
Projet personnel

Résolution des équations de Saint-Venant
2D par la méthode des volumes finis sur
un maillage non-structuré avec une
goutte perturbatrice initiale

Anthony CARBILLET

Mathématiques de la Modélisation
Méthodes numériques et Mécanique des fluides

2025

Table des matières

Introduction	2
1 Introduction des équations	3
1.1 Équation de Navier-Stokes	3
1.2 Équation de Saint-Venant	6
1.2.1 Hypothèses préalables	6
1.2.2 Passage de Navier-Stokes à Saint-Venant	7
2 Procédure numérique	11
2.1 Préambule	11
2.1.1 Motivation du choix de la méthode	11
2.1.2 Principe de la méthode	11
2.2 Méthode explicite	14
2.2.1 Schéma des volumes finis explicite par flux de type Rusanov	14
2.3 Méthode implicite	19
2.3.1 Schéma des volumes finis implicite linéarisé par itérations de point fixe . .	19
2.3.2 Schéma des volumes finis implicite avec résidu linéarisé et formulation ma- tricielle globale	21
3 Application au cas d'une goutte perturbatrice	26
3.1 Mise en place de la simulation numérique	26
3.1.1 Conditions et paramètres	26
3.1.2 Outils et fonctions auxiliaires	29
3.2 Analyse des résultats de la simulation numérique	33
3.2.1 Résultat des modélisations	33
3.2.2 Analyse et comparaison des schémas numériques	36
Bibliographie	44

Introduction

Ce document a pour but d'être une introduction pour une première approche de la méthode numérique des volumes finis avec pour support la résolution des équations de Saint-Venant. Il vise avant tout à proposer un parcours progressif, à la fois théorique et pratique, sans trop de formalisme, pour comprendre ces équations fondamentales de la mécanique des fluides et leur mise en œuvre computationnelle à travers un spectre de méthodes à la complexité variable. Ces méthodes sont très courantes dans la résolution de phénomènes fluides et les équations de Saint-Venant sont un cas très standard pour leur utilisation.

Dans un premier temps, un cheminement vers l'expression des équations de Saint-Venant va être mené. Le but est de remonter aux origines avec les équations de Navier-Stokes, notamment en présentant les hypothèses qui permettent de les établir. Puis, par étape avec encore une fois des hypothèses précises, amener la transition jusqu'aux équations de Saint-Venant avec les étapes clés. Ce processus permet de mieux saisir la portée physique de chaque paramètre présent dans les équations.

Le second chapitre va s'intéresser aux méthodes numériques permettant de résoudre ces équations. Il s'agira de s'intéresser à la méthode des volumes finis, en justifiant ce choix dans un premier temps, suivie d'une présentation générale du principe. Puis, à travers la discrétisation temporelle, présenter plusieurs schémas numériques par difficulté croissante. En premier, une méthode explicite, plus simple à appréhender, puis d'introduire progressivement des méthodes plus complexes, des méthodes implicites.

Enfin, le dernier chapitre va mettre en œuvre les concepts précédents à travers un exemple original : celui d'une goutte d'eau perturbatrice. Ce cas d'étude va permettre non seulement de comparer les performances des différentes approches numériques présentées en amont, mais aussi de mettre en lumière des éléments pratiques du code, comme les fonctions auxiliaires, souvent indispensables mais en marge de la méthode des volumes finis à proprement parler.

En somme, l'objectif du travail déployé à travers ce document est de présenter, analyser, évaluer et discuter la méthode des volumes finis à travers un cas concret.

Chapitre 1

Introduction des équations

Dans l'étude de la mécanique des fluides, les équations de Saint-Venant, publiées en 1871 par Adhémar Barré de Saint-Venant [2], forment un modèle couramment utilisé pour la modélisation de cours d'eau. Très présentes en océanographie comme en étude hydraulique fluviale, ces équations décrivent les écoulements à surface libre en eaux peu profondes (en anglais *shallow water*).

Leur origine vient d'une dérivation des équations de Navier-Stokes couplée à un certain nombre d'hypothèses qui seront présentées par la suite.

1.1 Équation de Navier-Stokes

Les équations de Navier-Stokes sont originaires de l'équation de continuité, qui exprime la conservation de la masse du fluide étudié, et des équations de quantité de mouvement, qui résultent en fait de la relation fondamentale de la dynamique des fluides. Viennent ensuite des manipulations mathématiques, qui vont être détaillées, qui permettent d'aboutir aux équations si bien connues.

- Les équations de départ

Soit $\vec{U} = (u, v, w)$ le vecteur vitesse et ses composantes cartésiennes u , v et w , ρ la densité et p la pression.

Pour un volume élémentaire Ω , la masse M à l'instant t est donnée par :

$$M(t) = \int_{\Omega} \rho \, d\mu \quad . \quad (1.1)$$

L'équation de continuité (loi de conservation de la masse) exige que le taux de variation de la masse soit nul,

$$\frac{d}{dt} M(t) = 0 \quad . \quad (1.2)$$

Puis, l'équation de la quantité de mouvement exprime la deuxième loi de Newton appliquée à M , ce qui donne,

$$\frac{d}{dt} (M\vec{U}) = \sum \vec{F} \quad . \quad (1.3)$$

Dans le cas d'un fluide newtonien, ces équations sont nommées par le nom des deux scientifiques qui les ont ainsi établies, Navier [6] et Stokes [9].

- Équations de continuité

Partant de (1.2) nous utilisons la relation entre l'écriture Lagrangienne et l'écriture Eulérienne de la dérivée. Ceci donne,

$$\frac{d}{dt}M = \frac{\partial M}{\partial t} + M_{\Omega} \quad , \quad (1.4)$$

où M_{Ω} représente le flux à travers la surface S_{Ω} . Puis en appliquant (1.1) il s'en suit,

$$\frac{\partial M}{\partial t} = \frac{\partial}{\partial t} \int_{\Omega} \rho \, d\mu = \int_{\Omega} \frac{\partial}{\partial t} \rho \, d\mu \quad , \quad (1.5)$$

et M_{Ω} peut se réécrire comme,

$$M_{\Omega} = \int_{S_{\Omega}} \rho \vec{U} \cdot \vec{n} \, d\sigma \quad . \quad (1.6)$$

Par le théorème de Gauss, le flux aux frontières est ramené à une intégrale de volume. En combinant ceci avec (1.5) et (1.6) l'équation (1.2) devient,

$$\int_{\Omega} \left(\frac{\partial}{\partial t} \rho + \vec{\nabla} \cdot (\rho \vec{U}) \right) d\mu = 0 \quad . \quad (1.7)$$

Puis comme le volume Ω est arbitraire, l'intégrale en tous points de Ω doit donc être nulle, ainsi

$$\frac{\partial}{\partial t} \rho + \vec{\nabla} \cdot (\rho \vec{U}) = 0 \quad . \quad (1.8)$$

Enfin, en utilisant l'écriture de dérivation particulaire, l'équation devient,

$$\frac{\partial}{\partial t} \rho + \vec{\nabla} \cdot (\rho \vec{U}) = \frac{\partial}{\partial t} \rho + (\vec{U} \cdot \vec{\nabla}) \rho + \rho (\vec{\nabla} \cdot \vec{U}) = \frac{\partial}{\partial t} \rho + \rho (\vec{\nabla} \cdot \vec{U}) = 0 \quad . \quad (1.9)$$

Maintenant, une nouvelle hypothèse est faite, celle de fluide incompressible, soit que sa variation de densité à travers le temps est nulle. Ceci implique que,

$$\frac{\partial}{\partial t} \rho = 0 \quad . \quad (1.10)$$

Ainsi, l'équation de continuité pour un fluide newtonien et incompressible est donnée par,

$$\vec{\nabla} \cdot \vec{U} = 0 \quad . \quad (1.11)$$

- Équations de quantité de mouvement

En partant du même processus effectué pour l'équation de continuité à travers les équations (1.4), (1.5) et (1.6), l'équation de la quantité de mouvement devient,

$$\frac{d}{dt} (M \vec{U}) = \frac{d}{dt} \int_{\Omega} (\rho \vec{U}) \, d\mu = \int_{\Omega} \frac{\partial}{\partial t} (\rho \vec{U}) \, d\mu + \int_{S_{\Omega}} \rho \vec{U} (\vec{U} \cdot \vec{n}) \, d\sigma \quad , \quad (1.12)$$

puis en réutilisant le théorème de Gauss l'équation de (1.3)

$$\frac{d}{dt} (M \vec{U}) = \int_{\Omega} \left(\frac{\partial}{\partial t} (\rho \vec{U}) + \vec{\nabla} \cdot (\rho \vec{U} \otimes \vec{U}) \right) d\mu = \sum \vec{F} \quad . \quad (1.13)$$

Les forces extérieures en présence sont les forces massiques \vec{F}_M et les forces surfaciques \vec{F}_S . Ces deux forces peuvent s'écrire sous forme intégrale en Ω . Le premier cas, la force massique.

Une force massique peut s'exprimer comme,

$$\vec{F}_M = \int_{\Omega} \rho \vec{\gamma} \, d\mu \quad , \quad (1.14)$$

où $\vec{\gamma}$ est homogène à une accélération, comme \vec{g} par exemple. L'exemple considéré se plaçant dans un référentiel de type laboratoire, les forces de Coriolis seront négligées, ainsi le choix sera $\vec{\gamma} = \vec{g}$.

Une force surfacique peut s'exprimer comme,

$$\vec{F}_S = \int_{S_\Omega} \vec{n} \cdot \tau \, d\sigma = \int_\Omega \vec{\nabla} \cdot \tau \, d\mu \quad , \quad (1.15)$$

où τ représente le tenseur des contraintes. Dans le cas d'un fluide newtonien et incompressible [1], le tenseur est donné par,

$$\tau = -p \mathbf{I}_d + 2\rho\nu D \quad , \quad (1.16)$$

où p est la pression, \mathbf{I}_d le tenseur unitaire, ν le coefficient de viscosité cinématique, considéré comme constant ici, et D le tenseur des taux de déformation.

Précisément,

$$\mathbf{I}_d = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad , \quad (1.17)$$

et

$$D = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) & \frac{1}{2} \left(\frac{\partial w}{\partial y} + \frac{\partial u}{\partial z} \right) \\ \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) & \frac{\partial v}{\partial y} & \frac{1}{2} \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial y} \right) \\ \frac{1}{2} \left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \right) & \frac{1}{2} \left(\frac{\partial v}{\partial z} + \frac{\partial v}{\partial y} \right) & \frac{\partial w}{\partial z} \end{pmatrix} \quad . \quad (1.18)$$

Ceci donne,

$$\int_\Omega \left(\frac{\partial}{\partial t} (\rho \vec{U}) + \vec{\nabla} \cdot (\rho \vec{U} \otimes \vec{U}) \right) d\mu = \int_\Omega -\vec{\nabla} p + \rho\nu \Delta \vec{U} + \rho \vec{\gamma} \, d\mu \quad . \quad (1.19)$$

De la même façon que le passage des équations (1.7) à (1.8), le volume Ω est arbitraire, et donc l'équation devient

$$\frac{\partial}{\partial t} (\rho \vec{U}) + \vec{\nabla} \cdot (\rho \vec{U} \otimes \vec{U}) = -\vec{\nabla} p + \rho\nu \Delta \vec{U} + \rho \vec{\gamma} \quad , \quad (1.20)$$

qui est la forme conservative de l'équation de quantité de mouvement. La forme non-conservative s'écrit,

$$\frac{\partial}{\partial t} \vec{U} + (\vec{U} \cdot \vec{\nabla}) \vec{U} = -\frac{1}{\rho} \vec{\nabla} p + \nu \Delta \vec{U} + \vec{\gamma} \quad . \quad (1.21)$$

En toute généralité, la masse volumique peut varier à cause de la température, de la présence de substances dans l'eau, comme le sel ou des sédiments. Mais dans tous les cas, cette variation est suffisamment faible pour pouvoir être négligée et utiliser seulement la forme incompressible de l'équation de continuité [14].

Finalement, l'écriture suivante des équations en coordonnées cartésiennes est obtenue.

$$\begin{cases} \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \end{cases} \quad (1.22)$$

$$\begin{cases} \frac{\partial u}{\partial t} + \frac{\partial u^2}{\partial x} + \frac{\partial uv}{\partial y} + \frac{\partial uw}{\partial z} = -\frac{1}{\rho} \frac{\partial p}{\partial x} + \nu \Delta u \end{cases} \quad (1.23)$$

$$\begin{cases} \frac{\partial v}{\partial t} + \frac{\partial vu}{\partial x} + \frac{\partial v^2}{\partial y} + \frac{\partial vw}{\partial z} = -\frac{1}{\rho} \frac{\partial p}{\partial y} + \nu \Delta v \end{cases} \quad (1.24)$$

$$\begin{cases} \frac{\partial w}{\partial t} + \frac{\partial wu}{\partial x} + \frac{\partial wv}{\partial y} + \frac{\partial w^2}{\partial z} = -\frac{1}{\rho} \frac{\partial p}{\partial z} + \nu \Delta w - g \end{cases} \quad (1.25)$$

1.2 Équation de Saint-Venant

Les équations de Navier-Stokes qui ont précédemment été exposées sont donc le point de départ pour aboutir aux équations de Saint-Venant. Pour les obtenir, les équations de Navier-Stokes vont être intégrées le long de la hauteur. Mais par souci de clarté, l'explication va être développée en deux temps.

D'abord, présenter les hypothèses nécessaires à l'écriture des équations puis développer les calculs pour aboutir aux équations de Saint-Venant.

1.2.1 Hypothèses préalables

Équilibre hydrostatique - densité

Dans le cadre des équations de Saint-Venant, ce sont des écoulements en eau peu profonde qui sont considérés. La répercussion immédiate de cette hypothèse est que la variation de densité est donc mineure. Pour une densité de référence ρ_{ref} nous avons alors

$$\rho = \rho_{ref} + \Delta\rho \approx \rho_{ref} \quad . \quad (1.26)$$

Cette hypothèse représente la conservation de l'équilibre hydrostatique le long de la hauteur, soit une distribution linéaire de la densité.

Équilibre hydrostatique - pression

Toujours dans la considération d'écoulement en eau peu profonde, l'accélération le long de la hauteur peut être négligée face à l'intensité de la pesanteur. Ainsi, l'équation (1.25) se simplifie comme

$$\frac{\partial p}{\partial z} = -\rho g \quad . \quad (1.27)$$

Comme pour la densité, cela traduit une distribution linéaire le long de la verticale. L'avantage de cette équation est qu'elle permet de pouvoir exprimer la pression en fonction de la hauteur. Par intégration, et en supposant que la pression à la surface est nulle, nous avons

$$p(x, y, z) = \rho g(Z_s - z) \quad . \quad (1.28)$$

Imperméabilité des bords horizontaux

L'hypothèse ici est qu'il est supposé qu'il n'y a pas de transfert de masse à travers la surface et le fond. De plus, si un point se trouve situé sur l'une de ces deux surfaces, il y restera au cours du temps. Avec ce qui précède et ces hypothèses, cela se traduit en équation de la façon suivante

$$\frac{\partial Z_s}{\partial t} + u_s \frac{\partial Z_s}{\partial x} + v_s \frac{\partial Z_s}{\partial y} - w_s = 0 \quad , \quad (1.29)$$

$$\frac{\partial Z_f}{\partial t} + u_f \frac{\partial Z_f}{\partial x} + v_f \frac{\partial Z_f}{\partial y} - w_f = 0 \quad , \quad (1.30)$$

où

$$\begin{array}{ll} u_s = u(x, y, Z_s) & u_f = u(x, y, Z_f) \\ v_s = v(x, y, Z_s) & v_f = v(x, y, Z_f) \\ w_s = w(x, y, Z_s) & w_f = w(x, y, Z_f) \end{array}$$

avec l'indice s qui désigne les valeurs au niveau de la surface, $h = Z_s$, et l'indice f les valeurs au fond, $h = Z_f$.

1.2.2 Passage de Navier-Stokes à Saint-Venant

Les hypothèses précédemment introduites sont le point de départ. Maintenant, elles vont être utilisées en présentant le passage entre les équations de Navier-Stokes aux équations de Saint-Venant.

Le point principal repose sur l'intégration le long de la hauteur, entre Z_s et Z_f , des équations. Pour cela, les variables suivantes sont posées :

$$\tilde{u} = \frac{1}{h} \int_{Z_f}^{Z_s} u \, dz \quad \tilde{v} = \frac{1}{h} \int_{Z_f}^{Z_s} v \, dz \quad , \quad (1.31)$$

où $h = Z_s - Z_f$ et \tilde{u} et \tilde{v} désignent les vitesses moyennées sur z .

Aussi l'outil qui sera ici le plus utile durant le développement de ces calculs est la formule de Leibniz pour la dérivation sous le signe d'intégration. Soit que pour une fonction f avec certaines hypothèses de régularité, nous avons

$$\frac{\partial}{\partial x} \int_{Z_f}^{Z_s} f \, dz = \int_{Z_f}^{Z_s} \frac{\partial f}{\partial x} \, dz + f(x, y, Z_s) \frac{\partial Z_s}{\partial x} - f(x, y, Z_f) \frac{\partial Z_f}{\partial x} \quad . \quad (1.32)$$

Avec tout ceci, tout est prêt pour commencer.

- Équations de continuité

Comme énoncé précédemment, l'équation (1.22) va être intégrée le long de la hauteur, ce qui donne

$$\int_{Z_f}^{Z_s} \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) dz = 0 \quad . \quad (1.33)$$

Puis, en utilisant la formule de Leibniz (1.32) l'équation devient

$$\begin{aligned} \int_{Z_f}^{Z_s} \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) dz &= \frac{\partial}{\partial x} \int_{Z_f}^{Z_s} u \, dz - u(x, y, Z_s) \frac{\partial Z_s}{\partial x} + u(x, y, Z_f) \frac{\partial Z_f}{\partial x} \\ &\quad + \frac{\partial}{\partial y} \int_{Z_f}^{Z_s} v \, dz - v(x, y, Z_s) \frac{\partial Z_s}{\partial y} + v(x, y, Z_f) \frac{\partial Z_f}{\partial y} \\ &\quad + w(Z_s) - w(Z_f) \\ &= 0 \end{aligned} \quad (1.34)$$

Avec l'hypothèse d'imperméabilité des bords des formules (1.29) et (1.30) nous avons

$$\begin{aligned} -u_s \frac{\partial Z_s}{\partial x} - v_s \frac{\partial Z_s}{\partial y} &= \frac{\partial Z_s}{\partial t} - w_s \quad , \\ u_f \frac{\partial Z_f}{\partial x} + v_f \frac{\partial Z_f}{\partial y} &= -\frac{\partial Z_f}{\partial t} + w_f \quad , \end{aligned}$$

ce qui permet de simplifier (1.34) en

$$\int_{Z_f}^{Z_s} \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) dz = \frac{\partial}{\partial x} \int_{Z_f}^{Z_s} u \, dz + \frac{\partial}{\partial y} \int_{Z_f}^{Z_s} v \, dz + \frac{\partial Z_s}{\partial t} - \frac{\partial Z_f}{\partial t} = 0 \quad , \quad (1.35)$$

puis finalement

$$\frac{\partial h}{\partial t} + \frac{\partial h \tilde{u}}{\partial x} + \frac{\partial h \tilde{v}}{\partial y} = 0 \quad . \quad (1.36)$$

- Équations de quantité de mouvement

Le procédé qui va être mis en œuvre ici est le même que pour l'équation de continuité et intégrer sur la hauteur les équations

$$\frac{\partial u}{\partial t} + \frac{\partial u^2}{\partial x} + \frac{\partial uv}{\partial y} + \frac{\partial uw}{\partial z} = -\frac{1}{\rho} \frac{\partial p}{\partial x} + \nu \Delta u \quad , \quad (1.23)$$

$$\frac{\partial v}{\partial t} + \frac{\partial vu}{\partial x} + \frac{\partial v^2}{\partial y} + \frac{\partial vw}{\partial z} = -\frac{1}{\rho} \frac{\partial p}{\partial y} + \nu \Delta v \quad . \quad (1.24)$$

Mais comme les calculs sont analogues pour la vitesse en x et la vitesse en y , seul les passages pour l'équation (1.23) seront traités.

La première étape donne donc

$$\int_{Z_f}^{Z_s} \left(\frac{\partial u}{\partial t} + \frac{\partial u^2}{\partial x} + \frac{\partial uv}{\partial y} + \frac{\partial uw}{\partial z} \right) dz = \int_{Z_f}^{Z_s} \left(-\frac{1}{\rho} \frac{\partial}{\partial x} (\rho g(Z_s - z)) + \nu \Delta u \right) dz \quad , \quad (1.37)$$

où p a été remplacée par son expression obtenue à (1.28).

Il y a plusieurs termes dans ce calcul d'intégration. Un terme de variation temporelle, un terme de flux et un terme source. Le premier a être traité va être celui de variation temporelle.

$$\begin{aligned} \int_{Z_f}^{Z_s} \frac{\partial u}{\partial t} dz &= \frac{\partial}{\partial t} \int_{Z_f}^{Z_s} u dz - u(x, y, Z_s) \frac{\partial Z_s}{\partial t} + u(x, y, Z_f) \frac{\partial Z_f}{\partial t} \\ &= \frac{\partial h \tilde{u}}{\partial t} - u_s \frac{\partial Z_s}{\partial t} + u_f \frac{\partial Z_f}{\partial t} \quad . \end{aligned} \quad (1.38)$$

Viens en suite le terme de flux,

$$\begin{aligned} \int_{Z_f}^{Z_s} \left(\frac{\partial u^2}{\partial x} + \frac{\partial uv}{\partial y} + \frac{\partial uw}{\partial z} \right) dz &= \frac{\partial}{\partial x} \int_{Z_f}^{Z_s} u^2 dz - u^2(x, y, Z_s) \frac{\partial Z_s}{\partial x} + u^2(x, y, Z_f) \frac{\partial Z_f}{\partial x} \\ &\quad + \frac{\partial}{\partial y} \int_{Z_f}^{Z_s} uv dz - u(x, y, Z_s) v(x, y, Z_s) \frac{\partial Z_s}{\partial y} \\ &\quad + u(x, y, Z_f) v(x, y, Z_f) \frac{\partial Z_f}{\partial y} + u(x, y, Z_s) w(x, y, Z_s) \\ &\quad - u(x, y, Z_f) w(x, y, Z_f) \end{aligned}$$

En utilisant la logique des valeurs moyennées sur la hauteur posée pour (1.32) et en compactant l'écriture nous avons

$$\begin{aligned} &= \frac{\partial h \tilde{u}^2}{\partial x} + \frac{\partial h \tilde{u} \tilde{v}}{\partial y} - u_s^2 \frac{\partial Z_s}{\partial x} + u_f^2 \frac{\partial Z_f}{\partial x} - u_s v_s \frac{\partial Z_s}{\partial y} \\ &\quad + u_f v_f \frac{\partial Z_f}{\partial y} + u_s w_s - u_f w_f \end{aligned} \quad (1.39)$$

En combinant (1.38) et (1.39) nous obtenons

$$\begin{aligned} \int_{Z_f}^{Z_s} \left(\frac{\partial u}{\partial t} + \frac{\partial u^2}{\partial x} + \frac{\partial uv}{\partial y} + \frac{\partial uw}{\partial z} \right) dz &= \frac{\partial h \tilde{u}}{\partial t} + \frac{\partial h \tilde{u}^2}{\partial x} + \frac{\partial h \tilde{u} \tilde{v}}{\partial y} - u_s \frac{\partial Z_s}{\partial t} + u_f \frac{\partial Z_f}{\partial t} \\ &\quad - u_s^2 \frac{\partial Z_s}{\partial x} + u_f^2 \frac{\partial Z_f}{\partial x} - u_s v_s \frac{\partial Z_s}{\partial y} + u_f v_f \frac{\partial Z_f}{\partial y} \\ &\quad + u_s w_s - u_f w_f \\ &= \frac{\partial h \tilde{u}}{\partial t} + \frac{\partial h \tilde{u}^2}{\partial x} + \frac{\partial h \tilde{u} \tilde{v}}{\partial y} \\ &\quad - u_s \left(\frac{\partial Z_s}{\partial t} + u_s \frac{\partial Z_s}{\partial x} + v_s \frac{\partial Z_s}{\partial y} - w_s \right) \\ &\quad + u_f \left(\frac{\partial Z_f}{\partial t} + u_f \frac{\partial Z_f}{\partial x} + v_f \frac{\partial Z_f}{\partial y} - w_f \right) \end{aligned}$$

Puis en utilisant l'hypothèse (1.30) et (1.31) nous avons finalement

$$\int_{Z_f}^{Z_s} \left(\frac{\partial u}{\partial t} + \frac{\partial u^2}{\partial x} + \frac{\partial uv}{\partial y} + \frac{\partial uw}{\partial z} \right) dz = \frac{\partial h\tilde{u}}{\partial t} + \frac{\partial h\tilde{u}^2}{\partial x} + \frac{\partial h\tilde{u}\tilde{v}}{\partial y} . \quad (1.40)$$

De façon analogue, pour la vitesse en y , nous avons

$$\int_{Z_f}^{Z_s} \left(\frac{\partial v}{\partial t} + \frac{\partial vu}{\partial x} + \frac{\partial v^2}{\partial y} + \frac{\partial vw}{\partial z} \right) dz = \frac{\partial h\tilde{v}}{\partial t} + \frac{\partial h\tilde{u}\tilde{v}}{\partial x} + \frac{\partial h\tilde{v}^2}{\partial y} . \quad (1.41)$$

La suite des calculs se concentre sur le terme source et son développement va être découpé en deux temps. D'abord l'intégration du terme de pression.

$$\begin{aligned} \int_{Z_f}^{Z_s} -\frac{1}{\rho} \frac{\partial}{\partial x} (\rho g(Z_s - z)) dz &= \frac{\partial}{\partial x} \int_{Z_f}^{Z_s} -g(Z_s - z) dz + 0 - g(Z_s - Z_f) \frac{\partial Z_f}{\partial x} \\ &= \frac{\partial}{\partial x} \left[-g \left(Z_s z - \frac{1}{2} z^2 \right) \right]_{Z_f}^{Z_s} - g(Z_s - Z_f) \frac{\partial Z_f}{\partial x} \\ &= \frac{\partial}{\partial x} \left(-g \left(Z_s^2 - \frac{1}{2} Z_s^2 - Z_s Z_f + \frac{1}{2} Z_f^2 \right) \right) - g(Z_s - Z_f) \frac{\partial Z_f}{\partial x} \\ &= \frac{\partial}{\partial x} \left(-g \left(\frac{1}{2} Z_s^2 - Z_s Z_f + \frac{1}{2} Z_f^2 \right) \right) - g(Z_s - Z_f) \frac{\partial Z_f}{\partial x} \\ &= -\frac{1}{2} g \frac{\partial}{\partial x} (Z_s - Z_f)^2 - g(Z_s - Z_f) \frac{\partial Z_f}{\partial x} \\ &= -\frac{1}{2} g \frac{\partial h^2}{\partial x} - gh \frac{\partial Z_f}{\partial x} \end{aligned} \quad (1.42)$$

Encore une fois, de façon analogue pour la vitesse en y nous avons

$$\int_{Z_f}^{Z_s} -\frac{1}{\rho} \frac{\partial}{\partial y} (\rho g(Z_s - z)) dz = -\frac{1}{2} g \frac{\partial h^2}{\partial y} - gh \frac{\partial Z_f}{\partial y} . \quad (1.43)$$

Enfin, l'intégration du terme de contrainte visqueuse

$$\begin{aligned} \int_{Z_f}^{Z_s} \nu \Delta u dz &= \int_{Z_f}^{Z_s} \nu \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) dz \\ &= \int_{Z_f}^{Z_s} \nu \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) dz + \nu \int_{Z_f}^{Z_s} \frac{\partial^2 u}{\partial z^2} dz \end{aligned}$$

En appliquant l'hypothèse de couche mince, la variation de vitesse en x et y est négligeable vis à vis de celle en z . Ainsi

$$= 0 + \nu \int_{Z_f}^{Z_s} \frac{\partial^2 u}{\partial z^2} dz .$$

Nous utilisons ensuite le théorème de la divergence,

$$\begin{aligned} &= \nu \left[\frac{\partial u}{\partial z} \right]_{Z_f}^{Z_s} \\ &= \nu \frac{\partial}{\partial z} u(x, y, Z_s) - \nu \frac{\partial}{\partial z} u(x, y, Z_f) \\ &= \frac{\tau_x^s}{\rho} - \frac{\tau_x^f}{\rho} \end{aligned} \quad (1.44)$$

où τ_x^s et τ_x^f désignent respectivement le cisaillement à la surface et au fond en x . En effet,

$$\tau_x^s = \rho \nu \frac{\partial}{\partial z} u(x, y, Z_s) \quad \tau_x^f = \rho \nu \frac{\partial}{\partial z} u(x, y, Z_f) \quad (1.45)$$

Analogiquement pour la vitesse en y nous avons

$$\int_{Z_f}^{Z_s} \nu \Delta v \, dz = \frac{\tau_y^s}{\rho} - \frac{\tau_y^f}{\rho} . \quad (1.46)$$

En mettant bout à bout les équations (1.40), (1.42) et (1.44) pour la vitesse en x et les équations (1.41), (1.43) et (1.46) pour la vitesse en y , les équations (1.23) et (1.24) deviennent

$$\frac{\partial h \tilde{u}}{\partial t} + \frac{\partial h \tilde{u}^2}{\partial x} + \frac{\partial h \tilde{u} \tilde{v}}{\partial y} = -\frac{1}{2} g \frac{\partial h^2}{\partial x} - gh \frac{\partial Z_f}{\partial x} + \frac{\tau_x^s}{\rho} - \frac{\tau_x^f}{\rho} \quad (1.47)$$

$$\frac{\partial h \tilde{v}}{\partial t} + \frac{\partial h \tilde{u} \tilde{v}}{\partial x} + \frac{\partial h \tilde{v}^2}{\partial y} = -\frac{1}{2} g \frac{\partial h^2}{\partial y} - gh \frac{\partial Z_f}{\partial y} + \frac{\tau_y^s}{\rho} - \frac{\tau_y^f}{\rho} \quad (1.48)$$

Tous les ingrédients pour écrire la formulation des équations de Saint-Venant sont maintenant réunis. Pour alléger l'écriture, l'écriture des vitesses, u et v , sera confondue avec les vitesses moyennées sur la hauteur, \tilde{u} et \tilde{v} . Ainsi, nous avons

$$\left\{ \begin{array}{l} \frac{\partial h}{\partial t} + \frac{\partial hu}{\partial x} + \frac{\partial hv}{\partial y} = 0 \\ \frac{\partial hu}{\partial t} + \frac{\partial}{\partial x} \left(hu^2 + \frac{1}{2} gh^2 \right) + \frac{\partial huv}{\partial y} = -gh \frac{\partial Z_f}{\partial x} + \frac{\tau_x^s}{\rho} - \frac{\tau_x^f}{\rho} \\ \frac{\partial hv}{\partial t} + \frac{\partial huv}{\partial x} + \frac{\partial}{\partial y} \left(hv^2 + \frac{1}{2} gh^2 \right) = -gh \frac{\partial Z_f}{\partial y} + \frac{\tau_y^s}{\rho} - \frac{\tau_y^f}{\rho} \end{array} \right. \quad (1.49)$$

$$\frac{\partial hu}{\partial t} + \frac{\partial}{\partial x} \left(hu^2 + \frac{1}{2} gh^2 \right) + \frac{\partial huv}{\partial y} = -gh \frac{\partial Z_f}{\partial x} + \frac{\tau_x^s}{\rho} - \frac{\tau_x^f}{\rho} \quad (1.50)$$

$$\frac{\partial hv}{\partial t} + \frac{\partial huv}{\partial x} + \frac{\partial}{\partial y} \left(hv^2 + \frac{1}{2} gh^2 \right) = -gh \frac{\partial Z_f}{\partial y} + \frac{\tau_y^s}{\rho} - \frac{\tau_y^f}{\rho} \quad (1.51)$$

De façon plus compacte, et pour simplifier l'écriture, la formulation matricielle suivante sera utilisée,

$$\frac{\partial}{\partial t} U + \frac{\partial}{\partial x} F(U) + \frac{\partial}{\partial y} G(U) = S , \quad (1.52)$$

avec,

$$U = \begin{pmatrix} h \\ hu \\ hv \end{pmatrix} \quad F(U) = \begin{pmatrix} hu \\ hu^2 + \frac{1}{2} gh^2 \\ huv \end{pmatrix} \quad G(U) = \begin{pmatrix} hv \\ huv \\ hv^2 + \frac{1}{2} gh^2 \end{pmatrix} \quad S = \begin{pmatrix} 0 \\ -gh \frac{\partial Z_f}{\partial x} + \frac{\tau_x^s}{\rho} - \frac{\tau_x^f}{\rho} \\ -gh \frac{\partial Z_f}{\partial y} + \frac{\tau_y^s}{\rho} - \frac{\tau_y^f}{\rho} \end{pmatrix}$$

Chapitre 2

Procédure numérique

Pour résoudre ces équations, le choix qui a été fait est d'utiliser une méthode de résolution classique dans le cas d'équations fluides, à savoir la méthode des volumes finis pour l'espace couplée à la méthode des différences finies pour le temps.

Dans ce chapitre, ce choix des volumes finis pour la résolution spatiale sera motivé dans une première partie, ainsi qu'une présentation du concept. Puis, différents schémas temporels aux différences finies seront vus entre les méthodes explicites et implicites, comparant leur fonctionnement pour voir les bénéfices de chacune.

2.1 Préambule

2.1.1 Motivation du choix de la méthode

La méthode de résolution spatiale qui a été choisie est donc la méthode des volumes finis (en anglais *Finite Volume Method* - FVM). Il va ici être expliqué pourquoi.

En matière de mécanique des fluides, hydraulique, aérodynamique, physique des plasmas, c'est une méthode de référence [12]. En effet, là où les méthodes éléments finis (en anglais *Finite Element Method* - FEM) ou différences finies (en anglais *Finite Difference Method* - FDM) vont se baser sur des approximations des opérateurs dérivés, la méthode FVM va reposer sur une formulation conservative naturelle, ce qui garantit le respect des lois de conservation de la masse et de la quantité de mouvement à l'échelle locale comme globale, là où les méthodes FEM ou FDM ont plus de mal à avoir une conservation stricte.

De plus, pour l'exemple qui est ici considéré, il s'agit d'une équation hyperbolique et d'un maillage non structuré, ce qui implique la modélisation de phénomènes souvent non lisses avec des discontinuités et une formulation plus complexe sur ce type de maillage.

Cependant, contrairement à la méthode FDM, qui peut rencontrer des difficultés en présence de discontinuités ou de variations brusques (comme les ondes de choc), la méthode FVM permet de traiter efficacement les régimes transitoires et non linéaires.

Aussi, par rapport à la méthode FEM, souvent plus complexe à mettre en œuvre pour des problèmes hyperboliques, les volumes finis offrent une structure plus simple et robuste, en particulier sur des maillages non structurés, ce qui en fait une méthode de choix pour la modélisation des écoulements à surface libre.

2.1.2 Principe de la méthode

La première chose va être de présenter le fonctionnement de la méthode des volumes finis. Elle repose sur l'intégration des équations de conservation sur un volume de contrôle (cellule du maillage), et sur l'application du théorème de Gauss (divergence).

Pour mieux illustrer, le point de départ sera l'équation de Saint-Venant établie au chapitre précédent.

$$\frac{\partial}{\partial t}U + \frac{\partial}{\partial x}F(U) + \frac{\partial}{\partial y}G(U) = S \quad . \quad (1.52)$$

L'équation commence par être intégrée sur un volume de contrôle Ω ,

$$\int_{\Omega} \frac{\partial}{\partial t}U \, d\mu + \int_{\Omega} \frac{\partial}{\partial x}F(U) + \frac{\partial}{\partial y}G(U) \, d\mu = \int_{\Omega} S \, d\mu \quad .$$

Elle est ensuite re-écrite sous forme de divergence,

$$\int_{\Omega} \frac{\partial}{\partial t}U \, d\mu + \int_{\Omega} \nabla \cdot \Phi(U) \, d\mu = \int_{\Omega} S \, d\mu \quad ,$$

où $\Phi(U) = (F(U), G(U))$. Puis le théorème de Gauss est appliqué pour obtenir,

$$\int_{\Omega} \frac{\partial}{\partial t}U \, d\mu + \int_{\partial\Omega} \Phi(U) \cdot \mathbf{n} \, d\mu = \int_{\Omega} S \, d\mu \quad . \quad (2.1)$$

L'étape suivante est de considérer un maillage non structuré $\mathcal{T} = \bigcup_i T_i$ (figure (2.1)) où chaque T_i représente un triangle du maillage tel que leur union forme la géométrie totale et leur intersection est nulle.

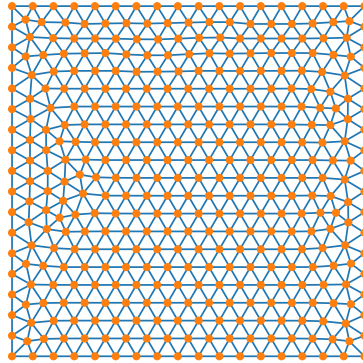


FIGURE 2.1 – Exemple de maillage triangulaire non structuré sur une cavité carrée

Sur un maillage de ce type, l'équation (2.1) peut se réécrire sur un triangle T_i du maillage et devient,

$$\int_{T_i} \frac{\partial}{\partial t}U_i \, d\mu + \int_{\partial T_i} \Phi(U_i) \cdot \mathbf{n}_i \, d\mu = \int_{T_i} S_i \, d\mu \quad .$$

Puis en développement d'avantage le calcul l'équation devient,

$$\frac{\partial}{\partial t}U_i \cdot \mu(T_i) + \sum_{j=1}^3 l_{ij} \cdot \mathfrak{F}(U_i, U_j) \cdot \mathbf{n}_{ij} = S_i \cdot \mu(T_i) \quad , \quad (2.2)$$

où $\mu(T_i)$ désigne l'air du triangle T_i , l_{ij} la longueur des l'arêtes j du triangle i , U_j les valeurs du triangle voisin de T_i sur l'arête j et \mathbf{n}_{ij} le vecteur normal unitaire sortant de T_i sur l'arête j .

La première chose à remarquer est la disparition du terme de flux $\Phi(U_i)$ par son remplacement par $\mathfrak{F}(U_i, U_j)$. L'explication derrière ce changement est assez simple. Le point principal est que, contrairement aux méthodes FEM et FDM où les valeurs d'états sont localisées sur les sommets, dans la méthode FVM les valeurs sont uniformisées sur chaque élément triangulaire T_i , comme illustré ci-dessous sur la figure (2.2).

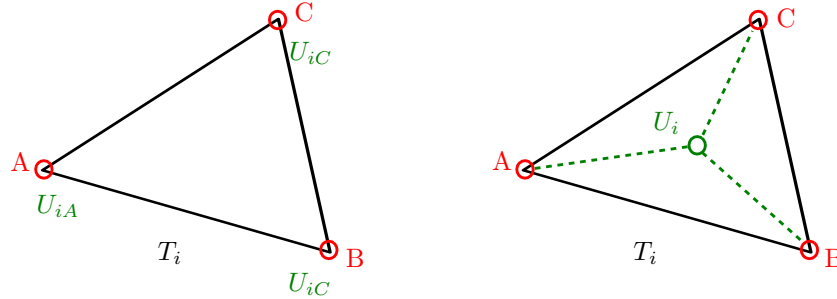


FIGURE 2.2 – Illustration de la différence entre la répartition nodale des méthodes FEM et FDM (gauche) et la répartition cellulaire de la méthode FVM (droite)

Cette répartition est souvent une moyenne des valeurs nodales. Il existe plusieurs façons de faire cette moyenne, mais dans le cas présent, c'est l'expression de la moyenne standard qui va être utilisée,

$$U_i = \frac{U_{iA} + U_{iB} + U_{iC}}{3} . \quad (2.3)$$

Cela a pour conséquence de provoquer des discontinuités le long des bords des éléments. Or, la méthode FVM repose justement sur les valeurs sur ces bords. Pour remédier à cette complication, un flux numérique $\mathfrak{F}(U_i, U_j)$ (voir figure (2.3)) va être introduit.

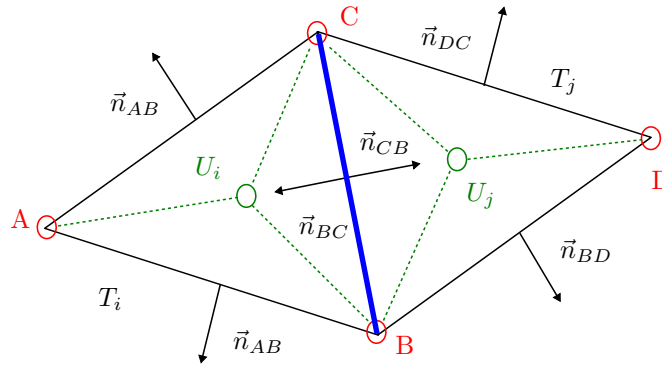


FIGURE 2.3 – Interaction entre deux triangles voisins avec l'introduction d'un flux numérique interfaciale (bleu)

L'idée est qu'il n'est pas possible d'utiliser directement $\Phi(U_i)$ dans la méthode des volumes finis parce qu'il s'agit d'un flux continu qui devient de fait inadapté entre T_i et T_j . La vraie solution à l'interface n'est pas connue avec la présence de discontinuités. Le flux numérique sert alors à approximer ce flux interfacial d'une manière stable, même en présence de chocs ou de discontinuités.

Pour définir ce flux, il existe plusieurs manières. Voici quelques exemples succincts des méthodes les plus répandues :

- Flux de Rusanov : C'est le flux le plus simple. Il prend la moyenne des flux physiques et ajoute une dissipation proportionnelle à la vitesse maximale d'onde. Très simple et robuste à mettre en place, mais fortement diffusif (adoucit trop les discontinuités, peu précis sur les fronts) [11].

- **Flux de Roe** : Flux qui approxime le système non linéaire par un système linéaire constant autour d'un état moyen de Roe, dont on peut calculer les valeurs propres et vecteurs propres. Précision plus forte, mais est plus complexe à implémenter et moins robuste avec des hauteurs d'eau h proches de 0 [7].
- **Flux Harten–Lax–van Leer (HLL)** : Flux qui approxime la solution par une onde lente S_L et une onde rapide S_R , ignorant les ondes intermédiaires. Cela donne un flux par morceaux constant. La méthode est plus précise encore que les deux précédentes, mais ne capture pas les ondes de contact (perte de finesse sur certaines structures) [4].
- **Flux HLL-Contact** : Flux qui est une extension du flux HLL qui ajoute une troisième onde : l'onde de contact S_M , pour mieux capturer les discontinuités dans u , v et h . Globalement plus performant que le flux HLL car il corrige ses lacunes, mais est très complexe à mettre en place et nécessite une formulation analytique spécifique au système étudié [13].
- **Flux de Godunov** : Flux qui résout exactement le problème de Riemann entre U_i et U_j . Il donne la connaissance des ondes générées, leur ordre, vitesse, etc. C'est une méthode de référence qui donne les valeurs exactes du système, mais a un coût très élevé et est difficile à mettre en place pour des systèmes non-linéaires (Euler, Saint-Venant) [3].

Pour synthétiser cette présentation de flux, le tableau (2.1) ci-dessous met en avant un résumé comparatif de ces méthodes.

Flux	Précision	Robustesse	Complexité	Remarques
Rusanov	★	★★★★	★	Simple et robuste, très diffusif
Roe	★★★★	★★	★★★★	Bon compromis précision/coût
HLL	★★	★★★★	★★	Robuste, mais néglige l'onde de contact
HLLC	★★★★	★★★★	★★★★	Précis et robuste, meilleur compromis
Godunov	★★★★★	★★	★★★★★	Référence, exacte mais coûteux

TABLE 2.1 – Comparaison qualitative des principaux flux numériques.

En résumé, le principe général de la méthode des volumes finis consiste à approximer les solutions moyennes dans des volumes de contrôle, en appliquant localement le principe de conservation sous forme intégrale, et en modélisant les échanges entre cellules par des flux numériques.

2.2 Méthode explicite

Après avoir présenté le cœur de la méthode générale des volumes finis, le détail va pouvoir être abordé, notamment avec la discrétisation temporelle. Comme dit précédemment, cette discrétisation sera faite par différences finies. Une méthode bien connue qui va être présentée succinctement à travers cette partie.

La première à être abordée va être une méthode simple, une méthode explicite.

2.2.1 Schéma des volumes finis explicite par flux de type Rusanov

Écriture du schéma explicite

Ce premier schéma qui va être présenté est l'un des plus accessibles et simples à mettre en place. Un schéma aux différences finies explicites avec un flux de Rusanov. Pour rappel, les différences finies visent à approximer des dérivés d'ordre k en utilisant des approximations de Taylor [10], de telle sorte qu'à l'instant n , la formule suivante

$$\frac{\partial U^n}{\partial t} \simeq \frac{U^{n+1} - U^n}{\Delta t} \quad , \quad (2.4)$$

où Δt est le pas temporel, soit une bonne approximation de l'ordre 1.

Également, un schéma explicite est un schéma où l'état futur $n + 1$ s'exprime comme une combinaison de l'état présent n ,

$$U^{n+1} = f(U^n) \quad . \quad (2.5)$$

Maintenant, en repartant de l'équation (2.2), et en appliquant les rappels des formules (2.4) et (2.5), l'équation à l'instant n s'écrit,

$$\frac{\partial}{\partial t} U_i^n \cdot \mu(T_i) + \sum_{j=1}^3 l_{ij} \cdot \mathfrak{F}(U_i^n, U_j^n) \cdot \mathbf{n}_{ij} = S(U_i^n) \cdot \mu(T_i) \quad , \quad (2.2)$$

puis devient,

$$\begin{aligned} \frac{U_i^{n+1} - U_i^n}{\Delta t} \cdot \mu(T_i) + \sum_{j=1}^3 l_{ij} \cdot \mathfrak{F}(U_i^n, U_j^n) \cdot \mathbf{n}_{ij} &= S(U_i^n) \cdot \mu(T_i) \quad , \\ U_i^{n+1} &= U_i^n - \frac{\Delta t}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \mathfrak{F}(U_i^n, U_j^n) \cdot \mathbf{n}_{ij} + \Delta t \cdot S(U_i^n) \quad . \end{aligned} \quad (2.6)$$

Flux de Rusanov

L'étape suivante est la présentation du flux de Rusanov $\mathfrak{F}(U_i^n, U_j^n)$, qui se définit par

$$\mathfrak{F}(U_i^n, U_j^n) = \frac{1}{2} \left(\Phi(U_i^n) + \Phi(U_j^n) \right) - \frac{\lambda_{\max ij}}{2} (U_j^n - U_i^n) \quad , \quad (2.7)$$

où $\lambda_{\max ij} = \max \left(|\Upsilon_i^n| + \sqrt{gh_i^n}; |\Upsilon_j^n| + \sqrt{gh_j^n} \right)$ avec $\Upsilon_i^n = \tilde{\Upsilon}_i^n \cdot \mathbf{n}_{ij}$ et $\tilde{\Upsilon}_i^n = \left(\frac{hu_i^n}{h_i^n}; \frac{hv_i^n}{h_i^n} \right)$.

Dans ce flux, deux parties distinctes peuvent être identifiées :

- $\frac{1}{2} \left(\Phi(U_i^n) + \Phi(U_j^n) \right)$, le terme conservatif.
- $\frac{\lambda_{\max ij}}{2} (U_j^n - U_i^n)$, le terme de diffusion numérique.

Il lisse les discontinuités pour stabiliser le schéma. Le terme $\lambda_{\max ij}$ représente la valeur maximale de la vitesse de propagation des ondes. Plus il est grand, plus le lissage est fort alors que plus $U_j^n - U_i^n$ est grande, plus la diffusion est forte.

Comme présenté en préambule, et en dépit de sa forte diffusion, c'est un flux qui permet un bon compromis entre simplicité de mise en œuvre et robustesse numérique pour une première approche.

Condition de stabilité

Enfin, comme la méthode traitée est une méthode explicite, il reste un dernier point à discuter. Il s'agit du critère de CFL (critère de Courant–Friedrichs–Lewy), qui vise à garantir la stabilité d'un schéma explicite [8]. Son but est de faire en sorte que l'information ne parcoure pas plus d'une maille pendant un pas de temps. Sinon, le calcul "manque" des données importantes, et des erreurs apparaissent et explosent.

Pour rappel, dans un cas très général, comme une équation de transport linéaire 2D,

$$\frac{\partial U}{\partial t} + \alpha \frac{\partial U}{\partial x} + \beta \frac{\partial U}{\partial y} = 0 \quad , \quad (2.8)$$

la condition CFL est donnée par :

$$\text{CFL} \simeq \Delta t \left(\frac{|\alpha|}{\Delta x} + \frac{|\beta|}{\Delta y} \right) \quad . \quad (2.9)$$

Dans ce cas, il y a stabilité et convergence du schéma si $\text{CFL} \leq 1$.

La problématique ici est plus complexe que ça, il n'y a pas de formule *a priori* aussi simple pour déterminer la CFL. Déjà, car l'équation traitée a un flux non linéaire, et aussi, parce que le maillage considéré est non-structuré, les valeurs Δx et Δy n'ont pas de sens dans ce cas. Pour déterminer la CFL c'est la méthode d'analyse spectrale qui va être utilisée. Elle consiste à linéariser l'opérateur du flux autour d'un état stationnaire et à étudier le spectre de l'opérateur qui en résulte. La condition de stabilité sera vérifiée si

$$\text{CFL} \simeq \rho(\mathcal{D}) < 1 \quad , \quad (2.10)$$

où $\rho(\mathcal{D})$ est le rayon spectral de l'opérateur \mathcal{D} , et implique la convergence.

Le point de départ commence par repartir du schéma numérique (2.6), pris sur un triangle T_i ,

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \mathfrak{F}(U_i^n, U_j^n) \cdot \mathbf{n}_{ij} + \Delta t \cdot S(U_i^n) \quad . \quad (2.6)$$

La première étape sera d'effectuer quelques simplifications utiles, en commençant par l'écriture,

$$U_i^{n+1} = U_i^n + \Delta t \left(\mathcal{L}(U_i^n) + S(U_i^n) \right) \quad , \quad (2.11)$$

où $\mathcal{L}(U_i^n) = \frac{-1}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \mathfrak{F}(U_i^n, U_j^n) \cdot \mathbf{n}_{ij}$.

Puis, une hypothèse concernant le terme source est introduite. L'hypothèse de source non raide, c'est-à-dire considérer que le terme source est "petit" devant le flux en termes de variation rapide. Il ne change pas significativement la stabilité, ni la condition CFL.

Ainsi, analyser la CFL de (2.11) est identique à analyser la CFL de l'équation

$$U_i^{n+1} = U_i^n + \Delta t \mathcal{L}(U_i^n) \quad , \quad (2.12)$$

ce qui va grandement simplifier l'analyse.

Maintenant vient le cœur de la méthode spectrale. Supposons avoir un état stationnaire \bar{U}_i sur un triangle T_i , tel que :

$$\bar{U}_i^{n+1} = \bar{U}_i^n = \bar{U}_i \implies \mathcal{L}(\bar{U}_i) = 0 \quad . \quad (2.13)$$

Nous introduisons ensuite une perturbation,

$$U_i^n = \bar{U}_i + \delta U_i^n \quad , \quad (2.14)$$

et linéarisons le terme $\mathcal{L}(U_i^n)$, ce qui donne,

$$\mathcal{L}(U_i^n) = \mathcal{L}(\bar{U}_i + \delta U_i^n) = \mathcal{L}(\bar{U}_i) + D\mathcal{L}(\bar{U}_i) \cdot \delta U_i^n + \mathcal{O}(\|\delta U_i^n\|^2) \quad . \quad (2.15)$$

Mais comme $\mathcal{L}(\bar{U}_i) = 0$, nous avons,

$$\mathcal{L}(U_i^n) = \mathcal{L}(\bar{U}_i + \delta U_i^n) \simeq D\mathcal{L}(\bar{U}_i) \cdot \delta U_i^n \quad . \quad (2.16)$$

En injectant les résultats précédents, l'équation (2.12) devient,

$$\bar{U}_i + \delta U_i^{n+1} \simeq \bar{U}_i + \delta U_i^n + \Delta t \cdot D\mathcal{L}(\bar{U}_i) \cdot \delta U_i^n \quad , \quad (2.17)$$

qui se simplifie en,

$$\delta U_i^{n+1} \simeq \delta U_i^n + \Delta t \cdot D\mathcal{L}(\bar{U}_i) \cdot \delta U_i^n \quad . \quad (2.18)$$

Et finalement,

$$\delta U_i^{n+1} \simeq \left(\text{Id} + \Delta t \cdot D\mathcal{L}(\bar{U}_i) \right) \cdot \delta U_i^n \quad . \quad (2.19)$$

Cette équation linéaire décrit l'évolution des petites perturbations autour de \bar{U} . Ainsi, deux cas peuvent être distingués. En posant $\mathcal{D} = \text{Id} + \Delta t \cdot \text{D}\mathcal{L}(\bar{U})$, il y a soit

- si $\rho(\mathcal{D}) < 1$, alors $\|\mathcal{D}^n\| \rightarrow 0$, donc les erreurs diminuent et le schéma est localement stable.
- si $\rho(\mathcal{D}) > 1$, alors les erreurs peuvent croître et le schéma est instable.

Dans le cas stable, il y a alors

$$\begin{aligned} \|\delta U_i^{n+1}\| &\leq \|\mathcal{D}\| \cdot \|\delta U_i^n\| + Cte \cdot \|\delta U_i^n\|^2 \\ &\leq \|\delta U_i^n\| + Cte \cdot \|\delta U_i^n\|^2 \end{aligned} \quad (2.20)$$

Ce qui montre que les perturbations restent bornées et même qu'elles décroissent. Donc, la stabilité du système linéarisé implique une stabilité locale du schéma non-linéarisé.

Après cette justification les éléments de la stabilité, elle peut être appliquée plus concrètement. L'idée se concentre donc sur l'analyse du spectre de \mathcal{D} , mais il est assez difficile de pouvoir le calculer en pratique. Cependant, il est plus facile d'approcher celui de $\text{D}\mathcal{L}(U)$. Pour rappel, sur un triangle T_i ,

$$\mathcal{L}(U_i^n) = \frac{-1}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \mathfrak{F}(U_i^n, U_j^n) \cdot \mathbf{n}_{ij} \quad , \quad (2.21)$$

et sans aller dans le détail pour éviter des calculs lourds, nous avons

$$\text{D}\mathcal{L}(U_i^n) = \frac{-1}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \text{D}\mathfrak{F}(U_i^n, U_j^n) \cdot \mathbf{n}_{ij} \quad , \quad (2.22)$$

avec

$$\text{D}\mathfrak{F}(U_i^n, U_j^n) \simeq \pm \frac{\lambda_{\max ij}}{2} \text{Id} \quad . \quad (2.23)$$

Rappel aussi que pour toute norme matricielle subordonnée, par la norme $\|\cdot\|_1$ ou $\|\cdot\|_\infty$ par exemple, nous avons

$$\rho(\text{D}\mathcal{L}) \leq \|\text{D}\mathcal{L}\|_p \quad . \quad (2.24)$$

Puis en combinant (2.22), (2.23) et (2.24) il s'ensuit que

$$\|\text{D}\mathcal{L}\|_\infty \leq \max_i \left(\frac{1}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \|\text{D}\mathfrak{F}(U_i^n, U_j^n)\| \right) \quad , \quad (2.25)$$

et

$$\rho(\text{D}\mathcal{L}) \leq \|\text{D}\mathcal{L}\|_\infty \leq \max_i \left(\frac{1}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \frac{\lambda_{\max ij}}{2} \right) = \Lambda_{\max} \quad . \quad (2.26)$$

En faisant la synthèse de toutes ces informations, il est désormais possible de conclure sur la condition CFL. En prenant $\xi_{L'}$ une valeur propre de $\text{D}\mathcal{L}$, nous avons,

$$\rho(\mathcal{D}) < 1 \quad \Longleftrightarrow \quad |1 - \Delta t \cdot \xi_{L'}| < 1 \quad ,$$

et par positivité et valeurs réelles strictes du spectre de $\text{D}\mathcal{L}$,

$$\Longleftrightarrow \quad 0 < \Delta t < \frac{2}{\xi_{L'}} \quad .$$

Par définition de $\rho(\mathcal{DL})$ et par l'inégalité (2.24), les valeurs précédentes peuvent être encadrées et aboutir à une majoration plus fine,

$$\xi_{L'} \leq \rho(\mathcal{DL}) \leq \Lambda_{\max} \iff \frac{2}{\Lambda_{\max}} \leq \frac{2}{\rho(\mathcal{DL})} \leq \frac{2}{\xi_{L'}} . \quad (2.27)$$

Après toutes ces étapes de calculs, il est finalement possible d'écrire la condition de CFL de façon explicite comme,

$$\Delta t < \frac{2}{\Lambda_{\max}} = \frac{2}{\max_i \left(\frac{1}{\mu(\mathcal{T}_i)} \sum_{j=1}^3 l_{ij} \cdot \frac{\lambda_{\max ij}}{2} \right)} ,$$

et de façon plus propre,

$$\Delta t < \min_i \left(\frac{4 \cdot \mu(\mathcal{T}_i)}{\sum_{j=1}^3 l_{ij} \cdot \lambda_{\max ij}} \right) . \quad (2.28)$$

Conclusion

En somme, le schéma des volumes finis explicite par flux de type Rusanov repose sur trois points. Il est centré sur les bases des différences finies et des volumes finis, profite d'un flux facile à appréhender et à implémenter et d'une condition CFL explicite (2.28) pour la stabilité. C'est la méthode idéale dans le cadre d'une initialisation aux méthodes FVM. Comme l'illustre le diagramme (2.4), il a également l'avantage d'avoir plusieurs variables qui peuvent être précalculées en amont et à chaque pas de temps. Ceci réduit les calculs à la simple résolution de l'équation triangle par triangle en bouclant sur le maillage.

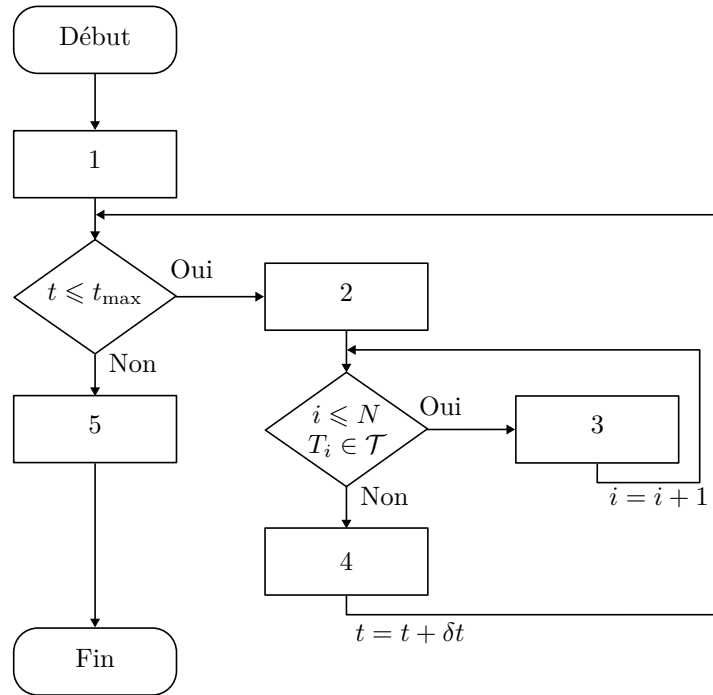


FIGURE 2.4 – Diagramme simplifié du schéma des volumes finis explicite par flux de type Rusanov

- 1) Calcul des données géométriques, aires des triangles, longueurs des arêtes, vecteurs normaux, voisins pour chaque arête.
- 2) Calcul du flux et valeurs sources pour tout les triangles T_i à l'instant t .
- 3) Résolution de l'équation (2.6) et calcul de l'énergie du système sur le triangle T_i à l'instant t .
- 4) Stockage de la solution complète de l'instant t et calcul de l'énergie totale du système.
- 5) Mise en forme des résultats.

Avec ces différents avantages, ce schéma remplit parfaitement son rôle pour une première approche de la résolution des équations de Saint-Venant. Dans le chapitre suivant, il sera mis en œuvre, comparé aux méthodes qui vont suivre et éprouvé à travers des évolutions.

2.3 Méthode implicite

Dans la deuxième partie de ce chapitre, c'est la philosophie des schémas implicites qui va être présentée. L'idée cette fois est d'exprimer l'état futur $n + 1$ comme une combinaison des états présent n et futur $n + 1$,

$$U_{n+1} = f(U_n, U_{n+1}) \quad . \quad (2.29)$$

La première chose qui apparaît à la vision de cette équation est la difficulté de devoir prendre en compte dans le calcul un état inconnu. Pour répondre à ce problème, deux méthodes vont ici être présentées.

2.3.1 Schéma des volumes finis implicite linéarisé par itérations de point fixe

Écriture du schéma implicite

Le premier schéma implicite qui va être présenté est un schéma qui est relativement proche du schéma des volumes finis explicite par flux de type Rusanov, mais qui s'adapte à la difficulté introduite par l'égalité (2.29). Il s'agit du schéma des volumes finis implicite linéarisé par itérations de point fixe.

Le point de départ est l'équation du cas explicite, toujours positionnée sur un triangle T_i du maillage,

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \mathfrak{F}(U_i^n, U_j^n) \cdot \mathbf{n}_{ij} + \Delta t \cdot S(U_i^n) \quad , \quad (2.6)$$

réécrite en tenant compte de la forme implicite,

$$U_i^{n+1} = U_i^n - \frac{\Delta t}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \mathfrak{F}(U_i^{n+1}, U_j^{n+1}) \cdot \mathbf{n}_{ij} + \Delta t \cdot S(U_i^{n+1}) \quad . \quad (2.30)$$

L'astuce pour débloquer la situation est de définir une fonction résidu de l'état U_i^{n+1} ,

$$\text{Res}_i(U^{n+1}) = U_i^{n+1} - U_i^n + \frac{\Delta t}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \mathfrak{F}(U_i^{n+1}, U_j^{n+1}) \cdot \mathbf{n}_{ij} - \Delta t \cdot S(U_i^{n+1}) \quad . \quad (2.31)$$

Le but est ensuite de rendre ce résidu nul, car dans ces conditions,

$$\text{Res}_i(U^{n+1}) = 0 \quad \Longleftrightarrow \quad U_i^{n+1} \text{ est solution de (2.30)}$$

Ceci permet de trouver une solution à l'écriture implicite qui semblait insolvable au premier abord, sans avoir à connaître précisément les valeurs $\mathfrak{F}(U_i^{n+1}, U_j^{n+1})$ et $S(U_i^{n+1})$.

L'objectif est donc,

$$\text{Res}_i(U^{n+1}) = 0 \quad . \quad (2.32)$$

Pour y arriver, l'équation va être linéarisée à travers la méthode itérative du point fixe. Les termes non linéaires de flux et de sources sont figés avec les valeurs de l'itération précédente m . De cette façon, l'équation obtenue devient linéaire en $U^{(m+1)}$, puis va pas à pas converger vers une solution U^{n+1} .

Voici comment se traduit le processus en pratique.

- Pour un pas de temps n , sur un triangle T_i , l'initialisation commence,

$$U_i^{(0)} = U_i^n \quad . \quad (2.33)$$

- Puis une itération construit l'état $U^{(m+1)}$,

$$U_i^{(m+1)} = U_i^n - \frac{\Delta t}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \mathfrak{F}(U_i^{(m)}, U_j^{(m)}) \cdot \mathbf{n}_{ij} + \Delta t \cdot S(U_i^{(m)}) \quad . \quad (2.34)$$

- L'opération est répétée jusqu'à,

$$\|U^{(m+1)} - U^{(m)}\| < \varepsilon \quad . \quad (2.35)$$

Conclusion

Sur le plan numérique, la différence fondamentale avec le schéma explicite est la construction des solutions. Contrairement au schéma explicite, dans lequel les flux numériques pouvaient être précalculés globalement à partir des états connus au temps n avant de procéder à la résolution cellule par cellule, la méthode implicite requiert une approche plus fine.

Comme l'illustre le diagramme (2.5), à chaque pas de temps, une solution globale estimée est construite "bloc par bloc", c'est-à-dire en mettant à jour simultanément l'ensemble des cellules en fonction des valeurs actuelles. Les itérations successives raffinent ensuite cette solution jusqu'à ce qu'un équilibre satisfaisant soit atteint.

- 1) Calcul des données géométriques, aires des triangles, longueurs des arêtes, vecteurs normaux, voisins pour chaque arête.
- 2) Initialisation du point fixe à l'instant n .
- 3) Calcul du flux et valeurs sources pour le triangle T_i à l'instant t .
- 4) Résolution de l'équation (2.34) sur le triangle T_i à l'instant t .
- 5) Calcul de la différence en norme de $U^{(m+1)}$ et $U^{(m)}$ avant l'évaluation (2.35).
- 6) Stockage de la solution complète de l'instant t et calcul de l'énergie totale du système.
- 7) Mise à jour de la solution globale estimée pour l'itération du point fixe suivante.
- 8) Mise en forme des résultats.

Ce que montre également le diagramme (2.5) c'est que ce schéma implicite est moins facile à mettre en place. En effet, le simple changement d'expression entre (2.5) et (2.29) aboutit à une grande évolution de la complexité. Il est moins intuitif à comprendre que sa version explicite et occasionne également un nombre de calculs bien plus élevé, dû aux itérations successives du point fixe.

Ceci dit, ces défauts ne viennent pas sans gains. Le schéma implicite détermine une solution au temps $n + 1$ en fonction d'elle-même. Le résultat est donc de base plus fidèle, avec un risque d'erreur grandement diminué, et permet de ne plus avoir besoin d'une condition CFL pour se

stabiliser. Il peut également avancer plus loin dans le temps, avec un pas Δt plus grand, sans perdre la cohérence de la solution.

En conclusion, le schéma des volumes finis implicite linéarisé par itérations de point fixe est plus stable et plus robuste que le précédent, bénéficie toujours d'un flux simple, celui de Rusanov, mais paie ses améliorations par un contrecoup notable en termes de calcul. C'est une option plus lourde mais plus précise pour la résolution des équations de Saint-Venant.

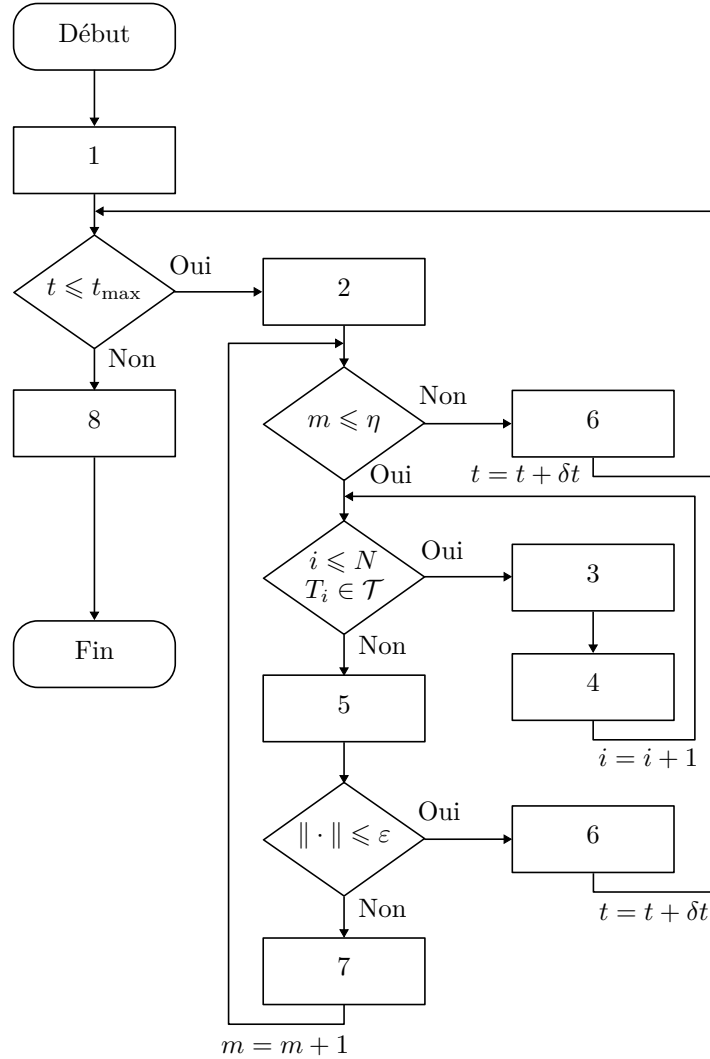


FIGURE 2.5 – Diagramme simplifié du schéma des volumes finis implicite linéarisé par itérations de point fixe

2.3.2 Schéma des volumes finis implicite avec résidu linéarisé et formulation matricielle globale

Écriture du schéma implicite

Le deuxième schéma implicite qui va être présenté est un schéma qui va encore plus loin que celui des volumes fins implicites linéarisés par itérations de point fixe. Bien plus robuste, il fait appel à des outils mathématiques plus pointus qui vont être présentés ici. Il s'agit du schéma des volumes finis implicites avec résidu linéarisé et formulation matricielle globale.

Le point de départ est, comme pour la méthode du point fixe, l'objectif

$$\text{Res}_i(U^{n+1}) = 0 \quad . \quad (2.32)$$

La différence cette fois est qu'au lieu de linéariser en figeant les termes non linéaires, la linéarisation va être faite par développement de Taylor autour d'un état connu U^n [15]. Ceci donne,

$$\text{Res}_i(U^{n+1}) \simeq \text{Res}_i(U^n) + \frac{\partial \text{Res}_i(U^n)}{\partial U} \cdot \delta U \quad . \quad (2.36)$$

où $\delta U = U^{n+1} - U^n$. Puis avec l'objectif (2.32),

$$\begin{aligned} \text{Res}_i(U^{n+1}) = 0 &\implies \text{Res}_i(U^n) + \frac{\partial \text{Res}_i(U^n)}{\partial U} \cdot \delta U = 0 \\ &\implies \frac{\partial \text{Res}_i(U^n)}{\partial U} \cdot \delta U = -\text{Res}_i(U^n) \quad . \end{aligned} \quad (2.37)$$

Ainsi, une solution de (2.32) sera issue de la résolution du système linéaire (2.37), obtenue elle aussi par itérations successives.

Voici comment cela se traduit,

- Pour un pas de temps n , l'initialisation commence,

$$U^{(0)} = U^n \quad . \quad (2.38)$$

- Puis le système (2.37) est résolue,

$$\delta U = - \left(\frac{\partial \text{Res}_i(U^{(m)})}{\partial U} \right)^{-1} \cdot \text{Res}_i(U^{(m)}) \quad , \quad (2.39)$$

avec $\delta U = U^{(m+1)} - U^{(m)}$.

- L'opération est répétée jusqu'à,

$$\|\delta U\| < \varepsilon \quad . \quad (2.40)$$

- Enfin, quand la condition (2.40) est vérifiée, U^{n+1} est mis à jour,

$$U^{n+1} = \delta U + U^{(m)} \quad . \quad (2.41)$$

Jacobien du résidu

Comme cela peut se voir, ce schéma implicite propose une manière plus sophistiquée de traiter le résidu et la résolution implicite elle-même. L'intervention du jacobien du résidu contribue à mieux contrôler la précision des solutions. Il n'est plus juste question de "deviner" la solution U^{n+1} comme avec la méthode précédente du point fixe, mais de l'atteindre en suivant la bonne direction mathématique.

Pour pouvoir exploiter ce jacobien dans l'implémentation de la méthode, il est nécessaire de pouvoir l'expliciter. Avec méthode et technique, l'expression (2.31) peut se dériver sans mal. Le résultat est le suivant,

$$\begin{aligned} \frac{\partial \text{Res}_i(U^{(m)})}{\partial U} &= \frac{\partial (U_i^{(m)} - U_i^n)}{\partial U} + \frac{\Delta t}{\mu(T_i)} \frac{\partial \left(\sum_{j=1}^3 l_{ij} \cdot \mathfrak{F}(U_i^{(m)}, U_j^{(m)}) \cdot \mathbf{n}_{ij} \right)}{\partial U} - \Delta t \frac{\partial (S(U_i^{(m)}))}{\partial U} \\ &= \text{Id}_3 + \frac{\Delta t}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \frac{\partial (\mathfrak{F}(U_i^{(m)}, U_j^{(m)}) \cdot \mathbf{n}_{ij})}{\partial U} - \Delta t \frac{\partial (S(U_i^{(m)}))}{\partial U} \end{aligned} \quad (2.42)$$

A partir d'ici, il va être nécessaire de procéder séparément. Le terme $\frac{\partial(S(U_i^{(m)}))}{\partial U}$ est assez complexe à écrire et dépend des choix faits pour les expressions internes de ses variables, il sera détaillé ultérieurement. Pour ce qui est du terme $\frac{\partial(\mathfrak{F}(U_i^{(m)}, U_j^{(m)}) \cdot \mathbf{n}_{ij})}{\partial U}$, il est lui aussi complexe mais peut se traiter par étapes.

Le point de départ se fait en rappelant que,

$$\mathfrak{F}(U_i^{(m)}, U_j^{(m)}) = \frac{1}{2}(\Phi(U_i^{(m)}) + \Phi(U_j^{(m)})) - \frac{\lambda_{\max ij}}{2}(U_j^{(m)} - U_i^{(m)}) \quad . \quad (2.7)$$

L'étape suivante est de dériver par rapport à U . Seulement, l'expression (2.7) fait intervenir U_i et U_j en même temps. La dérivation se fera alors en deux temps, une première fois par rapport à U_i puis une deuxième fois par rapport à U_j . Ceci donne,

$$\left. \frac{\partial \mathfrak{F}(U_i^{(m)}, U_j^{(m)})}{\partial U} \right|_{U_i} = \frac{1}{2} \frac{\partial \Phi(U_i^{(m)})}{\partial U_i} + \frac{\lambda_{\max ij}}{2} \text{Id}_3 \quad . \quad (2.43)$$

Puis,

$$\left. \frac{\partial \mathfrak{F}(U_i^{(m)}, U_j^{(m)})}{\partial U} \right|_{U_j} = \frac{1}{2} \frac{\partial \Phi(U_j^{(m)})}{\partial U_j} - \frac{\lambda_{\max ij}}{2} \text{Id}_3 \quad . \quad (2.44)$$

En injectant ces dérivés dans l'équation (2.42), et en arrangeant l'écriture de (2.37) sous forme matricielle pour mieux ordonner les interactions entre les valeurs, l'équation devient

$$\underbrace{\left(\text{Id}_3 + \frac{\Delta t}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \frac{\partial \mathfrak{F}_{ij}}{\partial U_i} - \Delta t \frac{\partial(S(U_i^{(m)}))}{\partial U} \right)}_{A_{ii}} \delta U_i + \underbrace{\left(\frac{\Delta t}{\mu(T_i)} \sum_{j=1}^3 l_{ij} \cdot \frac{\partial \mathfrak{F}_{ij}}{\partial U_j} \right)}_{A_{ij}} \delta U_j = -\text{Res}_i(U^{(m)}) \quad , \quad (2.45)$$

où

$$\frac{\partial \mathfrak{F}_{ij}}{\partial U_i} = \frac{1}{2} \left(\frac{\partial \Phi(U_i^{(m)})}{\partial U} \cdot \mathbf{n}_{ij} + \lambda_{\max ij} \text{Id}_3 \right) \quad , \quad (2.46)$$

et

$$\frac{\partial \mathfrak{F}_{ij}}{\partial U_j} = \frac{1}{2} \left(\frac{\partial \Phi(U_j^{(m)})}{\partial U} \cdot \mathbf{n}_{ij} - \lambda_{\max ij} \text{Id}_3 \right) \quad . \quad (2.47)$$

Avec la formule explicite de la dérivée du flux, sur un triangle T_i , qui s'écrit comme

$$\begin{aligned} \frac{\partial \Phi(U_i)}{\partial U} \cdot \mathbf{n}_{ij} &= \frac{\partial}{\partial U} (F(U_i) \cdot \mathbf{n}_x + G(U_i) \cdot \mathbf{n}_y) = \frac{\partial F(U_i)}{\partial U} \cdot \mathbf{n}_x + \frac{\partial G(U_i)}{\partial U} \cdot \mathbf{n}_y \\ &= \mathbf{n}_x \begin{pmatrix} 0 & 1 & 0 \\ gh - u^2 & 2u & 0 \\ -uv & v & u \end{pmatrix} + \mathbf{n}_y \begin{pmatrix} 0 & 0 & 1 \\ -uv & v & u \\ gh - u^2 & 0 & 2v \end{pmatrix} \quad . \end{aligned} \quad (2.48)$$

La matrice A_{ii} représente la sensibilité du résidu du triangle T_i à sa propre inconnue et la matrice A_{ij} décrit l'influence des triangles voisins T_j sur le triangle T_i .

Compacté, cela donne,

$$A_{ii}(U^{(m)}) \delta U_i + A_{ij}(U^{(m)}) \delta U_j = -\text{Res}_i(U^{(m)}) \quad . \quad (2.49)$$

L'explicitation du Jacobien du résidu met en évidence deux propriétés numériques essentielles de cette méthode implicite :

- La première est liée à la justesse et la fiabilité des résultats. La formule (2.37) montre que la correction suit une direction optimale vers l'état suivant à l'échelle globale. L'écriture locale (2.49) précise que cette optimisation s'applique aussi à chaque triangle T_i , où l'état est mis à jour en suivant une direction localement pertinente. Cela garantit à la fois une convergence rapide et le respect de la structure locale du maillage.
- La deuxième propriété est liée à l'efficacité de la résolution. L'écriture (2.49) révèle que le Jacobien est une matrice creuse (*sparse*), reflétant le couplage uniquement entre cellules voisines. Cette structure permet une faible consommation mémoire, une résolution accélérée via des méthodes itératives, et une excellente adaptabilité aux maillages non structurés, ce qui en fait un atout majeur pour les grandes simulations implicites.

Conclusion

Au final, le schéma des volumes finis implicite avec résidu linéarisé et formulation matricielle globale se distingue nettement par sa robustesse et sa stabilité, supérieures à celles des approches plus simples. Cet avantage repose sur l'introduction du jacobien du résidu, qui transforme la recherche de la solution au pas de temps suivant en un véritable processus d'optimisation directionnelle. Plutôt que d'avancer par approximations successives sans guide, cette méthode, appuyée sur le jacobien, oriente la mise à jour de U^{n+1} selon une direction mathématiquement optimale, ce qui réduit significativement le nombre d'itérations nécessaires à la convergence. Cette stratégie reste compatible avec la structure locale du maillage : chaque triangle T_i est corrigé en fonction de ses voisins directs uniquement, ce qui assure à la fois finesse locale et cohérence globale de la solution. La nature creuse du Jacobien rend possible l'utilisation de solveurs itératifs efficaces et permet de limiter la complexité malgré la taille du système.

Néanmoins, cette puissance a un coût. Le schéma reste implicite, et comme le montre le diagramme (2.6) chaque pas de temps exige plusieurs opérations lourdes : assemblage du Jacobien, résolution d'un système linéaire, et itérations internes pour la convergence. Malgré des optimisations structurelles (matrice creuse, solveurs dédiés) la méthode demeure exigeante en ressources mémoire et en temps de calcul. Ce compromis illustre bien le principe classique en calcul scientifique : gagner en précision et en stabilité nécessite d'accepter une complexité algorithmique accrue.

- 1) Calcul des données géométriques, airs des triangles, longueurs des arêtes, vecteurs normaux, voisins pour chaque arête.
- 2) Initialisation du point fixe à l'instant n .
- 3) Remplissage de la matrice du Jacobien et calcul du flux et valeurs sources du triangle T_i à l'instant t .
- 4) Résolution de l'équation (2.37) à l'instant t .
- 5) Calcul de la différence en norme de $U^{(m+1)}$ et $U^{(m)}$ avant l'évaluation (2.35).
- 6) Calcul de (2.41) et stockage de la solution complète de l'instant t et calcul de l'énergie totale du système.
- 7) Mise à jour de la solution globale estimée pour l'itération du point fixe suivante.
- 8) Mise en forme des résultats.

Pour conclure, le schéma des volumes finis implicite avec résidu linéarisé et formulation matricielle globale est redoutable en termes de précision. C'est la méthode la plus poussée et fiable de celles présentées jusqu'ici. Sa grande complexité de mise en place et son poids calculatoire ne rendent pas ce schéma accessible pour une première prise en main des méthodes de volumes finis, mais il est très adapté pour la résolution des équations de Saint-Venant.

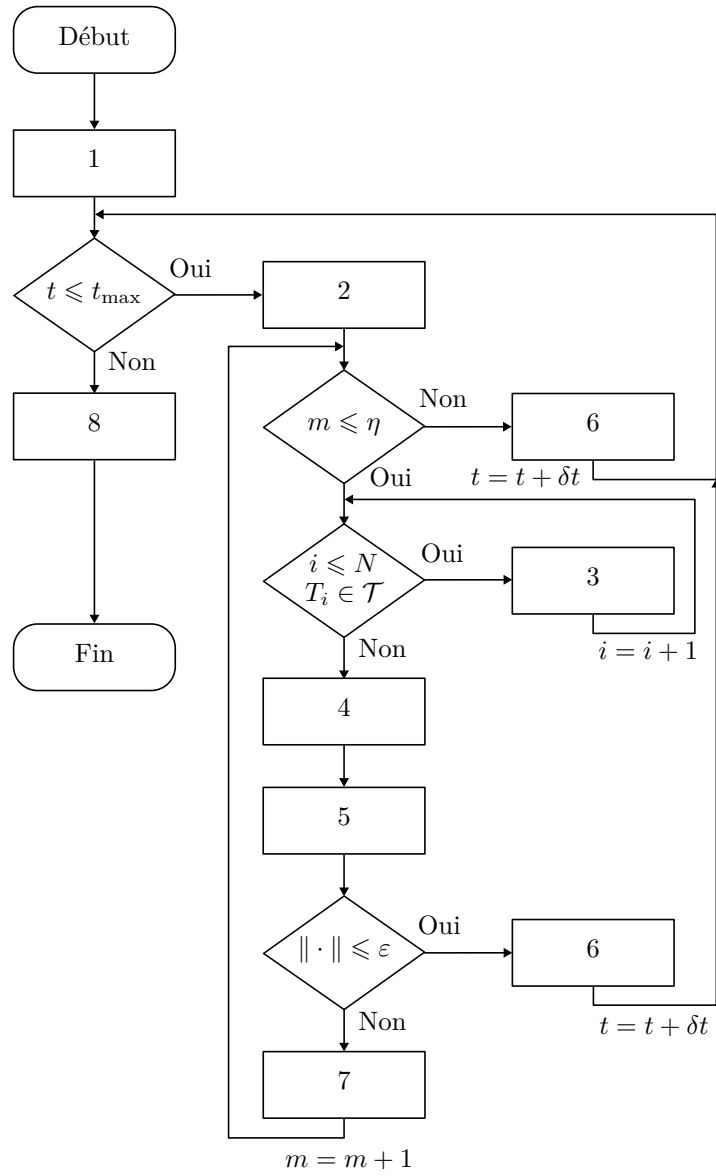


FIGURE 2.6 – Diagramme simplifié du schéma des volumes finis implicite avec résidu linéarisé et formulation matricielle globale

Chapitre 3

Application au cas d'une goutte perturbatrice

Après avoir introduit les équations de Saint-Venant et présenté les schémas numériques de résolution appropriés, il est maintenant venu le temps de la mise en œuvre.

Dans ce chapitre, la première partie va être consacrée à la présentation de tous les éléments de mise en place avant la résolution numérique. Cela permettra de pouvoir justifier des choix de modélisation ainsi que mettre en lumière des éléments pratiques de code. Puis, une analyse des résultats qui découlent des différents schémas numériques sera effectuée afin de pouvoir les discuter et les comparer.

3.1 Mise en place de la simulation numérique

Pour mener à bien une simulation, il est souvent requis de configurer un certain nombre de paramètres. Ils permettent de définir les conditions de la modélisation. Pour la bonne implémentation des schémas numériques, il est aussi nécessaire d'introduire des fonctions annexes qui viennent supporter le corps principal des méthodes.

Ce sont ces différents éléments qui vont être présentés ici.

3.1.1 Conditions et paramètres

En premier lieu, le point de départ va être de compléter l'équation (1.52), précédemment établie au chapitre 1.

$$\frac{\partial}{\partial t}U + \frac{\partial}{\partial x}F(U) + \frac{\partial}{\partial y}G(U) = S \quad , \quad (1.52)$$

avec,

$$U = \begin{pmatrix} h \\ hu \\ hv \end{pmatrix} \quad F(U) = \begin{pmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \\ huv \end{pmatrix} \quad G(U) = \begin{pmatrix} hv \\ huv \\ hv^2 + \frac{1}{2}gh^2 \end{pmatrix} \quad S = \begin{pmatrix} 0 \\ -gh \frac{\partial Z_f}{\partial x} + \frac{\tau_x^s}{\rho} - \frac{\tau_x^f}{\rho} \\ -gh \frac{\partial Z_f}{\partial y} + \frac{\tau_y^s}{\rho} - \frac{\tau_y^f}{\rho} \end{pmatrix}$$

Ces compléments vont être apportés progressivement par étapes.

Topologie du fond

C'est un élément qui n'apparaît que très peu à travers les équations et qui pourtant a un impact crucial. Le fond, avec sa géométrie influence directement l'accélération des flux et l'écoulement. Suivant sa complexité, il peut introduire des obstacles en surface qui influent sur les résultats.

Ce document vise à initier à la résolution des équations de Saint-Venant ; pour l'illustration, un fond plat est retenu,

$$z_f(x, y) = 0 \quad . \quad (3.1)$$

Termes de cisaillement

Le terme source avait volontairement été mis de côté dans les chapitres précédents, par souci de simplification. C'est ici qu'il va pouvoir être plus détaillé avec ses différentes composantes.

- Cisaillement à la surface :

Pour la simulation, la surface est considérée comme libre, c'est à dire qu'elle n'est soumise à aucune forme de pression. Aussi, le vent ambiant est négligeable face aux forces de la gravité, de frottement au fond ou de la pression. Pour ces raisons, l'hypothèse de surface libre implique que

$$\tau^s = (\tau_x^s, \tau_y^s) = (0, 0) \quad . \quad (3.2)$$

- Cisaillement dans le fond :

Contrairement à la surface, le frottement du fond ne peut pas être négligé. Il est fondamental pour modéliser de façon réaliste le comportement du fluide. Sans ça, les vitesses seraient non-amorties et les ondes qui en découlent ne rendraient compte d'aucune physique.

Pour adopter un comportement réaliste le cisaillement dans le fond est modélisé avec la loi de Manning [5].

$$\tau^f = (\tau_x^f, \tau_y^f) = \left(\rho g n^2 \frac{u \sqrt{u^2 + v^2}}{h^{\frac{4}{3}}}, \rho g n^2 \frac{v \sqrt{u^2 + v^2}}{h^{\frac{4}{3}}} \right) \quad , \quad (3.3)$$

où n correspond au coefficient de Manning.

Ces précisions apportées au terme source permettent aussi de pouvoir donner une expression au terme $\frac{\partial S(U)}{\partial U}$ qui avait été laissée en suspens durant la présentation du jacobien du résidu au chapitre précédent. Ceci donne,

$$\frac{\partial S(U)}{\partial U} = \begin{pmatrix} 0 & 0 & 0 \\ g n^2 \frac{10}{3} \frac{u \sqrt{u^2 + v^2}}{h^{\frac{7}{3}}} & -g \frac{n^2}{h^{\frac{7}{3}}} \frac{2u^2 + v^2}{\sqrt{u^2 + v^2}} & -g \frac{n^2}{h^{\frac{7}{3}}} \frac{uv}{\sqrt{u^2 + v^2}} \\ g n^2 \frac{10}{3} \frac{v \sqrt{u^2 + v^2}}{h^{\frac{7}{3}}} & -g \frac{n^2}{h^{\frac{7}{3}}} \frac{uv}{\sqrt{u^2 + v^2}} & -g \frac{n^2}{h^{\frac{7}{3}}} \frac{u^2 + 2v^2}{\sqrt{u^2 + v^2}} \end{pmatrix} \quad . \quad (3.4)$$

Condition initiale

Pour modéliser une goutte en 2D avec les équations de Saint-Venant, il faut représenter uniquement l'effet macroscopique de l'impact : une perturbation localisée de la surface et, éventuellement, une impulsion horizontale. Cette perturbation doit être lisse pour éviter les instabilités numériques, contrôler la masse et l'énergie initiales, et rester compatible avec les hypothèses du modèle. Une surélévation gaussienne, éventuellement associée à un champ de vitesse radial, répond à ces critères (voir figure 3.1) et permet une mise en œuvre simple et robuste.

$$U_0 = \begin{cases} h_0(x, y) = 1.2 e \left(-\frac{(x+1)^2 + (y+1)^2}{0.16} \right) & (3.5a) \\ u_0(x, y) = 0 & (3.5b) \\ v_0(x, y) = 0 & (3.5c) \end{cases}$$

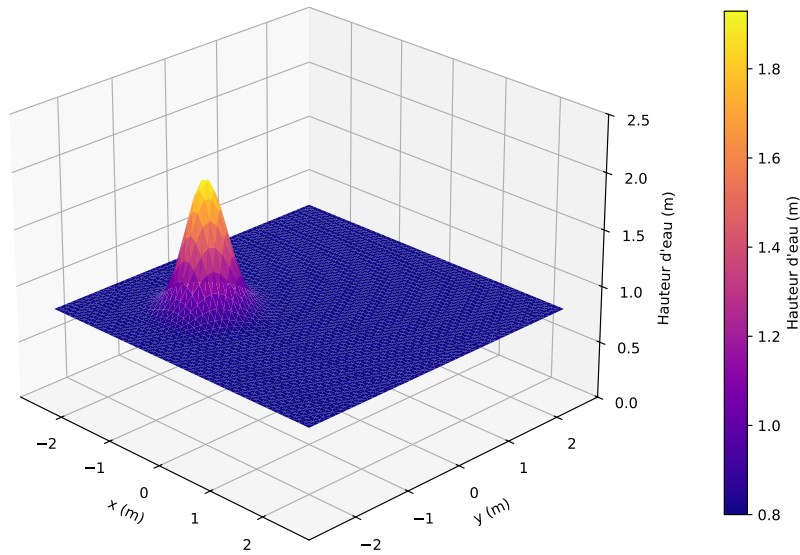


FIGURE 3.1 – Hauteur d’eau initiale choisie pour la modélisation d’une goutte d’eau perturbatrice

Ce choix de condition initiale est motivé par l’idée directrice de ce document, à savoir donner une introduction aux équations de Saint-Venant et la façon de les résoudre. Une perturbation provoquée par une goutte est une condition simple, qui reste assez visuelle, et qui permet de générer de légères vagues.

A la différence d’exemples comme la rupture de barrage, un écoulement trans-critique ou une vague de tsunami, c’est un choix qui reste abordable, tout en ayant une complexité minimale.

Condition aux limites

Afin de continuer sur la philosophie des conditions simples avec une légère complexité, le choix de la condition aux limites sera celui d’une paroi lisse, représentée par un bord miroir, comme l’illustre la figure (3.2). Cette configuration permet de générer des retours de vagues qui viennent se superposer aux autres, créant un défi numérique pour la modélisation.

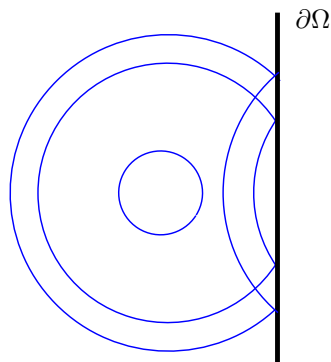


FIGURE 3.2 – Illustration de la condition aux limites pour une vague face à une paroi lisse

Plus généralement, cela revient à dire que l’eau ne traverse pas la paroi et que le mouvement est symétrique par réflexion. Cette symétrie impose que la composante de la vitesse normale à la paroi soit nulle ce qui, traduit en condition aux limites, s’écrit simplement

$$(u, v) \cdot \mathbf{n} = 0 \quad . \quad (3.6)$$

Toutes ces précisions apportées permettent de mettre à jour l'équation (1.52) et de former un système d'équations avec condition initiale et condition aux limites. C'est ce système qui va être résolu par les différents schémas présentés précédemment.

$$\begin{cases} \frac{\partial U}{\partial t} + \frac{\partial F(U)}{\partial x} + \frac{\partial G(U)}{\partial y} = S & , \text{ sur } \Omega \times [0, T] \\ U|_{t=0} = U_0 & , \text{ sur } \Omega \\ (u, v) \cdot \mathbf{n} = 0 & , \text{ sur } \partial\Omega \times [0, T] \end{cases} \quad (3.7)$$

avec,

$$U = \begin{pmatrix} h \\ hu \\ hv \end{pmatrix} \quad F(U) = \begin{pmatrix} hu \\ hu^2 + \frac{1}{2}gh^2 \\ huv \end{pmatrix} \quad G(U) = \begin{pmatrix} hv \\ huv \\ hv^2 + \frac{1}{2}gh^2 \end{pmatrix} \quad S = \begin{pmatrix} 0 \\ -\frac{\tau_x^f}{\rho} \\ -\frac{\tau_y^f}{\rho} \end{pmatrix}$$

3.1.2 Outils et fonctions auxiliaires

Dans un second temps, il y a un certain nombre de fonctions pratiques qui s'articulent autour du code des schémas vus au chapitre précédent et qui vont être présentées ici. Il ne s'agit pas de fonctions forcément complexes, mais elles mettent en forme ou structurent les idées de mise en œuvre pratique. Voici leur présentation.

Données géométriques

En premier lieu, il y a le calcul des données liées à la géométrie du maillage, à savoir l'aire des triangles et la longueur des arêtes. La seule chose à disposition pour ces calculs est la liste des coordonnées et la table de connectivité.

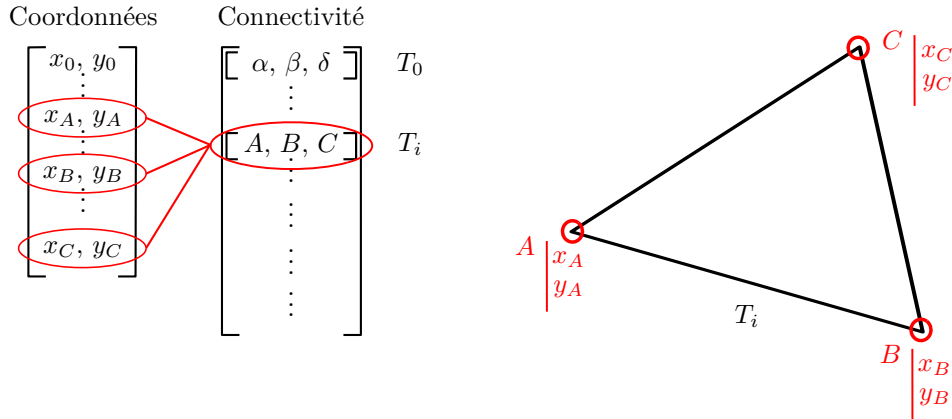


FIGURE 3.3 – Illustration de l'association d'un triangle T_i , représenté par ses sommets A , B , et C , avec les coordonnées correspondantes

Le raisonnement se fait en itérant sur la table de connectivité. Un triangle est associé à son triplet de coordonnées (voir figure (3.3)). Puis, les arêtes sont construites via les coordonnées,

$$\overrightarrow{AB} = \begin{pmatrix} x_B - x_A \\ y_B - y_A \end{pmatrix} \quad \overrightarrow{AC} = \begin{pmatrix} x_C - x_A \\ y_C - y_A \end{pmatrix} \quad \overrightarrow{BC} = \begin{pmatrix} x_C - x_B \\ y_C - y_B \end{pmatrix} . \quad (3.8)$$

Enfin, l'aire A_i du triangle T_i se calcule en utilisant le déterminant,

$$A_i = \frac{1}{2} \det(\overrightarrow{AB}, \overrightarrow{AC}) , \quad (3.9)$$

et $l_{\text{arête } i}$ la longueur des arêtes de T_i , se calcule avec la norme $\| \cdot \|_2$,

$$l_{\text{arête } i} = \left(\sqrt{x_{AB}^2 + y_{AB}^2}, \sqrt{x_{BC}^2 + y_{BC}^2}, \sqrt{x_{AC}^2 + y_{AC}^2} \right) . \quad (3.10)$$

Après avoir itéré, cela donne une liste des aires de tous les triangles et une liste pour la longueur de chaque arête (voir figure (3.4)). À noter que, pour faciliter l'utilisation dans le code, le calcul des arêtes doit toujours être effectué dans le même ordre.

Aires	Longueurs
$\begin{bmatrix} A_0 \\ \vdots \\ A_i \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$	$\begin{bmatrix} [l_{0,0}, l_{1,0}, l_{2,0}] \\ \vdots \\ [l_{AB}, l_{BC}, l_{CA}] \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$

FIGURE 3.4 – Listes des données géométriques d'aires et de longueurs d'arêtes obtenues après calcul

Structure de voisinage

Un autre point important pour les schémas numériques développés est de pouvoir déterminer les voisins de chaque triangle en fonction de chaque arête. Ce point est cruciale pour le calcul des flux intercellulaires.

Comme précédemment, le point de départ est la liste des coordonnées et la table de connectivité, mais la démarche va ici s'effectuer en deux étapes principales.

- Première étape, la construction d'une table des arêtes :

Chaque triangle est parcouru une fois et ses trois arêtes sont identifiées via ses sommets. Pour éviter les doublons elles sont représentées de façon ordonnée comme suit,

$$T_i = [A, B, C] \longrightarrow (A, B), (B, C), (A, C) .$$

Puis chaque arête est parcourue à son tour pour être ajoutée à un dictionnaire, adjoint avec le numéro du triangle auquel elle appartient. Une fois cette étape répétée pour chaque triangle, cela aboutit sur une table des arêtes illustrée par la figure (3.5).

$\begin{bmatrix} \vdots \\ (A, B) \\ (B, C) \\ (A, C) \\ \vdots \\ (K, L) \\ \vdots \end{bmatrix}$	$\begin{bmatrix} [i, j] \\ [i, k] \\ [i, r] \\ [z] \end{bmatrix}$
--	---

FIGURE 3.5 – Illustration d'exemple d'une table des arêtes

- Deuxième partie, association des voisins et arêtes

Chaque triangle est une nouvelle fois parcouru pour y identifier ses trois arêtes, comme à la première étape. Puis chaque arête est une nouvelle fois parcourue à son tour pour que à partir du dictionnaire précédemment établi, la liste des triangles associés à une arête soit récupérée.

Le triangle en cours d'examen est retiré de cette liste, et s'il reste un triangle il est identifié comme voisin et adjoint à l'arête en cours. Sinon, c'est une arête de bord et un NaN lui est adjoint. Ceci est répété pour chaque triangle, pour donner la liste des triangles voisins par arête pour chaque triangle, comme le montre la figure (3.6) ci-dessous.

$$T_i \rightarrow \begin{bmatrix} \vdots \\ [(j, (A, B)), (k, (B, C)), (r, (A, C))] \\ \vdots \\ [(\text{NaN}, (K, L)), (\text{NaN}, (L, M)), (w, (K, M))] \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$$

FIGURE 3.6 – Exemple d'une liste des voisins avec le cas exposé pour un triangle T_i

Enfin, pour que le calcul des flux soit complet, il manque l'association des vecteurs normaux à chaque arête. Pour les obtenir, c'est assez simple, et ils peuvent être calculés en parallèle durant le premier calcul des arêtes pour l'élaboration de la table des arêtes.

Pour une arête \overrightarrow{AB} d'un triangle $T_i = [A, B, C]$, son vecteur normal unitaire \mathbf{n}_{AB} s'obtient comme suit,

$$\overrightarrow{AB} = \begin{pmatrix} x_{AB} \\ y_{AB} \end{pmatrix} \quad \mathbf{n}_{AB} = \frac{1}{\|\overrightarrow{AB}\|} \begin{pmatrix} -y_{AB} \\ x_{AB} \end{pmatrix} . \quad (3.11)$$

Ensuite, il est nécessaire que le vecteur normal soit sortant. Voici un test qui permet de le vérifier et de le corriger si besoin.

Pour le triangle T_i deux points de référence sont construits. D'abord, le barycentre du triangle, β_{T_i} qui sera réutilisé pour toutes les arêtes. Puis, pour chaque vecteur normal testé, le milieu de l'arête relative est calculé, Mid_{AB} pour l'arête AB ,

$$\text{Mid}_{AB} = \frac{1}{2} \begin{pmatrix} x_A + x_B \\ y_A + y_B \end{pmatrix} \quad \beta_{T_i} = \frac{1}{3} \begin{pmatrix} x_A + x_B + x_C \\ y_A + y_B + y_C \end{pmatrix} . \quad (3.12)$$

Puis, un vecteur dit intérieur est construit, partant du milieu de l'arête en cours de traitement, AB , pointant vers le barycentre de T_i ,

$$\text{Vec}_{\text{int } AB} = \beta_{T_i} - \text{Mid}_{AB} . \quad (3.13)$$

L'angle formé par $\text{Vec}_{\text{int } AB}$ et \mathbf{n}_{AB} est ensuite testé pour déterminer l'orientation de \mathbf{n}_{AB} (voir figure (3.7)),

$$\begin{cases} \mathbf{n}_{AB} = -\mathbf{n}_{AB} & , \text{ si } \langle \text{Vec}_{\text{int } AB} ; \mathbf{n}_{AB} \rangle > 0 \\ \mathbf{n}_{AB} = \mathbf{n}_{AB} & , \text{ sinon} \end{cases} \quad (3.14)$$

Si l'angle est aigu, le vecteur opposé est le vecteur normal unitaire sortant, sinon le vecteur initial est conservé.

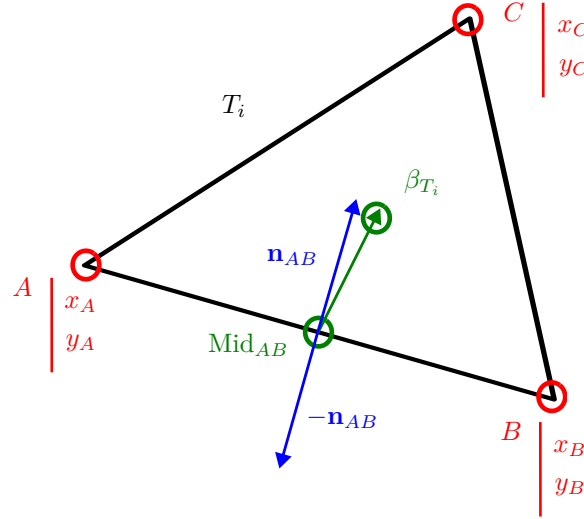


FIGURE 3.7 – Exemple du teste du vecteur normal \mathbf{n}_{AB} sur l'arête AB et de sa correction $-\mathbf{n}_{AB}$

Cette démarche est ainsi répétée sur toutes les arêtes, pour former un triplet de vecteurs normaux associés à chaque triangle, avec les vecteurs normaux classés dans le même ordre que les arêtes sont construites.

Une remarque pratique peut être faite à ce stade. Les outils présentés jusqu'à présent sont tous des éléments qui peuvent être pré-calculés en amont, car ils ne dépendent pas des données de la résolution elle-même. Ils sont également tous déterminés à partir de la même base, à savoir la liste des coordonnées et la table de connectivité.

Pour toutes ces raisons, et d'un point de vue pratique, il est tout à fait possible de pré-calculer tous ces éléments via une seule fonction globale.

Bords miroirs

Dans le calcul des flux il y a un autre point pratique qui s'impose durant la résolution. Pour pouvoir le calculer sur un triangle T_i il est nécessaire d'avoir accès aux données de ses voisins. La détermination des voisins a été présentée ci-dessus, mais il y a un cas qu'il reste à traiter : celui où il n'y a pas de voisin, le cas du bord.

Comme décrit dans la section dédiée aux conditions aux limites, le bord doit être réfléchissant. Cela veut dire que la hauteur d'eau au bord est inchangée mais que les vitesses sont, elles, réfléchies.

Voici le procédé. Pour un triangle T_i sur le bord les vitesses sont isolées dans un vecteur à part,

$$\mathcal{V} = (u_i, v_i) \quad . \quad (3.15)$$

Puis ce vecteur est décomposé sur le vecteur normal de l'arête concernée,

$$\mathcal{V} = \underbrace{(\mathcal{V} \cdot \mathbf{n})\mathbf{n}}_{\text{composante normale}} + \underbrace{(\mathcal{V} - (\mathcal{V} \cdot \mathbf{n})\mathbf{n})}_{\text{composante tangentielle}} \quad . \quad (3.16)$$

La composante tangentielle est parallèle au bord et correspond au glissement le long du bord. La composante normale est perpendiculaire au bord et mesure combien le fluide "pousse" dans la direction du bord. C'est cette dernière composante qui va être modifiée.

Pour modéliser la condition sur le bord la composante normale doit être inversée, ceci modélise un comportement de rebond sur le bord de la part du fluide.

Cela donne une vitesse miroir,

$$\begin{aligned}\mathcal{V}_{\text{mir}} &= -(\mathcal{V} \cdot \mathbf{n})\mathbf{n} + (\mathcal{V} - (\mathcal{V} \cdot \mathbf{n})\mathbf{n}) \\ &= \mathcal{V} - 2(\mathcal{V} \cdot \mathbf{n})\mathbf{n}\end{aligned}\tag{3.17}$$

Ainsi, pour un triangle T_i sur le bord avec ses valeurs d'état $U_i = (h_i, \mathcal{V})$, les valeurs d'état de son "voisin" T_j sont $U_j = U_{\text{mir}} = (h_i, \mathcal{V}_{\text{mir}})$. C'est avec ces valeurs que le flux intercellulaire est calculé sur le bord. Durant le calcul des flux un simple test permet de traiter les deux cas et déterminer les valeurs U_{voi} du voisin,

$$\begin{cases} U_{\text{voi}} = U_j & , \text{ si Voisin} \neq \text{NaN} \\ U_{\text{voi}} = U_{\text{mir}} & , \text{ sinon} \end{cases},$$

où j désigne l'indice du triangle voisin T_j dans le cas où il existe, et U_{mir} les valeurs d'état miroir dans le cas où il n'y a pas de voisin à T_i .

3.2 Analyse des résultats de la simulation numérique

En tenant compte des outils présentés précédemment et de leur inclusion dans le code, la modélisation peut être faite. De cette façon, les résultats des simulations utilisant les schémas introduits dans ce document peuvent être présentés. Plus exactement, et pour des raisons de lisibilité comparative, seul le schéma des volumes finis explicite par flux de Rusanov et le schéma des volumes finis implicite linéarisé par itération de point fixe vont être présentés.

Les principaux paramètres de modélisation utilisés vont d'abord être exposés pour préciser les conditions dans lesquelles les simulations ont été réalisées et garantir la reproductibilité des expériences numériques. Puis, les résultats et les performances des différents schémas numériques vont être illustrés au moyen de représentations et comparaisons graphiques pour être discutés.

3.2.1 Résultat des modélisations

Dans les deux schémas numériques, la simulation a été réalisée dans des conditions identiques, de manière à permettre une comparaison directe par la suite. Les calculs se déroulent dans une cavité carrée de 5 mètres de côté, centrée en $(0;0)$. Le tableau suivant (3.1) présente de façon synthétique les principaux paramètres utilisés au cours des simulations.

Paramètre	Valeur	Unité	Description
Δt	5.0e-6	seconde (s)	Pas de temps
Δx	1.0e-1	mètre (m)	Diamètre maximal des triangles du maillage
N_{temps}	652	NA	Nombre de pas temporel
T_{max}	3.12e-3	second (s)	Temps maximal de la simulation
ρ	1000	kg/m ³	Densité de l'eau
g	9.81	m/s ²	Accélération de la pesanteur
n_{Maning}	2.0e-2	s/m ^{$\frac{1}{3}$}	Coefficient de Maning
iter_{max}	10	NA	Nombre d'itération maximal pour la résolution du point fixe

TABLE 3.1 – Valeurs des paramètres des simulations

Voici les résultats pour le schéma des volumes finis explicite par flux de Rusanov, avec en premier, l'animation de l'évolution de la hauteur d'eau (voir vidéo (3.8)), suivie de l'évolution de l'énergie totale du système à travers le temps (voir figure (3.9)).

Animation - Schéma Explicite ($t = 0.00000$ s)

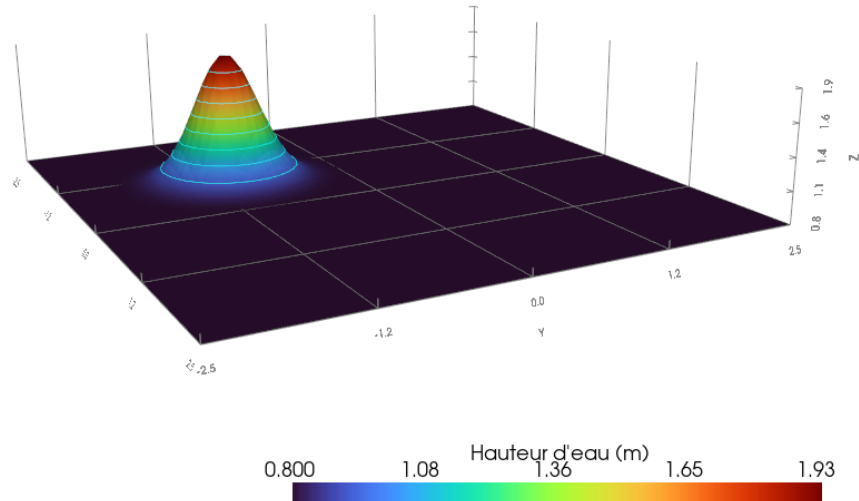


FIGURE 3.8 – Aperçu cliquable de l'animation du schéma explicite par flux de Rusanov

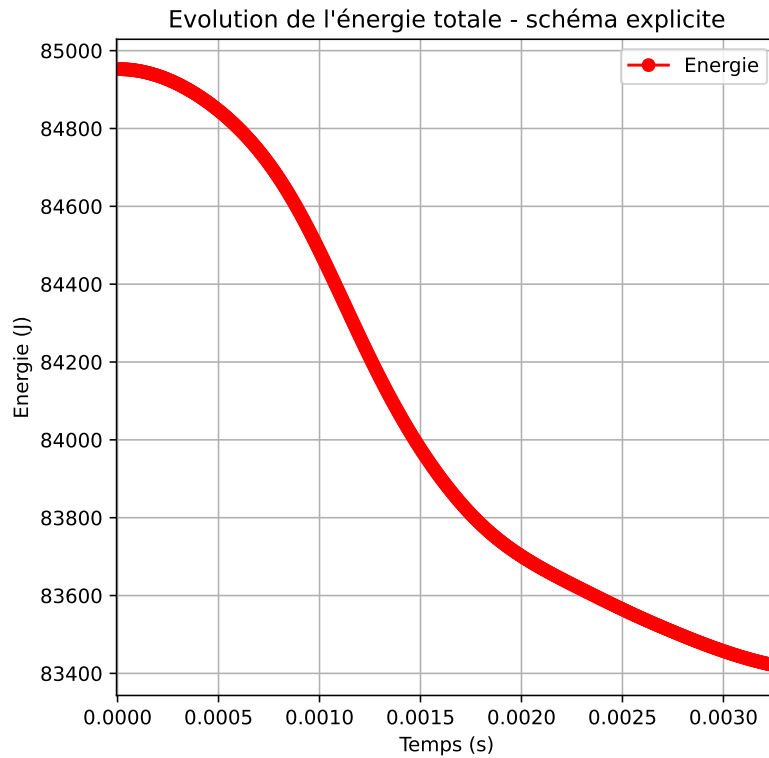


FIGURE 3.9 – Évolution de l'énergie totale du système pour le schéma explicite par flux de Rusanov

Puis, de la même façon, voici les résultats pour le schéma des volumes finis implicite linéarisé par itération de point fixe (voir vidéo (3.10) et figure (3.11)).

Animation - Schéma Implicite (Point Fixe)
($t = 0.00000$ s)

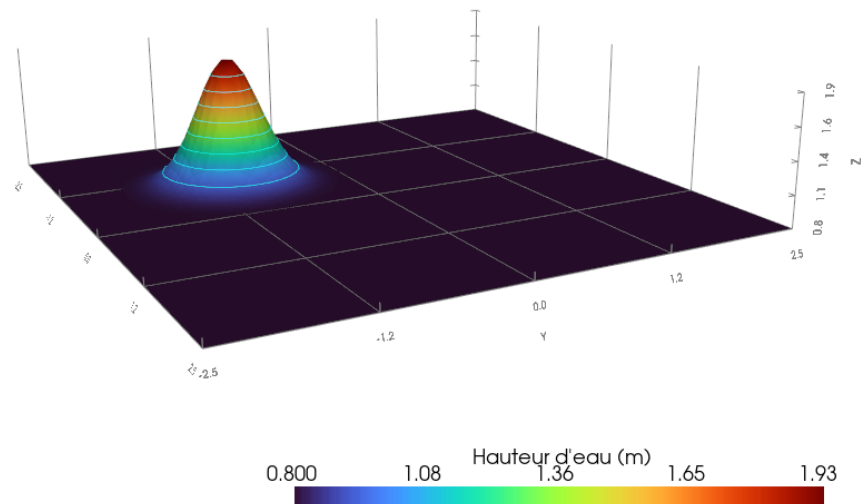


FIGURE 3.10 – Aperçu cliquable de l’animation du schéma implicite linéarisé par itération de point fixe

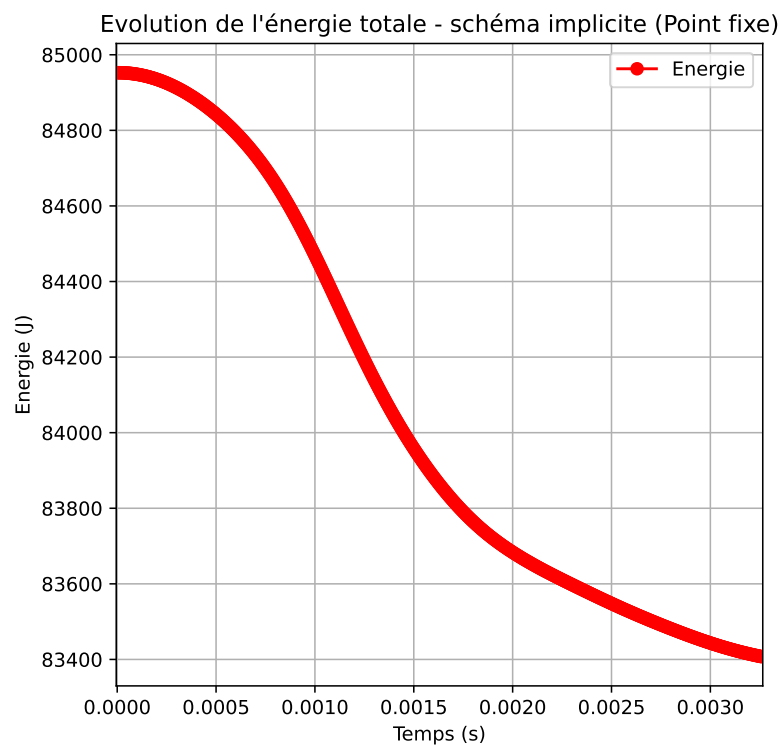


FIGURE 3.11 – Évolution de l’énergie totale du système pour le schéma implicite linéarisé par itération de point fixe

Pour le calcul de l'énergie totale du système modélisé par (3.7), la formule est la suivante,

$$E_{\text{tot}} = \underbrace{\frac{1}{2}h(u^2 + v^2)}_{\text{énergie cinétique}} + \underbrace{\frac{1}{2}gh^2}_{\text{énergie potentielle}} . \quad (3.18)$$

Cette formule est celle qui serait utilisée pour le calcul d'une solution analytique. Pour obtenir l'énergie totale numérique illustrée dans les graphiques (3.9) et (3.11), c'est une version discrète de l'équation (3.18) qui est utilisée. Il s'agit d'une somme des énergies locales de chaque triangle T_i , pondérée par l'aire A_i correspondante,

$$E_{\text{tot}} = \sum_i \frac{A_i}{2} \left(h_i(u_i^2 + v_i^2) + gh_i^2 \right) . \quad (3.19)$$

La formule (3.19) est utilisée à chaque pas de temps de la résolution pour chaque schéma afin d'aboutir aux résultats présentés dans les graphiques.

3.2.2 Analyse et comparaison des schémas numériques

Afin d'évaluer la pertinence et la robustesse des approches présentées, une comparaison des résultats numériques obtenus par les deux méthodes est réalisée. Cette analyse permettra de mettre en évidence leurs points communs et leurs écarts.

Cependant, une légère difficulté demeure. Dans le cas des équations de Saint-Venant il n'existe pas de solution analytique. Sans cela, il est impossible de mesurer directement l'erreur absolue de chaque méthode. Les résultats obtenus seront donc confrontés entre eux, de manière à mettre en évidence leurs différences et leur cohérence.

Pour commencer cette comparaison, c'est une différence point par point, à chaque pas de temps, qui va être examinée et faite de la façon suivante,

$$\text{Diff}_i(t) = h_i^{\text{exp}}(t) - h_i^{\text{Imp}}(t) . \quad (3.20)$$

Animation - Différence entre les schémas
(t = 0.00000 s)

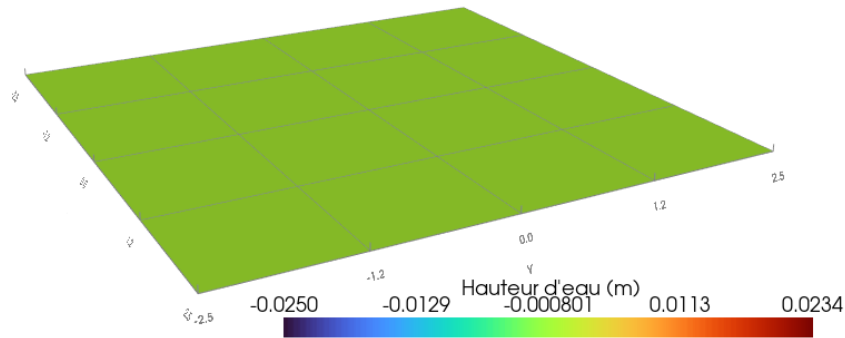


FIGURE 3.12 – Aperçu cliquable de l'animation de la différence point par point entre les deux schémas numériques présentés

Ci-dessus se trouve l'animation (3.12) qui correspond à l'évolution de (3.20) dans le temps. L'un des premiers constats est que les variations entre les deux méthodes sont relativement faibles, comprises entre ± 2.5 cm pour la hauteur d'eau. Cependant, cette fourchette est extrême. Les variations sont plus faibles en moyenne sur l'ensemble du maillage. Cette animation permet donc d'apprécier un comportement global des différences, mais reste imprécise comme méthode d'analyse.

Pour proposer une analyse plus fine, plusieurs calculs vont être effectués afin d'apporter des points concrets de comparaison. Qu'ils puissent être lisibles et clairement interprétables.

Moyenne et variations locales

Ce premier calcul a pour but d'éliminer les disparités parasites présentes sur le maillage. Comme le montre la formule (3.21), il consiste à moyenner l'ensemble des $\text{Diff}_i(t)$ à chaque pas de temps et de les pondérer par l'aire A_i des triangles correspondants.

$$\text{Diff}_{\text{Moy}}(t) = \frac{\sum_i \text{Diff}_i(t) \cdot A_i}{\sum_i A_i} . \quad (3.21)$$

Cette formule est plus efficace pour l'analyse des résultats que la simple différence point par point. Elle permet de donner une tendance globale, là où l'animation précédente est plus une carte de l'évolution des erreurs locales. Pour apporter des précisions supplémentaires, la dispersion de la différence va également être calculée en utilisant la variance,

$$\text{Var}(t) = \frac{\sum_i \left(\text{Diff}_i(t) - \text{Diff}_{\text{Moy}}(t) \right)^2 \cdot A_i}{\sum_i A_i} , \quad (3.22)$$

et l'écart-type,

$$\sigma(t) = \sqrt{\text{Var}(t)} . \quad (3.23)$$

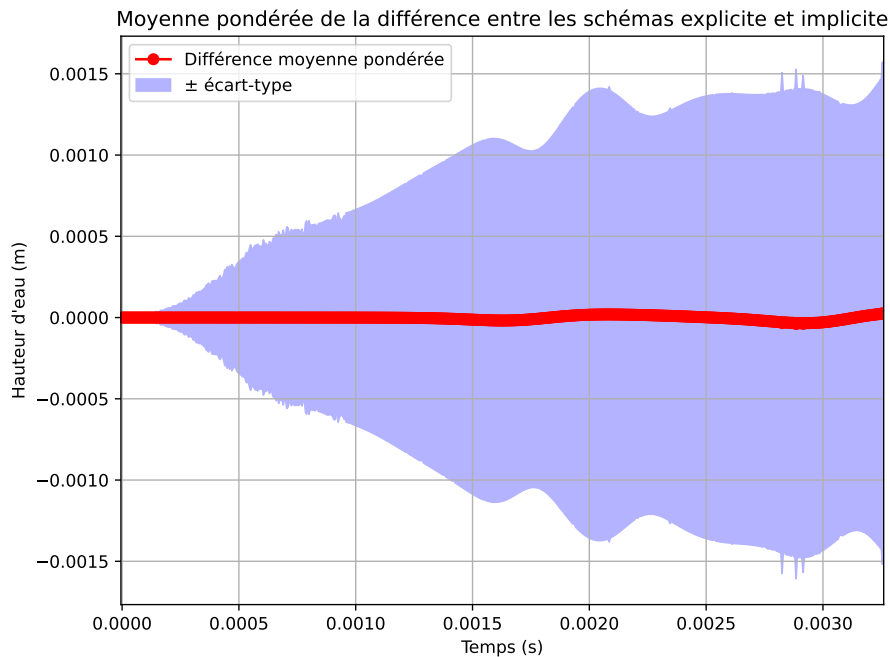


FIGURE 3.13 – Évolution de la moyenne pondérée sur la maillage de la différence entre le schéma explicite et implicite

Le signe de $\text{Diff}_{\text{Moy}}(t)$ permet de donner une indication sur le biais moyen entre les méthodes.

- si $\text{Diff}_{\text{Moy}}(t) > 0$, le schéma explicite tend à surestimer la hauteur d'eau vis à vis du schéma implicite.
- si $\text{Diff}_{\text{Moy}}(t) < 0$, le schéma explicite tend à sous-estimer la hauteur d'eau vis à vis du schéma implicite.
- si $\text{Diff}_{\text{Moy}}(t) \simeq 0$, il n'y a pas de biais global, mais il peut y avoir des différences très localisées.

Puis, l'écart-type $\sigma(t)$ mesure la dispersion des données autour de $\text{Diff}_{\text{Moy}}(t)$.

- Plus $\sigma(t)$ est petit, plus les données sont regroupées autour de la moyenne $\text{Diff}_{\text{Moy}}(t)$.
- Plus $\sigma(t)$ est grand, plus les données sont très dispersées.

Avec le graphique (3.13) il est clair que $\text{Diff}_{\text{Moy}}(t) \simeq 0$ et l'écart-type confirme que les disparités sont faibles localement. Les deux schémas produisent donc des résultats quasi similaires en moyenne.

Cependant, un comportement croissant au fil du temps peut être observé de la part de l'écart type. Même si cette croissance est faible et reste contenue, elle indique que la différence moyenne observée s'annule globalement toujours, mais tend à avoir des écarts localisés.

Cela peut s'expliquer par le fait d'une accumulation numérique normale. Une autre possibilité est que le schéma explicite nécessite des paramètres de modélisation Δt et Δx qui satisfassent la condition CFL précédemment présentée, et que ce schéma tendrait à devenir sensible sur des temps plus longs. Ceci impactant donc la différence faite avec le schéma implicite, qui, lui, est plus stable.

Évaluation en norme

Le calcul suivant vise à caractériser les différences spatiales entre les deux schémas. Pour être précis, son but est d'observer comment les schémas diffèrent à travers le temps.

Pour évaluer ces différences, il est classique de calculer les normes du champ de différences, $\text{Diff}(t)$. Ces normes sont utilisées parce qu'elles fournissent une mesure quantitative, reproductible et indépendante du maillage de l'écart entre solutions, ce qui permet de comparer objectivement les méthodes sans se limiter à une inspection visuelle des résultats.

Dans cette analyse, ce sont les normes suivantes qui vont être calculées.

- La norme L_1 ,

$$\|\text{Diff}(t)\|_{L_1} = \frac{\sum_i |\text{Diff}_i(t)| \cdot A_i}{\sum_i A_i} . \quad (3.24)$$

Elle mesure l'erreur absolue moyenne sur le domaine. Elle est robuste et directement liée à une idée intuitive de différence moyenne.

- La norme L_2 ,

$$\|\text{Diff}(t)\|_{L_2} = \sqrt{\frac{\sum_i \text{Diff}_i(t)^2 \cdot A_i}{\sum_i A_i}} . \quad (3.25)$$

Elle met davantage en valeur les écarts importants grâce au carré des différences, ce qui en fait une mesure sensible aux erreurs localement marquées et reliée à une notion d'énergie de l'écart.

- La norme L_∞ ,

$$\|\text{Diff}(t)\|_{L_\infty} = \max_i (|\text{Diff}_i(t)|) \quad . \quad (3.26)$$

Elle renseigne sur le pire cas local, c'est-à-dire l'écart maximal entre les deux solutions, ce qui permet de détecter des divergences ponctuelles mais potentiellement critiques.

En couvrant les principaux points de vue nécessaires : la tendance globale, l'importance typique et l'extrême local, ces trois normes sont privilégiées pour l'analyse qui suit.

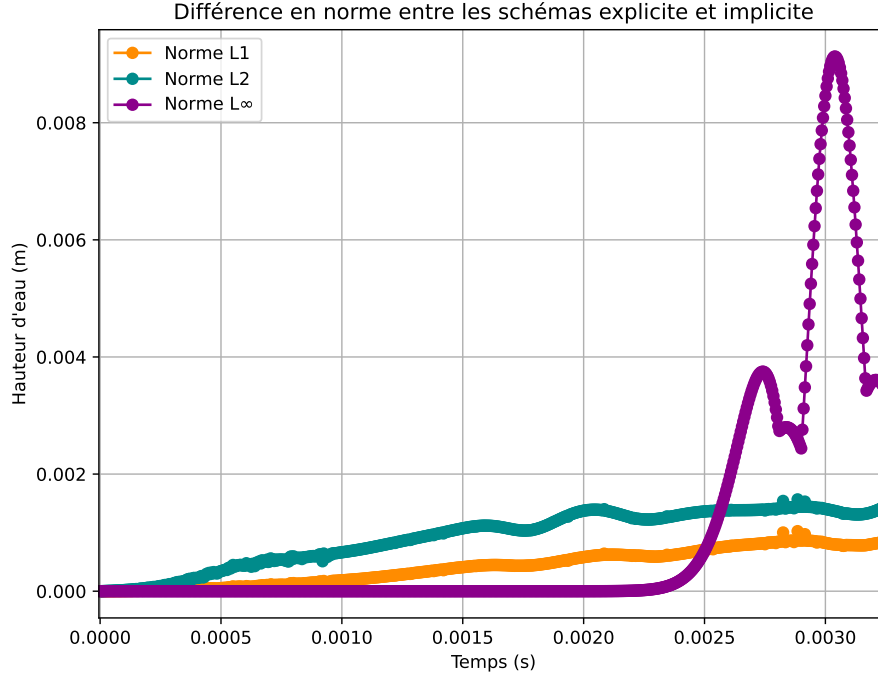


FIGURE 3.14 – Évolution des normes L_1 , L_2 et L_∞ de la différences entre le schéma explicite et implicite

Le graphique (3.14) ci-dessus expose les résultats. Un premier constat global est que, indépendamment de la norme, les valeurs se caractérisent par une faible intensité, de l'ordre de $1\text{e-}3\text{m}$, et une certaine stabilité, sans fluctuations critiques.

La norme L_1 et L_2 sont contenues entre 0.0m et $2.0\text{e-}3\text{m}$ sur toute la longueur de la simulation. Ceci confirme une fois de plus que l'écart entre les deux schémas numériques est très faible en valeur absolue, et qu'ils produisent des résultats pratiquement identiques sur l'ensemble du domaine et du temps. Également, un autre fait notable est que $L_2 > L_1$, ce qui est attendu en raison de la plus forte sensibilité aux écarts ponctuels de la norme L_2 . Cela précise aussi qu'il existe bien quelques zones plus sensibles où les schémas s'écartent localement.

Enfin, la norme L_∞ présente un comportement différent des deux précédentes. En effet, elle se découpe en deux temps. Un premier où elle est nulle sur les deux premiers tiers de la simulation, puis des oscillations croissantes fortes sur le dernier tiers. Ceci veut dire que jusqu'aux deux tiers de la simulation, les schémas ont produit des résultats concordants, mais qu'au-delà, il y a des différences très localisées qui se déplacent sur le domaine.

Ceci peut s'expliquer assez facilement. Dans la simulation, le phénomène de déplacement correspond à un effet de vague. En analysant l'instant de transition de la norme L_∞ et en le comparant avec l'animation (3.12), le constat montre qu'il s'agit vraisemblablement d'un effet de frontière. En effet, la réflexion des vagues sur les bords, selon le schéma numérique utilisé, peut engendrer un déphasage entre les deux schémas, qui se conserve et se propage au cours du temps.

Pour affiner l'analyse de la norme L_∞ et vérifier les hypothèses formulées, le signe des variations sera examiné en utilisant,

$$\text{Min}_{\text{Diff}}(t) = \min_i \left(\text{Diff}_i(t) \right) \quad \text{Max}_{\text{Diff}}(t) = \max_i \left(\text{Diff}_i(t) \right) . \quad (3.27)$$

Ces informations supplémentaires permettent de déterminer plusieurs comportements,

- Si $\text{Max}_{\text{Diff}}(t) > 0$, $\text{Min}_{\text{Diff}}(t) \simeq 0$, le schéma explicite surestime partout la hauteur d'eau (respectivement $\text{Max}_{\text{Diff}}(t) \simeq 0$, $\text{Min}_{\text{Diff}}(t) < 0$ le schéma implicite sous-estime partout la hauteur d'eau)
- Si $\text{Max}_{\text{Diff}}(t) > 0$, $\text{Min}_{\text{Diff}}(t) < 0$, parfois le schéma explicite est supérieur au schéma implicite et parfois l'inverse. C'est le signe d'une oscillation ou d'un déphasage entre les valeurs des deux schémas.

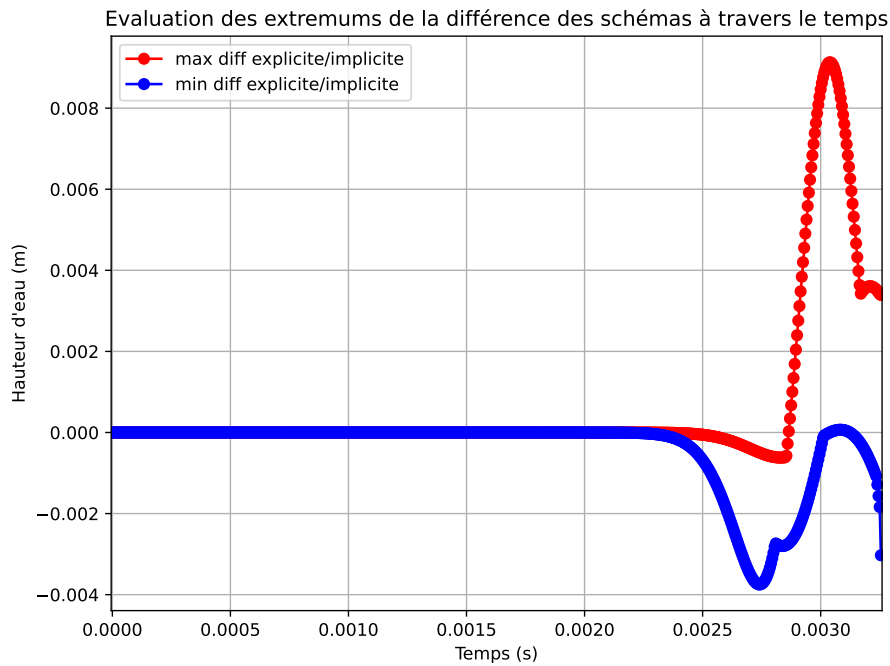


FIGURE 3.15 – Évolution du minimum et maximum de la différences entre le schéma explicite et implicite

Le graphique (3.15) illustre clairement que le système se situe dans le cas d'un déphasage, ce qui confirme les suppositions formulées. Néanmoins, il est observé que ces variations restent comprises entre un maximum de $8.0\text{e-}3\text{m}$ et un minimum de $4.0\text{e-}3\text{m}$. En confrontant ces résultats à ceux obtenus avec les normes L_1 et L_2 , il apparaît une nouvelle fois que les deux schémas fournissent des résultats proches, avec quelques écarts locaux et de déphasage pouvant être attribués à leurs différences de robustesse et de sensibilité.

Conservation de la masse

Le prochain calcul constitue un indicateur direct de la qualité de conservation de la masse par les schémas numériques. Il s'agit du calcul de l'écart ΔM de la masse par rapport à sa moyenne,

$$\Delta M(t) = \frac{M(t) - M_{\text{Moy}}}{M_{\text{Moy}}} \quad , \quad (3.28)$$

où

$$M(t) = \sum_i h_i(t) A_i \quad \quad M_{\text{Moy}} = \frac{1}{T} \sum_{k=0}^T M(t_k) \quad .$$

Ce calcul permet de vérifier si les schémas numériques reproduisent fidèlement la contrainte physique de conservation. Plusieurs cas sont alors possibles,

- $\Delta M \simeq 0$ pour tout t . La masse reste proche de la moyenne, la conservation est totale, le schéma est stable.
- ΔM oscille autour de 0 avec des amplitudes faibles. La masse globale est bien conservée, mais il existe de petites oscillations numériques (typiquement dues au schéma ou au pas de temps).
- ΔM dérive progressivement. Il y a une perte ou un gain systématique de masse. Si la dérive est monotone, une instabilité progresse.
- ΔM oscille grandement de façon irrégulière. Le schéma n'est pas bien équilibré et souffre de problèmes de stabilité importants.

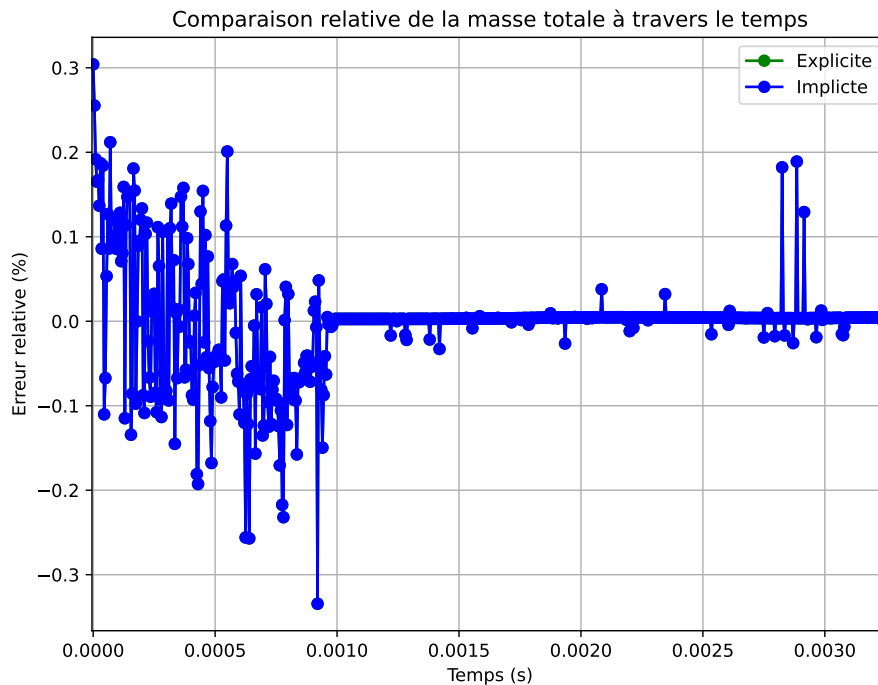


FIGURE 3.16 – Comparaison entre le schéma explicite et implicite de la conservation de la masse

Les résultats illustrés par le graphique (3.16) montrent deux choses. La première est que, pour le schéma explicite par flux de Rusanov ou le schéma implicite linéarisé par itération de point fixe, les résultats sont identiques. La deuxième est le comportement des ΔM .

Le premier tiers de la simulation présente de nombreuses oscillations autour de 0, mais d'amplitude limitée ($\pm 0.3\%$). Les deux tiers suivants se révèlent quant à eux majoritairement stables autour de 0. Ce comportement s'explique par une condition initiale non stationnaire, fortement éloignée de l'équilibre, qui engendre ces oscillations avant leur stabilisation progressive.

Au bilan, la conservation de la masse est respectée de manière équivalente par les deux schémas. Cet indicateur ne permet donc pas de discriminer leur qualité, mais confirme que chacun d'eux est bien conforme à la propriété physique attendue.

Analyse de l'énergie

Le dernier calcul tourne autour de l'interprétation de l'énergie totale pour les deux schémas. Tout d'abord, les graphiques (3.9) et (3.11) présentés en début de section montrent que les deux cas ont un comportement quasi identique, caractérisé par une décroissance amortie tout au long de la simulation. Ce résultat est attendu dans la mesure où les deux schémas reposent sur un flux de Rusanov, de nature dissipative, et où le terme source intègre un frottement de fond de type Manning.

Maintenant, afin d'évaluer objectivement la cohérence énergétique entre les deux schémas, en l'absence de modèle de référence et sans introduire de biais, l'erreur symétrisée sur l'énergie est calculée,

$$\Delta E_{\text{sym}}(t) = \frac{2|E_{\text{exp}}(t) - E_{\text{imp}}(t)|}{|E_{\text{exp}}(t)| + |E_{\text{imp}}(t)|} . \quad (3.29)$$

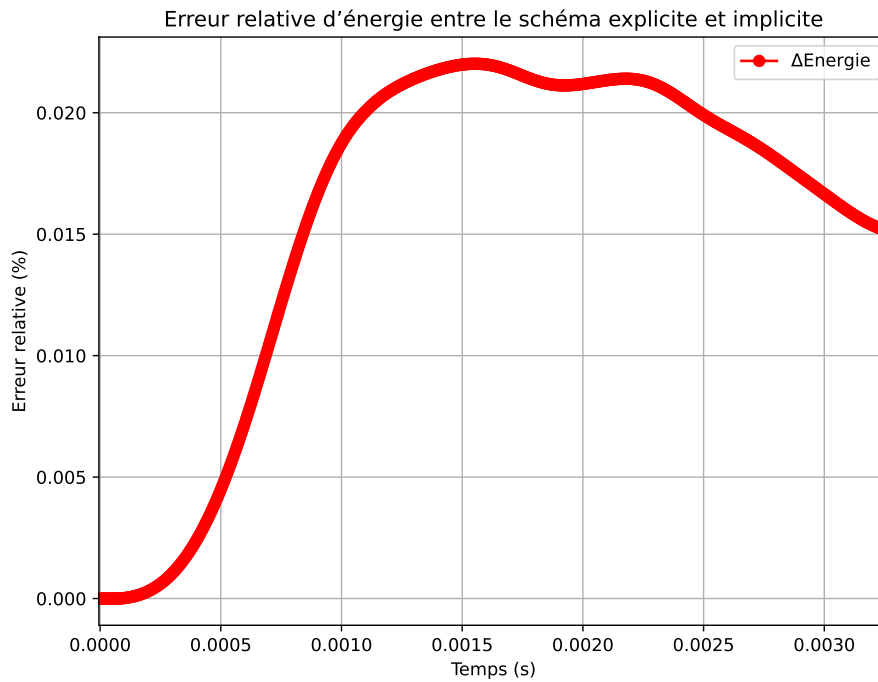


FIGURE 3.17 – Evolution de la différence symétrisée entre les énergies du schéma explicite et du schéma implicite

Ce calcul permet de quantifier la stabilité et la robustesse des schémas, ainsi que leur cohérence. Plusieurs cas sont donc possibles,

- $\Delta E_{\text{sym}}(t)$ reste proche de 0 de façon stable. Les deux méthodes produisent une évolution énergétique quasiment identique. Cela implique une grande cohérence entre schémas.

- $\Delta E_{\text{sym}}(t)$ croît de façon lente et régulière. L'écart se creuse entre les schémas. Traduit un potentiel effet d'accumulation numérique. Plus la pente est douce plus les schémas sont robustes.
- $\Delta E_{\text{sym}}(t)$ oscille autour d'une valeur. Les deux schémas suivent la même dynamique d'énergie mais avec des décalages de phase locaux
- $\Delta E_{\text{sym}}(t)$ augmente vite voir avec des pics marqués. Indique que, dans certaines configurations, un des deux schémas dissipe ou amplifie artificiellement l'énergie.

L'analyse de l'erreur symétrisée sur l'énergie montre avec le graphique (3.17) que l'écart relatif entre les deux méthodes numériques reste extrêmement faible, toujours compris entre 0 et 0,02 % sur toute la durée de la simulation. Une telle amplitude indique que les deux schémas conservent ou dissipent l'énergie de manière pratiquement identique. Ce résultat suggère que, du point de vue du bilan énergétique global, les deux approches sont équivalentes et qu'aucune ne présente de biais ou de dérive significative par rapport à l'autre.

Conclusion

L'ensemble des analyses menées, qu'il s'agisse des normes L_1 , L_2 et L_∞ , de la carte d'erreur des différences spatiales, de la conservation de la masse, ou encore du bilan énergétique, convergent toutes vers le même constat : les deux schémas numériques examinés, explicite et implicite, produisent des résultats pratiquement indiscernables à l'échelle globale. Les écarts observés demeurent faibles, généralement limités à de très petites variations locales. Ces différences peuvent s'expliquer par la nature des méthodes elles-mêmes (traitement explicite vs implicite), mais ne traduisent ni dérive numérique, ni défaut de conservation notable.

En l'absence de solution analytique de référence, cette cohérence entre deux approches indépendantes constitue un indicateur fort de la fiabilité des résultats. Autrement dit, le fait que deux schémas reposant sur des résolutions distinctes aboutissent à des solutions quasi identiques renforce la confiance dans la véracité physique et numérique des simulations.

Ainsi, la conclusion est que les deux schémas sont valides et robustes pour ce problème, et que le choix entre eux peut légitimement se fonder sur des considérations pratiques (stabilité CFL, coût de calcul, facilité d'implémentation) plutôt que sur la précision intrinsèque des solutions produites.

Cependant, une extension de ce travail est possible. Ces résultats encourageants ouvrent plusieurs pistes d'investigation. D'une part, il serait pertinent d'évaluer l'impact du choix du flux numérique : l'utilisation de schémas moins diffusifs (HLLC, Roe, ou autres flux approx.) pourrait révéler des différences plus marquées dans la conservation de l'énergie ou la précision locale. D'autre part, l'introduction d'un fond variable permettrait de tester la robustesse des deux approches dans un contexte plus réaliste et de mieux cerner leurs limites en présence de termes topographiques.

Enfin, la troisième méthode de résolution implicite, basée sur une linéarisation de type Newton et le calcul du jacobien du résidu, pourrait être mise en concurrence avec les deux précédentes, avec les pistes évoquées. Une telle extension fournirait une vision plus complète de la performance relative des différentes stratégies numériques pour la résolution des équations de Saint-Venant.

Bibliographie

- [1] R BSL. Byron bird & warren e. stewart & edwin n. lightfoot. *Transport Phenomena*,, 2002.
- [2] B De St Venant. Theorie du mouvement non-permanent des eaux avec application aux crues des rivers et a l'introduction des marees dans leur lit. *Academic de Sci. Comptes Redus*, 73(99) :148–154, 1871.
- [3] S Godounov. A difference method for numerical calculation of discontinuous solutions of the equation of hydrodynamics. *Matematicheskii Sbornik*, 47(89-3) :271–306, 1959.
- [4] Amiram Harten, Peter D Lax, and Bram van Leer. On upstream differencing and godunov-type schemes for hyperbolic conservation laws. *SIAM review*, 25(1) :35–61, 1983.
- [5] Robert Manning. On the flow of water in open channels and pipes. *Transactions, Institution of Civil Engineers of Ireland*, 20 :161, 1891.
- [6] CLMH Navier. Mémoire sur les lois du mouvement des fluides. *Mémoires de l'Académie Royale des Sciences de l'Institut de France*, 6(1823) :389–440, 1823.
- [7] Philip L Roe. Approximate riemann solvers, parameter vectors, and difference schemes. *Journal of computational physics*, 43(2) :357–372, 1981.
- [8] Abner J Salgado and Steven M Wise. *Classical Numerical Analysis*. Cambridge University Press, 2022.
- [9] George Gabriel Stokes. printed in 1847. *On the theories of the internal friction of fluids in motion and of the equilibrium and motion of elastic solids.* Transactions of the Cambridge Philosophical Society, 8 :287–319, 1845.
- [10] Endre Süli and David F Mayers. *An introduction to numerical analysis*. Cambridge university press, 2003.
- [11] Eleuterio F Toro. *Riemann Solvers and Numerical Methods for Fluid Dynamics : A Practical Introduction*. Springer Science & Business Media, 2009.
- [12] Eleuterio F Toro. *Riemann solvers and numerical methods for fluid dynamics : a practical introduction*. Springer Science & Business Media, 2013.
- [13] Eleuterio F Toro, Michael Spruce, and William Speares. Restoration of the contact surface in the hll-riemann solver. *Shock waves*, 4(1) :25–34, 1994.
- [14] D. J Tritton. *Physical fluid dynamics*. Springer Science & Business Media, 2012.
- [15] Edmund Taylor Whittaker and George Neville Watson. *A course of modern analysis*. Courier Dover Publications, 2020.