsilon - g_\varepsilon) = \int_{\mathbb{R}^d} h_\varepsilon (f-g) \le W_1(f,g)||\nabla h_\varepsilon||_{L^\infty},$$

where we used the identity $W_1(f,g) = \sup_{h\in \mathcal{C}^1/||\nabla h||_{L^\infty}\le 1} \int_{\mathbb{R}^d} (f-g)h$ (see [32]). Furthermore,

$$||\nabla h_\varepsilon||_{L^\infty} = ||h \star \nabla \rho_\varepsilon||_{L^\infty} \le ||h||_{L^\infty}||\nabla\rho_\varepsilon||_{L^1} \le \frac{1}{\varepsilon}||\nabla \rho||_{L^1}.$$

In the end, taking $C = \max(\int_{\mathbb{R}^d} |x|\rho(x)dx, \int_{\mathbb{R}^d} |\nabla\rho(x)|dx)$, one gets

$$\int_{\mathbb{R}^d} h(f-g) \le C\Big(\varepsilon |f-g|_{BV} + \frac{1}{\varepsilon} W_1(f,g)\Big).$$

"Optimizing" in ε, that is to say, taking $\varepsilon = W_1(f,g)^{1/2}/|f-g|_{BV}^{1/2}$ (without any loss of generality, we can assume $|f-g|_{BV}^{1/2} \neq 0$), one gets

$$||f-g||_{L^1} = \sup_{h\in L^\infty, ||h||_{L^\infty}=1} \int_{\mathbb{R}^d} h(f-g) \le C|f-g|_{BV}^{1/2} W_1(f,g)^{1/2},$$

which completes the proof. □

Acknowledgements. NV acknowledges partial support from the french "ANR blanche" project Kibord : ANR-13-BS01-0004.

References

[1] L. Ambrosio, N. Fusco, D. Pallara, **Functions of bounded variation and free discontinuity problems**, Oxford Mathematical Monographs, 2000.

[2] L. Ambrosio, N. Gigli, G. Savaré, **Gradient flows in metric space of probability measures**, Lectures in Mathematics, Birkäuser, 2005.

[3] J-P. Aubin, A. Cellina, Differential inclusions. Set-valued maps and viability theory. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], **264**. Springer-Verlag, Berlin, 1984.

[4] S. Bianchini, M. Gloyer, *An estimate on the flow generated by monotone operators*, Comm. Partial Diff. Eq., **36** (2011), n^o 5, 777–796.

[5] D. Bouche, J.-M. Ghidaglia, F. Pascal, *Error estimate and the geometric corrector for the upwind finite volume method applied to the linear advection equation*, SIAM J. Numer. Anal. **43** (2) (2005), 578–603.

[6] F. Bouchut, **Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources**, Frontiers in Mathematics, Birkhäuser Verlag, Basel, 2004.

[7] F. Bouchut, R. Eymard, A. Prignet, *Finite volume schemes for the approximation via characteristics of linear convection equations with irregular data*, J. Evol. Equ. **11** (2011), 687–724.

[8] F. Bouchut, F. James, *One-dimensional transport equations with discontinuous coefficients*, Nonlinear Analysis TMA, **32** (1998), n^o 7, 891–933.

[9] F. Bouchut, F. James, *Duality solutions for pressureless gases, monotone scalar conservation laws, and uniqueness*, Comm. Partial Differential Eq., **24** (1999), 2173–2189.

[10] F. Bouchut, F. James, S. Mancini, *Uniqueness and weak stability for multidimensional transport equations with one-sided Lipschitz coefficients*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (5), IV (2005), 1–25.

[11] F. Boyer, *Analysis of the upwind finite volume method for general initial and boundary value transport problems*, IMA Journal on Numerical Analysis, Vol 32 no 4 (2012), 1404–1439

[12] F. Delarue, F. Lagoutière, *Probabilistic analysis of the upwind scheme for the transport equations*, Arch. Rational Mech. Anal. **199** (2011) 229–268.

[13] N. Crouseilles, Th. Respaud, É. Sonnendrücker, *A forward semi-Lagrangian method for the numerical solution of the Vlasov equation*, Comput. Phys. Comm. **180** (2009), 1730?1745.

[14] J. Denavit, *Numerical simulation fo plasmas with periodic smoothing in phase space*, J. Comput. Phys. **9** (1972) 75–98.

[15] B. Després, *An explicit a priori estimate for a finite volume approximation of linear advection on non-Cartesian grid*, SIAM J. Numer. Anal. **42** (2) (2004), 484–504.

[16] B. Engquist, S. Osher, *One-sided difference approximations for nonlinear conservation laws*, Math. Comp. **36** (1981), 321–351.

[17] A.F. Filippov, *Differential Equations with Discontinuous Right-Hand Side*, A.M.S. Transl. (2) **42** (1964), 199–231.

[18] L. Gosse, F. James, *Numerical approximations of one-dimensional linear conservation equations with discontinuous coefficients*, Math. Comp. **69** (2000) 987–1015.

[19] L. Gosse, G. Toscani, *Identification of Asymptotic Decay to Self-Similarity for One-Dimensional Filtration Equations*, SIAM J. Numer. Anal. 43 (2006) 2590–2606.

[20] N. N. Kuznetsov, *The accuracy of certain approximate methods for the computation of weak solutions of a first order quasilinear equation*, Ž. Vyčisl. Mat. i Mat. Fiz. **16** (6) (1976) 1489–1502, 1627.

[21] R. J. LeVeque, **Numerical methods for conservation laws**, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 1992.

[22] R. J. LeVeque, **Finite volume methods for hyperbolic problems**, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2002.

[23] B. Merlet, *L^∞- and L^2-error estimates for a finite volume approximation of linear advection*, SIAM J. Numer. Anal. **46** (1) (2007), 124–150.

[24] B. Merlet, J. Vovelle, *Error estimate for finite volume scheme*, Numer. Math. **106** (2007), 129–155.

[25] V.V. Petrov, **Sums of independent random variables**, Springer-Verlag, New-York Heidelberg, 1975.

[26] F. Poupaud, M. Rascle, *Measure solutions to the linear multidimensional transport equation with discontinuous coefficients*, Comm. Partial Diff. Equ., **22** (1997), 337–358.

[27] S. T. Rachev and L. Rüschendorf, **Mass Transportation Problems. Vol. I. Theory**, Probab. Appl. (N. Y.), Springer-Verlag, New York, 1998.

[28] F. Santambrogio, *personal communication.*

[29] F. Santambrogio, **Optimal transport for applied mathematicians**, Progress in Nonlinear Differential Equations and their Applications, 87, Birkhäuser/Springer, Cham, 2015.

[30] A.N. Shiryaev, **Probability**, Springer-Verlag, Berlin Heidelberg, 1984.

[31] J.-P. Vila, Ph. Villedieu, *Convergence of an explicit finite volume scheme for first order symmetric systems*, Numer. Math, **94** (2003), 573–602.

[32] C. Villani, **Optimal transport, old and new**, Grundlehren der Mathematischen Wissenschaften 338, Springer, 2009.

[33] C. Villani, **Topics in optimal transportation**, Graduate Studies in Mathematics **58**, Amer. Math. Soc, Providence, 2003.