Numerical methods for hyperbolic equations of Saint-Venant type.

Chiara Simeoni

THÈSE de DOCTORAT
de l'UNIVERSITÉ PARIS VI

Spécialité : MATHÉMATIQUES

présentée par
Mlle CHIARA SIMEONI

pour obtenir le grade de
DOCTEUR de l'UNIVERSITÉ PARIS VI

Sujet de la thèse :
MÉTHODES NUMÉRIQUES POUR DES ÉQUATIONS HYPERBOLIQUES DE TYPE SAINT-VENANT

soutenue le 7 Novembre 2002 devant le jury composé de :

M. François Bouchut	Examinateur
Mme Marie-Odile Bristeau	Examinateur
M. Bruno Despres	Examinateur
M. Thierry Gallouët	Rapporteur
M. Pierangelo Marcati	Examinateur
M. Benoît Perthame	Directeur de thèse
M. Giovanni Russo	Rapporteur

Résumé

L'objet de cette thèse est de contribuer à l'étude numérique des lois de conservation hyperboliques avec termes sources, ce qui est motivé par les applications aux équations de Saint-Venant pour les eaux peu profondes.

La première partie traite des questions habituelles de l'analyse des approximations numériques des lois de conservation scalaires. On se concentre sur des schémas aux volumes finis semi-discrets, dans le cas général d'un maillage non-uniforme. Pour définir des discrétisations appropriées du terme source, on introduit le formalisme spécifique de la méthode "Upwind Interface Source" et on établit des conditions sur les fonctions numériques telles que le solveur discret préserve les solutions stationnaires. Une définition rigoureuse de consistance est ensuite formulée, adaptée aux "schémas équilibres", pour laquelle on est capable de prouver un théorème de convergence faible de type Lax-Wendroff.
La méthode considérée dans un premier temps est essentiellement d'ordre un en espace. Pour améliorer la précision, on développe des approches à haute résolution pour la méthode "Upwind Interface Source" et on montre que celles-ci sont un moyen efficace de dériver des schémas d'ordre plus élevé avec des propriétés convenables. On prouve une estimation d'erreur dans L^p, $1 \leq p < +\infty$, qui est un résultat optimal dans le cas d'un maillage uniforme. On conclut alors que les mêmes taux de convergence $\mathcal{O}(h)$ et $\mathcal{O}(h^2)$ que pour les systèmes homogènes correspondants sont valables.

La deuxième partie présente un schéma numérique pour approcher les équations de Saint-Venant, avec un terme source géométrique, qui vérifie les propriétés théoriques suivantes : il préserve les états stationnaires de l'eau au repos, vérifie une inégalité d'entropie discrète, préserve la positivité de la hauteur de l'eau et reste stable avec des profiles du fond discontinus. Cela est obtenu grce à une approche cinétique au système ; dans ce contexte, on utilise une description formelle du comportement microscopique du système pour définir les flux numériques aux interfaces d'un maillage nonstructuré. On utilise aussi le concept de variables conservatives centrées (typique de la méthode des volumes finis) et des termes sources décentrés aux interfaces.
Finalement, on présente des simulations numériques du système des équations de Saint-Venant modifiées pour prendre en compte le frottement et la viscosité, afin de retrouver les résultats de certaines études expérimentales.
Une application à la modélisation des termes de frottement pour les avalanches de neige est discutée dans l'Appendice.

Abstract

The purpose of this dissertation is to contribute to the numerical study of hyperbolic conservation laws with source terms, motivated by the application to the Saint-Venant equations for shallow waters.

The first part deals with usual questions in the analysis of numerical approximations for scalar conservation laws. We focus on semi-discrete finite volume schemes, in the general case of a nonuniform spatial mesh. To define appropriate discretizations of the source term, we introduce the formalism peculiar to the Upwind Interface Source method and we establish conditions on the numerical functions so that the discrete solver preserves the stationary solutions. Then we formulate a rigorous definition of consistency, adapted to well-balanced schemes, for which we are able to prove a Lax-Wendroff type convergence theorem.
The method first considered is essentially first order. To improve accuracy, we develop high resolution approaches for the Upwind Interface Source method and we show that these are efficient ways to derive higher order schemes with suitable properties. We prove an error estimate in L^p, $1 \leq p < +\infty$, which is an optimal result in the case of a uniform mesh. We thus conclude that the same convergence rates $\mathcal{O}(h)$ and $\mathcal{O}(h^2)$ hold as for the corresponding homogeneous systems.

The second part presents a numerical scheme to compute Saint-Venant equations, with a geometrical source term, which satisfies the following theoretical properties : it preserves the steady states of still water, it satisfies a discrete entropy inequality, it preserves the non-negativity of the height of water and remains stable with a discontinuous bottom. This is achieved by means of a kinetic approach to the system ; in this context, we use a natural description of the microscopic behaviour of the system to define numerical fluxes at the interfaces of an unstructured mesh. We also use the concept of cell-centered conservative quantities (typical of the finite volume method) and upwind interfacial sources.
Finally, we present some numerical simulations of the Saint-Venant system modified by including small friction and viscosity, in order to recover the results of experimental studies. An application to the numerical modelling of friction terms for debris avalanches is proposed in the Appendix.

Table des Matières

Appendice

Introduction

Cette introduction entend présenter les considérations générales et extraire les idées fondamentales qui sont à la base des résultats exposées dans la suite de la thèse, pour la formulation rigoureuse desquels on est invité à se reporter aux chapitres correspondants.

1 Présentation du problème

1.1 Les lois de conservation hyperboliques avec terme source géométrique

On considère le problème aux valeurs initiales pour une loi de conservation scalaire avec un terme source,

$$\frac{\partial u}{\partial t} + \frac{\partial A(u)}{\partial x} + B(x,u) = 0, \quad t \in \mathbb{R}_+,\ x \in \mathbb{R}, \tag{1.1}$$

$$u(0,x) = u_0(x) \in L^p(\mathbb{R}) \cap L^\infty(\mathbb{R}),\ 1 \leq p < +\infty, \tag{1.2}$$

avec $u(t,x) \in \mathbb{R}$ et une fonction de flux régulière A à valeurs réelles, avec

$$a(u) = A'(u) \in C^1(\mathbb{R}). \tag{1.3}$$

On se restreint ici aux termes sources dans l'équation (1.1) définis par

$$B(x,u) = z'(x)b(u), \quad z' \in L^p(\mathbb{R}),\ b \in C^1(\mathbb{R}). \tag{1.4}$$

L'équation (1.1) admet une famille d'inégalités d'entropie,

$$\frac{\partial S(u)}{\partial t} + \frac{\partial \eta(u)}{\partial x} + S'(u)B(x,u) \leq 0, \qquad \eta'(u) = S'(u)a(u), \tag{1.5}$$

pour toute paire d'une fonction d'entropie convexe S et flux d'entropie η correspondant (voir [37], [17] et [38]). Pour des hypothèses de régularité plus fortes sur le terme source, Kružkov [37] démontre existence et unicité de la solution entropique du problème (1.1)-(1.2), dans l'espace des fonctions $L^\infty([0,T); L^p(\mathbb{R}))$, pour tout $T \in \mathbb{R}_+$. Des résultats récents concernants les systèmes hyperboliques de lois de conservation avec termes sources (voir [4], par exemple) généralisent les travaux classiques [18] sur l'existence globale de solutions faibles à variation bornée (voir aussi [28] et [5]).
Pour le cas délicat d'un terme source singulier de la forme (1.4), notamment avec une fonction z discontinue, un résultat d'unicité est prouvé dans [55], en utilisant l'approche cinétique formulée rigoureusement dans [44] et [51].

Par comparaison au problème homogène, une différence significative est observée pour les solutions stationnaires de l'équation (1.1), qui sont définies

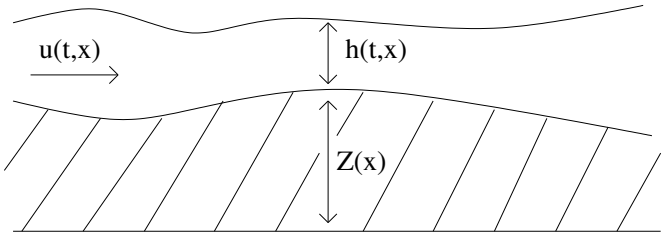
par l'équilibre entre terme source et forces internes ; en intégrant l'équation stationnaire associée à (1.1), on obtient alors la relation algébrique

$$D(u) + z(x) = C^{st}, \qquad D'(u) = \frac{a(u)}{b(u)}. \tag{1.6}$$

On suppose dans la suite que la fonction D est strictement monotone, ce qui garantit l'existence d'une solution unique et lipschitzienne de l'équation (1.6). Cette dernière hypothèse apparait restrictive et n'est pas toujours satisfaite dans des situations réalistes, néanmoins elle est utilisée pour simplifier la formulation de certaines théories numériques (voir [10] et [26], par exemple).

1.2 Le système des équations de Saint-Venant

L'analyse des solutions stationnaires des lois de conservation avec terme source géométrique est motivée par leur application aux écoulements permanents dans les rivières et les canaux avec une topographie complexe.



Les équations de Saint-Venant décrivent l'écoulement monodimensionnel dans un canal rectangulaire par l'intermédiaire de la hauteur d'eau $h(t,x) \geq 0$ et de sa vitesse moyenne $u(t,x) \in \mathbb{R}$, qui vérifient le système hyperbolique

$$\frac{\partial h}{\partial t} + \frac{\partial (hu)}{\partial x} = 0, \tag{1.7}$$

$$\frac{\partial (hu)}{\partial t} + \frac{\partial}{\partial x}(hu^2 + \frac{g}{2}h^2) + ghZ' = 0, \tag{1.8}$$

o g est la gravité et $Z(x)$ représente le profil longitudinal du fond du canal, donc $h+Z$ est la cote de la surface libre et hu la quantité de mouvement. La fonction d'entropie du système (1.7)-(1.8) est l'énergie physique,

$$E(h,u,Z) = h\frac{u^2}{2} + \frac{g}{2}h^2 + ghZ, \tag{1.9}$$

pour laquelle on démontre (voir [17] et [54]) l'inégalité d'entropie

$$\frac{\partial E}{\partial t} + \frac{\partial}{\partial x}[u(E + \frac{g}{2}h^2)] \leq 0. \tag{1.10}$$

La prise en compte d'une topographie variable introduit un terme source dans l'équation (1.8) sur la quantité de mouvement, qui intervient dans la définition des états stationnaires,

$$hu = C_1, \tag{1.11}$$

$$\frac{u^2}{2} + g(h + Z) = C_2, \tag{1.12}$$

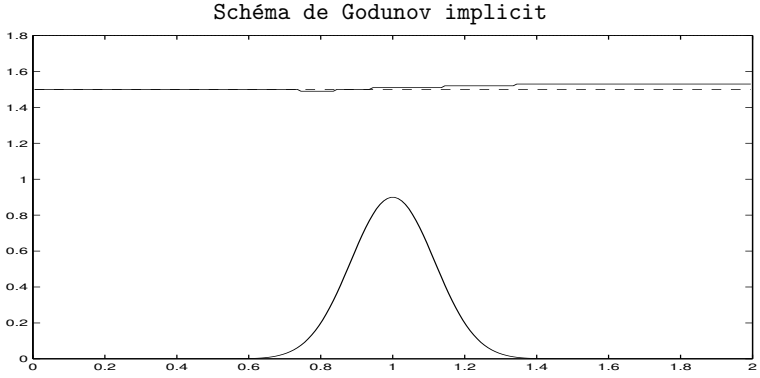
avec C_1 et C_2 constantes arbitraires. En particulier, l'équilibre simple d'un lac au repos est donné par $u=0$, $h+Z=C^{st}$.

Une dérivation formelle du système (1.7)-(1.8) à partir des équations de Navier-Stokes est obtenue dans [24], sous certaines hypothèses (hauteur d'eau faible, approximation hydrostatique de la pression, homogénéité verticale des vitesses horizontales). Des questions analogues sont discutées dans [1] et [2]. Des difficultés persistent dans l'analyse du système (1.7)-(1.8), qui sont liées à sa structure mathématique complexe : la preuve de l'existence globale de solutions faibles après apparition de singularités est dans [43] ; d'autres résultats sont présentés dans [38], [46], [47], [35], [16], [48], [13], [3] et [22].

1.3 La question numérique des solutions stationnaires

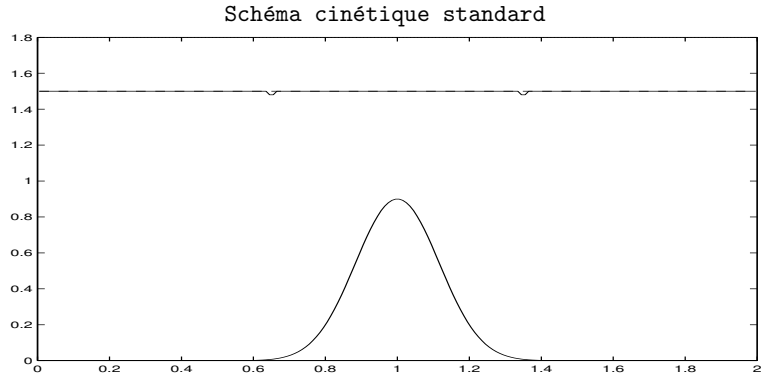
La présence du terme source dans l'équation (1.1) influence ses propriétés analytiques, par conséquent une approximation correcte de celui-ci apparait comme un point crucial de l'approche numérique du problème (1.1)-(1.2), en particulier pour préserver les solutions stationnaires au niveau discret.

Des nombreux travaux sur le sujet ont déjà été conduits, dans le cas des équations scalaires comme pour le système de Saint-Venant, initialement avec des objectifs différents (voir [11], [14], [25] et [39], par exemple).



La plupart des méthodes proposées se basent sur des schémas classiques pour les lois de conservation hyperboliques, avec un traitement spécifique du

terme source, ce qui assure une résolution précise des problèmes d'évolution mais se révèle quelquefois insuffisant pour préserver des états stationnaires : l'équilibre simple d'un lac au repos sur une topographie régulière n'est pas bien approché par des schémas centrés standards (même si les perturbations tendent à disparaître avec une diminution du pas d'espace, voir [57]).



Une formulation générale du problème des solutions stationnaires pour la méthode des volumes finis est examinée dans le Chapitre 1, o des définitions rigoureuses sont introduites et confirmées par les résultats correspondants.

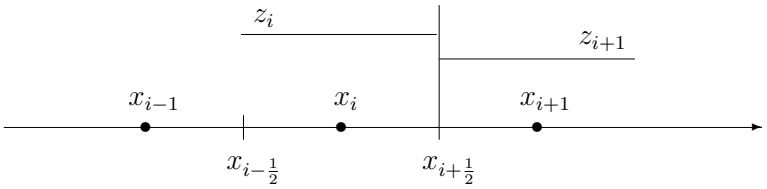
Des approches différentes de cette question ont également été développées dans la littérature, qui conduisent à la construction de schémas numériques adaptés et avec les propriétés requises de conservation des états stationnaires. On se réfère notamment à la méthode des éléments finis, appliquée aux schémas de relaxation (voir [34], [33] et [19]), ainsi qu'aux "schémas centraux" (voir [45], [52] et [49]), qui sont utilisés pour les applications au cas de termes sources singuliers (voir [41], [15] et [53], par exemple).
D'autres techniques sont basées sur les maillages adaptatifs (voir [8], [12] et [7]) et des questions diverses liées à l'analyse numérique du système des équations de Saint-Venant (1.7)-(1.8) sont considérées (voir [6], par exemple).

2 L'approche numérique

2.1 La méthode "Upwind Interface Source" pour les termes sources

La théorie numérique générale présentée dans la première partie de cette thèse utilise la notation classique de la méthode des volumes finis, pour caractériser les structures fondamentales des schémas numériques analysés.

On pose un maillage non-uniforme sur $\mathbb{R}$, composé de mailles $C_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$, de centre x_i, $i \in \mathbb{Z}$ et longueur Δx_i, donc les points $x_{i+\frac{1}{2}}$ sont des interfaces.



On considère une approximation, constante par maille, de la fonction z dans le terme source (1.4), avec des coefficients $z_i = \frac{1}{\Delta x_i} \int_{C_i} z(x)\, dx$ par exemple, et on remarque que la dérivée de cette fonction discrète est donnée par les différences des valeurs aux interfaces,

$$z_\Delta(x) = \sum_{i\in\mathbb{Z}} z_i \, \mathbb{1}_{C_i}(x), \qquad z'_\Delta(x) = \sum_{i\in\mathbb{Z}} \Delta z_{i+\frac{1}{2}} \, \delta_{i+\frac{1}{2}}, \tag{2.1}$$

o $\delta_{i+\frac{1}{2}}$ indique la fonction de Dirac du point $x_{i+\frac{1}{2}}$ et $\Delta z_{i+\frac{1}{2}} = z_{i+1} - z_i$.

La méthode des volumes finis, dans le cas semi-discret en espace, commence par une intégration de l'équation (1.1) sur chaque maille, pour obtenir

$$\Delta x_i \, \frac{d}{dt} u_i(t) + A(u(t, x_{i+\frac{1}{2}})) - A(u(t, x_{i-\frac{1}{2}})) + \int_{C_i} B(x, u(t,x))\, dx = 0,$$

avec $u_i(t) = \frac{1}{\Delta x_i} \int_{C_i} u(t,x)\, dx$ la moyenne par maille de la solution analytique. Un schéma volumes finis pour le problème (1.1)-(1.2) est une approximation de la relation précédente, sous la forme générale présentée dans le Chapitre 1.

S'agissant de lois de conservation hyperboliques, les flux numériques sont définis aux interfaces et la condition habituelle de consistance est imposée, pour retrouver les résultats théoriques du cas homogène (voir [36]).

La question de la discrétisation du terme source est loin d'être accessoire et un traitement inapproprié produit des résultats insatisfaisants, notamment pour la conservation des solutions stationnaires.
Une approche efficace et adaptée au terme source géometrique (1.4) est basée sur des approximations aux interfaces et décentrées, qui apparaissent comme les seules compatibles avec la méthode des volumes finis. En effet, on déduit directement de (2.1) les relations suivantes,

$$\begin{aligned} \int_{C_i} z'(x)\, b(u(t,x))\, dx &\approx \int_{C_i} z'_\Delta(x)\, b(u(t,x))\, dx \\ &\approx \frac{1}{2} \Delta z_{i-\frac{1}{2}} b(u(t, x_{i-\frac{1}{2}})) + \frac{1}{2} \Delta z_{i+\frac{1}{2}} b(u(t, x_{i+\frac{1}{2}})) \\ &\approx \frac{1}{2} \Delta z_{i-\frac{1}{2}} B(u_{i-1}, u_i) + \frac{1}{2} \Delta z_{i+\frac{1}{2}} B(u_i, u_{i+1}), \end{aligned} \tag{2.2}$$

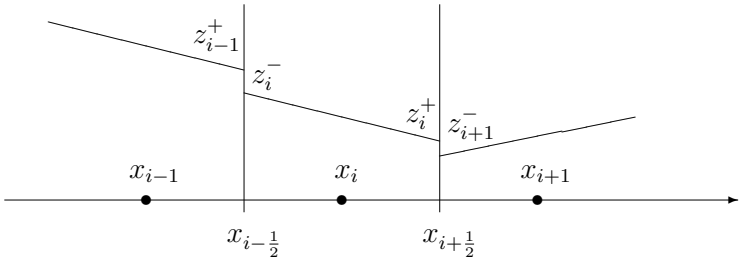
avec $B(u,u)=b(u)$ la condition de consistance conséquente à cette formulation, qui est généralisée rigoureusement par les définitions du Chapitre 1.

Dans les dernières années, la méthode Upwind Interface Source pour les lois de conservation hyperboliques avec terme source a pris un certain essor, en particulier pour son application au problème des états stationnaires du système de Saint-Venant (1.7)-(1.8), et jusqu'aux situations complexes des équations en deux dimensions d'espace (voir la bibliographie du Chapitre 1 et du Chapitre 3, aussi que [40], [23], [56], [32], [9] et [31], par exemple).

2.2 L'extension aux discrétisations d'ordre deux

La méthode décrite précédemment est essentiellement d'ordre un en espace, sans autres hypothèses sur les fonctions numériques que la simple consistance.

Une technique classique pour construire des schémas d'ordre deux consiste à remplacer l'approximation constante par maille des fonctions numériques avec des approximations linéaires, par exemple, qui fournissent des valeurs plus précises aux interfaces (se reporter au Chapitre 2 pour les détails).



La dérivée de l'approximation linéaire par maille de la fonction z dans (1.4) est alors définie par une partie centrée et les contributions aux interfaces,

$$z'_\Delta(x) = \sum_{i\in\mathbb{Z}} z'_i \mathbb{1}_{(x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})}(x) + \sum_{i\in\mathbb{Z}} \Delta^2 z_{i+\frac{1}{2}}\, \delta_{i+\frac{1}{2}}, \tag{2.3}$$

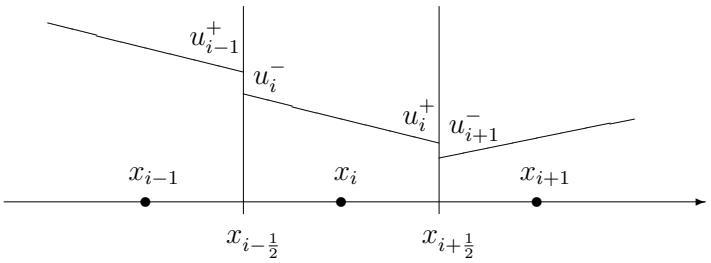
avec z'_i une valeur numérique pour la dérivée constante par maille, calculée en utilisant un opérateur discret spécifique, et dans ce cas $\Delta^2 z_{i+\frac{1}{2}} = z_{i+1}^- - z_i^+$. Un procédé analogue à (2.2) conduit, pour (2.3), à la relation

$$\begin{aligned}
&\int_{C_i} z'(x)\, b(u(t,x))\, dx \\
&\approx z'_i \int_{C_i} b(u(t,x))\, dx + \frac{1}{2}\Delta^2 z_{i-\frac{1}{2}} b(u(t, x_{i-\frac{1}{2}})) + \frac{1}{2}\Delta^2 z_{i+\frac{1}{2}} b(u(t, x_{i+\frac{1}{2}})) \\
&\approx z'_i\, \Delta x_i B(u_i) + \frac{1}{2}\Delta^2 z_{i-\frac{1}{2}} B(u_{i-1}^+, u_i^-) + \frac{1}{2}\Delta^2 z_{i+\frac{1}{2}} B(u_i^+, u_{i+1}^-),
\end{aligned}$$

o $B(u_i) \approx \frac{1}{\Delta x_i} \int_{C_i} b(u(t,x))\, dx$ et les valeurs aux interfaces de la reconstruction linéaire par maille de la solution numérique interviennent dans les termes décentrés de l'approximation précédente.

Quoique la question de préserver les états stationnaires au niveau discret ne soit traitée rigoureusement que pour des discrétisations d'ordre un (voir la théorie dans le Chapitre 1), les résultats numériques obtenus pour le système de Saint-Venant indiquent que l'extension d'ordre deux du schéma cinétique proposé dans le Chapitre 3 reproduit précisement l'équilibre d'un lac au repos (se reporter au Chapitre 4 pour plus de détails).

Pour construire des extensions d'ordre deux de la méthode "Upwind Interface Source", une alternative à l'approche que l'on vient d'illustrer consiste à formuler des critères de consistance plus restrictifs sur les discrétisations décentrées aux interfaces du type (2.2). Cette idée est suggérée par la forme du terme source géométrique (1.4), donné par le produit de fonctions avec un ordre de régularité différent. On utilise alors l'approximation constante par maille (2.1) de la fonction z et une reconstruction linéaire par maille de la solution numérique.



Par conséquent, les fonctions numériques de la discrétisation (2.2) sont dans le cas présent calculées sur des nouvelles valeurs aux interfaces,

$$\begin{aligned}\int_{C_i} z'(x)\, b(u(t,x))\, dx &\approx \frac{1}{2}\Delta z_{i-\frac{1}{2}} b(u(t, x_{i-\frac{1}{2}})) + \frac{1}{2}\Delta z_{i+\frac{1}{2}} b(u(t, x_{i+\frac{1}{2}})) \\ &\approx \frac{1}{2}\Delta z_{i-\frac{1}{2}} B(u_{i-1}^+, u_i^-) + \frac{1}{2}\Delta z_{i+\frac{1}{2}} B(u_i^+, u_{i+1}^-).\end{aligned}$$

La généralisation formelle de la relation précédente s'écrit

$$\int_{C_i} z'(x)\, b(u(t,x))\, dx \approx B^+(u_{i-1}^+, u_i^-, \Delta z_{i-\frac{1}{2}}) + B^-(u_i^+, u_{i+1}^-, \Delta z_{i+\frac{1}{2}}),$$

pour laquelle on considère la définition de consistance d'ordre deux suivante,

$$\lim_{\Delta z \to 0} \frac{B^+(u,u,\Delta z) + B^-(u,u,\Delta z)}{\Delta z} = b(u) + \mathcal{O}(|\Delta z|^2).$$

La théorie des approximations numériques d'ordre deux admet une formulation univoque, puisqu'on peut passer de l'une à l'autre des discrétisations d'ordre deux décrites précédemment par un simple calcul algébrique sur les quantités utilisées.

2.3 Les résultats de convergence

La question d'énoncer des critères de consistance appropriés pour les approximations numériques adoptées revêt une importance particulière dans la démonstration des théorèmes de convergence.

Dans le cas d'un maillage non-uniforme, vérifiant la condition géométrique de non-dégénérescence suivante,

$$\exists\, \alpha, \beta > 0 \;/\; \alpha \Delta x_{i+1} \leq \Delta x_i \leq \beta \Delta x_{i+1}, \quad \forall i \in \mathbb{Z}, \tag{2.4}$$

une extension du théorème de Lax-Wendroff est présentée dans le Chapitre 1, pour les lois de conservation hyperboliques avec terme source (1.4). Sous la seule hypothèse de consistance sur la discrétisation considérée, on prouve la convergence dans l'espace $\mathcal{D}'(\mathbb{R}_+ \times \mathbb{R})$ de la suite approchée obtenue à partir du schéma numérique vers une solution faible du problème (1.1)-(1.2).
Cependant, dériver des estimations d'erreur et en déduire la convergence forte dans un espace $L^p(\mathbb{R})$, $1 \leq p < +\infty$, reste difficile même dans la situation d'un maillage bien construit mais toujours non-uniforme.

On peut d'ailleurs démontrer par un contre-exemple qu'un tel résultat est faux pour la méthode (2.2), si la discrétisation associée ne prend pas en compte explicitement les variations de la taille des mailles.
Un modèle extrêmement simple de problème (1.1)-(1.2) est donné par

$$\frac{\partial u}{\partial t} = z'(x), \qquad u(0,x) = u_0(x), \tag{2.5}$$

pour lequel une intégration du terme source suivant la méthode des volumes finis conduit à la discrétisation (2.2), pour l'approximation

$$\frac{1}{\Delta x_i} \int_{C_i} z'(x)\, dx \approx \frac{1}{\Delta x_i} \left(\frac{z_i + z_{i+1}}{2} - \frac{z_{i-1} + z_i}{2} \right).$$

D'après un développement asymptotique des moyennes par maille dans la rélation précédente, utilisant aussi l'hypothèse (2.4), on obtient

$$\frac{1}{2\Delta x_i} (z_{i+1} - z_{i-1}) \approx \frac{1}{2\Delta x_i} z'(x_i) (x_{i+1} - x_{i-1}) + \mathcal{O}(\Delta x_i),$$

avec $|x_{i+1} - x_{i-1}| = \Delta x_i + \frac{\Delta x_{i-1}}{2} + \frac{\Delta x_{i+1}}{2}$. Il parait alors évident que cette

approximation converge vers la valeur $z'(x_i)$ souhaitée uniquement quand on se restreint au cas d'un maillage uniforme, $\Delta x_i = \Delta x$, $\forall i \in \mathbb{Z}$.

Les estimations d'erreur pour la méthode "Upwind Interface Source" sont détaillées dans le Chapitre 2, dans le cas d'un maillage uniforme, pour le schéma numérique d'ordre un et pour ses extensions d'ordre deux.
Pour les discrétisations d'ordre deux, certaines hypothèses de régularité sur les coefficients des dérivées discrètes dans la formulation (2.3) sont nécéssaires afin de garantir l'ordre de convergence attendu. Cela est confirmé par différents calculs numériques menés autour de l'équation (2.5).
La stabilité des méthodes considérées est, par contre, relativement simple à prouver même pour un maillage non-uniforme de la forme (2.4) en utilisant les arguments du Chapitre 2.

3 La méthode cinétique

3.1 Interprétation cinétique du système de Saint-Venant

Par analogie avec les équations d'Euler compressibles de la dynamique des gaz, on peut établir un lien mathématique entre les équations de la mécanique des fluides et la description microscopique du système de particules associé.

On considère une fonction χ à valeurs réelles, définie sur $\mathbb{R}$, qui vérifie

$$\chi(\omega) = \chi(-\omega) \geq 0, \quad \int_{\mathbb{R}} \chi(\omega)\, d\omega = 1, \quad \int_{\mathbb{R}} \omega^2 \chi(\omega)\, d\omega = \frac{g}{2}. \tag{3.1}$$

On introduit une densité de particules dans l'espace des phases, présentes au temps $t \geq 0$ à la position $x \in \mathbb{R}$ et ayant une vitesse $\xi \in \mathbb{R}$, définie par

$$f(t,x,\xi) = \sqrt{h(t,x)}\, \chi\left(\frac{\xi - u(t,x)}{\sqrt{h(t,x)}} \right). \tag{3.2}$$

D'après la définition (3.2) et les propriétés (3.1), on déduit les égalités

$$\begin{gathered} h = \int_{\mathbb{R}} f(t,x,\xi)\, d\xi, \qquad hu = \int_{\mathbb{R}} \xi f(t,x,\xi)\, d\xi, \\ hu^2 + \frac{g}{2} h^2 = \int_{\mathbb{R}} \xi^2 f(t,x,\xi)\, d\xi. \end{gathered} \tag{3.3}$$

Le système de Saint-Venant (1.7)-(1.8) est obtenu comme limite formelle de l'équation cinétique suivante pour la densité microscopique (3.2),

$$\frac{\partial f}{\partial t} + \xi \frac{\partial f}{\partial x} - g Z' \frac{\partial f}{\partial \xi} = Q(t,x,\xi), \tag{3.4}$$

avec $Q(t,x,\xi)$ un "terme de collision" qui vérifie

$$\int_{\mathbb{R}} Q\,d\xi = 0, \qquad \int_{\mathbb{R}} \xi Q\,d\xi = 0. \tag{3.5}$$

En intégrant l'équation (3.4) par rapport à $\xi \in \mathbb{R}$, simplement ou multipliée par ξ, on retrouve les équations du système et l'énergie (1.9) s'écrit

$$E(h,u,Z) = \int_{\mathbb{R}} \left[\frac{\xi^2}{2} f(t,x,\xi) + \frac{\pi^2 g^2}{6} f^3(t,x,\xi) + gZ f(t,x,\xi) \right] d\xi. \tag{3.6}$$

On remarque que, en l'absence de terme source externe, le modèle cinétique (3.4)-(3.5) est analogue à l'équation de Boltzmann de la théorie cinétique des gaz, pour laquelle les notations utilisées dans cette partie ont d'abord été introduites (se reporter à [27], par exemple).

3.2 La formulation générale du schéma cinétique

La démarche que l'on vient de présenter permet de construire une classe de schémas numériques, appelés "schémas cinétiques", particulièrement adaptés aux équations de la mécanique des fluides, compatibles avec la méthode de volumes finis et qui assurent la conservation au niveau discret des propriétés naturelles du système continu (voir [50] pour la théorie générale).

On applique à l'équation (3.4) une méthode de différences finies en temps, avec $t^n = n\Delta t$, $n \in \mathbb{N}$, et un schéma décentré classique en espace, avec le terme source pris en compte directement dans la définition des flux numériques,

$$f_i^{n+1}(\xi) - f_i^n(\xi) + \frac{\Delta t}{\Delta x_i} \xi \left(f_{i+\frac{1}{2}}^{n,-}(\xi) - f_{i-\frac{1}{2}}^{n,+}(\xi) \right) = 0. \tag{3.7}$$

L'intégration par rapport à $\xi \in \mathbb{R}$ de l'équation précédente fournit un schéma pour les quantités macroscopiques associées à (3.3), définies par

$$U_i^n = (h_i^n, (hu)_i^n), \qquad h_i^n = \int_{\mathbb{R}} f_i^n(\xi)\,d\xi, \quad (hu)_i^n = \int_{\mathbb{R}} \xi f_i^n(\xi)\,d\xi, \tag{3.8}$$

qui se présente sous la forme

$$U_i^{n+1} - U_i^n + \frac{\Delta t}{\Delta x_i} \left(\mathbb{A}_{i+\frac{1}{2}}^{n,-} - \mathbb{A}_{i-\frac{1}{2}}^{n,+} \right) = 0, \tag{3.9}$$

o les flux numériques sont donnés par les formules cinétiques

$$\mathbb{A}_{i+\frac{1}{2}}^{n,-} = \int_{\mathbb{R}} \xi \begin{pmatrix} 1 \\ \xi \end{pmatrix} f_{i+\frac{1}{2}}^{n,-}(\xi)\,d\xi, \tag{3.10}$$

$$\mathbb{A}_{i-\frac{1}{2}}^{n,+} = \int_{\mathbb{R}} \xi \begin{pmatrix} 1 \\ \xi \end{pmatrix} f_{i-\frac{1}{2}}^{n,+}(\xi)\,d\xi. \tag{3.11}$$

En toute rigueur, il faudrait remarquer que l'équation (3.7) est employée improprement pour représenter la discrétisation de l'équation cinétique (3.4), puisque le rôle du terme Q n'apparait qu'implicitement dans le procédé numérique. En effet, la présence de collisions introduit dans le cas général une discontinuité en temps sur la densité miscroscopique, ce qui devrait affecter la valeur f_i^{n+1} calculée par le schéma numérique. Cependant la définition de la densité discrète, correspondant à la formule (3.2),

$$f_i^n(\xi) = \sqrt{h_i^n}\,\chi\left(\frac{\xi - u_i^n}{\sqrt{h_i^n}}\right), \tag{3.12}$$

et les propriétés (3.1) de la fonction χ garantissent que les variables macroscopiques (3.8) restent continues en temps. On obtient donc un schéma (3.9) valable pour la résolution numérique du système de Saint-Venant (1.7)-(1.8).

Les flux cinétiques dans les intégrales de (3.10) et (3.11) peuvent être reécrits sous la forme suivante,

$$f_{i+\frac{1}{2}}^{n,\pm}(\xi) = f_{i+\frac{1}{2}}^{n}(\xi) + \left(f_{i+\frac{1}{2}}^{n,\pm}(\xi) - f_{i+\frac{1}{2}}^{n}(\xi)\right), \tag{3.13}$$

avec $f_{i+\frac{1}{2}}^n(\xi) \approx \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} f(t, x_{i+\frac{1}{2}}, \xi)\,dt$ une approximation consistante des valeurs moyennes aux interfaces, et par conséquent une formulation analogue s'étend aux flux macroscopiques correspondants $\mathbb{A}_{i+\frac{1}{2}}^{n,\pm} = \mathbb{A}_{i+\frac{1}{2}}^{n} + \left(\mathbb{A}_{i+\frac{1}{2}}^{n,\pm} - \mathbb{A}_{i+\frac{1}{2}}^{n}\right)$. Les quantités entre parenthèses sont alors des approximations aux interfaces et décentrées du terme source, pour l'équation (3.4) et le système (1.7)-(1.8) respectivement, en accord avec la théorie exposée au Chapitre 1.

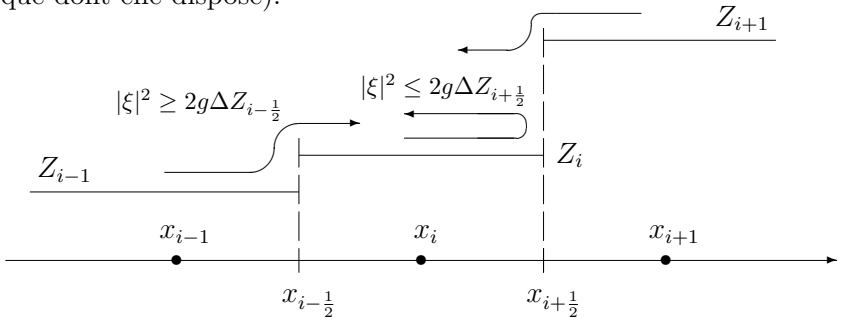
3.3 Le schéma cinétique pour la méthode "Upwind Interface Source"

Les propriétés du schéma (3.9) dépendent en général de l'approximation choisie pour les flux numériques (3.10)-(3.11), ainsi que du critère retenu pour sélectionner la fonction χ appropriée dans (3.2), et donc dans (3.12), parmi les nombreux choix compatibles avec les propriétés (3.1).

Pour la construction du schéma cinétique "avec réflexions" présenté dans le Chapitre 3, la définition de χ est basée sur la minimisation de la fonctionnelle d'énergie (3.6) et sur la conservation au niveau microscopique des états stationnaires (1.11)-(1.12) du système de Saint-Venant (se reporter au Chapitre 3 pour les détails).

La structure des flux numériques est liée à la description physique du système microscopique et du comportement des particules en présence de la barrière

de potentiel représentée par les variations de la topographie (une particule de fluide passe dans une maille voisine ou est réfléchie selon la quantité d'énergie cinétique dont elle dispose).



On montre, dans le Chapitre 3, que le schéma cinétique "avec réflexions" vérifie au niveau discret certaines propriétés importantes des équations de Saint-Venant, notamment la positivité de la hauteur d'eau et la conservation des états stationnaires d'un lac au repos, sous une condition de CFL adaptée. De plus, la discrétisation assure l'absence d'effets dus aux termes sources dans le cas du problème homogène (l'équation (1.8) avec $Z'=0$), o elle s'identifie avec le schéma cinétique standard. Il faut souligner que la condition de CFL introduite dans le Chapitre 3 ne dépend pas explicitement du terme source, ce qui permet de traiter numériquement le cas de termes sources singuliers. La consistance du schéma cinétique ne peut plus s'énoncer ici comme pour les équations scalaires (voir [10]) et une définition spécifique est requise (comme indiqué dans le Chapitre 4).
Enfin, une inégalité d'entropie discrète correspondant à (1.10) est démontrée. On considère l'énergie microscopique associée à l'interprétation cinétique du système de Saint-Venant dans (3.6), qui est définie par la fonctionnelle

$$H(f) = \frac{\xi^2}{2} f + \frac{\pi^2 g^2}{6} f^3 + gZf.$$

On multiplie l'équation cinétique (3.4) par H', pour obtenir la relation suivante vérifiée par les solutions faibles,

$$\frac{\partial H}{\partial t} + \xi \frac{\partial H}{\partial x} - gZ' \frac{\partial H}{\partial \xi} \leq 0.$$

Cette inégalité d'entropie au niveau microscopique a une structure analogue à l'équation cinétique (3.4) et une intégration par rapport à $\xi \in \mathbb{R}$ conduit à la formule macroscopique conservative (1.10), écrite sous la forme

$$\frac{\partial E}{\partial t} + \frac{\partial \eta}{\partial x} \leq 0. \tag{3.14}$$

On identifie alors facilement les fonctions numériques qui figurent dans le schéma dérivé pour (3.14) au sens des volumes finis,

$$\eta^n_{i+\frac{1}{2}} \approx \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} \eta(f)(t, x_{i+\frac{1}{2}}, \xi)\, dt.$$

Le schéma cinétique "avec réflexions" et l'inégalité d'entropie correspondante peuvent être interprétés par la méthode des caractéristiques relative à l'équation (3.4), en appliquant laquelle on retrouve exactement les formules données dans le Chapitre 3 pour les flux numériques cinétiques.

4 Conclusions et perspectives

4.1 Systèmes modifiés et comparaisons expérimentales

Les différentes approximations effectuées pour établir formellement le système des équations de Saint-Venant (1.7)-(1.8) restreignent évidemment son champ d'applications. Néanmoins ces équations fournissent un modèle assez réaliste pour décrire les écoulements de rivières ou de canaux et rendent également compte des mouvements d'eau dans des baies peu profondes, ce qui permet notamment d'étudier les problèmes de marées.

L'écoulement d'un fleuve est, sur la majeure partie de son cours (seules les zones de confluents ou les débordements nécessitent un traitement particulier), un problème pour lequel le système de Saint-Venant monodimensionel reste pertinent. Cependant, les hypothèses considérées habituellement d'écoulement irrotationnel et non-visqueux doivent être corrigées quand on s'intéresse à la simulation numérique à partir des données expérimentales. La prise en compte dans le modèle analytique de la viscosité du fluide et du frottement au fond du canal conduit alors à des systèmes modifiés capables de prévoir des comportements très divers de la surface libre (trains d'ondes, ressauts hydrauliques, déferlement).

Les simulations numériques présentées au Chapitre 4 s'inscrivent dans ce contexte. En particulier, on se concentre sur le profil du fond de la rivière pour analyser le cas de perturbations de l'écoulement dues à un obstacle unique placé au fond d'un canal rectangulaire.
Les données utilisées proviennent des expériences menées dans le cadre de l'ACI "Catastrophes Naturelles : modélisation de processus hydrauliques à surface libre en présence de singularités" (Ministère de la Recherche - France). Pour d'autres études expérimentales d'écoulements permanents au-dessus d'un obstacle, on peut se reporter aussi à [30], ainsi qu'à [20], [21] et [29] pour des simulations numériques de situations réelles.

4.2 Introduction à la page web

Les codes Fortran et les sous-programmes correspondants pour le schéma cinétique "avec réflexions" développé dans le Chapitre 3 sont disponibles à l'adresse Internet suivante :

http ://www-rocq.inria.fr/m3n/CatNat/

Des nombreux tests numériques sont présentés, traités par le schéma d'ordre un et ses extensions d'ordre deux, avec les indications détaillées sur les procédés à suivre pour les reproduire.

4.3 Une application aux écoulements granulaires

La validité des modèles considérés au Chapitre 4 dépend de la précision dans la définition des paramètres physiques utilisés, notamment de la forme du terme de frottement.

Les interactions entre les termes sources correspondant à la topographie du fond et au frottement dans le système de Saint-Venant sont particulièrement importantes pour déterminer les états stationnaires.

Différentes approches pour la discrétisation des termes de frottement doivent être mises au point selon la nature spécifique de ces termes. Si cette question parait relativement simple à résoudre dans le cas d'un frottement fluide, le problème devient beaucoup plus complexe lorsqu'il s'agit de simuler des écoulements granulaires. Pour ce dernier cas, l'analyse de la dynamique microscopique du système conduit à la formulation d'un terme de frottement discontinu, difficile à traiter par des méthodes classiques.

Une première solution du problème numérique des états stationnaires pour un modèle d'avalanche est proposée dans l'Appendice. La méthode est basée sur une représentation cinétique des équations de type Saint-Venant qui décrivent les phénomènes et qui sont traitées ensuite par des schémas cinétiques adaptés. Des progrès dans ce domaine devraient ouvrir un vaste champ d'applications nouvelles.

Bibliographie

[1] M. Abduraimov, K. Muzafarov, A.A. Puttiev, The motion of water in open channels (the Saint-Venant equation), *Mat. Model.*, **10** (1998), no. 6, 97-106

[2] V.I. Agoshkov, D. Ambrosi, V. Pennati, A. Quarteroni, F. Saleri, Mathematical and numerical modelling of shallow water flow, *Comput. Mech.*, **11** (1993), no. 5-6, 280-299

[3] F. Alcrudo, F. Benkhaldoun, Exact solutions to the Riemann problem of the shallow water equations with a bottom step, *Comput. & Fluids*, **30** (2001), no. 6, 643-671

[4] D. Amadori, G. Guerra, Global weak solutions for systems of balance laws, *Appl. Math. Lett.*, **12** (1999), no. 6, 123-127

[5] D. Amadori, L. Gosse, G. Guerra, Global BV entropy solutions and uniqueness for hyperbolic systems of balance laws, *Arch. Ration. Mech. Anal.* (2002)

[6] D.S. Bale, R.J. LeVeque, S. Mitran, J.A. Rossmanith, A wave propagation method for conservation laws and balance laws with spatial varying flux functions, *Appl. Math. Tech. Rep.*, Univ. of Washington, Seattle (2001)

[7] V.B. Barakhnin, TVD scheme of second-order approximation on a nonstationary adaptive grid for hyperbolic systems, *Russian J. Numer. Anal. Math. Modelling*, **16** (2001), no. 1, 1-17

[8] M. Berger, R.J. LeVeque, Adaptive mesh refinement using wave-propagation algorithms for hyperbolic systems, *SIAM J. Numer. Anal.*, **35** (1998), no. 6, 2298-2316

[9] A. Bermúdez, A. Dervieux, J.A. Desideri, M.E. Vázquez, Upwind schemes for the two-dimensional shallow water equations with variable depth using unstructured meshes, *Comput. Methods Appl.Mech.Engrg.*, **155** (1998), no. 1-2, 49-72

[10] R. Botchorishvili, B. Perthame, A. Vasseur, Equilibrium schemes for scalar conservation laws with stiff sources, *Math. Comp.*, to appear

[11] T. Buffard, T. Gallouet, J.M. Hérard, Un schéma simple pour les équations de Saint-Venant, *C. R. Acad. Sci. Paris Sér.I Math.*, **326** (1998), no. 3, 385-390

[12] M.J. Castro, P. Garcia-Navarro, The application of a conservative grid adaptation technique to 1D shallow water equations, *Math. Comput. Modelling*, **34** (2001), no. 1-2, 29-35

[13] J.J. Cauret, J.F. Colombeau, A.Y. LeRoux, Solutions généralisées discontinues de problèmes hyperboliques non conservatifs, *C. R. Acad. Sci. Paris Sér.I Math.*, **302** (1986), no. 12, 435-437

[14] A. Chalabi, Stable upwind schemes for hyperbolic conservation laws with source terms, *IMA J. Numer. Anal.*, **12** (1992), no. 2, 217-241

[15] A. Chalabi, Y. Qiu, Relaxation schemes for hyperbolic conservation laws with stiff source terms : application to reacting Euler equations, *J. Sci. Comput.*, **15** (2000), no. 4, 395-416

[16] G. Crasta, B. Piccoli, Viscosity solutions and uniqueness for systems of inhomogeneous balance laws, *Discrete Contin. Dynam. Systems*, **3** (1997), no. 4, 477-502

[17] C.M. Dafermos, *Hyperbolic conservation laws in continuum physics*, Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences) **325**, Springer-Verlag, Berlin, 2000

[18] C.M. Dafermos, L. Hsiao, Hyperbolic systems and balance laws with inhomogeneity and dissipation, *Indiana Univ. Math. J.*, **31** (1982), no. 4, 471-491

[19] A.I. Delis, T. Katsaounis, Relaxation schemes for the shallow water equations, *DMA-ENS Report* (2002)

[20] A.I. Delis, C.P. Skeels, TVD schemes for open channel flow, *Internat. J. Numer. Methods Fluids*, **26** (1998), 791-809

[21] A.I. Delis, C.P. Skeels, S.C. Ryrie, Evaluation of some approximate Riemann solvers for transient open channel flows, *J. Hydraulic Research*, **38** (2000), 217-232

[22] H. Frid, Uniqueness of solutions to hyperbolic balance laws in several space dimensions, *Comm. Partial Differential Equations*, **14** (1989), no. 8-9, 959-979

[23] P. Garcia-Navarro, M.E. Vázquez-Cendón, On numerical treatment of the source terms in the shallow water equations, *Comput. & Fluids*, **29** (2000), 951-979

[24] J.F. Gerbeau, B. Perthame, Derivation of viscous Saint-Venant system for laminar shallow water ; numerical validation, *Discrete Contin. Dyn. Syst.* Ser. B1 (2001), no. 1, 89-102

[25] P. Glaister, An efficient numerical method for subcritical and supercritical open channel flows, *Appl. Numer. Math.*, **11** (1993), no. 6, 497-508

[26] J.M. Greenberg, A.Y. LeRoux, A well-balanced scheme for the numerical processing of source terms in hyperbolic equations, *SIAM J. Numer. Anal.*, **33** (1996), 1-16

[27] E. Godlewski, P.A. Raviart, *Numerical approximation of hyperbolic systems of conservation laws*, Applied Mathematical Sciences **118**, New York, Springer-Verlag, 1996

[28] L. Gosse, Localization effects and measure source terms in numerical schemes for balance laws, *Math. Comp.*, **71** (2002), no. 238, 553-582

[29] N. Goutal, Finite element solution for the transcritical shallow-water equations, *Math. Methods Appl. Sci.*, **11** (1989), no. 4, 503-524

[30] N. Goutal, F. Maurel, Proceedings of the 2nd workshop on dam-break wave simulation, *EDF-DER Report* HE-43/97/016B (1997)

[31] M.E. Hubbard, M.J. Baines, Conservative multidimensional upwinding for the steady two-dimensional shallow water equations, *J. Comput. Phys.*, **138** (1997), no. 2, 419-448

[32] M.E. Hubbard, P. Garcia-Navarro, Flux difference splitting and the balancing of source terms and flux gradients, *J. Comput. Phys.*, **165** (2000), no. 1, 89-125

[33] T. Katsaounis, C. Makridakis, Relaxation models and finite element schemes for the shallow water equations, *DMA-ENS Report* (2002)

[34] T. Katsaounis, C. Makridakis, Adaptive finite element relaxation schemes for the Saint-Venant system, Dept. of Applied Mathematics, University of Crete, preprint (2001)

[35] C. Klingenberg, Y. Lu, Existence of solutions to hyperbolic conservation laws with a source, *Comm. Math. Phys.*, **187** (1997), no. 2, 327-340

[36] D. Kröner, *Numerical schemes for conservation laws*, Wiley-Teubner Series Advances in Numerical Mathematics, John Wiley & Sons, Ltd., Chichester ; B.G. Teubner, Stuttgart, 1997

[37] S.N. Kružkov, First order quasilinear equations in several independent space variables, *Math. USSR Sb.*, **10** (1970), 217-243

[38] P. Lax, Shock waves and entropy, *Contributions to nonlinear functional analysis* (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971), pp. 603-634, Academic Press, New York, 1971

[39] A.Y. Le Roux, Riemann solvers for some hyperbolic problems with a source term, Actes du 30ème Congrès d'Analyse Numérique : CANum '98 (Arles, 1998), pp. 75-90, *ESAIM Proc.*, **6**, Soc. Math. Appl. Indust., Paris, 1999

[40] R.J. LeVeque, D.S. Bale, Wave propagation methods for conservation laws with source terms, Hyperbolic problems : theory, numerics, applications, Vol. II (Zrich, 1998), *Internat. Ser. Numer. Math.*, **130**, pp. 609-618, Birkhuser, Basel, 1999

[41] R.J. LeVeque, H.C. Yee, A study of numerical methods for hyperbolic conservation laws with stiff source terms, *J. Comput. Phys.*, **86** (1990), no. 1, 187-210

[42] M. Lewicka, On the well posedness of a system of balance laws with L^∞ data, *Rend. Sem. Mat. Univ. Padova*, **102** (1999), 319-340

[43] P.L. Lions, B. Perthame, P.E. Souganidis, Existence and stability of entropy solutions for the hyperbolic systems of isentropic gas dynamics in Eulerian and Lagrangian coordinates, *Comm. Pure Appl. Math.*, **49** (1996), no. 6, 599-638

[44] P.L. Lions, B. Perthame, E. Tadmor, A kinetic formulation of multidimensional scalar conservation laws and related equations, *J. Amer. Math. Soc.*, **7** (1994), no. 1, 169-191

[45] S.F. Liotta, V. Romano, G. Russo, Central schemes for balance laws of relaxation type, *SIAM J. Numer. Anal.*, **38** (2000), no. 4, 1337-1356

[46] T.P. Liu, Nonlinear resonance for quasilinear hyperbolic equations, *J. Math. Phys.*, **28** (1987), no. 11, 2593-2602

[47] T.P. Liu, *Hyperbolic and viscous conservation laws*, CBMS-NSF Regional Conference Series in Applied Mathematics, **72**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000

[48] A. Nicolau, Le système de Saint-Venant unidimensionnel en distributions, *Rev. Roumaine Math. Pures Appl.*, **31** (1986), no. 10, 905-914

[49] L. Pareschi, Central differencing based numerical schemes for hyperbolic conservation laws with relaxation terms, *SIAM J. Numer. Anal.*, **39** (2001), no. 4, 1395-1417

[50] B. Perthame, An introduction to kinetic schemes for gas dynamics, An introduction to recent developments in theory and numerics for conservation laws (Freiburg/Littenweiler, 1997), *Lect. Notes Comput. Sci. Eng.*, **5**, pp. 1-27, Springer, Berlin, 1999

[51] B. Perthame, Uniqueness and error estimates in first order quasilinear conservation laws via the kinetic entropy defect measure, *J. Math. Pures Appl. (9)*, **77** (1998), no. 10, 1055-1064

[52] G. Russo, Central schemes for balance laws, Hyperbolic problems : theory, numerics, applications, Vol. I, II (Magdeburg, 2000), *Internat. Ser. Numer. Math.*, **140, 141**, pp. 821-829, Birkhuser, Basel, 2001

[53] J. Santos, P. de Oliveira, A converging finite volume scheme for hyperbolic conservation laws with source terms, Numerical methods for differential equations (Coimbra, 1998), *J. Comput. Appl. Math.*, **111** (1999), no. 1-2, 239-251

[54] D. Serre, *Systèmes hyperboliques de lois de conservation*, Tomes I et II, Diderot Editeur, Paris, 1996

[55] Vasseur A., Well-posedness of scalar conservation laws with singular sources, preprint

[56] M.E. Vázquez-Cendón, Improved treatment of source terms in upwind schemes for the shallow water equations in channels with irregular geometry, *J. Comput. Phys.*, **148** (1999), no. 2, 497-526

[57] J.G. Zhou, D.M. Causon, C.G. Mingham, D.M. Ingram, The surface gradient method for the treatment of source terms in the shallow-water equations, *J. Comput. Phys.*, **168** (2001), no. 1, 1-25

Première partie

LA MÉTHODE "UPWIND INTERFACE SOURCE" POUR LES LOIS DE CONSERVATION HYPERBOLIQUES

Chapitre 1

Convergence de la méthode "Upwind Interface Source" pour les lois de conservation hyperboliques

(soumis à *Proceedings of Hyp2002*)

Convergence of the Upwind Interface Source method for hyperbolic conservation laws

B. Perthame, C. Simeoni

Département de Mathématiques et Applications
École Normale Supérieure
45, rue d'Ulm - 75230 Paris Cedex 05 - France
e-mails: Benoit.Perthame@ens.fr, Chiara.Simeoni@ens.fr

Abstract

This paper deals with typical questions arising in the analysis of numerical approximations for scalar conservation laws with a source term. We focus our attention on semi-discrete finite volume schemes, in the general case of a nonuniform spatial mesh. To define appropriate discretizations of the source term, we introduce the formalism peculiar to the Upwind Interface Source method and we establish conditions on the numerical functions so that the discrete solver preserves the steady state solutions. Then we formulate a rigorous definition of consistency, adapted to the class of well-balanced schemes, for which we are able to prove a Lax-Wendroff type convergence theorem. Some examples of numerical methods are discussed, in order to validate the arguments we propose.

Key-words: hyperbolic conservation laws, source terms, upwind interfacial methods, well-balanced schemes, consistency, convergence.

1 Introduction

We consider a scalar conservation law with a source term, in one space dimension,

$$\frac{\partial u}{\partial t} + \frac{\partial A(u)}{\partial x} + B(x,u) = 0, \quad t \in \mathbb{R}_+,\ x \in \mathbb{R}, \tag{1.1}$$

with $u(t,x) \in \mathbb{R}$ and a real-valued flux function A, associated with a Cauchy problem by introducing the initial condition

$$u(0,x) = u^0(x) \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R}). \tag{1.2}$$

We set

$$a(u) = A'(u) \in \mathcal{C}^1(\mathbb{R}) \tag{1.3}$$

and we restrict our analysis to a particular form of source term,

$$B(x,u) = z'(x)b(u), \quad z' \in L^1(\mathbb{R}),\ b \in \mathcal{C}^1(\mathbb{R}). \tag{1.4}$$

This is suggested by the usual application of hyperbolic conservation laws as simple mathematical models in continuum mechanics: in the Saint-Venant system for shallow water, for instance, $z(x)$ describes the bottom topography. The equation (1.1) is endowed with the family of entropy inequalities

$$\frac{\partial S(u)}{\partial t} + \frac{\partial \eta(u)}{\partial x} + S'(u)B(x,u) \leq 0, \quad \eta'(u) = S'(u)a(u), \tag{1.5}$$

for any pair of a convex entropy function S and the corresponding entropy flux η (see [21] and [24]). Under stronger assumptions on the source term, Kružkov [21] proved existence and uniqueness of the entropy solution for the initial value problem (1.1)-(1.2), in the functional space $L^\infty([0,T); L^1(\mathbb{R}))$, for all $T \in \mathbb{R}_+$. Many results concerning the convergence of numerical approximations for the entropy solution of hyperbolic conservation laws are inspired by this fundamental theory. In the case of singular source terms (i.e. the function $z(x)$ is discontinuous), a remarkable uniqueness result has recently been proved by Vasseur [31].

The presence of source terms modifies the analytical properties of the equation (1.1), in comparison with the homogeneous case. More specifically, a fundamental change is the occurrence of other kinds of steady state solutions, resulting from the balance between source terms and internal forces, given by the formula

$$D(u) + z(x) = C^{st}, \quad D'(u) = \frac{a(u)}{b(u)}. \tag{1.6}$$

This fact also influences the numerical approach to the problem, as it was pointed out by several authors (refer to [15] and [27]), in order to investigate discrete approximations preserving the properties of the continuous system.

A well-known difficulty encountered in the numerical treatment of hyperbolic conservation laws with a source term relates to the approximation of such a source term, to assure that the scheme preserves the steady state solutions at discrete level.

Initially for scalar problems, Greenberg, LeRoux and others introduced the notion of well-balanced schemes (see [15],[16] for details). This definition has been further developed by Gosse and LeRoux [11], which used a reformulation of the source terms by means of non-conservative products to derive numerical fluxes at the interfaces of an unstructured mesh. A recent approach by LeVêque [27] is based on the Godunov scheme extended for an appropriately modified system. Botchorishvili, Perthame and Vasseur present in [2] a kinetic scheme, that maintains steady states and which is proved to converge when stiff source terms are considered. Using interfacial values, instead of the cell-averages, for the source term, Jin proposes in [17] a rather simple method for capturing steady state solutions with a high order accuracy. Previous schemes have also been modified for this target by Bermudez and Vasquez [1] and some different approaches are developed in [22],[23] and [18]. Quite recently, these kinds of numerical processing have been extended to hyperbolic systems of balance laws (like the Saint-Venant system for shallow water), to obtain stable schemes which preserve the steady states (see [8], [13],[14] and [28], for instance). In particular, one of the main conclusions in [28] is that, while preserving steady states, well-balanced schemes can also enjoy stability under the usual CFL condition (independent of z').

The aim of this paper is to present a general consistency condition for discrete approximations of equation (1.1). In fact, to analyze theoretical properties of numerical solvers for a conservation law with a source term, the only classical condition on the flux function is not enough and specific definitions for the discrete source term are required.

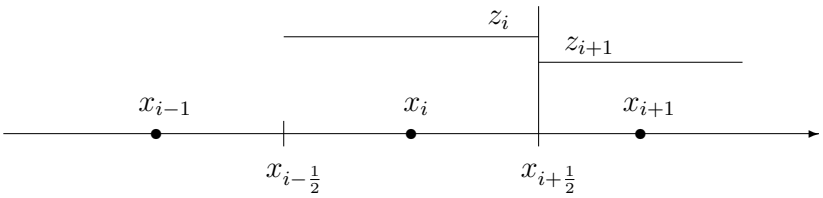
The outline of the paper is the following. In Section 2, we illustrate the Upwind Interface Source method, which consists in upwinding source terms at the interfaces of the mesh cells, as usual for the fluxes according to the finite volume formalism. Then, in Section 3, we consider discretizations which preserve steady state solutions (well-balanced schemes) and we review several classical methods to build such schemes. The question of consistency is addressed in Section 4 and we show that well-balanced schemes are consistent in the sense we have established. By using these arguments, in Section 5, we finally prove an extension of the Lax-Wendroff theorem.

2 Upwind Interface Source method

The finite volume method is possible for treating numerically hyperbolic systems of conservation laws, it is robust and presents the advantage to be conservative (we refer to [6] for a survey of its properties). We look at the semi-discrete scheme, called method of lines, where only a space discretization

of equation (1.1) is performed.

We consider a mesh of $\mathbb{R}$ made up of cells $\mathcal{C}_i$, with center x_i, $i \in \mathbb{Z}$ and nonuniform length Δx_i; we denote $x_{i+\frac{1}{2}}$ the cell interfaces, so that the control volume can be identified as $\mathcal{C}_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$ and $x_i = \frac{x_{i-\frac{1}{2}}+x_{i+\frac{1}{2}}}{2}$. Then, we construct a piecewise constant representation of the function $z(x)$ on the mesh, whose coefficients are $z_i = \frac{1}{\Delta x_i}\int_{\mathcal{C}_i} z(x)\,dx$ for example.



In this context, the discrete unknowns are expected to be approximations of the mean values of u on the mesh cells (the conservative quantities are cell centered),

$$u_i(t) \approx \frac{1}{\Delta x_i}\int_{\mathcal{C}_i} u(t,x)\,dx, \quad t \in \mathbb{R}_+,\ i \in \mathbb{Z},$$

while the numerical fluxes are defined at the interfaces of the mesh.

To correctly treat the source term is more difficult than it seems and centered schemes give unsatisfactory results, as it is well reported in the literature: a direct discretization of the source term by cell-averages, for instance, can not preserve the steady state solutions. A better approach is based on the Upwind Interface Source method (upwinding of external terms was originally formulated by Roe [29]), where the source term is also upwinded at the interfaces.

The general finite volume scheme for equation (1.1) can thus be written in the explicit form

$$\Delta x_i \frac{du_i}{dt} + (A_{i+\frac{1}{2}} - A_{i-\frac{1}{2}}) + \mathcal{B}^+_{i-\frac{1}{2}} + \mathcal{B}^-_{i+\frac{1}{2}} = 0, \tag{2.1}$$

dropping the time dependence of the numerical functions for simplicity.
We proceed to explain the notation in the previous formula. We first introduce the discrete fluxes

$$A_{i+\frac{1}{2}} = \mathcal{A}(u_i, u_{i+1}), \quad \mathcal{A} \in \mathcal{C}^1, \tag{2.2}$$

where the numerical function $\mathcal{A}$ is chosen as a consistent approximation of the analytical flux,

$$\mathcal{A}(u,u) = A(u). \tag{2.3}$$

Because of the choice of particular source terms (1.4), the function $z(x)$ is defined up to a constant. Therefore, without loss of generality, we suppose the source term is discretized at the cell interfaces by means of functions

$$\mathcal{B}^{\pm}_{i+\frac{1}{2}} = \mathcal{B}^{\pm}(u_i, u_{i+1}, z_{i+1} - z_i), \qquad \mathcal{B}^{\pm} \in \mathcal{C}^2, \tag{2.4}$$

and, in view of (1.4), it is natural to impose

$$\mathcal{B}^{+}(u, v, 0) = \mathcal{B}^{-}(u, v, 0) = 0. \tag{2.5}$$

The last condition refers to the interpretation of the numerical solver (2.1), applied to the model (1.1)-(1.4). According to the finite volume formalism, we can identify

$$\mathcal{B}^{+}_{i-\frac{1}{2}} + \mathcal{B}^{-}_{i+\frac{1}{2}} \approx \int_{\mathcal{C}_i} z'(x) b(u) dx; \tag{2.6}$$

formally, this leads to deduce

$$\mathcal{B}^{-}_{i+\frac{1}{2}} + \mathcal{B}^{+}_{i+\frac{1}{2}} \approx \int_{x_i}^{x_{i+\frac{1}{2}}} z'(x) b(u) dx + \int_{x_{i+\frac{1}{2}}}^{x_{i+1}} z'(x) b(u) dx, \tag{2.7}$$

that is a way to perform an interfacial approximation of zero order terms. Such a discretization is also upwinded, in the sense that $\mathcal{B}^{-}_{i+\frac{1}{2}}$ represents the contribution of the waves coming from the left of the interface $x_{i+\frac{1}{2}}$ and moving towards the cell $\mathcal{C}_i$ if they have a nonpositive velocity, while $\mathcal{B}^{+}_{i+\frac{1}{2}}$ represents the waves moving forwards from the right of $x_{i+\frac{1}{2}}$ and counted only if they have nonnegative velocity. Notice that, when the problem becomes homogeneous (for example, $z'(x) = 0$ in equation (1.4), motivated by the analogy with the Saint-Venant model), this scheme reduces to the usual finite volume approximation for a scalar conservation law.

We observe that all what is stated in this section and in the following is also valid for a fully explicit scheme (obtained, for instance, using a standard forward Euler method for the time discretization),

$$\frac{\Delta x_i}{\Delta t}(u_i^{n+1} - u_i^n) + (A^n_{i+\frac{1}{2}} - A^n_{i-\frac{1}{2}}) + \mathcal{B}^{n,+}_{i-\frac{1}{2}} + \mathcal{B}^{n,-}_{i+\frac{1}{2}} = 0,$$

where we introduce a time-step Δt and set $t_n = n\Delta t$, $n \in \mathbb{N}$. We then have to consider an additional restriction on the size of the ratio $\frac{\Delta t}{\Delta x_i}$, the usual CFL condition, to guarantee numerical stability.

3 Well-balanced schemes

We define general conditions on the discretizations $\mathcal{B}^{\pm}_{i+\frac{1}{2}}$ so that the numerical scheme (2.1) preserves the steady state solutions. Note that all the methods developed in the references mentioned above are compatible with the formalism introduced in Section 2 and can be put in form (2.1), as we will do later for some particular cases.

By integrating the stationary equation associated with (1.1)-(1.4), we obtain the algebraic relation (1.6) for smooth steady state solutions. A discrete version is given by

$$D(u_i) + z_i = C^{st}, \quad \forall\ i \in \mathbb{Z}. \tag{3.1}$$

We consider appropriate hypotheses on D, to ensure the existence of a unique Lipschitz continuous solution of that problem, namely that D is strictly monotonic. This assumption is restrictive and not always satisfied in realistic situations, but it is usually made (we are aware of that does use it in [2]).

For all initial data defined according to (3.1), a solver preserving steady states has to verify

$$(A_{i+\frac{1}{2}} - A_{i-\frac{1}{2}}) + \mathcal{B}^{+}_{i-\frac{1}{2}} + \mathcal{B}^{-}_{i+\frac{1}{2}} = 0.$$

This last statement can be formulated in terms of numerical functions, thanks to definition (2.2) and (2.4), so that it writes

$$\mathcal{A}(u_i, u_{i+1}) - \mathcal{A}(u_{i-1}, u_i) + \mathcal{B}^{+}(u_{i-1}, u_i, z_i - z_{i-1}) + \mathcal{B}^{-}(u_i, u_{i+1}, z_{i+1} - z_i) = 0,$$

for all u_{i-1}, u_i, u_{i+1}, such that $D(u_j) + z_j = H$, $j = i-1, i, i+1$. In particular, we choose alternatively $u_{i-1} = u_i$ and $u_i = u_{i+1}$ (then we deduce, respectively, $z_{i-1} = z_i$ and $z_i = z_{i+1}$), to obtain equivalent conditions at the interfaces, also exploiting properties (2.3) and (2.5),

$$\begin{aligned} \mathcal{A}(u_i, u_{i+1}) - A(u_i) + \mathcal{B}^{-}(u_i, u_{i+1}, z_{i+1} - z_i) = 0, \\ A(u_{i+1}) - \mathcal{A}(u_i, u_{i+1}) + \mathcal{B}^{+}(u_i, u_{i+1}, z_{i+1} - z_i) = 0. \end{aligned}$$

We summarize the previous statements in the following proposition.

Lemma 3.1. *A numerical scheme in form (2.1)-(2.5) is well-balanced, i.e. it preserves the steady state solutions (3.1), if and only if the equalities*

$$\mathcal{A}(u, v) - A(u) + \mathcal{B}^{-}(u, v, z_+ - z_-) = 0, \tag{3.2}$$

$$A(v) - \mathcal{A}(u, v) + \mathcal{B}^{+}(u, v, z_+ - z_-) = 0, \tag{3.3}$$

hold true, for all u, v, z_-, z_+ such that

$$D(u) + z_- = D(v) + z_+. \tag{3.4}$$

We call well-balanced or Steady State Preserving schemes the numerical solvers for the problem (1.1)-(1.2) which satisfy those conditions.
We present some numerical schemes to which Lemma 3.1 applies. We check these approaches enable to preserve the discrete steady state solutions, according to the result stated above.

B.P.V. method In [2], the authors introduce their solver in a compact form, taking into account the source term directly in the definition of the numerical fluxes,

$$\Delta x_i \frac{du_i}{dt} + (A^-_{i+\frac{1}{2}} - A^+_{i-\frac{1}{2}}) = 0, \tag{3.5}$$

with

$$A^-_{i+\frac{1}{2}} = \mathcal{A}(u_i, u^-_{i+1}), \quad A^+_{i-\frac{1}{2}} = \mathcal{A}(u^+_{i-1}, u_i). \tag{3.6}$$

The numerical flux used in [2] is given by a standard Engquist-Osher function, but one readily finds out that similar methods can be formulated with any consistent flux function $\mathcal{A}$. The points u^-_{i+1} and u^+_{i-1} are defined by means of the relations

$$D(u^-_{i+1}) + z_i = D(u_{i+1}) + z_{i+1}, \tag{3.7}$$
$$D(u^+_{i-1}) + z_i = D(u_{i-1}) + z_{i-1}. \tag{3.8}$$

To reproduce this scheme in form (2.1), we identify

$$\mathcal{B}^+_{i-\frac{1}{2}} = \mathcal{A}(u_{i-1}, u_i) - \mathcal{A}(u^+_{i-1}, u_i), \quad \mathcal{B}^-_{i+\frac{1}{2}} = \mathcal{A}(u_i, u^-_{i+1}) - \mathcal{A}(u_i, u_{i+1}).$$

For a steady state, we immediately deduce from (3.1), (3.7) and (3.8) that

$$u^-_{i+1} = u_i, \quad u^+_{i-1} = u_i,$$

then resulting in the conditions (3.2) and (3.3).

This method extends to more general classes of function D, such as quadratic functions, and it also applies to hyperbolic systems of conservation laws endowed with a kinetic interpretation (see [28]).

We remark that, combined with specific approximate Riemann solvers, the algorithm (3.5)-(3.6) can be interpreted as the well-balanced scheme derived by Greenberg and LeRoux [15] or by Gosse and LeRoux [11], for which the condition (3.2)-(3.4) is verified.

The quasi-steady wave-propagation algorithm The basic idea of the method developed by LeVêque [27] is to introduce a new Riemann problem in the center of each mesh cell, with values u_i^- on the left half of the cell

and u_i^+ on the right half, whose flux difference exactly cancels the effect of the source term. These artificial states are defined so that the cell-average is preserved,

$$u_i^- = u_i - \delta_i, \quad u_i^+ = u_i + \delta_i, \quad \frac{1}{2}(u_i^- + u_i^+) = u_i; \tag{3.9}$$

moreover, if δ_i is chosen according to the in-cell balance condition

$$A(u_i^+) - A(u_i^-) + z_i' b(u_i)\Delta x_i = 0, \tag{3.10}$$

then the jumps occurring at the cell interfaces will correspond to perturbations from the steady states. Note that (3.10) represents a discrete version of the stationary problem associated with equation (1.1). The explicit formula for the scheme thus obtained looks like the classical Godunov solver,

$$\Delta x_i \frac{du_i}{dt} + \left(\Delta^+ A(u_{i-1}^+, u_i^-) + \Delta^- A(u_i^+, u_{i+1}^-)\right) = 0,$$

with

$$\Delta^+ A(u_{i-1}^+, u_i^-) = A(u_i^-) - A(u_{i-\frac{1}{2}}^*), \tag{3.11}$$

$$\Delta^- A(u_i^+, u_{i+1}^-) = A(u_{i+\frac{1}{2}}^*) - A(u_i^+), \tag{3.12}$$

where $u_{i+\frac{1}{2}}^*$ now denotes the solution to the modified Riemann problem at the cell interfaces, between values u_i^+ and u_{i+1}^-. If the solution we are looking for is quasi-steady then we deduce from (3.9) and (3.10) that $u_i^+ \approx u_{i+1}^-$, as δ_i tends to 0, so that the steady states are asymptotically preserved.

As for previous examples, this method can extend to any consistent numerical flux function $\mathcal{A}$, by rewriting the flux differences (3.11)-(3.12) in the more general form

$$\Delta^+ A(u_{i-1}^+, u_i^-) = A(u_i^-) - \mathcal{A}(u_{i-1}^+, u_i^-),$$
$$\Delta^- A(u_i^+, u_{i+1}^-) = \mathcal{A}(u_i^+, u_{i+1}^-) - A(u_i^+).$$

Jin's formulas A simple scheme for handling hyperbolic systems of conservation laws with source terms is proposed in [17], which preserves the steady state solutions exactly at the cell interfaces. For methods based on a generalized Riemann solver, it takes the form (2.1) and the source term is discretized by

$$\mathcal{B}_{i-\frac{1}{2}}^+ + \mathcal{B}_{i+\frac{1}{2}}^- = \frac{A_{i+\frac{1}{2}} - A_{i-\frac{1}{2}}}{D_{i+\frac{1}{2}} - D_{i-\frac{1}{2}}}(z_{i+\frac{1}{2}} - z_{i-\frac{1}{2}}), \tag{3.13}$$

using interface values $A_{i+\frac{1}{2}} = A(u_{i+\frac{1}{2}})$ and $D_{i+\frac{1}{2}} = D(u_{i+\frac{1}{2}})$, rather than the cell-averages. Consequently, if one gets a steady state at the interfaces,

$$D_{i+\frac{1}{2}} + z_{i+\frac{1}{2}} = C^{st}, \quad \forall i \in \mathbb{Z},$$

a direct computation leads to verify that it is preserved, as we have

$$(A_{i+\frac{1}{2}} - A_{i-\frac{1}{2}}) + \frac{A_{i+\frac{1}{2}} - A_{i-\frac{1}{2}}}{D_{i+\frac{1}{2}} - D_{i-\frac{1}{2}}}(z_{i+\frac{1}{2}} - z_{i-\frac{1}{2}}) = 0.$$

A more generic scheme, again proposed in [17], is defined by

$$\Delta x_i \frac{du_i}{dt} + (A_{i+\frac{1}{2}} - A_{i-\frac{1}{2}}) + \frac{b_{i-\frac{1}{2}} + b_{i+\frac{1}{2}}}{2}(z_{i+\frac{1}{2}} - z_{i-\frac{1}{2}}) = 0. \tag{3.14}$$

Although it is not possible to derive an explicit expression of D for a general flux function A, some applications considered by the author (shallow water equations, for instance) show this method yields formally second order approximations to the steady states at the interfaces, as suggested by an asymptotic expansion of (3.14).

The numerical discretizations formulated by Jin are called Steady State Capturing schemes, that is a weaker definition since only the interface values are preserved. According to the idea to process the source term by making use explicitly of relations on the steady states, a Steady State Preserving variation of (3.13), which agrees with the general formalism (2.1)-(2.4), is given by

$$\mathcal{B}^+_{i-\frac{1}{2}} = \frac{\mathcal{A}(u_{i-1}, u_i) - A(u_i)}{D(u_{i-1}) - D(u_i)}(z_i - z_{i-1}),$$
$$\mathcal{B}^-_{i+\frac{1}{2}} = \frac{\mathcal{A}(u_i, u_{i+1}) - A(u_i)}{D(u_{i+1}) - D(u_i)}(z_{i+1} - z_i).$$

Again we can readily check the condition (3.2)-(3.4) for this method.

4 Consistency

In order to investigate theoretical properties of the Upwind Interface Source method, a crucial question to discuss is that equation (2.1) verifies the consistency with the continuous equation (1.1)-(1.4).

We indicate a rigorous definition of consistency, which also results to be satisfied by well-balanced schemes.

Definition 4.1. *A numerical scheme in form (2.1)-(2.5) is said to be consistent with (1.1) if the following limit is verified, locally uniformly in u,*

$$\lim_{\lambda \to 0} \frac{\mathcal{B}^+(u,u,\lambda) + \mathcal{B}^-(u,u,\lambda)}{\lambda} = b(u). \tag{4.1}$$

We point out that the above definition of consistency for the source term, in finite volume sense, does not imply that the consistency error vanishes just as for the flux terms. Indeed, because of the choice of an arbitrary nonuniform spatial mesh, the space-step Δx_i could be very different from the length of an interfacial interval $\Delta x_{i+\frac{1}{2}} = |x_{i+1} - x_i| = \Delta x_i/2 + \Delta x_{i+1}/2$; therefore, the corresponding interfacial discretizations (2.6) and (2.7) may have very different values. The condition (4.1) we have established is closer to (2.7), which is the most appropriate interpretation of the discrete source term for the general method illustrated in this paper.

As it will be seen clearly in next section, consistency plays an important role to achieve convergence properties of a numerical solver, in particular to prove that the strong limit of discrete approximations (as the mesh is refined) is the suitable weak solution of the continuous problem.

The following result guarantees consistency for the numerical schemes described in Section 3.

Lemma 4.2. *We assume D is monotonic. Let a numerical solver for the system (1.1)-(1.4) satisfy the conditions (3.2)-(3.4) in Lemma 3.1, then the property (4.1) is verified. In other words, all Steady State Preserving schemes are consistent.*

Proof. We perform a Taylor expansion of the relation (3.4) and we deduce that, for some $\xi \in (u,v)$,

$$D'(\xi)(u-v) = z_+ - z_-. \tag{4.2}$$

After adding equality (3.2) to (3.3), thanks to the definition (1.3) and (1.6), this leads to

$$\mathcal{B}^+(u,v,z_+-z_-)+\mathcal{B}^-(u,v,z_+-z_-) = A(u)-A(v) = \frac{a(\zeta)}{D'(\xi)}(z_+-z_-), \tag{4.3}$$

for some $\zeta \in (u,v)$. We also note that the regularity assumed for the numerical functions enables to perform the general approximations

$$\begin{aligned}\mathcal{B}^{\pm}(u,v,z_+-z_-) &= \mathcal{B}^{\pm}(u,u,z_+-z_-) \\ &\quad + \frac{\partial}{\partial v}\mathcal{B}^{\pm}(u,u,z_+-z_-)(v-u) + O(|v-u|^2).\end{aligned}$$

It thus follows from (2.5) that

$$\frac{\partial}{\partial v}\mathcal{B}^+(u,u,0) = \frac{\partial}{\partial v}\mathcal{B}^-(u,u,0) = 0$$

and, in view of (4.2), we obtain that

$$\begin{aligned} &\lim_{z_+ - z_- \to 0} \frac{\mathcal{B}^+(u,u,z_+ - z_-) + \mathcal{B}^-(u,u,z_+ - z_-)}{z_+ - z_-} \\ &= \lim_{z_+ - z_- \to 0} \frac{\mathcal{B}^+(u,v,z_+ - z_-) + \mathcal{B}^-(u,v,z_+ - z_-)}{z_+ - z_-}. \end{aligned} \tag{4.4}$$

By combining relation (4.3) with (4.4), since

$$\frac{a(\zeta)}{D'(\xi)} \longrightarrow \frac{a(u)}{D'(u)} = b(u),$$

we finally conclude that the property (4.1) is satisfied. □

5 A Lax-Wendroff type convergence theorem

We are now interested in the convergence of the numerical scheme (2.1), as the mesh size tends to zero, by analyzing the convergence properties of its solution $\{u_i(t)\}_{i\in\mathbb{Z}}$.
A discretization of the initial condition is given, for instance, by the sequence

$$u_i^0 = \frac{1}{|\mathcal{C}_i|}\int_{\mathcal{C}_i} u^0(x)dx, \quad i \in \mathbb{Z}.$$

As a measure of mesh refinement, we consider the parameter

$$h = \sup_{i\in\mathbb{Z}} \Delta x_i.$$

For our purpose, we introduce the piecewise constant function u_h, defined a.e. in $\mathbb{R}_+ \times \mathbb{R}$ by

$$u_h(t,x) = \sum_{i\in\mathbb{Z}} u_i(t)\mathbb{1}_{\mathcal{C}_i}(x), \tag{5.1}$$

and we study the convergence towards a solution to the problem (1.1)-(1.2), as h tends to 0.

Theorem 5.1. *Assume $z \in W^{1,1}$ for the source term (1.4). Let u_h be obtained from a numerical scheme in form (2.1)-(2.5), which satisfies the consistency condition (4.1). Suppose there exists a constant C such that, uniformly in h,*

$$||u_h||_{L^\infty_{loc}(\mathbb{R}_+\times\mathbb{R})} \leq C \tag{5.2}$$

and that u_h converges to a function u in $L^1_{loc}(\mathbb{R}_+\times\mathbb{R})$, as h tends to 0. Moreover, we assume either that, for all bounded intervals I of $\mathbb{R}$,

$$\sum_{i\in K} \Delta x_{i+\frac{1}{2}} |u_{i+1}(t) - u_i(t)| \overset{h\to 0}{\longrightarrow} 0, \quad \text{in } L^1_{loc}(\mathbb{R}_+), \tag{5.3}$$

where K denotes the set of indices such that $x_i \in I$; or a geometrical constraint on the spatial mesh, that is

$$\exists \ \alpha, \beta > 0 \ \text{ so that } \ \alpha\Delta x_{i+1} \leq \Delta x_i \leq \beta \Delta x_{i+1}, \quad \forall i \in \mathbb{Z}. \tag{5.4}$$

Then u is a weak solution to the initial value problem (1.1)-(1.2), i.e.

$$\frac{\partial u}{\partial t} + \frac{\partial A(u)}{\partial x} + z'(x)b(u) = 0, \quad u(0,x) = u^0(x), \quad \text{in } \mathcal{D}'(\mathbb{R}_+\times\mathbb{R}).$$

Proof. The proof is an adaptation of the classical Lax-Wendroff theorem [25] for homogeneous systems of conservation laws.
Let $\varphi \in \mathcal{C}^1_0(\mathbb{R}_+\times\mathbb{R})$ be a test function and set

$$\varphi_i(t) = \varphi(t, x_i), \quad i \in \mathbb{Z}. \tag{5.5}$$

After multiplying equation (2.1) by φ_i, we sum over i and integrate in dt, to obtain

$$\begin{aligned}\int_{\mathbb{R}_+} \sum_{i\in\mathbb{Z}} \Delta x_i \frac{du_i}{dt}\varphi_i dt + \int_{\mathbb{R}_+} \sum_{i\in\mathbb{Z}} \left(A_{i+\frac{1}{2}} - A_{i-\frac{1}{2}}\right)\varphi_i dt \\ + \int_{\mathbb{R}_+} \sum_{i\in\mathbb{Z}} \left(\mathcal{B}^+_{i-\frac{1}{2}} + \mathcal{B}^-_{i+\frac{1}{2}}\right)\varphi_i dt = 0.\end{aligned}$$

An integration by parts in the first term and a summation by parts in the other ones give

$$\begin{aligned}&\int_{\mathbb{R}_+} \sum_{i\in\mathbb{Z}} \Delta x_i u_i \frac{d\varphi_i}{dt} dt + \int_{\mathbb{R}_+} \sum_{i\in\mathbb{Z}} A_{i+\frac{1}{2}}(\varphi_{i+1} - \varphi_i) dt \\ &- \int_{\mathbb{R}_+} \sum_{i\in\mathbb{Z}} \left(\mathcal{B}^+_{i+\frac{1}{2}}\varphi_{i+1} + \mathcal{B}^-_{i+\frac{1}{2}}\varphi_i\right) dt + \sum_{i\in\mathbb{Z}} \Delta x_i u^0_i \varphi_i(0) = 0.\end{aligned} \tag{5.6}$$

We define a.e. in $\mathbb{R}_+ \times \mathbb{R}$ the piecewise constant functions A_h and $\mathcal{B}_h$, associated with the numerical flux and source term by

$$A_h(t,x) = \mathcal{A}(u_i, u_{i+1}), \tag{5.7}$$

$$\mathcal{B}_h(t,x) = \frac{1}{\Delta x_{i+\frac{1}{2}}} \left(\mathcal{B}^+(u_i, u_{i+1}, z_{i+1} - z_i) + \mathcal{B}^-(u_i, u_{i+1}, z_{i+1} - z_i) \right), \tag{5.8}$$

for $t \in \mathbb{R}_+$ and $x \in [x_i, x_{i+1})$, recalling that $\Delta x_{i+\frac{1}{2}} = |x_{i+1} - x_i| = \frac{\Delta x_i}{2} + \frac{\Delta x_{i+1}}{2}$. Next, according to (5.5), we introduce the piecewise constant approximation of the test function

$$\varphi_h(t,x) = \varphi_i(t), \quad t \in \mathbb{R}_+, \ x \in \mathcal{C}_i,$$

which converges to φ (together with $\frac{\partial \varphi_h}{\partial t}$ towards $\frac{\partial \varphi}{\partial t}$) uniformly in $\mathcal{C}_0(\mathbb{R}_+ \times \mathbb{R})$, as h tends to 0. We also consider a continuous piecewise linear function ψ_h such that

$$\begin{aligned} \psi_h(t, x_i) &= \varphi_i(t), \quad i \in \mathbb{Z}, \\ \frac{\partial \psi_h}{\partial x}(t,x) &= \frac{\varphi_{i+1}(t) - \varphi_i(t)}{\Delta x_{i+\frac{1}{2}}}, \quad t \in \mathbb{R}_+, \ x \in [x_i, x_{i+1}), \end{aligned}$$

so that ψ_h and $\frac{\partial \psi_h}{\partial x}$ converge respectively to φ and $\frac{\partial \varphi}{\partial x}$ in $\mathcal{C}_0^1(\mathbb{R}_+ \times \mathbb{R})$, as h tends to 0. As a direct consequence of these definitions, we have

$$\begin{aligned} \varphi_j &= \frac{\varphi_i + \varphi_{i+1}}{2} + O(h), \quad j = i, i+1, \\ \int_{x_i}^{x_{i+1}} \psi_h(t,x)dx &= \frac{\varphi_i + \varphi_{i+1}}{2} \ \Delta x_{i+\frac{1}{2}}. \end{aligned}$$

Taking into account all the relations stated above, we can put the discrete sum (5.6) into the integral form

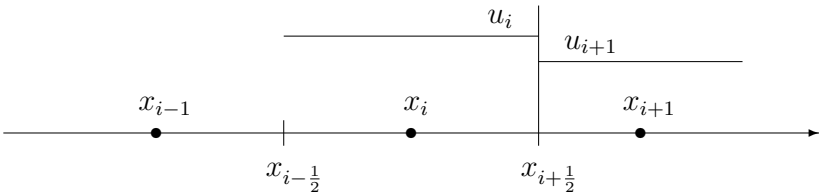
$$\int_{\mathbb{R}_+} \int_{\mathbb{R}} u_h(t,x) \frac{\partial \varphi_h}{\partial t}(t,x)dxdt + \int_{\mathbb{R}_+} \int_{\mathbb{R}} A_h(t,x) \frac{\partial \psi_h}{\partial x}(t,x)dxdt \tag{5.9}$$

$$- \int_{\mathbb{R}_+} \int_{\mathbb{R}} \mathcal{B}_h(t,x) \left(\psi_h(t,x) + O(h) \right) dxdt + \int_{\mathbb{R}} u_h^0(x) \varphi_h(0,x)dx = 0.$$

As h tends to 0, passing to the limit in (5.9), we claim that it turns out

$$\begin{aligned} \int_{\mathbb{R}_+} \int_{\mathbb{R}} \left[u(t,x) \frac{\partial \varphi}{\partial t}(t,x) + A(t,x) \frac{\partial \varphi}{\partial x}(t,x) - B(t,x)\varphi(t,x) \right] dxdt \\ + \int_{\mathbb{R}} u^0(x) \varphi(0,x)dx = 0. \end{aligned} \tag{5.10}$$

The computation is obvious for the first and the last terms, by exploiting the convergence properties of approximations u_h and φ_h.
For the other integrals of (5.9), the process is less straightforward: as we remarked in Section 4, due to the presence of a variable space-step, the interfacial interval $[x_i, x_{i+1})$ could be really different from the mesh cell $\mathcal{C}_i$ (where the conservative unknowns are discretized); so, standard techniques do not work in this case and proving convergence requires the additional hypotheses on the structures that we have imposed.



We need to characterize the functions $A(t,x)$ and $B(t,x)$ in equation (5.10) as the weak limits of $A_h(t,x)$ and $\mathcal{B}_h(t,x)$, for reproducing them in terms of the numerical unknowns.

We first observe that, thanks to (5.2) and the properties (2.2)-(2.3) of the numerical flux, A_h is locally bounded on $\mathbb{R}_+ \times \mathbb{R}$ (uniformly in h). Coming back to discrete notation (5.7), we decompose on the subintervals $[x_i, x_{i+\frac{1}{2}})$ and $[x_{i+\frac{1}{2}}, x_{i+1})$, rearranging terms as follows,

$$\begin{aligned} A_h(t,x) &= \sum_{i\in\mathbb{Z}} A(u_i)\mathbb{1}_{[x_{i-\frac{1}{2}},x_{i+\frac{1}{2}})}(x) \\ &+ \sum_{i\in\mathbb{Z}} \left[\mathcal{A}(u_i,u_{i+1}) - A(u_i)\right]\mathbb{1}_{[x_i,x_{i+\frac{1}{2}})}(x) \\ &+ \sum_{i\in\mathbb{Z}} \left[\mathcal{A}(u_{i-1},u_i) - A(u_i)\right]\mathbb{1}_{[x_{i-\frac{1}{2}},x_i)}(x) \\ &= A(u_h(t,x)) + \mathcal{R}_h^+(t,x) + \mathcal{R}_h^-(t,x). \end{aligned} \tag{5.11}$$

We conclude that $A(u(t,x))$ is the expected value for the limit $A(t,x)$, as h tends to 0, provided that the two remainders in (5.11) vanish.
We only treat with $\mathcal{R}_h^+$, the other one results in the same way. According to definition (5.1), since u_h converges to a function u in $L^1_{loc}(\mathbb{R}_+\times\mathbb{R})$, we also derive

$$\sum_{i\in\mathbb{Z}} |u_i(t) - \bar{u}_i(t)|\mathbb{1}_{\mathcal{C}_i}(x) \underset{L^1_{loc}(\mathbb{R}_+\times\mathbb{R})}{\longrightarrow} 0, \tag{5.12}$$

where the sequence $\{\bar{u}_i(t)\}_{i\in\mathbb{Z}}$ is defined by the cell-averages of u on the discretization mesh. Let I be any bounded interval of $\mathbb{R}$ and we denote by

$C_{\mathcal{A}}$ a Lipschitz constant for $\mathcal{A}$, then the assumptions on the numerical flux lead to estimate (with the notation set out above in the theorem)

$$\begin{aligned}\int_I \left|\mathcal{R}_h^+(t,x)\right| dx &\leq \int_I \sum_{i\in\mathbb{Z}} \left|\mathcal{A}(u_i,u_{i+1}) - A(u_i)\right| \mathbb{1}_{[x_i,x_{i+\frac{1}{2}})}(x)\ dx \\ &\leq C_{\mathcal{A}} \sum_{i\in K} \frac{\Delta x_i}{2} |u_{i+1}-u_i|.\end{aligned} \tag{5.13}$$

Under the hypothesis (5.3), this last term vanishes and the conclusion is done. Otherwise, we can further manipulate the previous bound by introducing appropriate quantities, that is

$$\begin{aligned}\sum_{i\in K} \Delta x_i |u_{i+1}-u_i| \leq &\sum_{i\in K} \Delta x_i |u_{i+1}-\bar{u}_{i+1}| \\ &+ \sum_{i\in K} \Delta x_i |\bar{u}_{i+1}-\bar{u}_i| + \sum_{i\in K} \Delta x_i |\bar{u}_i - u_i|.\end{aligned} \tag{5.14}$$

In particular, for the alternative hypothesis of nondegeneracy (5.4) made on the mesh, we have

$$\sum_{i\in K} \Delta x_i |u_{i+1}-\bar{u}_{i+1}| \leq \beta \sum_{i\in K} \Delta x_{i+1} |u_{i+1}-\bar{u}_{i+1}|,$$

so that property (5.12) ensures that first and last sum in the right-hand side of (5.14) vanish, as h tends to 0. It remains the second term to be studied, involving only the function u, which immediately converges to 0 if $u \in BV$ (or smooth enough); this result also holds for $u \in L^1$, by applying standard regularization arguments (we define $u^\epsilon \in BV$, $u^\epsilon \to u$ in L^1, then we perform an estimation on the cell-averages $\bar{u}_i$ and $\bar{u}_i^\epsilon$ like in (5.14) and we finally conclude by combining convergence properties).

We now pass to the source term, to which a similar procedure applies. Taking into account the definition (5.8), setting $\mathcal{B} = \mathcal{B}^+ + \mathcal{B}^-$, we can write

$$\mathcal{B}_h(t,x) = \sum_{i\in\mathbb{Z}} \frac{\mathcal{B}(u_i,u_{i+1},z_{i+1}-z_i)}{z_{i+1}-z_i} \cdot \frac{z_{i+1}-z_i}{\Delta x_{i+\frac{1}{2}}} \mathbb{1}_{[x_i,x_{i+1})}(x).$$

Notice that the hypothesis $z \in W^{1,1}$ guarantees that discrete differences converge to the derivative $z' \in L^1$; together with condition (4.1), this leads to justify the assertion that $\mathcal{B}_h(t,x)$ is $L^1 - weak$ bounded in $\mathbb{R}_+ \times \mathbb{R}$.

We proceed as in (5.11), by performing the following decomposition

$$\begin{aligned}\mathcal{B}_h(t,x) = &\sum_{i\in\mathbb{Z}} \frac{\mathcal{B}(u_i,u_i,z_{i+1}-z_i)}{z_{i+1}-z_i}\cdot\frac{z_{i+1}-z_i}{\Delta x_{i+\frac12}}\mathbb{1}_{[x_i,x_{i+\frac12})}(x)+\mathcal{E}_h^+(t,x)\\ &+\sum_{i\in\mathbb{Z}} \frac{\mathcal{B}(u_i,u_i,z_i-z_{i-1})}{z_i-z_{i-1}}\cdot\frac{z_i-z_{i-1}}{\Delta x_{i-\frac12}}\mathbb{1}_{[x_{i-\frac12},x_i)}(x)+\mathcal{E}_h^-(t,x).\end{aligned} \tag{5.15}$$

The sum of the two principal terms of (5.15) converges to $z'(x)b(u)$, in view of the arguments just mentioned and the strong convergence of u_h towards u, by means of Lebesgue's theorem.
For the remainders, we give details in the case of $\mathcal{E}_h^+(t,x)$, for instance. We then have

$$\begin{aligned}\mathcal{E}_h^+(t,x) = \sum_{i\in\mathbb{Z}} \frac{1}{\Delta x_{i+\frac12}}\,&(\mathcal{B}(u_i,u_{i+1},z_{i+1}-z_i)\\ &-\mathcal{B}(u_i,u_i,z_{i+1}-z_i))\,\mathbb{1}_{[x_i,x_{i+\frac12})}(x).\end{aligned} \tag{5.16}$$

The analogue of estimation (5.13) for (5.16) becomes

$$\begin{aligned}\int_I |\mathcal{E}_h^+(t,x)|dx &\le \sum_{i\in K} |\mathcal{B}(u_i,u_{i+1},z_{i+1}-z_i)-\mathcal{B}(u_i,u_i,z_{i+1}-z_i)|\\ &= \sum_{i\in K}\left|\int_{u_i}^{u_{i+1}} \frac{\partial\mathcal{B}}{\partial v}(u_i,v,z_{i+1}-z_i)dv\right|\\ &= \sum_{i\in K}\left|\int_{u_i}^{u_{i+1}} \left(\frac{\partial\mathcal{B}}{\partial v}(u_i,v,z_{i+1}-z_i)-\frac{\partial\mathcal{B}}{\partial v}(u_i,v,0)\right)dv\right|\\ &\le C_{\mathcal{B}}\sum_{i\in K}|u_{i+1}-u_i||z_{i+1}-z_i|,\end{aligned}$$

where we exploited property (2.5) and the regularity assumed in (2.4). Then we can regularize the function z by introducing $z^\epsilon\in W^{1,\infty}$, $z^\epsilon\to z$ in $W^{1,1}$, for which we write

$$\begin{aligned}\sum_{i\in K}|u_{i+1}-u_i||z_{i+1}-z_i| &\le \sum_{i\in K}|u_{i+1}-u_i||z_{i+1}-z^\epsilon_{i+1}|\\ &+\sum_{i\in K}|u_{i+1}-u_i||z^\epsilon_i-z_i|+\sum_{i\in K}|u_{i+1}-u_i||z^\epsilon_{i+1}-z^\epsilon_i|\\ &\le 4C\sum_{i\in K}|z^\epsilon_i-z_i|+C_\epsilon\sum_{i\in K}\Delta x_{i+\frac12}|u_{i+1}-u_i|,\end{aligned} \tag{5.17}$$

with C defined as in (5.2) and C_ϵ only depending on the regularization. The same procedure we have considered before allow us to conclude that the upper bound (5.17) vanishes, as h tends to 0.

Thanks to the previous computations, we have proved that the limit function u satisfies the weak formulation of the Cauchy problem (1.1)-(1.2),

$$\int_{\mathbb{R}_+}\int_{\mathbb{R}}\left[u(t,x)\frac{\partial\varphi}{\partial t}(t,x)+A(u(t,x))\frac{\partial\varphi}{\partial x}(t,x)-z'(x)b(u(t,x))\varphi(t,x)\right]dx\;dt + \int_{\mathbb{R}}u^0(x)\varphi(0,x)dx=0,$$

so that the proof of the theorem is completed. □

Remark 5.2. *For the particular case of uniform spatial mesh, i.e.* $\Delta x_i = h$, $\forall i \in \mathbb{Z}$, *sets of weaker hypotheses can be considered and the proof is simplified, due to the fact that the consistency error vanishes for the source term (refer to [9] and [26]). The version presented above is compatible with those general statements (see also [10]) and actually extends the classical Lax-Wendroff theorem to scalar conservation laws with a source term.*

6 Conclusion

We have proposed a consistency condition for hyperbolic conservation laws with a source term $z'(x)b(u)$, discretized according to the Upwind Interface Source method. We have proved that numerical schemes which preserve the steady state solutions are consistent in that sense. Moreover, a strong limit of discrete solutions satisfies the continuous equation, as the mesh is refined. Theorem 5.1 thus constitutes a fundamental result in the theoretical analysis of the numerical method.

Nevertheless, the conditions established in the previous sections do not guarantee that discrete approximations given by (2.1)-(2.5) do converge and we do not know whether the limit weak solution is the unique physical solution of the Cauchy problem (1.1)-(1.4). For that we need to precise some criteria of stability for the approximate solution and we have to consider further assumptions on the discrete functions to derive suitable error estimates.

In order to ensure that a weak solution obtained as limit of (5.1) satisfies the family of entropy inequalities (1.5), it suffices to show that a discrete entropy inequality,

$$\Delta x_i\frac{d}{dt}S(u_i)+\left(\eta^S_{i+\frac{1}{2}}-\eta^S_{i-\frac{1}{2}}\right)+\mathcal{B}^{S,+}_{i-\frac{1}{2}}+\mathcal{B}^{S,-}_{i+\frac{1}{2}}\leq 0, \tag{6.1}$$

holds for the numerical scheme, with the usual formalism

$$\eta^S_{i+\frac{1}{2}}=\eta^S(u_i,u_{i+1}),\quad \mathcal{B}^{S,\pm}_{i+\frac{1}{2}}=\mathcal{B}^{S,\pm}(u_i,u_{i+1},z_{i+1}-z_i), \tag{6.2}$$

where η^S and $\mathcal{B}^{S,\pm}$ are some numerical entropy flux function and source term, which must be consistent with $\eta(u)$ and $S'(u)B(x,u)$ in the same way that we required $\mathcal{A}$ and $\mathcal{B}^{\pm}$ to be consistent with $A(u)$ and $B(x,u)$ in (1.1). Therefore, mimicking the proof of the Lax-Wendroff theorem, we can prove that the weak form of the entropy inequality (1.5) is also verified.
Note that the possibility to write formulas (6.1)-(6.2) relies only on the characterization of the numerical flux (namely, at least the condition of *E*-scheme has to be assumed); for the source term, the definition is automatically made, since we have

$$\mathcal{B}^{S,+}_{i-\frac{1}{2}} = S'(u_i)\mathcal{B}^{+}_{i-\frac{1}{2}}, \qquad \mathcal{B}^{S,-}_{i+\frac{1}{2}} = S'(u_i)\mathcal{B}^{-}_{i+\frac{1}{2}}.$$

The question to determine the order of accuracy of the numerical scheme by means of error estimates is more delicate than for the homogeneous system, due to the presence of source terms. The approach formulated by Kruzkov is used henceforth in the literature (see [11],[12] and [19],[20] for instance), providing a method to convert any discrete entropy inequality into an error estimate. The general procedure consists in the following formulation for the approximate solution, in $\mathcal{D}'(\mathbb{R}_+ \times \mathbb{R})$,

$$\frac{\partial S(u_h)}{\partial t} + \frac{\partial}{\partial x}\eta^S(u_h) + S'(u_h)z'(x)b(u_h) \leq \frac{\partial}{\partial x}\mathrm{Err}_2(t,x) + \mathrm{Err}_1(t,x),$$

where we set

$$\frac{\partial}{\partial x}\mathrm{Err}_1(t,x) = \frac{\partial}{\partial x}\eta^S(u_h) - \sum_{i\in\mathbb{Z}} \frac{1}{\Delta x_i}\left(\eta^S_{i+\frac{1}{2}} - \eta^S_{i-\frac{1}{2}}\right)\mathbb{1}_{\mathcal{C}_i}(x),$$

$$\mathrm{Err}_2(t,x) = S'(u_h)z'(x)b(u_h) - \sum_{i\in\mathbb{Z}} \frac{1}{\Delta x_i}\left(\mathcal{B}^{S,+}_{i-\frac{1}{2}} + \mathcal{B}^{S,-}_{i+\frac{1}{2}}\right)\mathbb{1}_{\mathcal{C}_i}(x).$$

Then the results of [3] apply to this particular problem, to deduce stability bounds and conclude the convergence properties we have assumed in the Theorem 5.1. We remark that regularity hypotheses like (5.3) and (5.4) are necessary to control the variations of the numerical solution in comparison with the space-step (refer to [30] for the homogeneous problem).
In order to avoid BV estimates, which are not available in case of insufficiently smooth source terms and for multidimensional systems on an unstructured mesh, arguments based on the measure-valued method and the so-called weak BV estimates are developed (see [7] and [4],[5], for instance) or the kinetic approach presented in [2].

References

[1] Bermudez A., Vasquez M.E., Upwind methods for hyperbolic conservation laws with source terms, *Comput. & Fluids*, **23** (1994), no. 8, 1049-1071

[2] Botchorishvili R., Perthame B., Vasseur A., Equilibrium Schemes for Scalar Conservation Laws with Stiff Sources, *Math. Comp.*, to appear

[3] Bouchut F., Perthame B., Kružkov's estimates for scalar conservation laws revisited, *Trans. Amer. Math. Soc.*, **350** (1998), no. 7, 2847-2870

[4] Cockburn B., Coquel F., LeFloch P., An error estimate for finite volume methods for multidimensional conservation laws, *Math. of Comp.*, **63** (1994), 77-103

[5] Cockburn B., Coquel F., LeFloch P., Convergence of the finite volume method for multidimensional conservation laws, *SIAM J. Numer. Anal.*, **32** (1995), no. 3, 687-705

[6] Eymard R., Gallouët T., Herbin R., *Finite Volume Methods*, Handbook of numerical analysis, vol. VIII, P.G.Ciarlet and J.L.Lions editors, Amsterdam, North-Holland, 2000

[7] Eymard R., Gallouët T., Ghilani M., Herbin R., Error estimates for the approximate solutions of a nonlinear hyperbolic equation given by some finite volume schemes, *I.M.A. Journal of Numer. Anal.*, **18** (1998), 563-594

[8] Gallouët T., Hérard J.M., Seguin N., Some approximate Godunov schemes to compute shallow-water equations with topography, *AIAA-2001* (2000)

[9] Godlewski E., Raviart P.A., *Hyperbolic systems of conservation laws*, Mathématiques & Applications, n. 3/4, Paris, Ellipses, 1991

[10] Godlewski E., Raviart P.A., *Numerical approximation of hyperbolic systems of conservation laws*, Applied Mathematical Sciences **118**, New York, Springer-Verlag, 1996

[11] Gosse L., LeRoux A.Y., A well-balanced scheme designed for inhomogeneous scalar conservation laws, *C.R. Acad. Sci. Paris Sér.I Math.*, **323** (1996), no. 5, 543-546

[12] Gosse L., A priori error estimate for a well-balanced scheme designed for inhomogeneous scalar conservation laws, *C.R. Acad. Sci. Paris Sér.I Math.*, **327** (1998), no. 5, 467-472

[13] Gosse L., A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms, *Comput. Math. Appl.*, **39** (2000), no. 9-10, 135-159

[14] Gosse L., A well-balanced scheme using non-conservative products designed for hyperbolic systems of conservation laws with source terms, *Math. Models Methods Appl. Sci.*, **11** (2001), no. 2, 339-365

[15] Greenberg J.M., LeRoux A.Y., A well balanced scheme for the numerical processing of source terms in hyperbolic equations, *SIAM J. Numer. Anal.*, **33** (1996), 1-16

[16] Greenberg J.M., LeRoux A.Y., Baraille R., Noussair A., Analysis and approximation of conservation laws with source term, *SIAM J. Numer. Anal.*, **34** (1997), no. 5, 1980-2007

[17] Jin S., A steady-state capturing method for hyperbolic systems with geometrical source terms, *M2AN Math. Model. Numer. Anal.*, **35** (2001), no. 4, 631-645

[18] Karni S., Source linearization for conservation laws, submitted to *M2NA Math. Model. Numer. Anal.*

[19] Katsaounis T., Makridakis C., Finite volume relaxation schemes for multidimensional conservation laws, *Math. Comp.*, **70** (2001), n. 234, 533-553

[20] Katsoulakis M.A., Kossioris G., Makridakis C., Convergence and error estimates of relaxation schemes for multidimensional conservation laws, *Comm. Partial Differential Equations*, **24**(1999), n. 3-4, 395-424

[21] Kružkov S.N., First order quasilinear equations in several independent space variables, *Math. USSR Sb.*, **10** (1970), 217-243

[22] Kurganov A., Central-upwind schemes for balance laws. Application to the Broadwell model, *Proceedings of the Third International Symposium on Finite Volumes for Complex Applications* (2002), to appear

[23] Kurganov A., Levy D., Central-Upwind Schemes for the Saint-Venant system, *M2AN Math. Model. Numer. Anal.* (2001), to appear

[24] Lax P.D., Shock waves and entropy, in *Contributions to nonlinear functional analysis*, E.H.Zarantonello editor, New York, Academic Press, 1971, pp. 603-634

[25] Lax P.D., Wendroff B., Systems of conservations laws, *Comm. Pure Appl. Math.*, **13** (1960), 217-237

[26] LeVêque R.J., *Numerical Methods for Conservation Laws*, Lectures in Mathematics, ETH Zurich, Birkhauser, 1992

[27] LeVêque R.J., Balancing source terms and flux gradients in high-resolution Godunov methods: the quasi-steady wave-propagation algorithm, *J. Comput. Phys.*, **146** (1998), no. 1, 346-365

[28] Perthame B., Simeoni C., A kinetic scheme for the Saint-Venant system with a source term, *Calcolo*, **38** (2001), no. 4, 201-231

[29] Roe P.L., Upwind differencing schemes for hyperbolic conservation laws with source terms, in *Nonlinear Hyperbolic Problems*, C.Carasso, P.A.Raviart and D.Serre editors, Lecture Notes in Math., vol. 1270, Berlin, Springer-Verlag, 1987, pp. 41-51

[30] Sanders R., On Convergence of Monotone Finite Difference Schemes with Variable Spatial Differencing, *Math. Comp.*, **40** (1983), 91-106

[31] Vasseur A., Well-posedness of scalar conservation laws with singular sources, in preparation

Chapitre 2

Estimations d'erreur d'ordre un et deux pour la méthode "Upwind Interface Source"

(preprint)

First and second order error estimates for the Upwind Interface Source method

T. Katsaounis, C. Simeoni

Département de Mathématiques et Applications
École Normale Supérieure
45, rue d'Ulm - 75230 Paris Cedex 05 - France
e-mails: katsaoun@dma.ens.fr, simeoni@dma.ens.fr

Abstract

The Upwind Interface Source method for hyperbolic conservation laws presented in [30] is essentially first order accurate. Under appropriate hypotheses of consistency on the source discretization, we prove L^p-error estimates, for $1 \leq p < +\infty$, in the case of a uniform spatial mesh, for which an optimal result can be obtained. We thus conclude that the same convergence rate holds as in the corresponding homogeneous problem (refer to [8]). To improve the numerical accuracy, we develop two different approaches of treating the source term and we discuss the question to derive second order error estimates. Numerical evidence shows that those techniques produce high resolution schemes compatible with the Upwind Interface Source method.

Key-words: scalar conservation laws, source terms, upwind interfacial methods, consistency, error estimates.

1 Introduction

We consider the initial values problem for a transport equation with nonlinear source term, in one space dimension,

$$\partial_t u + \partial_x u = B(x,u), \quad t \in \mathbb{R}_+,\ x \in \mathbb{R}, \tag{1.1}$$

$$u(0,x) = u_0(x) \in L^p(\mathbb{R}) \cap L^\infty(\mathbb{R}),\ 1 \leq p < +\infty, \tag{1.2}$$

with $u(t,x) \in \mathbb{R}$ and the analytical source operator is given by

$$B(x,u) = z'(x)\, b(u), \quad z' \in L^p(\mathbb{R}),\ b \in C^1(\mathbb{R}). \tag{1.3}$$

The system (1.1)-(1.3) corresponds to the simplest model of scalar conservation law with a geometrical source term, extensively treated in [30].
The entropy inequalities associated to (1.1) are described by the equation

$$\partial_t S(u) + \partial_x S(u) + S'(u)B(x,u) \leq 0, \tag{1.4}$$

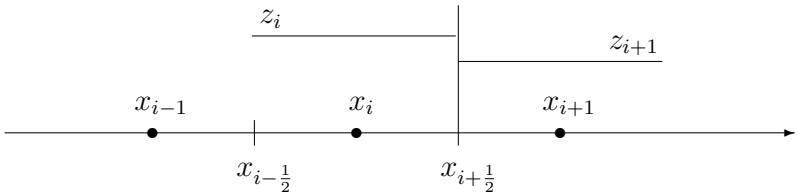
for any convex entropy function S (see [17] and [18]). Under stronger assumptions on the source term, Kružkov [17] proved existence and uniqueness of the entropy solution to the problem (1.1)-(1.2), in the functional space $L^\infty([0,T);L^p(\mathbb{R}))$, for all $T \in \mathbb{R}_+$. Another approach, based on convergence analysis for special approximations, is presented in [1]. In the case of singular source terms (namely, z discontinuous), a uniqueness result has recently been proved by Vasseur [35].

1.1 Formalism of the Upwind Interface Source method

We set up a uniform mesh on $\mathbb{R}$, whose vertices are $x_i, i \in \mathbb{Z}$ and with characteristic space-step h. We denote by $C_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$ the control volume (cell) centered on x_i, where $x_{i+\frac{1}{2}} = \frac{x_i+x_{i+1}}{2}$ are the cell interfaces, so that $h = length(C_i)$. Then we construct a piecewise constant approximation of the function z on the mesh, for example

$$z^h(x) = \sum_{i\in\mathbb{Z}} z_i \mathbb{1}_{C_i}(x), \qquad z_i = \frac{1}{h}\int_{C_i} z(x)\,dx, \tag{1.5}$$

where $\mathbb{1}_{C_i}$ is the characteristic function of the cell C_i.



We also introduce a piecewise constant approximation of the analytical solution to the problem (1.1)-(1.2), defined by

$$u^h(t,x) = \sum_{i\in\mathbb{Z}} u_i(t)\mathbb{1}_{C_i}(x), \qquad u_i(t) = \frac{1}{h}\int_{C_i} u(t,x)\,dx. \tag{1.6}$$

In the above framework, the numerical solution obtained from a finite volume scheme applied to (1.1)-(1.2) is a function $v^h(t,x)$, whose cell-averages

$$v_i(t) = \frac{1}{h}\int_{C_i} v^h(t,x)\,dx, \quad i \in \mathbb{Z}, \tag{1.7}$$

are interpreted as approximations of the cell-averages of the analytical solution, $v_i(t) \approx u_i(t)$, $i \in \mathbb{Z}$. The general scheme for (1.1) reads

$$\partial_t v^h + \partial_x v^h = B^N(x, v^h), \tag{1.8}$$

with initial data corresponding to the approximate initial condition

$$v_0^h(x) = \sum_{i\in\mathbb{Z}} u_{0i} \mathbb{1}_{C_i}(x), \qquad u_{0i} = \frac{1}{h}\int_{C_i} u_0(x)\,dx. \tag{1.9}$$

According to the Upwind Interface Source method in [30], appropriate discretizations of the source term in (1.8) are given by

$$B^N(x, v^h) = \sum_{i\in\mathbb{Z}} \frac{1}{h}\left[\mathcal{B}^+(v_{i-1}, v_i, \Delta z_{i-\frac{1}{2}}) + \mathcal{B}^-(v_i, v_{i+1}, \Delta z_{i+\frac{1}{2}})\right] \mathbb{1}_{C_i}, \tag{1.10}$$

where we set $\Delta z_{i+\frac{1}{2}} = z_{i+1} - z_i$ (we dropped the time and space dependence in the formula, for simplicity). We assume the following consistency properties for the numerical source operator (1.10), in respect of (1.3), which are fundamental to the convergence analysis,

$$\mathcal{B}^\pm \in C^2, \quad \mathcal{B}^\pm(u, v, 0) = 0, \quad \frac{\partial \mathcal{B}^\pm}{\partial u}(u, v, 0) = \frac{\partial \mathcal{B}^\pm}{\partial v}(u, v, 0) = 0, \tag{1.11a}$$

$$\lim_{\zeta\to 0} \frac{\mathcal{B}^+(u, u, \zeta) + \mathcal{B}^-(u, u, \zeta)}{\zeta} = b(u). \tag{1.11b}$$

The last limit holds uniformly in u, as specified by the further assumption

$$\left|\frac{\mathcal{B}^+(u, u, \zeta) + \mathcal{B}^-(u, u, \zeta)}{\zeta} - b(u)\right| \le K_B\,\zeta, \tag{1.12}$$

where K_B is a fixed constant (independent of u). Moreover, we denote by L_b and L_B any Lipschitz constant associated respectively to the continuous or discrete source operator.

1.2 What is a second order scheme for the Upwind Interface Source method?

In order to obtain second order extensions of the discrete solver (1.8)-(1.9), we apply a slope limiter method to the numerical functions: the basic idea is to replace the piecewise constant reconstruction on the mesh of the approximate solution by more accurate reconstructions, namely piecewise linear (refer to [10] and [22] for a survey of high resolution methods).

We associate to the numerical solution (1.7) some coefficients, defined as second order interpolation of the discrete unknowns,

$$\bar{v}_i(t,x) = v_i(t) + (x - x_i)v_i', \quad i \in \mathbb{Z},\ x \in C_i, \tag{1.13}$$

where v_i' indicates a generic numerical derivative (computed by means of an appropriate *limiter*, as it will be discussed more precisely later on).
From (1.6), analogous definitions are introduced for the analytical solution,

$$\bar{u}_i(t,x) = u_i(t) + (x - x_i)u_i', \quad i \in \mathbb{Z},\ x \in C_i. \tag{1.14}$$

The function z can also be represented in terms of piecewise linear approximations on the spatial mesh, departing from (1.5), with coefficients

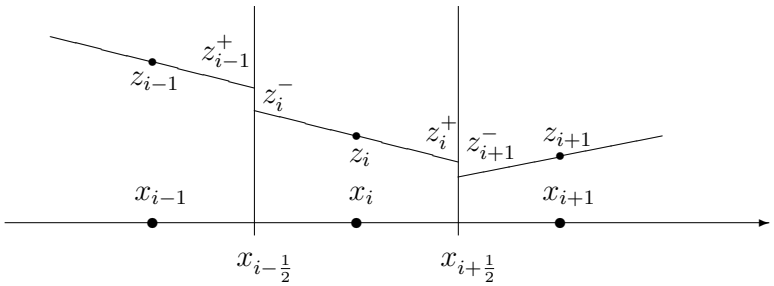
$$\bar{z}_i(x) = z_i + (x - x_i)z_i', \quad x \in C_i.$$

At the cell interfaces, the values of the numerical functions are given by

$$v_i^- = \bar{v}_i(x_{i-\frac{1}{2}}) = v_i - \frac{h}{2}v_i', \quad v_i^+ = \bar{v}_i(x_{i+\frac{1}{2}}) = v_i + \frac{h}{2}v_i', \tag{1.15a}$$

$$z_i^- = \bar{z}_i(x_{i-\frac{1}{2}}) = z_i - \frac{h}{2}z_i', \quad z_i^+ = \bar{z}_i(x_{i+\frac{1}{2}}) = z_i + \frac{h}{2}z_i', \tag{1.15b}$$

as represented in the figure below, so that $\Delta z_{i+\frac{1}{2}} = z_{i+1}^- - z_i^+$ in this case (we drop the time and space dependence when no mistake is possible).



Therefore, it is natural to perform a discretization of the source term (1.3) by using the interfacial values (1.15a)-(1.15b), as follows,

$$\begin{aligned} B^N(x, v^h) = &\sum_{i\in\mathbb{Z}} \frac{1}{h}\left[\mathcal{B}^+(v_{i-1}^+, v_i^-, \Delta z_{i-\frac{1}{2}}) + \mathcal{B}^-(v_i^+, v_{i+1}^-, \Delta z_{i+\frac{1}{2}})\right] \mathbb{1}_{C_i} \\ &+ \sum_{i\in\mathbb{Z}} z_i' \, b(v_i)\, \mathbb{1}_{C_i}, \end{aligned} \tag{1.16}$$

with an additional term in comparison to the discrete source operator (1.10), which depends on the cell-averages and is necessary to achieve second order estimates (see Section 4 for details).

An alternative approach to formulating second order extensions of the Upwind Interface Source method is based on improving the consistency properties of the numerical source operator.
We consider a piecewise constant approximation (1.5) of the function z and piecewise linear reconstructions (1.13) of the numerical solution on the mesh, to define the upwind interfacial discretization

$$B^N(x, v^h) = \sum_{i\in\mathbb{Z}} \frac{1}{h} \left[\mathcal{B}^+(v^+_{i-1}, v^-_i, \Delta z_{i-\frac{1}{2}}) + \mathcal{B}^-(v^+_i, v^-_{i+1}, \Delta z_{i+\frac{1}{2}})\right] \mathbb{1}_{C_i}, \quad (1.17)$$

where the numerical functions are computed on the interfacial values (1.15a) and $\Delta z_{i+\frac{1}{2}} = z_{i+1} - z_i$. To obtain second order accuracy, we need to assume that (1.11a) holds and the second order definition of consistency

$$\left|\frac{\mathcal{B}^+(u, u, \zeta) + \mathcal{B}^-(u, u, \zeta)}{\zeta} - b(u)\right| \leq K_B\, \zeta^2. \quad (1.18)$$

This is suggested by the particular form of the source term (1.3), given by the product of functions which exhibit different orders of derivative.

Remark 1.1. *In effect, the two discretizations* (1.16) *and* (1.17) *are strictly related, as formally verified by means of standard asymptotic expansions on the numerical functions and simple algebraic calculations with the discrete differences of values* (1.5) *or* (1.15b)*. We also note that many of the second order schemes proposed in the literature do not include the additional term in* (1.16)*, for the sake of simplicity (see [25], [2] and [29], for instance), but that is probably recovered implicitly in the formulation.*

1.3 Convergence and error estimates

To deal with the question of deriving error estimates for the approximation (1.8) to the equation (1.1), we introduce the error function

$$e(t, x) = u(t, x) - v^h(t, x), \quad (1.19)$$

which satisfies

$$\begin{aligned} \partial_t e + \partial_x e &= B(x, u) - B^N(x, v^h) \\ &= \big(B(x, u) - B^N(x, u^h)\big) + \big(B^N(x, u^h) - B^N(x, v^h)\big) \\ &:= \mathcal{C}(u; u^h) + \mathcal{S}(u^h; v^h). \end{aligned} \quad (1.20)$$

From (1.6) and (1.7), we obtain the usual expression for the cell-averages,

$$e_i(t) = \frac{1}{h}\int_{C_i} e(t, x)\, dx = u_i(t) - v_i(t), \quad i \in \mathbb{Z}. \quad (1.21)$$

The operators $\mathcal{C}(u;u^h)$ and $\mathcal{S}(u^h;v^h)$ in the formula (1.20) indicate the consistency and stability error term respectively.

The following result constitute the main stage of the convergence analysis for the Upwind Interface Source method.

Theorem 1.2. *We assume $z \in W^{2,p}$, $1 \leq p < +\infty$, and we consider the numerical source operator* (1.10) *in* (1.20)*. Then, for all $t \in \mathbb{R}_+$, the error function* (1.19) *verifies the first order estimate*

$$\|e(t)\|_{L^p} \leq C(t) \left(\|e_0\|_{L^p} + h\|z\|_{W^{2,p}} + h \int_0^t exp\{-Cs\} \|u(s)\|_{W^{1,p}} \, ds \right), \tag{1.22}$$

where $C(t)$ is a constant independent of h.

The convergence properties of second order schemes are notably affected by the technique used to construct piecewise linear approximations of the numerical functions, namely the choice of the slope *limiter*, as pointed out by several authors (see [27], [36] and [15]). Without appropriate hypotheses on the coefficients of such approximations, the proof of the consistency estimate given in Section 3.2 fails and numerical evidence shows that the discretization (1.16) loses second order accuracy (refer to Section 4 for details).

The following results extend the one which is established in Theorem 1.2 to the discretization (1.16) and (1.17).

Theorem 1.3. *We assume $z \in W^{3,p}$, $1 \leq p < +\infty$, and we consider the numerical source operator* (1.16)*, with discrete derivatives computed in the restricted class of slope limiters introduced in Section 3. Then, for all $t \in \mathbb{R}_+$, the error function* (1.19) *verifies the second order estimate*

$$\|e(t)\|_{L^p} \leq C(t) \left(\|e_0\|_{L^p} + h^2\|z\|_{W^{3,p}} + h^2 \int_0^t exp\{-Cs\} \|u(s)\|_{W^{2,p}} \, ds \right), \tag{1.23}$$

where $C(t)$ is a constant independent of h.

Theorem 1.4. *We assume $z \in W^{3,p}$, $1 \leq p < +\infty$, and we consider the numerical source operator* (1.17)*, with the consistency property* (1.18)*. Then, for all $t \in \mathbb{R}_+$, the error function* (1.19) *verifies the second order estimate*

$$\|e(t)\|_{L^p} \leq C(t) \left(\|e_0\|_{L^p} + h^2\|z\|_{W^{3,p}} + h^2 \int_0^t exp\{-Cs\} \|u(s)\|_{W^{2,p}} \, ds \right), \tag{1.24}$$

where $C(t)$ is a constant independent of h.

Because of the definition (1.6) and (1.9), we have $v_0^h = u_0^h$ and then we deduce from (1.19) that $e_0(x) = u_0(x) - u_0^h(x)$, $x \in \mathbb{R}$. Besides, the following statements are classical and not difficult to prove (see [5], for instance),

$$\begin{aligned} &\|u_0^h\|_{L^p} \leq \|u_0\|_{L^p}, \quad 1 \leq p < +\infty, \\ &\|e_0\|_{L^p} \leq Ch \quad \text{if} \quad u_0 \in W^{1,p}, \\ &\|e_0\|_{L^p} \leq Ch^2 \quad \text{if} \quad u_0 \in W^{2,p}. \end{aligned}$$

The convergence of initial data in (1.22), (1.23) and (1.24), as the mesh size tends to zero, is thus guaranteed by the first and second order convergence of piecewise constant approximations.
The detailed proofs of Theorem 1.2, Theorem 1.3 and Theorem 1.4, with the corresponding intermediate stages, are presented in Section 2 and Section 3.

Remark 1.5. *The same approach as described above applies to nonlinear scalar conservation laws with a source term, also to define numerical fluxes in semi-discrete methods (refer to [30] for specific notations). Therefore, the arguments developed in this paper might extend to the general case, to derive complete error estimates for the Upwind Interface Source method.*

2 Error estimates for first order schemes

Before giving details about the estimates, we introduce some relations on the discrete differences of numerical functions, we will frequently use later on the proofs.

We consider a generic function $w \in C^1$, whose cell-averages on the spatial mesh are denoted by $w_i = \frac{1}{h}\int_{C_i} w(x)\,dx$, $i \in \mathbb{Z}$. By performing appropriate expansions, we obtain

$$w_{i+1} - w_i = \int_{C_i} w'(\xi(x))\,dx \tag{2.1a}$$

$$= hw'(x_i) + \int_{C_i} w''(\eta(x))(x - x_i)\,dx, \tag{2.1b}$$

$$w_{i+1} - 2w_i + w_{i-1} = h\int_{C_i} w''(\vartheta(x))\,dx, \tag{2.1c}$$

for some $\xi(x), \eta(x), \vartheta(x) \in C_i$. We also recall the classical *Taylor's formula*,

$$w(x) = \sum_{k=0}^{n} \frac{1}{k!} w^k(x_i)(x - x_i)^k + \frac{1}{n!}\int_{x_i}^{x} (x - s)^n w^{n+1}(s)\,ds, \tag{2.2}$$

in the particular form with an integral expression for the remainder.

2.1 Stability estimate

We begin by dealing with the stability error term $\mathcal{S}(u^h; v^h)$ in (1.20), to test the stability of the numerical source operator.

Lemma 2.1. *For the assumptions of Theorem 1.2, together with* (1.11a)*, there exists a constant $C := C(L_B, \|z'\|_{L^\infty})$, independent of h, such that*

$$\left| \int_{\mathbb{R}} \mathcal{S}(u^h; v^h)\, |e|^{p-1} sgn(e)\, dx \right| \leq C \|e\|_{L^p}^p. \tag{2.3}$$

Proof. From (1.20), we have

$$\int_{\mathbb{R}} \mathcal{S}(u^h; v^h)\, |e|^{p-1} sgn(e)\, dx = \int_{\mathbb{R}} \left[B^N(x, u^h) - B^N(x, v^h) \right] |e|^{p-1} sgn(e)\, dx.$$

Then, according to (1.10) and (1.6), we deduce

$$\begin{aligned}
&\int_{\mathbb{R}} \mathcal{S}(u^h; v^h)\, |e|^{p-1} sgn(e)\, dx \\
&= \int_{\mathbb{R}} \left\{ \sum_{i \in \mathbb{Z}} \frac{1}{h} \left[\mathcal{B}^+(u_{i-1}, u_i, \Delta z_{i-\frac{1}{2}}) + \mathcal{B}^-(u_i, u_{i+1}, \Delta z_{i+\frac{1}{2}}) \right] \mathbb{1}_{C_i} \right. \\
&\qquad \left. - \sum_{i \in \mathbb{Z}} \frac{1}{h} \left[\mathcal{B}^+(v_{i-1}, v_i, \Delta z_{i-\frac{1}{2}}) + \mathcal{B}^-(v_i, v_{i+1}, \Delta z_{i+\frac{1}{2}}) \right] \mathbb{1}_{C_i} \right\} |e|^{p-1} sgn(e)\, dx \\
&\qquad = \sum_{i \in \mathbb{Z}} \left[\mathcal{B}^+(u_{i-1}, u_i, \Delta z_{i-\frac{1}{2}}) - \mathcal{B}^+(v_{i-1}, v_i, \Delta z_{i-\frac{1}{2}}) \right] e_i^{p-1} \\
&\qquad + \sum_{i \in \mathbb{Z}} \left[\mathcal{B}^-(u_i, u_{i+1}, \Delta z_{i+\frac{1}{2}}) - \mathcal{B}^-(v_i, v_{i+1}, \Delta z_{i+\frac{1}{2}}) \right] e_i^{p-1} \\
&\qquad = \sum_{i \in \mathbb{Z}} \left[\mathcal{B}^+(u_i, u_{i+1}, \Delta z_{i+\frac{1}{2}}) - \mathcal{B}^+(v_i, v_{i+1}, \Delta z_{i+\frac{1}{2}}) \right] e_{i+1}^{p-1} \\
&\qquad + \sum_{i \in \mathbb{Z}} \left[\mathcal{B}^-(u_i, u_{i+1}, \Delta z_{i+\frac{1}{2}}) - \mathcal{B}^-(v_i, v_{i+1}, \Delta z_{i+\frac{1}{2}}) \right] e_i^{p-1} := S_1 + S_2,
\end{aligned}$$

where we set $e_i^{p-1} = \frac{1}{h} \int_{C_i} |e|^{p-1} sgn(e)\, dx$ (as usually, we dropped the time and space dependence in the above formulas for simplicity).
We estimate the terms S_1 and S_2 separately. For S_1, we have

$$\begin{aligned}
S_1 = &\sum_{i \in \mathbb{Z}} \left[\mathcal{B}^+(u_i, u_{i+1}, \Delta z_{i+\frac{1}{2}}) - \mathcal{B}^+(v_i, u_{i+1}, \Delta z_{i+\frac{1}{2}}) \right] e_{i+1}^{p-1} \\
&+ \sum_{i \in \mathbb{Z}} \left[\mathcal{B}^+(v_i, u_{i+1}, \Delta z_{i+\frac{1}{2}}) - \mathcal{B}^+(v_i, v_{i+1}, \Delta z_{i+\frac{1}{2}}) \right] e_{i+1}^{p-1}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i\in\mathbb{Z}} \left(\int_{v_i}^{u_i} \frac{\partial \mathcal{B}^+}{\partial u}(u, u_{i+1}, \Delta z_{i+\frac{1}{2}})\, du \right) e_{i+1}^{p-1} \\
&+ \sum_{i\in\mathbb{Z}} \left(\int_{v_{i+1}}^{u_{i+1}} \frac{\partial \mathcal{B}^+}{\partial v}(v_i, v, \Delta z_{i+\frac{1}{2}})\, dv \right) e_{i+1}^{p-1},
\end{aligned}
\tag{2.4}
$$

so that, in view of (1.11a), we get

$$
\begin{aligned}
S_1 &= \sum_{i\in\mathbb{Z}} \left(\int_{v_i}^{u_i} \left[\frac{\partial \mathcal{B}^+}{\partial u}(u, u_{i+1}, \Delta z_{i+\frac{1}{2}}) - \frac{\partial \mathcal{B}^+}{\partial u}(u, u_{i+1}, 0) \right] du \right) e_{i+1}^{p-1} \\
&+ \sum_{i\in\mathbb{Z}} \left(\int_{v_{i+1}}^{u_{i+1}} \left[\frac{\partial \mathcal{B}^+}{\partial v}(v_i, v, \Delta z_{i+\frac{1}{2}}) - \frac{\partial \mathcal{B}^+}{\partial v}(v_i, v, 0) \right] dv \right) e_{i+1}^{p-1} \\
&\leq L_B \sum_{i\in\mathbb{Z}} |\Delta z_{i+\frac{1}{2}}| \left(|u_i - v_i| + |u_{i+1} - v_{i+1}| \right) |e_{i+1}^{p-1}|.
\end{aligned}
\tag{2.5}
$$

We proceed in similar way for S_2 and we establish the relations corresponding to (2.4) and (2.5). Therefore, also recalling (1.21), we conclude

$$
S_1 \leq L_B \sum_{i\in\mathbb{Z}} |\Delta z_{i+\frac{1}{2}}| \left(|e_i| + |e_{i+1}| \right) |e_{i+1}^{p-1}|, \tag{2.6}
$$

$$
S_2 \leq L_B \sum_{i\in\mathbb{Z}} |\Delta z_{i+\frac{1}{2}}| \left(|e_i| + |e_{i+1}| \right) |e_i^{p-1}|. \tag{2.7}
$$

Because of $|sgn(e)| \leq 1$, by using the *Hölder's inequality* for $1 \leq p < +\infty$, simple computations lead to obtain $|e_i^{p-1}| \leq |e_i|^{p-1}$. This implies, after rearranging terms in (2.6) and (2.7), that

$$
\begin{aligned}
&\left| \int_{\mathbb{R}} \mathcal{S}(u^h; v^h)\, |e|^{p-1} sgn(e)\, dx \right| \\
&\leq L_B \sum_{i\in\mathbb{Z}} |\Delta z_{i+\frac{1}{2}}| \left(|e_i|^p + |e_i|^{p-1}|e_{i+1}| + |e_i||e_{i+1}|^{p-1} + |e_{i+1}|^p \right).
\end{aligned}
\tag{2.8}
$$

Now the *Young's inequality*, $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$, $\frac{1}{p} + \frac{1}{q} = 1$, applied to (2.8) and the immediate property $|e_i|^p \leq \frac{1}{h} \int_{C_i} |e|^p\, dx$, $i \in \mathbb{Z}$, provide

$$
\left| \int_{\mathbb{R}} \mathcal{S}(u^h; v^h)\, |e|^{p-1} sgn(e)\, dx \right| \leq 2L_B \sum_{i\in\mathbb{Z}} \frac{|\Delta z_{i+\frac{1}{2}}|}{h} \left(\int_{C_i} |e|^p\, dx + \int_{C_{i+1}} |e|^p\, dx \right). \tag{2.9}
$$

In the case of (1.5), according to (2.1a), a direct estimate yields the first order approximation $\frac{|\Delta z_{i+\frac{1}{2}}|}{h} \leq \|z'\|_{L^\infty}$. The proof of (2.3) is thus completed. □

2.2 Consistency estimate

We turn our attention to the consistency error term $\mathcal{C}(u;u^h)$ in (1.20), for which an optimal result in terms of the rate of convergence is obtained.

Lemma 2.2. *For the assumptions of Theorem 1.2, together with* (1.11a), (1.11b) *and* (1.12), *there exists a constant independent of* h *such that*

$$\left|\int_{\mathbb{R}} \mathcal{C}(u;u^h)\,|e|^{p-1}sgn(e)\,dx\right| \le Ch\left(\|z\|_{W^{2,p}}+\|u\|_{W^{1,p}}\right)\|e\|_{L^p}^{p-1}. \tag{2.10}$$

Proof. From (1.20), we have

$$\int_{\mathbb{R}} \mathcal{C}(u;u^h)\,|e|^{p-1}sgn(e)\,dx = \int_{\mathbb{R}} \left[B(x,u)-B^N(x,u^h)\right]|e|^{p-1}sgn(e)\,dx.$$

We compute the integral of the discrete source operator,

$$\begin{aligned}
&\int_{\mathbb{R}} B^N(x,u^h)\,|e|^{p-1}sgn(e)\,dx\\
&=\int_{\mathbb{R}}\left\{\sum_{i\in\mathbb{Z}}\frac{1}{h}\left[\mathcal{B}^+(u_{i-1},u_i,\Delta z_{i-\frac12})+\mathcal{B}^-(u_i,u_{i+1},\Delta z_{i+\frac12})\right]\mathbb{1}_{C_i}\right\}|e|^{p-1}sgn(e)\,dx\\
&=\sum_{i\in\mathbb{Z}}\left[\mathcal{B}^+(u_{i-1},u_i,\Delta z_{i-\frac12})+\mathcal{B}^-(u_i,u_{i+1},\Delta z_{i+\frac12})\right]e_i^{p-1},
\end{aligned}$$

where $e_i^{p-1}=\frac{1}{h}\int_{C_i}|e|^{p-1}sgn(e)\,dx$. Then we decompose as follows,

$$\begin{aligned}
&\int_{\mathbb{R}} B^N(x,u^h)\,|e|^{p-1}sgn(e)\,dx\\
&=\sum_{i\in\mathbb{Z}}\left[\mathcal{B}^+(u_i,u_{i+1},\Delta z_{i+\frac12})+\mathcal{B}^-(u_i,u_{i+1},\Delta z_{i+\frac12})\right]e_i^{p-1}+\\
&+\sum_{i\in\mathbb{Z}}\left[\mathcal{B}^+(u_{i-1},u_i,\Delta z_{i-\frac12})-\mathcal{B}^+(u_i,u_{i+1},\Delta z_{i+\frac12})\right]e_i^{p-1}:=T_1+T_2.
\end{aligned}\tag{2.11}$$

We estimate each T_j, $j=1,2$ separately. Setting $\mathcal{B}=\mathcal{B}^+ +\mathcal{B}^-$, we write

$$\begin{aligned}
T_1&=\sum_{i\in\mathbb{Z}}\left[\mathcal{B}(u_i,u_{i+1},\Delta z_{i+\frac12})-\mathcal{B}(u_i,u_i,\Delta z_{i+\frac12})\right]e_i^{p-1}\\
&+\sum_{i\in\mathbb{Z}}\left[\mathcal{B}^+(u_i,u_i,\Delta z_{i+\frac12})+\mathcal{B}^-(u_i,u_i,\Delta z_{i+\frac12})\right]e_i^{p-1}:=T_1^1+T_1^2.
\end{aligned}\tag{2.12}$$

For the remainder T_1^1, thanks to (1.11a), we get

$$\begin{aligned} T_1^1 &= \sum_{i\in\mathbb{Z}} \left(\int_{u_i}^{u_{i+1}} \left[\frac{\partial \mathcal{B}}{\partial v}(u_i, v, \Delta z_{i+\frac{1}{2}}) - \frac{\partial \mathcal{B}}{\partial v}(u_i, v, 0) \right] dv \right) e_i^{p-1} \\ &\leq L_B \sum_{i\in\mathbb{Z}} |\Delta z_{i+\frac{1}{2}}| |u_{i+1} - u_i| |e_i^{p-1}|, \end{aligned}$$

so that (2.1a) applied to (1.5) and (1.6) leads to conclude

$$T_1^1 \leq L_B \|z'\|_{L^\infty} \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |\partial_x u| \, dx \right) \left(\int_{C_i} |e|^{p-1} \, dx \right). \tag{2.13}$$

The second term T_1^2 of (2.12) can be further decomposed into three parts,

$$\begin{aligned} T_1^2 &= \sum_{i\in\mathbb{Z}} \frac{\mathcal{B}(u_i, u_i, \Delta z_{i+\frac{1}{2}})}{\Delta z_{i+\frac{1}{2}}} \frac{\Delta z_{i+\frac{1}{2}}}{h} h e_i^{p-1} \\ &= \sum_{i\in\mathbb{Z}} \frac{\mathcal{B}(u_i, u_i, \Delta z_{i+\frac{1}{2}})}{\Delta z_{i+\frac{1}{2}}} \left[\frac{\Delta z_{i+\frac{1}{2}}}{h} - z'(x_i) \right] h e_i^{p-1} \\ &+ \sum_{i\in\mathbb{Z}} \left[\frac{\mathcal{B}(u_i, u_i, \Delta z_{i+\frac{1}{2}})}{\Delta z_{i+\frac{1}{2}}} - b(u_i) \right] z'(x_i) \, h e_i^{p-1} \\ &+ \sum_{i\in\mathbb{Z}} z'(x_i) b(u_i) \, h e_i^{p-1} := T_1^{2,1} + T_1^{2,2} + T_1^{2,3}. \end{aligned} \tag{2.14}$$

We give details for each part. By using (1.11a), from (2.1b) we deduce that

$$\begin{aligned} T_1^{2,1} &\leq \sum_{i\in\mathbb{Z}} \left| \frac{\mathcal{B}(u_i, u_i, \Delta z_{i+\frac{1}{2}}) - \mathcal{B}(u_i, u_i, 0)}{\Delta z_{i+\frac{1}{2}}} \right| \left| \frac{\Delta z_{i+\frac{1}{2}}}{h} - z'(x_i) \right| h |e_i^{p-1}| \\ &\leq L_B \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |z''| \, dx \right) \left(\int_{C_i} |e|^{p-1} \, dx \right). \end{aligned} \tag{2.15}$$

Because of the consistency property (1.12) and (2.1a) for (1.5), we derive

$$T_1^{2,2} \leq K_B \|z'\|_{L^\infty} \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |z'| \, dx \right) \left(\int_{C_i} |e|^{p-1} \, dx \right). \tag{2.16}$$

Finally, the third term in (2.14) is equivalent to the integral of the analytical source operator (1.3). Indeed, by means of Taylor's expansions in (1.6), we obtain the *midpoint formula* (dropping the time dependence, for simplicity)

$$u_i = u(x_i) + R_i, \quad R_i = \frac{1}{h} \int_{C_i} \partial_x u(\xi(x))(x - x_i) \, dx,$$

for some $\xi(x) \in C_i$, and the regularity assumed in (1.3) guarantees that

$$b(u_i) = b(u(x_i)) + b'(\nu_i)R_i, \quad |b'(\nu_i)| \leq L_b, \ \forall i \in \mathbb{Z}.$$

We thus write

$$T_1^{2,3} = \sum_{i\in\mathbb{Z}} z'(x_i)b(u(x_i))\, he_i^{p-1} + R_1, \tag{2.17}$$

where the remainder satisfies

$$R_1 \leq L_b\|z'\|_{L^\infty} \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |\partial_x u|\, dx\right)\left(\int_{C_i} |e|^{p-1}\, dx\right). \tag{2.18}$$

The *Taylor's formula* (2.2), applied to the source term (1.3), yields

$$\int_{\mathbb{R}} B(x,u)\,|e|^{p-1}sgn(e)\,dx = \sum_{i\in\mathbb{Z}} \int_{C_i} z'(x_i)b(u(x_i))\,|e|^{p-1}sgn(e)\,dx + R_2$$

and we readily check that the remainder satisfies

$$\begin{aligned} R_2 \leq\ & L_b \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |z''|\, dx\right)\left(\int_{C_i} |e|^{p-1}\, dx\right) \\ & + L_b\|z'\|_{L^\infty} \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |\partial_x u|\, dx\right)\left(\int_{C_i} |e|^{p-1}\, dx\right). \end{aligned} \tag{2.19}$$

Therefore, from (2.17) we have

$$T_1^{2,3} = \int_{\mathbb{R}} B(x,u)\,|e|^{p-1}sgn(e)\,dx + R_1 - R_2, \tag{2.20}$$

with the remainders (2.18) and (2.19), which are conveniently bounded. Coming back to decomposition (2.11), using (1.11a) we have for the last term

$$\begin{aligned} T_2 =& \sum_{i\in\mathbb{Z}} \left[\mathcal{B}^+(u_{i-1}, u_i, \Delta z_{i-\frac{1}{2}}) - \mathcal{B}^+(u_i, u_i, \Delta z_{i-\frac{1}{2}})\right] e_i^{p-1} \\ &+ \sum_{i\in\mathbb{Z}} \left[\mathcal{B}^+(u_i, u_i, \Delta z_{i-\frac{1}{2}}) - \mathcal{B}^+(u_i, u_i, \Delta z_{i+\frac{1}{2}})\right] e_i^{p-1} \\ &+ \sum_{i\in\mathbb{Z}} \left[\mathcal{B}^+(u_i, u_i, \Delta z_{i+\frac{1}{2}}) - \mathcal{B}^+(u_i, u_{i+1}, \Delta z_{i+\frac{1}{2}})\right] e_i^{p-1} \\ =& \sum_{i\in\mathbb{Z}} \left(\int_{u_i}^{u_{i-1}} \left[\frac{\partial \mathcal{B}^+}{\partial u}(u, u_i, \Delta z_{i-\frac{1}{2}}) - \frac{\partial \mathcal{B}^+}{\partial u}(u, u_i, 0)\right] du\right) e_i^{p-1} \\ &+ \sum_{i\in\mathbb{Z}} \left[\mathcal{B}^+(u_i, u_i, \Delta z_{i-\frac{1}{2}}) - \mathcal{B}^+(u_i, u_i, \Delta z_{i+\frac{1}{2}})\right] e_i^{p-1} \\ &+ \sum_{i\in\mathbb{Z}} \left(\int_{u_{i+1}}^{u_i} \left[\frac{\partial \mathcal{B}^+}{\partial v}(u_i, v, \Delta z_{i+\frac{1}{2}}) - \frac{\partial \mathcal{B}^+}{\partial v}(u_i, v, 0)\right] dv\right) e_i^{p-1} \end{aligned}$$

$$\leq L_B \sum_{i\in\mathbb{Z}} \Big(|\Delta z_{i-\frac{1}{2}}||u_i - u_{i-1}| + |\Delta z_{i+\frac{1}{2}} - \Delta z_{i-\frac{1}{2}}| \\ + |\Delta z_{i+\frac{1}{2}}||u_{i+1} - u_i| \Big) |e_i^{p-1}|,$$

so that we conclude by means of (2.1a) and (2.1c),

$$\begin{aligned} T_2 &\leq 2L_B \|z'\|_{L^\infty} \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |\partial_x u| \, dx \right) \left(\int_{C_i} |e|^{p-1} \, dx \right) \\ &+ L_B \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |z''| \, dx \right) \left(\int_{C_i} |e|^{p-1} \, dx \right). \end{aligned} \tag{2.21}$$

We put (2.11), (2.12), (2.14) and (2.20) together, with the corresponding estimates stated in (2.13), (2.15), (2.16), (2.18), (2.19) and (2.21). We apply the discrete *Hölder's inequality*, $\sum_{i\in\mathbb{Z}} a_i b_i \leq \left(\sum_{i\in\mathbb{Z}} a_i^p\right)^{\frac{1}{p}} \cdot \left(\sum_{i\in\mathbb{Z}} b_i^q\right)^{\frac{1}{q}}$, $\frac{1}{p} + \frac{1}{q} = 1$, to the products and then the usual continuous inequality to each integral on the mesh cells. This provides a coefficient h in front of all expressions and the result in (2.10) thus follows, with $C := C(L_B, K_B, L_b, \|z'\|_{L^\infty})$. □

2.3 Proof of Theorem 1.2

We multiply equation (1.20) by $|e|^{p-1} sgn(e)$ and we integrate as follows,

$$\begin{aligned} &\int_{\mathbb{R}} (\partial_t e + \partial_x e) \, |e|^{p-1} sgn(e) \, dx \\ &\quad = \int_{\mathbb{R}} \mathcal{C}(u; u^h) \, |e|^{p-1} sgn(e) \, dx + \int_{\mathbb{R}} \mathcal{S}(u^h; v^h) \, |e|^{p-1} sgn(e) \, dx. \end{aligned} \tag{2.22}$$

An integration by parts shows $\int_{\mathbb{R}} |e|^{p-1} sgn(e) \, \partial_x e \, dx = 0$, then we deduce from (2.22), (2.3) and (2.10) that

$$\frac{1}{p} \partial_t \|e(t)\|_{L^p}^p \leq C \|e(t)\|_{L^p}^p + Ch \left(\|z\|_{W^{2,p}} + \|u(t)\|_{W^{1,p}} \right) \|e(t)\|_{L^p}^{p-1}. \tag{2.23}$$

Let $t^* \in \mathbb{R}_+$ be such that $\|e(t^*)\|_{L^p} = \max_{t\in\mathbb{R}_+} \|e(t)\|_{L^p}$. By integrating in time from 0 to t^*, we get

$$\begin{aligned} \|e(t^*)\|_{L^p}^p &\leq \|e(0)\|_{L^p}^p + Cp \int_0^{t^*} \|e(s)\|_{L^p}^p \, ds \\ &+ Cp\, h \|z\|_{W^{2,p}} \int_0^{t^*} \|e(s)\|_{L^p}^{p-1} \, ds + Cp\, h \int_0^{t^*} \|u(s)\|_{W^{1,p}} \|e(s)\|_{L^p}^{p-1} \, ds \\ &\leq \|e(0)\|_{L^p} \|e(t^*)\|_{L^p}^{p-1} + Cp \|e(t^*)\|_{L^p}^{p-1} \int_0^{t^*} \|e(s)\|_{L^p} \, ds \\ &+ Cp\, h\, t^* \|z\|_{W^{2,p}} \|e(t^*)\|_{L^p}^{p-1} + Cp\, h \|e(t^*)\|_{L^p}^{p-1} \int_0^{t^*} \|u(s)\|_{W^{1,p}} \, ds, \end{aligned}$$

which implies that

$$\begin{aligned}\|e(t^*)\|_{L^p} \leq \|e(0)\|_{L^p} + Cp \int_0^{t^*} \|e(s)\|_{L^p}\, ds \\ + Cp\, h\, t^* \|z\|_{W^{2,p}} + Cp\, h \int_0^{t^*} \|u(s)\|_{W^{1,p}}\, ds.\end{aligned} \tag{2.24}$$

Finally, a straightforward extension of *Gronwall's inequality* yields the desired result (1.22), where $C(t) := C(t; p, L_B, K_B, L_b, \|z'\|_{L^\infty})$ is any positive constant depending on time by the factor $exp\{-Ct\}$.

3 Error estimates for second order schemes

The convergence properties of the approximation (1.16) are shown by mimicking the proof of the analogous results for (1.10), provided in Section 2.

As in the case of first order approximations, we derive some preliminary estimates on the discrete differences of numerical functions.
For a function $w \in C^2$, with cell-averages $w_i = \frac{1}{h}\int_{C_i} w(x)\, dx$, $i \in \mathbb{Z}$, we construct piecewise linear approximations on the spatial mesh by means of the coefficients

$$\bar{w}_i(x) = w_i + (x - x_i) w_i', \quad i \in \mathbb{Z},\ x \in C_i, \tag{3.1}$$

where the numerical derivatives are defined as appropriate interpolations of the discrete increments between neighboring cells,

$$w_i' = lmtr\left\{\frac{w_{i+1} - w_i}{h}, \frac{w_i - w_{i-1}}{h}\right\}, \quad i \in \mathbb{Z}. \tag{3.2}$$

We consider a general representation of the slope *limiter* introduced in the above formula, i.e. if $M = lmtr\{\alpha, \beta\}$, then $M = \kappa\alpha + \lambda\beta$, with $\kappa, \lambda \in [0, 1]$ and $\kappa + \lambda = 1$ or $\kappa + \lambda = 0$. In particular, we restrict our analysis to the special class of operators which satisfy the condition $\kappa_i + \lambda_i = 1$, $\forall i \in \mathbb{Z}$ (that excludes, for instance, the classical *minmod limiter* in the case of nonmonotonic numerical functions). We also assume that the numerical application (3.2) relating the cell-averages w_j, $j = i-1, i, i+1$, to the discrete derivative w_i' is Lipschitz continuous on its arguments, with constant $\frac{C}{h}$. Several examples of slope *limiter* which satisfies these properties have been formulated in the literature (refer to [12], [13], [14], [26], [32] and [33]).

We deduce from those definitions that

$$w_i' = \kappa_i \frac{w_{i+1} - w_i}{h} + \lambda_i \frac{w_i - w_{i-1}}{h}, \quad i \in \mathbb{Z}. \tag{3.3}$$

The interfacial values of the reconstruction (3.1) are given by

$$w_i^- = \bar{w}_i(x_{i-\frac{1}{2}}) = w_i - \frac{h}{2}w_i', \quad w_i^+ = \bar{w}_i(x_{i+\frac{1}{2}}) = w_i + \frac{h}{2}w_i', \tag{3.4}$$

and we are interested in evaluating the jumps at the interfaces, i.e. $w_{i+1}^- - w_i^+$. Taking into account (3.4) and (3.3), we have

$$\begin{aligned} w_{i+1}^- - w_i^+ &= w_{i+1} - w_i - \frac{h}{2}\left(w_i' + w_{i+1}'\right) \\ &= \left(1 - \frac{\kappa_i}{2} - \frac{\lambda_{i+1}}{2}\right) W_1 - \frac{\lambda_i}{2} W_2 - \frac{\kappa_{i+1}}{2} W_3, \end{aligned} \tag{3.5}$$

where we indicate

$$W_1 = w_{i+1} - w_i, \quad W_2 = w_i - w_{i-1}, \quad W_3 = w_{i+2} - w_{i+1}. \tag{3.6}$$

Consequently, to deal with (3.5) we use the same procedure as (2.1a)-(2.1b), based on high order Taylor expansions applied to the various terms in (3.6). Besides, we observe that

$$w_{i+1} - w_i = \frac{1}{h}\int_{C_i} [w(x+h) - w(x)]\, dx = \frac{1}{h}\int_{C_i}\int_0^h w'(x+s)\, ds\, dx,$$

then the following result holds,

$$|W_j| \leq \|w'\|_{L^1(C_i)}, \quad i \in \mathbb{Z}, \quad j = 1, 2, 3. \tag{3.7}$$

The simplest first order approximation reads

$$\begin{aligned} w_{i+1}^- - w_i^+ = &\left(1 - \frac{\kappa_i}{2} - \frac{\lambda_{i+1}}{2}\right)\int_{C_i} w'(\xi(x))\, dx \\ &- \frac{\lambda_i}{2}\int_{C_i} w'(\eta(x))\, dx - \frac{\kappa_{i+1}}{2}\int_{C_i} w'(\vartheta(x))\, dx, \end{aligned} \tag{3.8}$$

for some $\xi(x), \eta(x), \vartheta(x) \in C_i$, so it follows also from (3.7) that

$$|w_{i+1}^- - w_i^+| \leq D_{i+\frac{1}{2}}\|w\|_{W^{1,1}} \quad \text{or} \quad |w_{i+1}^- - w_i^+| \leq D_{i+\frac{1}{2}} h \|w'\|_{L^\infty}. \tag{3.9}$$

Recalling that $\kappa_i + \lambda_i = 1$, $\forall i \in \mathbb{Z}$, we obtain the second order approximation

$$\begin{aligned} w_{i+1}^- - w_i^+ = &\left(1 - \frac{\kappa_i}{2} - \frac{\lambda_{i+1}}{2}\right)\int_{C_i} w''(\xi(x))(x - x_i)\, dx \\ &+ \frac{\lambda_i}{2}\int_{C_i} w''(\eta(x))(x - x_i)\, dx - \frac{3}{2}\,\kappa_{i+1}\int_{C_i} w''(\vartheta(x))(x - x_i)\, dx, \end{aligned}$$

for some $\xi(x), \eta(x), \vartheta(x) \in C_i$, and then it follows

$$|w_{i+1}^- - w_i^+| \leq D_{i+\frac{1}{2}} h \|w\|_{W^{2,1}} \quad \text{or} \quad |w_{i+1}^- - w_i^+| \leq D_{i+\frac{1}{2}} h^2 \|w''\|_{L^\infty}. \tag{3.10}$$

Remark 3.1. *We note that the constant in* (3.9) *and* (3.10) *satisfies, uniformly for* $i \in \mathbb{Z}$*, the estimate* $D_{i+\frac{1}{2}} \leq \max\left\{\left(1-\frac{\kappa_i}{2}-\frac{\lambda_{i+1}}{2}\right), \frac{\lambda_i}{2}, \frac{3}{4}\kappa_{i+1}\right\} \leq 1$*. Moreover, for any set of values* $(\kappa_i, \lambda_i)_{i\in\mathbb{Z}}$*, the bounds on these quantities are always not degenerate.*

Finally, a long but straightforward computation, involving also the third order expansions, leads to conclude that

$$\begin{aligned} w_{i+1}^- - w_i^+ &= (\lambda_{i+1}-\kappa_i)\,\frac{h^2}{2}\,w''(x_i) \\ &+ \left(1-\frac{\kappa_i}{2}-\frac{\lambda_{i+1}}{2}\right)\int_{C_i} w'''(\xi(x))(x-x_i)^2\,dx \\ &- \frac{\lambda_i}{2}\int_{C_i} w'''(\eta(x))(x-x_i)^2 dx - \frac{\kappa_{i+1}}{2}\int_{C_i} w'''(\vartheta(x))(x-x_i)^2 dx, \end{aligned} \tag{3.11}$$

for some $\xi(x), \eta(x), \vartheta(x) \in C_i$.

Remark 3.2. *According to the piecewise linear reconstruction* (3.1)*, discrete interfacial jumps approximate the second derivative of the numerical functions, as it can be roughly deduced from* (3.5)*.*

3.1 Stability estimate

The following result corresponds to that presented in Section 2.1 and then we adapt the proof of Lemma 2.1 in the case of the numerical source operator (1.16).

Lemma 3.3. *For the assumptions of Theorem 1.3, together with* (1.11a)*, there exists a constant* $C := C(L_B, L_b, \|z'\|_{L^\infty}, \|z''\|_{L^\infty})$*, independent of* h*, such that*

$$\left|\int_{\mathbb{R}} \mathcal{S}(u^h; v^h)\,|e|^{p-1} sgn(e)\,dx\right| \leq C\|e\|_{L^p}^p. \tag{3.12}$$

Proof. From (1.20), (1.16) and (1.6), we deduce

$$\begin{aligned} &\int_{\mathbb{R}} \mathcal{S}(u^h; v^h)\,|e|^{p-1} sgn(e)\,dx \\ &= \int_{\mathbb{R}} \left\{ \sum_{i\in\mathbb{Z}} \frac{1}{h}\left[\mathcal{B}^+(u_{i-1}^+, u_i^-, \Delta z_{i-\frac{1}{2}}) + \mathcal{B}^-(u_i^+, u_{i+1}^-, \Delta z_{i+\frac{1}{2}})\right] \mathbb{1}_{C_i} \right. \\ &\qquad \left. - \sum_{i\in\mathbb{Z}} \frac{1}{h}\left[\mathcal{B}^+(v_{i-1}^+, v_i^-, \Delta z_{i-\frac{1}{2}}) + \mathcal{B}^-(v_i^+, v_{i+1}^-, \Delta z_{i+\frac{1}{2}})\right] \mathbb{1}_{C_i} \right\} |e|^{p-1} sgn(e)\,dx \\ &+ \int_{\mathbb{R}} \left\{ \sum_{i\in\mathbb{Z}} z_i'\,b(u_i)\,\mathbb{1}_{C_i} - \sum_{i\in\mathbb{Z}} z_i'\,b(v_i)\,\mathbb{1}_{C_i} \right\} |e|^{p-1} sgn(e)\,dx \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i\in\mathbb{Z}} \left[\mathcal{B}^+(u_i^+, u_{i+1}^-, \Delta z_{i+\frac{1}{2}}) - \mathcal{B}^+(v_i^+, v_{i+1}^-, \Delta z_{i+\frac{1}{2}})\right] e_{i+1}^{p-1} \\
&+ \sum_{i\in\mathbb{Z}} \left[\mathcal{B}^-(u_i^+, u_{i+1}^-, \Delta z_{i+\frac{1}{2}}) - \mathcal{B}^-(v_i^+, v_{i+1}^-, \Delta z_{i+\frac{1}{2}})\right] e_i^{p-1} \qquad (3.13) \\
&+ \sum_{i\in\mathbb{Z}} z_i' \left[b(u_i) - b(v_i)\right] h e_i^{p-1} := S_1 + S_2 + S_3,
\end{aligned}
$$

where $e_i^{p-1} = \frac{1}{h}\int_{C_i} |e|^{p-1} sgn(e)\, dx$.
To deal with S_1 and S_2, we proceed exactly as in (2.4) and (2.5). Because of the definition (3.3) and (3.4), together with the Lipschitz property of the application (3.2), simple computations lead to verify that

$$
|u_i^+ - v_i^+| \leq \max\left(|u_i - v_i| + |u_{i+1} - v_{i+1}| + |u_{i-1} - v_{i-1}|\right),
$$

and the same relation is satisfied by $|u_i^- - v_i^-|$. So, we can establish for the second order methods similar estimates to (2.6) and (2.7).
On the other hand, a direct treatment of the last term in (3.13) yields

$$
S_3 \leq L_b \sum_{i\in\mathbb{Z}} |z_i'|\, |u_i - v_i|\, h |e_i^{p-1}|. \qquad (3.14)
$$

We give some details about the estimate of numerical derivatives (3.3), for the particular case of (1.5), we will use later on the proofs.
By performing appropriate expansions, also recalling that $\kappa_i + \lambda_i = 1$, $\forall i \in \mathbb{Z}$, we obtain

$$
\begin{aligned}
z_i' &= \frac{\kappa_i}{h}\int_{C_i} z'(\xi(x))\, dx + \frac{\lambda_i}{h}\int_{C_i} z'(\eta(x))\, dx \\
&= z'(x_i) + (\kappa_i - \lambda_i)\frac{h}{2} z''(x_i) \qquad (3.15) \\
&+ \frac{\kappa_i}{3}\, h \int_{C_i} z'''(\vartheta(x))\, dx + \frac{\lambda_i}{3}\, h \int_{C_i} z'''(\varrho(x))\, dx,
\end{aligned}
$$

for some $\xi(x), \eta(x), \vartheta(x), \varrho(x) \in C_i$, which implies that $|z_i'| \leq \|z'\|_{L^\infty}$ in (3.14). Thanks to the arguments used for passing to (2.8) and (2.9), with the first order approximation (3.9) applied to (1.15b), we conclude (3.12). □

3.2 Consistency estimate

The proof of the following result is also an extension of that of Lemma 2.2.

Lemma 3.4. *For the assumptions of Theorem 1.3, together with* (1.11a), (1.11b) *and* (1.12), *there exists a constant independent of* h *such that*

$$\left| \int_{\mathbb{R}} \mathcal{C}(u;u^h)\,|e|^{p-1} sgn(e)\,dx \right| \leq Ch^2 \left(\|z\|_{W^{3,p}} + \|u\|_{W^{2,p}} \right) \|e\|_{L^p}^{p-1}. \tag{3.16}$$

Proof. We consider the integral of the source operator (1.16), computed on the approximation (1.6) of the analytical solution,

$$\begin{aligned}
&\int_{\mathbb{R}} B^N(x,u^h)\,|e|^{p-1} sgn(e)\,dx \\
&= \int_{\mathbb{R}} \left\{ \sum_{i\in\mathbb{Z}} \frac{1}{h} \left[\mathcal{B}^+(u_{i-1}^+, u_i^-, \Delta z_{i-\frac{1}{2}}) + \mathcal{B}^-(u_i^+, u_{i+1}^-, \Delta z_{i+\frac{1}{2}}) \right] \mathbb{1}_{C_i} \right\} |e|^{p-1} sgn(e)\,dx \\
&+ \int_{\mathbb{R}} \left\{ \sum_{i\in\mathbb{Z}} z_i' b(u_i) \mathbb{1}_{C_i} \right\} |e|^{p-1} sgn(e)\,dx \\
&= \sum_{i\in\mathbb{Z}} \left[\mathcal{B}^+(u_{i-1}^+, u_i^-, \Delta z_{i-\frac{1}{2}}) + \mathcal{B}^-(u_i^+, u_{i+1}^-, \Delta z_{i+\frac{1}{2}}) \right] e_i^{p-1} + \sum_{i\in\mathbb{Z}} z_i' b(u_i)\, h e_i^{p-1},
\end{aligned}$$

where we set $e_i^{p-1} = \frac{1}{h}\int_{C_i} |e|^{p-1} sgn(e)\,dx$. In the sequel, we neglect the dependence on time of the numerical functions to simplify the notation.
We decompose the first part of the above formula, similarly to (2.11), into two terms T_j, $j=1,2$ treated separately. The remainder can be rewritten as

$$\begin{aligned}
T_2 = &\sum_{i\in\mathbb{Z}} \left[\mathcal{B}^+(u_{i-1}^+, u_i^-, \Delta z_{i-\frac{1}{2}}) - \mathcal{B}^+(u_i^+, u_i^-, \Delta z_{i-\frac{1}{2}}) \right] e_i^{p-1} \\
&+ \sum_{i\in\mathbb{Z}} \left[\mathcal{B}^+(u_i^+, u_i^-, \Delta z_{i-\frac{1}{2}}) - \mathcal{B}^+(u_i^+, u_i^-, \Delta z_{i+\frac{1}{2}}) \right] e_i^{p-1} \\
&+ \sum_{i\in\mathbb{Z}} \left[\mathcal{B}^+(u_i^+, u_i^-, \Delta z_{i+\frac{1}{2}}) - \mathcal{B}^+(u_i^+, u_{i+1}^-, \Delta z_{i+\frac{1}{2}}) \right] e_i^{p-1}
\end{aligned}$$

and the usual procedures for the differences, by using (1.11a), leads to deduce

$$\begin{aligned}
T_2 \leq L_B \sum_{i\in\mathbb{Z}} \Big(&|\Delta z_{i-\frac{1}{2}}||u_i^+ - u_{i-1}^+| + |\Delta z_{i+\frac{1}{2}} - \Delta z_{i-\frac{1}{2}}| \\
&+ |\Delta z_{i+\frac{1}{2}}||u_{i+1}^- - u_i^-| \Big) |e_i^{p-1}|.
\end{aligned} \tag{3.17}$$

According to the definition (3.3) and (3.4), concerning (1.14), we easily obtain

$$|u_i^+ - u_{i-1}^+| = u_i - u_{i-1} + \frac{h}{2}\left(u_i' - u_{i-1}'\right) \leq \int_{C_i} |\partial_x u|\,dx,$$

with an analogous estimate for $|u^-_{i+1} - u^-_i|$, while a second order approximation is needed for the central term in (3.17), that is

$$|\Delta z_{i+\frac{1}{2}} - \Delta z_{i-\frac{1}{2}}| = (z_{i+1} - 2z_i + z_{i-1}) - \frac{h}{2}\left(z'_{i+1} - z'_{i-1}\right) \le h^2 \int_{C_i} |z'''|\, dx.$$

These estimates and (3.10) for (1.15b) provide an analogous inequality to (2.21),

$$\begin{aligned} T_2 \le{}& 2L_B\, h\|z''\|_{L^\infty} \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |\partial_x u|\, dx\right)\left(\int_{C_i} |e|^{p-1}\, dx\right) \\ &+ L_B h \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |z'''|\, dx\right)\left(\int_{C_i} |e|^{p-1}\, dx\right). \end{aligned} \tag{3.18}$$

For the term corresponding to (2.12), setting $\mathcal{B} = \mathcal{B}^+ + \mathcal{B}^-$, we thus have

$$\begin{aligned} T_1 ={}& \sum_{i\in\mathbb{Z}} \left[\mathcal{B}(u_i^+, u^-_{i+1}, \Delta z_{i+\frac{1}{2}}) - \mathcal{B}(u_i^+, u_i^+, \Delta z_{i+\frac{1}{2}})\right] e_i^{p-1} \\ &+ \sum_{i\in\mathbb{Z}} \mathcal{B}(u_i^+, u_i^+, \Delta z_{i+\frac{1}{2}})\, e_i^{p-1} := T_1^1 + T_1^2. \end{aligned} \tag{3.19}$$

We use again the property (1.11a) and we deduce

$$T_1^1 \le L_B \sum_{i\in\mathbb{Z}} |\Delta z_{i+\frac{1}{2}}||u^-_{i+1} - u_i^+||e_i^{p-1}|,$$

to conclude from (3.8) and (3.10) that

$$T_1^1 \le L_B\, h\|z''\|_{L^\infty} \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |\partial_x u|\, dx\right)\left(\int_{C_i} |e|^{p-1}\, dx\right). \tag{3.20}$$

The second term of (3.19) is further decomposed, also thanks to (3.11),

$$\begin{aligned} T_1^2 ={}& \sum_{i\in\mathbb{Z}} \frac{\mathcal{B}(u_i^+, u_i^+, \Delta z_{i+\frac{1}{2}})}{\Delta z_{i+\frac{1}{2}}} \left[\frac{\Delta z_{i+\frac{1}{2}}}{h} - Q_{i+\frac{1}{2}} \frac{h}{2} z''(x_i)\right] h e_i^{p-1} \\ &+ \sum_{i\in\mathbb{Z}} \left[\frac{\mathcal{B}(u_i^+, u_i^+, \Delta z_{i+\frac{1}{2}})}{\Delta z_{i+\frac{1}{2}}} - b(u_i^+)\right] Q_{i+\frac{1}{2}} \frac{h}{2} z''(x_i)\, h e_i^{p-1} \\ &+ \sum_{i\in\mathbb{Z}} Q_{i+\frac{1}{2}} \frac{h}{2} z''(x_i) b(u_i^+)\, h e_i^{p-1} := T_1^{2,1} + T_1^{2,2} + T_1^{2,3}, \end{aligned} \tag{3.21}$$

where $Q_{i+\frac{1}{2}} = \lambda_{i+1} - \kappa_i \le 1$, $\forall i \in \mathbb{Z}$, for the properties of coefficients in (3.3). We give a few details of the estimate for each part. We proceed as in (2.15), by means of (1.11a) and (3.11), to obtain

$$T_1^{2,1} \le L_B h \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |z'''|\, dx\right)\left(\int_{C_i} |e|^{p-1}\, dx\right). \tag{3.22}$$

From the consistency bound (1.12), together with the approximation (3.8), we derive that

$$T_1^{2,2} \leq K_B\, h \|z''\|_{L^\infty} \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |z'|\, dx \right) \left(\int_{C_i} |e|^{p-1}\, dx \right). \tag{3.23}$$

Then we pass to the crucial point of the proof, to show the convergence towards the integral of the analytical source operator (1.3). On the one hand, by applying to that function classical Taylor's expansions, we have

$$\begin{aligned} \int_{\mathbb{R}} B(x,u)\, |e|^{p-1} sgn(e)\, dx = & \sum_{i\in\mathbb{Z}} \int_{C_i} z'(x_i) b(u(x_i))\, |e|^{p-1} sgn(e)\, dx \\ & + \sum_{i\in\mathbb{Z}} \int_{C_i} \left(z' b(u)\right)' (\xi(x_i))(x - x_i)\, |e|^{p-1} sgn(e)\, dx, \end{aligned} \tag{3.24}$$

for some $\xi(x_i) \in C_i$. On the other hand, recalling the definition of interfacial values (1.15a) and by the regularity assumed in (1.3), we can write

$$b(u_i^+) = b(u_i) + b'(\nu_i) \frac{h}{2} u_i', \quad |b'(\nu_i)| \leq L_b,\ \forall i \in \mathbb{Z},$$

so that from (3.21) we deduce

$$T_1^{2,3} = \sum_{i\in\mathbb{Z}} Q_{i+\frac{1}{2}} \frac{h}{2} z''(x_i) b(u_i)\, h e_i^{p-1} + R_1 \tag{3.25}$$

and we use analogous approximations to (3.15) for the numerical derivatives of the analytical solution to obtain

$$R_1 \leq L_b h \|z''\|_{L^\infty} \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |\partial_x u|\, dx \right) \left(\int_{C_i} |e|^{p-1}\, dx \right). \tag{3.26}$$

To conclude the announced result, we need to take into account the contribution of the additional term in the numerical source operator, neglected in the first part of the proof. The second order approximation of cell-averages,

$$u_i = u(x_i) + R_i, \quad R_i = \frac{1}{h} \int_{C_i} \partial_{xx} u(\xi(x)) \frac{(x - x_i)^2}{2}\, dx, \tag{3.27}$$

for some $\xi(x) \in C_i$, together with the usual Taylor's expansion

$$b(u_i) = b(u(x_i)) + b'(\nu_i) R_i, \quad |b'(\nu_i)| \leq L_b,\ \forall i \in \mathbb{Z}, \tag{3.28}$$

after proceeding according to (3.15), setting $P_{i+\frac{1}{2}} = \kappa_i - \lambda_i$, $i \in \mathbb{Z}$, leads to

$$\begin{aligned}\sum_{i\in\mathbb{Z}} z_i' b(u_i)\, h e_i^{p-1} = &\sum_{i\in\mathbb{Z}} z'(x_i) b(u(x_i))\, h e_i^{p-1} + R_2 \\ &+ \sum_{i\in\mathbb{Z}} P_{i+\frac{1}{2}} \frac{h}{2} z''(x_i) b(u_i)\, h e_i^{p-1} + R_3,\end{aligned} \tag{3.29}$$

with the following estimates for the remainders,

$$R_2 \le L_b\, h \|z'\|_{L^\infty} \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |\partial_{xx} u|\, dx \right) \left(\int_{C_i} |e|^{p-1}\, dx \right), \tag{3.30}$$

$$R_3 \le L_b\, h \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |z'''|\, dx \right) \left(\int_{C_i} |e|^{p-1}\, dx \right). \tag{3.31}$$

Therefore, up to the bounded remainders (3.26), (3.30) and (3.31), by combining (3.24), (3.25) and (3.29), we have

$$\begin{aligned}&\int_{\mathbb{R}} B(x,u)\, |e|^{p-1} sgn(e)\, dx - T_1^{2,3} - \sum_{i\in\mathbb{Z}} z_i' b(u_i)\, h e_i^{p-1} \\ &= \sum_{i\in\mathbb{Z}} \int_{C_i} z''(\xi(x_i))\, b(u(\xi(x_i)))\, (x - x_i)\, |e|^{p-1} sgn(e)\, dx \\ &+ \sum_{i\in\mathbb{Z}} \int_{C_i} z'(\xi(x_i))\, b'(u(\xi(x_i)))\, u'(\xi(x_i))\, (x - x_i)\, |e|^{p-1} sgn(e)\, dx \\ &- \sum_{i\in\mathbb{Z}} (P_{i+\frac{1}{2}} + Q_{i+\frac{1}{2}}) \frac{h}{2} z''(x_i) b(u(x_i))\, h e_i^{p-1} \\ &- \sum_{i\in\mathbb{Z}} (P_{i+\frac{1}{2}} + Q_{i+\frac{1}{2}}) \frac{h}{2} z''(x_i) b'(\nu_i) R_i\, h e_i^{p-1},\end{aligned} \tag{3.32}$$

where again we used (3.27)-(3.28) and $P_{i+\frac{1}{2}} + Q_{i+\frac{1}{2}} = \lambda_{i+1} - \lambda_i$, $i \in \mathbb{Z}$.
We introduce an appropriate hypothesis on the slope *limiter* (3.2)-(3.3), as discussed in Section 1.3, namely an additional property for its coefficients,

$$\exists\, \Lambda_0 > 0 \quad \text{such that} \quad \lambda_{i+1} - \lambda_i \ge \Lambda_0,\ \forall i \in \mathbb{Z}. \tag{3.33}$$

This condition and general properties of the numerical functions allow us to rewrite the difference between first and third term in the right-hand side of (3.32) in integral form, to conclude that

$$\begin{aligned}&\sum_{i\in\mathbb{Z}} \int_{C_i} z''(\xi(x_i))\, b(u(\xi(x_i)))\, (x - x_i)\, |e|^{p-1} sgn(e)\, dx \\ &- \sum_{i\in\mathbb{Z}} (P_{i+\frac{1}{2}} + Q_{i+\frac{1}{2}}) \frac{h}{2} z''(x_i) b(u(x_i))\, h e_i^{p-1}\end{aligned}$$

$$\leq h \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |\,(z''b(u))'\,|\,dx \right) \left(\int_{C_i} |e|^{p-1}\,dx \right)$$
$$\leq L_b\, h \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |z'''|\,dx \right) \left(\int_{C_i} |e|^{p-1}\,dx \right)$$
$$+ L_b\, h \|z''\|_{L^\infty} \sum_{i\in\mathbb{Z}} \left(\int_{C_i} |\partial_x u|\,dx \right) \left(\int_{C_i} |e|^{p-1}\,dx \right).$$

An analogous estimate is proved for the difference between second and fourth term in the right-hand side of (3.32), which also involves the second derivative of the analytical solution, similarly to the remainder (3.30).
We apply to the previous computations the arguments in the conclusion of Lemma 2.2, to obtain (3.16) with $C = C(L_B, K_B, L_b, \|z'\|_{L^\infty}, \|z''\|_{L^\infty})$. □

3.3 Proof of Theorem 1.3 and Theorem 1.4

With the stability estimate (3.12) and the consistency estimate (3.16), we proceed as in Section 2.3 to conclude the second order error estimate (1.23).

The proof of Theorem 1.4 is obtained by using the main tools introduced for proving the results of Section 2 and Section 3.
Because of the consistency hypotheses (1.11a)-(1.18), the same techniques as in Lemma 2.2 extend to the numerical source operator (1.17), while we apply the arguments formulated in Lemma 3.3 to deduce stability estimates.

4 Remarks and numerical evidence

The principal issue in the proofs of Theorem 1.2 and Theorem 1.3 is to establish the consistency estimates (2.10) and (3.16) respectively, in particular to show the convergence of the numerical source operators towards the analytical source term (1.3) from the relations (2.17) and (3.25).

We note that, due to the introduction of piecewise linear reconstructions of the function z, the differences of interfacial values approximate the second order derivative and the upwind part of the discretization (1.16) "overtakes" the desired result; an additional term is thus needed to recover the first order derivatives in the Taylor's expansion of the source term. Moreover, some restrictions on the definition of the slope *limiter* are also required, to guarantee the occurrence of suitable error estimates (refer also to [7], [23] and [24]). Without these assumptions, only suboptimal results are derived (see [19] and [20], for instance).

These considerations can also be justified numerically: the tables above reproduce the convergence rates observed when the Upwind Interface Source

	$\|\|e(t)\|\|_{L^1}$		$\|\|e(t)\|\|_{L^2}$		$\|\|e(t)\|\|_{L^\infty}$	
	Error	Rate	Error	Rate	Error	Rate
50	0.323743E-02		0.772868E-02		0.368811E-01	
100	0.816610E-03	1.987	0.270992E-02	1.512	0.184893E-01	0.996
200	0.207343E-03	1.982	0.951254E-03	1.511	0.921217E-02	1.001
400	0.516765E-04	1.990	0.336009E-03	1.508	0.461648E-02	0.999
800	0.128919E-04	1.993	0.118745E-03	1.506	0.231075E-02	0.999
1600	0.321149E-05	1.995	0.419888E-04	1.505	0.115659E-02	0.999

Table 1

	$\|\|e(t)\|\|_{L^1}$		$\|\|e(t)\|\|_{L^2}$		$\|\|e(t)\|\|_{L^\infty}$	
	Error	Rate	Error	Rate	Error	Rate
50	0.129563E-02		0.148671E-02		0.324608E-02	
100	0.326417E-03	1.989	0.368537E-03	2.012	0.814699E-03	1.994
200	0.819300E-04	1.992	0.917495E-04	2.009	0.203996E-03	1.996
400	0.205217E-04	1.993	0.228873E-04	2.007	0.509962E-04	1.997
800	0.513540E-05	1.995	0.571567E-05	2.006	0.127479E-04	1.998
1600	0.128447E-05	1.996	0.142815E-05	2.005	0.318655E-05	1.998

Table 2

method illustrated in this paper is applied to the simplified problem

$$\partial_t u = z'(x), \qquad u(0,x) = u_0(x),$$

with $z(x) = sin(\pi * x)$, $x \in [0,1]$, for which an analytical solution is available to make direct comparisons, $u(t,x) = u_0(x) + z'(x)\, t$.
The results plotted correspond to the discretization (1.16), for the standard *VanLeer limiter*, with a simple TVD reconstruction (see [34]) in Table 1 and with an appropriate ENO reconstruction (see [12]) in Table 2.

The problems just discussed do not arise in the case of discretization (1.17), for which stronger consistency hypotheses are made, to compensate reduced regularity of the reconstructions.

Some classical convergence results for numerical approximations of hyperbolic conservation laws are presented in [28], [37], [11], [6] and its references. Further applications of these methods to different situations are proposed in [3] and [4].

Although the question of preserving stationary states at the discrete level is only handled rigorously for discretizations of the first order (refer to [30]), the numerical results obtained for the Saint-Venant system indicate that the

discretization (1.16) exactly simulates simple equilibria (refer to [16]). As far as we know, similar issues are only addressed in [9] and [21].

References

[1] D. Amadori, L. Gosse, G. Guerra, Global BV entropy solutions and uniqueness for hyperbolic systems of balance laws, *Arch. Ration. Mech. Anal.* (2002), to appear

[2] V.B. Barakhnin, TVD scheme of second-order approximation on a non-stationary adaptive grid for hyperbolic systems, *Russian J. Numer. Anal. Math. Modelling*, **16** (2001), no. 1, 1-17

[3] M. Ben-Artzi, J. Falcovitz, A high-resolution upwind scheme for quasi 1-D flows, Numerical methods for the Euler equations of fluid dynamics (Rocquencourt, 1983), pp. 66-83, *SIAM*, Philadelphia, PA, 1985

[4] M. Ben-Artzi, J. Falcovitz, An upwind second-order scheme for compressible duct flows, *SIAM J. Sci. Statist. Comput.*, **7** (1986), no. 3, 744-768

[5] H. Brezis, *Analyse fonctionnelle. Thorie et applications*, Collection Mathmatiques Appliques pour la Matrise, Masson, Paris, 1983

[6] C. Chainais-Hillairet, S. Champier, Finite volume schemes for nonhomogeneous scalar conservation laws: error estimate, *Numer. Math.*, **88** (2001), no. 4, 607-639

[7] A. Chalabi, On convergence of numerical schemes for hyperbolic conservation laws with stiff source terms, *Math. Comp.*, **66** (1997), no. 218, 527-545

[8] M.G. Crandall, A. Majda, Monotone difference approximations for scalar conservation laws, *Math. Comp.*, **34** (1980), no.149, 1-21

[9] L. Gascón, J.M. Corberán, Construction of second-order TVD schemes for nonhomogeneous hyperbolic conservation laws, *J. Comput. Phys.*, **172** (2001), no. 1, 261-297

[10] E. Godlewski, P.A. Raviart, *Hyperbolic systems of conservation laws*, Mathmatiques & Applications, **3/4**, Ellipses, Paris, 1991

[11] L. Gosse, Sur la stabilit des approximations implicites des lois de conservation scalaires non homognes, *C. R. Acad. Sci. Paris Sr.I Math.*, **329** (1999), no. 1, 79-84

[12] A. Harten, S. Osher, Uniformly high-order accurate nonoscillatory schemes I, *SIAM J. Numer. Anal.*, **24** (1987), no. 2, 279-309

[13] A. Harten, S. Osher, B. Engquist, S.R. Chakravarthy, Some results on uniformly high-order accurate essentially nonoscillatory schemes, *Appl. Numer. Math.*, **2** (1986), no. 3-5, 347-377

[14] A. Harten, B. Engquist, S. Osher, S.R. Chakravarthy, Uniformly high-order accurate essentially nonoscillatory schemes III, *J. Comput. Phys.*, **71** (1987), no. 2, 231-303

[15] M.E. Hubbard, Multidimensional slope limiters for MUSCL-type finite volume schemes on unstructured grids, *J. Comput. Phys.*, **155** (1999), no. 1, 54-74

[16] T. Katsaounis, C. Simeoni, Second order approximation of the viscous Saint-Venant system and comparison with experiments, submitted to *Hyp2002*

[17] S.N. Kružkov, First order quasilinear equations in several independent space variables, *Math. USSR Sb.*, **10** (1970), 217-243

[18] P.D. Lax, Shock waves and entropy, in *Contributions to nonlinear functional analysis*, E.H.Zarantonello editor, New York, Academic Press, 1971, pp. 603-634

[19] A.Y. Le Roux, Convergence d'un schma quasi d'ordre deux pour une quation quasi linaire du premier ordre, *C. R. Acad. Sci. Paris Sr. A-B*, **289** (1979), no. 11, A575-A577

[20] A.Y. Le Roux, Convergence of an accurate scheme for first order quasilinear equations, *RAIRO Anal. Numr.*, **15** (1981), no. 2, 151-170

[21] A.Y. Le Roux, M.N. Le Roux, Convergence d'un schma profils stationnaires pour les quations quasi linaires du premier ordre avec termes sources, *C. R. Acad. Sci. Paris Sr.I Math.*, **333** (2001), no. 7, 703-706

[22] R.J. LeVeque, *Numerical methods for conservation laws*, Lectures in Mathematics ETH Zrich, Birkhuser Verlag, Basel, 1990

[23] D. Levy, G. Puppo, G. Russo, Compact central WENO schemes for multidimensional conservation laws, *SIAM J. Sci. Comput.*, **22** (2000), no. 2, 656-672

[24] D. Levy, G. Puppo, G. Russo, Central WENO schemes for hyperbolic systems of conservation laws, *M2AN Math. Model. Numer. Anal.*, **33** (1999), no. 3, 547-571

[25] M. Louaked, L. Hanich, Un schma TVD-multirsolution pour les quations de Saint-Venant, *C. R. Acad. Sci. Paris Sr.I Math.*, **331** (2000), no. 9, 745-750

[26] H. Nessyahu, E. Tadmor, Nonoscillatory central differencing for hyperbolic conservation laws, *J. Comput. Phys.*, **87** (1990), no. 2, 408-463

[27] S. Osher, P.K. Sweby, Recent developments in the numerical solution of nonlinear conservation laws, The state of the art in numerical analysis (Birmingham, 1986), pp. 681-701, *Inst. Math. Appl. Conf. Ser. New Ser.*, **9**, Oxford Univ. Press, New York, 1987

[28] S. Osher, E. Tadmor, On the convergence of difference approximations to scalar conservation laws, *Math. Comp.*, **50** (1988), no. 181, 19-51

[29] P. de Oliveira, J. Santos, On a class of high resolution methods for solving hyperbolic conservation laws with source terms, *Applied nonlinear analysis*, pp. 403-416, Kluwer/Plenum, New York, 1999

[30] B. Perthame, C. Simeoni, Convergence of the Upwind Interface Source method for hyperbolic conservation laws, submitted to *Hyp2002*

[31] R. Sanders, On convergence of monotone finite difference schemes with variable spatial differencing, *Math. Comp.*, **40** (1983), no.161, 91-106

[32] C.W. Shu, Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws, Advanced numerical approximation of nonlinear hyperbolic equations (Cetraro, 1997), pp. 325-432, *Lecture Notes in Math.*, **1697**, Springer, Berlin, 1998

[33] B. van Leer, Towards the ultimate conservative difference scheme V. A second-order sequel to Godunov's method, *J. Comput. Phys.*, **32** (1979), no. 1, 101-136

[34] P.K. Sweby, TVD schemes for inhomogeneous conservation laws, Nonlinear hyperbolic equations: theory, computation methods and applications (Aachen, 1988), pp. 599-607, *Notes Numer. Fluid Mech.*, **24**, Vieweg, Braunschweig, 1989

[35] A. Vasseur A., Well-posedness of scalar conservation laws with singular sources, preprint

[36] J.P. Vila, An analysis of a class of second-order accurate Godunov-type schemes, *SIAM J. Numer. Anal.*, **26** (1989), no. 4, 830-853

[37] J.P. Vila, High-order schemes and entropy condition for nonlinear hyperbolic systems of conservation laws, *Math. Comp.*, **50** (1988), no. 181, 53-73

Deuxième partie

L'APPROXIMATION NUMÉRIQUE DES ÉQUATIONS DE SAINT-VENANT ET APPLICATIONS AUX ÉTUDES EXPÉRIMENTALES

Chapitre 3

Un schéma cinétique pour le système de Saint-Venant avec un terme source

(publié dans *Calcolo*, **38** (2001), no. 4, 201-231)

A kinetic scheme for the Saint-Venant system with a source term

B. Perthame, C. Simeoni

Département de Mathématiques et Applications
École Normale Supérieure
45, rue d'Ulm - 75230 Paris Cedex 05 - France
e-mails: Benoit.Perthame@ens.fr, Chiara.Simeoni@ens.fr

Abstract

The aim of this paper is to present a numerical scheme to compute Saint-Venant equations with a source term, due to the bottom topography, in a one-dimensional framework, which satisfies the following theoretical properties: it preserves the steady state of still water, satisfies an entropy inequality, preserves the non-negativity of the height of water and remains stable with a discontinuous bottom. This is achieved by means of a kinetic approach to the system, which is the departing point of the method developed here. In this context, we use a natural description of the microscopic behaviour of the system to define numerical fluxes at the interfaces of an unstructured mesh. We also use the concept of cell-centered conservative quantities (as usual in the finite volume method) and upwind interfacial sources as advocated by several authors. We show, analytically and also by means of numerical results, that the above properties are satisfied.

Key-words: Saint-Venant system, finite volume method, upwind interfacial sources, kinetic schemes.

1 Introduction

The Saint-Venant equations, a particular case of shallow water equations, are commonly used to describe physical situations such as flows in rivers or coastal areas. The one-dimensional version is well adapted for ideal rectangular rivers. It allows to describe the flow, at time $t \geq 0$ and at point $x \in \mathbb{R}$,

through the height of water $h(t,x) \geq 0$ and its velocity $u(t,x) \in \mathbb{R}$, by the hyperbolic system

$$\frac{\partial h}{\partial t} + \frac{\partial (hu)}{\partial x} = 0, \tag{1.1}$$

$$\frac{\partial (hu)}{\partial t} + \frac{\partial}{\partial x}(hu^2 + \frac{g}{2}h^2) + gh\frac{\partial Z}{\partial x} = 0, \tag{1.2}$$

where g denotes the gravity intensity and $Z(x)$ is the bottom height; therefore $h + Z$ is the level of the water surface (in what follows, we also denote the discharge by $q = hu$).

These equations were originally written by A. de Saint-Venant in [21] and more complete systems can be derived from the Navier-Stokes equations (see [8] and its references). In fact, the system (1.1)-(1.2) corresponds to a particularly simple case: other terms can be added to the right-hand side in order to take into account frictions on the bottom and the surface; a more general system can also be stated for rivers with variable sections.

The bottom topography introduces a source term in equation (1.2), influencing the unknowns of the problem. Hence, analytical properties of the system of isentropic Euler equations are deeply modified in comparison with the homogeneous model. For instance, a well-known problem is the occurrence of different kinds of steady states.

Several methods for solving hyperbolic systems of conservation laws with source terms have been investigated. The main problem is related to the approximation of such a source term, to assure the numerical preservation of properties fulfilled by the continuous model. A classical approach consists in using finite volume schemes (refer to [9], [16] and [6]), the finite volume method displaying the remarkable property of being water height conservative. But of course other methods are possible, such as finite elements (see [2] and the references therein).

A difficulty arising specifically with the Saint-Venant system is that of forcing the scheme to preserve steady states given by a lake at rest ($u = 0$, $h + Z = C^{st}$), as has been pointed out by several authors. To treat these particular problems, a specific numerical approach is needed. Here it is based on two concepts: first, the conservative quantities are cell-centered, as usual for finite volume schemes; second, as introduced by Roe [20], the source terms are upwinded at the cell interfaces. Initially for scalar conservation laws, Greenberg and LeRoux [14], then Gosse and LeRoux [10],[11] introduced the notion of well-balanced scheme and they use a reformulation of the source terms by means of non-conservative products to derive their numerical fluxes; this kind of numerical processing has been recently extended by Gosse [12],[13] to hyperbolic systems of balance laws. A kinetic scheme, which introduces

the notion of reflexions on bottom jumps and maintains steady states, is presented in Botchorishvili, Perthame and Vasseur [4] , which is proved to be convergent. Still using the interface values, instead of the cell-averages, for the source terms, Jin proposes in [15] a rather simple method for capturing the steady state solutions with a high order accuracy. Quite recently, various approaches appeared to build stable schemes which preserve the steady states: previous schemes have been modified for this purpose by Bermudez and Vasquez [3]; the Godunov scheme for an appropriately extended system is developed in [17]; an appropriate linearized Godunov scheme preserving all steady states has also been obtained by Gallouët, Hérard and Seguin [7].

But to our knowledge none of these methods are proved to satisfy all the stability properties: (i) water height remains nonnegative, (ii) the energy (entropy) inequality is satisfied, (iii) it preserves the steady states of still water. Certainly, Godunov schemes as modified in [14] or [7] can do that, at the expense of a fixed point, but this has not been proved.

In this paper, we consider a particular class of numerical schemes to compute the Saint-Venant equations, based on the kinetic interpretation of the system, which is presented in [1]. These kinetic schemes have many good properties that other solvers have difficulty in achieving; in particular, they are able to treat the case of a vacuum ($h = 0$ here, corresponding to dry soils, when the system loses hyperbolicity) and satisfy the properties (i), (ii), (iii) above. We refer to Perthame [19] for a survey of the theoretical properties of these schemes. We only note that kinetic schemes are a simple way to generate efficient building blocks (interpolations at the interfaces) in finite volume methods; they do not involve rarefied flows except for the technique used in proving their theoretical properties.
We present a numerical scheme for the system (1.1)-(1.2) by using a new kinetic solver, which exhibits all the advantages of this specific approach. We propose a way to take the source term directly into account in the definition of the numerical fluxes, whose structure relies on a natural description of the microscopic behaviour of the system through a potential barrier which is the bottom $Z(x)$.

The paper is organized in four sections. In the second section, we recall some properties of the Saint-Venant equations and we explain the kinetic approach to this system. In the third, we illustrate the kinetic scheme "with reflexions" and we demonstrate the properties (i), (ii), (iii). Several test problems for the flat and non-flat bottom cases are reported in the last section. We leave the extension of this method to higher order accuracy for a future work. Implementation in two spatial dimensions is also in progress (see [1] for a first attempt).

2 Preliminaries about the Saint-Venant equations

We recall here some well-known properties of the shallow water system. We take them into account to develop our numerical method so as to be coherent with the physical model. Then, by analogy with the Euler equations of compressible gas dynamics, we link the macroscopic Saint-Venant system to a microscopic description of the fluid, on which the method proposed in this paper is based.

2.1 Properties of the system

First of all, the system is naturally posed for $h(t,x) \geq 0$ and the water height h can indeed vanish (flooding zones, dry soils, tidal flats); this fact leads to a theoretical and numerical difficulty, because the system loses hyperbolicity at $h = 0$.

Another fundamental property is related to the *entropy inequality* of the Saint-Venant system, satisfied by the weak solutions, defined as in the following theorem.

Theorem 2.1. *The system (1.1)-(1.2) is strictly hyperbolic for $h > 0$. It admits a mathematical entropy, which is also the physical energy,*

$$E(h,u,Z) = h\frac{u^2}{2} + \frac{g}{2}h^2 + gZh, \tag{2.1}$$

which satisfies the "entropy inequality"

$$\frac{\partial E}{\partial t} + \frac{\partial}{\partial x}[u(E + \frac{g}{2}h^2)] \leq 0. \tag{2.2}$$

We do not prove this theorem, which relies on the classical theory of hyperbolic equations (see Serre [22], Dafermos [5]) and simple algebraic calculations. We remark only that for smooth solutions the inequality (2.2) becomes an equality.

Also, the system admits a family of smooth steady states characterized by the relations

$$hu = C_1, \tag{2.3}$$

$$\frac{u^2}{2} + g(h+Z) = C_2, \tag{2.4}$$

where C_1 and C_2 are two arbitrary constants. In particular, the simplest is the traditional steady state of a lake at rest, given by $u = 0$, $h + Z = C^{st}$.

2.2 Kinetic approach

We pass to explain how it is possible to introduce a kinetic approach for the Saint-Venant system. We describe it for the one-dimensional problem, but the construction is similar in two dimensions.

We consider a real function χ defined on $\mathbb{R}$, with the following properties,

$$\chi(\omega) = \chi(-\omega) \geq 0, \quad \int_{\mathbb{R}} \chi(\omega) d\omega = 1, \quad \int_{\mathbb{R}} \omega^2 \chi(\omega) d\omega = \frac{g}{2} \tag{2.5}$$

and define the density of particles $M(t, x, \xi)$ by a so-called *Gibbs equilibrium*

$$M(t, x, \xi) = M(h, \xi - u) = \sqrt{h(t,x)} \chi \left(\frac{\xi - u(t,x)}{\sqrt{h(t,x)}} \right). \tag{2.6}$$

These definitions allow to obtain a kinetic representation of the system.

Theorem 2.2. *The pair of functions (h, hu) is a strong solution of the Saint-Venant system (1.1)-(1.2) if and only if $M(h, \xi - u)$ satisfies the kinetic equation*

$$\frac{\partial M}{\partial t} + \xi \cdot \frac{\partial M}{\partial x} - g \frac{\partial Z}{\partial x} \cdot \frac{\partial M}{\partial \xi} = Q(t, x, \xi), \tag{2.7}$$

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

$$\int_{\mathbb{R}} Q d\xi = 0, \qquad \int_{\mathbb{R}} \xi Q d\xi = 0. \tag{2.8}$$

Proof. The proof relies on a very obvious computation. The two Saint-Venant equations are obtained by taking the moments of the kinetic equation (2.7) in $d\xi$, against 1, ξ and ξ^2 respectively: the right-hand side vanishes according to (2.8) and the left-hand sides coincide exactly thanks to hypothesis (2.5). These are consequences of the easy relations,

$$h = \int_{\mathbb{R}} M(h, \xi - u) d\xi, \tag{2.9}$$

$$hu = \int_{\mathbb{R}} \xi M(h, \xi - u) d\xi, \tag{2.10}$$

$$hu^2 + \frac{g}{2} h^2 = \int_{\mathbb{R}} \xi^2 M(h, \xi - u) d\xi, \tag{2.11}$$

directly obtained from the microscopic equilibrium (2.6). □

This theorem produces a very useful consequence: the non-linear shallow water system can be viewed as a single linear equation on a non-linear quantity M, for which it is easier to find simple numerical schemes with good theoretical properties.
We note that this form is much weaker than the *kinetic formulation* proposed by Lions, Perthame and Tadmor in [18], which represents all the entropies of the system.

We characterize the function χ which defines the density of particles $M(t,x,\xi)$ in the kinetic approach; in particular, we justify the interpretation of such a density as the microscopic equilibrium of the system, the *Gibbs equilibrium*. These facts are stated in the following propositions.

Lemma 2.3. *The minimum of the energy*

$$\mathcal{E}(f) = \int_{\mathbb{R}} \left[\frac{\xi^2}{2} f(\xi) + \frac{\pi^2 g^2}{6} f^3(\xi) + gZf(\xi) \right] d\xi, \tag{2.12}$$

under the constraints

$$f \geq 0, \quad \int_{\mathbb{R}} f(\xi) d\xi = h, \quad \int_{\mathbb{R}} \xi f(\xi) d\xi = hu,$$

is attained by the function $M(h, \xi - u) = \sqrt{h}\chi\left(\frac{\xi-u}{\sqrt{h}}\right)$, *with* χ *defined by*

$$\chi(\omega) = \frac{\sqrt{2}}{\pi\sqrt{g}} \left(1 - \frac{\omega^2}{2g}\right)_+^{\frac{1}{2}}. \tag{2.13}$$

Remark 2.4. The cubic term in the functional (2.12) takes the internal energy into account. In one dimension, it results from the transverse translational energy. Indeed the corresponding two-dimensional variational problem, for $Z = 0$, gives

$$\frac{1}{2}h(u^2+v^2) + \frac{g}{2}h^2 = \min \left\{ \int_{\mathbb{R}^2} \frac{|\xi|^2}{2} f(\xi) d\xi; \right.$$
$$\left. f \geq 0, \int_{\mathbb{R}^2} f(\xi) d\xi = h, \int_{\mathbb{R}^2} \xi f(\xi) d\xi = (hu, hv) \right\}$$

and we deduce $f(\xi_1) = \int_{\mathbb{R}} g(\xi_1, \xi_2)\, d\xi_2$.

Proof. Because of the constraints, it is sufficient to minimize the functional

$$\mathcal{E}_0(f) = \int_{\mathbb{R}} \left[\xi^2 f(\xi) + \frac{\pi^2 g^2}{3} f^3(\xi) \right] d\xi.$$

Since $\mathcal{E}_0(f)$ is a convex functional, the formula for M (and thus for χ) follows directly from the Euler-Lagrange equation associated to the minimization problem, for $f > 0$,

$$\xi^2 + \pi^2 g^2 f^2 = \lambda + \mu\xi,$$

where $\lambda(h,u)$ and $\mu(h,u)$ are Lagrange multipliers. One readily checks by convexity that it is a strict minimizer. □

Recalling the formula (2.1), we see that the minimum considered in Lemma 2.3 is given by

$$\mathcal{E}(M(h,\xi-u)) = E(h,u,Z),$$

again an immediate consequence of the relations stated in (2.9)-(2.11) and by the choice of the specific value $\frac{\pi^2 g^2}{6}$ in the energy. Hence, the properties of the function χ are consistent with the kinetic approach to the system, as introduced above.

We conclude this section by pointing out another motivation for the choice of χ in Lemma 2.3.

Lemma 2.5. *The function* $\chi(\omega) = \frac{\sqrt{2}}{\pi\sqrt{g}}\left(1 - \frac{\omega^2}{2g}\right)_+^{\frac{1}{2}}$ *is the only choice such that* $M(h,\xi-u) = \sqrt{h}\chi\left(\frac{\xi-u}{\sqrt{h}}\right)$ *satisfies the equation*

$$\xi \cdot \frac{\partial M}{\partial x} - g\frac{\partial Z}{\partial x} \cdot \frac{\partial M}{\partial \xi} = 0 \tag{2.14}$$

on all steady states given by a lake at rest,

$$u(t,x) = 0, \quad h(t,x) + Z(x) = H, \quad \forall t \geq 0.$$

Proof. Exploiting the hypotheses, we compute

$$\begin{aligned}
\frac{\partial M}{\partial x} &= \frac{1}{2\sqrt{h}}\frac{\partial h}{\partial x}\left[\chi\left(\frac{\xi}{\sqrt{h}}\right) - \frac{\xi}{\sqrt{h}}\chi'\left(\frac{\xi}{\sqrt{h}}\right)\right], \\
\frac{\partial Z}{\partial x} &= -\frac{\partial h}{\partial x}, \quad \frac{\partial M}{\partial \xi} = \chi'\left(\frac{\xi}{\sqrt{h}}\right),
\end{aligned}$$

so that the equation (2.14) becomes

$$\frac{\xi}{2\sqrt{h}}\frac{\partial h}{\partial x}\chi\left(\frac{\xi}{\sqrt{h}}\right) - \frac{\xi^2}{2h}\frac{\partial h}{\partial x}\chi'\left(\frac{\xi}{\sqrt{h}}\right) + g\frac{\partial h}{\partial x}\chi'\left(\frac{\xi}{\sqrt{h}}\right) = 0.$$

Let $\omega = \frac{\xi}{\sqrt{h}}$, the last relation can be rewritten as

$$\frac{1}{2}\frac{\partial h}{\partial x}\left[\omega\chi(\omega) + (2g - \omega^2)\chi'(\omega)\right] = 0.$$

A characterization for χ is therefore given by the equation

$$\omega\chi(\omega) + (2g - \omega^2)\chi'(\omega) = 0,$$

that admits, under the constraints (2.5), the unique solution

$$\chi(\omega) = \left(2g - \omega^2\right)_+^{\frac{1}{2}}.$$

□

Remark 2.6. *We note that the difficulty in preserving such steady states at the kinetic level might explain why the Maxwellian case does not work well (see Xu [24]).*

3 The kinetic scheme with reflections

We present a finite volume scheme for the one-dimensional Saint-Venant system, based on the kinetic approach described in Section 2, which has the property of preserving the steady state of a lake at rest. Also, it preserves the stability properties of the usual kinetic solvers and satisfies a precise in-cell entropy inequality.

3.1 The formulas

We consider a uniform mesh of $\mathbb{R}$, whose vertices are denoted x_i, $i \in \mathbb{Z}$. Let $C_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$ be the control volume (cell), with $x_{i+\frac{1}{2}} = \frac{x_{i+1}+x_i}{2}$, and we denote the space-step by $\Delta x = length(C_i)$, so that $x_i = i\Delta x$, $i \in \mathbb{Z}$. We also consider a discretization in time by introducing a time-step Δt and we set $t_n = n\Delta t$, $n \in \mathbb{N}$.
If $Z(x)$ is the function describing the bottom height, its piecewise constant representation is given by $\bar{Z}(x) = Z_i \mathbb{1}_{C_i}(x)$, with $Z_i = \frac{1}{\Delta x}\int_{C_i} Z(x)dx$, for example.

We start from the microscopic equation (2.7), to perform a discretization directly on the density of particles

$$f_i^{n+1}(\xi) - M_i^n(\xi) + \frac{\Delta t}{\Delta x}\xi\left(M^-_{i+\frac{1}{2}}(\xi) - M^+_{i-\frac{1}{2}}(\xi)\right) = 0, \tag{3.1}$$

where the interface equilibrium densities $M^{\pm}_{i+\frac{1}{2}}$ are defined later. As usual, the "collision term" $Q(t,x,\xi)$ in the kinetic representation (2.7) of Saint-Venant equations, which relaxes the kinetic density to an equilibrium M, is neglected in the numerical scheme; at each time-step we project $f^n_i(\xi)$ on $M^n_i(\xi)$, which is a way of performing all collisions at once and to recover the *Gibbs equilibrium* without computing it.

Note that the fluxes can also be written as

$$M^-_{i+\frac{1}{2}}(\xi) = M_{i+\frac{1}{2}}(\xi) + \left(M^-_{i+\frac{1}{2}}(\xi) - M_{i+\frac{1}{2}}(\xi)\right)$$

and the quantity $\delta M^-_{i+\frac{1}{2}}(\xi) = M^-_{i+\frac{1}{2}}(\xi) - M_{i+\frac{1}{2}}(\xi)$ holds for the discrete contribution of the force term $h\frac{\partial Z}{\partial x}$ in the system, for negative velocities; indeed, $\delta M^-_{i+\frac{1}{2}}(\xi) = 0$ for $\xi \geq 0$ in the scheme to be presented below. This is the principle of the Interfacial Upwind Sources method: the source is not treated as a volumic term but at the interfaces and it is upwinded.

Now, we integrate the equation (3.1) in $d\xi$ against 1 and ξ, with notation

$$U^{n+1}_i = (h^{n+1}_i, (hu)^{n+1}_i), \tag{3.2}$$

$$h^{n+1}_i = \int_{\mathbb{R}} f^{n+1}_i(\xi)d\xi, \quad (hu)^{n+1}_i = \int_{\mathbb{R}} \xi f^{n+1}_i(\xi)d\xi \tag{3.3}$$

and we obtain the macroscopic scheme

$$U^{n+1}_i - U^n_i + \frac{\Delta t}{\Delta x}\left[\mathbb{F}^-_{i+\frac{1}{2}} - \mathbb{F}^+_{i-\frac{1}{2}}\right] = 0. \tag{3.4}$$

The numerical fluxes are thus given by the kinetic fluxes

$$\mathbb{F}^-_{i+\frac{1}{2}} = \int_{\mathbb{R}} \xi \begin{pmatrix} 1 \\ \xi \end{pmatrix} M^-_{i+\frac{1}{2}}(\xi)d\xi, \tag{3.5}$$

$$\mathbb{F}^+_{i-\frac{1}{2}} = \int_{\mathbb{R}} \xi \begin{pmatrix} 1 \\ \xi \end{pmatrix} M^+_{i-\frac{1}{2}}(\xi)d\xi. \tag{3.6}$$

In order to take the neighboring cells into account by means of a natural interpretation of the microscopic features of the system, we formulate a peculiar discretization for the fluxes in (3.1), computed by the upwind formulas

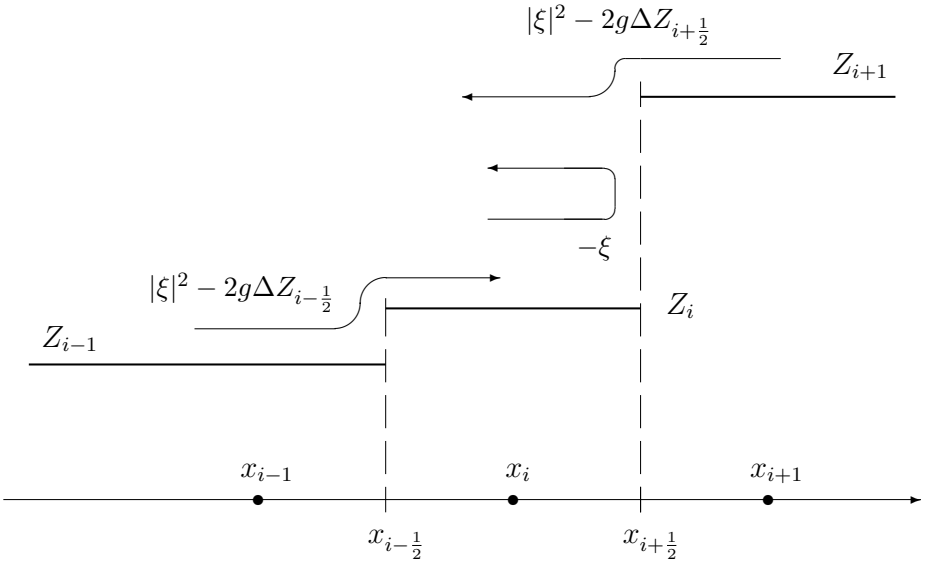
$$M^-_{i+\frac{1}{2}}(\xi) = M^n_i(\xi)\mathbb{1}_{\xi\geq 0} + M^n_{i+\frac{1}{2}}(\xi)\mathbb{1}_{\xi\leq 0}, \tag{3.7}$$

$$M^+_{i-\frac{1}{2}}(\xi) = M^n_{i-\frac{1}{2}}(\xi)\mathbb{1}_{\xi\geq 0} + M^n_i(\xi)\mathbb{1}_{\xi\leq 0}, \tag{3.8}$$

where we define

$$M^n_{i+\frac{1}{2}}(\xi) = M^n_i(-\xi)\mathbb{1}_{|\xi|^2\leq 2g\Delta Z_{i+\frac{1}{2}}} + M^n_{i+1}\left(-\sqrt{|\xi|^2 - 2g\Delta Z_{i+\frac{1}{2}}}\right)\mathbb{1}_{|\xi|^2\geq 2g\Delta Z_{i+\frac{1}{2}}},$$
$$M^n_{i-\frac{1}{2}}(\xi) = M^n_i(-\xi)\mathbb{1}_{|\xi|^2\leq 2g\Delta Z_{i-\frac{1}{2}}} + M^n_{i-1}\left(\sqrt{|\xi|^2 - 2g\Delta Z_{i-\frac{1}{2}}}\right)\mathbb{1}_{|\xi|^2\geq 2g\Delta Z_{i-\frac{1}{2}}}.$$

The figure below illustrates the typical situation occurring in a cell C_i of the mesh, centered at the point $x_i \in \mathbb{R}$; without loss of generality, we consider here the case of an increasing bottom slope, so that the bottom jumps are positive and negative for the neighboring cells C_{i-1} and C_{i+1} respectively.



The effect of the source term is made explicit by treating it as a physical potential. The definitions (3.7)-(3.8) are thus a mathematical formalization to describe the physical microscopic behaviour of the system: contributions to the value $f_i^{n+1}(\xi)$ are also given by particles in C_{i+1} and in C_{i-1} at time t_n, with kinetic energy sufficient to surpass the potential difference (speeded up or down through the potential jump) and by particles coming at velocity $-\xi$, reflected on the bottom jumps according to classical mechanics, when their energy is too small (i.e. $|\xi|^2 \leq 2g\Delta Z_{i+\frac{1}{2}}$).

Remark 3.1. We see immediately that *the kinetic scheme (3.4)-(3.6) is water height conservative*. In fact, still referring to the figure, we compute the first component of the numerical fluxes at the interface $x_{i+\frac{1}{2}}$ of the mesh

by the formulas (3.5)-(3.6),

$$(\mathbb{F}_h)^-_{i+\frac{1}{2}} = \int_{\xi\geq 0} \xi M_i^n(\xi)d\xi + \int_{\xi\leq 0} \xi M_i^n(-\xi)\mathbb{1}_{|\xi|^2\leq 2g\Delta Z_{i+\frac{1}{2}}}d\xi + \int_{\xi\leq 0} \xi M_{i+1}^n\left(-\sqrt{|\xi|^2-2g\Delta Z_{i+\frac{1}{2}}}\right)\mathbb{1}_{|\xi|^2\geq 2g\Delta Z_{i+\frac{1}{2}}}d\xi$$

and

$$(\mathbb{F}_h)^+_{i+\frac{1}{2}} = \int_{\xi\leq 0} \xi M_{i+1}^n(\xi)d\xi + \int_{\xi\geq 0} \xi M_i^n\left(\sqrt{|\xi|^2-2g\Delta Z_{i+\frac{1}{2}}}\right)d\xi,$$

so that a simple change of variable $|\xi'|^2 = |\xi|^2 - 2g\Delta Z_{i+\frac{1}{2}}$, $\xi' d\xi' = \xi d\xi$ allows to conclude that

$$(\mathbb{F}_h)^-_{i+\frac{1}{2}} = (\mathbb{F}_h)^+_{i+\frac{1}{2}}, \quad \forall i \in \mathbb{Z}.$$

The conservation of water height and momentum is also obvious for the system with a flat bottom: the continuous system (1.1)-(1.2) becomes homogeneous ($\frac{\partial Z}{\partial x} = 0$) and we obtain a conservative scheme, with the flux-splitting form of the standard kinetic scheme.
We emphasize that to show the property of the numerical scheme (3.4)-(3.6) to be consistent cannot be achieved in the classical manner. Because of the presence of the source term and the choice to process it implicitly, this question is much more delicate here.

3.2 Properties of the numerical scheme

We establish some theoretical properties of the numerical scheme introduced in the previous subsection, which represent the discrete analogue of the main properties of the Saint-Venant system stated in Section 2.

Theorem 3.2. *We assume the CFL condition*

$$\Delta t \ max\left(|u_i^n| + \sqrt{2gh_i^n}\right) \leq \Delta x. \tag{3.9}$$

Then, (i) the kinetic scheme (3.4)-(3.6) keeps the water height positive, i.e. $h_i^n \geq 0$ if this is the case initially; (ii) it satisfies the conservative in-cell entropy inequality,

$$E_i^{n+1} - E_i^n + \frac{\Delta t}{\Delta x}\left[\eta_{i+\frac{1}{2}}^n - \eta_{i-\frac{1}{2}}^n\right] \leq 0,$$

with the discrete entropy fluxes given in the formulas (3.11)-(3.12) below and the discrete energy

$$E_i^n = h_i^n\frac{|u_i^n|^2}{2} + \frac{g}{2}(h_i^n)^2 + gZ_ih_i^n;$$

(iii) the scheme (3.4)-(3.6) preserves the steady states of the system given by a lake at rest,

$$u_i^n = 0, \quad h_i^n + Z_i = H, \qquad \forall i \in \mathbb{Z}, \quad \forall n \in \mathbb{N}.$$

Proof. To prove the first stability property (i) of the scheme, we come back to the kinetic interpretation and we proceed by induction. We assume that $h_i^n \geq 0$, $\forall i \in \mathbb{Z}$, and we prove that $h_i^{n+1} \geq 0$, $\forall i \in \mathbb{Z}$; since

$$h_i^{n+1} = \int_{\mathbb{R}} f_i^{n+1}(\xi) d\xi,$$

it is sufficient to prove that $f_i^{n+1}(\xi) \geq 0$. We introduce the quantities

$$\xi_+ = max(0, \xi), \quad \xi_- = max(-\xi, 0), \quad \sigma = \frac{\Delta t}{\Delta x},$$

so that we can rewrite the microscopic scheme (3.1), (3.7)-(3.8) in the form

$$\begin{aligned} f_i^{n+1}(\xi) &= M_i^n(\xi) - \sigma\xi \left(M^-_{i+\frac{1}{2}}(\xi) - M^+_{i-\frac{1}{2}}(\xi) \right) \\ &= (1 - \sigma|\xi|)\, M_i^n(\xi) + \sigma\xi_- \Big[M_i^n(-\xi) \mathbb{1}_{|\xi|^2 \leq 2g\Delta Z_{i+\frac{1}{2}}} \\ &\qquad + M_{i+1}^n \left(-\sqrt{|\xi|^2 - 2g\Delta Z_{i+\frac{1}{2}}} \right) \mathbb{1}_{|\xi|^2 \geq 2g\Delta Z_{i+\frac{1}{2}}} \Big] \qquad (3.10) \\ &+ \sigma\xi_+ \Big[M_i^n(-\xi) \mathbb{1}_{|\xi|^2 \leq 2g\Delta Z_{i-\frac{1}{2}}} \\ &\qquad + M_{i-1}^n \left(\sqrt{|\xi|^2 - 2g\Delta Z_{i-\frac{1}{2}}} \right) \mathbb{1}_{|\xi|^2 \geq 2g\Delta Z_{i-\frac{1}{2}}} \Big]. \end{aligned}$$

Since the function χ has a compact support, it follows that

$$M_j^n(\xi) = 0 \qquad \text{if} \quad |\xi - u_j^n| \geq \sqrt{2gh_j^n};$$

we deduce that

$$f_i^{n+1}(\xi) \geq 0 \qquad \text{if} \quad |\xi - u_j^n| \geq \sqrt{2gh_j^n}, \quad \forall j \in \mathbb{Z},$$

as a sum of non-negative quantities.
Now, if $M_i^n(\xi) \neq 0$ then $|\xi - u_i^n| \leq \sqrt{2gh_i^n}$ and

$$|\xi| \leq |\xi - u_i^n| + |u_i^n| \leq \sqrt{2gh_i^n} + |u_i^n|.$$

We use the CFL condition to conclude that $\sigma|\xi| \leq 1$, so that $f_i^{n+1}(\xi)$ is a convex combination of three non-negative quantities and thus $f_i^{n+1}(\xi) \geq 0$,

$\forall \xi \in \mathbb{R}$, $\forall i \in \mathbb{Z}$.
For the entropy inequality (ii), the conclusion results from the relation (3.10), which describes $f_i^{n+1}(\xi)$ as a convex combination of density functions. Recalling the definition (2.12) for the energy convex functional $\mathcal{E}(f)$, we calculate it on the previous convex formula and, thanks to the relations stated in the proof of Theorem 2.2, we obtain

$$\mathcal{E}(f_i^{n+1}) - E_i^n + \frac{\Delta t}{\Delta x}\left[\eta_{i+\frac{1}{2}}^n - \eta_{i-\frac{1}{2}}^n\right] \leq 0,$$

where the entropy fluxes have the expressions

$$\begin{aligned}\eta_{i+\frac{1}{2}}^n = &\int_{\xi\geq 0}\left[\frac{\xi^3}{2}M_i^n(\xi) + \frac{\pi^2 g^2}{6}\xi\left[M_i^n(\xi)\right]^3 + gZ_i\,\xi M_i^n(\xi)\right]d\xi \\ &+ \int_{\xi\leq 0}\left[\frac{\xi^3}{2}M_{i+\frac{1}{2}}^n(\xi) + \frac{\pi^2 g^2}{6}\xi\left[M_{i+\frac{1}{2}}^n(\xi)\right]^3 + gZ_i\,\xi M_{i+\frac{1}{2}}^n(\xi)\right]d\xi,\end{aligned} \tag{3.11}$$

$$\begin{aligned}\eta_{i-\frac{1}{2}}^n = &\int_{\xi\geq 0}\left[\frac{\xi^3}{2}M_{i-\frac{1}{2}}^n(\xi) + \frac{\pi^2 g^2}{6}\xi\left[M_{i-\frac{1}{2}}^n(\xi)\right]^3 + gZ_i\,\xi M_{i-\frac{1}{2}}^n(\xi)\right]d\xi \\ &+ \int_{\xi\leq 0}\left[\frac{\xi^3}{2}M_i^n(\xi) + \frac{\pi^2 g^2}{6}\xi\left[M_i^n(\xi)\right]^3 + gZ_i\,\xi M_i^n(\xi)\right]d\xi.\end{aligned} \tag{3.12}$$

Next, we use Lemma 2.3 to deduce

$$E_i^{n+1} = \mathcal{E}(M_i^{n+1}) \leq \mathcal{E}(f_i^{n+1}).$$

Finally, we can again give a direct proof of the last statement (iii) at the microscopic level. We emphasize that this approach is also justified by the result stated in Lemma 2.5. From the formula (3.1) for the numerical scheme, it is enough to prove that

$$M_{i+\frac{1}{2}}^-(\xi) = M_{i-\frac{1}{2}}^+(\xi), \quad \forall \xi \in \mathbb{R}$$

and (iii) follows: indeed, this implies $f_i^{n+1}(\xi) = M_i^n(\xi)$, which also gives $h_i^{n+1} = h_i^n$, $u_i^{n+1} = u_i^n$, $\forall i \in \mathbb{Z}$.
According to the definition (3.7)-(3.8), we can distinguish two cases of the previous equality, for $\xi \geq 0$ and $\xi \leq 0$; since these cases present the same difficulty, we only consider the case $\xi \geq 0$. We also remark that, exploiting the hypothesis $u_i = 0$, we have

$$M_i^n(\xi) = \frac{\sqrt{2}}{\pi\sqrt{g}}\sqrt{h_i^n}\left(1 - \frac{\xi^2}{2gh_i^n}\right)^{\frac{1}{2}}$$

and

$$M^n_{i\pm1}\left(\mp\sqrt{|\xi|^2-2g\Delta Z_{i\pm\frac{1}{2}}}\right)=\frac{\sqrt{2}}{\pi\sqrt{g}}\sqrt{h^n_{i\pm1}}\left(1-\frac{\xi^2-2g\Delta Z_{i\pm\frac{1}{2}}}{2gh^n_{i\pm1}}\right)^{\frac{1}{2}}.$$

Next, for the case $\xi\geq 0$ and $|\xi|^2\leq 2g\Delta Z_{i-\frac{1}{2}}$, the result is obvious. There remains the case $\xi\geq 0$ and $|\xi|^2\geq 2g\Delta Z_{i-\frac{1}{2}}$, for which the statement reduces to verifying the equality

$$\sqrt{h^n_i}\left(1-\frac{\xi^2}{2gh^n_i}\right)^{\frac{1}{2}}=\sqrt{h^n_{i-1}}\left(1-\frac{\xi^2-2g\Delta Z_{i-\frac{1}{2}}}{2gh^n_{i-1}}\right)^{\frac{1}{2}}.$$

Thanks to the hypothesis $h^n_i+Z_i=h_{i-1}+Z_{i-1}$, $\forall i\in\mathbb{Z}$, it follows that $\Delta Z_{i-\frac{1}{2}}=h^n_i-h^n_{i-1}$, so that a simple algebraic computation completes the proof. □

4 Numerical implementation

To proceed to the actual implementation of the scheme (3.4)-(3.6), we have to compute the numerical fluxes explicitly. Since their expressions are not always immediate to calculate, it is necessary to use an approximation technique for some of them; in this section we indicate some fundamental properties and we give a possible appropriate approximation.

4.1 Computation of the integrals

According to the kinetic representation of the Saint-Venant system, the density of particles $M^n_i(\xi)$ in the formulas (3.7)-(3.8) is defined by

$$M^n_i(\xi)=\sqrt{h^n_i}\chi\left(\frac{\xi-u^n_i}{\sqrt{h^n_i}}\right),$$

which represents the discrete analogue of the microscopic *Gibbs equilibrium* considered in Section 2. With this definition, the formula (3.5) becomes

$$\begin{aligned}\mathbb{F}^-_{i+\frac{1}{2}}&=\int_{\xi\geq0}\xi\begin{pmatrix}1\\ \xi\end{pmatrix}M^n_i(\xi)d\xi+\int_{\xi\leq0}\xi\begin{pmatrix}1\\ \xi\end{pmatrix}M^n_{i+\frac{1}{2}}(\xi)d\xi\\&=\sqrt{h^n_i}\int_{\xi\geq0}\xi\begin{pmatrix}1\\ \xi\end{pmatrix}\chi\left(\frac{\xi-u^n_i}{\sqrt{h^n_i}}\right)d\xi\\&+\sqrt{h^n_i}\int_{\xi\leq0}\xi\begin{pmatrix}1\\ \xi\end{pmatrix}\chi\left(\frac{-\xi-u^n_i}{\sqrt{h^n_i}}\right)\mathbb{1}_{|\xi|^2\leq2g\Delta Z_{i+\frac{1}{2}}}d\xi\end{aligned}$$

$$+\sqrt{h_{i+1}^n}\int_{\xi\le 0}\xi\begin{pmatrix}1\\ \xi\end{pmatrix}\chi\left(\frac{-\sqrt{|\xi|^2-2g\Delta Z_{i+\frac12}}-u_{i+1}^n}{\sqrt{h_{i+1}^n}}\right)\mathbb{1}_{|\xi|^2\ge 2g\Delta Z_{i+\frac12}}d\xi.$$

We consider a change of variable $\xi'=-\sqrt{|\xi|^2-2g\Delta Z_{i+\frac12}}$, $\xi' d\xi'=\xi d\xi$, in the third term to obtain

$$\begin{aligned}&\int_{\xi\le 0}\xi\begin{pmatrix}1\\ \xi\end{pmatrix}\chi\left(\frac{-\sqrt{|\xi|^2-2g\Delta Z_{i+\frac12}}-u_{i+1}^n}{\sqrt{h_{i+1}^n}}\right)\mathbb{1}_{|\xi|^2\ge 2g\Delta Z_{i+\frac12}}d\xi\\ &=\int_{\xi'\le 0}\xi'\begin{pmatrix}1\\ -\sqrt{|\xi'|^2+2g\Delta Z_{i+\frac12}}\end{pmatrix}\chi\left(\frac{\xi'-u_{i+1}^n}{\sqrt{h_{i+1}^n}}\right)\mathbb{1}_{|\xi'|^2\ge -2g\Delta Z_{i+\frac12}}d\xi';\end{aligned}$$

then, a simple computation allows to conclude that

$$\begin{aligned}\mathbb{F}_{i+\frac12}^- &= h_i^n\int_{\omega\ge -\frac{u_i^n}{\sqrt{h_i^n}}}(\omega\sqrt{h_i^n}+u_i^n)\begin{pmatrix}1\\ \omega\sqrt{h_i^n}+u_i^n\end{pmatrix}\chi(\omega)d\omega\\ &\quad -h_i^n\int(\omega\sqrt{h_i^n}+u_i^n)\begin{pmatrix}1\\ -(\omega\sqrt{h_i^n}+u_i^n)\end{pmatrix}\\ &\qquad\times\chi(\omega)\mathbb{1}_{0\le\omega\sqrt{h_i^n}+u_i^n\le\sqrt{2g(\Delta Z_{i+\frac12})_+}}d\omega\\ &\quad +h_{i+1}^n\int_{\omega\le\frac{-\sqrt{2g(\Delta Z_{i+\frac12})_-}-u_{i+1}^n}{\sqrt{h_{i+1}^n}}}(\omega\sqrt{h_{i+1}^n}+u_{i+1}^n)\\ &\qquad\times\begin{pmatrix}1\\ -\sqrt{|\omega\sqrt{h_{i+1}^n}+u_{i+1}^n|^2+2g\Delta Z_{i+\frac12}}\end{pmatrix}\chi(\omega)d\omega,\end{aligned}$$

where we have used the classical algebraic notations

$$(\Delta Z_{i+\frac12})_+=max(0,\Delta Z_{i+\frac12}),\quad (\Delta Z_{i+\frac12})_-=max(-\Delta Z_{i+\frac12},0).$$

Similar manipulations in the formula (3.6) lead to

$$\begin{aligned}\mathbb{F}_{i-\frac12}^+ &= -h_i^n\int(\omega\sqrt{h_i^n}+u_i^n)\begin{pmatrix}1\\ -(\omega\sqrt{h_i^n}+u_i^n)\end{pmatrix}\\ &\qquad\times\chi(\omega)\mathbb{1}_{-\sqrt{2g(\Delta Z_{i-\frac12})_+}\le\omega\sqrt{h_i^n}+u_i^n\le 0}d\omega\\ &\quad +h_{i-1}^n\int_{\omega\ge\frac{\sqrt{2g(\Delta Z_{i-\frac12})_-}-u_{i-1}^n}{\sqrt{h_{i-1}^n}}}(\omega\sqrt{h_{i-1}^n}+u_{i-1}^n)\\ &\qquad\times\begin{pmatrix}1\\ \sqrt{|\omega\sqrt{h_{i-1}^n}+u_{i-1}^n|^2+2g\Delta Z_{i-\frac12}}\end{pmatrix}\chi(\omega)d\omega\end{aligned}$$

$$+ h_i^n \int_{\omega \le -\frac{u_i^n}{\sqrt{h_i^n}}} (\omega\sqrt{h_i^n} + u_i^n) \left(\frac{1}{\omega\sqrt{h_i^n} + u_i^n} \right) \chi(\omega) d\omega.$$

Remark 4.1. *As observed earlier, we are not able to compute all the integrals in the previous formulas. In fact, the choice of function χ in Section 2, which is necessary to achieve the properties of the numerical scheme stated in Theorem 3.2, leads to integrals which do not generically have an explicit primitive function.*

We distinguish three terms, for each component, in the formula of $\mathbb{F}^-_{i+\frac{1}{2}}$ stated above. We point out that similar integrals characterize the expression of $\mathbb{F}^+_{i-\frac{1}{2}}$, only with changes of sign in the domains of integration; thus we proceed to describe the flux $\mathbb{F}^-_{i+\frac{1}{2}}$.
Recalling that the function χ is defined as in (2.13), some elementary manipulations lead to obtain

$$\mathbb{F}^h_{i+\frac{1}{2}} = \frac{2\sqrt{2g}}{\pi} \left[(h_i^n)^{\frac{3}{2}} \mathbb{I}_1^h(Fr_i) + (h_i^n)^{\frac{3}{2}} \mathbb{I}_2^h(Fr_i, K^-_{i+\frac{1}{2}}) + (h_{i+1}^n)^{\frac{3}{2}} \mathbb{I}_3^h(Fr_{i+1}, K^+_{i+\frac{1}{2}}) \right]$$

and

$$\mathbb{F}^q_{i+\frac{1}{2}} = \frac{4g}{\pi} \left[(h_i^n)^2 \mathbb{I}_1^q(Fr_i) + (h_i^n)^2 \mathbb{I}_2^q(Fr_i, K^-_{i+\frac{1}{2}}) + (h_{i+1}^n)^2 \mathbb{I}_3^q(Fr_{i+1}, K^+_{i+\frac{1}{2}}) \right],$$

where we introduce the dimensionless numbers

$$Fr_i = \frac{u_i^n}{\sqrt{2gh_i^n}}, \quad K^-_{i+\frac{1}{2}} = \frac{\Delta Z_{i+\frac{1}{2}}}{h_i^n}, \quad K^+_{i+\frac{1}{2}} = \frac{\Delta Z_{i+\frac{1}{2}}}{h_{i+1}^n},$$

with, dropping the indices when no ambiguity is possible,

$$\mathbb{I}_1^h = \int_{\omega \ge -Fr} \omega \left(1 - \omega^2\right)_+^{\frac{1}{2}} d\omega + Fr \int_{\omega \ge -Fr} \left(1 - \omega^2\right)_+^{\frac{1}{2}} d\omega,$$

$$\mathbb{I}_2^h = - \int_{-Fr \le \omega \le \sqrt{(K^-)_+} - Fr} \omega \left(1 - \omega^2\right)_+^{\frac{1}{2}} d\omega - Fr \int_{-Fr \le \omega \le \sqrt{(K^-)_+} - Fr} \left(1 - \omega^2\right)_+^{\frac{1}{2}} d\omega,$$

$$\mathbb{I}_3^h = \int_{\omega \le -\sqrt{(K^+)_-} - Fr} \omega \left(1 - \omega^2\right)_+^{\frac{1}{2}} d\omega + Fr \int_{\omega \le -\sqrt{(K^+)_-} - Fr} \left(1 - \omega^2\right)_+^{\frac{1}{2}} d\omega,$$

$$\mathbb{I}_1^q = \int_{\omega\geq -Fr} \omega^2 \left(1-\omega^2\right)_+^{\frac{1}{2}} d\omega + 2\ Fr \int_{\omega\geq -Fr} \omega \left(1-\omega^2\right)_+^{\frac{1}{2}} d\omega + Fr^2 \int_{\omega\geq -Fr} \left(1-\omega^2\right)_+^{\frac{1}{2}} d\omega,$$

$$\mathbb{I}_2^q = \int_{-Fr\leq\omega\leq\sqrt{(K^-)_+}-Fr} \omega^2 \left(1-\omega^2\right)_+^{\frac{1}{2}} d\omega + 2\ Fr \int_{-Fr\leq\omega\leq\sqrt{(K^-)_+}-Fr} \omega \left(1-\omega^2\right)_+^{\frac{1}{2}} d\omega + Fr^2 \int_{-Fr\leq\omega\leq\sqrt{(K^-)_+}-Fr} \left(1-\omega^2\right)_+^{\frac{1}{2}} d\omega$$

and, finally,

$$\mathbb{I}_3^q = -\int_{\omega\leq -\sqrt{(K^+)_-}-Fr} (\omega+Fr)\sqrt{(\omega+Fr)^2+K^+}\left(1-\omega^2\right)_+^{\frac{1}{2}} d\omega.$$

Note that almost all the previous terms reduce to the same basic forms, except the last term which is quite different from the others and we will treat it later on. Classical techniques of integration, along with the usual goniometric equalities, allow us to conclude that

$$\int_a^b \omega(1-\omega^2)^{\frac{1}{2}} d\omega = -\frac{1}{3}(1-\omega^2)^{\frac{3}{2}}\Big|_a^b,$$

$$\int_a^b (1-\omega^2)^{\frac{1}{2}} d\omega = \frac{1}{2}\left(\arccos\omega - \omega\sqrt{1-\omega^2}\right)\Big|_b^a,$$

$$\int_a^b \omega^2(1-\omega^2)^{\frac{1}{2}} d\omega = -\frac{1}{3}\omega(1-\omega^2)^{\frac{3}{2}}\Big|_a^b + \frac{1}{12}\left[\frac{3}{2}\arccos\omega + \omega\sqrt{1-\omega^2}\left(\omega^2-\frac{5}{2}\right)\right]\Big|_b^a.$$

The choice of limits in these integrals is made according to the support of the function χ, so that

$$-1 \leq a = a(Fr, K^-, K^+, \pm 1) \leq 1,$$
$$-1 \leq b = b(Fr, K^-, K^+, \pm 1) \leq 1;$$

we specify the values of a and b at the moment of writing the final procedures in the actual implementation of the numerical method.
Then, a short computation leads to the following results:

$$\mathbb{I}_1^h = \frac{1}{3}(1-\alpha^2)^{\frac{3}{2}} + \frac{1}{2}Fr\arccos\alpha - \frac{1}{2}Fr\alpha\sqrt{1-\alpha^2},$$

$$\mathbb{I}_1^q = \frac{1}{3}(1-\alpha^2)^{\frac{3}{2}}(2Fr+\alpha) + \frac{1}{2}\arccos\alpha\left(\frac{1}{4}+Fr^2\right) + \frac{1}{2}\alpha\sqrt{1-\alpha^2}\left(\frac{1}{6}\alpha^2 - \frac{5}{12} - Fr^2\right),$$

where $\alpha = min\{1, max\{-1, -Fr\}\}$;

$$\mathbb{I}_2^h = \frac{1}{3}(1-\beta^2)^{\frac{3}{2}} - \frac{1}{3}(1-\alpha^2)^{\frac{3}{2}} + \frac{1}{2}Fr(\arccos\beta - \arccos\alpha) + \frac{1}{2}Fr(\alpha\sqrt{1-\alpha^2} - \beta\sqrt{1-\beta^2}),$$

$$\mathbb{I}_2^q = \frac{1}{3}(1-\alpha^2)^{\frac{3}{2}}(2Fr+\alpha) - \frac{1}{3}(1-\beta^2)^{\frac{3}{2}}(2Fr+\beta) + \frac{1}{2}\left(\frac{1}{4}+Fr^2\right)(\arccos\alpha - \arccos\beta) + \frac{1}{2}\beta\sqrt{1-\beta^2}\left(\frac{5}{12} - \frac{1}{6}\beta^2 + Fr^2\right) - \frac{1}{2}\alpha\sqrt{1-\alpha^2}\left(\frac{5}{12} - \frac{1}{6}\alpha^2 + Fr^2\right),$$

where $\alpha = min\{1, max\{-1, -Fr\}\}$ and $\beta = max\left\{-1, min\{\sqrt{(K^-)_+} - Fr, 1\}\right\}$;

$$\mathbb{I}_3^h = -\frac{1}{3}(1-\beta^2)^{\frac{3}{2}} + \frac{1}{2}Fr(\pi - \arccos\beta) - \frac{1}{2}Fr\beta\sqrt{1-\beta^2},$$

where $\beta = max\left\{-1, min\{-\sqrt{(K^+)_-} - Fr, 1\}\right\}$.

We now return to the most complicated term $\mathbb{I}_3^q$. Setting $\alpha = -1$ and $\beta = max\left\{-1, min\{-\sqrt{(K^+)_-} - Fr, 1\}\right\}$, we can rewrite it as

$$\mathbb{I}_3^q = \int_\alpha^\beta f(\omega, Fr, K^+)d\omega,$$

with

$$f(\omega, Fr, K^+) = -(\omega+Fr)\sqrt{(\omega+Fr)^2 + K^+}\left(1-\omega^2\right)_+^{\frac{1}{2}}.$$

The presence of the square root makes it impossible to compute immediately; we need to formulate a suitable approximation, preserving the main theoretical features of the real integral. We propose a rather natural choice, based on a numerical method of integration, by means of a quadrature formula: in particular, comparison tests lead us to prefer a classical *repeated midpoint formula*,

$$\mathbb{I}_3^q(Fr, K^+) \simeq \mathbb{I}^*(Fr, K^+) = \frac{\beta-\alpha}{N}\sum_{j=1}^{N} f\left(\alpha + (j-\frac{1}{2})\frac{\beta-\alpha}{N}\right),$$

where N is chosen in order to assure the best compromise between the accuracy of the numerical method and a reasonable computating time. Of course faster algorithms are possible but at this level we are only interested in testing the method and we leave it for further extensions to improve the computational performance.

4.2 Some numerical tests

We conclude these notes with some numerical examples, that illustrate the results stated in the previous sections, in order to confirm that the properties of the Saint-Venant system (1.1)-(1.2) are preserved by the numerical scheme (3.4)-(3.6) introduced in this paper and to evaluate its performance on other classical test cases.
We check the properties of the scheme on different test cases for which analytical solutions of the equations are available and on more realistic applications (most of the experiments simulated here come from a workshop on dam-break wave simulation).

4.2.1 We begin with a non-stationary test case, a dam-break problem in a rectangular channel with flat bottom ($Z = 0$). The initial conditions are

$$u(0,x) = 0,$$
$$h(0,x) = \begin{cases} h_l & \text{for } x \leq 0 \\ h_r & \text{for } x > 0, \end{cases}$$

where $h_l > h_r$ in order to be consistent with the physical phenomenon of a dam-break from the left to the right.
Note that this case corresponds to a Riemann problem for the simpler homogeneous model of system (1.1)-(1.2) and we can compare the numerical solution with the exact solution (plotted with a dotted line), computed by the classical theory (see Dafermos [5], Serre [22]).
The channel length is L = $2000m$ and the computational domain is chosen to be symmetric around the point x = 0; the mesh size is Ds = L/100 and the time-step Dt is computed according to the CFL condition (3.9), in order to verify numerically that the water height positivity is preserved.
The Figures (1)-(2) and (3)-(4) present respectively the results observed at time T = $200s$ for a dam-break on a wet bed ($h_l = 1m$, $h_r = 0.5m$) and at time T = $150s$ for a dam-break on a dry soil ($h_l = 1m$, $h_r = 0m$).

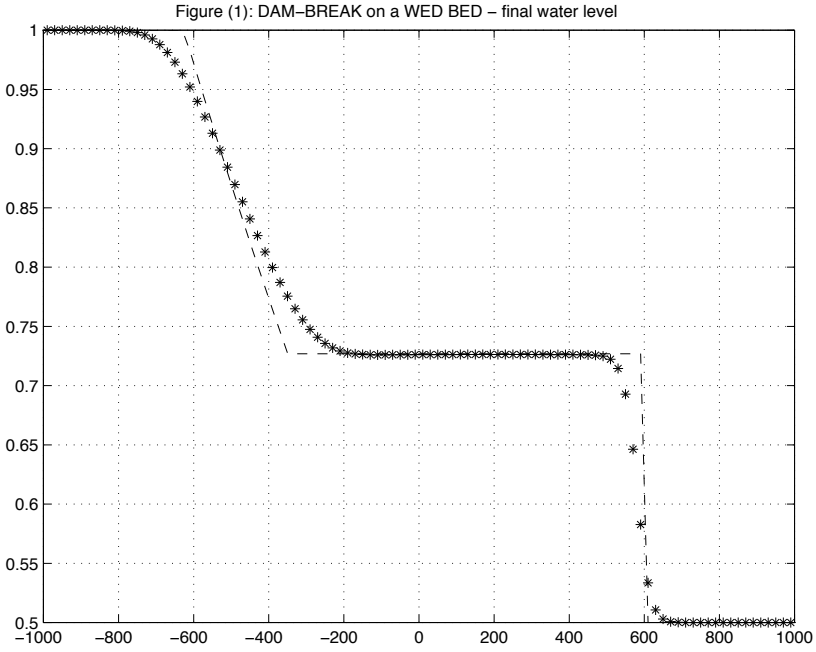


Figure (1): DAM–BREAK on a WED BED – final water level
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
−1000
−800
−600
−400
−200
0
200
400
600
800
1000

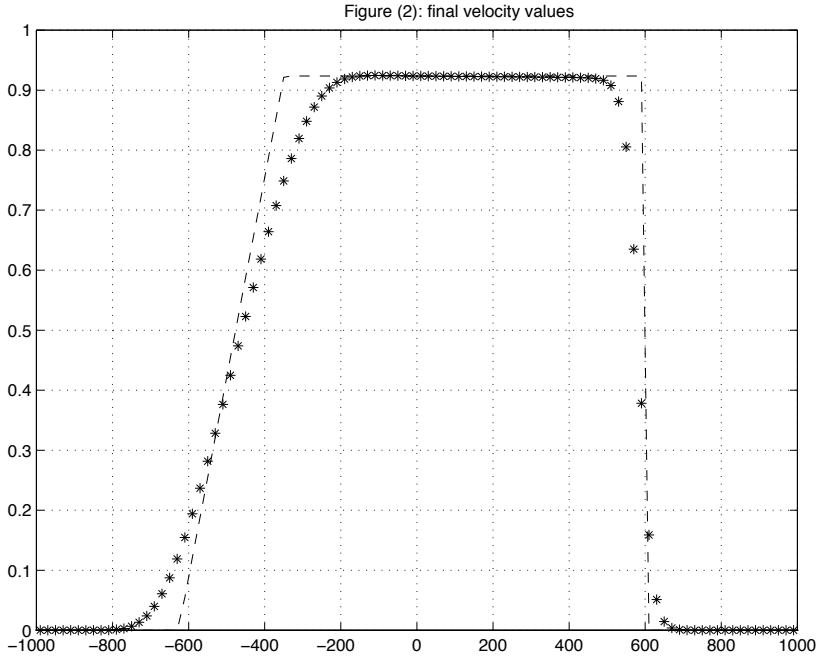
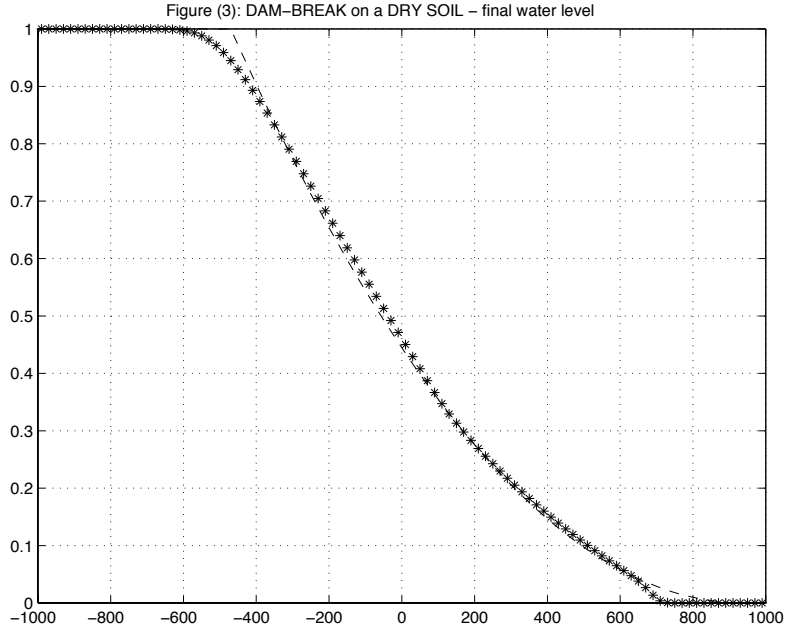
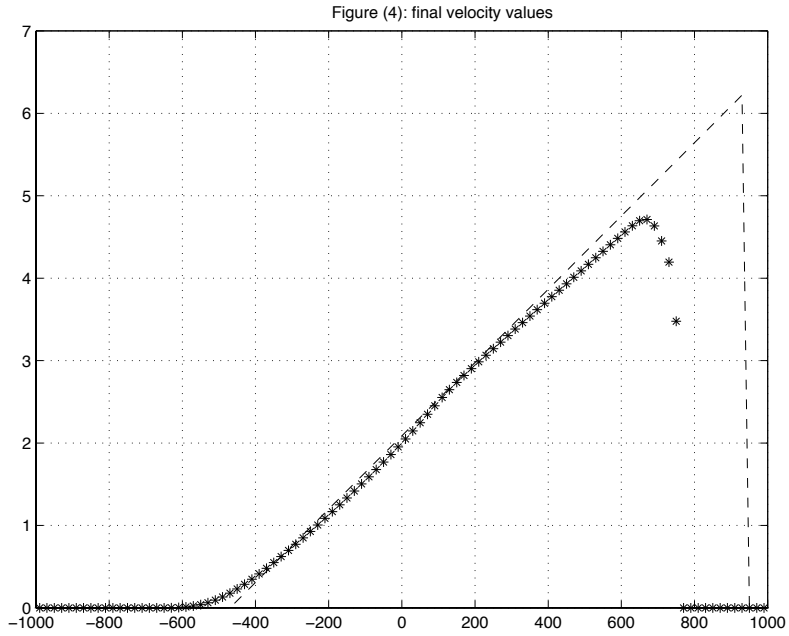


Figure (2): final velocity values
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−1000
−800
−600
−400
−200
0
200
400
600
800
1000





4.2.2 We now consider a test case concerning the steady state of a lake at rest, on a non-trivial topography, in order to validate the numerical scheme on a steady flow: we show that this steady state is preserved, up to the accuracy in the approximation of the integral we have discussed at the end of Subsection 4.1.

Figure (5): LAKE AT REST – final water level

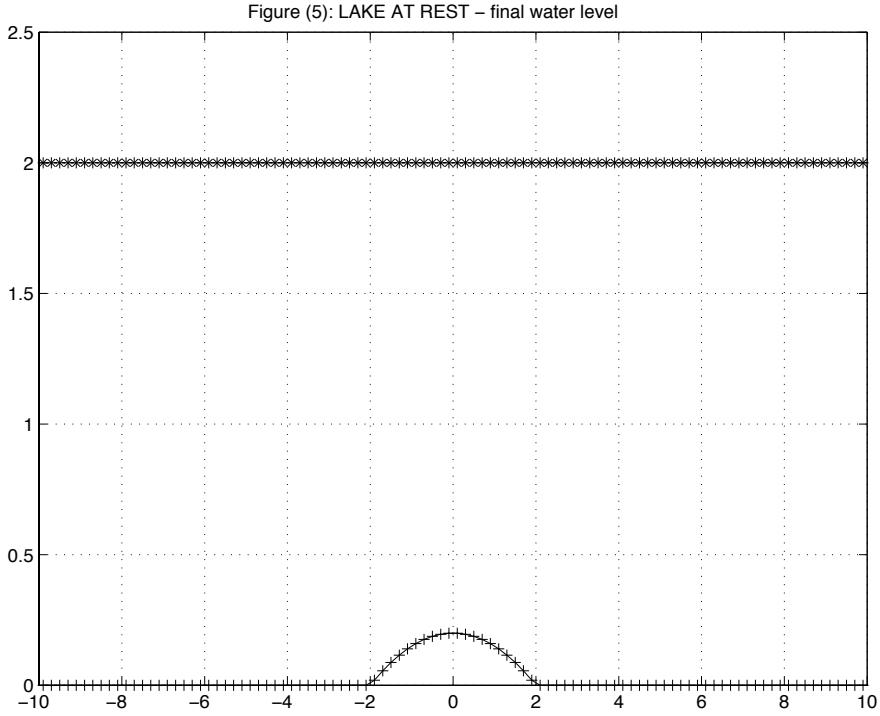


Figure (6): final velocity values

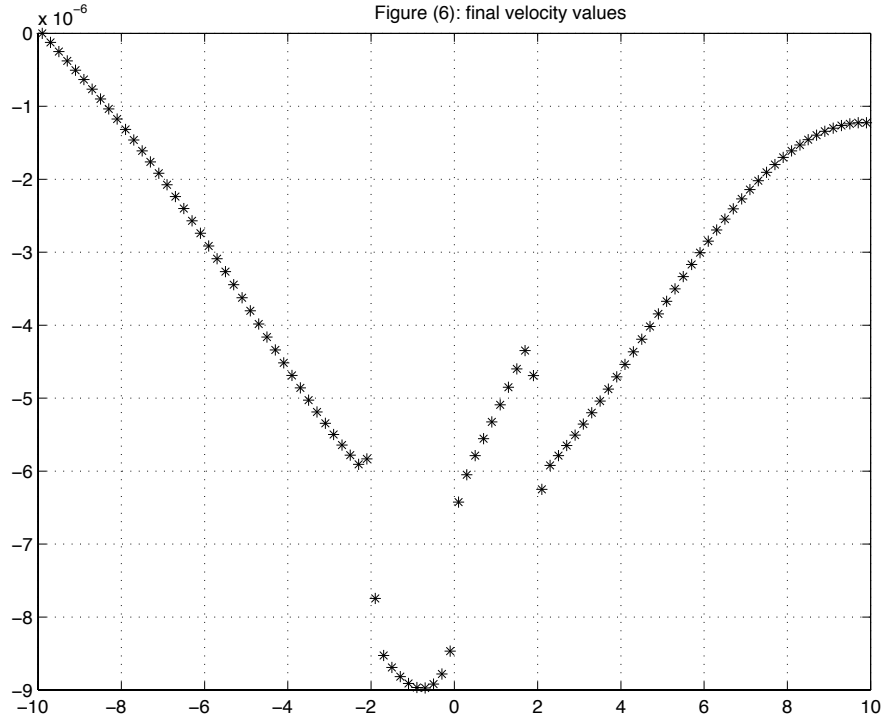


Figure (7): LAKE AT REST – final water level

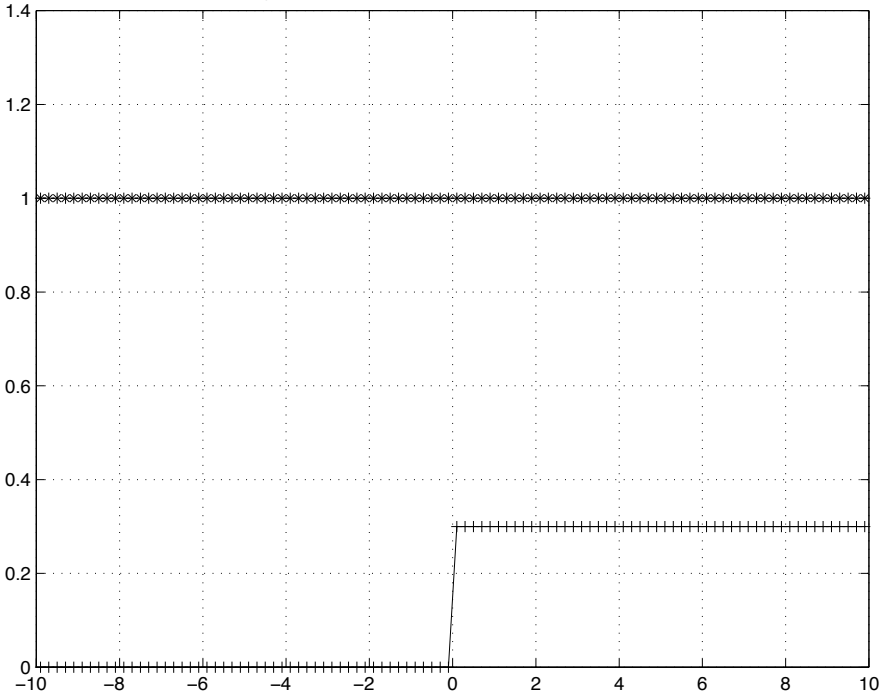
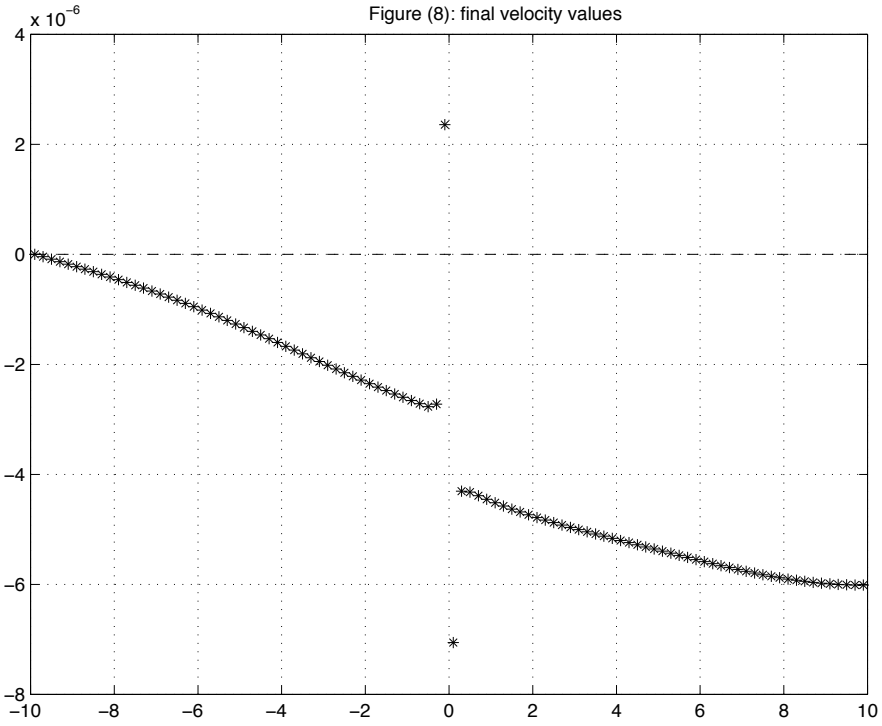


Figure (8): final velocity values



The initial conditions are

$$u(0,x) = 0,$$
$$h(0,x) + Z(x) = H,$$

with $H = 2m$ and we set the same computational parameters as in the first test case, except for the channel length L $= 20m$. The solution plotted in the Figures (5)-(6) corresponds to a channel with a parabolic bump on the bottom, described by the function

$$Z(x) = \left(0.2 - 0.05 * x^2\right)_+ ;$$

in the Figures (7)-(8) we present the same test case on a different topography, given by a discontinuous step

$$Z(x) = \begin{cases} Z_l & \text{for } x \leq 0 \\ Z_r & \text{for } x > 0, \end{cases}$$

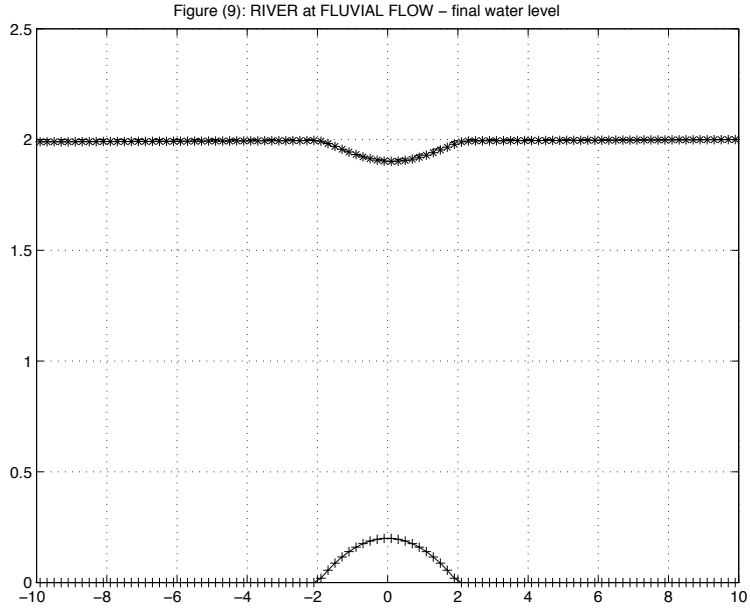
with $Z_l = 0m$ and $Z_r = 0.3m$ (note that the geometry of the source term is not regular here, which is not in agreement with the assumptions of the classical theory).

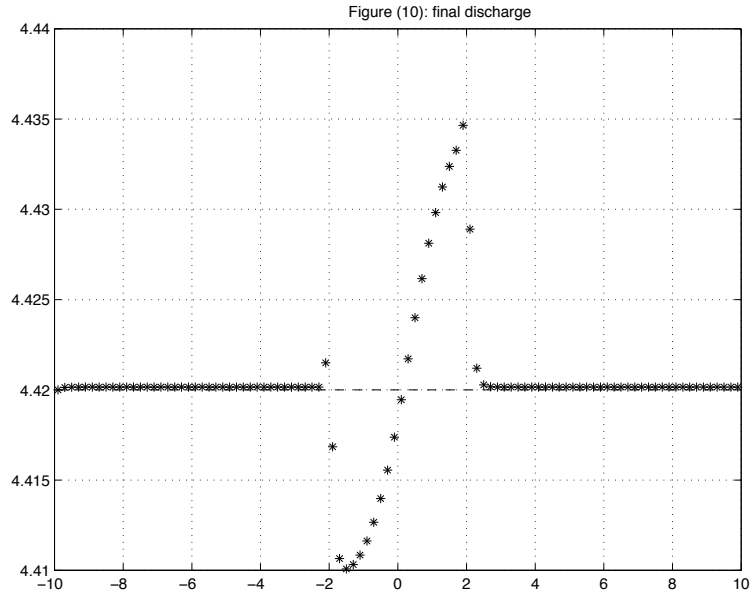
4.2.3 Our purpose in the following test cases is to study the convergence in time towards a more general steady state. We consider a rectangular channel with the same geometry as in Figure (5) and we compute the steady state occurring since a constant discharge is imposed at the upstream boundary condition. We compare the numerical solution with the analytical solution, provided by the means of the formulas (2.3)-(2.4) in Section 1.
According to the boundary and initial conditions, the flow may be subcritical (or fluvial), transcritical without shock (the flow becomes torrential at the top of the bump and the outflow is torrential) and transcritical with shock (the flow becomes torrential at the top of the bump and the outflow is fluvial). We impose an upstream boundary condition Q_{in} on the discharge and a downstream boundary condition H_{out} on the water level, as follows:

- subcritical flow $\begin{cases} Q_{in} = 4.42m^2/s \\ H_{out} = 2m, \end{cases}$
- transcritical flow without shock $\begin{cases} Q_{in} = 1.53m^2/s \\ H_{out} = 0.66m, \end{cases}$

if the outflow is subcritical (remark that no condition is imposed on the downstream limit when the outflow becomes supercritical),

- transcritical flow with shock $\begin{cases} Q_{in} = 0.18m^2/s \\ H_{out} = 0.33m. \end{cases}$



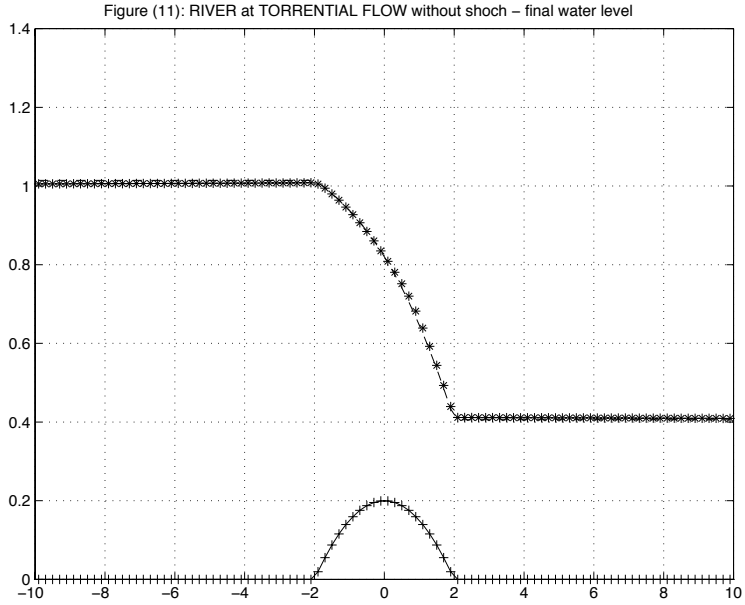


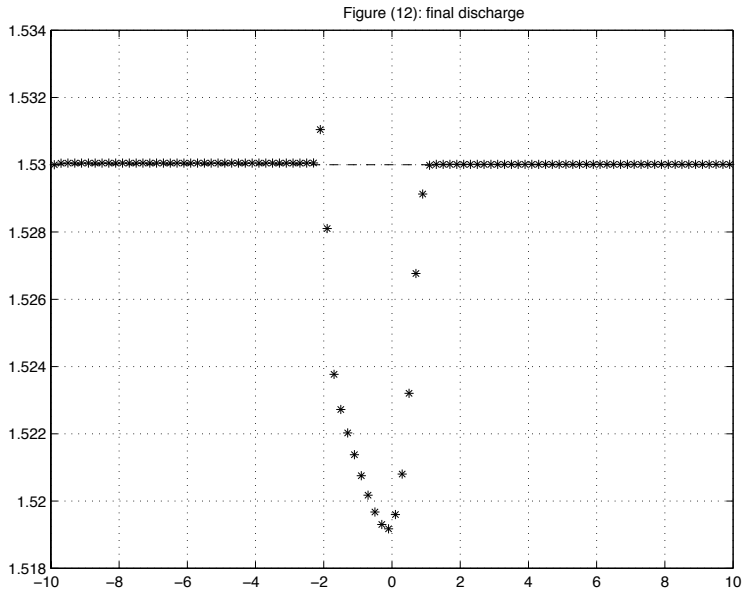
For all these cases, the initial conditions are

$$u(0,x) = 0,$$
$$h(0,x) + Z(x) = H,$$

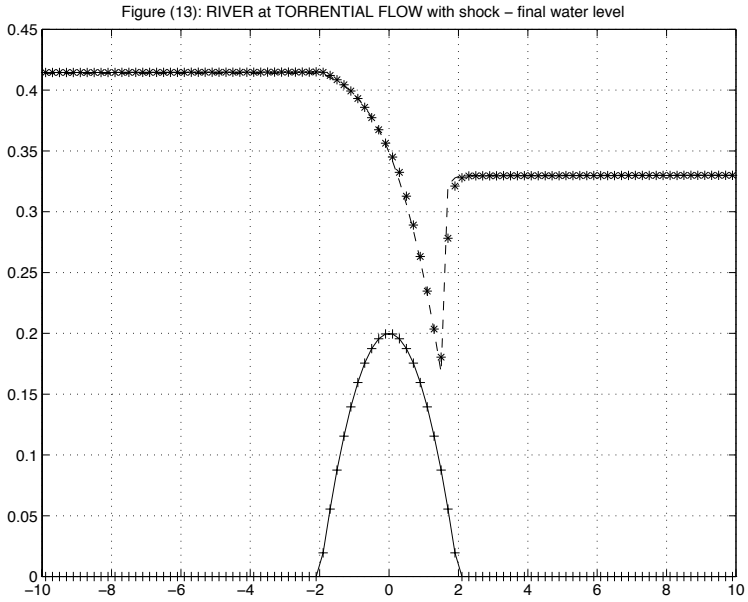
where H is the constant level of the water surface prescribed downstream. All solutions are plotted at time T = 200s (analytical solutions are also plotted

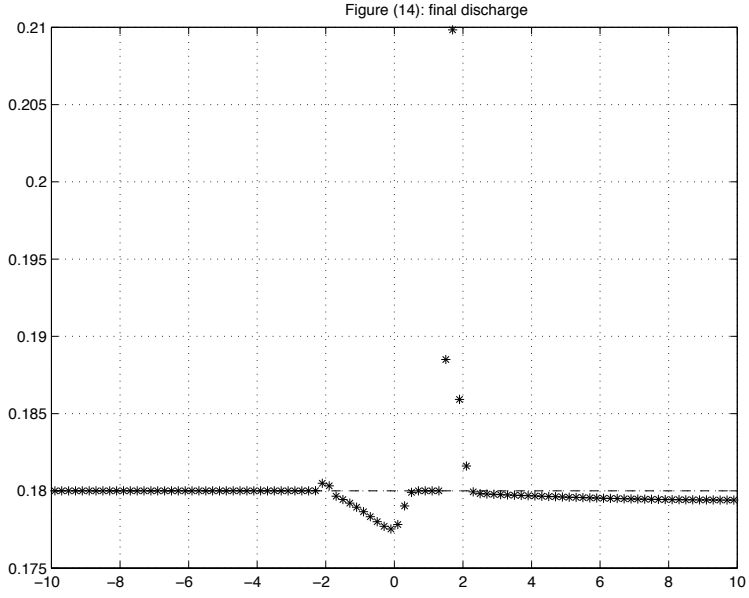
with a dotted line) and a mesh with Ds = L/100 is used, but of course some results can improve according to the mesh refinement.





Notice that Figures (10), (12) and (14) refer to steady states generally not at rest ($u \neq 0$) and thus property (iii) does not apply. To the best of our knowledge, only the results in [7] are comparable to these tests.





4.2.4 The last test case we deal with in this paper is the quasi-stationary case proposed by LeVeque in [17], to compute small perturbations of the steady state of a lake at rest. According to the parameters fixed by the author, as bottom topography we take

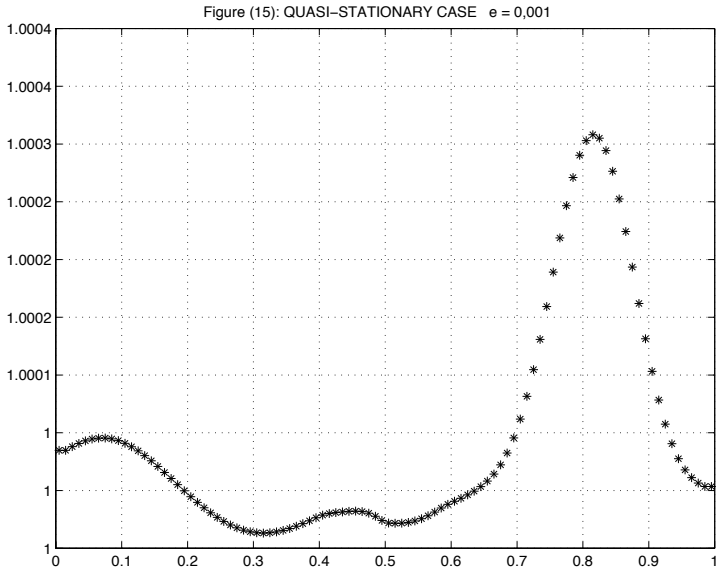
$$Z(x) = \Big(0.25 \left(\cos\left(\pi(x - 0.5)/0.1\right) + 1\right) \Big)_+ ,$$

centered in a rectangular channel of length L = $1m$ and the mesh size is Ds = L/100. The initial conditions are

$$u(0,x) = 0,$$
$$h(0,x) + Z(x) = H + \epsilon(x),$$

with $H = 1m$ and we consider a perturbation of the water surface

$$\epsilon(x) = \text{e } \mathbb{1}_{0.1<x<0.2}.$$



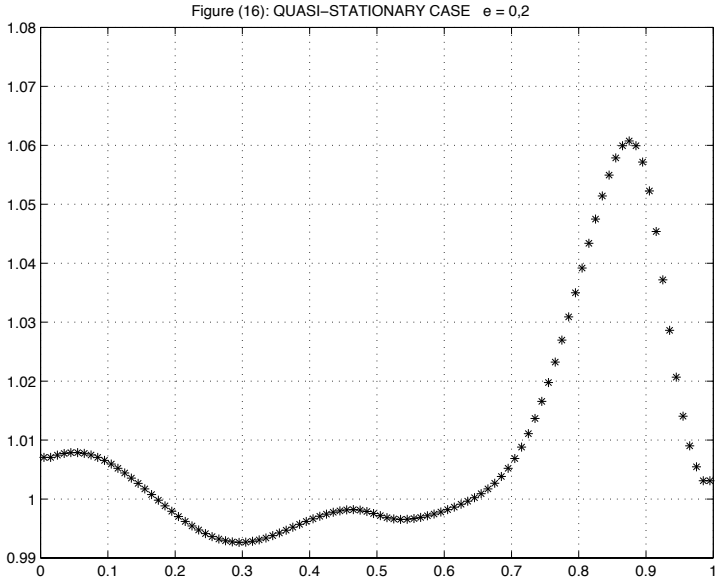


Figure (15) and Figure (16) show the water surface given by the numerical solution at time T = $0.7s$ (the usual simplification g=1 is also assumed in this example), respectively for e = 10^{-3} and e = 0.2.
We only remark that perturbations in quasi-steady problems are computed by our scheme with the same resolution as would be expected if calculating small perturbations about constant states for the homogeneous system ($\frac{\partial Z}{\partial x} = 0$).

References

[1] Audusse E., Bristeau M.O. and Perthame B., Kinetic schemes for Saint-Venant equations with source terms on unstructured grids, *INRIA Report* RR-3989 (2000)

[2] Arvanitis C., Katsaounis T. and Makridakis C., Adaptive finite element relaxation schemes for hyperbolic conservation laws, *M2AN Math. Model. Numer. Anal.*, **35** (2001), no. 1, 17-33

[3] Bermudez A. and Vasquez M.E., Upwind methods for hyperbolic conservation laws with source terms, *Comput. Fluids*, **23** (1994), no. 8, 1049-1071

[4] Botchorishvili R., Perthame B. and Vasseur A., Equilibrium Schemes for Scalar Conservation Laws with Stiff Sources, *INRIA Report* RR-8931 (2000)

[5] Dafermos C.M., *Hyperbolic conservation laws in continuum physics*, vol. GM 325 Springer-Verlag, Berlin, 1999

[6] Eymard R., Gallouët T. and Herbin R., *Finite volume methods*, Handbook of numerical analysis, vol. VIII, P.G.Ciarlet and J.L.Lions editors, Amsterdam, North-Holland, to appear

[7] Gallouët T., Hérard J.M. and Seguin N., Some approximate Godunov schemes to compute shallow-water equations with topography, *AIAA-2001* (2000)

[8] Gerbeau J.F. and Perthame B., Derivation of viscous Saint-Venant system for laminar shallow water; numerical validation, *Discrete Contin. Dynam. Systems*, to appear

[9] Godlewski E. and Raviart P.A., *Numerical approximation of hyperbolic systems of conservation laws*, Applied Mathematical Sciences **118**, New York, Springer-Verlag, 1996

[10] Gosse L. and LeRoux A.Y., A well-balanced scheme designed for inhomogeneous scalar conservation laws, *C.R. Acad. Sci. Paris Sér.I Math.*, **323** (1996), no. 5, 543-546

[11] Gosse L., A priori error estimate for a well-balanced scheme designed for inhomogeneous scalar conservation laws, *C.R. Acad. Sci. Paris Sér.I Math.*, **327** (1998), no. 5, 467-472

[12] Gosse L., A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms, *Comput. Math. Appl.*, **39** (2000), no. 9-10, 135-159

[13] Gosse L., A well-balanced scheme using non-conservative products designed for hyperbolic systems of conservation laws with source terms, *Math. Mod. Meth. Appl. Sci.*, to appear

[14] Greenberg J.M. and LeRoux A.Y., A well balanced scheme for the numerical processing of source terms in hyperbolic equations, *SIAM J. Num. Anal.*, **33** (1996), 1-16

[15] Jin S., A steady-state capturing method for hyperbolic systems with geometrical source terms, *M2AN Math. Model. Numer. Anal.*, to appear

[16] Le Vêque R.J., *Numerical Methods for Conservation Laws*, Lectures in Mathematics, ETH Zurich, Birkhauser, 1992

[17] Le Vêque R.J., Balancing source terms and flux gradients in high-resolution Godunov methods: the quasi-steady wave-propagation algorithm, *J. Comput. Phys.*, **146** (1998), no. 1, 346-365

[18] Lions P.L., Perthame B. and Tadmor E., Kinetic formulation of the Isentropic Gas Dynamics and p-Systems, *Commun. Math. Phys.*, **163** (1994), no. 2, 415-431

[19] Perthame B., An introduction to kinetic schemes for gas dynamics, *An introduction to recent developments in theory and numerics for conservation laws*, L.N. in Computational Sc. and Eng. **5**, D.Kroner, M.Ohlberger and C.Rohde editors, Springer, 1998

[20] Roe P.L., Upwind differenced schemes for hyperbolic conservation laws with source terms, *Proc. Conf. Hyperbolic Problems*, Carasso, Raviart and Serre editors, Springer, 1986, pp. 41-51

[21] de Saint-Venant A.J.C., Théorie du mouvement non-permanent des eaux, avec application aux crues des rivière at à l'introduction des marées dans leur lit, *C.R. Acad. Sci. Paris*, **73** (1871), 147-154

[22] Serre D., *Systèmes hyperboliques de lois de conservation*, *Tomes I et II*, Diderot Eds, Paris 1996

[23] Stroud A.H., *Numerical Quadrature and Solution of Ordinary Differential Equations*, Applied Mathematical Sciences **10**, New York-Heidlberg, Springer-Verlag, 1974, pp. 106-122

[24] Xu K., Unsplitting BGK-type schemes for the shallow water equations, *Internat. J. Modern Phys. C.*, **10** (1999), no.4, 505-516

Chapitre 4

Approximation d'ordre deux du système de Saint-Venant visqueux et comparaison avec les expériences

(soumis à *Proceedings of Hyp2002*)

Second order approximation of the viscous Saint-Venant system and comparison with experiments

T. Katsaounis, C. Simeoni

Département de Mathématiques et Applications
École Normale Supérieure
45, rue d'Ulm - 75230 Paris Cedex 05 - France
...
Unité de Recherche INRIA Rocquencourt - Projet M3N
Domaine de Voluceau-Rocquencourt
B.P. 105 - 78153 Le Chesnay Cedex - France
e-mails: katsaoun@dma.ens.fr, simeoni@dma.ens.fr

Abstract

We present numerical simulations of the Saint-Venant equations for shallow water, including small friction and viscosity, motivated by the interest in recovering the results of experimental studies on the free-surface flows over an obstacle. We use the kinetic scheme "with reflexions" formulated in [17], appropriately extended to obtain second order accuracy according to the theory developed in [11].

1 Introduction

The Saint-Venant equations for shallow waters were originally written by A. de Saint-Venant in 1871 from heuristic considerations about the mechanisms governing physical phenomena such as the flows in rivers or coastal areas (see [20]).

In a one-dimensional framework, these equations constitute simple mathematical model for the flow in ideal rectangular rivers, described at time $t \geq 0$

and at point $x \in \mathbb{R}$ through the height of water $h(t,x) \geq 0$ and its velocity $u(t,x) \in \mathbb{R}$ by means of the hyperbolic system

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(hu) = 0, \tag{1.1}$$

$$\frac{\partial}{\partial t}(hu) + \frac{\partial}{\partial x}\Big(hu^2 + \frac{g}{2}h^2\Big) + ghZ' = 0, \tag{1.2}$$

where g denotes the gravity intensity and $Z(x)$ is the bottom topography; therefore $h+Z$ is the level of the water surface and, in what follows, we also denote the discharge by $q=hu$.
Besides, other terms can be added to the right-hand side of equation (1.2) to take into account further natural features of the physical context, for instance friction on the bottom and viscosity inside the fluid.

The question to introduce more complete systems becomes crucial when one deals with the experimental verification of situations typically occurring in hydraulics, to provide a classification of different flow regimes in presence of an obstacle, for which the description based on the model (1.1)-(1.2) is unsatisfactory. Indeed, for dam breaks or hydraulic jumps, it does not allow to recover mathematically the right position with respect to the topography (see [12], for example) and the interaction of the source terms corresponding to bottom slope and friction in the shallow water equations is dominant for characterizing the steady states.

Despite its simple configuration, the shallow water flow in channels with nontrivial topography present a wide variety of regimes, producing some peculiar behaviours of the free-surface (wave trains, hydraulic jumps, turbulent profiles), which have not yet been fully examined. In [1], the results of classical analytical theories are reviewed, even though these models are validated only in the weakly nonlinear and weakly dispersive limits. Based on two-dimensional numerical simulations of the nonlinear uniform potential flow around moving obstacles, an accurate description of the different types of breaking waves is proposed in [14].
However, due to the inherent limitations of the theoretical formulations, their predictions have to be interpreted in terms of experimental verifications.

We refer to [26] for a survey of the experimental study of free-surface flows over an obstacle: the behaviour of an incident subcritical channel flow is investigated for various blocking factors (namely, obstacle shape and stationary water depth); the results are analyzed in comparison with the classification schemes proposed in the previous works (see the references in that paper).

We present in this paper some numerical simulations of the Saint-Venant system, according to the experimental configuration set in [26].
To take into account dissipative effects in the physical phenomenon, we con-

sider a modified equation for the momentum including small friction and viscosity, well then the system under analysis reads

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(hu) = 0, \tag{1.3}$$

$$\frac{\partial}{\partial t}(hu) + \frac{\partial}{\partial x}\left(hu^2 + \frac{g}{2}h^2\right) + ghZ' = -\frac{g}{K^2}\frac{u|u|}{h^{1/3}} + \mu\frac{\partial}{\partial x}\left(h\frac{\partial u}{\partial x}\right), \tag{1.4}$$

where K is the Strickler's coefficient of the Manning's equation and μ denotes the kinematic viscosity of the fluid. The particular form of the source terms is suggested by empirical laws, which were originally obtained for steady state flows (refer to [25]). Similar models have been derived from the Navier-Stokes system for incompressible flows with a free moving boundary (see [5] and its references) and more complex laws for the friction term can also be formulated to modelize analogous problems in the case of granular media (snow avalanches, for example, as referred in [15]).

The numerical approximation of system (1.3)-(1.4) is carried out by means of the kinetic scheme proposed and extensively studied in [17]. In fact, for all experimental tests performed in the works quoted above, the flow was observed to be stationary and a zone of supercritical flow downstream the obstacle always occurs, though the cases examined concern the subcritical regime. So, we use a method which preserves the free-surface profile of steady states with nontrivial bottom topography and which is able to deal with transcritical regimes.
The discretization of the additional source terms in (1.4) is rather standard, based on a semi-implicit approach for the friction term and direct integrations by the finite volume method for the diffusive term. Moreover, when friction and viscosity are neglected, we recover the system (1.1)-(1.2).
Appropriate extensions of the primitive algorithm are thus considered, in order to improve the numerical accuracy.

The paper is organized as follows. In Section 2, we recall some specific notations of the Upwind Interface Source method for hyperbolic conservation laws with geometrical source term, illustrated in [18], by extending its general formalism to the system (1.1)-(1.2). We also introduce the approximation of the dissipative terms, for treating the system (1.3)-(1.4). Connected with the numerical approach developed in [17], the question to derive a second order scheme is addressed in Section 3. We describe the experimental configuration underling our analysis in Section 4 and we present the results of numerical simulations made according to the experimental data provided in [26]. Some remarks are discussed to justify theoretical tests in comparison with the experiments.

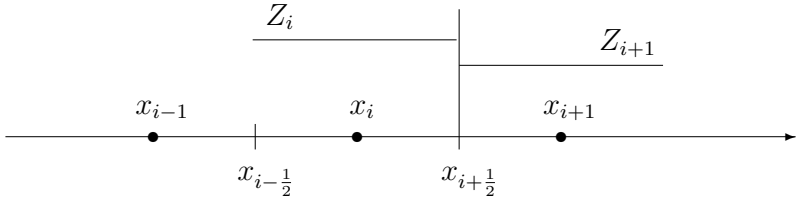
2 Formalism of the numerical method

The Saint-Venant system for shallow waters (1.1)-(1.2) belongs to the class of the hyperbolic systems of balance laws, with a geometrical source term, and can be written in the equivalent form

$$\frac{\partial \mathcal{U}}{\partial t} + \frac{\partial}{\partial x} A(\mathcal{U}) = B(x,\mathcal{U}), \tag{2.1}$$

where $\mathcal{U} = (h, hu)$ represents the vector of conservative variables, the flux function is given by $A(\mathcal{U}) = \left(hu, hu^2 + \frac{g}{2}h^2\right)$ and $B(x,\mathcal{U}) = (0, -ghZ')$ indicates the external term.
The equation (2.1) reproduces the general formalism introduced in [18] for the particular case of scalar conservation laws, so the numerical theory stated in that context formally extends to system (1.1)-(1.2), to characterize approximations with suitable theoretical properties.

We set up a mesh on $\mathbb{R}$, whose central vertices are x_i, $i \in \mathbb{Z}$, made of cells $\mathcal{C}_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$ with nonuniform length Δx_i and the points $x_{i+\frac{1}{2}}$ indicate the cell interfaces. We also consider a time discretization t^n, $n \in \mathbb{N}$, with variable time-step Δt. Then we construct a piecewise constant representation of the function $Z(x)$ on the mesh, with coefficients $Z_i = \frac{1}{\Delta x_i} \int_{\mathcal{C}_i} Z(x)\, dx$ for example.



A classical approach to nonlinear hyperbolic problems consists in using finite volume methods, which are designed for computations with arbitrary meshes (refer to [4], for instance). Taking into account the source term directly in the definition of the numerical fluxes, the fully explicit finite volume scheme for equation (2.1) is written in the compact form

$$\mathcal{U}_i^{n+1} - \mathcal{U}_i^n + \frac{\Delta t}{\Delta x_i}\left(A_{i+\frac{1}{2}}^{n,-} - A_{i-\frac{1}{2}}^{n,+}\right) = 0, \tag{2.2}$$

where $A_{i+\frac{1}{2}}^{n,\pm} = \mathcal{A}^{\pm}(\mathcal{U}_i^n, \mathcal{U}_{i+1}^n, \Delta Z_{i+\frac{1}{2}})$, with $\Delta Z_{i+\frac{1}{2}} = Z_{i+1} - Z_i$, are defined by means of appropriate numerical functions $\mathcal{A}^{\pm} = \left(\mathcal{A}_h^{\pm}, \mathcal{A}_q^{\pm}\right)$ for each component of (2.1) and the following consistency properties are required,

$$\mathcal{A}_h^+(\mathcal{U},\mathcal{V},\Delta Z) = \mathcal{A}_h^-(\mathcal{U},\mathcal{V},\Delta Z), \tag{2.3a}$$

$$\mathcal{A}_q^+(\mathcal{U},\mathcal{V},\Delta Z) - \mathcal{A}_q^-(\mathcal{U},\mathcal{V},\Delta Z) = -gh\Delta Z + \mathcal{O}\left(|\Delta Z| + |\mathcal{U}-\mathcal{V}|\right), \tag{2.3b}$$

$$\mathcal{A}^+(\mathcal{U},\mathcal{U},0) = \mathcal{A}^-(\mathcal{U},\mathcal{U},0) = A(\mathcal{U}). \tag{2.3c}$$

We note that condition (2.3a) ensures that the numerical fluxes are conservative for the equation (1.1). Moreover, we have $\mathcal{A}^{n,\pm}_{i+\frac{1}{2}} = \mathcal{A}^{n}_{i+\frac{1}{2}} + (\mathcal{A}^{n,\pm}_{i+\frac{1}{2}} - \mathcal{A}^{n}_{i+\frac{1}{2}})$ and the quantity in brackets holds for discrete contributions of the source term at the cell interfaces, according to the Upwind Interface Source method; therefore, the relation (2.3b) guarantees consistency with the continuous model, as readily obtained by standard asymptotic expansions. We deduce from (2.3c) that the numerical scheme (2.2) satisfies the classical definitions for homogeneous problems.

The kinetic scheme for the Saint-Venant system proposed in [17] is compatible with the above formalism and endowed with further stability properties associated to the physical model (it preserves the steady state of still water, satisfies a discrete entropy inequality and makes non-negative water height).

In order to perform numerical simulations with experimental data, we consider the modified shallow water equations (1.3)-(1.4), for which that scheme applies to terms corresponding to the hyperbolic system (1.1)-(1.2). The discretization of the friction term is implicit (see [16], for example) and splitted into two steps, only concerning the equation (1.4), which include the approximation of the viscous term,

$$q_i^{n+\frac{1}{2}} - q_i^n + \frac{\Delta t}{2\Delta x_i}\left(A^{n,-}_{q,\,i+\frac{1}{2}} - A^{n,+}_{q,\,i-\frac{1}{2}}\right) = \mu\frac{\Delta t}{2\Delta x_i}\left(V^n_{i+\frac{1}{2}} - V^n_{i-\frac{1}{2}}\right), \tag{2.4}$$

$$q_i^{n+1} - q_i^{n+\frac{1}{2}} = -\frac{\Delta t}{2}\,\frac{g}{K^2}\,\frac{q_i^{n+1}|q_i^{n+\frac{1}{2}}|}{\left(h_i^{n+1}\right)^{\frac{7}{3}}}, \tag{2.5}$$

where the numerical formulas used for calculating viscosity,

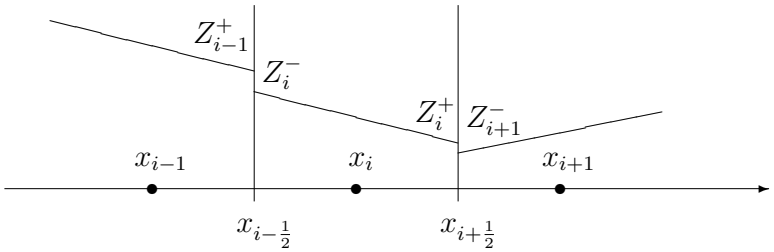
$$V^n_{i+\frac{1}{2}} = \frac{h_i + h_{i+1}}{2}\,\frac{u_{i+1} - u_i}{\Delta x_{i+\frac{1}{2}}}, \tag{2.6}$$

with $\Delta x_{i+\frac{1}{2}} = \frac{\Delta x_i}{2} + \frac{\Delta x_{i+1}}{2}$, are derived by means of simple finite volume integrations on the mesh cells and appropriate approximations of the resulting interfacial values (we note that the discretization (2.6) can be reinterpreted according to the classical first order finite element method). Some different methods for processing friction terms are proposed in [2], [6] and [7].

3 Second order schemes

To obtain second order extensions of finite volume schemes in form (2.2), a rather geometrical approach is based on *slope limiter* techniques.
We construct a piecewise linear approximation of the function $Z(x)$ on the

mesh, whose coefficients are $Z_i+(x-x_i)Z_i'$, $x\in\mathcal{C}_i$, and we denote by Z_i' the numerical derivatives computed by applying an appropriate *slope limiter* (we refer to [8] and [13] for a survey of these discrete operators).



According to the arguments in [11], a second order scheme for the Upwind Interface Source method formally reads

$$\mathcal{U}_i^{n+1}-\mathcal{U}_i^n+\frac{\Delta t}{\Delta x_i}\left(A_{i+\frac{1}{2}}^{n,-}-A_{i-\frac{1}{2}}^{n,+}\right)+\Delta t\,B_i^n=0, \tag{3.1}$$

where the numerical fluxes $A_{i+\frac{1}{2}}^{n,\pm}=\mathcal{A}^{\pm}(\mathcal{U}_i^{n,+},\mathcal{U}_{i+1}^{n,-},\Delta Z_{i+\frac{1}{2}})$ use the interfacial values of piecewise linear reconstructions of the numerical functions,

$$\mathcal{U}_i^{n,\pm}=\mathcal{U}_i^n\pm\frac{\Delta x_i}{2}\,\mathcal{U}_i^{n\prime},\qquad Z_i^{\pm}=Z_i\pm\frac{\Delta x_i}{2}\,Z_i', \tag{3.2}$$

and therefore $\Delta Z_{i+\frac{1}{2}}=Z_{i+1}^--Z_i^+$ in this case.
The additional discrete source term is defined by $B_i^n=(0,gZ_i'h_i^n)$. Although other methods exhibit noticeable improvements, it was shown in [11] that the centered term B_i^n is necessary to achieve second order accuracy for the Upwind Interface Source method, if the *slope limiter* used to construct the values (3.2) is correctly defined (see [21], [22] and [23]).

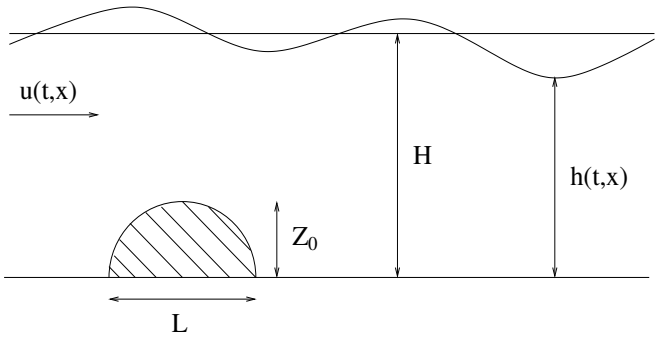
For the sake of simplicity, we consider in (3.1) only the first order discretization in time. It is standard to obtain higher order accuracy by applying Runge-Kutta methods for instance (see [9], [10] and its references).
The second order scheme (3.1) is validated by the numerical results obtained for the steady states of the Saint-Venant system (1.1)-(1.2).

4 Experimental configuration and numerical results

The situation studied in [26] is the one-dimensional flow of an incompressible fluid over an obstacle on the bottom of a smooth rectangular channel. The parameters of the problem are the stationary water depth H and the

F_0	Z_0	α	β	Re
0.62	$1.7cm$	0.179	0.58	$2.37 * 10^4$
0.66	$1.7cm$	0.26	0.58	$1.51 * 10^4$
0.64	$4.1cm$	0.68	0.58	$1.32 * 10^4$

mean velocity U, the kinematic viscosity of the fluid μ, the characteristic length of the obstacle L and its height Z_0. By combining these values, we define some specific dimensionless numbers: the Froude number $F_0 = \frac{U}{\sqrt{gH}}$, which relates the depth-averaged flow velocity to the characteristic wave propagation speed, in the long waves approximation; the blocking factor $\alpha = \frac{Z_0}{H}$, controlling the flow linearity in the absence of other perturbations (when α tends to zero, the flow becomes linear); the obstacle ratio $\beta = \frac{Z_0}{L}$, which can be interpreted as a control on the flow hydrostaticity (by analogy with the long waves approximation); the Reynolds number $Re = \frac{HU}{\mu}$, for the simple case of an ideal rectangular channel.
We note that the definitions above correspond to simplifications adapted to the theoretical system described by (1.3)-(1.4), when the fluid density and the channel width are formally reduced.



The experiments have been carried out in a channel of length $12\,m$, inclined at slope 0.002 and entirely made of glass. Two obstacle shapes are considered, a smooth Gaussian bump given by $Z(x) = Z_0 \exp\left(-\frac{x^2}{2L^2}\right)$, with ratio $\beta = 0.23$, and a semi-circular bump with ratio $\beta = 0.58$ (this last shape is commonly used for experimental analyses). Two sizes $Z_0 = 1.7\,cm$ and $Z_0 = 4.1\,cm$ are fixed for each shape, which allow to access a wider range of α values ($0.147 \leq \alpha \leq 0.7$). The tests are performed in subcritical regime ($F_0 < 1$) and the obstacles are placed in the channel so that fully developed turbulent flow conditions are attained before the obstacle.

We present some numerical results corresponding to the test cases illustrated in the above table, for which experimental data are available to make

direct comparisons. The pairs of figures reproduce the level of the water surface and the local Froude number in the vicinity of the obstacle, obtained at time $T = 150\,s$ (when the flow has become completely stationary), normalized with respect to the obstacle size. According to the experiments, the flow is classified by means of the free-surface profile and three different regimes are observed as function of the blocking factor. The boundaries of each regime, in terms of α values, are essentially independent of the obstacle shape.
Several tests have been performed also varying the Strickler's coefficient K introduced in (1.4), to evaluate the physical adequacy of the friction law (refer to [19] for a more precise discussion). The nonuniform mesh used for the numerical simulations is refined around the obstacle or in regions of stiff topographical variations, with a minimal mesh size of $5 * 10^{-4}\,m$.
The subcritical flow downstream of the obstacle displays many of the features of a supercritical flow behind a sluice gate with ensuing hydraulic jumps, implying a certain independence of the upstream conditions. The treatment of the boundary conditions, which turns out to be crucial for the numerical accuracy, is provided by the method developed in [3].

Regime I $(\alpha \leq 0.25)$

The graphics of the local Froude number show a transition from supercritical to subcritical flow downstream of the obstacle, through a hydraulic jump (Figure 2); the flow remains subcritical downstream the jump. The lower boundary of this regime is expected to be the value of α below which a classical subcritical regime should occur.

The results of the experimental test reveal the presence of a wave train on the free-surface profile (Figure 1). Moreover, some photographs taken during the experiments show bubbles arising on the crest of the hydraulic jump, as consequence of the turbulent motion. These phenomena are not taken into account in the mathematical model (1.3)-(1.4) and cannot be obtained numerically (we refer to [24] for further analysis).

Regime II $(0.26 \leq \alpha < 0.68)$

The graph of the local Froude number shows that a region of supercritical flow appears downstream the obstacle, followed by a transition from supercritical to subcritical flow through a hydraulic jump (Figure 4); however, in this regime, the local Froude number undergoes further transitions and its values remain close to $F = 1$ downstream the jump.

In the experimental tests, the local Froude number is oscillating around the critical value and the wave train on the free-surface profile becomes a

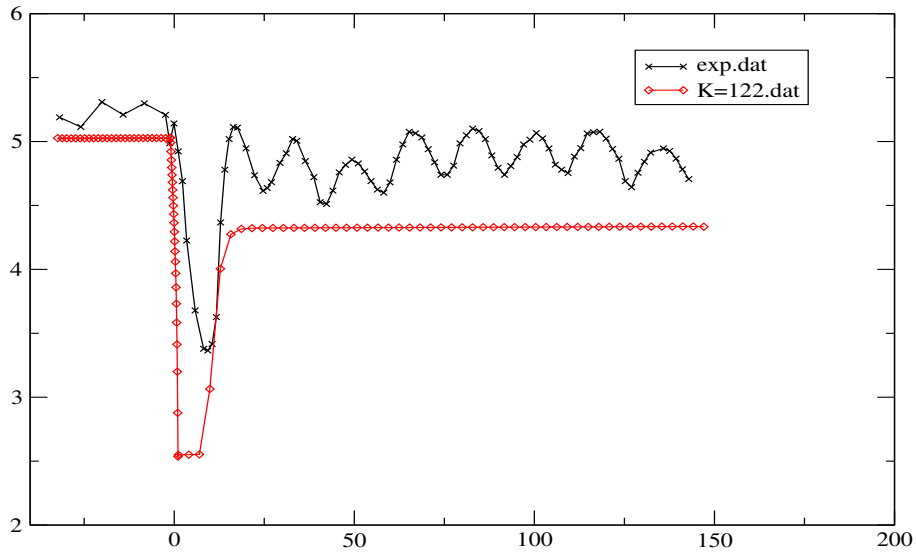


Figure 1: water surface ($\alpha = 0.179$)

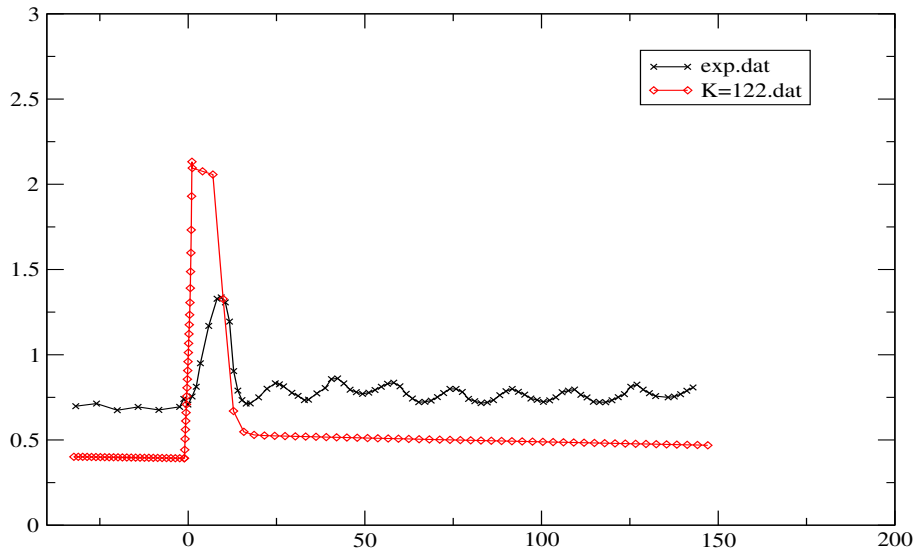


Figure 2: local Froude number

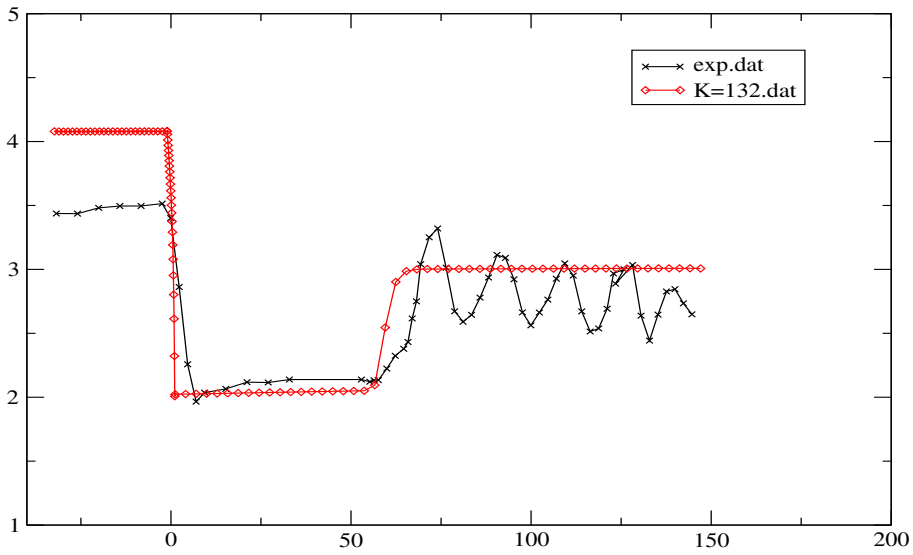


Figure 3: water surface ($\alpha = 0.26$)

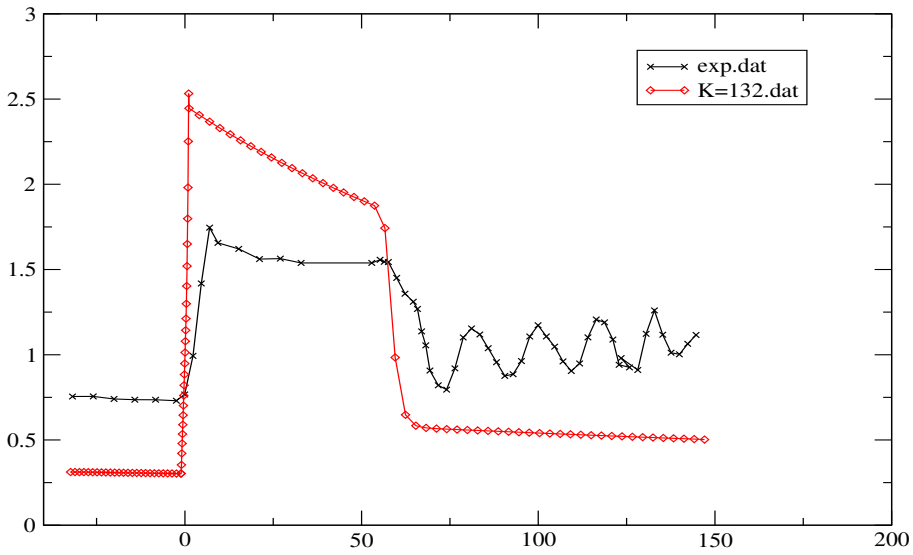


Figure 4: local Froude number

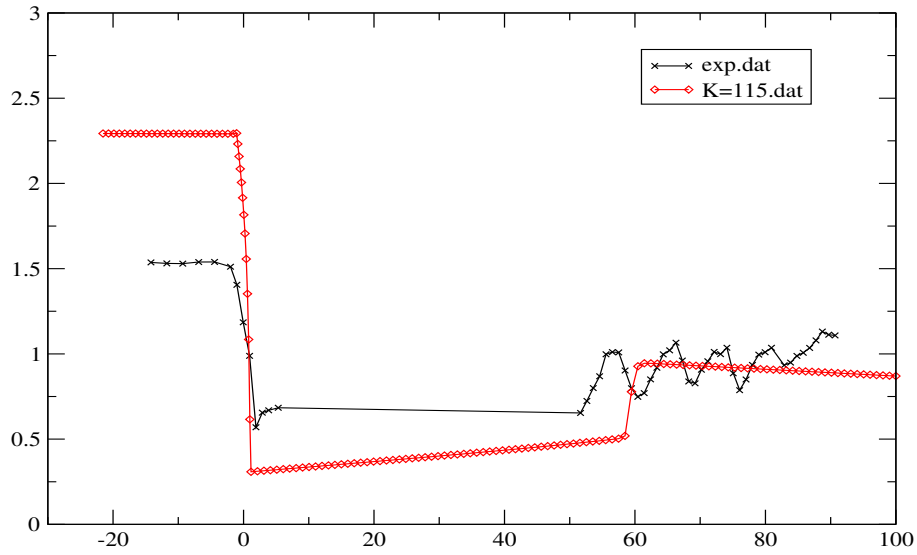


Figure 5: water surface ($\alpha = 0.68$)

series of triangular hydraulic jumps. This flow regime is essentially three-dimensional, the friction on the channel walls plays a significant role for the energy dissipation across the jumps (inducing water deceleration) and confinement effects seem to control the behaviour of the flow. Theoretical values obtained for the water depth (Figure 3) are higher than the experimental ones, suggesting that the numerical results are necessarily inaccurate.
Nevertheless, as two-dimensional side wall effects have not been considered in the model (1.3)-(1.4), it seems impossible to reproduce all the phenomena with one-dimensional simulations.

Regime III ($\alpha \geq 0.68$)

The experiments performed until $\alpha = 0.7$ show that the flow remains supercritical downstream the obstacle; in other words, no hydraulic jumps arise in this regime. Small perturbations appear on the free-surface profile, in form of a supercritical wave train located downstream the obstacle.
We remark a rather good agreement with the experimental results concerning the position of hydraulic jumps and the length of the supercritical region predicted by the shallow water equations (1.3)-(1.4), when friction and viscosity are included.

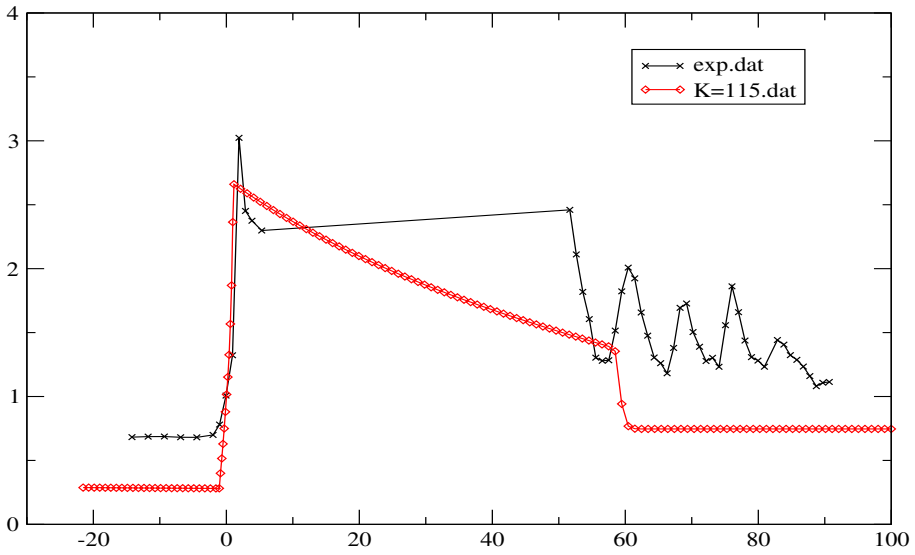


Figure 6: local Froude number

Acknowledgments

This work was partially supported by the ACI "Modélisation de processus hydrauliques à surface libre en présence de singularités" (Ministre de la Recherche - France).

We would like to thank François Bouchut (DMA-ENS) and Marie-Odile Bristeau (INRIA) for several helpful discussions.

References

[1] P.G. Baines, *Topographic effects in stratified flows*, Cambridge Monographs on Mechanics, Cambridge University Press, Cambridge, 1995

[2] A. Chinnayya, A.Y. Le Roux, A new general Riemann solver for the shallow water equations with friction and topography, preprint

[3] B. Coussin, M.O. Bristeau, Boundary conditions for the shallow water equations solved by kinetic schemes, *INRIA Report*, RR-4282 (2001)

[4] R. Eymard, T. Gallouët, R. Herbin, *Finite Volume Methods*, Handbook of numerical analysis, vol. VIII, P.G.Ciarlet and J.L.Lions editors, Amsterdam, North-Holland, 2000

[5] J.F. Gerbeau, B. Perthame, Derivation of viscous Saint-Venant system for laminar shallow water; numerical validation, *Discrete Contin. Dyn. Syst. Ser.B*, **1** (2001), no. 1, 89-102

[6] P. Glaister, The efficient prediction of shallow water flows I. Theory, *Comput. Math. Appl.*, **32** (1996), no. 12, 129-143

[7] P. Glaister, The efficient prediction of shallow water flows II. Application, *Comput. Math. Appl.*, **33** (1997), no. 9, 115-141

[8] E. Godlewski, P.A. Raviart, *Hyperbolic systems of conservation laws*, Mathmatiques & Applications, no. 3/4, Ellipses, Paris, 1991

[9] S. Gottlieb, C.W. Shu, Total variation diminishing Runge-Kutta schemes, *Math. Comp.*, **67** (1998), no. 221, 73-85

[10] S. Gottlieb, C.W. Shu, E. Tadmor, Strong stability-preserving high-order time discretization methods, *SIAM Rev.*, **43** (2001), no. 1, 89-112

[11] T. Katsaounis, C. Simeoni, First and second order error estimates for the Upwind Interface Source method, preprint

[12] G.A. Lawrence, Steady flow over an obstacle, *J. Hydraul. Engng*, ASCE, **113** (1987), no. 8, 981-991

[13] R.J. LeVeque, *Numerical methods for conservation laws*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 1990

[14] K. Lowery, S. Liapis, Free-surface flow over a semi-circular obstruction, *Internat. J. Numer. Methods Fluids*, **30** (1999), 43-63

[15] A. Mangeney-Castelnau, J.P. Vilotte, M.O. Bristeau, B. Perthame, C. Simeoni, S. Yernini, Numerical modelling of avalanches based on Saint-Venant equations using kinetic schemes, submitted to *Journal of Geophysical Research*

[16] A. Paquier, *Modélisation et simulation de la propagation de l'onde de rupture de barrage*, PhD Thesis, Université J. Monnet, St-Étienne (France), 1995

[17] B. Perthame, C. Simeoni, A kinetic scheme for the Saint-Venant system with a source term, *Calcolo*, **38** (2001), no. 4, 201-231

[18] B. Perthame, C. Simeoni, Convergence of the Upwind Interface Source method for hyperbolic conservation laws, preprint

[19] Pukhnacheva, T.P., The problem of determining the roughness coefficient for a flow in an open channel, *Prikl. Mekh. Tekhn. Fiz.*, **38** (1997), no. 3, 93-98; translation in *J. Appl. Mech. Tech. Phys.*, **38** (1997), no. 3, 412-416

[20] A. J.C. de Saint-Venant, Théorie du mouvement non-permanent des eaux, avec application aux crues des rivières at à l'introduction des marées dans leur lit, *C.R. Acad. Sci. Paris*, **73** (1871), 147-154

[21] C.W. Shu, S. Osher, Efficient implementation of essentially nonoscillatory shock-capturing schemes, *J. Comput. Phys.*, **77** (1988), no. 2, 439-471

[22] C.W. Shu, S. Osher, Efficient implementation of essentially nonoscillatory shock-capturing schemes II, *J. Comput. Phys.*, **83** (1989), no. 1, 32-78

[23] C.W. Shu, High order ENO and WENO schemes for computational fluid dynamics. High-order methods for computational physics, pp. 439-582, *Lect. Notes Comput. Sci. Eng.*, **9**, Springer, Berlin, 1999

[24] S. Soares-Frazão, Y. Zech, Undular bores and secondary waves. Experiments and hybrid finite-volume modelling, *Journal of Hydraulic Research*, 2002

[25] P.L. Viollet, J.P. Chabard, P. Esposito, D. Laurance, *Mécanique des fluides appliquée*, Presses des Ponts et Chaussées, Paris, 1998

[26] F. Vigie, O. Eiff, D. Astruc, Experimental study of the flow regimes of subcritical channel flow over an obstacle, submitted to *Riverflow 2002*

Appendice

Modélisation numérique des avalanches basée sur les équations de Saint-Venant et utilisant des schémas cinétiques

(soumis à *Journal of Geophysical Research*)

Numerical modelling of avalanches based on Saint-Venant equations using kinetic scheme

A. Mangeney-Castelnau, J.P. Vilotte
Département de Modélisation Physique et Numérique, IPGP
F75252 - Paris Cedex 05 - France
e-mails: mangeney@ipgp.jussieu.fr, vilotte@ipgp.jussieu.fr

M.O. Bristeau
Unité de Recherche INRIA Rocquencourt - Projet M3N
B.P. 105 - 78153 Le Chesnay Cedex - France
e-mail: Marie-Odile.Bristeau@inria.fr

B. Perthame, C. Simeoni
Département de Mathématique et Applications, ENS
45, rue d'Ulm - 75230 Paris Cedex 05 - France
e-mails: perthame@dma.ens.fr, simeoni@dma.ens.fr

S. Yernini
Center for Development of Advanced Computing, Pune University Campus
Ganesh Khind - Pune 411 007 - India
e-mail: sudhakar@cdac.ernet.in

Abstract

Numerical modelling of debris avalanches is presented here. The model uses the long waves approximation, based on the small aspect ratio of debris avalanches, as in classical Saint-Venant models for shallow water. Depth-averaged equations using this approximation are derived in a reference frame linked to the topography. Debris avalanches are treated here as a dry granular flow with Coulomb-type behavior.

The numerical finite volume method uses a kinetic scheme, based on the description of the microscopic behavior of the system, to define numerical fluxes at the interfaces of a finite element mesh. The main advantage of this method is to preserve the height positivity. The originality of the present scheme stands in the introduction of a Dirac distribution of particles at the microscopic scale, in order to describe the stopping of a granular mass when the driving forces are under the Coulomb threshold. Comparisons with analytical solutions for dam-break problems show the efficiency of the method to deal with significant discontinuities. The ability of the model to describe debris avalanche behavior is illustrated here by schematic 1D numerical simulations of an avalanche over simplified topography. Coulomb-type behavior with constant and variable friction angle are compared in the framework of this simple example. Numerical tests show that such approach does not only provide insights into the flowing and stopping stage of the granular mass but it also allows us to observe interesting behaviors, such as the existence of a fluidized zone behind a stopped granular mass in specific situations, suggesting the presence of horizontal surfaces in the deposited mass.

Key-words: avalanche modelling, Coulomb friction, Saint-Venant equations, finite volume kinetic scheme.

1 Introduction

Granular avalanches such as rock or debris flows regularly cause large amounts of human and material damages. The numerical simulation of granular avalanches should provide a useful tool for investigating, within realistic geological contexts, the dynamics of these flows and their arrest phase and for improving the risk assessment of such natural hazards. Computational models must however be able to correctly capture several features such as the formation of interacting surges [*Iverson*, 1997].

The physics and rheology of granular avalanches are indeed challenging problems and the subject of an active research [e.g. *Hunt*, 1994; *Laigle and Coussot*, 1997; *Arattano and Savage*, 1994; *Macedonio and Pareschi*, 1992; *Cheng-Lun et al.*, 1996; *Whipple*, 1997; *Iverson*, 1997]. Despite the lack of a clear physical understanding of avalanche flows, useful basic behaviors of granular avalanches can be derived from experimental approaches [e.g. *Pouliquen*, 1999; *Douady et al.*, 1999]. During a granular avalanche, the characteristic length in the flowing direction is generally much larger than the vertical one, i.e. the avalanche thickness. Such a long waves scaling argument has been widely used in the derivation of continuum flow models for

granular avalanches [e.g. *Hunt*, 1985; *Iverson*, 1997, *Iverson and Denlinger*, 2001; *Jenkins*, 1999; *Jenkins et al.*, 1999; *Savage and Hutter*, 1989; *Hutter et al.*, 1995; *Harbitz*, 1998; *Douady et al.*, 1999]. This leads to depth-averaged models governed by generalized Saint-Venant equations. These models provide a fruitful paradigm for investigating the dynamics and the extent of granular avalanches in the presence of smooth topography [e.g. *Hutter et al.*, 1995; *Naaim et al.*, 1997; *Pouliquen*, 1999]. It is worth to mention that by construction these flow models do not address the problem of the initiation and destabilization phases of an avalanche, see *Aranson and Tsimring* [2001] for models describing these processes. Granular surface flow models are closely related to other Saint-Venant models used in ocean and hydraulic engineering to describe both wave propagation, hydraulic jumps and open channel flows among others.

Without going into detailed rheological assumptions, which would be rather uncertain due to the lack of a physical understanding of the actual forces acting in debris avalanches, it is of interest here to emphasize some of the characteristics that make such flows quite specific.

The first characteristic is that granular media have the ability to remain static (solid) even along an inclined surface. This observation is related by Coulomb to some macroscopic solid-like friction and the system is able to flow only when the driving force reaches a critical value. In classical Coulomb's friction, the friction coefficient remains constant [e.g. *Hutter et al.*, 1995; *Naaim et al.*, 1997]. More evolved friction models, which assume a friction coefficient that depends on both the avalanche mean velocity and thickness, has been recently proposed [e.g. *Pouliquen*, 1999; *Douady et al.*, 1999] based on laboratory experiments and theoretical assumptions. These models have been shown quite useful to explain the geometry of the flow in the presence of topography as well as the observed runout of granular avalanches. In both cases, the existence of a macroscopic friction threshold leads to nonsmooth dynamics that has to be handled within appropriate mathematical and numerical formulations.

The second characteristic is that topography along which the avalanche is flowing can be quite steep and rough. Long waves approximation has therefore to be derived in a reference frame locally tangent to the bedrock or to the free surface of the flow, in contrast with the Galilean reference frame used in classical Saint-Venant models for hydraulic engineering. The definition of such a tangent frame of reference is not obvious for a realistic earth topography and is still a challenge problem. Strong variations of the bottom topography introduce a stiff source term in the governing flow equations, that strongly influences the properties of the models and leads to the occurrence of new steady states. When taking into account a Coulomb-type friction and

a realistic bottom topography, the source term becomes not only stiff during the flow but also nonsmooth and shocks are expected to develop in finite time regardless of the initial conditions. These difficulties have long hindered the development of realistic models for debris avalanches.

Computational methods developed in geophysics for solving the governing conservation laws of debris avalanches have mostly focused on the resolution of shock waves and surges. They are often based on fractional step methods and high resolution approximate Riemann solvers, like the Harten-Lax-vanLeer (HLL) solver [e.g. *Toro*, 1997]. Most of these methods are based on conservative nonoscillatory finite differences [e.g. *Gray et al.*, 1999; *Wieland et al.*, 1999; *Tai*, 2000; *Tai et al.*, 2002] or finite volumes which have the nice property of being conservative with respect to the flow height [e.g. *Naaim et al.*, 1997; *Laigle and Coussot*, 1997; *Denlinger and Iverson*, 2001]. They are based on an Eulerian formulation, a Lagrangian formulation [e.g. *Zwinger*, 2000] or a Lagrangian-Eulerian operator splitting [e.g. *Mangeney et al.*, 2000]. Even though these Riemann methods present significant improvements over the early Lagrangian finite difference methods [e.g. *Savage and Hutter*, 1989, 1991; *Greve et al.*, 1994], they do not preserve height positivity. Specific numerical development has to be introduced in the wetting-drying transition, where the system loses hyperbolicity, and generally an artificial small height has to be introduced in the regions where no fluid is present (see *Heinrich et al.*, 2001).

We consider here an alternative numerical scheme to compute debris avalanches, based on the kinetic interpretation of the system. Kinetic schemes have been proposed by *Audusse et al.* [2000] and *Bristeau et al.* [2001] to compute Saint-Venant equations in hydraulic problems. A survey of the theoretical properties of these schemes can be found in *Perthame* [2002]. Recently, kinetic schemes have been extended to include stiff source terms [e.g. *Botchorishvilli et al.*, 2000; *Perthame and Simeoni*, 2001]. Kinetic schemes have been shown to preserve the height positivity and to be able to treat the wetting-drying transition. However, classical kinetic schemes do not allow liquid-solid transitions, associated with a nonsmooth friction. The idea of the present scheme is to introduce a "zero-temperature" kinetic approximation when the driving force is under the Coulomb threshold.

We first present the basic equations and the conservation laws which govern the evolution of granular avalanches along a realistic topography. In particular, by using classical scaling arguments for surface flows, we derive the depth-averaged Saint-Venant equations in a reference frame linked to the bed topography. Then we review some minimal assumptions, inspired from experiments, on the characteristics of the frictional behavior of granular avalanches. Then we present a numerical scheme based on a finite volume

approximation of the governing set of conservation laws. At this stage, we introduce a kinetic solver which takes into account the existence of a friction threshold. The accuracy of this kinetic scheme is assessed against the classical dam-break problem over an inclined plane. Finally, some of the potentialities of the kinetic scheme are illustrated by simulating a debris avalanche over a simple bed topography. Comparisons between models with constant and nonconstant friction are discussed based on the runout, the shape of the deposit and the mechanism of the stopping phase.

2 Equations

Debris avalanches are described here within a continuum theoretical framework, as an incompressible material with constant density [e.g. *Savage and Hutter*, 1989; *Iverson and Denlinger*, 2001]. The evolution is therefore governed at time $t \geq 0$ by the mass and momentum conservation laws,

$$\boldsymbol{\nabla} \cdot \mathbf{u} = 0, \tag{2.1}$$

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \boldsymbol{\nabla} \mathbf{u} \right) = -\boldsymbol{\nabla} \cdot \boldsymbol{\sigma} + \rho \mathbf{g}, \tag{2.2}$$

where $\mathbf{u}(x, y, z, t) = (u(x, y, z, t), v(x, y, z, t), w(x, y, z, t))$ denotes the three-dimensional velocity vector inside the avalanche in a (x, y, z)-coordinate system that will be discussed later, $\boldsymbol{\sigma}(x, y, z, t)$ is the Cauchy stress tensor, ρ is the mass density and $\mathbf{g}$ the gravitational acceleration. The bottom boundary, or bed, is described by a surface $\psi_b(x, y, z, t) = z - b(x, y) = 0$ and the free surface of the flow by $\psi_s(x, y, z, t) = z - s(x, y, t) = z - b(x, y) - h(x, y, t) = 0$, where $h(x, y, t)$ is the depth of avalanche layer.
A kinematic boundary condition is imposed on the free and bed surfaces, that specifies that mass neither enters nor leaves the free surface or the base,

$$\frac{\mathrm{d}\psi_s}{\mathrm{d}t}|_s = \left(\frac{\partial \psi_s}{\partial t} + \mathbf{u} \cdot \boldsymbol{\nabla} \psi_s \right) |_s = 0, \tag{2.3}$$

$$\frac{\mathrm{d}\psi_b}{\mathrm{d}t}|_b = \left(\frac{\partial \psi_b}{\partial t} + \mathbf{u} \cdot \boldsymbol{\nabla} \psi_b \right) |_b = 0, \tag{2.4}$$

as well as a stress free-boundary condition at the surface, neglecting the atmospheric pressure,

$$\boldsymbol{\sigma} \cdot \mathbf{n}_s = 0, \tag{2.5}$$

where $\mathbf{n_s}$ denotes the unit vector normal to the free surface.

Depth-averaging of these equations and some shallow flow assumptions require the choice of an appropriate coordinate system. During the flow, the

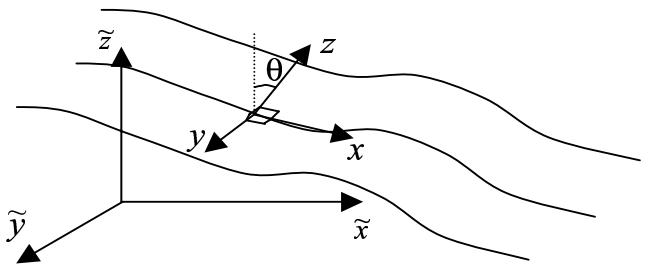


Figure 1: Reference frame (x, y, z) linked to the topography and Galilean reference frame $(\tilde{x}, \tilde{y}, \tilde{z})$ with θ the steepest slope angle.

avalanche thickness is much smaller than its extent parallel to the bed. In the case of significant slopes, the shallow flow assumption is more significant in a reference frame linked to the topography and the classical shallow water approximation relating horizontal and vertical direction is not appropriate. As in *Denlinger and Iverson* [2001], the equations are written here in terms of a local orthogonal Cartesian coordinate system in which the z-coordinate is normal to the local topography. We define a local x-axis corresponding to the projection of an arbitrary fixed $\tilde{x}$-direction in the local tangent plane to the topography and $\mathbf{y} = \mathbf{z} \wedge \mathbf{x}$ (Figure 1).
Note that the choice of an appropriate reference frame is not straightforward when dealing with real complex topography and may lead to nonorthonormal coordinate systems as in *Heinrich et al.* [2001], *Assier-Rzadkiewicz et al.* [2000] and *Sabot et al.* [1998]. The variation in space of a local coordinate system introduces errors in the calculation of the derivatives and require slow variation of the bedrock. The equations developed in a coordinate system linked to the topography are not directly applicable in a fixed reference frame as it was performed by *Naaim et al.* [1997] and *Naaim and Gurer* [1998]: appropriate rotations have to be used to transform properly topography-linked equations in a fixed reference frame [see *Douady et al.*, 1999].

In the reference frame linked to the topography (Figure 1), the equations of mass and momentum in the x- and y-direction, derived by integration of the Navier-Stokes equations (2.1)-(2.2) with boundary conditions (2.3)-(2.4)

and (2.5), read

$$\frac{\partial h}{\partial t} + \operatorname{div}(h\bar{\mathbf{u}}) = 0, \tag{2.6}$$

$$\frac{\partial}{\partial t}(h\bar{u}) + \frac{\partial}{\partial x}(h\overline{u^2}) + \frac{\partial}{\partial y}(h\overline{uv}) = \gamma_x gh + \frac{1}{\rho}\frac{\partial}{\partial x}(h\overline{\sigma}_{xx}) + \frac{1}{\rho}\frac{\partial}{\partial y}(h\overline{\sigma}_{xy}) + \frac{1}{\rho}T_{tx}, \tag{2.7}$$

$$\frac{\partial}{\partial t}(h\bar{v}) + \frac{\partial}{\partial x}(h\overline{uv}) + \frac{\partial}{\partial y}(h\overline{v^2}) = \gamma_y gh + \frac{1}{\rho}\frac{\partial}{\partial x}(h\overline{\sigma}_{xy}) + \frac{1}{\rho}\frac{\partial}{\partial y}(h\overline{\sigma}_{yy}) + \frac{1}{\rho}T_{ty}, \tag{2.8}$$

where $\bar{\mathbf{u}} = (\bar{u}, \bar{v})$ denotes the depth-averaged horizontal flow velocity in the reference frame (x, y, z) defined below, h is the fluid depth, γ_i are coefficients (function of the local slope) defining the projection of the gravity vector along the i-direction and $T_{ti} = \sigma_{iz}|_b$ represents the traction at the base of the flow. A small aspect ratio $\epsilon = H/L$, where H and L are characteristic dimensions along the z-axis and in the xOy plane respectively, is then introduced in the depth-averaged x- and y-equations (2.7)-(2.8) and in the nonaveraged z-equation obtained from the z-projection of equation (2.2). An asymptotic analysis with respect to ϵ [e.g. *Gray et al.*, 1999] leads to neglect the acceleration normal to the topography in the z-equation, leading to

$$\sigma_{zz} = \rho g \gamma_z (h - z), \tag{2.9}$$

where $\gamma_z = \cos\theta$, with θ defined as the angle between the vertical axis and the normal to the topography (Figure 1). The shape of the vertical profile of the horizontal velocity in debris avalanche flows is still an open question. The conservation of the initial stratigraphy, sometimes observed in the deposits of a debris avalanche, has led to the assumption that all the deformation is essentially located in a fine boundary layer near the bed surface, so that the horizontal velocity is approximately constant over the depth [e.g. *Savage and Hutter*, 1989; *Naaim et al.*, 1997]. More recently, laboratory experiments on granular flows suggest a linear profile of the horizontal velocity [e.g. *Azanza*, 1998; *Douady et al.*, 1999]. A weak influence of the vertical profile of the horizontal velocity has been observed by *Pouliquen and Forterre* [2002] for granular flows over inclined plane. We note that, in the locally tangent frame of reference, simple assumptions for the velocity profile (i.e. constant or linear profile) can be made unlike in the Galilean fixed reference frame. We assume here a vertically constant velocity so that $\overline{u_i u_j} = \bar{u}_i \bar{u}_j$. In the following, the overline will be dropped and (u, v) will represent the mean velocity field.

3 Flow and friction law

3.1 Simple friction law

A relation deduced from the mechanical behavior of the material has to be imposed between $\overline{\sigma}_{ij}$, T_{ti}, $\mathbf{u}$ and h in order to close equations (2.6)-(2.7) and (2.8). We consider here the minimal model, by using the hydrostatic assumption, i.e. $\overline{\sigma}_{ij} = 0,\ i \neq j$ and $\sigma_{xx} = \sigma_{yy} = \sigma_{zz}$. The depth-averaged mass is then considered as an effective material submitted to empirical frictions, introduced in the traction term T_{ti} in a way similar to the experimental approach by *Pouliquen* [1999].

Dissipation in granular materials is generally described by means of a Coulomb-type friction law, relating the tangential traction T_t on the bed to the normal stress $T_n = \sigma_{zz}|_b$ through a factor $\mu = \tan\delta$ involving the dynamic bed friction angle δ, namely

$$\|T_t\| \leq \sigma_c = \mu \|T_n\|,$$

which is acting opposite to the velocity. The value of σ_c defines the upper bound of the admissible stresses. In the coordinate system considered above, using the equation (2.9), we have

$$\sigma_c = \mu \rho g \gamma_z h.$$

The resulting Coulomb-type behavior can be summarized as follows,

$$\|T_t\| \geq \sigma_c \ \Rightarrow\ T_{ti} = -\sigma_c \frac{u_i}{\|\mathbf{u}\|}, \tag{3.1}$$

$$\|T_t\| < \sigma_c \ \Rightarrow\ \mathbf{u} = 0, \tag{3.2}$$

with $i = x, y$.

3.2 Flow variable friction law

Laboratory experiments [see *Pouliquen*, 1999] have shown that laws involving constant friction angle are restricted to granular flows over smooth inclined planes or flows over rough bed with high inclination angles. The assumption of constant friction angle seems to fail for granular flows over rough bedrock in a range of inclination angles for which steady uniform flows can be observed [*Pouliquen*, 1999]. In this range, the frictional force is able to balance the gravity force, indicating a shear rate dependence.
Pouliquen [1999] proposed to choice an empirical friction coefficient μ as

function of the Froude number $\|u\|/\sqrt{gh}$ and the thickness h of the granular layer, in the form

$$\mu(\|u\|, h) = \tan\delta_1 + (\tan\delta_2 - \tan\delta_1)\exp\left(-\beta\frac{h}{d}\frac{\sqrt{gh}}{\|u\|}\right), \tag{3.3}$$

where δ_1, δ_2 and d are characteristics of the material which can be measured from the deposit properties. In particular, d is a characteristic length of the friction law, which is scaled on the mean diameter of particles; in the case of spherical glass particles used in these laboratory experiments d is of the order of the diameter of the beads and $\beta = 0.136$ [*Pouliquen*, 1999]. Equation (3.3) provides a friction angle, ranging between two values δ_1 and δ_2, depending on the values of the velocity and the flow thickness. The friction is higher for small values of the thickness and high values of the velocity, contrary to the function proposed by *Gray et al.* [1999] where lowest elevations (i.e. the rear and the front) are subject to small friction. What this empirical law means in terms of microscopic forces is still an open problem. Hydraulic models using this flow law has been shown to be able to predict the spreading of a granular mass from released to deposit [*Pouliquen and Forterre*, 2002].

Finally, if $\|T_t\| \geq \sigma_c$, the granular mass is flowing following the dynamical equations

$$\frac{\partial}{\partial t}(h\bar{u}) + \frac{\partial}{\partial x}\left(h\bar{u}^2\right) + \frac{\partial}{\partial y}(h\bar{u}\bar{v}) = \gamma_x gh + \frac{\partial}{\partial x}(g\gamma_z\frac{h^2}{2}) - \mu g\gamma_z h\frac{u_x}{\|\mathbf{u}\|}, \tag{3.4}$$

$$\frac{\partial}{\partial t}(h\bar{v}) + \frac{\partial}{\partial x}(h\bar{u}\bar{v}) + \frac{\partial}{\partial y}\left(h\bar{v}^2\right) = \gamma_y gh + \frac{\partial}{\partial y}(g\gamma_z\frac{h^2}{2}) - \mu g\gamma_z h\frac{u_y}{\|\mathbf{u}\|}, \tag{3.5}$$

and, if $\|T_t\| < \sigma_c$, the granular mass stops, i.e. $\mathbf{u} = 0$.

4 Numerical Model

4.1 Finite volume method

The model developed here is based on the classical finite volume approach for solving hyperbolic systems, using the concept of cell-centered conservative quantities. This type of methods requires the formulation of the equations in terms of conservation laws. The system of equations (2.6) and (3.4)-(3.5) can be rewritten as

$$\frac{\partial \mathbf{U}}{\partial t} + \operatorname{div}\mathbf{F}(\mathbf{U}) = \mathbf{B}(\mathbf{U}), \tag{4.1}$$

with

$$\mathbf{U} = \begin{pmatrix} h \\ q_x \\ q_y \end{pmatrix}, \quad \mathbf{F}(\mathbf{U}) = \begin{pmatrix} q_x & q_y \\ \frac{q_x^2}{h} + \frac{g}{2}h^2 & \frac{q_x q_y}{h} \\ \frac{q_x q_y}{h} & \frac{q_y^2}{h} + \frac{g}{2}h^2 \end{pmatrix}, \tag{4.2}$$

$$\mathbf{B}(\mathbf{U}) = \begin{pmatrix} 0 \\ gh\gamma_x - \sigma_{xz}|_b \\ gh\gamma_y - \sigma_{yz}|_b \end{pmatrix}, \tag{4.3}$$

where $\mathbf{q} = h\mathbf{u}$ is the material flux.

The equations are discretized on general triangular grids with a finite element data structure, using a particular control volume which is the median based dual cell (Figure 2a). The finite element grid is appropriate to describe variable topography and refinement is performed when strong topographic gradients occur. Dual cells C_i are obtained by joining the centers of mass of the triangles surrounding each vertex P_i. We also use the following notations:

- K_i, set of nodes P_j surrounding P_i,
- $|C_i|$, area of C_i,
- Γ_{ij}, boundary edge belonging to cells C_i and C_j,
- L_{ij}, length of Γ_{ij},
- $\mathbf{n}_{ij}$, unit normal to Γ_{ij}, outward to C_i.

If P_i is a node belonging to the boundary Γ of the numerical domain, we join the centers of mass of the triangles adjacent to the boundary to the middle of the edge belonging to Γ (see Figure 2b).

Let Δt denote the time-step, $\mathbf{U}_i^n$ is the approximation of the cell-average of the exact solution at time $t^n = n\Delta t$, $n \in \mathbb{N}$, i.e.

$$\mathbf{U}_i^n \simeq \frac{1}{|C_i|} \int_{C_i} \mathbf{U}(t^n, x) dx,$$

and $\mathcal{B}(\mathbf{U}_i^n)$ is the approximation of the cell-average of the exact source term,

$$\mathcal{B}(\mathbf{U}_i^n) \simeq \frac{1}{|C_i|} \int_{C_i} \mathbf{B}(\mathbf{U}(t^n, x)) dx.$$

Then, the finite volume scheme reads

$$\mathbf{U}_i^{n+1} = \mathbf{U}_i^n - \sum_{j \in K_i} \alpha_{ij} \, \mathcal{F}\left(\mathbf{U}_i^n, \mathbf{U}_j^n, \mathbf{n}_{ij}\right) - \Delta t \mathcal{B}\left(\mathbf{U}_i^n\right), \tag{4.4}$$

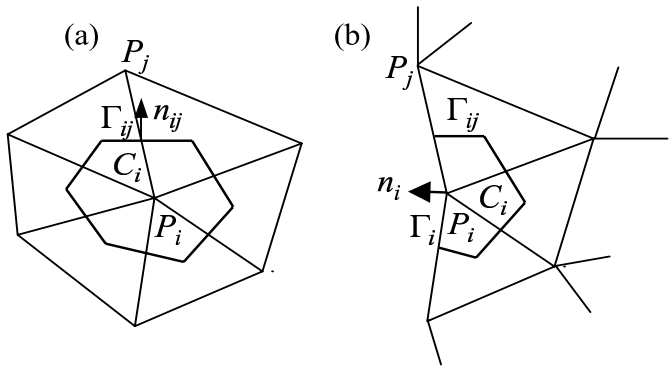


Figure 2: Triangular finite element mesh: (a) dual inner cell C_i, (b) dual boundary cell C_i.

with

$$\alpha_{ij} = \frac{\Delta t L_{ij}}{|C_{ij}|}, \tag{4.5}$$

and $\mathcal{F}\left(\mathbf{U}_i^n, \mathbf{U}_j^n, \mathbf{n}_{ij}\right)$ denotes an interpolation of normal components of the flux $\mathbf{F}(\mathbf{U})\cdot\mathbf{n}_{ij}$ along the edge Γ_{ij}. The treatment of the boundary conditions, namely the calculation of the boundary fluxes, using a Riemann invariant is addressed in *Bristeau et al.* [2001].

The main difficulty is to compute numerical fluxes at the control volume interfaces Γ_{ij} and the overall stability of the method requires some upwinding in the interpolation of the fluxes [see *Audusse et al.*, 2000]. The computation of these fluxes constitutes the major difference between the kinetic scheme used here and Godunov-type methods, which are usually very accurate for shock-capturing but not well suited to deal with vacuum front at the margins of the avalanche where the system looses hyperbolicity ($h = 0$ corresponding to dry soils). This drawback results from the lack of definable wave speeds in advance of a flow front. Many shock-capturing upwind schemes produce negative heights of water at these points and subsequently they break down or become unstable. An artificial small height of fluid in the whole domain has to be imposed to stabilize the scheme [e.g. *Mangeney et al.*, 2000]. *Tai* [2000] and *Tai et al.* [2002] overcome this imperfection by tracking the vacuum front. *Denlinger and Iverson* [2001] calculate the theoretical speed of a flow front using the Riemann invariant of the wave emanating from the front directed in the inner part of the mass.

We follow here an alternative approach to solve Saint-Venant equations

by using a kinetic solver, which is intrinsically able to treat vacuum and is appropriate to handle discontinuous solutions. These properties are of the highest importance for gravitational flow modelling. One further important property of this scheme is that it does not require any dimensional splitting. Kinetic schemes might be one of the best compromise between accuracy, stability and efficiency for the resolution of Saint-Venant equations [see *Audusse et al.*, 2000]. To our knowledge, this type of schemes has never been applied to avalanche flow modelling over slopping topography.

4.2 Kinetic formulation

The kinetic approach consists in using a fictitious description of the microscopic behavior of the system, in order to define numerical fluxes at the interface of an unstructured mesh. In fact, the macroscopic discontinuities disappear at the microscopic scale. We introduce here the main concepts of the kinetic scheme used for this model; a more complete description and deatils about its numerical implementation are done in *Audusse et al.* [2000] and *Bristeau et al.* [2001]. The scheme will be discussed by omitting the friction term, which is further introduced by using a semi-implicit scheme (see Section 4.3). In this method, fictitious particles are introduced and the equations are considered at the microscopic scale, where no discontinuities occur. A distribution function $M(t, x, y, \xi)$ of fictitious particles with microscopic velocity ξ is introduced to obtain a linear microscopic kinetic equation equivalent to macroscopic equation (4.1), with (4.2)-(4.3). The microscopic density of particles present at time t in the vicinity $\Delta x\, \Delta y$ of the position (x, y) and with velocity ξ is given by

$$M(t, x, y, \xi) = \frac{h(t, x, y)}{c^2}\, \chi\left(\frac{\xi - \mathbf{u}(t, x, y)}{c}\right), \tag{4.6}$$

with "fluid density" h, "fluid temperature" proportional to

$$c^2 = \frac{gh}{2}, \tag{4.7}$$

and $\chi(\omega)$ a positive even function defined on $\Re^2$ and satisfying

$$\int_{\Re^2} \chi(\omega) d\omega = 1, \quad \int_{\Re^2} \omega_i \omega_j \chi(\omega) d\omega = \delta_{ij}, \tag{4.8}$$

with δ_{ij} the Kronecker symbol and $\omega = (\omega_i, \omega_j)$. This function χ is assumed to be compactly supported, i.e.

$$\exists\, \omega_M \in \Re \ \text{ such that } \ \chi(\omega) = 0, \quad \text{for } |\omega| \geq \omega_M. \tag{4.9}$$

Note that the rectangular shape of the distribution function χ imposed for the fictitious particles would change in time if real particles where considered.

Simple calculations show that the macroscopic quantities are linked to the microscopic density function by the following relations,

$$\mathbf{U} = \int_{\Re^2} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M(t,x,y,\xi) d\xi, \tag{4.10}$$

$$\mathbf{F}(\mathbf{U}) = \int_{\Re^2} \begin{pmatrix} \xi \\ \xi \otimes \xi \end{pmatrix} M(t,x,y,\xi) d\xi, \tag{4.11}$$

$$\mathbf{B}_i(\mathbf{U}) = g\gamma_i \int_{\Re^2} \begin{pmatrix} 1 \\ \xi \end{pmatrix} \bigtriangledown_\xi M(t,x,y,\xi) d\xi, \tag{4.12}$$

with $i = x, y$. These relations imply that the nonlinear system (2.6) and (3.4)-(3.5) is equivalent to the linear transport equation for the quantity M, for which it is easier to find a simple numerical scheme with good properties,

$$\frac{\partial M}{\partial t} + \xi \cdot \bigtriangledown_{\mathbf{x}} M - g\,\boldsymbol{\gamma} \cdot \bigtriangledown_\xi M = Q(t,x,y,\xi), \tag{4.13}$$

for some collision term $Q(t,x,y,\xi)$ which satisfies

$$\int_{\Re^2} \begin{pmatrix} 1 \\ \xi \end{pmatrix} Q(t,x,y,\xi) d\xi = 0. \tag{4.14}$$

As usual, the collision term $Q(t,x,y,\xi)$ in this kinetic representation of the Saint-Venant equations, which relaxes the kinetic density to the microscopic equilibrium M, is neglected in the numerical scheme, i.e. at each time-step we project the kinetic density on M, which is a way to perform all the collisions at once and to recover the Gibbs equilibrium without computing it [see *Perthame and Simeoni*, 2001].

Finally, the discretization of this simple kinetic equation allows us to deduce an appropriate discretization of the macroscopic system. A simple upwind scheme is applied to the microscopic equation (4.13), leading to the formulation of the fluxes defined in equation (4.4),

$$\mathcal{F}(\mathbf{U}_i, \mathbf{U}_j, \mathbf{n}_{ij}) = \mathbf{F}^+(\mathbf{U}_i, \mathbf{n}_{ij}) + \mathbf{F}^-(\mathbf{U}_j, \mathbf{n}_{ij}), \tag{4.15}$$

$$\mathbf{F}^+(\mathbf{U}_i, \mathbf{n}_{ij}) = \int_{\xi \cdot \mathbf{n}_{ij} \geq 0} \xi \cdot \mathbf{n}_{ij} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M_i(\xi) d\xi, \tag{4.16}$$

$$\mathbf{F}^-(\mathbf{U}_j, \mathbf{n}_{ij}) = \int_{\xi \cdot \mathbf{n}_{ij} \leq 0} \xi \cdot \mathbf{n}_{ij} \begin{pmatrix} 1 \\ \xi \end{pmatrix} M_j(\xi) d\xi. \tag{4.17}$$

The simple form of the density function (here a rectangle-type function Π) allows analytical resolution of integrals (4.16)-(4.17) and gives the possibility

to write directly a finite volume formula, which therefore avoids using the extra variable ξ in the implementation of the code. The resulting numerical scheme is consistent and conservative. Furthermore, it is proved that the water height positivity is preserved under the Courant-Friedrichs-Levy condition [see *Audusse et al.*, 2000],

$$\Delta t \max\left(|u_i^n| + \omega_M c_i^n\right) \leq \frac{|C_i|}{\sum_{j\in K_i} L_{ij}}. \tag{4.18}$$

In comparison with flood modelling, avalanche modelling introduces a further difficulty relating to the property of granular media to be able to remain static (solid) even with an inclined free surface. This equilibrium is not intrinsically preserved by finite volume schemes and specific processing has to be introduced in the numerical scheme for the particular case of kinetic scheme, as it will be developed in the next section.

4.3 Friction

The friction is introduced in two steps. A first estimation of the numerical fluxes $\tilde{\mathbf{q}}_i^{n+1}$ is obtained by solving equation (4.4) without any friction term and the flow thickness h_i^{n+1} is calculated by solving explicitly the mass conservation (2.6). As friction does not change the direction of the velocity, we impose that the corrected flux $\mathbf{q}_i^{n+1}$ has the same direction of $\tilde{\mathbf{q}}_i^{n+1}$.
If the norm of the driving force $\tilde{\mathbf{q}}_i^{n+1}/\Delta t$ is lower than the Coulomb threshold $\sigma_c/\rho = \mu g \gamma_z\, h_i^{n+1}$, then the mass stops, i.e.

$$\|\tilde{\mathbf{q}}_i^{n+1}\| - \mu g \gamma_z\, h_i^{n+1}\Delta t \leq 0 \;\Rightarrow\; \mathbf{q}_i^{n+1} = 0. \tag{4.19}$$

On the other hand, if the driving force $\tilde{\mathbf{q}}_i^{n+1}/\Delta t$ is higher than the Coulomb threshold, then the norm of the friction term $\sigma|_b$ is equal to σ_c.
Following *Bristeau et al.* [2001], we introduce a semi-implicit treatment of the friction term. Equation (4.4), written in terms of the variable $\mathbf{q}$, reads

$$\mathbf{q}_i^{n+1} = \left(\|\tilde{\mathbf{q}}_i^{n+1}\| - \mu g \gamma_z\, h_i^{n+1}\Delta t\right) \frac{\tilde{\mathbf{q}}_i^{n+1}}{\|\tilde{\mathbf{q}}_i^{n+1}\|}. \tag{4.20}$$

This threshold-type behavior is generally not taken into account in numerical models, due to the resulting discontinuity in the velocity field. Generally, the magnitude of active and Coulomb friction forces is compared only for parts of the flow where $\mathbf{u} \neq 0$ [e.g. *Mangeney et al.*, 2000; *Eglit*, 1983].

Due to the possible space variations of h, classical kinetic schemes do not allow the mass stopping even though its velocity is equal to zero. In fact,

for the kinetic scheme based on a rectangle-type distribution function χ as in equation (4.6), perturbations propagate at velocity $\tilde{c} = \sqrt{gh}$ even though the fluid is at rest because the "temperature" is not equal to zero.
In our case, perturbations linked to the h-gradient of a nonflat free surface generate fluxes and the fluid never stops. On the opposite, the Coulomb criterium imposes that, under a given threshold, a perturbation (of the surface elevation, for example) do not propagates. It can be represented by a fluid at "temperature" equal to zero, so that the local speed of propagation of the disturbance relative to the moving stream is equal to zero. It can be obtained by using a Dirac distribution for the function χ.

The idea of the present scheme is to introduce a "zero-temperature fluid" with a Dirac-type density of particles M when the fluid is under the Coulomb threshold and a "nonzero-temperature fluid" using a rectangular-type density of particles when the fluid is over the Coulomb threshold, so we have

$$\|\tilde{\mathbf{q}}_i^{n+1}\| - \mu g \gamma_z \, h^{n+1} \Delta t < 0 \;\Rightarrow\; M(t,x,y,\xi) = h(t,x,y)\delta\left(\xi - \mathbf{u}(t,x,y)\right), \tag{4.21}$$

$$\|\tilde{\mathbf{q}}_i^{n+1}\| - \mu g \gamma_z \, h^{n+1} \Delta t \geq 0 \;\Rightarrow\; M(t,x,y,\xi) = \frac{h(t,x,y)}{c^2}\,\chi\left(\frac{\xi - \mathbf{u}(t,x,y)}{c}\right), \tag{4.22}$$

where the rectangular function χ given by *Bristeau et al.* [2001] reads

$$\chi(\omega) = \frac{1}{12}\Pi_{|\omega_i| \leq \sqrt{3}}, \quad i = 1, 2. \tag{4.23}$$

The expression of the flux related to the edge Γ_{ij} in the mass conservation equation using (4.16) then reads

$$\|\tilde{\mathbf{q}}_i^{n+1}\| - \mu g \gamma_z \, h^{n+1} \Delta t < 0 \;\Rightarrow\; \mathbf{F}_h^+(\mathbf{U}_i, \mathbf{n}_{ij}) = h_i u_{i,n} Y(u_{i,n}), \tag{4.24}$$

$$\|\tilde{\mathbf{q}}_i^{n+1}\| - \mu g \gamma_z \, h^{n+1} \Delta t \geq 0, \;\Rightarrow\; \mathbf{F}_h^+(\mathbf{U}_i, \mathbf{n}_{ij}) = \frac{1}{2} h_i u_{i,n} + \frac{\sqrt{3}}{4} h_i c_i + \frac{1}{4\sqrt{3}} h_i \frac{u_{i,n}^2}{c_i}, \tag{4.25}$$

where Y is the Heaviside distribution and $u_{i,n}$ is the velocity in the normal direction of the edge Γ_{ij}. Similar expressions are obtained for $\mathbf{F}^-(\mathbf{U}_j, \mathbf{n}_{ij})$.
Note that the Dirac distribution does not allow us to recover the momentum equation. In fact, the flux calculated for the momentum equation using this function reads

$$\mathbf{F}_m^+(\mathbf{U}_i, \mathbf{n}_{ij}) = h_i u_{i,n}^2 Y(u_{i,n}), \tag{4.26}$$

without the pressure gradient due to the zero-temperature fluid. However, when the fluid is under the Coulomb threshold, the momentum equation

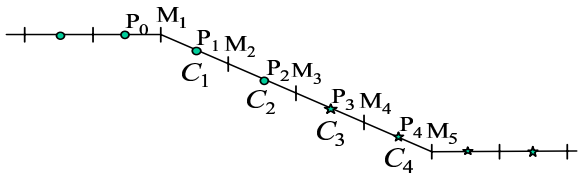


Figure 3: 1D mesh and dual cell C_i with center P_i (full circles denote the points under the Coulomb threshold and stars the points above the Coulomb threshold).

is replaced by $\mathbf{q} = 0$, so that the Dirac-type function is only used for the calculation of the fluxes in the mass conservation equation.

The first step of the numerical scheme is to evaluate the grid points that are under the Coulomb threshold using $\tilde{\mathbf{q}}_i^{n+1}$. We look at the simple 1D case (Figure 3) where the points P_0, P_1 and P_2 are under the Coulomb threshold (full circles) and the points P_3 and P_4 are above this threshold (stars). In order to obtain the flux $F_{h,i} = F_h^+(P_{i-1}) + F_h^-(P_i)$ at the interface M_i allowing to satisfy conservation laws, the same distribution function has to be used in both side of the interface: a rectangular distribution is imposed if one of the two points P_i or P_{i-1} is above the Coulomb threshold and a Dirac distribution elsewhere. As a result, the flux through the interface M_3 is calculated using a rectangular function whereas the flux through the interface M_2 is calculated using the Dirac function. The propagation of the h-gradient is then allowed to the right where the fluid is above the Coulomb threshold and forbidden to the left where the fluid is under the Coulomb threshold, recalling the typical solid-fluid transition of granular material. Numerical tests show that this method is mass conservative.

The resulting 2D scheme consists in evaluating at time t the points under the Coulomb threshold and in calculating at time $t + dt$ the flux F_h through the interface M_{ij} of a cell C_i

- using the rectangular distribution if one of the two points P_i and P_j situated on both sides of this interface is above the Coulomb threshold;
- using a Dirac distribution if the two points P_i and P_j are under the Coulomb threshold.

The numerical method is illustrated on the 2D mesh presented in Figure 4 where the points M_1, M_2, M_3, P_2, M_{10}, M_{11} surrounding the point P_1 are under the Coulomb threshold. The fluxes F_h through the interfaces of the

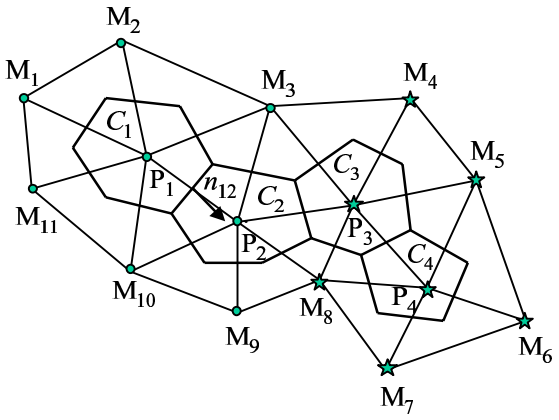


Figure 4: Triangular mesh and dual cell C_1, C_2, C_3 and C_4 (full circles denote the points under the Coulomb threshold and stars the points above the Coulomb threshold).

cell C_1 is then calculated using the Dirac distribution, whereas in cell C_4 all the fluxes are calculated using the rectangular distribution. For the cell C_2, the surrounding points P_3 and M_8 being above the Coulomb threshold, the fluxes F_h through the edges cutting P_2M_8, P_2P_3 are calculated using the rectangular distribution, while the fluxes F_h through the edges cutting P_2P_1, P_2M_3, P_2M_9, P_2M_{10} are calculated using the Dirac distribution. With this scheme, verifying the mass conservation at the machine accuracy, the fluid is able to stop.

5 Validation

The precision and performance of the numerical model is tested by comparing numerical results with those of an analytical solution, which takes into account a Coulomb-type friction on the base of the flow, provided the angle of friction is smaller than the slope angle and the fluid never stops on the inclined plane [see *Mangeney et al.*, 2000].

The test case consists of the instantaneous release of a fluid mass of $1\,m$ high on a dry flat bottom, infinitely long in the negative x-direction. The numerical domain ranges from $0\,m$ to $2000\,m$. Note that the aspect ratio of the geometry considered here is $\epsilon = 10^{-3}$, so that the long waves approximation is valid. All 1D numerical experiments are carried out with

the 2D model using the same number of points in the transversal direction (101 points with the same space-step as in the flow direction).
From Figure 5 and Figure 6, showing the comparison between analytical and numerical solution for two grid steps ($dx = 20\,m$ and $dx = 2\,m$), it can be observed that the numerical model provides a good representation of the dam-break problem as well without and with friction law. The main difference between analytical and numerical results is located at the front position and at the corner of the dam, as it was observed in *Mangeney et al.* [2000] with a Godunov-type numerical model. Note that the deviation from the analytical solution is qualitatively the same with the Godunov-type model and the kinetic model: the corner at the left discontinuity is rounded and the position of the front is lower than the position of the analytical front after a few seconds: the shock is smoothed, as usual with a first order scheme, as it was observed in *Audusse et al.* [2000].
Finally, the results are expressed in terms of the mean relative error

$$\mathrm{d}h = \frac{\Sigma\,(h - h_a)^2}{\Sigma h_a^2}, \tag{5.1}$$

where h_a is the analytical solution for h and Σ represents the sum over a fixed interval including the points where $0 < h < h_0$. Figure 7 shows that, when the space-step is reduced by a factor 10, the mean relative error is reduced by a factor about 4, which is compatible with other general convergence rates that can be proved for simple models in presence of singularities (e.g. *Perthame* [2002]). Similar results are obtained when the error on hu is considered.

6 One-dimensional simulation over simplified topography

To illustrate the potentiality of the numerical model, we have performed a series of numerical experiments using the friction laws described above over simplified 1D geometry. As an example, let us consider an exponential shape for the topography $Z(x)$, with characteristic dimensions of the order of the real topography of White River Valley in Montserrat island (Lesser Antilles) where an extensively studied debris avalanche occurred 26 December 1997 [*Sparks et al.*, 2002]. This debris avalanche with an estimated volume of about $40 - 45 \times 10^6\,m^3$ was caused by the failure of the upper south flank of the Soufriere Hills Volcano. Geological and numerical studies of this event have been performed and estimation of the thickness, velocity and

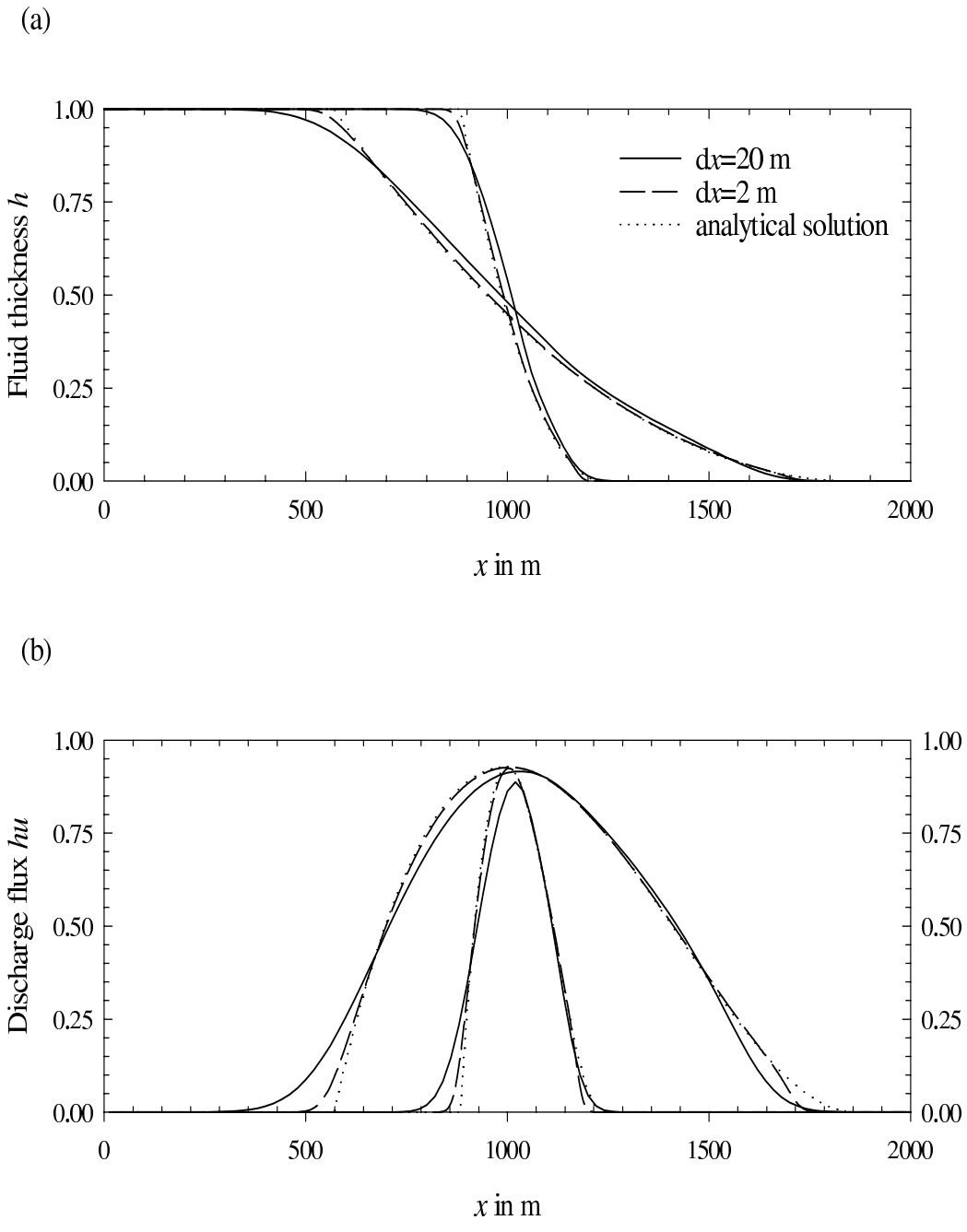


Figure 5: (a) Fluid thickness h and (b) discharge flux hu versus distance, obtained for $\delta = 0°$ and $\theta = 0°$ at time $t = 37\,s$ and $t = 137\,s$, calculated with the analytical solution (dotted lines) and with the numerical model for $dx = 20\,m$ (solid lines) and $dx = 2\,m$ (dashed lines).

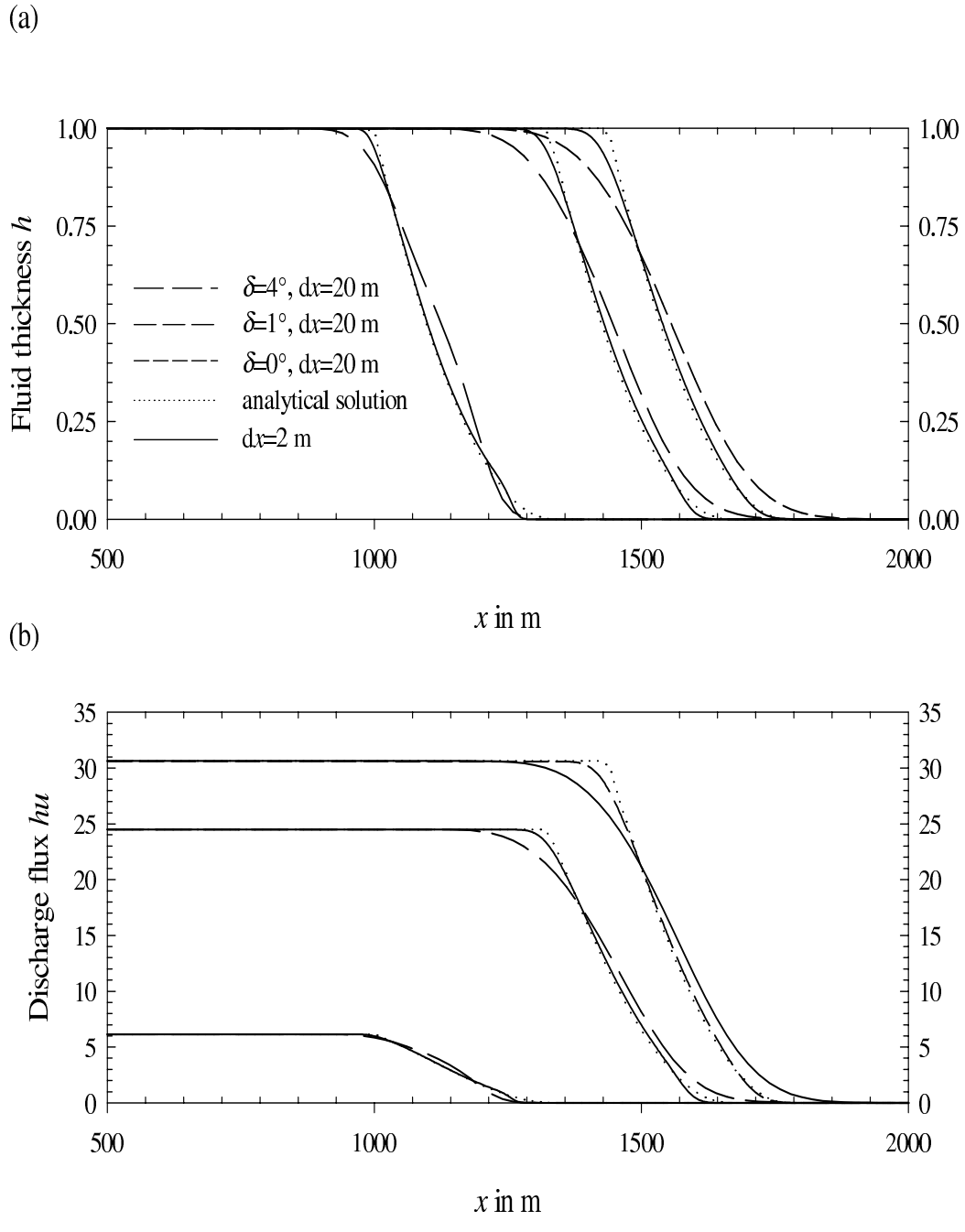


Figure 6: (a) Fluid thickness h and (b) discharge flux hu versus distance, obtained for $\delta = 0°$, $\delta = 1°$ and $\delta = 4°$ for inclination angle $\theta = 5°$ at time $t = 35\,s$, calculated with the analytical solution (dotted lines) and with the numerical model for $dx = 20\,m$ and $dx = 2\,m$ (solid lines).

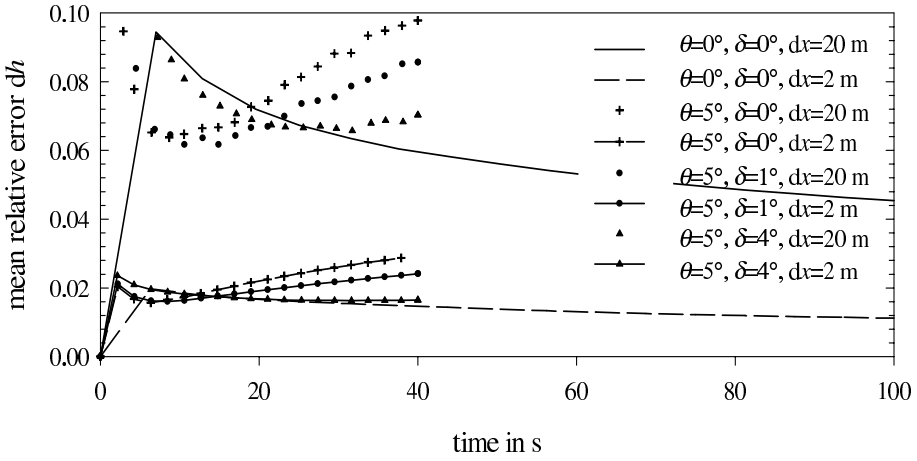


Figure 7: Mean relative error Δh for the dam-break problem for $dx = 20\,m$ (symbols) and and $dx = 2\,m$ (solid lines with symbols) for inclination angle of the bottom $\theta = 5°$ with various angle of friction $\delta = 0°$, $\delta = 1°$ and $\delta = 4°$ and mean relative error Δh for $\theta = 0°$ ($dx = 20\,m$ solid lines and $dx = 2\,m$ dashed lines).

runout distance of this debris avalanche are now available. The maximum deposit thicknesses range from $60\,m$ to $100\,m$. Front heights of about $20\,m$ are observed at a distance of $200\,m$ from the shoreline. It can be inferred from the observations that the avalanche travels approximately $3.5\,km$ from the center of the destabilized mass, in the reference frame linked to the topography.

Let us investigate the influence of the various flow laws in the range of parameters allowing the mass to stop around the position $x = 4500\,m$, corresponding approximately to the observed runout of the Boxing Day debris avalanche down the White River Valley. In the White River Valley, the altitude decreases from $900\,m$ at the top of the avalanche with a maximum slope inclination of 35° to the sea, with slope inclination of a few degrees at the shore. The corresponding angle is defined by

$$\theta(x) = \theta_0 \exp\left(-\frac{x}{a}\right), \tag{6.1}$$

with $\theta_0 = 35°$ and $a = 1750\,m$ (Figure 8b). The summit is located at an altitude of $950\,m$ with an initial slope of 35°, the topography being almost horizontal in the right part (Figure 8a). The results are presented in the coordinate system (x,z) linked to the topography. The initial conditions

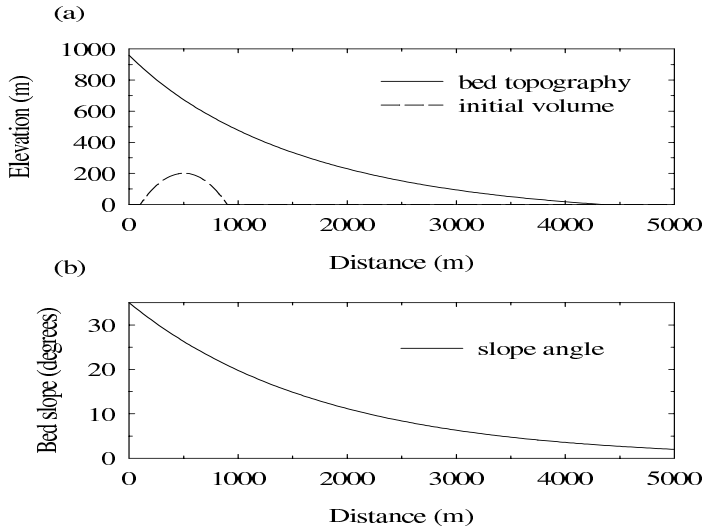


Figure 8: (a) Bed topography in Cartesian coordinates $(\tilde{x}, \tilde{z})$ and initial volume of the granular mass in topography-linked coordinates (x, z); (b) slope angle $\theta(x)$ of the bed in degrees in the topography-linked coordinates (x, z).

are defined by the instantaneous release of a parabolic mass over a rigid topography, represented in Figure 8a in the coordinate system (x,z) where

$$h(x, t = 0) = K\left(l - (x - x_0)^2\right), \tag{6.2}$$

$$u(x, t = 0) = 0, \tag{6.3}$$

with $K = 1.26 \times 10^{-3}\, m^{-1}$, $l = 1.6 \times 10^5\, m^2$ and $x_0 = 500\, m$. Initially, the maximal thickness of the mass is $200\, m$ in the direction perpendicular to the topography with a length of $800\, m$, close to the estimations of the Boxing Day debris avalanche destabilized mass. The numerical domain is discretized using 880 points in the x-direction with a space-step of $6.25\, m$.

6.1 Curvature effects

We note that equations (3.4)-(3.5) are obtained by neglecting the first order curvature terms. At first order in 1D, curvature effects lead to an additional friction force linked to centrifugal acceleration. According to the scale analysis of *Savage and Hutter* [1989], this first order curvature effect is taken into account by a term involving the curvature radius R of the bed profile in the momentum equation

$$\frac{\partial}{\partial t}\left(h\bar{u}\right) + \frac{\partial}{\partial x}(h\bar{u}^2) + \frac{\partial}{\partial y}\left(h\bar{u}\bar{v}\right) = \gamma_x g h + \frac{\partial}{\partial x}(g\gamma_z \frac{h^2}{2}) - \mu h(g\gamma_z + \frac{u^2}{R})\frac{u}{|u|}. \tag{6.4}$$

When either μ or $\lambda = L/R_c$, where R_c is a characteristic value of the curvature radius, or both are smaller than $O(\epsilon^{\frac{1}{2}})$ and when u does not become too large, then this term may be dropped in comparison with the others terms [see *Greve and Hutter*, 1993].

Numerical tests confirm that the first order curvature effects involved in the last term of equation (6.4) is not too large in our case, where the radius of curvature is relatively high. Note that, in the present case, ϵ is of order 0.1, $\mu = 0.27$ for $\delta = 15°$ is of order $\epsilon^{\frac{1}{2}}$ and λ is lower than 4×10^{-3}.
Figure 9 shows that the results with and without this curvature term are close to each other for a simple friction law with $\delta = 15°$, especially during the flow. Furthermore, the fluid stops almost at the same time ($t = 86.4\,s$ without curvature effects and $t = 86\,s$ with curvature effects) and the maximum elevation of the deposit is the same ($h_{max} = 67.8\,m$ without curvature and $h_{max} = 68\,m$ with curvature). However, a difference of $156\,m$ (5% of the deposit length) is observed in the runout distance.
When curvature effects are not taken into account, i.e. when the exponential shape does not slow down the granular mass, the front is located further away. The empirical nature of the friction angle in such a model is well illustrated in this example. In fact, curvature effects are difficult to take into account in 2D experiments. Dropping these effects leads to unverifiable error in the determination of the well-fitted friction angle. In the following 1D simulations, first order curvature effects have been also taken into account.

6.2 The Coulomb friction law

We first look at the results obtained by using the friction law with constant angle. Sensitivity study is performed just by varying the value of this angle. The avalanche deposit extends further for lower values of δ, as shown in Figure 10, where the geometry of the deposits is obtained when the flow comes to rest. A difference of approximatively $740\,m$ on the front position is obtained when δ varies from 14° to 16°, while a difference of approximatively $1060\,m$ when δ varies from 16° to 20°. Furthermore, the length of the deposit is larger and the maximum elevation lower when the friction angle decreases. The deposit extends along $2900\,m$ when $\delta = 14°$ with a maximum elevation $h = 65\,m$, while the extension is only $2290\,m$ when $\delta = 20°$ with a maximum elevation $h = 75\,m$. It appears that only low values of the friction angle around 15° are appropriate to reproduce the great mobility of real debris avalanches, as it was observed in 2D simulation [see *Heinrich et al.*, 2001].

The low value of δ is a consequence of the widely observed ability of large avalanches to travel distances much larger than expected from classical models of slope failure. Note that, dispite of the extreme simplification of this

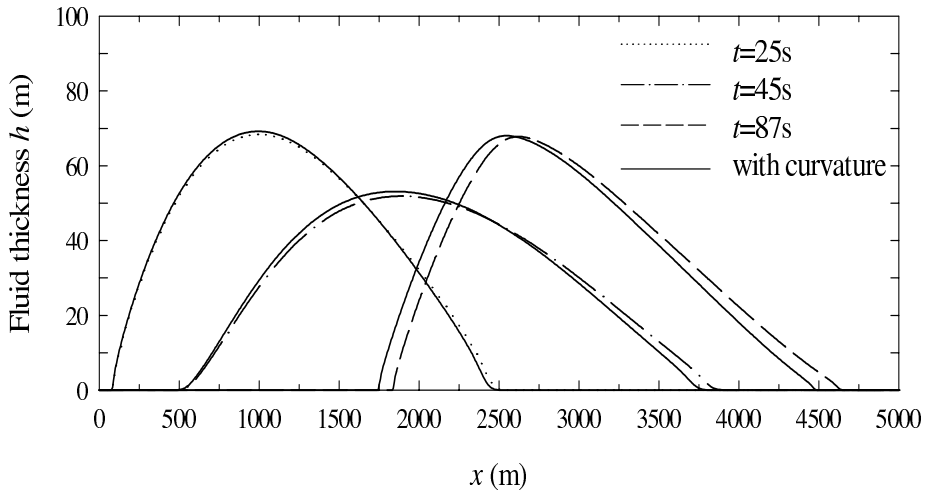


Figure 9: Fluid thickness h at $t = 25\,s$, $t = 45\,s$ and $t = 87\,s$ (i.e. when the fluid stops) with and without the curvature term for a simple friction law with $\delta = 15°$. The dash-dotted lines represent the result without curvature effect and the corresponding full lines those with curvature effect at the same time. Note that the fluid stops approximatively at the same time.

test, the calculated values are in the range of the deposit elevation estimated from geological observation [see *Sparks et al.*, 2002]. The x-position of the maximum elevation is situated toward the rear of the mass. In fact, with a constant friction angle, in the accelerating stage, the fluid flows with higher velocity near the front than near the rear due to a driving negative h-gradient. The positive h-gradient near the downhill rear of the fluid plays a braking role in the balance of forces, as it is illustrated in Figure 11. It is worth pointing out that the force due to the pressure gradient (i.e. the h-gradient) is relatively small compared to the other forces as well, at $t = 25\,s$ as at $t = 65\,s$ in the rest of the mass. This feature may explain the weak effect of the parameter $k_{actpass}$ involved in the pressure gradient when non-isotropy of normal stresses is assumed [see *Pouliquen and Forterre*, 2002].

6.3 Pouliquen's friction law

We propose to use here the more recent law developed empirically by *Pouliquen* [1999] (see Section 3.2). Contrary to the one-parameter simple friction law, three parameters have to be determined: two friction angles δ_1, δ_2 and the coefficient d. Debris avalanches are composed of particles with

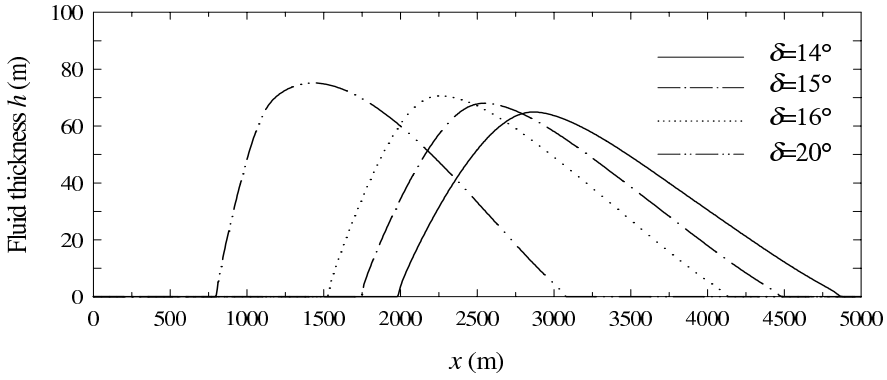


Figure 10: Profile of the mass at the time when the fluid stops, for various values of the friction angle δ using the simple friction law.

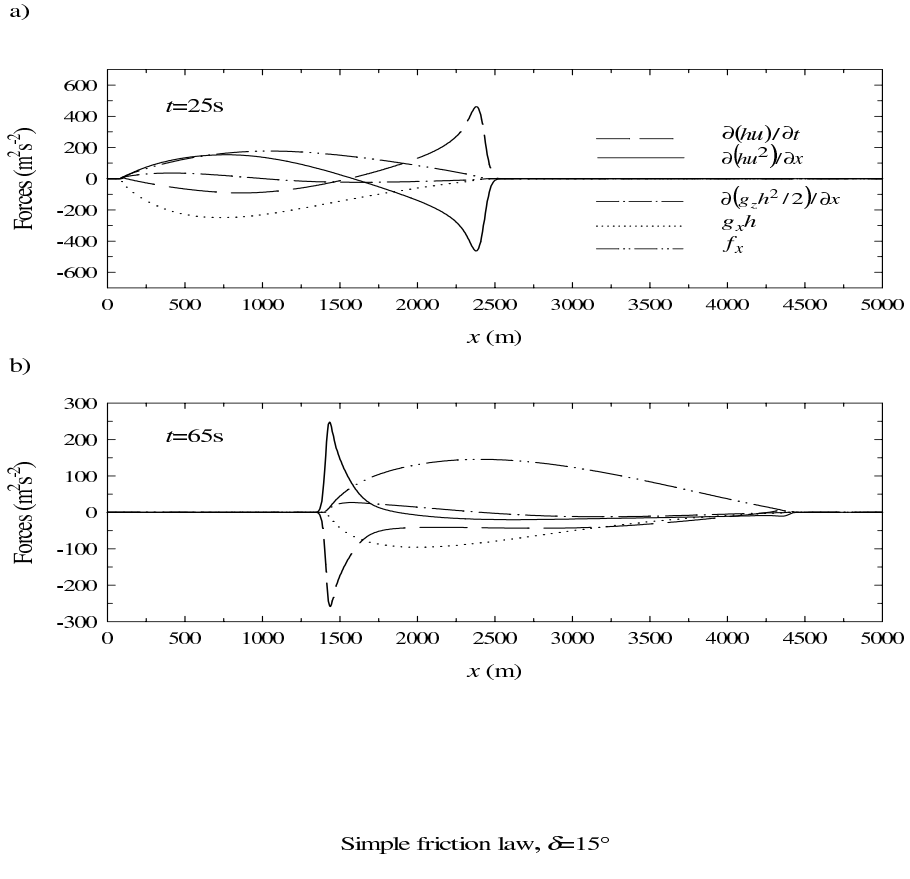


Figure 11: Forces involved in the x-momentum equation for a simple friction law with $\delta = 15°$ versus distance (a) at time $t = 25\,s$ and (b) at $t = 65\,s$.

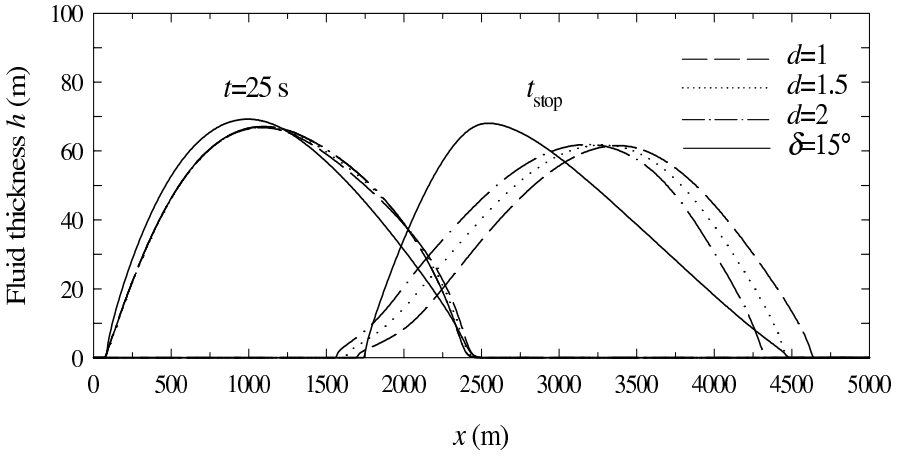


Figure 12: Fluid thickness at $t = 25\,s$ and when the fluid stops for various values of d in the Pouliquen's flow law, with $\delta_1 = 13°$ and $\delta_2 = 20°$, and for a simple friction law with $\delta = 15°$. The dash-dotted lines represent the result for the Pouliquen's flow law and the corresponding full lines those for the simple friction law.

sizes varying from less than one millimeter to tens of meters. It is therefore difficult to estimate the value of d in the model. However, a value of $d = 1.5\,m$ allows the mass reaching $x = 4500\,m$ for $\delta_1 = 13°$ and $\delta_2 = 20°$ (Figure 12). The variation of δ with the position is represented in Figure 13 at the instants $t = 25\,s$ and $t = 65\,s$ for $d = 1$, $d = 1.5$ and $d = 2$.

Note that for low value of d the results are similar to those obtained for simple friction law with $\delta = 13°$ and for high values of d the results are close to those obtained using a simple friction law with $\delta = 20°$. In this range of values, the flow is governed by δ_2 near the front and the rear of the flow and by δ_1 in the inner part of the mass. The friction angle evolves in time as a function of the flow parameters (h, hu) as in Figure 13. Differences of more than 1° are observed on δ when d-value goes from 1 to 2, leading to strong differences in the deposit (Figure 12). Figure 12 also shows that the shapes of the flowing mass at $t = 25\,s$ are similar for both various values of d and for the simple friction law.

During the flowing stage, the friction force does not play a leading role, as it is illustrated in Figure 14a at $t = 25\,s$. During the deceleration stage, the importance of the friction forces increases (Figure 14b) to the stopping stage, where the friction forces balanced by the gravity force dominate the other forces. Concerning the deposit, not only the runout distance changes

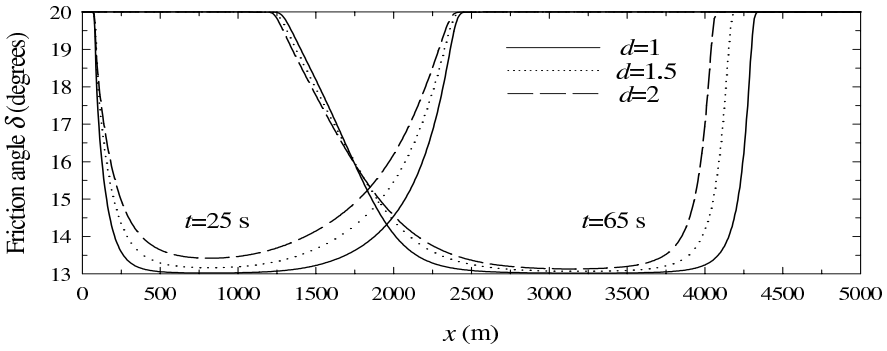


Figure 13: Friction angle δ versus position x in Pouliquen's flow law with $\delta_1 = 13°$ and $\delta_2 = 20°$ and $d = 1$ (full lines), $d = 1.5$ (dotted lines) and $d = 2$ (dashed lines) at time $t = 25\,s$ and $t = 65\,s$. At the rear and the front, i.e. for small values of h, the friction angle tends to δ_2.

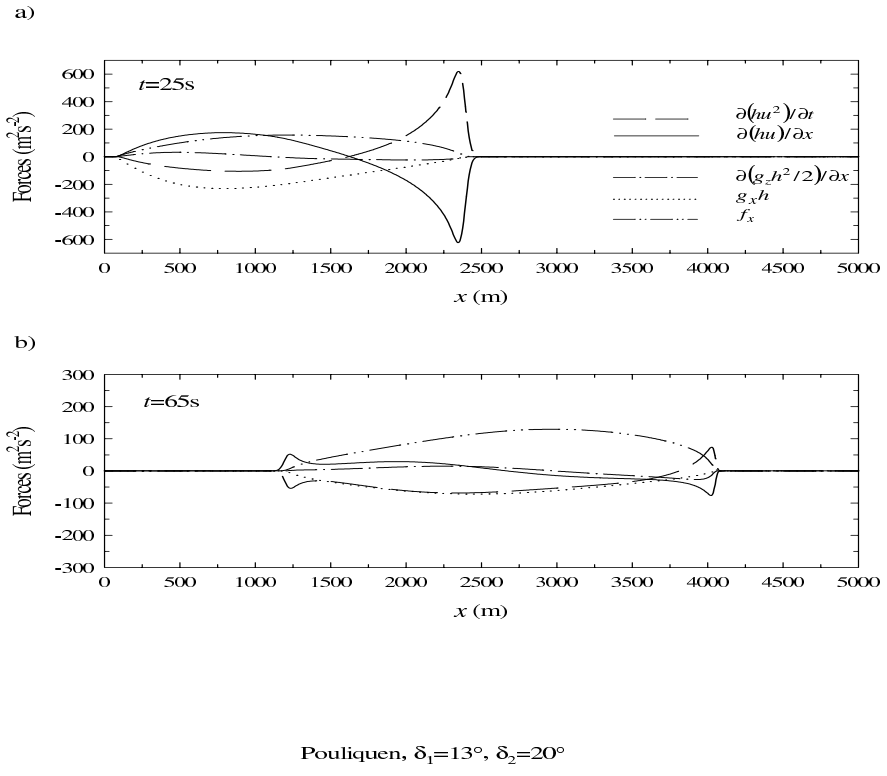


Figure 14: Forces involved in the x-momentum equation for Pouliquen's flow law with $\delta_1 = 13°$, $\delta_2 = 20°$ and $d = 1.5$ versus distance at time (a) $t = 25\,s$ and (b) $t = 65\,s$.

with d but also the shape. As d increases, the front becomes more marked and the rear finer. Such a shape seems to be closer to real observed front of avalanches. The shape of the deposit using Pouliquen's friction law with $d = 1.5$ is quite different from that obtained by simple friction law with $\delta = 15°$ even though the runout distance is the same and the extension of the deposit is similar (see Figure 12). The downhill part of the deposit using this variable friction angle is $18\,m$ high at $250\,m$ from the rear and $35\,m$ high for constant friction angle. Contrary to simple friction law, the maximum thickness is situated near the front for Pouliquen's flow law, due to low friction for high elevation in the inner part of the avalanche. In this example, contrary to simple friction law, Pouliquen's flow law can describe front height of approximatively $20\,m$ at a distance of $200\,m$ from the runout distance. As it was observed for simple friction law, the force due to the pressure gradient is relatively small compared with the other forces as well at $t = 25\,s$ as at $t = 65\,s$, except at the front (Figure 14).

These simple 1D simulations are in agreement with the results obtained using 2D simulations by *Heinrich et al.* [2001], where comparisons between flows calculated by Coulomb and Pouliquen's friction laws have shown the importance of the dependence of the friction angle on the Froude number and the flow height, suggesting a rate dependence in the mechanical behavior of debris avalanches.

6.4 Mass stopping

The major originality of the model presented in this paper consists in the introduction of the stopping mechanism in kinetic schemes. Let us look with more details at this stopping stage, illustrated in Figure 15 and Figure 16 for simple friction law and Pouliquen's flow law respectively.

For the simple friction law ($\delta = 15°$) the mass stops at $t = 86\,s$ and for the Pouliquen's flow law ($\delta_1 = 13°$, $\delta_2 = 20°$ and $d = 1.5$) at $t = 97.6\,s$. With these rheological parameters, the runout distance for both simple friction law and Pouliquen's flow law is approximatively $4500\,m$. With the constant angle friction law, the front encountering low slope begins to stop. The stopping propagates toward the rear of the mass until the whole fluid stops. The asymmetric shape becomes more pronounced when the fluid stops, due to this downward propagation of the stopping stage. Note that, with this topography and with this initial released mass, the Coulomb threshold is never reached in the rear of the flow for friction angle higher than $\delta = 23°$. For such high friction, the front stops and this stopping propagates toward the rear. However, the driving force and in particular the gravity near the rear of the flow is still higher than the Coulomb threshold due to high slope

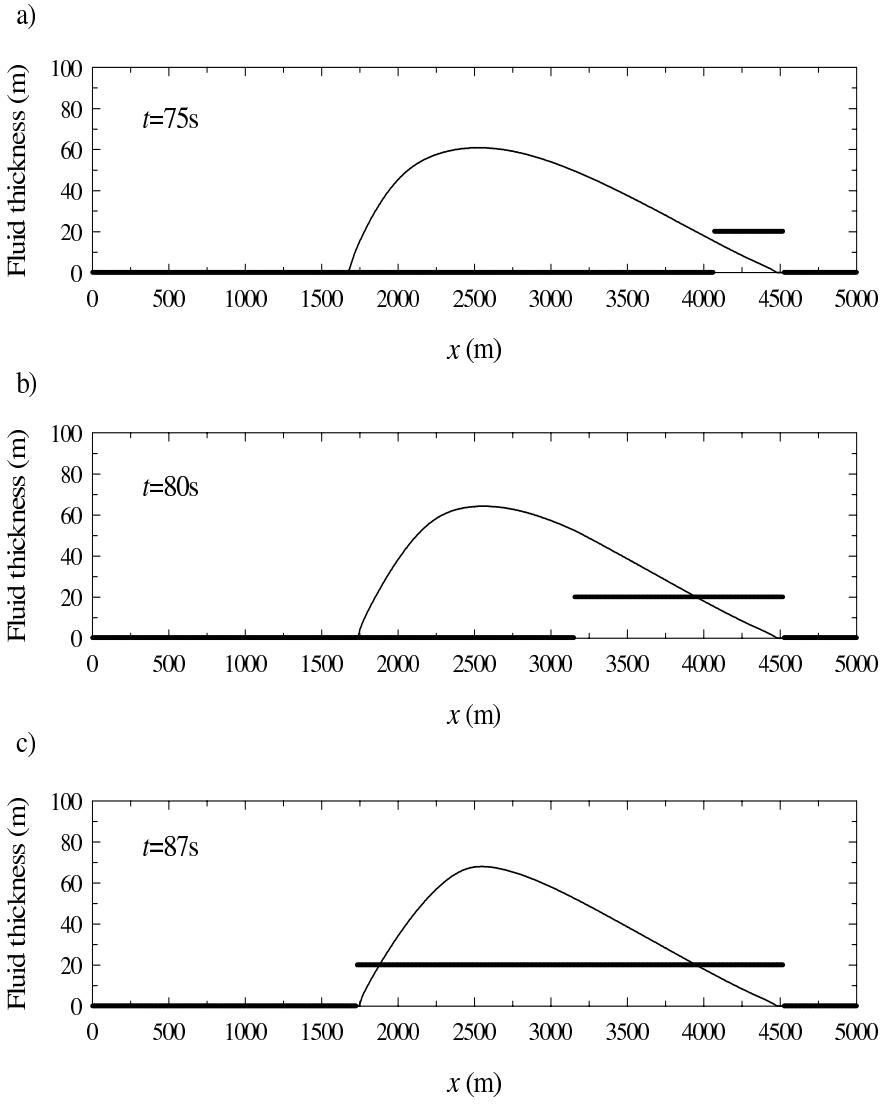


Figure 15: Fluid thickness (full lines) versus distance at (a) $t = 75\,s$, (b) $t = 80\,s$ and (c) $t = 87\,s$ during the stopping stage for simple friction law with $\delta = 15°$. A value of 0 is allocated to the fluid under the Coulomb threshold and a value of 20 to the fluid above the Coulomb threshold.

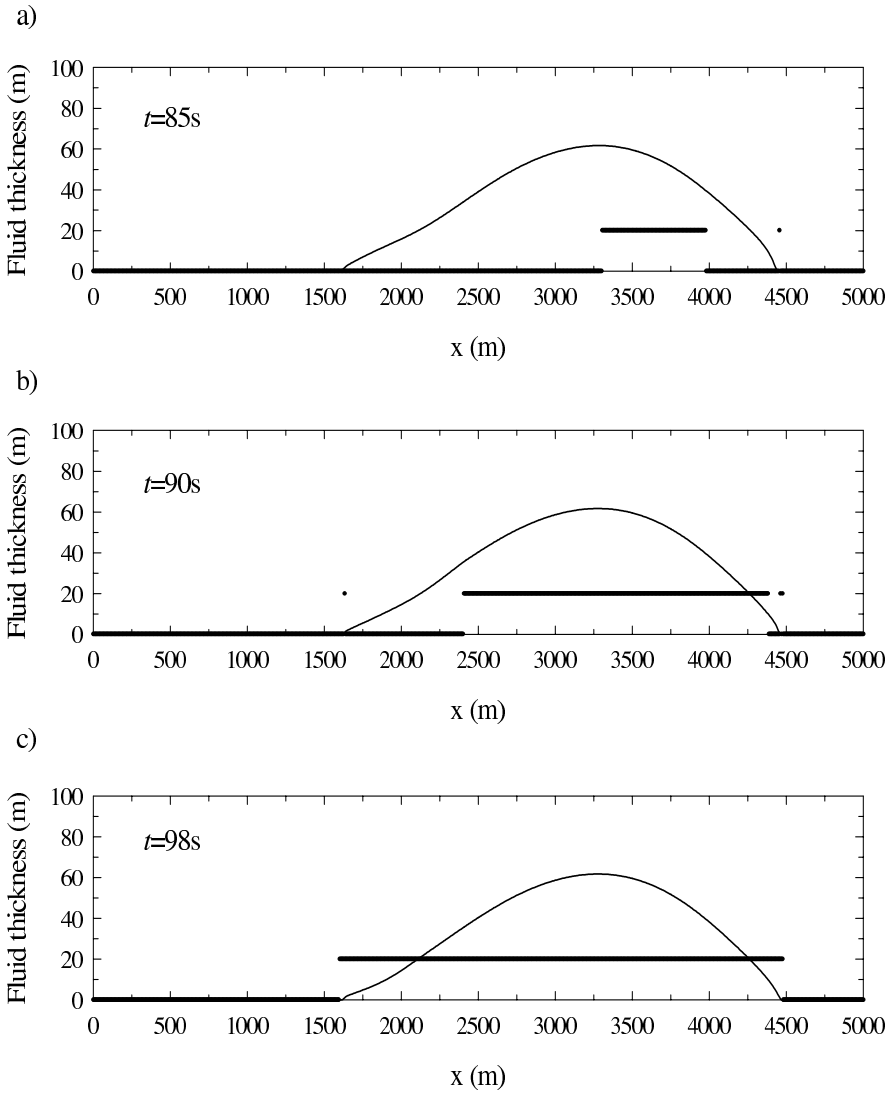


Figure 16: Fluid thickness (full lines) versus distance at (a) $t = 85\,s$, (b) $t = 90\,s$ and (c) $t = 98\,s$ during the stopping stage for Pouliquen's flow law with $\delta_1 = 13°$, $\delta_2 = 20°$ and $d = 1.5$. A value of 0 is allocated to the fluid under the Coulomb threshold and a value of 20 to the fluid above the Coulomb threshold.

of the topography. In this case, the h-gradient may play a significant role in controlling the balance of forces. As an example, at $t = 70\,s$ for $\delta = 24°$, the whole fluid is stopping except a $150\,m$ long part in the rear of the mass. In this region, the fluid is stopped by the downhill mass which is under the Coulomb threshold. The presence of a fluidized zone behind a rigid mass would be an interesting point to verify by comparing numerical results derived from mathematical models with empirical or geological observation of deposits. The stopping scenario is not the same for Pouliquen's flow law, for which the central part of the fluid is stopping first. In this case, the friction angle is not constant, as it was observed in the previous section.
The difference in the stopping behavior of a debris mass controlled by simple friction law or Pouliquen's flow law can be a useful test to determine the more appropriate flow law.

The presence of a fluidized zone behind a rigid mass is also observed, for example, with rheological parameters $\delta_1 = 12°$, $\delta_2 = 20°$ and $d = 10$, suggesting the existence of horizontal surfaces in the deposit. Further analysis of this phenomenon requires the development of a model reproducing the equilibrium of a fluid at rest [e.g. *Perthame and Simeoni*, 2001].

7 Conclusion

Numerical modelling of debris avalanches has been presented here based on Savage and Hutter's equations. Granular avalanche behavior has been described using a Coulomb-type friction law with constant and flow variable friction angle.

The numerical model is based on a kinetic scheme. The main idea is to introduce two different descriptions of the microscopic behavior of the system, suggested by the ambivalence of the fluid-solid behavior of granular material. The resulting solver appears to be stable and preserves height positivity, contrary to several Godounov-type methods. Efficiency of this model has been tested by comparisons with analytical solution of dam-break problems. The numerical scheme remains stable in spite of the introduction of the discontinuous Coulomb criterium. Furthermore, the discretization on a finite element mesh is well suited to simulate avalanches over real complex topographies.

Preliminary 1D simulations on a simplified geometry have allowed us to test the capacity of the numerical model and to compare constant and variable angle friction laws. The shape of the deposit strongly depends on the used friction law. Pouliquen's flow law, with a friction angle depending on the height and velocity, leads to steepest front of the granular deposit with

more elongating rear. Furthermore the stopping stage differs depending on the flow law. While the stopping propagates from the front to the rear when a constant friction angle is used, the inner part of the mass begins to stop when Pouliquen's flow law is used. This feature may be a useful tool to determine the best fitted flow law when comparing with experimental results. In these oversimplified numerical tests, Pouliquen's friction law appears to be more appropriate to describe debris avalanches than a simple Coulomb friction law, suggesting that frictional effects may play a significant role in debris avalanche mechanics.
Numerical modelling of debris avalanches provides the only way to estimate typical velocities and relative weight of the involved forces. The above analysis shows that the h-gradient force does not play a significant role in the examples studied here, except at the rear and front of the granular mass. The friction force begins to be a leading force only when the granular mass approaches the stopping stage.

The numerical tests show the possible existence of a fluidized zone in the deposit, under particular conditions. In such situations, part of the fluid remains over the Coulomb threshold, subjected for example to high gravity forces, and it is still blocked by the down slope deposit suggesting the existence of horizontal zones in the deposit. Observation of such features in real or experimental deposits would be interesting and may provide information on the mechanical behavior of a granular mass.

Acknowledgments

This work was supported by the Action Concerte Incitative (CNRS).

References

[1] Aranson, I.S., Tsimring, L.S., *Phys. Rev. E*, **64**, 020301 (R), 2001

[2] Arattano, M., Savage, W.Z., Modelling debris flows as kinematic waves, *Bull. Int. Assoc. Eng. Geol.*, **49**, 3-13, 1994

[3] Assier-Rzadkiewicz, S., Heinrich, P., Sabatier, P.C., Savoye, B., Bourillet, J.F., Numerical Modelling of a Landslide-generated Tsunami: The 1979 Nice Event, *Pure Appl. Geophys.*, **157**, 1707-1727, 2000

[4] Audusse, E., Bristeau, M.O., Perthame, B., Kinetic Schemes for Saint-Venant Equations with Source Terms on Unstructured Grids, *INRIA Report*, **3989**, 2000

[5] Azanza, E., Ecoulements granulaires bidimensionnels sur un plan incliné, *PhD of cole Nationale des Ponts et Chaussées*, 1998

[6] Botchorishvili, R., Perthame, B., Vasseur, A., Equilibrium Schemes for Scalar Conservation Laws with Stiff Sources, *INRIA Report*, **8931**, 2000

[7] Bristeau, M.O., Coussin, B., Boundary Conditions for the Shallow Water Equations solved by Kinetic Schemes, *INRIA Report*, **4282**, 2001

[8] Cheng-Lun, S., Chyan-Deng, J., Yuan-Fan, T., A Numerical Simulation of Debris Flow and Its Application, *Natural Hazards*, **13**, 39-54, 1996

[9] Denlinger, R.P., Iverson, R.M., Flow of variably fluidized granular masses across three-dimensional terrain 2. Numerical predictions and experimental tests, *J. Geophys. Res.*, **106**(B1), 553-566, 2001

[10] Douady, S., Andreotti, B., Daerr, A., On granular surface flow equations, *Eur. Phys. J. B.*, **11**, 131-142, 1999

[11] Eglit, E.M., Some mathematical models of snow avalanches, *Advances in the Mechanics and the Flow of Granular Materials*, M. Shahinpoor editor, **2**, 1983

[12] Gray, J.M.N.T., Wieland, M., Hutter, K., Gravity driven free surface flow of granular avalanches over complex basal topography, *Proc. Roy. Soc. Lond. A*, **455**, 1841-1874, 1999

[13] Greve, R., Hutter, K., Motion of a granular avalanche in a convex and concave chute: experiments and theoretical predictions *Proc. Roy. Soc. Lond. A*, **342**, 573-600, 1993

[14] Greve, R., Koch, T., Hutter, K., Unconfined flow of granular avalanches along a partly curved surface. I. Theory, *Proc. Roy. Soc. Lond. A*, **445**, 399-413, 1994

[15] Harbitz, C.B., Snow Avalanche Modelling, Mapping, and Warnig in Europe, *Report of the Fourth European Framework Programme Environment and Climate*, 1998

[16] Heinrich, Ph., Boudon, G., Komorowvski, J.C., Sparks, R.S.J., Herd, R., Voight, B., Numerical simulation of the December 1997 debris avalanche in Montserrat, Lesser Antilles, *Geophys. Lett.*, **13**, 2529-2532, 2001

[17] Hunt, B., Asymptotic Solution for Dam-Break Problem, *J. Hydraul. Eng.*, **110**(8), 1985

[18] Hunt, B., Newtonian fluid mechanics treatment of debris flows and avalanches, *J. Hydraul. Eng.*, **120**, 1350-1363, 1994

[19] Hutter, K., Koch, T., Pluss, C., Savage, S.B., The dynamics of avalanches of granular materials from initiation to runout. Part II. Experiments, *Acta Mech.*, **109**, 127-165, 1995

[20] Iverson, R.M., The physics of debris flows, *Rev. Geophys.*, **35**(3), 1997

[21] Iverson, R.M., Denlinger, R.P., Flow of variably fluidized granular masses across three-dimensional terrain, 1. Coulomb mixture theory, *J. Geophys. Res.*, **106**(B1), 537-552, 2001

[22] Jenkins, J.T., Askari, E., Hydraulic theory for a debris flow supported on a collisional shear layer, *CHAOS* **9**, 654-658, 1999

[23] Laigle, D., Coussot, Ph., Numerical Modeling of Mudflows, *J. Hydraul. Eng.*, **123**(7), 617-623, 1997

[24] Macedonio, G., Pareschi, M.T., Numerical simulation of some lahars from Mount St. Helens, *J. Volcanol. Geotherm. Res.*, **54**, 65-80, 1992

[25] Mangeney, A., Heinrich, Ph., Roche, R., Analytical Solution for Testing Debris Avalanche Numerical Models, *Pure Appl. Geophys.*, **157**, 1081-1096, 2000

[26] Naaim, M., Vial, S., Couture, R., Saint-Venant approach for rock avalanches modelling, *Saint Venant Symposium*, Paris, August 1997

[27] Naaim, M., Gurer, I., Two-phase Numerical Model of Powder Avalanche Theory and Application, *Natural Hazards*, **117**, 129-145, 1998

[28] Perthame, B., Simeoni, C., A kinetic scheme for the Saint-Venant system with a source term, *Calcolo*, **38(4)**, 201-231, 2001

[29] Perthame, B., Kinetic formulation of conservation laws, *Oxford Univ. Press*, 2002, to appear

[30] Pouliquen, O., Scaling laws in granular flows down rough inclined planes, *Phys. Fluids*, **11**(3), 542-548, 1999

[31] Pouliquen, O., Forterre, Y., Friction law for dense granular flows: application to the motion of a mass down a rough inclined plane, *J. Fluid Mech.*, **453**, 133-151, 2002

[32] Sabot, F., Naaim, M., Granada, F., Surinach, E., Planet, P., Furdada, G., Study of avalanche dynamics by seismic methods, image-processing techniques and numerical models, *Ann. Glaciol.*, **26**, 319-323, 1998

[33] Savage, S. B., Hutter, K., The motion of a finite mass of granular material down a rough incline, *J. Fluid Mech.*, **199**, 177-215, 1989

[34] Savage, S.B., Hutter, K., The dynamics of avalanches of granular materials from initiation to runout. Part I: Analysis, *Acta Mech.*, **86**, 201-223, 1991

[35] Sparks, R.S.J. et al., Generation of a debris avalanche and violent pyroclastic density current: the Boxing Day eruption of 26 December 1997 at the Soufriere Hills Volcano, Montserrat, *Geological Society Memoir*, 2001

[36] Tai, Y.C., Dynamics of Granular Avalanches and their Simulations with Shock-Capturing and Front-Tracking Numerical Schemes. PhD of Technische Universität of Darmstadt, 2000

[37] Tai, Y.C., Noelle, S., Gray, J.M.N.T., Hutter, K., Shock-Capturing and Front-Tracking Methods for Granular Avalanches, *J. Comp. Phys.*, 2002, to appear

[38] Toro, E.F., *Riemann Solvers and Numerical Methods for Fluid Dynamics*, 492 pp., Springer-Verlag, New York, 1997

[39] Whipple, K.X., Open-Channel Flow of Bingham Fluids: Applications in Debris-Flow Research, *J. Geol.*, **105**, 243-262, 1997

[40] Wieland, M., Gray, J.M.N.T., Hutter, K., Channelized free surface flow of cohesionless granular avalanches in a chute with shallow lateral curvature, *J. Fluid Mech.*, **392**, 73-100, 1999

[41] Zwinger, T., Dynamik einer Trockenscheelawine auf beliebig geformten Berghängen, PhD of the Technischen Universität Wien, 2000