Kinetic entropy inequality and hydrostatic reconstruction scheme for the Saint-Venant system

Emmanuel Audusse, François Bouchut, Marie-Odile Bristeau, Jacques Sainte-Marie

KINETIC ENTROPY INEQUALITY AND HYDROSTATIC RECONSTRUCTION SCHEME FOR THE SAINT-VENANT SYSTEM

EMMANUEL AUDUSSE, FRANÇOIS BOUCHUT, MARIE-ODILE BRISTEAU, AND JACQUES SAINTE-MARIE

Abstract. A lot of well-balanced schemes have been proposed for discretizing the classical Saint-Venant system for shallow water flows with non-flat bottom. Among them, the hydrostatic reconstruction scheme is a simple and efficient one. It involves the knowledge of an arbitrary solver for the homogeneous problem (for example Godunov, Roe, kinetic...). If this solver is entropy satisfying, then the hydrostatic reconstruction scheme satisfies a semi-discrete entropy inequality. In this paper we prove that, when used with the classical kinetic solver, the hydrostatic reconstruction scheme also satisfies a fully discrete entropy inequality, but with an error term. This error term tends to zero strongly when the space step tends to zero, including solutions with shocks. We prove also that the hydrostatic reconstruction scheme does not satisfy the entropy inequality without error term.

1. Introduction

The classical Saint-Venant system for shallow water describes the height of water $h(t,x) \geq 0$, and the water velocity $u(t,x) \in \mathbb{R}$ (x denotes a coordinate in the horizontal direction) in the direction parallel to the bottom. It assumes a slowly varying topography $z(x)$, and reads

$$\begin{aligned} &\partial_t h + \partial_x(hu) = 0, \\ &\partial_t(hu) + \partial_x(hu^2 + g\frac{h^2}{2}) + gh\partial_x z = 0, \end{aligned} \tag{1.1}$$

where $g > 0$ is the gravity constant. This system is completed with an entropy (energy) inequality

$$\partial_t\Big(h\frac{u^2}{2} + g\frac{h^2}{2} + ghz\Big) + \partial_x\Big(\big(h\frac{u^2}{2} + gh^2 + ghz\big)u\Big) \leq 0. \tag{1.2}$$

We shall denote $U = (h, hu)^T$ and

$$\eta(U) = h\frac{u^2}{2} + g\frac{h^2}{2}, \qquad G(U) = \big(h\frac{u^2}{2} + gh^2\big)u \tag{1.3}$$

the entropy and entropy fluxes without topography.

The derivation of an efficient, robust and stable numerical scheme for the Saint-Venant system has received an extensive coverage. The issue involves the notion

2000 *Mathematics Subject Classification.* 65M12, 74S10, 76M12, 35L65.

Key words and phrases. Shallow water equations, well-balanced schemes, hydrostatic reconstruction, kinetic solver, fully discrete entropy inequality.

of well-balanced schemes, and we refer the reader to [10, 19, 17, 25] and references therein.

The hydrostatic reconstruction (HR), introduced in [1], is a general and efficient method that evaluates an arbitrary solver for the homogeneous problem, like Roe, relaxation, or kinetic solvers on *reconstructed states* built with the steady state relations. It leads to a consistent, well-balanced, positive scheme satisfying a semi-discrete entropy inequality, in the sense that the inequality holds only in the limit when the timestep tends to zero. The method has been generalized to balance all subsonic steady-states in [11], and to multi-layer shallow water in [12] with the source-centered variant of the hydrostatic reconstruction. Generic extensions are provided in [14], and a case of moving water is treated in [20]. The HR technique enables second-order computations on unstructured meshes, see [2]. It has also been used to derive efficient and robust numerical schemes approximating the incompressible Euler and Navier-Stokes equations with free surface [5, 3], i.e. non necessarily shallow water flows.

The aim of this paper is to prove that the hydrostatic reconstruction, when used with the classical kinetic solver [8, 4, 24, 9, 2, 18, 13], satisfies a fully discrete entropy inequality, stated in Corollary 3.7. However, as established in Proposition 3.8, this inequality necessarily involves an error term. The main result of this paper is that this error term is in the square of the topography increment, ensuring that it tends to zero strongly as the space step tends to zero, for solutions that can include shocks. The topography needs however to be Lipschitz continuous.

In general, to satisfy an entropy inequality is a criterion for the stability of a scheme. In the fully discrete case, it enables in particular to get an *a priori* bound on the total energy. In the time-only discrete case and without topography, the single energy inequality that holds for the kinetic scheme ensures the convergence [7]. The fully discrete case (still without topography) has been treated in [6]. Another approach to get a scheme satisfying a fully discrete entropy inequality is proposed in [15]. Following our results, the proof of convergence of the hydrostatic reconstruction scheme with kinetic numerical flux will be performed in a forthcoming paper.

The outline of the paper is as follows. We recall in Section 2 the kinetic scheme without topography and its entropy analysis, in both the discrete and semi-discrete cases. We show in particular how one can see that the fully discrete inequality is always less dissipative than the semi-discrete one, see Lemma 2.1. In Section 3 we propose a kinetic interpretation of the hydrostatic reconstruction and we give its properties. We analyze in detail the entropy inequality. The semi-discrete scheme is considered first. Our main result Theorem 3.6 concerning the fully discrete scheme is finally proved.

We end this section by recalling the classical kinetic approach, used in [24] for example, and its relation with numerical schemes. The kinetic Maxwellian is given by

$$M(U,\xi) = \frac{1}{g\pi}\Big(2gh - (\xi - u)^2\Big)_+^{1/2}, \tag{1.4}$$

where $U = (h, hu)^T$, $\xi \in \mathbb{R}$ and $x_+ \equiv \max(0, x)$ for any $x \in \mathbb{R}$. It satisfies the following moment relations,

$$\begin{aligned} \int_{\mathbb{R}} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M(U, \xi)\, d\xi &= U, \\ \int_{\mathbb{R}} \xi^2 M(U, \xi)\, d\xi &= hu^2 + g\frac{h^2}{2}. \end{aligned} \tag{1.5}$$

These definitions allow us to obtain a *kinetic representation* of the Saint-Venant system.

Lemma 1.1. *If the topography $z(x)$ is Lipschitz continuous, the pair of functions (h, hu) is a weak solution to the Saint-Venant system (1.1) if and only if $M(U, \xi)$ satisfies the kinetic equation*

$$\partial_t M + \xi \partial_x M - g(\partial_x z)\partial_\xi M = Q, \tag{1.6}$$

for some "collision term" $Q(t, x, \xi)$ that satisfies, for a.e. (t, x),

$$\int_{\mathbb{R}} Q d\xi = \int_{\mathbb{R}} \xi Q d\xi = 0. \tag{1.7}$$

Proof. If (1.6) and (1.7) are satisfied, we can multiply (1.6) by $(1, \xi)^T$, and integrate with respect to ξ. Using (1.5) and (1.7) and integrating by parts the term in $\partial_\xi M$, we obtain (1.1). Conversely, if (h, hu) is a weak solution to (1.1), just define Q by (1.6); it will satisfy (1.7) according to the same computations. □

The standard way to use Lemma 1.1 is to write a kinetic relaxation equation [21, 22, 16, 8, 9], like

$$\partial_t f + \xi \partial_x f - g(\partial_x z)\partial_\xi f = \frac{M - f}{\epsilon}, \tag{1.8}$$

where $f(t, x, \xi) \geq 0$, $M = M(U, \xi)$ with $U(t, x) = \int (1, \xi)^T f(t, x, \xi) d\xi$, and $\epsilon > 0$ is a relaxation time. In the limit $\epsilon \to 0$ we recover formally the formulation (1.6), (1.7). We refer to [8] for general considerations on such kinetic relaxation models without topography, the case with topography being introduced in [24]. Note that the notion of *kinetic representation* as (1.6), (1.7) differs from the so called *kinetic formulations* where a large set of entropies is involved, see [23]. For systems of conservation laws, these kinetic formulations include non-advective terms that prevent from writing down simple approximations. In general, kinetic relaxation approximations can be compatible with just a single entropy. Nevertheless this is enough for proving the convergence as $\epsilon \to 0$, see [7].

Apart from satisfying the moment relations (1.5), the particular form (1.4) of the Maxwellian is taken indeed for its compatibility with a kinetic entropy, that ensures energy dissipation in the relaxation approximation (1.8). Consider the kinetic entropy

$$H(f, \xi, z) = \frac{\xi^2}{2} f + \frac{g^2 \pi^2}{6} f^3 + gzf, \tag{1.9}$$

where $f \geq 0$, $\xi \in \mathbb{R}$ and $z \in \mathbb{R}$, and its version without topography

$$H_0(f, \xi) = \frac{\xi^2}{2} f + \frac{g^2 \pi^2}{6} f^3. \tag{1.10}$$

Then one can check the relations

$$\int_{\mathbb{R}} H\big(M(U,\xi),\xi,z\big)\,d\xi = \eta(U) + ghz, \tag{1.11}$$

$$\int_{\mathbb{R}} \xi H\big(M(U,\xi),\xi,z\big)\,d\xi = G(U) + ghzu. \tag{1.12}$$

One has the following subdifferential inequality and entropy minimization principle.

Lemma 1.2. *(i) For any $h \geq 0$, $u \in \mathbb{R}$, $f \geq 0$ and $\xi \in \mathbb{R}$*

$$H_0(f,\xi) \geq H_0\big(M(U,\xi),\xi\big) + \eta'(U)\begin{pmatrix}1\\ \xi\end{pmatrix}\big(f - M(U,\xi)\big). \tag{1.13}$$

(ii) For any $f(\xi) \geq 0$, setting $h = \int f(\xi)d\xi$, $hu = \int \xi f(\xi)d\xi$ (assumed finite), one has

$$\eta(U) = \int_{\mathbb{R}} H_0\big(M(U,\xi),\xi\big)\,d\xi \leq \int_{\mathbb{R}} H_0\big(f(\xi),\xi\big)\,d\xi. \tag{1.14}$$

Proof. This approach by the subdifferential inequality has been introduced in [8]. The property (ii) easily follows from (i) by taking $f = f(\xi)$ and integrating (1.13) with respect to ξ. For proving (i), notice first that

$$\eta'(U) = \big(gh - u^2/2, u\big), \tag{1.15}$$

where prime denotes differentiation with respect to $U = (h, hu)^T$. Thus

$$\eta'(U)\begin{pmatrix}1\\ \xi\end{pmatrix} = gh - u^2/2 + \xi u = \frac{\xi^2}{2} + gh - \frac{(\xi-u)^2}{2}. \tag{1.16}$$

Observe also that

$$\partial_f H_0(f,\xi) = \frac{\xi^2}{2} + \frac{g^2\pi^2}{2} f^2. \tag{1.17}$$

The formula defining M in (1.4) yields that

$$gh - \frac{(\xi-u)^2}{2} = \begin{cases} \dfrac{g^2\pi^2}{2} M(U,\xi)^2 & \text{if } M(U,\xi) > 0, \\ \text{is nonpositive} & \text{if } M(U,\xi) = 0, \end{cases} \tag{1.18}$$

thus

$$\partial_f H_0\big(M(U,\xi),\xi\big) = \begin{cases} \eta'(U)\begin{pmatrix}1\\ \xi\end{pmatrix} & \text{if } M(U,\xi) > 0, \\ \geq \eta'(U)\begin{pmatrix}1\\ \xi\end{pmatrix} & \text{if } M(U,\xi) = 0. \end{cases} \tag{1.19}$$

We conclude using the convexity of H_0 with respect to f that

$$\begin{aligned} H_0(f,\xi) &\geq H_0\big(M(U,\xi),\xi\big) + \partial_f H_0\big(M(U,\xi),\xi\big)\big(f - M(U,\xi)\big) \\ &\geq H_0\big(M(U,\xi),\xi\big) + \eta'(U)\begin{pmatrix}1\\ \xi\end{pmatrix}\big(f - M(U,\xi)\big), \end{aligned} \tag{1.20}$$

which proves the claim. □

For numerical purposes it is usual to replace the right-hand side in the kinetic relaxation equation (1.8) by a time discrete projection to the Maxwellian state. When space discretization is present it leads to flux-vector splitting schemes, see [9] for the case without topography, [24] for the case with topography, and [2] for the 2d case on unstructured meshes.

Here we consider more general schemes. We would like to approximate the solution $U(t,x)$, $x \in \mathbb{R}$, $t \geq 0$ of the system (1.1) by discrete values U_i^n, $i \in \mathbb{Z}$, $n \in \mathbb{N}$. In order to do so, we consider a grid of points $x_{i+1/2}$, $i \in \mathbb{Z}$,

$$\dots < x_{i-1/2} < x_{i+1/2} < x_{i+3/2} < \dots,$$

and we define the cells (or finite volumes) and their lengths

$$C_i =]x_{i-1/2}, x_{i+1/2}[, \qquad \Delta x_i = x_{i+1/2} - x_{i-1/2}.$$

We consider discrete times t^n with $t^{n+1} = t^n + \Delta t^n$, and we define the piecewise constant functions $U^n(x)$ corresponding to time t^n and $z(x)$ as

$$U^n(x) = U_i^n, \quad z(x) = z_i, \quad \text{for } x_{i-1/2} < x < x_{i+1/2}. \tag{1.21}$$

A finite volume scheme for solving (1.1) is a formula of the form

$$U_i^{n+1} = U_i^n - \sigma_i(F_{i+1/2-} - F_{i-1/2+}), \tag{1.22}$$

where $\sigma_i = \Delta t^n / \Delta x_i$, telling how to compute the values U_i^{n+1} knowing U_i^n and discretized values z_i of the topography. Here we consider first-order explicit three points schemes where

$$F_{i+1/2-} = \mathcal{F}_l(U_i^n, U_{i+1}^n, z_{i+1} - z_i), \qquad F_{i+1/2+} = \mathcal{F}_r(U_i^n, U_{i+1}^n, z_{i+1} - z_i). \tag{1.23}$$

The functions $\mathcal{F}_{l/r}(U_l, U_r, \Delta z) \in \mathbb{R}^2$ are the numerical fluxes, see [10].

Indeed the method used in [24] in order to solve (1.1) can be viewed as solving

$$\partial_t f + \xi \partial_x f - g(\partial_x z)\partial_\xi f = 0 \tag{1.24}$$

for the unknown $f(t,x,\xi)$, over the time interval (t^n, t^{n+1}), with initial data

$$f(t^n, x, \xi) = M(U^n(x), \xi). \tag{1.25}$$

Defining the update as

$$U_i^{n+1} = \frac{1}{\Delta x_i} \int_{x_{i-1/2}}^{x_{i+1/2}} \int_{\mathbb{R}} \begin{pmatrix} 1 \\ \xi \end{pmatrix} f(t^{n+1-}, x, \xi)\, dx d\xi, \tag{1.26}$$

and

$$f_i^{n+1-}(\xi) = \frac{1}{\Delta x_i} \int_{x_{i-1/2}}^{x_{i+1/2}} f(t^{n+1-}, x, \xi)\, dx, \tag{1.27}$$

the formula (1.26) can then be written

$$U_i^{n+1} = \int_{\mathbb{R}} \begin{pmatrix} 1 \\ \xi \end{pmatrix} f_i^{n+1-}(\xi)\, d\xi. \tag{1.28}$$

This formula can in fact be written under the form (1.22), (1.23) for some numerical fluxes $\mathcal{F}_{l/r}$ computed in [24], involving nonexplicit integrals.

A main idea in this paper is to use simplified formulas, and it will be done by defining a suitable approximation of $f_i^{n+1-}(\xi)$. We shall often denote U_i instead of U_i^n, whenever there is no ambiguity.

2. Kinetic entropy inequality without topography

In this section we consider the problem (1.1) without topography, and the unmodified kinetic scheme (1.24), (1.25), (1.27), (1.28). This problem is classical, and we recall here how the entropy inequality is analyzed in this case, in the fully discrete and semi-discrete cases.

2.1. **Fully discrete scheme.** Without topography, the kinetic scheme is an entropy satisfying *flux vector splitting* scheme [9]. The update (1.27) of the solution of (1.24),(1.25) simplifies to the discrete kinetic scheme

$$f_i^{n+1-} = M_i - \sigma_i \xi \Big(\mathbb{1}_{\xi>0} M_i + \mathbb{1}_{\xi<0} M_{i+1} - \mathbb{1}_{\xi<0} M_i - \mathbb{1}_{\xi>0} M_{i-1} \Big), \tag{2.1}$$

with $\sigma_i = \Delta t^n / \Delta x_i$ and with short notation (we omit the variable ξ). One can write it

$$f_i^{n+1-} = \begin{cases} (1+\sigma_i\xi) M_i - \sigma_i \xi M_{i+1} & \text{if } \xi < 0, \\ (1-\sigma_i\xi) M_i + \sigma_i \xi M_{i-1} & \text{if } \xi > 0. \end{cases} \tag{2.2}$$

Then under the CFL condition that

$$\sigma_i |\xi| \leq 1 \text{ in the supports of } M_i, M_{i-1}, M_{i+1}, \tag{2.3}$$

f_i^{n+1-} is a convex combination of M_i and M_{i+1} if $\xi < 0$, of M_i and M_{i-1} if $\xi > 0$. Thus $f_i^{n+1-} \geq 0$, and recalling the kinetic entropy $H_0(f,\xi)$ from (1.10), we have

$$H_0(f_i^{n+1-},\xi) \leq \begin{cases} (1+\sigma_i\xi) H_0(M_i,\xi) - \sigma_i \xi H_0(M_{i+1},\xi) & \text{if } \xi < 0, \\ (1-\sigma_i\xi) H_0(M_i,\xi) + \sigma_i \xi H_0(M_{i-1},\xi) & \text{if } \xi > 0. \end{cases} \tag{2.4}$$

This can be also written as

$$\begin{aligned} H_0(f_i^{n+1-},\xi) \leq H_0(M_i,\xi) - \sigma_i \xi \Big(& \mathbb{1}_{\xi>0} H_0(M_i,\xi) + \mathbb{1}_{\xi<0} H_0(M_{i+1},\xi) \\ & - \mathbb{1}_{\xi<0} H_0(M_i,\xi) - \mathbb{1}_{\xi>0} H_0(M_{i-1},\xi) \Big), \end{aligned} \tag{2.5}$$

which can be interpreted as a conservative kinetic entropy inequality. Note that with (1.28) and (1.14),

$$\eta(U_i^{n+1}) \leq \int_{\mathbb{R}} H_0(f_i^{n+1-}(\xi),\xi) d\xi, \tag{2.6}$$

which by integration of (2.5) yields the macroscopic entropy inequality.

The scheme (2.1) and the definition (1.28) allow to complete the definition of the macroscopic scheme (1.22), (1.23) with the numerical flux $\mathcal{F}_l = \mathcal{F}_r \equiv \mathcal{F}$ given by the *flux vector splitting* formula [9]

$$\mathcal{F}(U_l,U_r) = \int_{\xi>0} \xi \begin{pmatrix} 1 \\ \xi \end{pmatrix} M(U_l,\xi)\, d\xi + \int_{\xi<0} \xi \begin{pmatrix} 1 \\ \xi \end{pmatrix} M(U_r,\xi)\, d\xi, \tag{2.7}$$

where M is defined in (1.4).

2.2. **Semi-discrete scheme.** Assuming that the timestep is very small (i.e. σ_i very small), we have the linearized approximation of the entropy variation from (2.1)

$$\begin{aligned} H_0(f_i^{n+1-},\xi) \simeq H_0(M_i,\xi) - \sigma_i \xi \partial_f H_0(M_i,\xi) \Big(& \mathbb{1}_{\xi>0} M_i + \mathbb{1}_{\xi<0} M_{i+1} \\ & - \mathbb{1}_{\xi<0} M_i - \mathbb{1}_{\xi>0} M_{i-1} \Big). \end{aligned} \tag{2.8}$$

This linearization with respect to Δt^n (or equivalently with respect to $\sigma_i = \Delta t^n/\Delta x_i$) represents indeed the entropy in the semi-discrete limit $\Delta t^n \to 0$ (divide (2.8) by Δt^n and let formally $\Delta t^n \to 0$). The entropy inequality attached to this linearization can be estimated as follows.

Lemma 2.1. *The linearized term from (2.8) is dominated by the conservative difference from (2.5),*

$$\begin{aligned} &-\sigma_i\xi\partial_f H_0(M_i,\xi)\Big(\mathbb{1}_{\xi>0}M_i + \mathbb{1}_{\xi<0}M_{i+1} - \mathbb{1}_{\xi<0}M_i - \mathbb{1}_{\xi>0}M_{i-1}\Big)\\ &\le -\sigma_i\xi\Big(\mathbb{1}_{\xi>0}H_0(M_i,\xi) + \mathbb{1}_{\xi<0}H_0(M_{i+1},\xi)\\ &\qquad\qquad - \mathbb{1}_{\xi<0}H_0(M_i,\xi) - \mathbb{1}_{\xi>0}H_0(M_{i-1},\xi)\Big). \end{aligned} \tag{2.9}$$

In particular, the semi-discrete scheme is more dissipative than the fully discrete scheme.

Proof. It is enough to prove two inequalities,

$$\xi\partial_f H_0(M_i)(\mathbb{1}_{\xi>0}M_i+\mathbb{1}_{\xi<0}M_{i+1}-M_i) \ge \xi(\mathbb{1}_{\xi>0}H_0(M_i)+\mathbb{1}_{\xi<0}H_0(M_{i+1})-H_0(M_i)) \tag{2.10}$$

and

$$\xi\partial_f H_0(M_i)(\mathbb{1}_{\xi<0}M_i+\mathbb{1}_{\xi>0}M_{i-1}-M_i) \le \xi(\mathbb{1}_{\xi<0}H_0(M_i)+\mathbb{1}_{\xi>0}H_0(M_{i-1})-H_0(M_i)). \tag{2.11}$$

We observe that (2.10) is trivial for $\xi > 0$, and (2.11) is trivial for $\xi < 0$. The two conditions can therefore be written

$$\begin{aligned} &\partial_f H_0(M_i)(M_{i+1} - M_i) \le H_0(M_{i+1}) - H_0(M_i) \quad \text{for } \xi < 0,\\ &\partial_f H_0(M_i)(M_{i-1} - M_i) \le H_0(M_{i-1}) - H_0(M_i) \quad \text{for } \xi > 0. \end{aligned} \tag{2.12}$$

These last inequalities follow from the convexity of H_0. □

3. Kinetic interpretation of the hydrostatic reconstruction scheme

The hydrostatic reconstruction scheme (HR scheme for short) for the Saint-Venant system (1.1), has been introduced in [1], and can be written as follows,

$$U_i^{n+1} = U_i - \sigma_i(F_{i+1/2-} - F_{i-1/2+}), \tag{3.1}$$

where $\sigma_i = \Delta t^n/\Delta x_i$,

$$\begin{aligned} F_{i+1/2-} &= \mathcal{F}(U_{i+1/2-}, U_{i+1/2+}) + \begin{pmatrix} 0 \\ g\frac{h_i^2}{2} - \frac{gh_{i+1/2-}^2}{2} \end{pmatrix},\\ F_{i+1/2+} &= \mathcal{F}(U_{i+1/2-}, U_{i+1/2+}) + \begin{pmatrix} 0 \\ g\frac{h_{i+1}^2}{2} - \frac{gh_{i+1/2+}^2}{2} \end{pmatrix}, \end{aligned} \tag{3.2}$$

$\mathcal{F}$ is a numerical flux for the system without topography, and the reconstructed states

$$U_{i+1/2-} = (h_{i+1/2-}, h_{i+1/2-}u_i), \qquad U_{i+1/2+} = (h_{i+1/2+}, h_{i+1/2+}u_{i+1}), \tag{3.3}$$

are defined by

$$h_{i+1/2-} = (h_i + z_i - z_{i+1/2})_+, \qquad h_{i+1/2+} = (h_{i+1} + z_{i+1} - z_{i+1/2})_+, \tag{3.4}$$

and

$$z_{i+1/2} = \max(z_i, z_{i+1}). \tag{3.5}$$

We would like here to propose a kinetic interpretation of the HR scheme, which means to interpret the above numerical fluxes as averages with respect to the kinetic variable of a scheme written on a kinetic function f. More precisely, we would like to approximate the solution to (1.24) by a kinetic scheme such that the associated macroscopic scheme is exactly (3.1)-(3.2) with homogeneous numerical flux $\mathcal{F}$ given by (2.7). We denote $M_i = M(U_i, \xi)$, $M_{i+1/2\pm} = M(U_{i+1/2\pm}, \xi)$, $f_i^{n+1-} = f_i^{n+1-}(\xi)$, and we consider the scheme

$$
\begin{aligned}
f_i^{n+1-} = M_i - \sigma_i \Big(& \xi \mathbb{1}_{\xi<0} M_{i+1/2+} + \xi \mathbb{1}_{\xi>0} M_{i+1/2-} + \delta M_{i+1/2-} \\
& - \xi \mathbb{1}_{\xi>0} M_{i-1/2-} - \xi \mathbb{1}_{\xi<0} M_{i-1/2+} - \delta M_{i-1/2+} \Big).
\end{aligned} \tag{3.6}
$$

In this formula, $\delta M_{i+1/2\pm}$ depend on ξ, U_i, U_{i+1}, $\Delta z_{i+1/2} = z_{i+1} - z_i$, and are assumed to satisfy the moment relations

$$
\int_{\mathbb{R}} \delta M_{i+1/2-} \, d\xi = 0, \quad \int_{\mathbb{R}} \xi \, \delta M_{i+1/2-} \, d\xi = g \frac{h_i^2}{2} - g \frac{h_{i+1/2-}^2}{2}, \tag{3.7}
$$

$$
\int_{\mathbb{R}} \delta M_{i-1/2+} \, d\xi = 0, \quad \int_{\mathbb{R}} \xi \, \delta M_{i-1/2+} \, d\xi = g \frac{h_i^2}{2} - g \frac{h_{i-1/2+}^2}{2}. \tag{3.8}
$$

Using again (1.28), the integration of (3.6) multiplied by $\begin{pmatrix} 1 \\ \xi \end{pmatrix}$ with respect to ξ then gives the HR scheme (3.1)-(3.2) with (3.3)-(3.5), (2.7). Thus as announced, (3.6) is a kinetic interpretation of the HR scheme. The remainder of this section is devoted to its analysis.

3.1. **Analysis of the semi-discrete scheme.** Assuming that the timestep is very small (i.e. σ_i very small), we have the linearized approximation of the entropy variation from (3.6),

$$
\begin{aligned}
H(f_i^{n+1-}, z_i) \simeq H(M_i, z_i) - \sigma_i \partial_f H(M_i, z_i) \Big(& \xi \mathbb{1}_{\xi<0} M_{i+1/2+} + \xi \mathbb{1}_{\xi>0} M_{i+1/2-} \\
& + \delta M_{i+1/2-} - \xi \mathbb{1}_{\xi>0} M_{i-1/2-} - \xi \mathbb{1}_{\xi<0} M_{i-1/2+} - \delta M_{i-1/2+} \Big),
\end{aligned} \tag{3.9}
$$

where the kinetic entropy $H(f, \xi, z)$ is defined in (1.9). As in Subsection 2.2, this linearization with respect to $\sigma_i = \Delta t^n / \Delta x_i$ represents indeed the entropy in the semi-discrete limit $\Delta t^n \to 0$. Its dissipation can be estimated as follows.

Proposition 3.1. *We assume that the extra variations $\delta M_{i+1/2\pm}$ satisfy (3.7), (3.8), and also*

$$
M(U_i, \xi) = 0 \ \Rightarrow \delta M_{i+1/2-}(\xi) = 0 \text{ and } \delta M_{i-1/2+}(\xi) = 0. \tag{3.10}
$$

Then the linearized term from (3.9) is dominated by a quasi-conservative difference,

$$
\begin{aligned}
& \partial_f H(M_i, z_i) \Big(\xi \mathbb{1}_{\xi<0} M_{i+1/2+} + \xi \mathbb{1}_{\xi>0} M_{i+1/2-} \\
& \qquad + \delta M_{i+1/2-} - \xi \mathbb{1}_{\xi>0} M_{i-1/2-} - \xi \mathbb{1}_{\xi<0} M_{i-1/2+} - \delta M_{i-1/2+} \Big) \\
& \geq \widetilde{H}_{i+1/2-} - \widetilde{H}_{i-1/2+},
\end{aligned} \tag{3.11}
$$

where

$$\begin{aligned}\widetilde{H}_{i+1/2-} &= \xi \mathbb{1}_{\xi<0} H(M_{i+1/2+}, z_{i+1/2}) + \xi \mathbb{1}_{\xi>0} H(M_{i+1/2-}, z_{i+1/2}) \\ &\quad + \xi H(M_i, z_i) - \xi H(M_{i+1/2-}, z_{i+1/2}) \\ &\quad + \Big(\eta'(U_i)\begin{pmatrix}1\\ \xi\end{pmatrix} + g z_i\Big)\big(\xi M_{i+1/2-} - \xi M_i + \delta M_{i+1/2-}\big),\end{aligned} \tag{3.12}$$

$$\begin{aligned}\widetilde{H}_{i-1/2+} &= \xi \mathbb{1}_{\xi<0} H(M_{i-1/2+}, z_{i-1/2}) + \xi \mathbb{1}_{\xi>0} H(M_{i-1/2-}, z_{i-1/2}) \\ &\quad + \xi H(M_i, z_i) - \xi H(M_{i-1/2+}, z_{i-1/2}) \\ &\quad + \Big(\eta'(U_i)\begin{pmatrix}1\\ \xi\end{pmatrix} + g z_i\Big)\big(\xi M_{i-1/2+} - \xi M_i + \delta M_{i-1/2+}\big).\end{aligned} \tag{3.13}$$

Moreover, the integral with respect to ξ of the last two lines of (3.12) (respectively of (3.13)) vanishes. In particular,

$$\int_{\mathbb{R}} \big(\widetilde{H}_{i+1/2-} - \widetilde{H}_{i-1/2+}\big)\, d\xi = \widetilde{G}_{i+1/2} - \widetilde{G}_{i-1/2}, \tag{3.14}$$

with

$$\widetilde{G}_{i+1/2} = \int_{\xi<0} \xi H(M_{i+1/2+}, z_{i+1/2})\, d\xi + \int_{\xi>0} \xi H(M_{i+1/2-}, z_{i+1/2})\, d\xi. \tag{3.15}$$

Proof. The value of the integral with respect to ξ of the two last lines of (3.12) is

$$\begin{aligned}&\big(h_i \frac{u_i^2}{2} + g h_i^2 + g h_i z_i\big) u_i - \big(h_{i+1/2-} \frac{u_i^2}{2} + g h_{i+1/2-}^2 + g h_{i+1/2-} z_{i+1/2}\big) u_i \\ &\quad + (g h_i + g z_i - u_i^2/2) u_i (h_{i+1/2-} - h_i) + u_i^3 (h_{i+1/2-} - h_i) \\ &= u_i g h_{i+1/2-} (-h_{i+1/2-} - z_{i+1/2} + z_i + h_i) \\ &= 0,\end{aligned} \tag{3.16}$$

because of the definition of $h_{i+1/2-}$ in (3.4). The computation for (3.13)) is similar. In order to prove (3.11), it is enough to prove the two inequalities

$$\begin{aligned}&\partial_f H(M_i, z_i)\Big(\xi \mathbb{1}_{\xi<0} M_{i+1/2+} + \xi \mathbb{1}_{\xi>0} M_{i+1/2-} + \delta M_{i+1/2-} - \xi M_i\Big) \\ &\geq \widetilde{H}_{i+1/2-} - \xi H(M_i, z_i),\end{aligned} \tag{3.17}$$

and

$$\begin{aligned}&\partial_f H(M_i, z_i)\Big(\xi \mathbb{1}_{\xi>0} M_{i-1/2-} + \xi \mathbb{1}_{\xi<0} M_{i-1/2+} + \delta M_{i-1/2+} - \xi M_i\Big) \\ &\leq \widetilde{H}_{i-1/2+} - \xi H(M_i, z_i).\end{aligned} \tag{3.18}$$

We note that the definitions of $h_{i+1/2\pm}$ in (3.4)-(3.5) ensure that $h_{i+1/2-} \leq h_i$, and $h_{i+1/2+} \leq h_{i+1}$. Therefore, because of (1.4) one has

$$0 \leq M_{i+1/2-} \leq M_i, \quad 0 \leq M_{i+1/2+} \leq M_{i+1}, \tag{3.19}$$

thus

$$M(U_i, \xi) = 0 \;\Rightarrow M(U_{i+1/2-}, \xi) = 0 \text{ and } M(U_{i-1/2+}, \xi) = 0. \tag{3.20}$$

Taking into account (3.10), with (1.19) we get

$$\begin{aligned}&\Big(\eta'(U_i)\begin{pmatrix}1\\ \xi\end{pmatrix} + g z_i\Big)\big(\xi M_{i+1/2-} - \xi M_i + \delta M_{i+1/2-}\big) \\ &= \partial_f H(M_i, z_i)\big(\xi M_{i+1/2-} - \xi M_i + \delta M_{i+1/2-}\big),\end{aligned} \tag{3.21}$$

and

$$\begin{aligned}(3.22)\qquad &\Big(\eta'(U_i)\begin{pmatrix}1\\ \xi\end{pmatrix}+gz_i\Big)\big(\xi M_{i-1/2+}-\xi M_i+\delta M_{i-1/2+}\big)\\ &=\partial_f H(M_i,z_i)\big(\xi M_{i-1/2+}-\xi M_i+\delta M_{i-1/2+}\big).\end{aligned}$$

Therefore, the inequalities (3.17)-(3.18) simplify to

(3.23)
$$\begin{aligned}&\partial_f H(M_i,z_i)\Big(\xi\mathbb{1}_{\xi<0}M_{i+1/2+}+\xi\mathbb{1}_{\xi>0}M_{i+1/2-}-\xi M_{i+1/2-}\Big)\\ &\geq \xi\mathbb{1}_{\xi<0}H(M_{i+1/2+},z_{i+1/2})+\xi\mathbb{1}_{\xi>0}H(M_{i+1/2-},z_{i+1/2})-\xi H(M_{i+1/2-},z_{i+1/2}),\end{aligned}$$

(3.24)
$$\begin{aligned}&\partial_f H(M_i,z_i)\Big(\xi\mathbb{1}_{\xi>0}M_{i-1/2-}+\xi\mathbb{1}_{\xi<0}M_{i-1/2+}-\xi M_{i-1/2+}\Big)\\ &\leq \xi\mathbb{1}_{\xi<0}H(M_{i-1/2+},z_{i-1/2})+\xi\mathbb{1}_{\xi>0}H(M_{i-1/2-},z_{i-1/2})-\xi H(M_{i-1/2+},z_{i-1/2}).\end{aligned}$$

The first inequality (3.23) is trivial for $\xi>0$, and the second inequality (3.24) is trivial for $\xi<0$. Therefore it is enough to satisfy the two inequalities

(3.25)
$$\partial_f H(M_i,z_i)\Big(M_{i+1/2+}-M_{i+1/2-}\Big)\leq H(M_{i+1/2+},z_{i+1/2})-H(M_{i+1/2-},z_{i+1/2}),$$

(3.26)
$$\partial_f H(M_i,z_i)\Big(M_{i-1/2-}-M_{i-1/2+}\Big)\leq H(M_{i-1/2-},z_{i-1/2})-H(M_{i-1/2+},z_{i-1/2}).$$

But as in Subsection 2.2, we have according to the convexity of H with respect to f,

$$\begin{aligned}(3.27)\qquad H(M_{i+1/2+},z_{i+1/2})\geq\ & H(M_{i+1/2-},z_{i+1/2})\\ &+\partial_f H(M_{i+1/2-},z_{i+1/2})(M_{i+1/2+}-M_{i+1/2-}),\end{aligned}$$

$$\begin{aligned}(3.28)\qquad H(M_{i-1/2-},z_{i-1/2})\geq\ & H(M_{i-1/2+},z_{i-1/2})\\ &+\partial_f H(M_{i-1/2+},z_{i-1/2})(M_{i-1/2-}-M_{i-1/2+}).\end{aligned}$$

In order to prove (3.25), we observe that if $M_i(\xi)=0$ then $M_{i+1/2-}(\xi)=0$ also, thus $\partial_f H(M_{i+1/2-},z_{i+1/2})-\partial_f H(M_i,z_i)=g(z_{i+1/2}-z_i)\geq 0$ because of (3.5), and the inequality (3.25) follows from (3.27). Next, if $M_i(\xi)>0$, one has

$$\begin{aligned}(3.29)\qquad &\partial_f H(M_i,z_i)(M_{i+1/2+}-M_{i+1/2-})\\ &=\big(\eta'(U_i)\begin{pmatrix}1\\ \xi\end{pmatrix}+gz_i\big)(M_{i+1/2+}-M_{i+1/2-}),\end{aligned}$$

and as in (1.20)

$$\begin{aligned}(3.30)\qquad &\partial_f H(M_{i+1/2-},z_{i+1/2})(M_{i+1/2+}-M_{i+1/2-})\\ &\geq\big(\eta'(U_{i+1/2-})\begin{pmatrix}1\\ \xi\end{pmatrix}+gz_{i+1/2}\big)(M_{i+1/2+}-M_{i+1/2-}).\end{aligned}$$

Taking the difference between (3.30) and (3.29), we obtain

(3.31)
$$\begin{aligned}&\partial_f H(M_{i+1/2-},z_{i+1/2})(M_{i+1/2+}-M_{i+1/2-})-\partial_f H(M_i,z_i)(M_{i+1/2+}-M_{i+1/2-})\\ &\geq\big(gh_{i+1/2-}-gh_i+gz_{i+1/2}-gz_i\big)(M_{i+1/2+}-M_{i+1/2-})\geq 0,\end{aligned}$$

because of the definition (3.4) of $h_{i+1/2-}$. Therefore we conclude that in any case ($M_i(\xi)$ being zero or not), one has

(3.32)
$$\begin{aligned}
&\partial_f H(M_i,z_i)(M_{i+1/2+}-M_{i+1/2-}) - H(M_{i+1/2+},z_{i+1/2}) + H(M_{i+1/2-},z_{i+1/2})\\
&\le H(M_{i+1/2-},z_{i+1/2}) - H(M_{i+1/2+},z_{i+1/2})\\
&\qquad\qquad +\partial_f H(M_{i+1/2-},z_{i+1/2})(M_{i+1/2+}-M_{i+1/2-})\\
&\le 0
\end{aligned}$$

because of (3.27), and this proves (3.25). Similarly one gets

(3.33)
$$\begin{aligned}
&\partial_f H(M_i,z_i)(M_{i-1/2-}-M_{i-1/2+}) - H(M_{i-1/2-},z_{i-1/2}) + H(M_{i-1/2+},z_{i-1/2})\\
&\le H(M_{i-1/2+},z_{i-1/2}) - H(M_{i-1/2-},z_{i-1/2})\\
&\qquad\qquad +\partial_f H(M_{i-1/2+},z_{i-1/2})(M_{i-1/2-}-M_{i-1/2+})\\
&\le 0,
\end{aligned}$$

proving (3.26). This concludes the proof, and we observe that we have indeed a dissipation estimate slightly stronger than (3.11),

$$\begin{aligned}
&\partial_f H(M_i,z_i)\Big(\xi\mathbb{1}_{\xi<0}M_{i+1/2+}+\xi\mathbb{1}_{\xi>0}M_{i+1/2-}\\
&\qquad +\delta M_{i+1/2-}-\xi\mathbb{1}_{\xi>0}M_{i-1/2-}-\xi\mathbb{1}_{\xi<0}M_{i-1/2+}-\delta M_{i-1/2+}\Big)\\
&\ge \widetilde{H}_{i+1/2-}-\widetilde{H}_{i-1/2+}\\
&-\xi\mathbb{1}_{\xi<0}\Big(H(M_{i+1/2+},z_{i+1/2})-H(M_{i+1/2-},z_{i+1/2})\\
&\qquad -\partial_f H(M_{i+1/2-},z_{i+1/2})(M_{i+1/2+}-M_{i+1/2-})\Big)\\
&+\xi\mathbb{1}_{\xi>0}\Big(H(M_{i-1/2-},z_{i-1/2})-H(M_{i-1/2+},z_{i-1/2})\\
&\qquad -\partial_f H(M_{i-1/2+},z_{i-1/2})(M_{i-1/2-}-M_{i-1/2+})\Big).
\end{aligned}\tag{3.34}$$

□

Remark 3.2. The numerical entropy flux (3.15) can be written

$$\widetilde{G}_{i+1/2}=\mathcal{G}(U_{i+1/2-},U_{i+1/2+})+gz_{i+1/2}\mathcal{F}^0(U_{i+1/2-},U_{i+1/2+}),\tag{3.35}$$

where $\mathcal{G}$ is the numerical entropy flux of the scheme without topography, and $\mathcal{F}^0$ is the first component of $\mathcal{F}$. This formula is in accordance of the analysis of the semi-discrete entropy inequality in [1].

Remark 3.3. At the kinetic level, the entropy inequality (3.11) is not in conservative form. The entropy inequality becomes conservative only when taking the integral with respect to ξ, as is seen on (3.14). This is also the case in [24]. Indeed we have written the macroscopic conservative entropy inequality as an integral with respect to ξ of the sum of a nonpositive term (the one in (3.11)), a kinetic conservative term (the difference of the first lines of (3.12) and (3.13)), and a term with vanishing integral (difference of the two last lines of (3.12) and (3.13)). However, such a decomposition is not unique.

3.2. **Analysis of the fully discrete scheme.** We still consider the scheme (3.6), and we make the choice

$$\begin{aligned}
\delta M_{i+1/2-}&=(\xi-u_i)(M_i-M_{i+1/2-}),\\
\delta M_{i-1/2+}&=(\xi-u_i)(M_i-M_{i-1/2+}),
\end{aligned}\tag{3.36}$$

that satisfies the assumptions (3.7), (3.8) and (3.10). The scheme (3.6) is therefore a kinetic interpretation of the HR scheme (3.1)-(3.5).

Lemma 3.4. *The scheme (3.6) with the choice (3.36) is "kinetic well-balanced" for steady states at rest, and consistent with (1.24).*

Proof. The expression kinetic well-balanced means that we do not only prove that

$$\int_{\mathbb{R}} \begin{pmatrix} 1 \\ \xi \end{pmatrix} f_i^{n+1-} \, d\xi = \int_{\mathbb{R}} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M_i \, d\xi, \tag{3.37}$$

at rest, but the stronger property

$$f_i^{n+1-}(\xi) = M_i(\xi), \quad \forall \xi \in \mathbb{R}, \tag{3.38}$$

when $u_i = 0$ and $h_i + z_i = h_{i+1} + z_{i+1}$ for all i. Indeed in this situation one has $U_{i+1/2-} = U_{i+1/2+}$ for all i, thus the first three terms between parentheses in (3.6) give ξM_i, and the last three terms give $-\xi M_i$, leading to (3.38).

The consistency of the HR scheme has been proved in [1], but here the statement is the consistency of the kinetic update (3.6) with the kinetic equation (1.24). We proceed as follows. Using (1.25) and (1.4), the topography source term in (1.24) reads

$$-g(\partial_x z)\partial_\xi M = g(\partial_x z)\frac{\xi - u}{2gh - (\xi - u)^2} M. \tag{3.39}$$

This formula is valid for $2gh - (\xi - u)^2 \neq 0$, i.e. when $\xi \neq u \pm \sqrt{2gh}$ or in $L^1(\xi \in \mathbb{R})$. Assuming that $h_i > 0$ (otherwise the consistency is obvious), one has that $h_{i+1/2-} = h_i + z_i - z_{i+1/2}$ for $z_{i+1} - z_i$ small enough, and an asymptotic expansion of $M_{i+1/2-}$ gives

$$M_{i+1/2-} = M_i + (z_i - z_{i+1/2})(\partial_{h_i} M_i)_{|u_i} + o(z_{i+1} - z_i), \tag{3.40}$$

with

$$(\partial_{h_i} M_i)_{|u_i} = g\frac{M_i}{2gh_i - (\xi - u_i)^2}. \tag{3.41}$$

Thus

$$\frac{\delta M_{i+1/2-}}{\Delta x_i} = g\frac{z_{i+1/2} - z_i}{\Delta x_i}\frac{\xi - u_i}{2gh_i - (\xi - u_i)^2} M_i + o(1). \tag{3.42}$$

Similarly, one has

$$\frac{\delta M_{i-1/2+}}{\Delta x_i} = g\frac{z_{i-1/2} - z_i}{\Delta x_i}\frac{\xi - u_i}{2gh_i - (\xi - u_i)^2} M_i + o(1). \tag{3.43}$$

With the usual shift of index i due to the distribution of the source to interfaces, the difference (3.42) minus (3.43) appears as a discrete version of (3.39). The other four terms in parentheses in (3.6) are conservative, and are classically consistent with $\xi \partial_x f$ in (1.24). □

Remark 3.5. The scheme (3.6) can be viewed as a consistent well-balanced scheme for (1.24), except that the notion of consistency is true here only for Maxwellian initial data. On the contrary, the exact solution used in [24] is consistent for initial data of arbitrary shape. The role of the special form of the Maxwellian (1.4) is seen here by the fact that for initial data U_i at rest, one has that $M(U_i, \xi)$ is a steady state of (1.24) (this results from (3.39) and (3.41)).

When writing the entropy inequality for the fully discrete scheme, the difficulty is to estimate the positive part of the entropy dissipation by something that tends to zero when Δx_i tends to zero, at constant Courant number σ_i, and assuming only that $\Delta z/\Delta x$ is bounded (Lipschitz topography), but not that $\Delta U/\Delta x$ is bounded (the solution can have discontinuities). Here Δz stands for a quantity like $z_{i+1}-z_i$, and ΔU stands for a quantity like $U_{i+1}-U_i$.

The principle of proof of such entropy inequality is that we use the dissipation of the semi-discrete scheme proved in Proposition 3.1, under the strong form (3.34). This inequality involves the terms linear in σ_i. Under a CFL condition, the higher order terms (quadratic in σ_i or higher) are either treated as errors if they are of the order of Δz^2 or $\Delta z\Delta U$, or must be dominated by the dissipation if they are of the order of ΔU^2. Note that the dissipation in (3.34), i.e. the two last expressions in factor of $\mathbb{1}_{\xi<0}$ and $\mathbb{1}_{\xi>0}$ respectively, are of the order of $(M_{i+1/2+}-M_{i+1/2-})^2$ and $(M_{i-1/2+}-M_{i-1/2-})^2$ respectively, and thus neglecting the terms in Δz, they control $(M_{i+1}-M_i)^2$ and $(M_i-M_{i-1})^2$ respectively. However, the Maxwellian (1.4) is not Lipschitz continuous with respect to U, thus a sharp analysis has to be performed in order to use the dissipation.

We consider a velocity $v_m \geq 0$ such that for all i,

$$M(U_i,\xi)>0 \Rightarrow |\xi|\leq v_m. \tag{3.44}$$

This means equivalently that $|u_i|+\sqrt{2gh_i}\leq v_m$. We consider a CFL condition strictly less than one,

$$\sigma_i v_m \leq \beta < 1 \quad \text{for all } i, \tag{3.45}$$

where $\sigma_i=\Delta t^n/\Delta x_i$, and β is a given constant.

Theorem 3.6. *Under the CFL condition (3.45), the scheme (3.6) with the choice (3.36) verifies the following properties.*
(i) The kinetic function remains nonnegative $f_i^{n+1-}\geq 0$.
(ii) One has the kinetic entropy inequality

$$\begin{aligned}
&H(f_i^{n+1-},z_i)\\
&\leq H(M_i,z_i)-\sigma_i\Big(\widetilde{H}_{i+1/2-}-\widetilde{H}_{i-1/2+}\Big)\\
&\quad-\nu_\beta\,\sigma_i|\xi|\frac{g^2\pi^2}{6}\bigg(\mathbb{1}_{\xi<0}\,(M_{i+1/2+}+M_{i+1/2-})(M_{i+1/2+}-M_{i+1/2-})^2\\
&\quad+\mathbb{1}_{\xi>0}\,(M_{i-1/2-}+M_{i-1/2+})(M_{i-1/2+}-M_{i-1/2-})^2\bigg)\\
&\quad+C_\beta(\sigma_i v_m)^2\frac{g^2\pi^2}{6}M_i\Big((M_i-M_{i+1/2-})^2+(M_i-M_{i-1/2+})^2\Big),
\end{aligned} \tag{3.46}$$

where $\widetilde{H}_{i+1/2-}$, $\widetilde{H}_{i-1/2+}$ are defined by (3.12),(3.13), $\nu_\beta>0$ is a dissipation constant depending only on β, and $C_\beta\geq 0$ is a constant depending only on β. The term proportional to C_β is an error, while the term proportional to ν_β is a dissipation that reinforces the inequality.

Theorem 3.6 has the following corollary.

Corollary 3.7. *Under the CFL condition (3.44), (3.45), integrating the estimate (3.46) with respect to ξ, using (1.14), (1.28), (3.14) (neglecting the dissipation*

proportional to ν_β) and Lemma 3.11 yields that

(3.47)
$$\eta(U_i^{n+1}) + g z_i h_i^{n+1} \leq \eta(U_i) + g z_i h_i - \sigma_i\Big(\widetilde{G}_{i+1/2} - \widetilde{G}_{i-1/2}\Big) + C_\beta(\sigma_i v_m)^2\bigg(g(h_i - h_{i+1/2-})^2 + g(h_i - h_{i-1/2+})^2\bigg),$$

where $\widetilde{G}_{i+1/2}$ is defined in (3.15) or equivalently (3.35), and $C_\beta \geq 0$ depends only on β. This is the discrete entropy inequality associated to the HR scheme (3.1)-(3.5) with kinetic homogeneous numerical flux (2.7). With (3.3)-(3.5) one has

$$0 \leq h_i - h_{i+1/2-} \leq |z_{i+1} - z_i|, \quad 0 \leq h_i - h_{i-1/2+} \leq |z_i - z_{i-1}|. \tag{3.48}$$

We conclude that the quadratic error terms proportional to C_β in the right-hand side of (3.47) (divide (3.47) by Δt^n to be consistent with (1.2)) has the following key properties: it vanishes identically when $z = cst$ (no topography) or when $\sigma_i \to 0$ (semi-discrete limit), and as soon as the topography is Lipschitz continuous, it tends to zero strongly when the grid size tends to 0 (consistency with the continuous entropy inequality (1.2)), even if the solution contains shocks.

We state now a counter result saying that it is not possible to remove the error term in (3.47). It is indeed true for the HR scheme even if the homogeneous flux used is not the kinetic one.

Proposition 3.8. *The HR scheme (3.1)-(3.5) does not satisfy the fully-discrete entropy inequality (3.47) without quadratic error term, whatever restrictive is the CFL condition.*

Proof of Theorem 3.6. Using (3.6) and (3.36), one has for $\xi \leq 0$

(3.49)
$$\begin{aligned} f_i^{n+1-} &= M_i - \sigma_i\Big(\xi M_{i+1/2+} - \xi M_{i-1/2+} + (\xi - u_i)(M_{i-1/2+} - M_{i+1/2-})\Big) \\ &= M_i - \sigma_i\Big(\xi(M_{i+1/2+} - M_{i+1/2-}) + u_i(M_{i+1/2-} - M_{i-1/2+})\Big), \end{aligned}$$

while for $\xi \geq 0$,

(3.50)
$$\begin{aligned} f_i^{n+1-} &= M_i - \sigma_i\Big(\xi M_{i+1/2-} - \xi M_{i-1/2-} + (\xi - u_i)(M_{i-1/2+} - M_{i+1/2-})\Big) \\ &= M_i - \sigma_i\Big(\xi(M_{i-1/2+} - M_{i-1/2-}) + u_i(M_{i+1/2-} - M_{i-1/2+})\Big). \end{aligned}$$

But because of (3.19), one has $0 \leq M_{i+1/2-}, M_{i-1/2+} \leq M_i$. Thus for all ξ we get from (3.49)-(3.50) that $f_i^{n+1-} \geq (1 - \sigma_i(|u_i| + |\xi - u_i|))M_i \geq 0$ under the CFL condition (3.45), proving (i).

Then, we write the linearization of H around the Maxwellian M_i

$$H(f_i^{n+1-}, z_i) = H(M_i, z_i) + \partial_f H(M_i, z_i)\big(f_i^{n+1-} - M_i\big) + L_i, \tag{3.51}$$

where L_i is a remainder. The linearized term $\partial_f H(M_i, z_i)\big(f_i^{n+1-} - M_i\big)$ in (3.51) is nothing but the dissipation of the semi-discrete scheme, that has been estimated in Proposition 3.1. Thus, multiplying (3.34) by $-\sigma_i$, using the form (1.9) of H and the identity

$$b^3 - a^3 - 3a^2(b-a) = (b+2a)(b-a)^2, \tag{3.52}$$

we get

$$
\begin{aligned}
&\partial_f H(M_i,z_i)\big(f_i^{n+1-}-M_i\big)\\
&\le -\sigma_i\big(\widetilde{H}_{i+1/2-}-\widetilde{H}_{i-1/2+}\big)\\
&+\sigma_i\xi\mathbb{I}_{\xi<0}\frac{g^2\pi^2}{6}\big(M_{i+1/2+}+2M_{i+1/2-}\big)\big(M_{i+1/2+}-M_{i+1/2-}\big)^2\\
&-\sigma_i\xi\mathbb{I}_{\xi>0}\frac{g^2\pi^2}{6}\big(M_{i-1/2-}+2M_{i-1/2+}\big)\big(M_{i-1/2-}-M_{i-1/2+}\big)^2.
\end{aligned}
\tag{3.53}
$$

Then, using again the form of H and (3.52), the quadratic term L_i in (3.51) can be expressed as

$$
L_i=\frac{g^2\pi^2}{6}(2M_i+f_i^{n+1-})(f_i^{n+1-}-M_i)^2. \tag{3.54}
$$

We notice that in (3.51), the time variation of the kinetic entropy H is estimated by a term linearized in Δt^n, that is itself estimated in (3.53) by a space integrated-conservative difference and nonpositive dissipations, and nonnegative errors L_i which are merely quadratic in Δt^n. These errors L_i do not vanish when the topography is constant, and moreover do not tend to zero strongly for discontinuous data U. The remainder of the argument is to prove that under a CFL condition, the quadratic terms L_i are dominated by the dissipation terms, up to errors that are directly estimated in terms of the variations of the topography z.

Using (3.49), we have for any $\alpha>0$

$$
\begin{aligned}
L_i\le\frac{g^2\pi^2}{6}\sigma_i^2(2M_i+f_i^{n+1-})\Big((1+\alpha)\xi^2\big(M_{i+1/2+}-M_{i+1/2-}\big)^2\\
+(1+1/\alpha)u_i^2\big(M_{i+1/2-}-M_{i-1/2+}\big)^2\Big),\quad\text{for all }\xi\le 0,
\end{aligned}
\tag{3.55}
$$

and similarly with (3.50)

$$
\begin{aligned}
L_i\le\frac{g^2\pi^2}{6}\sigma_i^2(2M_i+f_i^{n+1-})\Big((1+\alpha)\xi^2\big(M_{i-1/2+}-M_{i-1/2-}\big)^2\\
+(1+1/\alpha)u_i^2\big(M_{i+1/2-}-M_{i-1/2+}\big)^2\Big),\quad\text{for all }\xi\ge 0.
\end{aligned}
\tag{3.56}
$$

Therefore, adding the estimates (3.51), (3.53), (3.55), (3.56) yields

$$
H(f_i^{n+1-},z_i)\le H(M_i,z_i)-\sigma_i\Big(\widetilde{H}_{i+1/2-}-\widetilde{H}_{i-1/2+}\Big)+d_i, \tag{3.57}
$$

where

$$
\begin{aligned}
d_i=\sigma_i\xi\mathbb{I}_{\xi<0}\frac{g^2\pi^2}{6}\big(M_{i+1/2+}+2M_{i+1/2-}+(1+\alpha)\sigma_i\xi(2M_i+f_i^{n+1-})\big)\\
\times(M_{i+1/2+}-M_{i+1/2-})^2\\
-\sigma_i\xi\mathbb{I}_{\xi>0}\frac{g^2\pi^2}{6}\big(M_{i-1/2-}+2M_{i-1/2+}-(1+\alpha)\sigma_i\xi(2M_i+f_i^{n+1-})\big)\\
\times(M_{i-1/2+}-M_{i-1/2-})^2\\
+\sigma_i^2u_i^2\frac{g^2\pi^2}{6}(1+1/\alpha)(2M_i+f_i^{n+1-})(M_{i+1/2-}-M_{i-1/2+})^2,
\end{aligned}
\tag{3.58}
$$

and $\alpha>0$ is an arbitrary parameter. The first two lines in (3.58) are generically nonpositive for σ_i small enough (recall the bound (3.44) on ξ), whereas the third line is nonnegative.

Before going further in the proof of Theorem 3.6, i.e. upper bounding d_i by a sum of a dissipation term and an error, let us state a lemma, that gives another expression for d_i, in which the nonpositive contributions appear clearly.

Lemma 3.9. *The term d_i from (3.58) can also be written*

$$\begin{aligned} d_i &= \sigma_i\xi \mathbb{I}_{\xi<0}\,\gamma^-_{i+1/2}(M_{i+1/2+}-M_{i+1/2-})^2 \\ &\quad -\sigma_i\xi \mathbb{I}_{\xi>0}\,\gamma^+_{i-1/2}(M_{i-1/2+}-M_{i-1/2-})^2 \\ &\quad +\sigma_i^2\frac{g^2\pi^2}{6}\Big((1+1/\alpha)u_i^2(2M_i+f_i^{n+1-})(M_{i+1/2-}-M_{i-1/2+})^2 \\ &\qquad +(1+\alpha)\xi^2\big(\mathbb{I}_{\xi<0}\,\mu^-_{i+1/2}+\mathbb{I}_{\xi>0}\,\mu^+_{i-1/2}\big)\Big), \end{aligned} \tag{3.59}$$

with

$$\begin{aligned} \gamma^-_{i+1/2} &= \frac{g^2\pi^2}{6}\Big(\big(1-(1+\alpha)(\sigma_i\xi)^2\big)M_{i+1/2+} \\ &\qquad +\Big(2+(1+\alpha)(\sigma_i\xi)^2+3(1+\alpha)\sigma_i\xi\Big)M_{i+1/2-}\Big), \\ \gamma^+_{i-1/2} &= \frac{g^2\pi^2}{6}\Big(\big(1-(1+\alpha)(\sigma_i\xi)^2\big)M_{i-1/2-} \\ &\qquad +\Big(2+(1+\alpha)(\sigma_i\xi)^2-3(1+\alpha)\sigma_i\xi\Big)M_{i-1/2+}\Big), \end{aligned} \tag{3.60}$$

$$\begin{aligned} \mu^-_{i+1/2} &= (M_{i+1/2+}-M_{i+1/2-})^2\Big(3(M_i-M_{i+1/2-}) \\ &\qquad -\sigma_i u_i(M_{i+1/2-}-M_{i-1/2+})\Big), \\ \mu^+_{i-1/2} &= (M_{i-1/2+}-M_{i-1/2-})^2\Big(3(M_i-M_{i-1/2+}) \\ &\qquad -\sigma_i u_i(M_{i+1/2-}-M_{i-1/2+})\Big). \end{aligned} \tag{3.61}$$

Proof of Lemma 3.9. The expression (3.49) of f_i^{n+1-} for $\xi\le 0$ allows to precise the value of d_i in (3.58), and gives for $\xi\le 0$

$$\begin{aligned} & M_{i+1/2+}+2M_{i+1/2-}+(1+\alpha)\sigma_i\xi(2M_i+f_i^{n+1-}) \\ = \;& (1-(1+\alpha)(\sigma_i\xi)^2)M_{i+1/2+}+(2+(1+\alpha)(\sigma_i\xi)^2)M_{i+1/2-} \\ & +(1+\alpha)\sigma_i\xi\left(3M_i-\sigma_i u_i(M_{i+1/2-}-M_{i-1/2+})\right) \\ = \;& (1-(1+\alpha)(\sigma_i\xi)^2)M_{i+1/2+}+(2+(1+\alpha)(\sigma_i\xi)^2+3(1+\alpha)\sigma_i\xi)M_{i+1/2-} \\ & +(1+\alpha)\sigma_i\xi\left(3(M_i-M_{i+1/2-})-\sigma_i u_i(M_{i+1/2-}-M_{i-1/2+})\right). \end{aligned}$$

Using (3.50) we obtain analogously for $\xi\ge 0$

$$\begin{aligned} & M_{i-1/2-}+2M_{i-1/2+}-(1+\alpha)\sigma_i\xi(2M_i+f_i^{n+1-}) \\ = \;& (1-(1+\alpha)(\sigma_i\xi)^2)M_{i-1/2-}+(2+(1+\alpha)(\sigma_i\xi)^2)M_{i-1/2+} \\ & -(1+\alpha)\sigma_i\xi\left(3M_i-\sigma_i u_i(M_{i+1/2-}-M_{i-1/2+})\right) \\ = \;& (1-(1+\alpha)(\sigma_i\xi)^2)M_{i-1/2-}+(2+(1+\alpha)(\sigma_i\xi)^2-3(1+\alpha)\sigma_i\xi)M_{i-1/2+} \\ & -(1+\alpha)\sigma_i\xi\left(3(M_i-M_{i-1/2+})-\sigma_i u_i(M_{i+1/2-}-M_{i-1/2+})\right). \end{aligned}$$

These expressions yield the formulas (3.59)-(3.61). □

Continuation of the proof of Theorem 3.6. One would like the first two lines of (3.59) to be nonpositive. In order to get nonnegative coefficients $\gamma^-_{i+1/2}$, $\gamma^+_{i-1/2}$

in (3.59), it is enough that

$$1-(1+\alpha)(\sigma_i|\xi|)^2\geq 0,\quad 2+(1+\alpha)(\sigma_i|\xi|)^2-3(1+\alpha)\sigma_i|\xi|\geq 0, \tag{3.62}$$

for all ξ in the supports of M_{i-1}, M_i, M_{i+1}. But since both expressions in (3.62) are decreasing with respect to $|\xi|$ for $\sigma_i|\xi|\leq 1$ and because of the CFL condition (3.45), they are lower bounded respectively by

$$1-(1+\alpha)\beta^2,\quad 2+(1+\alpha)\beta^2-3(1+\alpha)\beta. \tag{3.63}$$

But since $\beta<1$, one can choose $\alpha>0$ such that

$$1+\alpha<\frac{2}{\beta(3-\beta)}, \tag{3.64}$$

and then the coefficients (3.63) are positive, and $\gamma^-_{i+1/2},\gamma^+_{i-1/2}\geq 0$. We denote

$$c_{\alpha,\beta}=\min\Big(1-(1+\alpha)\beta^2,2+(1+\alpha)\beta^2-3(1+\alpha)\beta\Big)>0. \tag{3.65}$$

Then we have

$$\mathbb{1}_{\xi<0}\gamma^-_{i+1/2}\geq \mathbb{1}_{\xi<0}\frac{g^2\pi^2}{6}c_{\alpha,\beta}(M_{i+1/2+}+M_{i+1/2-}), \tag{3.66}$$

and

$$\mathbb{1}_{\xi>0}\gamma^+_{i-1/2}\geq \mathbb{1}_{\xi>0}\frac{g^2\pi^2}{6}c_{\alpha,\beta}(M_{i-1/2-}+M_{i-1/2+}). \tag{3.67}$$

Next we write using (3.49), (3.50) and (3.19)

(3.68)
$$\begin{aligned}
&2M_i+f_i^{n+1-}\\
&\leq 3M_i-\sigma_i\xi\mathbb{1}_{\xi<0}(M_{i+1/2+}-M_{i+1/2-})_+\\
&\quad+\sigma_i\xi\mathbb{1}_{\xi>0}(M_{i-1/2-}-M_{i-1/2+})_+ +\sigma_i|u_i||M_{i+1/2-}-M_{i-1/2+}|\\
&\leq 4M_i-\sigma_i\xi\mathbb{1}_{\xi<0}(M_{i+1/2+}-M_{i+1/2-})_+ +\sigma_i\xi\mathbb{1}_{\xi>0}(M_{i-1/2-}-M_{i-1/2+})_+.
\end{aligned}$$

We can estimate the first quadratic error term from (3.59) as

$$\begin{aligned}
&(2M_i+f_i^{n+1-})(M_{i+1/2-}-M_{i-1/2+})^2\\
&\leq 4M_i(M_{i+1/2-}-M_{i-1/2+})^2\\
&\quad-\sigma_i\xi\mathbb{1}_{\xi<0}M_i|M_{i+1/2+}-M_{i+1/2-}||M_{i+1/2-}-M_{i-1/2+}|\\
&\quad+\sigma_i\xi\mathbb{1}_{\xi>0}M_i|M_{i-1/2-}-M_{i-1/2+}||M_{i+1/2-}-M_{i-1/2+}|.
\end{aligned} \tag{3.69}$$

Finally we estimate

$$\begin{aligned}
&|\mu^-_{i+1/2}|\\
&\leq 4(M_{i+1/2+}-M_{i+1/2-})^2\big(|M_i-M_{i+1/2-}|+|M_i-M_{i-1/2+}|\big)\\
&\leq 2|M_{i+1/2+}-M_{i+1/2-}|\Big(\epsilon(M_{i+1/2+}-M_{i+1/2-})^2\\
&\qquad\qquad+\epsilon^{-1}\big(|M_i-M_{i+1/2-}|+|M_i-M_{i-1/2+}|\big)^2\Big)\\
&\leq 2\epsilon(M_{i+1/2+}+M_{i+1/2-})(M_{i+1/2+}-M_{i+1/2-})^2\\
&\quad+4\epsilon^{-1}M_i|M_{i+1/2+}-M_{i+1/2-}|\big(|M_i-M_{i+1/2-}|+|M_i-M_{i-1/2+}|\big),
\end{aligned} \tag{3.70}$$

and similarly

$$\begin{aligned}
&|\mu^+_{i-1/2}|\\
&\leq 2\epsilon(M_{i-1/2-}+M_{i-1/2+})(M_{i-1/2+}-M_{i-1/2-})^2\\
&\quad+4\epsilon^{-1}M_i|M_{i-1/2+}-M_{i-1/2-}|\big(|M_i-M_{i+1/2-}|+|M_i-M_{i-1/2+}|\big),
\end{aligned} \tag{3.71}$$

where $\epsilon > 0$ is arbitrary. Putting together in (3.59) the estimates (3.66), (3.67), (3.70), (3.71), we get

$$
\begin{aligned}
d_i \leq\ & \sigma_i \xi \mathbb{1}_{\xi<0} \frac{g^2\pi^2}{6} \big(c_{\alpha,\beta} - 2\epsilon(1+\alpha)\sigma_i|\xi|\big) \\
& \qquad \times (M_{i+1/2+} + M_{i+1/2-})(M_{i+1/2+} - M_{i+1/2-})^2 \\
& - \sigma_i \xi \mathbb{1}_{\xi>0} \frac{g^2\pi^2}{6} \big(c_{\alpha,\beta} - 2\epsilon(1+\alpha)\sigma_i|\xi|\big) \\
& \qquad \times (M_{i-1/2-} + M_{i-1/2+})(M_{i-1/2+} - M_{i-1/2-})^2 \\
& + \sigma_i^2 \frac{g^2\pi^2}{6} \bigg((1+1/\alpha) u_i^2 (2M_i + f_i^{n+1-})(M_{i+1/2-} - M_{i-1/2+})^2 \\
& \quad + 4\epsilon^{-1}(1+\alpha)\xi^2 M_i \big(|M_i - M_{i+1/2-}| + |M_i - M_{i-1/2+}| \big) \\
& \qquad \times \big(\mathbb{1}_{\xi<0} |M_{i+1/2+} - M_{i+1/2-}| + \mathbb{1}_{\xi>0} |M_{i-1/2+} - M_{i-1/2-}| \big) \bigg).
\end{aligned}
\tag{3.72}
$$

We set

$$
\nu_\beta^0 = c_{\alpha,\beta} - 2\epsilon(1+\alpha)\beta,
\tag{3.73}
$$

which is positive if ϵ is taken small enough (recall that $\alpha > 0$ has been chosen so as to satisfy (3.64), and hence depends only on β). Then using (3.57) and (3.72), the two first lines in the right-hand side of (3.72) give a dissipation as stated in (3.46), while the last lines give an error. From (3.72) and (3.69), for $\xi < 0$ the typical error terms take the form

(3.74)

$$
\begin{aligned}
& M_i |M_{i+1/2+} - M_{i+1/2-}| |M_i - M_{i-1/2+}| \\
&= \big(\mathbb{1}_{M_i \leq M_{i+1/2+}} + \mathbb{1}_{M_i > M_{i+1/2+}} \big) M_i |M_{i+1/2+} - M_{i+1/2-}| |M_i - M_{i-1/2+}| \\
&\leq \mathbb{1}_{M_i \leq M_{i+1/2+}} M_i \Big(\epsilon_2 |M_{i+1/2+} - M_{i+1/2-}|^2 + \epsilon_2^{-1} |M_i - M_{i-1/2+}|^2 \Big) \\
&\quad + \mathbb{1}_{M_i > M_{i+1/2+}} \Big(M_{i+1/2-} |M_{i+1/2+} - M_{i+1/2-}| |M_i - M_{i-1/2+}| \\
&\qquad + |M_i - M_{i+1/2-}| |M_{i+1/2+} - M_{i+1/2-}| |M_i - M_{i-1/2+}| \Big) \\
&\leq \epsilon_2 M_{i+1/2+} |M_{i+1/2+} - M_{i+1/2-}|^2 + \epsilon_2^{-1} M_i |M_i - M_{i-1/2+}|^2 \\
&\quad + M_{i+1/2-} \Big(\epsilon_2 |M_{i+1/2+} - M_{i+1/2-}|^2 + \epsilon_2^{-1} |M_i - M_{i-1/2+}|^2 \Big) \\
&\quad + M_i |M_i - M_{i+1/2-}| |M_i - M_{i-1/2+}| \\
&\leq \epsilon_2 \big(M_{i+1/2+} + M_{i+1/2-} \big) |M_{i+1/2+} - M_{i+1/2-}|^2 \\
&\quad + 3\epsilon_2^{-1} M_i |M_i - M_{i-1/2+}|^2 + \epsilon_2 M_i |M_i - M_{i+1/2-}|^2.
\end{aligned}
$$

The term proportional to ϵ_2 can therefore be absorbed by ν_β^0. Since a similar estimate holds for $\xi > 0$, diminishing slightly ν_β^0 by something proportional to ϵ_2 (taken small enough), we get a coefficient $\nu_\beta > 0$. The only remaining error terms finally take the form stated in the last line of (3.46). This completes the proof of (ii) in Theorem 3.6. □

Remark 3.10. Consider the situation when for some i_0 one has

$$
u_{i_0-1} = u_{i_0} = u_{i_0+1} \neq 0 \text{ and } h_{i_0-1} + z_{i_0-1} = h_{i_0} + z_{i_0} = h_{i_0+1} + z_{i_0+1},
$$

with $z_{i_0-1} \neq z_{i_0}$ or $z_{i_0} \neq z_{i_0+1}$. Then by (3.3), (3.4), the reconstructed states satisfy $U_{i+1/2-} = U_{i+1/2+}$ for $i = i_0 - 1, i_0$. We observe that then, in the formula (3.58) for d_i, the dissipative terms vanish for $i = i_0$, for all ξ. Thus $d_{i_0} \geq 0$ and $\int d_{i_0}(\xi) d\xi > 0$, which means that the extra term d_i in (3.57) gives a dissipation with the wrong sign, in agreement with Proposition 3.8.

Proof of Proposition 3.8. It has been proved in [1] that the semi-discrete HR scheme (limit $\sigma_i \to 0$) satisfies the entropy inequality without error term. Here we prove that the fully-discrete scheme does not, whatever restrictive is the CFL condition. This result holds for an arbitrary numerical flux $\mathcal{F}$ taken for the homogeneous Saint-Venant system. The argument is as follows.

Consider the local dissipation

$$\mathcal{D}_i^n = \eta(U_i^{n+1}) + g z_i h_i^{n+1} - \eta(U_i) - g z_i h_i + \sigma_i \Big(\widetilde{G}_{i+1/2} - \widetilde{G}_{i-1/2} \Big), \tag{3.75}$$

where U_i^{n+1} is given by (3.1), $F_{i+1/2\pm}$ are defined by (3.2)-(3.5), and

$$\widetilde{G}_{i+1/2} = \mathcal{G}(U_{i+1/2-}, U_{i+1/2+}) + g z_{i+1/2} \mathcal{F}^0(U_{i+1/2-}, U_{i+1/2+}), \tag{3.76}$$

where $\mathcal{G}$ is the numerical entropy flux associated to $\mathcal{F}$, and $\mathcal{F}^0$ is the first (density) component of $\mathcal{F}$. Then, taking into account that $h_i^{n+1} = h_i - \sigma_i(\mathcal{F}^0(U_{i+1/2-}, U_{i+1/2+}) - \mathcal{F}^0(U_{i-1/2-}, U_{i-1/2+}))$, one has

$$\begin{aligned} \frac{\mathcal{D}_i^n}{\sigma_i} &= \frac{\eta\big(U_i - \sigma_i(F_{i+1/2-} - F_{i-1/2+})\big) - \eta(U_i)}{\sigma_i} \\ &\quad - g z_i \big(\mathcal{F}^0(U_{i+1/2-}, U_{i+1/2+}) - \mathcal{F}^0(U_{i-1/2-}, U_{i-1/2+}) \big) \\ &\quad + \widetilde{G}_{i+1/2} - \widetilde{G}_{i-1/2}. \end{aligned} \tag{3.77}$$

The entropy η being strictly convex, the function

$$\sigma_i \mapsto \eta\big(U_i - \sigma_i(F_{i+1/2-} - F_{i-1/2+})\big) \tag{3.78}$$

is convex, and strictly convex if

$$F_{i+1/2-} - F_{i-1/2+} \neq 0. \tag{3.79}$$

Assuming that this condition holds, we get that the right-hand side of (3.77) is strictly increasing with respect to σ_i. In particular, it will be strictly positive if the limit as $\sigma_i \to 0$ of this quantity vanishes. This limit is nothing else than the dissipation of the semi-discrete scheme

$$\begin{aligned} &-\eta'(U_i)(F_{i+1/2-} - F_{i-1/2+}) + \widetilde{G}_{i+1/2} - \widetilde{G}_{i-1/2} \\ &-g z_i \big(\mathcal{F}^0(U_{i+1/2-}, U_{i+1/2+}) - \mathcal{F}^0(U_{i-1/2-}, U_{i-1/2+}) \big). \end{aligned} \tag{3.80}$$

Consider data such that

$$U_i = U_l,\ z_i = z_l \text{ for } i \leq i_0, \quad U_i = U_r,\ z_i = z_r \text{ for } i > i_0, \tag{3.81}$$

for left and right states $U_l = (h_l, h_l u_l)$, $U_r = (h_r, h_r u_r)$ such that

$$u_l = u_r \neq 0, \qquad h_l + z_l = h_r + z_r, \qquad z_r - z_l > 0. \tag{3.82}$$

Then one checks easily that (3.79) holds for $i = i_0$, and that (3.80) vanishes for all i. Therefore, $\mathcal{D}_{i_0}^n > 0$, which proves the claim. □

The following lemma establishes a kind of L^2-Lipschitz dependency of the Maxwellian with respect to U, that allows to estimate the integral of the error terms in (3.46). Note that the Maxwellian (1.4) is only 1/2-Hölder continuous at fixed ξ.

Lemma 3.11. *Let $U_k = (h_k, h_k u_k)$ for $k = 1, 2, 3$ with $h_k \geq 0$. Then*

$$\begin{aligned} &\int_{\mathbb{R}} M(U_1,\xi)\Big(M(U_1,\xi) - M(U_2,\xi)\Big)^2 d\xi \\ &\leq \frac{3}{g^2\pi^2}\Big(g(h_2-h_1)^2 + \min(h_1,h_2)(u_2-u_1)^2\Big), \end{aligned} \tag{3.83}$$

and

$$\begin{aligned} &\int_{\mathbb{R}} M(U_3,\xi)\Big(M(U_1,\xi) - M(U_2,\xi)\Big)^2 d\xi \\ &\leq \frac{6}{g^2\pi^2}\Big(g(h_3-h_1)^2 + g(h_3-h_2)^2 \\ &\quad + \min(h_1,h_3)(u_3-u_1)^2 + \min(h_2,h_3)(u_3-u_2)^2\Big). \end{aligned} \tag{3.84}$$

Proof. One has

$$\begin{aligned} &\int_{\mathbb{R}} M(U_1,\xi)\Big(M(U_1,\xi) - M(U_2,\xi)\Big)^2 d\xi \\ &\leq \frac{1}{2}\int_{\mathbb{R}} \Big(2M(U_1,\xi) + M(U_2,\xi)\Big)\Big(M(U_1,\xi) - M(U_2,\xi)\Big)^2 d\xi \\ &= \frac{3}{g^2\pi^2}\int_{\mathbb{R}} \Big(H_0(M(U_2,\xi),\xi) - H_0(M(U_1,\xi),\xi) \\ &\qquad\qquad -\partial_f H_0(M(U_1,\xi),\xi)(M(U_2,\xi) - M(U_1,\xi))\Big) d\xi \\ &\leq \frac{3}{g^2\pi^2}\int_{\mathbb{R}} \Big(H_0(M(U_2,\xi),\xi) - H_0(M(U_1,\xi),\xi) \\ &\qquad\qquad -\eta'(U_1)\begin{pmatrix}1\\ \xi\end{pmatrix}(M(U_2,\xi) - M(U_1,\xi))\Big) d\xi \\ &= \frac{3}{g^2\pi^2}\Big(\eta(U_2) - \eta(U_1) - \eta'(U_1)(U_2 - U_1)\Big) \\ &= \frac{3}{g^2\pi^2}\Big(g\frac{(h_2-h_1)^2}{2} + h_2\frac{(u_2-u_1)^2}{2}\Big). \end{aligned} \tag{3.85}$$

We can also estimate $M(U_1,\xi)$ by $M(U_1,\xi) + 2M(U_2,\xi)$, giving the same estimate as (3.85) with U_1 and U_2 exchanged and with an extra factor 2. This proves (3.83). Then, denoting $M_k \equiv M(U_k,\xi)$, according to the Minkowsky inequality,

$$\begin{aligned} &\left(\int_{\mathbb{R}} M_3\big(M_1 - M_2\big)^2 d\xi\right)^{1/2} \\ &\leq \left(\int_{\mathbb{R}} M_3\big(M_1 - M_3\big)^2 d\xi\right)^{1/2} + \left(\int_{\mathbb{R}} M_3\big(M_3 - M_2\big)^2 d\xi\right)^{1/2}, \end{aligned} \tag{3.86}$$

Using (3.83), we obtain (3.84). □

4. Conclusion

We have established that the unmodified hydrostatic reconstruction scheme for the Saint Venant system with topography satisfies a fully discrete entropy inequality (3.47) with error term, in the case when the homogeneous numerical flux is the kinetic one with the Maxwellian (1.4). This inequality is obtained as the integral with respect to the kinetic variable ξ of a discrete kinetic entropy inequality (3.46) with error term. These error terms are not present in the case when the entropy dissipation is linearized with respect to the timestep Δt (or equivalently in the semi-discrete case). They come from the less dissipative nature of explicit schemes

with respect to their semi-discrete versions, as appears clearly in the case without topography of Lemma 2.1. In the case with topography, the identity (3.51) enables to write the entropy dissipation as a sum of the one for the semi-discrete scheme plus an error term L_i which has the wrong sign, and which is merely quadratic in Δt, see (3.54)-(3.56). In general, the second-order in Δt terms appearing in the entropy dissipation are dominated (under a CFL condition) by the linear in Δt dissipation terms. However, since here we have a well-balanced scheme, these first-order terms degenerate at the steady states at rest, and cannot dominate the second-order terms. This is why error terms remain in (3.47). Nevertheless, these errors are estimated in the square of the topography jumps, and do not involve jumps in the unknown U, that would not be small in the case of shocks. This property enables to proceed with a proof of convergence of the scheme, that will be provided in a forthcoming paper.

An open problem that remains however is to establish the fully discrete entropy inequality with error (3.47) for a HR scheme with general (non kinetic) homogeneous numerical flux $\mathcal{F}$ satisfying a fully discrete entropy inequality.

Acknowledgments

The authors wish to express their warm thanks to Carlos Parés Madroñal for many fruitful discussions. This work has been partially funded by the ANR contract ANR-11-BS01-0016 LANDQUAKES.

References

[1] E. Audusse, F. Bouchut, M.-O. Bristeau, R. Klein, B. Perthame, *A fast and stable well-balanced scheme with hydrostatic reconstruction for shallow water flows*, SIAM J. Sci. Comp. 25 (2004), 2050-2065.

[2] E. Audusse, M.-O. Bristeau, *A well-balanced positivity preserving second-order scheme for Shallow Water flows on unstructured meshes*, J. Comput. Phys. 206 (2005), 311-333.

[3] E. Audusse, M.-O. Bristeau, M. Pelanti, J. Sainte-Marie, *Approximation of the hydrostatic Navier-Stokes system for density stratified flows by a multilayer model. Kinetic interpretation and numerical validation*, J. Comp. Phys. 230 (2011), 3453-3478.

[4] E. Audusse, M.-O. Bristeau, B. Perthame, *Kinetic schemes for Saint Venant equations with source terms on unstructured grids*, Technical Report 3989, INRIA, Unité de recherche de Rocquencourt, France, 2000. http://www.inria.fr/rrrt/rr-3989.html.

[5] E. Audusse, M.-O. Bristeau, B. Perthame, J. Sainte-Marie, *A multilayer Saint-Venant system with mass exchanges for Shallow Water flows. Derivation and numerical validation*, ESAIM: M2AN 45 (2011), 169-200.

[6] F. Berthelin, *Convergence of flux vector splitting schemes with single entropy inequality for hyperbolic systems of conservation laws*, Numer. Math. 99 (2005), 585-604.

[7] F. Berthelin, F. Bouchut, *Relaxation to isentropic gas dynamics for a BGK system with single kinetic entropy*, Meth. and Appl. of Analysis 9 (2002), 313-327.

[8] F. Bouchut, *Construction of BGK models with a family of kinetic entropies for a given system of conservation laws*, J. Stat. Phys. 95 (1999), 113-170.

[9] F. Bouchut, *Entropy satisfying flux vector splittings and kinetic BGK models*, Numer. Math. 94 (2003), 623-672.

[10] F. Bouchut, *Nonlinear stability of finite volume methods for hyperbolic conservation laws, and well-balanced schemes for sources*, Birkhäuser, 2004.

[11] F. Bouchut, T. Morales, *A subsonic-well-balanced reconstruction scheme for shallow water flows*, Siam J. Numer. Anal. 48 (2010), 1733-1758.

[12] F. Bouchut, V. Zeitlin, *A robust well-balanced scheme for multi-layer shallow water equations*, Discrete and Continuous Dynamical Systems - Series B, 13 (2010), 739-758.

[13] M.-O. Bristeau, N. Goutal, J. Sainte-Marie, *Numerical simulations of a non-hydrostatic Shallow Water model*, Computers & Fluids 47 (2011), 51-64.

[14] M.J. Castro, A. Pardo Milanés, C. Parés, *Well-balanced numerical schemes based on a generalized hydrostatic reconstruction technique*, Math. Models Methods Appl. Sci. 17 (2007), 2055-2113.
[15] F. Coquel, K. Saleh, N. Seguin, *A robust and entropy-satisfying numerical scheme for fluid flows in discontinuous nozzles*, Math. Models Meth. Appl. Sci. 24 (2014), 2043.
[16] F. Coron, B. Perthame, *Numerical passage from kinetic to fluid equations*, SIAM J. Numer. Anal. 28 (1991), 26-42.
[17] L. Gosse, *Computing qualitatively correct approximations of balance laws. Exponential-fit, well-balanced and asymptotic-preserving*, SIMAI Springer Series, 2. Springer, Milan, 2013.
[18] N. Goutal, J. Sainte-Marie, *A kinetic interpretation of the section-averaged Saint-Venant system for natural river hydraulics*, Int. J. Numer. Meth. Fluids 67 (2011), 914-938.
[19] S. Jin, *Asymptotic preserving (AP) schemes for multiscale kinetic and hyperbolic equations: a review*, Lecture Notes for Summer School on "Methods and Models of Kinetic Theory" (M&MKT), Porto Ercole (Grosseto, Italy), June 2010. Rivista di Matematica della Universite di Parma 3 (2012), 177-216.
[20] S. Noelle, Y. Xing, C.-W. Shu, *High-order well-balanced finite volume WENO schemes for shallow water equation with moving water*, J. Comput. Phys. 226 (2007), 29-58.
[21] B. Perthame, *Boltzmann type schemes for gas-dynamics and the entropy property*, SIAM J. Numer. Anal. 27 (1990), 1405-1421.
[22] B. Perthame, *2nd-order Boltzmann schemes for compressible Euler equations in one and 2 space dimensions*, SIAM J. Numer. Anal. 29 (1992), 1-19.
[23] B. Perthame, *Kinetic formulation of conservation laws*, Oxford Lecture Series in Mathematics and its Applications, 21. Oxford University Press, Oxford, 2002.
[24] B. Perthame, C. Simeoni, *A kinetic scheme for the Saint Venant system with a source term*, Calcolo 38 (2001), 201-231.
[25] Y. Xing, C.-W. Shu, *A survey of high order schemes for the shallow water equations*, Journal of Mathematical Study 47 (2014), 221-249.

Université Paris 13, Laboratoire d'Analyse, Géométrie et Applications, 99 av. J.-B. Clément, F-93430 Villetaneuse, France - Inria, ANGE project-team, Rocquencourt - B.P. 105, F78153 Le Chesnay cedex, France - CEREMA, ANGE project-team, 134 rue de Beauvais, F-60280 Margny-Lès-Compiègne, France - Sorbonne University, UPMC University Paris VI, ANGE project-team, UMR 7958 LJLL, F-75005 Paris, France
E-mail address: eaudusse@yahoo.fr

Université Paris-Est, Laboratoire d'Analyse et de Mathématiques Appliquées (UMR 8050), CNRS, UPEM, UPEC, F-77454, Marne-la-Vallée, France
E-mail address: Francois.Bouchut@u-pem.fr

Inria, ANGE project-team, Rocquencourt - B.P. 105, F78153 Le Chesnay cedex, France - CEREMA, ANGE project-team, 134 rue de Beauvais, F-60280 Margny-Lès-Compiègne, France - Sorbonne University, UPMC University Paris VI, ANGE project-team, UMR 7958 LJLL, F-75005 Paris, France
E-mail address: Marie-Odile.Bristeau@inria.fr

Inria, ANGE project-team, Rocquencourt - B.P. 105, F78153 Le Chesnay cedex, France - CEREMA, ANGE project-team, 134 rue de Beauvais, F-60280 Margny-Lès-Compiègne, France - Sorbonne University, UPMC University Paris VI, ANGE project-team, UMR 7958 LJLL, F-75005 Paris, France
E-mail address: Jacques.Sainte-Marie@inria.fr