Problématiques d'analyse numérique et de modélisation pour écoulements de fluides environnementaux

Mathieu Cathala

THÈSE DE DOCTORAT DE
l'UNIVERSITÉ MONTPELLIER 2

Spécialité

Mathématiques et modélisation

École doctorale Information, Structures et Systèmes

Présentée par

Mathieu CATHALA

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ MONTPELLIER 2

Problématiques d'analyse numérique et de modélisation pour écoulements de fluides environnementaux

---

soutenue le 18 octobre 2013
devant le jury composé de :

M. Franck BOYER	Rapporteur
M. Didier BRESCH	Rapporteur
M. Daniele DI PIETRO	Président
M. David LANNES	Codirecteur de thèse
M. Fabien MARCHE	Codirecteur de thèse
M. Bijan MOHAMMADI	Directeur de thèse

Remerciements

Ces trois années de thèse ont été pour moi l'occasion d'user un nombre impressionnant d'encadrants, tous déjà fort occupés par ailleurs. Ma gratitude envers leur disponibilité n'a d'égal que la honte qui me ronge d'avoir abusé d'une grande partie de leur "temps de cerveau disponible". En premier lieu, je tiens à remercier Jérôme Droniou pour m'avoir d'abord gentiment accueilli durant mes différents stages puis pour m'avoir offert la possibilité de réaliser cette thèse. Je n'aurais peut-être jamais eu envie de goûter à l'analyse numérique s'il n'avait pris le temps de m'y initier et de m'en faire partager sa conception. Au gré des aventures australes de Jérôme, j'ai ensuite eu la chance d'aller tremper les pieds dans les eaux (peu profondes) de Fabien Marche et David Lannes. Avec toute la gentillesse qui le caractérise, Fabien a d'abord su trouver du temps pour me proposer un beau projet, m'offrant l'opportunité d'aller naviguer en toute liberté parmi un large choix de thématiques. Avec Fabien vint David, autre modèle de simplicité et de gentillesse. Le monde de David Lannes est composé de journées qui durent quatre-vingt-seize heures. C'est, du moins, l'unique solution que je trouve pour expliquer comment on peut être à ce point occupé et sollicité de tous bords tout en restant aussi zen et disponible, même pour un thésard débarqué de nulle part. Un grand merci à vous deux ! Pour terminer cette longue liste d'encadrants, je tiens à remercier très sincèrement Bijan Mohammadi pour avoir gardé un œil attentif sur mon évolution après le départ de Jérôme. Merci à lui pour m'avoir systématiquement accordé du temps lors de ses courts passages sur Montpellier et ce malgré de lourdes responsabilités administratives.

Je n'aurais pas pu soutenir cette thèse sans le travail de Franck Boyer et Didier Bresch, l'un des deux ayant même dû menacer de me sequestrer au beau milieu des Landes pour m'inciter à achever mon travail, en dépit de mes envies d'ailleurs. C'est un honneur pour moi qu'ils aient lu avec tant d'attention mon manuscrit. Je remercie également Daniele Di Pietro d'avoir eu l'amabilité d'accepter de faire partie de mon jury. J'ai été agréablement surpris et très honoré de l'intérêt qu'il a porté à mes quelques travaux depuis son arrivée à Montpellier.

J'ai eu l'occasion de collaborer avec Christophe Le Potier et Clément Cancès au cours de ma thèse. N'ayant pas vraiment l'âme d'un chercheur, je reconnaîs néanmoins que j'ai pris beaucoup de plaisir à travailler avec eux et je les en remercie. En premier lieu, merci à Christophe pour m'avoir offert la possibilité de découvrir puis de partager avec lui sa foultitude d'idées ingénieuses et innovantes ; cet homme est le Géo Trouvetou des schémas

monotones ! Merci ensuite à Clément, d'abord pour nous avoir offert son aide, ensuite pour son accueil à Paris (et pour le resto mexicain !), puis tout simplement pour sa gentillesse.

J'ai ici une pensée pour tous les étudiants auprès desquels j'ai eu l'occasion de faire mes classes comme enseignant durant ces trois dernières années, ces quelques heures ont été comme une grande bouffée d'air frais. Au sein de l'équipe ACSIOM, je remercie en particulier Vanessa Lleras, Matthieu Alfaro (et le club qu'il supporte), Simon Mendez, Pascal Azerad et Rémi Carles. Merci aussi à Bernadette Lacan pour tout ce qu'elle accomplit, ainsi qu'à Eric Hugounenq. Je tiens évidemment à remercier tous les doctorants que j'ai côtoyés. Merci à mes co-bureaux de tous horizons : Carine, Claudia, Camilo et Olivier. Merci à Christophe pour nos débrifs séries (et son "T'as vu Dexter ? !" du lundi). Merci à Afaf pour sa fraîcheur. Merci à Vincent pour son dynamisme et sa bonne humeur (et le club qu'il supporte). Merci à Yousri pour sa "Yousrité". Enfin, un énorme merci à Pierre pour tous nos (longs !) débats socio-mathématiques (qui ne servent à rien puisqu'on est d'accord de toutes façons ...).

S'il m'est aujourd'hui donné l'opportunité de faire mes preuves dans l'enseignement, c'est en grande partie grâce aux nombreuses démarches de Nicolas Saby en ma faveur : merci Nicolas ! Merci aussi à ma nouvelle collègue, Marielle Fritz, pour m'avoir si aimablement accueilli dans le monde de l'enseignement. Je n'oublie pas Guillaume Bulteau, tour à tour colleur puis collègue, et je le remercie pour son aide et ses conseils.

Je me dois de remercier le département de mathématiques de l'antenne de Bretagne de l'ENS Cachan, dans lequel j'ai eu le privilège d'étudier. Je remercie en particulier Michel Pierre et Grégory Vial, d'abord pour la qualité de leurs enseignements, ensuite pour avoir si gentiment pris de mes nouvelles au cours de ma thèse. J'en profite ici pour remercier Philippe et Julien, alias Michel et Michel, pour tous les bons moments passés durant cette période bretonne.

Mon contrat doctoral m'a aussi permis de découvrir la section football de l'association sportive corporative de l'UM2. Les moments passés avec cette joyeuse bande ont été autant de sources d'évasions et de franches rigolades. J'y ai aussi appris quelques nouveaux concepts footbalistiques comme le fameux "marquage du joueur de touche". Je salue en particulier mes trois comparses : Mathieu, Matthias et Yago. Merci à eux pour les soirées passées chez Francis en attendant pendant trois heures notre pizza au "korizo" soi-disant prête dans vingt-cinq minutes.

Merci aux amis Montpelliérains, notamment Christelle, David (et le club qu'il supporte), Gaëlle et Phi.

L'heure est venue de remercier ma famille. Merci ainsi à mes parents pour leur confiance sans faille et leur exemplarité. Merci à mon frère (et au club qu'il supporte), merci pour sa disponibilité à toute heure du jour et de la nuit, pour ses conseils plus que précieux, merci pour tout, tout simplement. Merci à toute sa petite famille, Olivia (et le club qu'elle supporte), Ghjulia et Mila. Merci aussi à Christian, Virginie et leurs petits, c'est un plaisir que de passer quelques jours avec vous chaque année ! Merci enfin à Jean-Jacques et Suzy pour leur soutien et leurs attentions.

Je termine bien sûr en remerciant celle qui, depuis plus de onze ans, se coltine, jour après jour, un grand stressé, insomniaque, hypocondriaque et un peu autiste sur les bords, ceci avec une patience, une tendresse et une compréhension sans égal. Il va sans dire que rien de ce que j'ai pu modestement accomplir durant cette thèse ne l'aurait été sans ta présence réconfortante à mes côtés (et le club que tu supportes).

Résumé

Le cadre général de ce travail est celui de l'étude mathématique d'écoulements de fluides environnementaux. Deux aspects de cette étude y sont abordés, à travers deux contextes d'application différents.

Dans une première partie, on s'intéresse à la discrétisation d'opérateurs de diffusion anisotropes hétérogènes par des méthodes de volumes finis sur des maillages généraux. Le contexte sous-jacent est celui des écoulements en milieux poreux, étudiés par exemple en industrie pétrolière ou dans le cadre de l'enfouissement de déchets nucléaires. Une simulation efficace de ce type d'écoulements nécessite en effet une discrétisation adéquate des termes elliptiques anisotropes et hétérogènes. Aussi, dans le but d'obtenir des solutions approchées qui respectent les bornes physiques des modèles, il est crucial de conserver un analogue discret du principe du maximum dont jouissent ces opérateurs elliptiques. Nous exposons ainsi des mécanismes généraux permettant de corriger tout schéma de type volumes finis centré aux mailles afin d'obtenir un principe du maximum discret, ceci en conservant certaines de ses propriétés principales. On s'attache en particulier à l'étude des propriétés de coercivité et de convergence des schémas corrigés. Des simulations numériques sont présentées afin d'illustrer l'efficacité de notre méthode.

La deuxième partie est consacrée à la construction de modèles approchés pour la propagation des vagues en eaux peu profondes. Le point de départ de ce travail est le constat selon lequel la présence de topographies non régulières peut poser problème dans les modèles classiques utilisés en océanographie (équations de Saint-Venant, système de Boussinesq ...). On se propose ainsi de mettre en place de nouveaux modèles spécifiques pour prendre en compte de telles topographies. Se limitant dans un premier temps à des écoulements bidimensionnels sur des fonds de formes polygonales, nous proposons une adaptation de la démarche d'étude classique en recourant à des outils d'analyse complexe. Nous développons dans un second temps une approche formelle, valable pour des écoulements tridimensionnels et pour des topographies de formes plus générales. Cette approche débouche sur une alternative non locale aux équations de Saint-Venant contenant des termes régularisants pour les contributions du fond. Sous des hypothèses supplémentaires sur l'amplitude des vagues, nous construisons également des modèles de type Boussinesq à même de prendre en compte des topographies peu régulières.

Mots clés : Equations elliptiques – Schémas volumes finis – Anisotropie – Principe du maximum – Corrections non linéaires – Equations d'Euler à surface libre – Modèles shallow water – Topographies irrégulières – Opérateur de Dirichlet-Neumann.

Abstract

This work investigates two research questions associated with environmental flows and their mathematical modeling.

The first part is devoted to the development of finite volume methods for anisotropic and heterogeneous diffusion operators on general meshes. Diffusion operators of this kind appear in numerous models of flows in porous media, in particular the ones involved in petroleum engineering or underground waste repository. In these contexts, working with numerical simulations requires proper numerical approximations of these diffusion operators. To ensure that the approximate solutions remain within physical bounds, it is crucial to maintain a discrete analogous of the maximum principle for elliptic operators. Starting from any given cell-centered finite volume scheme, we present a general approach to devise non-linear corrections providing a discrete maximum principle while retaining some main properties of the scheme. In particular, we study the coercivity and convergence properties of the modified schemes. We also provide numerical tests showing the efficiency of our method.

The second part of this work focuses on the derivation of approximate models for shallow water wave propagation. It has been known for quite some time that the presence of strongly varying topographies introduces special problems in the traditional shallow water models commonly used in oceanography (Saint-Venant equations, Boussinesq system ...). Based on this observation, we present new models to describe shallow water flows over non smooth topographies. In the particular case of one-dimensional polygonal bottom profiles, we first propose an adaptation of the usual derivation method using complex analysis tools. We then develop a formal approach to account for more general (possibly two-dimensional) topographies. We derive an alternative to the Saint-Venant equations which only involves smoothing contributions of the bottom. Under additional small amplitude assumptions, Boussinesq-type systems are also derived. All these alternative models account for non smooth topographies.

Keywords: Elliptic equations – Finite volume schemes – Anisotropy – Maximum principle – Nonlinear corrections – Water waves – Shallow water models – non smooth topographies – Dirichlet-Neumann operator.

Table des matières

Avant-propos

Ce manuscrit est divisé en deux parties, correspondant aux deux problématiques différentes abordées dans ce travail. Ces problématiques sont elles-même issues de deux aspects distincts de l'étude mathématique d'écoulements de fluides environnementaux. L'une s'attache ainsi à la mise en place de modèles d'écoulements à surface libre, l'autre se concentre sur l'analyse de schémas numériques pour des équations qui apparaissent dans l'étude des écoulements en milieux poreux. Un point commun à ces deux aspects est celui de l'approximation : approximation d'un phénomène réel par un modèle mathématique en amont de l'étude ; approximation numérique des solutions d'équations en aval de celle-ci. Dans un cas comme dans l'autre, notre travail consiste à examiner le devenir d'un aspect particulier du problème considéré au cours du procédé d'approximation :

- Dans la première partie nous étudions la discrétisation de problèmes de diffusion qui interviennent dans des modèles d'écoulements en milieu poreux. On s'attache ainsi plus précisément à la construction de schémas numériques conservant un analogue discret du principe du maximum.
- Dans la seconde partie, on s'intéresse à la modélisation asymptotique d'écoulements à surface libre dans un cadre océanographique. Il s'agit ici de savoir comment tenir compte de la présence de topographies irrégulières lors de la mise en place d'un modèle d'écoulement en eaux peu profondes.

Les différents chapitres correspondent à des travaux soumis ou publiés :

- Les chapitres 1 et 2 sont issus d'un travail réalisé en collaboration avec Clément Cancès et Christophe Le Potier, *Monotone corrections for generic cell-centered Finite Volume approximations of anisotropic diffusion equations* [27], à paraître dans Numerische Mathematik. Dans un souci de clarté, nous avons choisi de présenter ce travail sous la forme de deux chapitres distincts.
- Le chapitre 3 correspond à l'article *Shallow water waves over polygonal bottoms* [29], soumis pour publication.
- Le chapitre 4 est la reproduction de l'article *Asymptotic shallow water models with non smooth topographies* [28], soumis pour publication.

Introduction générale

Sommaire

Parie I. Principe du maximum discret pour des problèmes de diffusion anisotrope

Ecoulements en milieux poreux

Les écoulements et le transport en milieux poreux constituent un domaine de recherche très actif compte tenu du large champs de leurs applications. D'un point de vue environnemental, on peut par exemple citer les études liées au stockage de déchets radioactifs. Devant l'importance des enjeux qui l'entourent[1], la question de la gestion de ces déchets implique en effet de nombreuses disciplines de recherche. Concernant le stockage géologique profond, les mathématiques appliquées entrent notamment en jeu en vue d'améliorer la compréhension des différents phénomènes physiques liés au stockage. Du côté industriel, l'étude d'écoulements en milieux poreux trouve naturellement des applications en industrie pétrolière, dans la modélisation ainsi que la simulation de bassins et de réservoirs d'hydrocarbures. Notons que l'étendue des applications des études liées aux écoulements en milieux poreux dépasse le cadre des géosciences, on retrouve ainsi ce type d'écoulements en biologie ou encore en médecine (pour les modèles de culture osseuse [114] par exemple).

Le modèle de Darcy. La loi de Darcy permet de décrire l'écoulement d'un fluide monophasique dans un milieux poreux en reliant la vitesse de filtration du fluide, ou vitesse de Darcy, au gradient de pression. Dans le cas stationnaire et en négligeant les effets de la gravité, l'écoulement est ainsi régi par les équations :

$$\begin{cases} \nabla \cdot \mathbf{v} = 0, \\ \mathbf{v} = -D\nabla \bar{u}, \end{cases} \tag{1}$$

où $\mathbf{v}$ désigne la vitesse de Darcy, $\bar{u}$ la pression et D le tenseur de perméabilité. Ce système est assorti de conditions aux limites sur la pression ou le débit normal.

Des modèles complexes. Dans le contexte des applications précédentes, le problème elliptique (1) se trouve au cœur de modèles beaucoup plus complexes. Ainsi, dès lors que l'on s'intéresse au transport miscible en milieu poreux, ce problème doit être couplé à une nouvelle équation pour décrire l'évolution de la concentration de la quantité transportée. Dans le cadre de l'enfouissement de déchets radioactifs, un exemple est donné par l'étude du transfert de radionucléides dans les installations industrielles ou le milieu naturel, suite à un relâchement depuis un site de stockage. Ce transfert est en général modélisé par une équation de transport-réaction-diffusion tenant compte à la fois des mécanismes de migration des éléments radioactifs, par convection ou par diffusion naturelle, et des modifications physico-chimiques des éléments transportés [42]. Sous une forme élémentaire, qui ne tient compte que du phénomène de décroissance radioactive de l'élément en question, cette équation s'écrit

$$\omega \partial_t c - \nabla \cdot (S(\mathbf{v})\nabla c) + \nabla \cdot (c\mathbf{v}) = -\omega \lambda c,$$

où c est la concentration de l'élément transporté, ω désigne la porosité du milieu, S le tenseur de diffusion-dispersion (qui dépend de la vitesse de Darcy) et λ la constante de décroissance radioactive de l'élément considéré. La complexité du modèle est accrue si l'on tient compte

[1] En France, la production annuelle de déchets radioactifs est d'environ deux kilogrammes par habitant. Les hypothèses actuelles sur la production de ces déchets dans les prochaines décennies estiment que leur volume pourrait atteindre près de 1,9 millions de m^3 en 2020 puis 2,7 millions en 2030 (`http://www.dechets-radioactifs.com`)

des interactions des éléments transportés avec la phase immobile, notamment les interactions de nature géochimique. Une autre source de complexité dans l'étude des écoulements en milieux poreux apparaît lorsque l'on considère un fluide multiphasique. A titre d'exemple, on peut citer les écoulements du mélange eau/huile considérés en modélisation de bassins sédimentaires. Dans ce cas, le modèle le plus simple consiste en un couplage entre une loi de Darcy généralisée pour la pression et une loi de conservation non linéaire sur la saturation en eau ([30]).

Discrétisation d'opérateurs de diffusion anisotropes hétérogènes

Au regard des équations précédentes, la discrétisation des opérateurs de diffusion apparaît comme une brique de base essentielle dans la simulation numérique d'écoulements en milieux poreux régis par la loi de Darcy. De telles simulations requièrent ainsi la mise au point de méthodes spécifiques pour résoudre numériquement la problème elliptique modèle

$$\begin{cases} -\nabla \cdot (D\nabla \bar{u}) = f & \text{sur } \Omega, \\ \bar{u} = 0 & \text{sur } \partial\Omega. \end{cases} \tag{2}$$

Dans le contexte des écoulements souterrains, les difficultés inhérentes à la discrétisation d'un tel problème sont essentiellement de deux types :

- Le caractère fortement *hétérogène* et *anisotrope* du tenseur de diffusion D. Les hétérogénéités du milieu géologique se traduisent au niveau du tenseur D par des discontinuités au passage d'un type de roche à un autre, tandis que son caractère anisotrope résulte de la stratification du sous-sol, laquelle influe beaucoup sur les directions d'écoulements.
- La nécessité de considérer des *maillages généraux*. Le maillage est souvent conçu en amont de l'étude numérique, ses mailles étant adaptées aux couches géologiques. Les mailles en questions sont donc de formes très générales, et deux mailles voisines peuvent avoir plusieurs faces en commun.

Face à ces contraintes importantes, les schémas numériques classiques ne sont pas suffisants pour approcher l'équation (2). Plusieurs méthodes ont ainsi été développées dans le but de proposer une discrétisation efficace de problèmes de diffusion hétérogène et anisotrope sur des maillages généraux : méthodes de volumes finis à flux multi-points [1, 59, 7, 5], de volumes finis en dualité discrète [76, 49, 21], de Galerkine discontinues [46], de différences finies mimétiques [22, 23], de volumes finis hybrides [63, 64, 65], de volumes finis mixtes [52, 53].

Discrétisation et principe du maximum

Du point de vue du modèle d'écoulement de Darcy, le respect des bornes physiques est imposé par le principe du maximum. Pour le problème (2) cette propriété assure notamment que, si le second membre f est positif, alors il en va de même pour la solution $\bar{u}$. L'absence d'une telle propriété au niveau discret peut ainsi conduire à l'apparition d'oscillations non physiques sur la solution numérique, comme des concentrations négatives [70]. Outre le fait que l'apparition de telles valeurs n'est jamais souhaitable, celles-ci peuvent avoir des conséquences importantes lorsque l'on aborde la discrétisation de systèmes plus complexes couplant l'évolution de plusieurs quantités. C'est par exemple le cas pour les modèles d'écoulements multiphasiques, comme le mélange eau gaz étudié dans le cadre de la récupération d'hydrocarbures [104], ou pour les modèles couplant chimie et transport en lien avec le stockage de déchets radioactifs [69, 70].

Limites des schémas linéaires. Si le schéma volumes finis standard conçu pour la discrétisation d'opérateurs de diffusion isotropes sur des maillages orthogonaux [61] vérifie une version discrète du principe du maximum, il n'en va malheureusement pas de même des différents schémas linéaires proposés pour la discrétisation d'opérateurs diffusifs anisotropes sur des maillages généraux (voir par exemple les résultats des tests présentés lors des congrès FVCA5 et FVCA6 [75, 66]). Pour de forts ratios d'anisotropie ou pour maillages très déformées, on sait même que la construction d'un schéma linéaire consistant à neuf point n'est pas possible [24, 82].

Des schémas spécifiques respectant le principe du maximum. Plusieurs schémas de discrétisation non linéaires ont ainsi vu le jour au cours des dernières années dans le but d'obtenir un principe du maximum discret. Dans [16], une approximation non linéaire, formellement d'ordre deux en espace, est proposée pour des problèmes de convection-diffusion dans le cas d'une diffusion isotrope. Le cas général d'un opérateur anisotrope a été traité par Le Potier dans [88] où une discrétisation volumes finis est proposée sur des maillages de triangles. Ce schéma a ensuite été généralisé dans [81, 93, 95, 117, 90, 110] avec notamment une extension à des maillages généraux en dimension deux ainsi qu'une extension en dimension trois sur des maillages de tétraèdres. Les différents tests numériques présentés dans ces références confirment leur respect du principe du maximum et tendent à montrer que ces schémas sont généralement d'ordre deux (bien qu'aucune preuve théorique de convergence ne soit proposée). Cependant, pour toutes ces méthodes numériques, l'obtention d'un principe du maximum semble se faire au détriment de la coercivité, autre propriété fondamentale de l'opérateur de diffusion du problème (2), qui n'est alors plus assurée au niveau discret. Le défaut de coercivité pour ce type de méthodes est relevé dans [56]. Dans ce travail, les auteurs développent une nouvelle méthode non linéaire qui satisfait une version discrète du principe du maximum sur des maillages généraux et en toutes dimensions. Pour cette méthode, ils proposent une preuve théorique de convergence valable sous une hypothèse de coercivité discrète. Si cette dernière hypothèse n'est pas garantie d'un point de vue théorique, les différents résultats numériques proposés montrent qu'elle semble vérifiée en pratique.

L'approche non linéaire de Le Potier. Plutôt que de chercher à construire entièrement une nouvelle méthode de discrétisation respectant le principe du maximum, une approche différente consiste à "forcer" cette propriété en modifiant des méthodes existantes. Cette idée a d'abord été proposée dans [25] où un terme de stabilisation non linéaire est introduit afin de garantir un principe du maximum discret sur des maillages de simplexes pour une discrétisation de type éléments finis de l'équation de Poisson. Dans le cas d'un opérateur de diffusion général, ce concept de correction non linéaire a été introduit par Le Potier [91]. Ce dernier propose un terme de correction non linéaire pouvant s'appliquer à tout schéma volumes finis centré et permettant d'obtenir un principe du maximum discret. Outre son caractère général, l'un des principaux avantages de ce terme correctif est qu'il assure le respect du principe du maximum tout en préservant la coercivité du schéma de départ. Notons qu'un travail récent [43] établit un parallèle entre cette approche et les techniques classiques de limitation des flux permettant la construction de schémas volumes finis du second ordre non oscillants pour la discrétisation de lois de conservation scalaire (voir [113] par exemple). Dans ce travail, Després développe cette analogie en adaptant la méthode de Le Potier pour construire des schémas non linéaires d'ordre élevé respectant le principe du maximum pour la discrétisation de l'équation de la chaleur unidimensionnelle.

Chapitres 1 et 2 : sur la construction de corrections monotones

L'approche non linéaire précédente est à l'origine du travail exposé dans la première partie de cette thèse et réalisé en collaboration avec Christophe Le Potier et Clément Cancès. L'une des motivations de ce travail est l'identification de mécanismes généraux permettant de corriger un schéma volumes finis afin d'obtenir un principe du maximum discret. Nous avons ainsi mis au point une approche abstraite expliquant comment construire des corrections non linéaires qui garantissent un principe du maximum discret tout en préservant certaines des propriétés principales du schéma de départ, en particulier la coercivité et la convergence de la solution numérique vers la solution exacte de (2) lorsque le maillage devient de plus en plus fin. La présentation de cette approche est l'objet de la partie I. Dans un souci de clarté, nous commençons par définir au chapitre 1 le cadre de travail abstrait qui sert de point de départ à la construction de corrections non linéaires. Ce cadre est celui de la discrétisation du problème de diffusion modèle (2) par une méthode de volume finis centrée aux mailles. Après un bref rappel des principes de la discrétisation au sens des volumes finis, nous précisons quelques propriétés essentielles partagées par nombre de schémas. Plus précisément, nous considérons les trois propriétés suivantes :

(i) *Conservativité* au sens des volumes finis, c'est à dire le fait que les équations du schéma s'écrivent sous la forme de bilans discrets sur des flux approchés associés à l'opérateur continu $\bar{u} \mapsto \nabla \cdot (D\nabla \bar{u})$.

(ii) *Coercivité* de l'opérateur de diffusion discret. Il s'agit de l'analogue, au niveau discret, de la propriété de coercivité de la forme bilinéaire associée à la formulation variationnelle du problème (2).

(iii) *Consistance* de l'approximation avec l'opérateur de diffusion $\bar{u} \mapsto -\nabla \cdot (D\nabla \bar{u})$.

Un récapitulatif des schémas centrés vérifiant ces trois propriétés est donné, sous la forme d'un tableau, à la fin du chapitre 1.

En utilisant ce cadre de travail, nous nous posons, dans le chapitre 2, la question de savoir comment corriger un schéma volumes finis centré générique pour assurer un principe du maximum discret. Le point de départ est ainsi la donnée d'un schéma linéaire centré aux mailles. Etant donné un maillage de l'ouvert Ω, c'est à dire une partition de $\overline{\Omega} = \cup_{K\in\mathcal{M}} \overline{K}$ formée à partir d'ouverts polygonaux, ce schéma consiste en la recherche d'une approximation discrète $u = (u_K)_{K\in\mathcal{M}}$ de $\bar{u}$ solution du système

$$\forall K \in \mathcal{M}, \quad -\mathcal{A}_K(u) = \int_K f,$$

dans lequel la discrétisation initiale, notée $\mathcal{A}_K(u)$, approche la quantité exacte $A(\bar{u}) := \nabla\cdot(D\nabla\bar{u})$ sur la maille K. Afin de savoir comment corriger ce schéma initial, il s'agit en premier lieu de délimiter une classe particulière de schémas vérifiant le principe du maximum. Suivant la définition introduite dans [56], nous considérons ainsi une structure discrète de principe du maximum local. Il s'agit des schémas dont les équations peuvent se mettre sous la forme

$$\sum_{L\in\mathcal{M}} \tau_{K,L}(u)(u_K - u_L) = \int_K f, \tag{3}$$

avec des coefficients $\tau_{K,L}$ positifs (définition 2.1). L'idée est alors de corriger les équations du schéma initial d'une façon conforme à la structure précédente. Plus précisément, on choisit

de modifier les équations par l'ajout d'une quantité corrective comme suit :

$$-\mathcal{A}_K(u) + \sum_{L\in V(K)} \beta_{K,L}(u)(u_K - u_L) = \int_K f, \tag{4}$$

où $V(K)$ correspond au stencil du schéma initial et où les coefficients correctifs $\beta_{K,L}$ sont à déterminer. Partant de cette classe très générale de corrections, le cœur du travail consiste alors à affiner petit à petit leur forme pour concilier les deux objectifs suivants :

- l'obtention d'un schéma corrigé s'écrivant sous la forme requise pour le respect du principe du maximum,
- la préservation des propriétés éventuelles du schéma initial parmi les propriétés (i)–(iii).

L'atteinte du premier objectif repose sur le choix d'une famille de coefficients $(\gamma_{K,L})$ dépendant de u et vérifiant les identités

$$\sum_{Z\in V(K)} \gamma_{K,L}(u)|u_K - u_L| = 1.$$

Ceci permet d'écrire les équations corrigées sous la forme de combinaisons non linéaires :

$$\sum_{Z\in V(K)} \{\gamma_{K,L}(u)\mathrm{sgn}(u_Z - u_K)\mathcal{A}_K(u) + \beta_{K,Z}(u)\}(u_K - u_Z) = \int_K f,$$

faisant ainsi apparaître une structure proche de (3). De cette façon nous en déduisons, une fois la famille $(\gamma_{K,L})$ choisie, une condition suffisante sur les coefficients correctifs $(\beta_{K,L})$ pour que le schéma corrigé respecte le principe du maximum (proposition 2.4).

Nous nous penchons ensuite sur le devenir des propriétés du schéma initial. Dans un premier temps, nous montrons que la préservation des propriétés de conservativité et de coercivité est assurée sous des hypothèses simples sur la forme des coefficients $\beta_{K,L}$:

- hypothèse de symétrie pour la conservativité : $\beta_{K,L} = \beta_{L,K}$ (proposition 2.7) ,
- hypothèse de positivité pour la coercivité : $\beta_{K,L} \geq 0$ (proposition 2.9).

Faisant le compromis des conditions ainsi déterminées, nous en déduisons une démarche pratique permettant, à partir d'un schéma volumes finis centré, de construire des corrections qui garantissent un principe du maximum discret tout en préservant conservativité et coercivité (voir les règles 1–4 de la section 2.2.2.4). Nous proposons comme application de cette démarche la construction de deux exemples de corrections non linéaires. La dernière partie de l'étude théorique du procédé de correction non linéaire concerne la convergence des schémas corrigés. Sous l'hypothèse que le schéma choisi initialement est consistant avec l'opérateur de diffusion continu, la question est de savoir sous quelles conditions la solution approchée calculée par le schéma modifié converge vers la solution exacte lorsque la taille du maillage tend vers 0. Il s'agit, autrement dit, de s'assurer que le terme correctif introduit dans l'équation (4) ne détruit pas la consistance du schéma initial. Nous formulons pour cela un critère général qui assure le maintient de cette propriété de convergence (proposition 2.11). Pour chacun des deux exemples de correction proposé, nous explicitons une condition numérique sur la solution du schéma sous laquelle le critère précédent est satisfait (propositions 2.14 et 2.17). Sur les différentes simulations numériques que nous avons effectuées et dont nous présentons les résultats en conclusion de cette partie, nous avons observé que, pour les deux exemples de corrections proposés, ces conditions numériques semblent effectivement respectées.

Partie II. Prise en compte de topographies irrégulières dans des modèles shallow water

Modélisation d'écoulements en eaux peu profondes

Comprendre et modéliser la circulation de masses d'eau en milieu littoral est aujourd'hui un défi d'importance, aussi bien du point de vue de l'ingénierie côtière que de celui, plus général, de la gestion intégrée des zones côtières. Au sein des processus hydrodynamiques qui régissent les transformations non linéaires de la houle à l'approche de la côte, les variations topographiques du fond jouent un rôle non négligeable. Aussi, dans le cadre des écoulements en eaux "peu profondes", c'est à dire lorsque la profondeur de l'eau est petite devant la longueur d'onde des vagues, la présence de variations brusques de la topographie peut avoir des conséquences importantes sur la transformation des vagues. Il n'est pas rare d'avoir affaire à de telles variations lorsque l'on s'intéresse à la propagation de vagues de type tsunami (à titre d'exemple la figure 1 représente la topographie de la zone de Sumatra, touchée par le tsunami du 26 décembre 2004 dans l'océan Indien). Dans un autre registre,

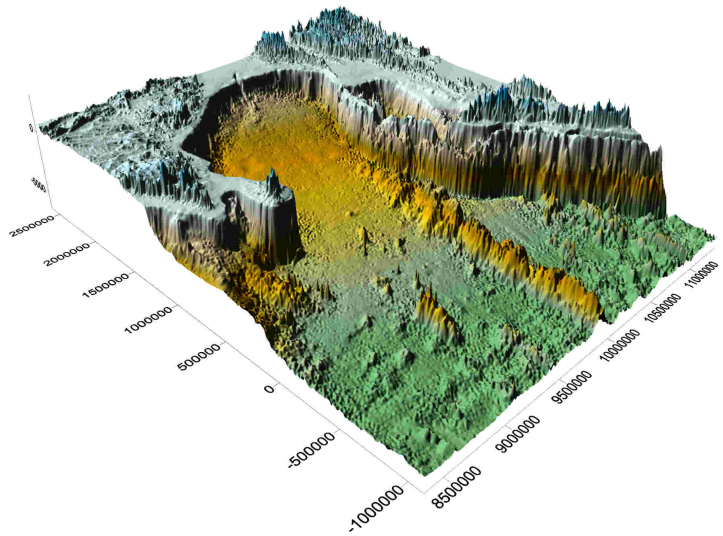

FIGURE 1 – Topographie de la baie de Sumatra

la présence d'aménagements artificiels peut également constituer une source de variation irrégulière de la topographie [68], à même d'influencer la transformation des vagues.

Le problème d'évolution pour les vagues

L'objet de l'étude de la propagation des vagues à la surface de l'océan est de déterminer l'évolution de la surface libre de l'eau, sur une zone éloignée de la zone de formation de ces vagues, de sorte que l'eau y est soumise à la seule force de gravitation. Cette zone d'étude exclut par ailleurs la zone de déferlement ce qui permet, en particulier, de supposer que l'écoulement y est irrotationnel.

Dans toute la suite, d désignera la dimension horizontale de l'écoulement ($d = 1$ ou 2). On notera x la variable horizontale ($x \in \mathbb{R}^d$) et z la coordonnée verticale ($z \in \mathbb{R}$). On supposera

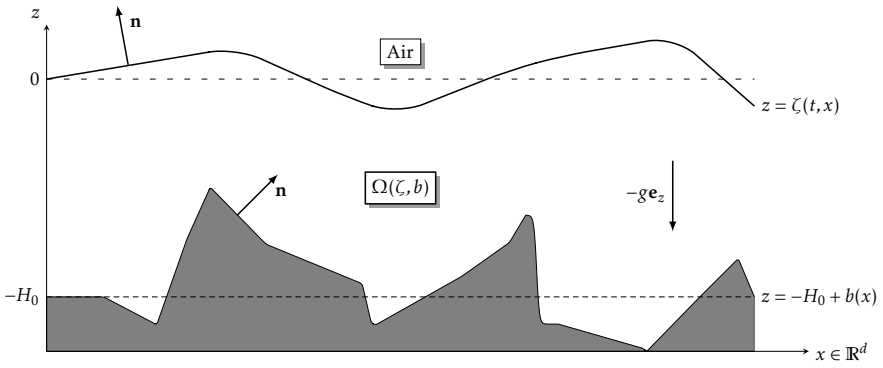

FIGURE 2 – Le domaine fluide.

que la surface est un graphe paramétré par $z = \zeta(t,x)$, le niveau $z = 0$ correspondant au niveau de l'eau au repos. Notant H_0 la profondeur moyenne de l'eau, on supposera de même que la topographie est paramétrée par $z = -H_0 + b(x)$, la quantité $b(x)$ désignant ainsi les variations du fond par rapport à la profondeur moyenne (voir la figure 2). Le "domaine fluide", c'est à dire le domaine occupé par l'eau et noté $\Omega(\zeta,b)$, est ainsi donné à l'instant t par

$$\Omega(\zeta,b) = \Big\{(x,z) \in \mathbb{R}^d \times \mathbb{R} \ ; \ -H_0 + b(x) < z < \zeta(t,x)\Big\}.$$

Il sera supposé que ce domaine fluide ne comporte pas de zone "sèche", c'est à dire qu'il existe une constante $H_{\min}$ strictement positive telle que

$$H_0 + \zeta - b \geq H_{\min}.$$

Les équations de Bernoulli. L'hypothèse d'écoulement irrotationnel assure l'existence d'un potentiel Φ tel que le champ de vitesse $\mathbf{U}$ du fluide s'écrive $\mathbf{U} = \nabla_{x,z}\Phi$, où $\nabla_{x,z}$ désigne le gradient pris par rapport à x et à z. En négligeant la viscosité de l'eau, il est possible de décrire l'évolution de ce potentiel des vitesses à l'aide de l'équation de Bernoulli. Précisément, notant P le champ de pression du fluide et ρ sa masse volumique, supposée constante, cette équation s'écrit

$$\partial_t\Phi + \frac{1}{2}\left|\nabla_{x,z}\Phi\right|^2 + gz = -\frac{1}{\rho}P, \tag{5}$$

où g est l'accélération de la pesanteur. En considérant en outre l'eau comme un fluide incompressible, ce potentiel vérifie également l'équation de Laplace à l'intérieur du fluide :

$$\Delta_{x,z}\Phi = 0 \quad \text{sur } \Omega(\zeta,b),$$

où l'on a noté $\Delta_{x,z}$ le Laplacien pris par rapport à (x,z). Aux bords du domaine occupé par l'eau, il est convenu qu'aucune particule de fluide ne traverse la surface ni ne pénètre le fond. Notant $\mathbf{n}$ le vecteur normal à la frontière dans la direction ascendante (voir figure 2), ces conditions cinématiques s'écrivent

$$\partial_t\zeta - \sqrt{1+|\nabla\zeta|^2}\,\partial_{\mathbf{n}}\Phi = 0 \qquad \text{sur } \{z = \zeta(t,x)\}, \tag{6}$$

$$\partial_{\mathbf{n}}\Phi = 0 \qquad \text{sur } \{z = -H_0 + b(x)\}, \tag{7}$$

où l'on a noté ∇ l'opérateur gradient horizontal et où $\partial_{\mathbf{n}} = \mathbf{n} \cdot \nabla_{x,z}$ désigne la dérivée dans la direction du vecteur $\mathbf{n}$. Enfin, les effets de la tension de surface à l'interface eau/air sont négligés et la pression de l'air est supposée constante. Quitte à renormaliser celle-ci, on peut supposer cette constante nulle.

Formulation du problème en fonction de variables localisées à la surface. Comme remarqué par Zakharov [118], la connaissance de la position de la surface, donnée par ζ, et de la trace du potentiel à la surface $\psi = \Phi_{|z=\zeta}$ détermine les valeurs du potentiel des vitesses dans tout le fluide, ceci en voyant le potentiel comme l'*unique* solution du problème de Laplace

$$\begin{cases} \Delta_{x,z}\Phi = 0 & \text{sur } \Omega(\zeta, b), \\ \Phi = \psi & \text{sur } \{z = \zeta\}, \\ \partial_{\mathbf{n}}\Phi = 0 & \text{sur } \{z = -H_0 + b\}. \end{cases} \tag{8}$$

L'écoulement est ainsi caractérisé par l'évolution des seules quantités ζ et ψ localisées à la surface du fluide. Pour décrire cette évolution, l'idée, due à Craig et Sulem [40], est d'introduire un opérateur de Dirichlet-Neumann afin d'exprimer la donnée de Neumann $\partial_{\mathbf{n}}\Phi$ à la surface en fonction de la donnée de Dirichlet ψ sur cette même surface. Cet opérateur, noté $G[\zeta, b]$, est ainsi défini par

$$G[\zeta, b] : \psi \mapsto \sqrt{1 + |\nabla\zeta|^2}\,\partial_{\mathbf{n}}\Phi_{|z=\zeta},$$

où Φ est vue comme l'extension harmonique de ψ c'est à dire l'unique solution du problème (8). L'introduction de cet opérateur permet d'exprimer les dérivées de Φ à la surface uniquement en fonction des variables ζ et ψ. Dès lors, en utilisant l'équation de Bernouilli (5) écrite à la surface et la condition cinématique (6), il est possible de formuler le problème d'évolution pour les vagues en fonction des seules variables (ζ, ψ) comme un système d'équations d'évolutions posées sur $\mathbb{R}^d$:

$$\begin{cases} \partial_t\zeta - G[\zeta, b]\psi = 0, \\ \partial_t\psi + g\zeta + \dfrac{1}{2}\left|\nabla\psi\right|^2 - \dfrac{(G[\zeta, b]\psi + \nabla\zeta \cdot \nabla\psi)^2}{2(1 + |\nabla\zeta|^2)} = 0. \end{cases} \tag{9}$$

Régimes et modèles asymptotiques pour l'évolution des vagues

Malgré les hypothèses simplificatrices faites sur la nature du fluide, les équations des vagues (9) ont une structure particulièrement riche et admettent des solutions aux comportements radicalement différents. Face à cette complexité, une approche intéressante, aussi bien du point de vue de l'analyse mathématique que de celui de la simulation numérique, consiste à rechercher des modèles plus simples permettant de décrire le mouvement des vagues dans des régimes spécifiques d'écoulement.

Le problème adimensionné. L'identification de régimes particuliers pour l'écoulement repose elle-même sur la comparaison de certaines de ses caractéristiques physiques. Aussi, cette identification est-elle facilitée par la mise des équations sous forme non dimensionnelle. Pour ce faire, on utilise certaines grandeurs caractéristiques du système (voir figure 3) :

- La profondeur moyenne de l'eau H_0 ;
- Une longueur d'onde typique λ de l'écoulement ;

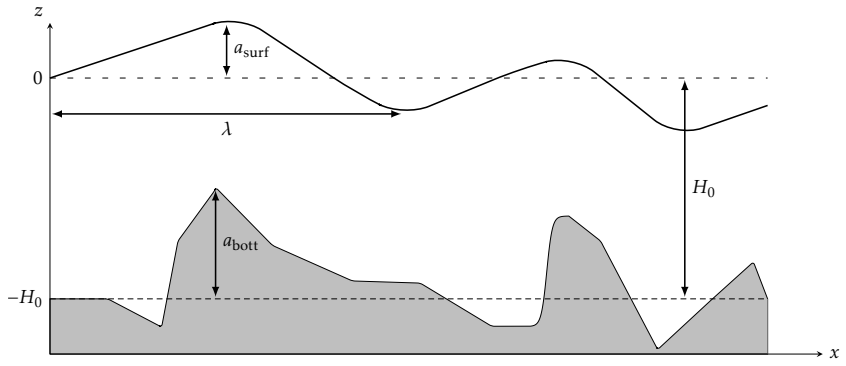

FIGURE 3 – Grandeurs caractéristiques de l'écoulement.

- Un ordre grandeur a_{surf} de l'amplitude des vagues ;
- Un ordre grandeur a_{bott} des déformations du fond.

On définit alors variables et inconnues sans dimension en posant[2] :

$$x' = \frac{x}{\lambda}, \quad z' = \frac{z}{H_0}, \quad t' = \frac{\sqrt{gH_0}}{\lambda} t,$$

et

$$\zeta' = \frac{\zeta}{a_{\text{surf}}}, \quad b' = \frac{b}{a_{\text{bott}}}, \quad \Phi' = \frac{\Phi}{a_{\text{surf}} \lambda \sqrt{g/H_0}},$$

où le choix des quantités utilisées pour adimensionner le temps et le potentiel des vitesses peut être mis en évidence par l'étude du linéarisé de (9) autour de $(\zeta, \psi) = (0, 0)$ et sur fond plat (voir par exemple [85, Section 1.3.2]). Les grandeurs précédentes permettent également de définir trois paramètres sans dimension

$$\mu = \frac{H_0^2}{\lambda^2}, \quad \varepsilon = \frac{a_{\text{surf}}}{H_0}, \quad \beta = \frac{a_{\text{bott}}}{H_0}.$$

Parmi les paramètres ci-dessus, μ mesure le caractère d'écoulement en eaux peu profondes ou "shallowness" alors que ε et β mesurent respectivement les amplitudes des vagues et des variations topographiques. L'introduction de ces différentes quantités sans dimensions permet d'écrire les équations des vagues sous une forme non dimensionnelle. En omettant les primes pour plus de lisibilité, ces équations sont maintenant posées sur le domaine

$$\Omega(\varepsilon\zeta, \beta b) = \Big\{ (x, z) \in \mathbb{R}^d \times \mathbb{R} \ ; \ -1 + \beta b(x) < z < \varepsilon\zeta(t, x) \Big\}.$$

Sur ce dernier, le potentiel adimensionné Φ est solution d'une équation de Laplace "tordue" :

$$\begin{cases} \mu\Delta\Phi + \partial_\Phi z^2 = 0 & \text{sur } \Omega(\varepsilon\zeta, \beta b), \\ \Phi = \psi & \text{sur } \{z = \varepsilon\zeta\}, \\ \partial_{\mathbf{n}}\Phi = 0 & \text{sur } \{z = -1 + \beta b\}, \end{cases} \tag{10}$$

[2] L'adimensionnement de la variable horizontale proposé ici suppose implicitement que la longueur d'onde typique de l'écoulement est la même dans toutes les directions horizontales. De même, l'adimensionnement du fond suppose que la longueur d'onde des variations topographiques du même ordre que la longueur d'onde des vagues.

où Δ désigne le Laplacien relatif aux seules variables horizontales. Voyant Φ comme la solution de l'équation précédente, on peut également définir un opérateur de Dirichlet-Neumann adimensionné

$$\mathcal{G}_\mu[\varepsilon\zeta, \beta b] : \psi \mapsto \sqrt{1 + \varepsilon^2 |\nabla\zeta|^2}\, \partial_{\mathbf{n}} \Phi_{|z=\varepsilon\zeta}. \tag{11}$$

Enfin, la formulation de Zakharov/Craig-Sulem du problème d'évolution pour les vagues s'écrit, sous forme non dimensionnelle,

$$\begin{cases} \partial_t \zeta - \dfrac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = 0, \\ \partial_t \psi + \zeta + \dfrac{\varepsilon}{2}\left|\nabla\psi\right|^2 - \varepsilon\mu \dfrac{\left(\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi + \varepsilon\nabla\zeta\cdot\nabla\psi\right)^2}{2(1+\varepsilon^2\mu|\nabla\zeta|^2)} = 0, \end{cases} \tag{12}$$

Modèles asymptotiques en régime shallow water. Comme nous l'avons déjà mentionné, le régime shallow water correspond à un écoulement pour lequel la profondeur moyenne H_0 est petite devant la longueur d'onde λ des vagues. Avec les paramètres définis ci-dessus, cela revient à supposer $\mu \ll 1$. Ce régime asymptotique s'applique non seulement à l'étude des écoulements côtiers mais aussi, par exemple, à la propagation d'ondes de type tsunami[3]. Aussi, la recherche de modèles simplifiés, permettant de décrire la dynamique des vagues dans ce régime particulier, est une entreprise déjà largement entamée. L'un des modèles les plus simples consiste à approcher les équations des vagues par le système de Saint-Venant (voir ci-après). Si la première apparition de ce modèle date du XIX$^{\text{e}}$ siècle, la justification rigoureuse de sa pertinence en tant qu'approximation des équations des vagues est beaucoup plus récente. Les premières analyses de ce type remontent ainsi aux travaux d'Ovsjannikov [106, 107] et de Kano et Nishida [80] dans le cas d'une surface unidimensionnelle, sur fond plat et sous certaines restrictions sur les conditions initiales. La justification complète de ce modèle est maintenant établie, avec notamment la contribution de Li [92], toujours pour un domaine fluide bidimensionnel ($d = 1$) à fond plat, puis les travaux de Iguchi [79] et de Alvarez-Samaniego et Lannes [10] dans le cas général ($d = 1$ ou 2) avec un fond non plat (mais régulier). Partant de la formulation de Zakharov/Craig-Sulem (12), la démarche générale adoptée par ses deux derniers travaux consiste notamment à retrouver les équations de Saint-Venant à partir d'une analyse asymptotique de l'opérateur de Dirichlet-Neumann lorsque $\mu \ll 1$. Il s'agit plus précisément de construire un développement asymptotique de $\mathcal{G}_\mu[\varepsilon\zeta\beta b]\psi$ par rapport à μ puis de remplacer dans (12) l'opérateur de Dirichlet-Neumann par l'approximation ainsi obtenue. Lorsque les variations du fond sont régulières, on est à même d'établir (voir par exemple la proposition 3.8 de [10]) que

$$\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = -\nabla\cdot\Big((1 + \varepsilon\zeta - \beta b)\nabla\psi\Big) + O(\mu). \tag{13}$$

Partant de cette approximation, on montre alors qu'à des termes d'ordre $O(\mu)$ près, le couple $(\zeta, \mathbf{v})$, où $\mathbf{v} = \nabla\psi$, est solution des équations de Saint-Venant :

$$\begin{cases} \partial_t \zeta + \nabla\cdot((1 + \varepsilon\zeta - \beta b)\mathbf{v}) = 0, \\ \partial_t \mathbf{v} + \nabla\zeta + \varepsilon(\mathbf{v}\cdot\nabla)\mathbf{v} = 0. \end{cases} \tag{14}$$

[3] Sur le plateau côtier, on a affaire à des profondeurs de l'ordre de la dizaine de mètres, pour une houle d'une longueur d'onde d'environ cent mètres, ce qui correspond à $\mu \approx 10^{-2}$. Concernant le tsunami de 2004 dans l'océan Indien, d'une profondeur de l'ordre du kilomètre, les longueurs d'onde observées étaient d'une centaine de mètres environ, soit $\mu \approx 10^{-4}$.

Des problèmes de fond (cas des topographies irrégulières)

La prise en compte de topographies peu régulières dans la construction de modèles shallow water est un obstacle aussi bien du point de vue de l'analyse asymptotique de l'opérateur de Dirichlet-Neumann que de celui de l'écriture même d'un modèle.

Limites du modèle de Saint-Venant. On sait que la présence de topographies très irrégulières ne pose pas de problème concernant l'écriture des équations des vagues. Un examen heuristique de la contribution de la topographie dans le problème des vagues permet en effet de constater que la paramétrisation du fond b entre uniquement en jeu dans le problème elliptique (10) à travers la condition de Neumann. Aussi, en dehors du fond du domaine fluide, la régularité de la solution Φ ne dépend pas de la régularité de cette frontière. En tant que quantité définie à partir des valeurs de Φ à la surface (et donc loin du fond), la régularité de $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ est donc indépendante de celle de b. De façon rigoureuse, Alazard *et al.* [9] ont prouvé que le problème des vagues (12) est en particulier bien posé sans hypothèse de régularité sur la topographie (la seule hypothèse étant que le fond et la surface doivent être séparés par une bande de largeur fixe). En revanche, du point de vue du modèle de Saint-Venant, la présence de topographies irrégulières peut faire apparaître des termes singuliers dans la première équation de (14), laissant ainsi penser que ce modèle n'est, dans ce cas, plus une approximation satisfaisante des équations des vagues en régime shallow water. Les limites des modèles shallow water usuels pour la prise en compte de topographies irrégulières sont en fait connues depuis longtemps. Dans [74], Hamilton relève déjà les problèmes posés par la présence de fonds comportant de fortes variations dans les modèles faiblement non linéaires (*i.e.* sous l'hypothèse $\varepsilon \sim \mu \ll 1$) de Mei et Le Méhauté [96] et de Peregrine [108]. Il explique plus exactement que la démarche formelle utilisée par ces derniers[4] n'est plus valable lorsque les variations topographiques sont trop grandes. Se restreignant au cas d'écoulements bidimensionnels ($d = 1$), il propose une adaptation de cette démarche formelle, ceci en exploitant le fait que le domaine fluide est alors conformément équivalent à une bande horizontale (selon une idée due à Kreisel [83]). Cette technique conduit à un nouveau modèle faiblement non-linéaire, mais dont l'écriture nécessite la connaissance de l'application conforme transformant le domaine occupé par le fluide (au repos) en une bande. Dans le cas où le fond est de forme polygonale, la théorie de Schwarz-Christoffel fournit un moyen pratique d'accéder à cette application conforme (voir par exemple [99]). Le théorème de Schwarz-Christoffel réduit en effet la recherche d'une telle application à la résolution d'un système non linéaire, permettant ainsi d'aborder efficacement cette recherche du point de vue de l'approximation numérique [50]. Partant de ce constat et s'inspirant de la démarche d'Hamilton, Nachbin [97] utilise la théorie de Schwarz-Christoffel pour construire un modèle de type Boussinesq à même de prendre en compte des profils topographiques polygonaux très généraux.

Limites de l'étude asymptotique de l'opérateur de Dirichlet-Neumann. Les difficultés relevées par Hamilton concernant la prise en compte de topographies irrégulières dans des modèles shallow water se retrouvent également lorsque l'on aborde la construction d'un tel modèle par la technique plus récente basée sur l'analyse asymptotique de l'opérateur de Dirichlet-Neumann. Cet opérateur étant explicitement défini en fonction du potentiel des vitesses Φ, une façon naturelle d'effectuer une telle analyse consiste précisément à étudier le comportement asymptotique de ce même potentiel lorsque $\mu \ll 1$. La principale difficulté

[4] Démarche qui repose sur un développement analytique du potentiel des vitesses en la coordonnée verticale à partir du niveau du fond $z = -1 + \beta b(x)$.

réside alors dans le fait que l'équation de Laplace (10), qui donne le potentiel, est posée sur un domaine qui est lui-même une inconnue du problème des vagues. Une approche efficace pour contourner cette difficulté est de redresser le domaine fluide $\Omega(\varepsilon\zeta, \beta b)$ en une bande horizontale, transformant du même coup l'équation de Laplace posée sur ce domaine en mouvement en une équation elliptique à coefficients variables, mais posée sur un domaine fixe (voir par exemple [20, 10, 79] ou encore [101]). Sachant qu'il est possible d'exprimer l'opérateur de Dirichlet-Neumann en fonction de la solution de cette équation transformée, on ramène ainsi l'étude asymptotique de l'opérateur à la construction d'une solution approchée d'une équation elliptique posée sur une bande horizontale. Comme peut le laisser transparaître le bref résumé précédant, le choix du difféomorphisme permettant de redresser le domaine fluide a beaucoup d'influence dans cette approche. Celui-ci détermine en effet les coefficients du problème elliptique transformé et conditionne, par suite, l'analyse asymptotique de sa solution. Plus précisément, étant donné un difféomorphisme Σ entre la bande $\mathcal{S} = \mathbb{R}^d \times (-1, 0)$ et le domaine $\Omega(\varepsilon\zeta, \beta b)$, on montre (voir la proposition 2.7 de [84]) que le potentiel transformé $\phi = \Phi \circ \Sigma$ vérifie sur $\mathcal{S}$ l'équation elliptique

$$\nabla_{x,z} \cdot P[\Sigma] \nabla_{x,z} \phi = 0,$$

où $P[\Sigma]$ est la matrice symétrique définie positive donnée, en fonction de la matrice Jacobienne J_Σ de Σ, par

$$P[\Sigma] = |\det J_\Sigma| J_\Sigma^{-1} \begin{bmatrix} \mu & 0 \\ 0 & 1 \end{bmatrix} (J_\Sigma^{-1})^T.$$

Un exemple simple est fourni par le difféomorphisme suivant :

$$\forall (x, z) \in \mathcal{S}, \quad \Sigma(x, z) = \Big(x, \varepsilon\zeta(x) + z(1 + \varepsilon\zeta(x) - \beta b(x)) \Big), \tag{15}$$

transformation qui laisse invariantes les coordonnées horizontales. Un tel choix requiert néanmoins une certaine régularité sur la paramétrisation b puisque les coefficients de $P[\Sigma]$ s'écrivent notamment en fonction de ∇b. Par conséquent, comme remarqué par Lannes [85, Section 2.5.3], il n'est plus possible d'utiliser une telle transformation dans le cas où la paramétrisation de la topographie n'est pas régulière[5].

Chapitre 3 : le cas d'une topographie polygonale

Compte tenu des considérations précédentes et au vu du travail de Nachbin, les applications de Schwarz-Christoffel apparaissent comme des candidats naturels pour redresser un domaine fluide bidimensionnel à fond polygonal. Utilisant ces outils d'analyse complexe, nous nous proposons, au chapitre 3, d'adapter la démarche classique d'analyse asymptotique de l'opérateur de Dirichlet-Neumann au cas d'une topographie polygonale.

Redresser un domaine à fond polygonal. Suivant la démarche présentée ci-dessus, il s'agit dans un premier temps de construire un difféomorphisme entre un domaine fixe et le domaine à fond polygonal occupé par le fluide. En identifiant le domaine bidimensionnel à une partie du plan complexe, on utilise la théorie de Schwarz-Christoffel pour construire cette application et transformer ainsi le problème de Laplace (10) en une équation elliptique définie sur le domaine redressé $\mathcal{S} = \mathbb{R} \times (-1, 0)$. Pour ce faire, on procède en deux temps :

[5] Si l'on suit par exemple l'établissement de l'approximation $\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = -\nabla \cdot ((1 + \varepsilon\zeta - \beta b)\nabla\psi) + O(\mu)$ dans [85], on constate que le contrôle en norme H^s du reste dans cette approximation, établi par la proposition 3.37, vaut pour une paramétrisation du fond de régularité supérieure à H^{s+3}.

1. On commence par redresser le fond polygonal du domaine occupé par le fluide, au moyen d'un difféomorphisme noté $\Sigma_{\text{bott}}^{-1}$. Ce faisant, on transporte l'équation de Laplace définie sur le domaine à fond polygonal en *la même* équation posée sur un domaine à fond plat.

2. Partant de ce nouveau problème de Laplace avec fond plat, la deuxième étape consiste alors à redresser la surface libre, via un difféomorphisme noté $\Sigma_{\text{surf}}^{-1}$, transformant ainsi l'équation de Laplace en une équation elliptique sur $\mathcal{S}$ à coefficients variables.

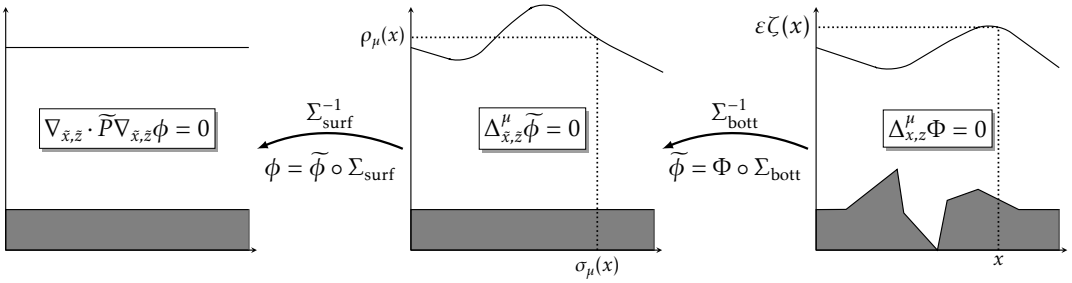

FIGURE 4 – Transformation en deux étapes du domaine fluide en une bande horizontale, et transformations associées de l'équation de Laplace.

La première étape repose entièrement sur l'utilisation de la théorie de Schwarz-Christoffel. L'intérêt d'un tel recours est qu'il permet de redresser le fond à l'aide d'une application régulière tout en conservant l' harmonicité du potentiel des vitesses[6] : le potentiel transporté $\widetilde{\phi} = \Phi \circ \Sigma_{\text{bott}}$ vérifie donc la même équation que le potentiel Φ. De cette façon, on est ramené, à l'issu de cette première étape, à une situation d'écoulement à surface libre dans un cadre d'étude classique. En effet, le potentiel transformé $\widetilde{\phi}$ est maintenant solution d'une équation de Laplace définie sur un domaine fluide dont le fond est plat. On peut alors utiliser un difféomorphisme trivial (défini comme en (15)) pour redresser ce nouveau domaine et transformer par là même l'équation de Laplace en une équation elliptique posée sur la bande horizontale $\mathcal{S}$.

Une nouvelle approximation de l'opérateur de Dirichlet-Neumann. Le transport du problème de Laplace posé sur le domaine à fond polygonal en un problème elliptique posé sur $\mathcal{S}$ peut ensuite être mis à profit pour construire un développement asymptotique de l'opérateur de Dirichlet-Neumann. Pour ce faire, on suit pas à pas la démarche évoquée plus haut, utilisée dans le cas où la topographie est régulière (voir par exemple [10] ou [20, 87, 31]). Celle-ci peut être décomposée en trois étapes :

1. On commence par exprimer l'opérateur de Dirichlet-Neumann en fonction du potentiel transporté ϕ, solution du problème elliptique finalement obtenu sur $\mathcal{S}$.

2. On construit ensuite une solution approchée de ce problème elliptique posé sur $\mathcal{S}$ sous la forme d'un développement BKW.

[6] Pour être précis, notons qu'en raison de l'adimensionnement, le potentiel des vitesses Φ n'est pas à proprement parler harmonique puisqu'il vérifie une équation de Laplace "tordue par μ" : $\mu\partial_x^2\Phi + \partial_z^2\Phi = 0$. Afin de préserver l'équation précédente lors de la transformation du domaine, on n'utilise donc pas directement des applications conformes mais des applications dites "μ-conformes" (voir la section 3.2.1). Ces dernières s'obtiennent à partir d'applications conformes en composant par un changement de variable simple [97].

3. L'approximation de l'opérateur de Dirichlet-Neumann résulte alors du remplacement, dans l'expression obtenue à l'étape 1, du potentiel transporté ϕ par la solution approchée calculée à l'étape 2.

En notant $(\sigma_\mu(x), \rho_\mu(x)) = \Sigma_{\text{bott}}^{-1}(x, \varepsilon\zeta(x))$ la paramétrisation de la frontière libre dans le domaine à fond plat (voir la figure 4), on obtient finalement

$$\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = -\partial_x\left(\frac{1+\rho_\mu}{\sigma_\mu'}\partial_x\psi\right) + O(\mu). \tag{16}$$

Nous établissons par ailleurs une estimation du reste dans ce développement asymptotique, ceci en utilisant successivement une estimation elliptique à l'étape 2 et un théorème de trace à l'étape 3 (voir la proposition 3.9 pour un énoncé détaillé).

Un nouveau modèle de type shallow water pour des topographies polygonales. En remplaçant l'opérateur de Dirichlet-Neumann par cette nouvelle approximation dans les équations des vagues, on en déduit que le couple (ζ, v), où $v = \partial_x\psi$, vérifie à des termes d'ordre $O(\mu)$ près le système

$$\begin{cases} \partial_t\zeta + \partial_x(M_\mu v) = 0, \\ \partial_t v + \partial_x\zeta + \varepsilon v_s \partial_x v = 0, \end{cases} \tag{17}$$

dans lequel la quantité M_μ est définie en fonction de la surface transformée (σ_μ, ρ_μ) par

$$M_\mu = \frac{1+\rho_\mu}{\sigma_\mu'}.$$

Le principal défaut du nouveau modèle shallow water ainsi obtenu est que l'on ne dispose pas d'expression explicite de la quantité M_μ en fonction des paramétrisations ζ et b. Ce défaut tient au fait que les fonctions σ_μ et ρ_μ sont définies via le difféomorphisme $\Sigma_{\text{bott}}^{-1}$ de redressement du fond. En effet, comme on l'a dit, la construction de ce dernier repose sur la résolution d'un système non linéaire à travers l'application du théorème de Schwarz-Christoffel. Si ce système n'admet pas en général de solution analytique, il existe toutefois des méthodes numériques efficaces pour résoudre celui-ci et par suite évaluer un tel difféomorphisme (voir par exemple le chapitre 3 de [50]). En annexe du chapitre 4, nous expliquons ainsi comment utiliser ce type de méthode pour construire un schéma de discrétisation pour (17).

Notons enfin que le modèle shallow water (17) s'inscrit naturellement dans la lignée des travaux de Nachbin (voir notamment [97] ainsi que la contribution récente [68]), venant ainsi compléter le modèle de type Boussinesq obtenu par ce dernier dans le cas faiblement non linéaire (c'est à dire sous l'hypothèse supplémentaire $\varepsilon \sim \mu$).

Chapitre 4 : une approche formelle pour des topographies plus générales

Devant les limites du modèle obtenu en suivant l'approche classique envisagée au chapitre 3, nous développons au chapitre 4 une approche formelle basée sur l'analyticité de l'opérateur de Dirichlet-Neumann par rapport au couple (ζ, b) définissant la frontière du domaine fluide.

Analyticité de l'opérateur de Dirichlet-Neumann par rapport aux frontières du domaine. Sur fond plat, la dépendance analytique de l'opérateur de Dirichlet-Neumann par rapport à la paramétrisation de la surface est connue depuis les travaux de Calderón [26] et de Coifman et Meyer [35] (voir aussi les travaux plus récents de Craig *et al.* [39], Craig et Sulem [38]

et de Hu et Nicholls [78]). Dans le cas d'un fond non plat, l'analyticité de l'opérateur de Dirichlet-Neumann par rapport au couple (ζ, b) a été plus récemment étudiée, notamment par Nicholls et Taber [103] et Lannes [85, Théorème A.11]. Il est en outre possible de calculer explicitement les expressions des dérivées de formes de l'opérateur de Dirichlet-Neumann à tout ordre autour du point $(\zeta, b) = (0, 0)$. Une formule de récurrence a été établie à cet effet par Craig *et al.* [36] (voir également [40]). Aussi, Guyenne et Nicholls [73] ont remarqué que ces relations de récurrence font intervenir un opérateur régularisant pour les contributions du fond, laissant ainsi entrevoir la possibilité de considérer des paramétrisations générales de la topographie dans les termes du développement analytique.

Développement analytique de l'opérateur de Dirichlet-Neumann en régime shallow water. Motivée par la remarque précédente, l'idée à la base des travaux présentés au chapitre 4 est de construire formellement une approximation de l'opérateur de Dirichlet-Neumann en régime shallow water en deux étapes successives :

1. Le premier pas de cette construction consiste à remplacer l'opérateur de Dirichlet-Neumann $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]$ par sa série de Taylor par rapport à $\varepsilon\zeta$ et βb.

2. Le deuxième pas est une étude asymptotique des termes individuels de ce développement en régime shallow water, c'est à dire lorsque $\mu \ll 1$.

L'approximation de l'opérateur de Dirichlet-Neumann par un développement de Taylor a déjà été largement utilisée en régime ondes longues ($\varepsilon \ll 1$) et pour de faibles variations topographiques (voir par exemple [40, 72, 73]). Néanmoins, l'idée d'approcher l'opérateur de Dirichlet-Neumann par un tel développement en régime shallow water est beaucoup moins naturelle. En effet, s'il est légitime d'approcher la fonction $\mathcal{G}_\mu[\cdot,\cdot]\psi$ au voisinage de $(\varepsilon\zeta,\beta b) = (0,0)$ à l'aide d'un développement de Taylor, la série de Taylor en question n'a par contre aucune raison *a priori* de converger en dehors d'un voisinage de (0,0) (c'est à dire sans hypothèse de petitesse sur ε et β). Ceci étant, en écrivant les versions adimensionnalisées des relations de récurrence donnant les termes de cette série, nous constatons que la énième dérivée de forme de $\mathcal{G}_\mu[\cdot,\cdot]\psi$ en $(0,0)$ est formellement d'ordre au moins $O(\sqrt{\mu}^{n+1})$. Ainsi, en régime shallow water, la série de Taylor de l'opérateur de Dirichlet-Neumann converge (au moins formellement[7]) sans hypothèse de petitesse sur ε ou β, ce qui permet d'écrire $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ comme somme de sa série de Taylor à l'origine. L'assurance de ce que l'ordre en μ des termes du développement analytique de l'opérateur de Dirichlet-Neumann s'élève conjointement avec l'ordre des dérivées de formes facilite dans un second temps la construction d'un opérateur approché lorsque $\mu \ll 1$. En vue d'obtenir un modèle de type Saint-Venant, nous cherchons en effet à approcher l'opérateur de Dirichlet-Neumann à des termes d'ordre $O(\mu^2)$ près[8]. L'observation précédente permet ainsi de ne considérer que les premiers termes du développement de Taylor de $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ pour déterminer cette approximation (précisément, nous montrons qu'il suffit d'arrêter le développement à l'ordre 1 en $\varepsilon\zeta$ et à l'ordre 2 en βb). La construction effective d'un opérateur de Dirichlet Neumann approché découle alors de l'analyse asymptotique de chacun des termes conservés en régime shallow water.

[7] Au sens où nous ne sommes pas capables d'expliciter une norme dans laquelle les dérivées de formes sont effectivement des $O(\sqrt{\mu}^{n+1})$ et par suite une topologie pour laquelle la série de Taylor converge.

[8] Car il s'agit d'approcher $\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ dans (12) à des termes d'ordre $O(\mu)$ près.

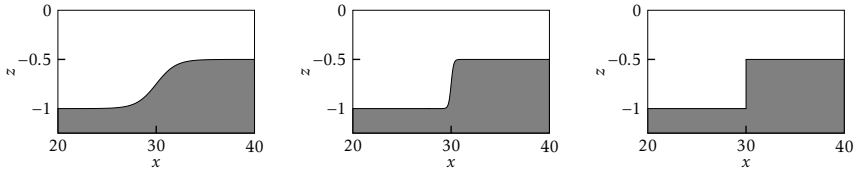

FIGURE 5 – Topographies considérées pour les tests numériques

Des alternatives aux modèles classiques pour prendre compte de topographies générales. L'approximation fournie par la méthode précédente consiste à remplacer la fonction βb par un opérateur régularisant $b_\mu[\beta b]$ dans le développement asymptotique classique de l'opérateur de Dirichlet-Neumann :

$$\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = -\nabla\cdot\Big((1+\varepsilon\zeta - b_\mu[\beta b])\nabla\psi\Big) + O(\mu), \tag{18}$$

(la définition précise de l'opérateur $b_\mu[\beta b]$ en fonction de βb est donnée à la section 4.3.1, équations (4.25)-(4.26)). Cette nouvelle approximation débouche à son tour sur un système non local de type Saint-Venant :

$$\begin{cases} \partial_t\zeta + \nabla\cdot((1+\varepsilon\zeta - b_\mu[\beta b])\mathbf{v}) = 0, \\ \partial_t\mathbf{v} + \nabla\zeta + \varepsilon(\mathbf{v}\cdot\nabla)\mathbf{v} = 0. \end{cases} \tag{19}$$

La présence du terme régularisant dans ce dernier permet de prendre en compte des topographies plus générales qu'il n'est permis avec le modèle de Saint-Venant classique[9].

L'approche précédente permet par ailleurs de construire des modèles asymptotiques précis à des termes d'ordre $O(\mu^2)$ près. Il suffit pour cela de considérer des termes supplémentaires dans le développement de Taylor de l'opérateur de Dirichlet-Neumann. Sous des hypothèses supplémentaires sur l'amplitude des vagues (moyenne amplitude $\varepsilon \sim \sqrt{\mu}$, ou petite amplitude $\varepsilon \sim \mu$), nous obtenons de cette manière des alternatives non locales à certains modèles classiques, à savoir les équations de Serre (dans le cas $\varepsilon \sim \sqrt{\mu}$) et le système de Boussinesq (dans le cas $\varepsilon \sim \mu$). Tous ces modèles alternatifs sont à même de prendre en compte des topographies plus générales.

Nous présentons enfin quelques simulations numériques afin d'évaluer le comportement de ces nouveaux modèles. Lorsque la paramétrisation du fond est régulière, les résultats obtenus semblent montrer que les différents modèles alternatifs sont consistants avec leurs homologues classiques. Nous simulons ensuite numériquement le passage d'une vague au dessus d'un fond en forme de marche, en considérant une pente de plus en plus abrupte (voir la figure 5). Pour chaque fond considéré, nous évaluons l'erreur de consistance entre les modèles classiques et les nouveaux modèles non locaux. Il s'agit ainsi d'estimer approximativement le prix à payer, en termes de précision du modèle, lorsque l'on utilise un système shallow water classique en présence d'une topographie comportant une forte variation. En

[9] Précisément on montre que le teme $b_\mu[\beta b]\mathbf{v}$ est une fonction infiniment dérivable dès que b est lipschitzienne. Pour des paramétrisations $b \in L^\infty(\mathbb{R}^d)$ plus générales, certaines difficultés apparaissent quant à la définition de certains termes figurant dans l'expression de l'opérateur $b_\mu[\beta b]$ (voir la remarque 4.1). Dans certains cas particuliers (voir par exemple la remarque 4.8 dans le cas d'une topographie en forme de marche d'escalier), nous sommes capables de donner un sens à cet opérateur pour des topographies discontinues. Par ailleurs, la plupart des simulations numériques ont été effectuées avec de telles topographies, pour lesquelles le modèle (19) donne des résultats prometteurs.

annexe, nous considérons le cas d'une topographie polygonale et nous comparons numériquement le modèle de type Saint-Venant construit par la présente approche au modèle obtenu au chapitre 3. Les simulations sont réalisées avec un fond présentant une "bosse rectangulaire". Dans ce cas, les résultats obtenus laissent supposer que ces deux approches différentes conduisent à des modèles consistants entre eux.

PARTIE

I

Principe du maximum discret pour des problèmes de diffusion anisotrope

CHAPITRE

1 Méthodes de volumes finis pour la discrétisation d'équations elliptiques

Ce chapitre établit le cadre de travail abstrait qui servira de point de départ au chapitre suivant consacré à la construction de corrections qui garantissent un principe du maximum pour des dicrétisations de type volumes finis d'opérateurs de diffusion. Après un bref état des lieux des méthodes existantes pour la discrétisation de problèmes diffusifs, nous nous concentrons sur les méthodes volumes finis centrées aux mailles. Nous précisons quelques propriétés importantes partagées par nombre de ces schémas (conservativité, coercivité et consistance) et dont il nous importera de tenir compte lorsque nous aborderons la construction de corrections non linéaires.

Sommaire

1.1 Enjeux de la discrétisation

Soit Ω un ouvert borné polygonal de $\mathbb{R}^d$. On considère le problème elliptique modèle suivant, d'inconnue $\bar{u}$,

$$\begin{cases} -\nabla\cdot(D\nabla\bar{u}) = f & \text{sur } \Omega, \\ \bar{u} = 0 & \text{sur } \partial\Omega\,; \end{cases} \tag{1.1}$$

où :

- $f \in L^2(\Omega)$;

- $D : \Omega \to \mathcal{M}_d(\mathbb{R})$ est une fonction mesurable bornée telle que $D(x)$ est symétrique pour presque tout $x \in \Omega$ et telle qu'il existe un réel $\lambda > 0$ pour lequel $D(x)\xi \cdot \xi \geq \lambda |\xi|^2$ pour presque tout $x \in \Omega$ et tout $\xi \in \mathbb{R}^d$.

Comme nous l'avons mentionné dans le chapitre introductif, la discrétisation de l'opérateur de diffusion à la base du problème (1.1) revêt une importance particulière dès lors que l'on aborde l'étude d'écoulements en milieux poreux, par exemple dans le domaine des géosciences. Aussi, dans ce domaine spécifique, la simulation de ce type d'écoulements est soumise à des contraintes importantes inhérentes au contexte de l'étude. Le champs de diffusion D est ainsi susceptible de présenter des hétérogénéités importantes, ces dernières étant liées aux hétérogénéités du milieu géologique lui-même. Par exemple, dans le cadre de la simulation de réservoirs en industrie pétrolière, le champs de perméabilité peut être fortement discontinu au passage d'une roche à l'autre. Cette même perméabilité est par ailleurs souvent anisotrope dans les réservoirs de pétrole avec de forts contrastes entre perméabilité horizontale et perméabilité verticale. Outre cette nature fortement hétérogène et anisotrope du tenseur de diffusion, une autre particularité de ce contexte est que la simulation des écoulements doit se faire sur des maillages dont la géométrie est adaptée aux couches géologiques, ce qui oblige à considérer des mailles de forme très générale pour la discrétisation de (1.1). Pour reprendre l'exemple de l'industrie pétrolière, la structure des réservoirs peut ainsi être fortement déformée par des phénomènes de compaction qui accompagnent la production d'hydrocarbures. Les schémas de discrétisation standards n'étant pas adaptés à de telles contraintes, l'utilisation de ces méthodes pour résoudre numériquement ce type de problème peut aboutir à des oscillations et/ou des solutions non-physiques [60]. Un grand nombre de schémas ont ainsi vu le jour au cours de ces dernières décennies dans le but de proposer une discrétisation spécifique de (1.1) adaptée à ces contraintes particulières. Sans prétendre à l'exhaustivité, mentionnons parmi cette vaste zoologie de méthodes numériques :

- Les méthodes de volumes finis à flux numériques multi-points ou MPFA pour "Multi-Point Flux Approximation". Ces méthodes ont initialement été développées dans les années 90 de façon indépendante par Aavatsmark *et al.* [1, 2], et Edwards et Rogers [59] comme extension à des maillages non-orthogonaux et pour des tenseurs de diffusion anisotropes hétérogènes de la méthode de volumes finis classique de Eymard, Gallouët et Herbin [61]. Aavatsmark *et al.* [3, 4] ont ensuite proposé un schéma MPFA à stencil compact, dit MPFA L, méthode récemment généralisée par Agélas *et al.* dans [5].

- Les méthodes variationnelles. Il s'agit des méthodes de différences finies mimétiques de Brezzi *et al.* [22, 23], de volumes finis hybrides de Eymard *et al.* [63, 64, 65], de volumes finis mixtes de Droniou et Eymard [52, 53] et des méthodes de Galerkine centrées aux mailles de Di Pietro et Vohralík [44, 45]. Ces différents schémas, regroupés dans [48] sous le nom de "Variational lowest-order methods", ont ceci en commun que leur écriture s'inspire de la formulation variationnelle de (1.1). L'existence de liens, parfois très étroits, entre ces derniers à récemment fait l'objet de plusieurs travaux. Citons par exemple l'analyse menée par Droniou *et al.* [54], lesquels prouvent l'équivalence algébrique entre des versions généralisées des méthodes hybrides, mimétiques et mixtes. Notons enfin le travail récent de Droniou *et al.* [55] qui définit une classe plus large, dite de schémas gradients, englobant les schémas précédents.

- Les méthodes de volumes finis en dualité discrète ou DDFV pour "Discrete Duality Finite Volume". Ces méthodes ont été introduites par Hermeline [76] et Domelevo et Omnes [49] pour la discrétisation de l'équation de Laplace sur des maillages généraux.

Elles ont ensuite été étendues notamment dans [11] puis dans [21] par Boyer et Hubert pour la discrétisation d'opérateurs très généraux de type Leray-Lions (qui inclut notamment l'opérateur de diffusion de (1.1)). De manière générale, ces schémas se basent sur la construction d'un gradient discret et d'un opérateur de divergence discrète qui, tout comme leurs homologues continus, sont en dualité (à travers une formule de Green discrète).

- Les méthodes de Galerkine discontinues. L'utilisation de méthodes de Galerkine discontinues pour la discrétisation d'opérateurs de diffusion remonte aux travaux de Baker [13], Wheeler [116] et Arnold [12]. Plus récemment, celles-ci ont été notamment utilisées pour la simulation du transport réactif (voir par exemple les travaux de Sun et Wheeler [112, 111], Bastian *et al.* [14], Houston *et al.* [77], Di Pietro *et al.* [47]).

Cet état des lieux, bien que partiel, atteste néanmoins par la diversité des méthodes existantes des difficultés que suscite la discrétisation du problème de diffusion (1.1) (on renvoie à [51] pour une présentation détaillée des discrétisations de (1.1) par des méthodes de volumes finis et à [75, 66] pour un comparatif complet des performances des différents schémas cités ci-dessus). Un des obstacles réside dans la préservation de deux propriétés fondamentales de l'équation (1.1) :

(i) La coercivité de l'opérateur de diffusion. Conserver un analogue discret de la propriété de coercivité permet notamment d'établir des estimations d'énergie sur la solution discrète, estimations qui, comme dans le cas continu, sont un gage de compacité pour cette solution approchée.

(ii) Le principe du maximum. La conservation de ce dernier lors de la discrétisation garantit le respect des bornes physiques du modèle ; par exemple, dans le cas où l'inconnue est une concentration, le fait d'obtenir une concentration numérique effectivement comprise entre 0 et 1.

Aussi, si la plupart des schémas de discrétisation linéaires classiques pour (1.1) sont coercifs[1], le respect du principe du maximum n'est pour autant pas toujours garanti au niveau discret, en particulier pour des maillages très déformés ou pour de forts ratios d'anisotropie [75, 66, 104]. Les travaux de Buet et Cordier [24] et de Keilegavlen *et al.* [82] prouvent même qu'il n'existe pas de schéma volumes finis linéaire à neuf points consistant qui respecte le principe du maximum, que ce soit dans le cas de forts ratios d'anisotropie (sur un maillage de carrés) ou sur des maillages très déformés (avec un tenseur de diffusion isotrope). Plusieurs schémas de discrétisation non-linéaires ont ainsi vu le jour au cours des dernières années dans le but d'obtenir un principe du maximum discret (on renvoie au chapitre introductif pour une présentation succincte de ces travaux). Néanmoins pour nombre de ces méthodes, telles celles développées dans [88, 81, 93, 95, 90, 117, 110, 56], la propriété de coercivité n'est pas assurée sans condition sur la géométrie du maillage ou le ratio d'anisotropie. L'antagonisme qui semble ainsi exister du point de vue discret entre coercivité et principe du maximum donne un aperçu de l'enjeu de la construction d'une méthode de discrétisation respectant simultanément ces deux propriétés.

[1] Sous certaines conditions sur le maillage et/ou le tenseur de diffusion D pour les schémas à flux numériques multi-points.

1.2 Un schéma parfait

Sous certaines conditions d'admissibilité sur la géométrie du maillage de l'ouvert Ω, la discrétisation de (1.1) par la méthode de volumes finis standard développée par Eymard, Herbin et Gallouët [61] préserve à la fois la coercivité de l'opérateur elliptique et le principe du maximum. Bien entendu, ces conditions géométriques sont souvent impossibles à vérifier en pratique et sont en outre d'autant plus restrictives que le tenseur D est hétérogène. Néanmoins, compte tenu de sa simplicité d'écriture et de la richesse de ses propriétés, l'étude de ce schéma est un bon point de départ à la fois pour comprendre les principes de la discrétisation de (1.1) au sens des volumes finis mais aussi, ce qui nous intéressera par la suite, pour identifier une structure de schéma qui assure le respect du principe du maximum. Pour cela, avant d'aborder le cas d'un schéma volumes finis centré générique, nous décrivons brièvement dans cette section la discrétisation du problème (1.1) au sens des volumes finis en nous appuyant sur ce schéma particulier.

1.2.1 Principe de la discrétisation au sens des volumes finis

La discrétisation de (1.1) au sens des volumes finis s'appuie sur la nature conservative du problème. Il s'agit plus précisément de traduire, du point de vue discret, le bilan de conservation vérifié par la quantité $\bar{u}$

$$-\int_{\partial V} D\nabla\bar{u}\cdot\mathbf{n} = \int_V f,$$

valable sur tout volume de contrôle V inclus dans l'ouvert Ω ($\mathbf{n}$ désignant ici la normale extérieure à ∂V). Cela revient en un sens à faire le chemin inverse[2] de celui qui conduit à l'équation (1.1). Ainsi étant donné un maillage de l'ouvert Ω, c'est à dire une famille $\mathcal{M}$ de volumes de contrôle polygonaux telle que $\overline{\Omega} = \cup_{K\in\mathcal{M}}\overline{K}$ (voir Figure 1.1), le bilan de

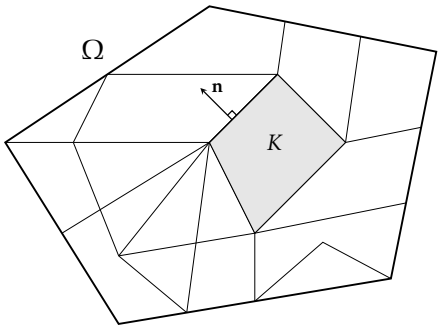

FIGURE 1.1 – Maillage du domaine polygonal Ω.

conservation précédent, écrit sur chacune de ces mailles, fournit un nombre fini d'identités (une par maille) vérifiées par la solution exacte :

$$\forall K\in\mathcal{M},\quad -\sum_{\sigma}\int_{\sigma} D\nabla\bar{u}\cdot\mathbf{n} = \int_K f, \tag{1.2}$$

[2]Chemin qui consiste, partant de l'équation (1.1), à intégrer celle-ci sur le volume de contrôle considéré puis à appliquer le théorème de Stokes à l'intégrale volumique sur $\bar{u}$ ainsi obtenue.

où la somme porte sur les arêtes du volume de contrôle K. Construire un schéma de discrétisation revient alors à spécifier une règle pour approcher les flux surfaciques $\int_\sigma D\nabla\bar{u}\cdot\mathbf{n}$ en fonction des valeurs de $\bar{u}$ que l'on cherche à déterminer. Le schéma de discrétisation proprement dit consiste en la recherche d'une famille d'inconnues $u=(u_K)_{K\in\mathcal{M}}$ (une inconnue u_K par maille destinée à approcher $\bar{u}$ sur la maille K) solution du système obtenu à partir de (1.2) en remplaçant les flux surfaciques suivant la règle ainsi spécifiée.

1.2.2 Le schéma à flux deux points

Considérons le cas d'une diffusion homogène et isotrope $D=\mathrm{Id}$. Dans ce cas, un choix simple pour calculer $\nabla\bar{u}\cdot\mathbf{n}$ sur une arête située entre deux mailles K et L est le suivant :

$$\nabla\bar{u}\cdot\mathbf{n}\approx\frac{\bar{u}(x_L)-\bar{u}(x_K)}{|x_L-x_K|},\tag{1.3}$$

où x_K et x_L sont des points bien choisis dans les mailles K et L (les "centres de mailles"). Précisément, on montre que l'approximation précédente est consistante à condition que la

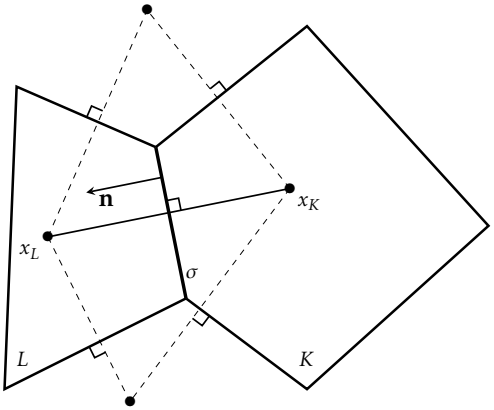

FIGURE 1.2 – Condition d'admissibilité pour le schéma deux-points.

droite (x_Kx_L) soit perpendiculaire à la face commune à K et L (voir Figure 1.2)[3]. Le schéma volumes finis obtenu en suivant ce principe simple d'approximation consiste ainsi à trouver des inconnues discrètes $(u_K)_{K\in\mathcal{M}}$ vérifiant le système de bilans discrets

$$\forall K\in\mathcal{M},\quad -\sum_\sigma F_{K,\sigma}=\int_K f,\tag{1.4}$$

où le flux discret $F_{K,\sigma}$, approximation de $\int_\sigma\nabla\bar{u}\cdot\mathbf{n}$ sur l'arête σ entre les mailles K et L, est donné suivant la règle "à deux points" (1.3) par

$$F_{K,\sigma}=|\sigma|\,\frac{u_L-u_K}{|x_L-x_K|}.$$

L'élégance du schéma précédant réside notamment dans le fait que le choix simple sur lequel repose sa construction permet de conserver à la fois coercivité et principe du

[3] Cela tient au fait que, sous cette condition d'orthogonalité, la direction $\mathbf{n}$ dans laquelle on souhaite approcher le gradient est aussi celle du vecteur x_L-x_K.

maximum au niveau discret (voir [61] pour plus de détails). Malheureusement, comme nous l'avons noté précédemment, ce schéma est inutilisable sur des maillages déformés pour lesquels la condition d'orthogonalité n'est pas vérifiée. En outre, dans le cas plus général où le tenseur de diffusion est anisotrope, la condition géométrique assurant la consistance d'une approximation deux points devient une condition d'orthogonalité au sens du produit scalaire défini par les valeurs moyennes de D sur les volumes de contrôle (voir [61, Définition 3.8]), condition d'autant plus contraignante que le tenseur est hétérogène. Néanmoins, l'observation de la structure de ce schéma est utile en vue de la construction de schémas de discrétisation respectant le principe du maximum.

1.2.3 Une structure préservant le principe du maximum

Dans le cas de la méthode deux points, le respect du principe du maximum résulte de ce que l'opérateur de diffusion discret s'écrit la forme monotone suivante :

$$-\sum_{\sigma} F_{K,\sigma} = \sum_{L\in\mathcal{M}} \tau_{K,L}(u_K - u_L),$$

où les transmissivités $\tau_{K,L}$ sont positives (voir la définition 2.1 du Chapitre 2 pour un énoncé précis). Aussi, la question qui va nous occuper dans le chapitre suivant va être de savoir comment modifier un schéma de discrétisation générique pour (1.1) afin d'assurer le respect du principe du maximum. Dans cette optique, l'identification de cette structure particulière (appelée "Local Maximum Principle structure" dans [56]) permet de dégager un critère explicite pour corriger le schéma de départ. En d'autres termes, on cherchera à corriger celui-ci dans le but de lui conférer cette structure monotone. Mais il s'agira également de s'intéresser au devenir de certaines des propriétés initiales de ce même schéma. En particulier, dans le cas où ce dernier est coercif, les modifications apportées se devront de ne pas altérer cette propriété de coercivité. Par conséquent, avant d'aborder la question de la correction proprement dite, nous dressons dans la partie suivante le portrait type d'un schéma générique pour (1.1) ainsi que des propriétés éventuelles de celui-ci que nous chercherons à préserver.

1.3 Forme générique d'un schéma de type volumes finis centré aux mailles

La donnée d'un schéma de discrétisation pour le problème de diffusion anisotrope (1.1) sera le point de départ du chapitre suivant où nous aborderons la construction de corrections monotones. Dans un souci de généralité, nous considèrerons un schéma volumes finis centré écrit sous forme générique en nous concentrant sur certaines propriétés essentielles que l'on retrouve chez de nombreux schémas de discrétisation.

1.3.1 Cadre de la discrétisation

1.3.1.1 Maillages généraux

Dans tout ce qui suit, nous souhaitons pouvoir considérer des maillages généraux de l'ouvert Ω. Les hypothèses géométriques minimales que nous imposerons sur la discrétisation sont précisées dans la définition suivante.

Définition 1.1. Un maillage admissible de Ω est un triplet $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ vérifiant les conditions suivantes :

- $\mathcal{M}$ est une famille disjointe d'ouverts non vides et polygonaux de Ω (les *volumes de contrôles*) telle que $\overline{\Omega} = \cup_{K\in\mathcal{M}}\overline{K}$.

- $\mathcal{E}$ est une famille finie disjointe de parties de Ω (les *arêtes* du maillage) telle que, pour tout élément σ de $\mathcal{E}$, il existe un hyperplan affine E de $\mathbb{R}^d$ et un volume de contrôle K pour lesquels $\sigma \subset \partial K \cap \mathcal{E}$ et σ est un ouvert non vide de E. En outre, pour tout $K \in \mathcal{M}$, $\mathcal{E}$ admet un sous-ensemble $\mathcal{E}_K$ tel que $\partial K = \cup_{\sigma\in\mathcal{E}_K}\overline{\sigma}$. Les éléments de $\mathcal{E}$ sont enfin supposés vérifier l'alternative suivante : soit $\sigma \in \partial\Omega$ soit il existe $K, L \in \mathcal{M}$ distincts tels que $\sigma \in \overline{K} \cap \overline{L}$.

- $\mathcal{P} = \{x_K\}_{K\in\mathcal{M}}$ est un ensemble de points de Ω (les *centres de mailles*) tels que, pour tout $K \in \mathcal{M}$, K est étoilé par rapport à x_K.

Remarque 1.2. Il convient de noter que les éléments de $\mathcal{E}_K$ ne sont pas les arêtes à proprement parler de la maille K. Une arête entière de la frontière polygonale de K peut en effet être divisée en plusieurs "arêtes" de la discrétisation. Notons par ailleurs que les volumes de contrôle ne sont pas supposés convexes, de sorte que deux mailles voisines peuvent avoir plusieurs arêtes en commun.

Etant donné une discrétisation admissible $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ de Ω, nous adopterons dans ce chapitre ainsi que dans le suivant les notations suivantes :

- La mesure d'un volume de contrôle K est notée $|K|$.

- La mesure $(d-1)$-dimensionnelle d'une arête σ est notée $|\sigma|$.

- Pour tous $K, L \in \mathcal{M}$, on note $K|L = \mathcal{E}_K \cap \mathcal{E}_L$ l'ensemble (éventuellement vide) des arêtes communes à K et L et on pose $|K|L| = \sum_{\sigma\in K|L} |\sigma|$.

- On définit l'ensemble des arêtes internes par $\mathcal{E}_{\text{int}} = \{\sigma \in \mathcal{E} \ ; \ \sigma \not\subset \partial\Omega\}$.

- On définit l'ensemble des arêtes externes par $\mathcal{E}_{\text{ext}} = \{\sigma \in \mathcal{E} \ ; \ \sigma \subset \partial\Omega\}$.

- Pour tout $K \in \mathcal{M}$, $\mathcal{N}_K$ désigne l'ensemble des volumes de contrôle voisins de K *i.e.* $\mathcal{N}_K = \{L \in \mathcal{M} \setminus \{K\} \ ; \ K|L \neq \emptyset\}$.

- Pour tout $x_K \in \mathcal{P}$, si $\sigma \in \mathcal{E}_K$, on note $d_{K,\sigma}$ la distance de x_K à l'hyperplan contenant σ.

- Pour toute arête $\sigma \in \mathcal{E}$, on pose $d_\sigma = d_{K,\sigma} + d_{L,\sigma}$ si $\sigma \in K|L$ est une arête interne et $d_\sigma = d_{K,\sigma}$ si $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}}$.

Enfin, afin d'étudier la convergence des schémas de discrétisation nous définissons d'une part la taille du maillage

$$\text{size}(\mathcal{D}) = \sup_{K\in\mathcal{M}} \text{diam}(K),$$

d'autre part nous aurons besoin de contrôler la régularité de la discrétisation

$$\text{regul}(\mathcal{D}) = \sup_{\substack{K\in\mathcal{M}\\ \sigma\in\mathcal{E}_K}} \left\{\frac{\text{diam}(K)}{d_{K,\sigma}}\right\} + \sup_{\substack{K,L\in\mathcal{M}\\ \sigma\in K|L}} \left\{\frac{d_{L,\sigma}}{d_{K,\sigma}}\right\}.$$

1.3.1.2 Cadre fonctionnel discret

Etant donné un maillage admissible $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ de Ω, on note $\mathcal{H}_\mathcal{M}$ l'espace des fonctions de Ω dans $\mathbb{R}$ constantes sur chacun des volumes de contrôle de $\mathcal{M}$. Nous identifierons tout élément u de $\mathcal{H}_\mathcal{M}$ avec la famille $(u_K)_{K\in\mathcal{M}}$ de $\mathbb{R}^{\mathrm{Card}\,\mathcal{M}}$ formée des valeurs de u sur chaque volume de contrôle. L'espace $\mathcal{H}_\mathcal{M}$ est muni d'une norme H_0^1 discrète définie par

$$\forall u \in \mathcal{H}_\mathcal{M}, \quad \|u\|_\mathcal{D}^2 = \sum_{\sigma\in\mathcal{E}} |\sigma| \frac{|u_K - u_L|^2}{d_\sigma},$$

où si $\sigma \in \mathcal{E}_{\mathrm{int}}$, K et L sont les deux volumes de contrôle qui partagent σ et si $\sigma \in \mathcal{E}_{\mathrm{ext}}$, $u_L = 0$ et K est le volume de contrôle tel que $\sigma \in \mathcal{E}_K$.

On est à même d'établir un analogue discret de l'inégalité de Poincaré pour la norme précédente comme le rappelle le lemme qui suit, que l'on peut par exemple déduire du lemme 5.3 de [65].

Lemme 1.3. *Soit $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$ un maillage admissible de Ω et soit $\theta \in \mathbb{R}_+^*$ tel que* $\mathrm{regul}(\mathcal{D}) \le \theta$. *Il existe une constante C, qui ne dépend que de Ω et θ, telle que*

$$\forall u \in \mathcal{H}_\mathcal{M}, \quad \|u\|_{L^2(\Omega)} \le C \|u\|_\mathcal{D}. \tag{1.5}$$

On dispose également d'un résultat de compacité dans $L^2(\Omega)$ pour les familles bornées de $\mathcal{H}_\mathcal{M}$ au sens de la norme H_0^1 discrète (voir par exemple les lemmes 5.6 et 5.7 de [65] dans le cas $p = 2$).

Lemme 1.4. *Soit $(\mathcal{D}^n)_{n\ge1}$ une suite de maillages admissibles de Ω telle que* $\mathrm{size}(\mathcal{D}^n) \to 0$ *lorsque $n \to \infty$ et telle que* $(\mathrm{regul}(\mathcal{D}^n))_{n\ge1}$ *est bornée. Soit $(u^n)_{n\ge1}$ telle que $u^n \in \mathcal{H}_{\mathcal{D}^n}$ pour tout $n \ge 1$. On suppose qu'il existe une constante M telle que*

$$\forall n \ge 1, \quad \|u\|_{\mathcal{D}^n} \le M.$$

Alors $\{u^n\}_{n\ge1}$ admet dans $L^2(\Omega)$ une valeur d'adhérence u telle que $u \in H_0^1(\Omega)$.

1.3.2 Schémas de type volumes finis centrés aux mailles pour des problèmes de diffusion anisotrope et hétérogène

Nous nous limitons à des schémas numériques centrés aux mailles pour la discrétisation du problème de diffusion (1.1). De façon très générale, étant donné un maillage admissible $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$, il s'agit d'un système d'équations sur les inconnues $(u_K)_{K\in\mathcal{M}}$ aux centres de mailles. Un tel schéma est donc caractérisé par la donnée d'une application

$$\begin{aligned} \mathcal{S}^\mathcal{D} : \mathcal{H}_\mathcal{M} &\longrightarrow \mathcal{H}_\mathcal{M} \\ u &\longmapsto (\mathcal{S}_K(u))_{K\in\mathcal{M}}, \end{aligned}$$

et consiste plus précisément à trouver $u \in \mathcal{H}_\mathcal{M}$ telle que

$$\forall K \in \mathcal{M}, \quad \mathcal{S}_K(u) = |K| f_K, \tag{1.6}$$

où l'on a noté f_K la valeur moyenne du second membre f de (1.1) sur la maille K.

En vue d'étudier l'effet d'une modification apportée sur un tel schéma de discrétisation, il convient de préciser les propriétés principales dont le devenir nous importe.

1.3.2.1 Conservativité

D'après la définition générale ci-dessus, un schéma numérique pour (1.1) et donc donné par une famille de fonctions $\mathcal{S}_K : \mathcal{H}_\mathcal{M} \to \mathbb{R}$ au sens où, pour tout $K \in \mathcal{M}$, l'équation sur le volume de contrôle K s'écrit $\mathcal{S}_K(u) = |K| f_K$. Un tel schéma est conservatif (au sens des volumes finis) si chacune de ces équations peut s'écrire sous la forme d'un bilan discret sur des flux approchés.

Définition 1.5 (Conservativité)**.** Soit $\mathcal{D}$ un maillage admissible de Ω et soit $\mathcal{S}^\mathcal{D}$ un schéma numérique pour (1.1). Ce schéma est dit conservatif s'il existe une famille $(F_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$ de fonctions $F_{K,\sigma} : \mathcal{H}_\mathcal{M} \to \mathbb{R}$ (les *flux numériques*) telle que :

$$\forall K \in \mathcal{M}, \forall L \in \mathcal{N}_K, \forall \sigma \in K|L, \quad F_{K,\sigma} + F_{L,\sigma} = 0, \tag{1.7}$$

$$\forall K \in \mathcal{M}, \quad \mathcal{S}_K = -\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}. \tag{1.8}$$

1.3.2.2 Coercivité

On peut définir une notion de coercivité pour un schéma numérique, analogue discret de la coercivité de la forme bilinéaire associée à la formulation variationnelle de (1.1).

Définition 1.6 (Coercivité)**.** Soit $\mathcal{D}$ un maillage admissible de Ω. Un schéma numérique $\mathcal{S}^\mathcal{D}$ pour (1.1) est dit coercif s'il existe $\zeta \in \mathbb{R}_+^*$ tel que

$$\forall u \in \mathcal{H}_\mathcal{M}, \quad \sum_{K\in\mathcal{M}} \mathcal{S}_K(u) u_K \geq \zeta \, \|u\|_\mathcal{D}^2. \tag{1.9}$$

Comme nous l'avons mentionné précédemment, l'intérêt de cette propriété est qu'elle assure un contrôle des solutions du schéma en norme H_0^1 discrète.

Proposition 1.7 (Estimation *a priori*)**.** *Soit $\mathcal{D}$ un maillage admissible de Ω et soit $\mathcal{S}^\mathcal{D}$ un schéma coercif pour* (1.1) *de constante de coercivité ζ. Si $\theta \geq \mathrm{regul}(\mathcal{D})$, alors il existe une constante C dépendant uniquement de Ω, ζ et θ et telle que pour toute solution discrète u de* (1.6)

$$\|u\|_\mathcal{D} \leq C \, \|f\|_{L^2(\Omega)}. \tag{1.10}$$

Démonstration. La preuve est calquée sur la preuve de l'homologue continu de cette estimation. Précisément, en multipliant l'équation sur K dans (1.6) par u_K, en sommant ensuite les égalités obtenues sur l'ensemble des volumes de contrôle puis en utilisant enfin l'hypothèse de coercivité, on obtient

$$\zeta \, \|u\|_\mathcal{D}^2 \leq \int_\Omega f u. \tag{1.11}$$

L'inégalité (1.10) résulte alors de l'application successive de l'inégalité de Cauchy-Schwarz au second membre de (1.11) puis de l'inégalité de Poincaré discrète. □

1.3.2.3 Consistance

Le contrôle en norme H_0^1 discrète dont on dispose pour un schéma coercif est un gage de compacité pour la solution discrète (ceci grâce au critère du Lemme 1.4). Il assure ainsi la convergence de la solution numérique vers un élément de $H_0^1(\Omega)$ lorsque la taille du maillage tend vers 0. Afin de prouver que cette limite est solution (faible) du problème continu (1.1), il s'agit ensuite de s'assurer que l'on peut passer à la limite dans les équations du schéma numérique.

Définition 1.8 (Consistance)**.** Soit $(\mathcal{D}^n)_{n\geq 1}$ une suite de maillages admissibles de Ω telle que $\text{size}(\mathcal{D}^n) \to 0$ lorsque $n \to \infty$. Soit $(\mathcal{S}^n)_{n\geq 1}$ telle que, pour tout $n \geq 1$, $\mathcal{S}^n = (\mathcal{S}^n_K)_{K\in\mathcal{M}^n}$ définit un schéma pour (1.1) associé à la discrétisation $\mathcal{D}^n = (\mathcal{M}^n, \mathcal{E}^n, \mathcal{P}^n)$. La famille de schémas $(\mathcal{S}^n)_{n\geq 1}$ est dite consistante avec l'équation (1.1) si, pour toute famille $(u^n)_{n\geq 1}$ de fonctions discrètes vérifiant les conditions suivantes :

- Pour tout $n \geq 1$, $u^n \in \mathcal{H}_{\mathcal{M}^n}$,
- il existe $M > 0$ tel que, pour tout $n \geq 1$, $\|u^n\|_{\mathcal{D}^n} \leq M$,
- il existe $\bar{u} \in H^1_0(\Omega)$ tel que $u^n \to \bar{u}$ dans $L^2(\Omega)$ lorsque $n \to \infty$,

on a

$$\forall \varphi \in \mathcal{C}^\infty_c(\Omega), \quad \lim_{n\to\infty} \sum_{K\in\mathcal{M}^n} \mathcal{S}^n_K(u^n)\varphi(x_K) = \int_\Omega D\nabla\bar{u}\cdot\nabla\varphi. \tag{1.12}$$

1.3.2.4 Tour d'horizon des discrétisations centrées aux mailles pour les opérateurs de diffusion

Dans le tableau 1.1 donné ci-dessous, nous avons listé quelques uns des schémas volumes finis centrés existants pour la discrétisation du problème de diffusion (1.1). Pour chacun d'eux nous précisons quelles propriétés sont respectées parmi les trois définies ci-dessus et, le cas échéant, si le respect de ces propriétés est soumis à des conditions spécifiques.

Table 1.1 – Quelques schémas volumes finis centrés pour la discrétisation d'opérateurs de diffusion anisotropes hétérogènes

Schéma	Conservativité	Coercivité	Consistance
MPFA O ([1, 7])	✓	✓[a]	✓[a]
Schéma Dioptre [b] ([6])	✓	✓	✓
Schéma de Eymard *et al.* [62][c]	✓	✓	✓
SUSHI (barycentrique) ([65, 67])	pas au sens usuel	✓	✓
VFSYM[d] ([89])	✓	✓	✓
Schéma de Lipnikov *et al.* [94]	✓	✓[e]	✓[e]

[a] sur un maillage de parallélogrammes/parallélépipèdes
[b] avec conditions sur le maillage et la ratio d'anisotropie
[c] sur des maillages admissibles
[d] sur des maillages de simplexes ou de parallélogrammes/parallélépipèdes
[e] sur des maillages de simplexes

CHAPITRE

2 Monotone corrections for generic cell-centered Finite Volume approximations of anisotropic diffusion equations

In this chapter, we present a nonlinear technique to correct a generic Finite Volume scheme for anisotropic diffusion problems, which provides a discrete maximum principle. We prove the proposed corrections preserve the general properties defined in Chapter 1. We then study two specific corrections proving, under numerical assumptions, that the corresponding approximate solutions converge to the continuous one as the size of the mesh tends to 0. Finally we present numerical results showing that these corrections suppress local minima produced by the original Finite Volume scheme.

Contents

2.1 Statement of the problem

Let Ω be an open bounded connected polygonal subset of $\mathbb{R}^d$. We shall consider numerical approximations to the anisotropic heterogeneous problem

$$\begin{cases} -\nabla\cdot(D\nabla\bar{u}) = f & \text{in } \Omega, \\ \bar{u} = 0 & \text{on } \partial\Omega; \end{cases} \tag{2.1}$$

under the following assumptions:

- $f \in L^2(\Omega)$;
- $D : \Omega \to \mathcal{M}_d(\mathbb{R})$ is a bounded measurable function such that $D(x)$ is symmetric for a.e. $x \in \Omega$ and that there exists $\lambda > 0$ satisfying $D(x)\xi \cdot \xi \geq \lambda |\xi|^2$ for a.e. $x \in \Omega$ and all $\xi \in \mathbb{R}^d$.

As mentioned in the previous chapter, it is well known that classical linear methods discretizing diffusion operators do not always satisfy maximum principle for distorted meshes or with high anisotropy ratio [75, 104]. That is the reason why the question of constructing numerical methods for (2.1) ensuring the approximate solution satisfies a discrete maximum principle has been investigated. In [25], a non-linear stabilization term is introduced to design a Galerkin approximation of the Laplacian, but heterogeneous anisotropic tensors are not considered. More recently, a few non-linear finite volume schemes have been proposed to discretize elliptic problems [70, 81, 93, 90, 95, 117, 110, 56]. For theses methods, the authors obtained the desired properties and accurate results which are generally second order in space. Unfortunately, none of these methods can ensure that they are coercive without conditions on the geometry or on the anisotropy ratio.

Starting from any given cell-centered finite volume scheme, our goal, in the present work, is to elaborate, in the spirit of methods described in [91], a general approach to construct non-linear corrections providing a discrete maximum principle while retaining some main properties of the scheme, in particular coercivity and convergence toward the solution of (2.1) as the size of the mesh tends to zero. To do so, we proceed step by step, beginning with a general correction and then refining it by considering successively the required properties. The corrections we obtain give nonoscillating solutions and can be applied, for example, to the cell-centered finite volume schemes developed in [1, 8, 6, 62, 89, 94] (see Table 1.1 of Chapter 1). Let us notice that these new corrections are quite easy to implement since they conserve the data structure used for the original linear scheme that has been corrected.

It is also worth mentioning the recent contribution [43] where a closely related question is investigated, i.e. the convergence of nonlinearly corrected Finite Volume schemes towards the solution of the unidimensional heat equation.

This chapter is organized as follows. In the rest of this introduction we define a specific structure of schemes (the so-called *LMP structure*, cf. Definition 2.1), which yields a discrete version of the local maximum principle. Using the abstract framework developed in Chapter 1, we address in section 2.2 the problem of correcting a generic convergent cell centered finite volume scheme in order to enforce the LMP structure. In section 2.2.1 we state the main assumptions that can be made on the generic original scheme to be corrected, and that we expect to keep on its corrected version. Section 2.2.2 then establishes sufficient conditions for the corrections to bring the desired structure while retaining conservation property and coercivity. Section 2.2.3 is devoted to the convergence of the corrected scheme. In section 2.3, we detail two examples of non-linear corrections and we perform for both a theoretical study of the corrected scheme. The convergence proofs rely in both cases on numerical assumptions on the approximate solutions to these schemes. The numerical results we present in section 2.4 confirm these assumptions seems to be actually fulfilled even for strongly anisotropic permeabilities.

2.1.1 Local Maximum Principle structure

The main problem we address is to modify a cell-centered finite volume scheme in order to enforce the preservation of the maximum-principle. More precisely, using the terminology

of [56], we focus on the following class of schemes.

Definition 2.1 (LMP structure)**.** Let $\mathcal{D}$ be an admissible mesh of Ω. A scheme $\mathcal{S}^{\mathcal{D}}$ for (2.1) has the Local Maximum Principle structure (LMP structure for short) if it can be written

$$\forall K \in \mathcal{M}, \quad \mathcal{S}_K(u) = \sum_{L\in\mathcal{M}} \tau_{K,L}(u)(u_K - u_L) + \sum_{\sigma\in\mathcal{E}_{ext}} \tau_{K,\sigma}(u)u_K, \tag{2.2}$$

with functions $\tau_{K,L} : \mathbb{R}^{\mathrm{Card}(\mathcal{M})} \to \mathbb{R}_+$ (for $K, L \in \mathcal{M}$) and $\tau_{K,\sigma} : \mathbb{R}^{\mathrm{Card}(\mathcal{M})} \to \mathbb{R}_+$ (for $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_{\mathrm{ext}}$) satisfying, for all $u \in \mathbb{R}^{\mathrm{Card}M}$,

$$\forall K \in \mathcal{M}, \forall L \in \mathcal{N}_K, \quad \tau_{K,L}(u) > 0, \tag{2.3a}$$

$$\forall K \in \mathcal{M}, \forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\mathrm{ext}}, \quad \tau_{K,\sigma}(u) > 0. \tag{2.3b}$$

The schemes having the LMP structure meet a discrete version of the maximum principle as stated by the following proposition, whose proof is given in [56].

Proposition 2.2 (Discrete Maximum Principle)**.** *Assume that $f \geq 0$ on Ω. If $u = (u_K)_{K\in\mathcal{M}}$ is a solution to a scheme having the LMP structure, then* $\min_{K\in\mathcal{M}} u_K \geq 0$.

2.2 Non-linear corrections of a generic cell-centered finite volume scheme

Starting from a cell-centered scheme (for instance one from Table 1.1), we describe in this section how to construct a non-linear correction which gives the LMP structure while paying attention not to lose the main properties of the original scheme, namely conservativity, coercivity or consistency. We first state the main assumptions on the original scheme. Then we detail some general guidelines about the construction of such corrections.

2.2.1 The original scheme

Let us denote by A the continuous operator from problem (2.1) defined by $A(\bar{u}) = \nabla\cdot(D\nabla\bar{u})$.

In the following, we consider a generic discrete approximation $\mathcal{A}^{\mathcal{D}} : \mathcal{H}_{\mathcal{M}} \to \mathcal{H}_{\mathcal{M}}$ of the operator A. $\mathcal{A}^{\mathcal{D}}$ defines a scheme for (2.1) that writes

$$-\mathcal{A}^{\mathcal{D}}(u) = f_{\mathcal{D}}, \tag{2.4}$$

where we let $f_{\mathcal{D}} = (|K| f_K)_{K\in\mathcal{M}} \in \mathcal{H}_{\mathcal{M}}$. We assume that $\mathcal{A}^{\mathcal{D}} : \mathcal{H}_{\mathcal{M}} \to \mathcal{H}_{\mathcal{M}}$ is linear and invertible so that the original scheme (2.4) has a unique solution.

For the sake of clarity, it is convenient to introduce, for any $u \in \mathcal{H}_{\mathcal{M}}$, additional (trivial) values $(u_\sigma)_{\sigma\in\mathcal{E}_{\mathrm{ext}}}$ which we all take equal to zero. We denote by $V(K) \subset \mathcal{M} \cup \mathcal{E}_{\mathrm{ext}}$ the sets corresponding to the stencil of this scheme and we suppose the discrete linear operator $\mathcal{A}^{\mathcal{D}}$ writes in the following form[1]:

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \forall K \in \mathcal{M}, \quad \mathcal{A}_K(u) = \sum_{Z\in V(K)} \alpha_{K,Z}(u_Z - u_K) \tag{2.5}$$

[1] Using additional unknowns u_σ playing the role of approximation of $\bar{u}$ on the boundary edges, assuming (2.5) is nothing but assuming that the scheme is exact when applied to constant families: $\mathcal{A}^{\mathcal{D}}(u) = 0$ if $u = ((u_K)_{K\in\mathcal{M}}, (u_\sigma)_{\sigma\in\mathcal{E}_{\mathrm{ext}}}) = \text{constant}$.

(where with the previous convention $u_Z = 0$ if $Z = \sigma \in \mathcal{E}_{\text{ext}}$). If need be by adding some null coefficients, we further suppose the stencil contains the neighboring cells (that is $V(K) \supset \mathcal{N}_K$) and that it is symmetric in the following sense:

$$\forall (K,L) \in \mathcal{M}^2, \quad L \in V(K) \implies K \in V(L). \tag{2.6}$$

In the following we address the problem of correcting this original scheme in order to provide it the LMP structure. Except from this property we want to reach, we focus on preserving any of the following additional properties:

(A1) The scheme defined by $\mathcal{A}^{\mathcal{D}}$ is conservative with numerical fluxes denoted by $F^{\mathcal{D}} = (F_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$:

$$\forall K \in \mathcal{M}, \quad \mathcal{A}_K = \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}.$$

(A2) There exists $\zeta > 0$, independent of the mesh $\mathcal{D}$, such that the scheme defined by $\mathcal{A}^{\mathcal{D}}$ is coercive with constant ζ:

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \quad -\sum_{K\in\mathcal{M}} \mathcal{A}_K(u) u_K \geq \zeta \, \|u\|_{\mathcal{D}}^2$$

(A3) Let $(\mathcal{D}^n)_{n\geq 1}$ be a sequence of admissible meshes such that $\text{size}(\mathcal{D}^n) \to 0$ as $n \to \infty$. Assume that $(\text{regul}(\mathcal{D}^n))_{n\geq 1}$ and $(\max_{K\in\mathcal{M}^n} \text{Card } V(K))_{n\geq 1}$ are bounded. Then the family of schemes defined by $(\mathcal{A}^{\mathcal{D}^n})_{n\geq 1}$ is consistent with problem (2.1).

2.2.2 General construction of non-linear corrections

Driven by the LMP structure, we consider corrections having the following form.

Definition 2.3 (Correction)**.** Let $\mathcal{D}$ be an admissible mesh of Ω. A correction for the scheme (2.4) defined by $\mathcal{A}^{\mathcal{D}}$ is a family $\beta^{\mathcal{D}} = (\beta_{K,Z})_{K\in\mathcal{M},Z\in V(K)}$ of functions $\beta_{K,Z} : \mathcal{H}_{\mathcal{M}} \to \mathbb{R}$. Given a correction β:

- the corrected scheme $\mathcal{S}^{\mathcal{D}}$ (from (2.4)) is defined by

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \forall K \in \mathcal{M}, \quad \mathcal{S}_K(u) = -\mathcal{A}_K(u) + \sum_{Z\in V(K)} \beta_{K,Z}(u)(u_K - u_Z), \tag{2.7}$$

- the corrective term is the function $\mathcal{R}^{\mathcal{D}} : \mathcal{H}_{\mathcal{M}} \to \mathcal{H}_{\mathcal{M}}$ defined by

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \forall K \in \mathcal{M}, \quad \mathcal{R}_K(u) = \sum_{Z\in V(K)} \beta_{K,Z}(u)(u_K - u_Z). \tag{2.8}$$

2.2.2.1 Monotone corrections

The corrections defined above lead to a scheme having the LMP structure if they match the following conditions.

Proposition 2.4 (Monotone correction)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω and $\beta^{\mathcal{D}}$ be a correction for* (2.4)*. Let $(\gamma_{K,Z})_{K\in\mathcal{M},Z\in V(K)}$ be a family of functions $\gamma_{K,Z} : \mathcal{H}_{\mathcal{M}} \to \mathbb{R}_+$ such that, for all $u \in \mathcal{H}_{\mathcal{M}}$ and all $K \in \mathcal{M}$,*

$$\text{if } \sum_{Z\in V(K)} |u_K - u_Z| \neq 0 \text{ then } \sum_{Z\in V(K)} \gamma_{K,Z}(u)|u_K - u_Z| = 1. \tag{2.9}$$

Assume that $\beta^{\mathcal{D}} = (\beta_{K,Z})_{K\in\mathcal{M},Z\in V(K)}$ *satisfies, for all* $u \in \mathcal{H}_{\mathcal{M}}$ *and all* $K \in \mathcal{M}$,

$$\forall Z \in V(K), \quad \beta_{K,Z}(u) \geq \gamma_{K,Z}(u)|\mathcal{A}_K(u)|, \tag{2.10a}$$

$$\forall L \in \mathcal{N}_K, \quad \beta_{K,L}(u) > \gamma_{K,L}(u)|\mathcal{A}_K(u)|, \tag{2.10b}$$

$$\forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}}, \quad \beta_{K,\sigma}(u) > \gamma_{K,\sigma}(u)|\mathcal{A}_K(u)|. \tag{2.10c}$$

Then the corrected scheme has the LMP structure.

Proof. Let $u \in \mathcal{H}_{\mathcal{M}}$. Using condition (2.9), the coordinate K of the original scheme (2.4) can be written

$$-\mathcal{A}_K(u) = -\sum_{Z\in V(K)} \gamma_{K,Z}(u)|u_K - u_Z|\mathcal{A}_K(u),$$

that is

$$-\mathcal{A}_K(u) = \sum_{Z\in V(K)} \{\gamma_{K,Z}(u)\text{sgn}(u_Z - u_K)\mathcal{A}_K(u)\}(u_K - u_Z). \tag{2.11}$$

Thus the coordinate K of the corrected scheme reads

$$\mathcal{S}_K(u) = \sum_{Z\in V(K)} \{\gamma_{K,Z}(u)\text{sgn}(u_Z - u_K)\mathcal{A}_K(u) + \beta_{K,Z}(u)\}(u_K - u_Z). \tag{2.12}$$

Letting, for $K \in \mathcal{M}$ and $Z \in V(K)$,

$$\tau_{K,Z}(u) = \gamma_{K,Z}(u)\text{sgn}(u_Z - u_K)\mathcal{A}_K(u) + \beta_{K,Z}(u),$$

the corrected scheme takes the form of (2.2)

$$\mathcal{S}_K(u) = \sum_{Z\in V(K)} \tau_{K,Z}(u)(u_K - u_Z),$$

with $\tau_{K,Z} \geq 0$ according to (2.10a). To verify the corrected scheme has the LMP structure, it remains to check that these coefficients meet conditions (2.3), that is, they are positive whenever $Z \in \mathcal{N}_K$ or $Z = \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}}$. This last condition is precisely ensured by the strict inequalities (2.10b) and (2.10c). □

Remark 2.5. Actually, the main condition we have to focus on when building a correction is condition (2.10a). Indeed, assume a correction $\tilde{\beta}^{\mathcal{D}}$ matches condition (2.10a), then, following the above calculus, we can see that the corresponding corrected scheme has the form of (2.2) with the non negative coefficients $\tau_{K,Z}$ given by

$$\tau_{K,Z}(u) = \gamma_{K,Z}(u)\text{sgn}(u_K - u_Z)\mathcal{A}_K(u) + \tilde{\beta}_{K,Z}(u).$$

Now, from $\tilde{\beta}^{\mathcal{D}}$, take positive numbers $(\nu_K)_{K\in\mathcal{M}}$ and define a new correction $\beta^{\mathcal{D}}$ by setting, for $u \in \mathcal{H}_{\mathcal{M}}$, $K \in \mathcal{M}$ and $Z \in V(K)$

$$\beta_{K,Z}(u) = \tilde{\beta}_{K,Z}(u) + \nu_K \frac{|K|Z|}{\text{diam}(K)},$$

where we have extended the notation $K|Z$ to the edges $Z = \sigma \in V(K) \cap \mathcal{E}_{\text{ext}}$ by setting $K|Z = \{\sigma\}$. Then the correction $(\beta^{\mathcal{D}})$ matches all the conditions of (2.10) so that the scheme

corrected with $\beta^{\mathcal{D}}$ has the LMP structure. Note that if we define a discrete Laplacian operator $\Delta^{\mathcal{D}} : \mathcal{H}_{\mathcal{M}} \to \mathcal{H}_{\mathcal{M}}$ by

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \forall K \in \mathcal{M}, \quad \Delta_K(u) = \sum_{\sigma \in \mathcal{E}_K} \frac{|\sigma|}{\mathrm{diam}(K)}(u_L - u_K), \tag{2.13}$$

then using the correction $\beta^{\mathcal{D}}$ amounts to adding some numerical diffusion to the scheme corrected by $\tilde{\beta}^{\mathcal{D}}$. Indeed the scheme corrected with $\beta^{\mathcal{D}}$ writes, in terms of the correction $\tilde{\beta}^{\mathcal{D}}$, for $u \in \mathcal{H}_{\mathcal{M}}$ and $K \in \mathcal{M}$,

$$\mathcal{S}_K(u) = -\mathcal{A}_K(u) + \sum_{Z \in V(K)} \tilde{\beta}_{K,Z}(u)(u_K - u_Z) - \nu_K \Delta_K(u). \tag{2.14}$$

Then, provided that $\tilde{\beta}^{\mathcal{D}}$ has been chosen so that the corresponding corrected scheme is still consistent with the continuous operator A, we see that $\mathcal{S}_K(u)$ formally approximate the integral on K of $-A(u) - \nu_K \Delta u$. Taking ν_K to be of order size($\mathcal{D}$), the effect of the last term can be expected[2] to vanish as size($\mathcal{D}$) $\to 0$. For instance, if $\nu_K = \mathrm{diam}(K)$, the correction turns to

$$\beta_{K,Z}(u) = \tilde{\beta}_{K,Z}(u) + |K|Z|.$$

Remark 2.6. Conditions (2.10) ensures that the terms $\beta_{K,Z}$ are large enough to compensate the discrete maximum principle weakening contributions of $-\mathcal{A}^{\mathcal{D}}$, namely the coefficients in the right-hand side sum in (2.11) which correspond to elements $Z \in V(K)$ such that $\mathcal{A}_K(u)(u_Z - u_K) < 0$. Actually this condition entails that the compensation does not only happen on these weakening contribution but on all contributions. Thus the results remains true if we compensate only the weakening contributions. More precisely, setting

$$V(K,u)^+ = \{Z \in V(K) \ ; \ \mathcal{A}_K(u)(u_Z - u_K) > 0\}$$

and

$$V(K,u)^- = \{Z \in V(K) \ ; \ \mathcal{A}_K(u)(u_Z - u_K) < 0\},$$

we can take $\beta_{K,Z}(u) = 0$ if $Z \in V(K,u)^+$ and change (2.10) into

$$\begin{aligned}
&\forall Z \in V(K,u)^-, \quad \beta_{K,Z}(u) \geq \gamma_{K,Z}(u)|\mathcal{A}_K(u)|, \\
&\forall L \in \mathcal{N}_K \cap V(K,u)^-, \quad \beta_{K,L}(u) > \gamma_{K,L}(u)|\mathcal{A}_K(u)|, \\
&\forall \sigma \in \mathcal{E}_{\mathrm{ext}} \cap V(K,u)^-, \quad \beta_{K,\sigma}(u) > \gamma_{K,\sigma}(u)|\mathcal{A}_K(u)|.
\end{aligned}$$

One first drawback of this choice is that it could lead to consider corrections that are not continuous functions of $u \in \mathcal{H}_{\mathcal{M}}$, which is due to the fact the partition $V(K) = V(K,u)^+ \cup V(K,u)^-$ depends on u. Moreover, this would also break the symmetry of the correction that plays an important role as will be shown in Sections 2.2.2.2 and 2.2.2.3

There are various ways to choose functions $\gamma_{K,Z}$ satisfying condition (2.9):

i) Taking, for $K \in \mathcal{M}$ and $Z \in V(K)$,

$$\gamma_{K,Z}(u) = \frac{1}{\sum_{Y \in V(K)} |u_K - u_Y|} \tag{2.15}$$

[2] This expectation is rigorously demonstrated in Remark 2.13.

if $\sum_{Y\in V(K)}|u_K-u_Y|\neq 0$ and $\gamma_{K,Z}(u)=0$ else, condition (2.10a) writes

$$\beta_{K,Z}(u)\geq \frac{|\mathcal{A}_K(u)|}{\sum_{Y\in V(K)}|u_K-u_Y|}. \tag{2.16}$$

ii) For $u\in\mathcal{H}_\mathcal{M}$, let us define $V(K,u)^*=\{Z\in V(K)\ ;\ u_Z-u_K\neq 0\}$. Taking, for $K\in\mathcal{M}$ and $Z\in V(K)$,

$$\gamma_{K,Z}(u)=\frac{1}{\mathrm{Card}V(K,u)^*|u_Z-u_K|} \tag{2.17}$$

if $u_Z-u_K\neq 0$ and $\gamma_{K,Z}(u)=0$ else, condition (2.10a) writes

$$\beta_{K,Z}(u)\geq \frac{|\mathcal{A}_K(u)|}{\mathrm{Card}V(K,u)^*|u_Z-u_K|}. \tag{2.18}$$

2.2.2.2 Conservation preserving corrections

Even if the original scheme is a Finite Volume scheme in the sense that it matches assumption *(A1)*, this is not automatically the case of the corrected scheme. However a simple symmetry assumption on the correction ensures that the conservative structure is preserved.

The statement of this condition needs to introduce polygonal paths in the mesh as in [91]. Given an admissible mesh $\mathcal{D}$ of Ω we fix, for any pair $(I,J)\in\mathcal{M}^2$ such that $I\in V(J)$ (or equivalently $J\in V(I)$) a polygonal path IJ that does not include any edge or vertex of the mesh and that crosses any edge at most one time. Then, assuming the control volumes are sorted out, we denote by $\mathcal{C}$ the set $\mathcal{C}=\{IJ\ ;\ I\leq J\}$ and we let, for any edge $\sigma\in\mathcal{E}$, $\mathrm{ch}(\sigma)$ be the set of the polygonal paths IJ with $I\leq J$ and such that IJ crosses σ (see Figure 2.1). Finally,

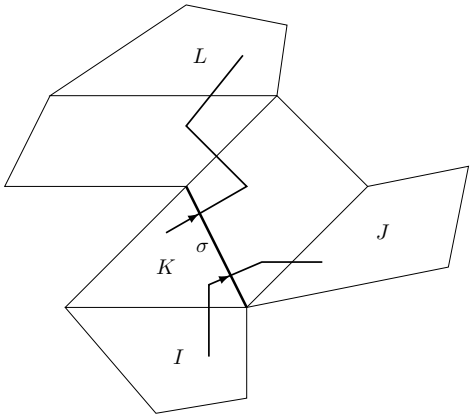

Figure 2.1 – Paths IJ and KL in $\mathrm{ch}(\sigma)$, $J\in V(I)$, $L\in V(K)$.

given a path $IJ\in\mathrm{ch}(\sigma)$ with $\sigma\in\mathcal{E}_K$, we set $\varepsilon_{K,\sigma,IJ}=1$ if, from I to J, the path IJ enters the cell K through σ and $\varepsilon_{K,\sigma,IJ}=-1$ if it leaves K through σ.

Proposition 2.7 (Conservative corrections)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω and $\beta^\mathcal{D}=(\beta_{K,Z})_{K\in\mathcal{M},Z\in V(K)}$ be a correction for* (2.4)*. Assume the family $\beta^\mathcal{D}$ is symmetric:*

$$\forall K\in\mathcal{M},\forall L\in V(K)\cap\mathcal{M},\quad \beta_{K,L}=\beta_{L,K}. \tag{2.19}$$

If the original scheme is conservative, then so is the corrected one, with numerical fluxes $F'_{K,\sigma}$ given, for all $u \in \mathcal{H}_\mathcal{M}$ and all $K \in \mathcal{M}$, by

$$\forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\rm int}, \quad F'_{K,\sigma}(u) = F_{K,\sigma}(u) - \sum_{IJ \in {\rm ch}(\sigma)} \varepsilon_{K,\sigma,IJ} \beta_{I,J}(u)(u_J - u_I) \tag{2.20a}$$

$$\forall \sigma \in \mathcal{E}_K \cap \mathcal{E}_{\rm ext}, \quad F'_{K,\sigma}(u) = F_{K,\sigma}(u) - \beta_{K,\sigma}(u) u_K \tag{2.20b}$$

Remark 2.8. In case the correction $\beta^\mathcal{D}$ is symmetric (in the sense of (2.19)) the previous proposition states that correcting the original scheme with $\beta^\mathcal{D}$ amounts to correct the original fluxes $F_{K,\sigma}$ with the corrective fluxes $R_{K,\sigma}$ defined, for all $u \in \mathcal{H}_\mathcal{M}$, all $K \in \mathcal{M}$ and all interior edge $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\rm int}$ by

$$R_{K,\sigma}(u) = - \sum_{IJ \in {\rm ch}(\sigma)} \varepsilon_{K,\sigma,IJ} \beta_{I,J}(u)(u_J - u_I), \tag{2.21}$$

and for all boundary edge $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\rm ext}$ by

$$R_{K,\sigma}(u) = -\beta_{K,\sigma}(u) u_K. \tag{2.22}$$

Proof of Proposition 2.7. We proceed as in the proof of Proposition 4.1 from [91]. Let us first remark that the corrective fluxes defined by (2.21) satisfy the conservativity condition (1.7) (this follows from the fact that, by definition, the quantity $\varepsilon_{K,\sigma,IJ}$ itself is conservative). Consequently the fluxes $F'_{K,\sigma}$ also satisfy this condition.

It remains to check that the corrective term $\mathcal{R}_K$ in (2.7) matches with the balance $-\sum_{\sigma \in \mathcal{E}_K} R_{K,\sigma}$ of the corrective fluxes. On that account note that, for $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$, if $IJ \in {\rm ch}(\sigma)$ is such that $K \notin \{I, J\}$ (*i.e.* the path crosses the cell K) and if IJ enters (resp. leaves) K across σ, then there exists $\sigma' \in \mathcal{E}_K$ such that IJ leaves (resp. enters) K across, this means $\varepsilon_{K,\sigma,IJ} = -\varepsilon_{K,\sigma',IJ}$. Thus, in the sum below, the terms corresponding to σ and σ' cancel so that we can state:

$$\forall u \in \mathcal{H}_\mathcal{M}, \forall K \in \mathcal{M}, \quad \sum_{\sigma \in \mathcal{E}_K} \sum_{\substack{IJ \in {\rm ch}(\sigma) \\ K \notin \{I,J\}}} \varepsilon_{K,\sigma,IJ} \beta_{I,J}(u)(u_J - u_I) = 0.$$

Consequently, for any $u \in \mathcal{H}_\mathcal{M}$ and any $K \in \mathcal{M}$, the balance reduces to

$$- \sum_{\sigma \in \mathcal{E}_K \cap \mathcal{E}_{int}} R_{K,\sigma}(u) = \sum_{\sigma \in \mathcal{E}_K} \sum_{\substack{IJ \in {\rm ch}(\sigma) \\ K \in \{I,J\}}} \varepsilon_{K,\sigma,IJ} \beta_{I,J}(u)(u_J - u_I)$$

which writes, in view of the definition of ${\rm ch}(\sigma)$ and $\varepsilon_{K,\sigma,IJ}$,

$$- \sum_{\sigma \in \mathcal{E}_K \cap \mathcal{E}_{int}} R_{K,\sigma}(u) = \sum_{L \in V(K) \cap \mathcal{M}} \beta_{K,L}(u)(u_K - u_L)$$

and then

$$- \sum_{\sigma \in \mathcal{E}_K} R_{K,\sigma}(u) = \sum_{Z \in V(K)} \beta_{K,Z}(u)(u_K - u_Z) = \mathcal{R}_K(u).$$

□

2.2.2.3 Coercivity preserving corrections

If the correction is symmetric (in the sense of Proposition 2.7) it further suffices for the corrective functions to be non-negative to preserve the coercivity of the original scheme.

Proposition 2.9 (Coercivity preserving corrections)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω and $\beta^{\mathcal{D}} = (\beta_{K,Z})_{K\in\mathcal{M},Z\in V(K)}$ be a symmetric correction for* (2.4)*. Assume the family $\beta^{\mathcal{D}}$ is non-negative:*

$$\forall K \in \mathcal{M}, \forall Z \in V(K), \quad \beta_{K,Z} \geq 0. \tag{2.23}$$

If the original scheme is coercive, then so is the corrected one, with the same constant.

Proof. Let $u \in \mathcal{H}_{\mathcal{M}}$. Assume the original scheme is coercive with constant ζ. Then

$$\sum_{K\in\mathcal{M}} \mathcal{S}_K(u)u_K \geq \zeta \|u\|_{\mathcal{D}}^2 + \sum_{K\in\mathcal{M}} u_K \sum_{Z\in V(K)} \beta_{K,Z}(u)(u_K - u_Z).$$

Let us denote by $\mathcal{T}$ the last term of the inequality and remark that provided $\mathcal{T} \geq 0$, the coercivity of the original scheme is preserved. Now gathering by polygonal paths and using symmetry assumption (2.19) on $\beta^{\mathcal{D}}$ and assumption (2.6) on the stencil yields

$$\mathcal{T} = \sum_{IJ\in\mathcal{C}} \beta_{I,J}(u)(u_I - u_J)^2 + \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{ext}} \beta_{K,\sigma}(u)u_K^2$$

which proves, with (2.23), that $\mathcal{T} \geq 0$. □

Provided coefficients $\mathcal{R}_K$ of the corrective term are continuous functions of the unknown u, coercivity assumption also guaranties that there exists at least one solution to the corrected scheme.

Proposition 2.10 (Existence of a solution)**.** *Let $\mathcal{D}$ be an admissible mesh of Ω and let $\beta^{\mathcal{D}}$ be a correction for* (2.4) *satisfying* (2.19) *and* (2.23)*. Assume that the original scheme is coercive and that the corrective term $\mathcal{R}^{\mathcal{D}} : \mathcal{H}_{\mathcal{M}} \to \mathcal{H}_{\mathcal{M}}$ is continuous. Then there exists one solution to the corrected scheme.*

Proof. The proof relies on Brower's topological degree. According to the hypothesis made on $\mathcal{R}^{\mathcal{D}}$, the application $h_t = -\mathcal{A}^{\mathcal{D}} + t\mathcal{R}^{\mathcal{D}}$ is continuous for all $t \in [0,1]$. Then it is sufficient to show that, for R large enough, any solution to $h_t(u) = f_{\mathcal{D}}$ is bounded by R in $\mathcal{H}_{\mathcal{M}}$ to ensure that the degree of $h_1 = \mathcal{S}^{\mathcal{D}}$ on the ball of radius R at the point $f_{\mathcal{D}}$ is the same as the degree of $h_0 = -\mathcal{A}^{\mathcal{D}}$ which is not zero (since $\mathcal{A}^{\mathcal{D}}$ is invertible), and consequently to prove the existence of one solution to the corrected scheme $\mathcal{S}^{\mathcal{D}}(u) = f_{\mathcal{D}}$. The expected *a priori* estimate on the solution to $h_t(u) = f_{\mathcal{D}}$ is based on the coercivity of $-\mathcal{A}^{\mathcal{D}}$ and $\mathcal{S}^{\mathcal{D}}$. Indeed noting that $h_t = -(1-t)\mathcal{A}^{\mathcal{D}} + t\mathcal{S}^{\mathcal{D}}$ and denoting by ζ the coercivity constant of $-\mathcal{A}^{\mathcal{D}}$, Proposition 2.9 guarantees that the scheme defined by h_t is coercive with constant ζ. From Proposition 1.7 and the discrete Poincaré inequality (1.5) we deduce that any solution to $h_t(u) = f_{\mathcal{D}}$ is bounded in L^2 norm by $R = C\|f\|_{L^2(\Omega)}$ where the constant C does not depend on t or u. □

2.2.2.4 How to build monotone, conservative and coercive corrections

Assume the original scheme to be conservative and coercive. A simple way to construct corrections that match all the previous conditions ensuring the corrected scheme has the LMP structure and is still conservative and coercive is to take the following steps:

1. Choose a family $\gamma^{\mathcal{D}}$ such that (2.9) holds (for instance take $\gamma^{\mathcal{D}}$ as in (2.15) or (2.17));

2. Define the correction $b^{\mathcal{D}}$ by

$$\forall K \in \mathcal{M}, \forall Z \in V(K), \quad b_{K,Z} = \gamma_{K,Z} |\mathcal{A}_K|. \tag{2.24}$$

   This correction matches condition (2.10a)

3. a) For $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K \cap \mathcal{E}_{\text{ext}}$, define $\tilde{\beta}_{K,\sigma} = b_{K,\sigma}$,

   b) For $(K,L) \in \mathcal{M}^2$ such that $L \in V(K)$, define $\tilde{\beta}_{K,L}$ as a symmetric combination of $b_{K,L}$ and $b_{L,K}$ such that $\tilde{\beta}_{K,L} \geq b_{K,L}$. For instance one can take $\tilde{\beta}_{K,L} = b_{K,L} + b_{L,K}$ or $\tilde{\beta}_{K,L} = \max(b_{K,L}, b_{L,K})$.

   The correction $\tilde{\beta}^{\mathcal{D}}$ is thus symmetric, non-negative and satisfies condition (2.10a).

4. Augment $\tilde{\beta}^{\mathcal{D}}$ to match conditions (2.10b) and (2.10c): for instance define (see remark 2.5) $\beta^{\mathcal{D}}$ by

$$\forall K \in \mathcal{M}, \forall Z \in V(K), \quad \beta_{K,Z} = \tilde{\beta}_{K,Z} + |K|Z|.$$

The correction $\beta^{\mathcal{D}} = (\beta_{K,Z})_{K\in\mathcal{M}, Z\in V(K)}$ we obtain from these guidelines is thus symmetric, non-negative, hence it yields a scheme having the LMP structure.

As an example let us consider the following correction $\beta^{\mathcal{D}}$, similar to the non-linear correction proposed in [91], and defined, for all $u \in \mathcal{H}_{\mathcal{M}}$, all $K \in \mathcal{M}$ and all $Z \in V(K)$, by:

- If $Z = \sigma \in \mathcal{E}_{\text{ext}}$, then

$$\beta_{K,\sigma}(u) = \frac{|\mathcal{A}_K(u)|}{\sum_{Y\in V(K)} |u_Y - u_K|} + |\sigma|. \tag{2.25}$$

- If $Z = L \in \mathcal{M}$, then

$$\beta_{K,L}(u) = \frac{|\mathcal{A}_K(u)|}{\sum_{Y\in V(K)} |u_Y - u_K|} + \frac{|\mathcal{A}_L(u)|}{\sum_{Y\in V(L)} |u_Y - u_L|} + |K|L|. \tag{2.26}$$

If one of the quantities $\sum_{Y\in V(K)} |u_Y - u_K|$ or $\sum_{Y\in V(L)} |u_Y - u_L|$ is zero, we define $\beta_{K,Z}(u)$ in that case by dropping the corresponding term in (2.25) or (2.26). Note that each function $\beta_{K,Z} : \mathcal{H}_{\mathcal{M}} \to \mathbb{R}$ is continuous outside the set $\{u \in \mathcal{H}_{\mathcal{M}} \ ; \ u_K - u_Z \neq 0\}$ and bounded on $\mathcal{H}_{\mathcal{M}}$ according to assumption (2.5) on the structure of the original scheme. Hence the corrective term $\mathcal{R}^{\mathcal{D}} : \mathcal{H}_{\mathcal{M}} \to \mathcal{H}_{\mathcal{M}}$ defined through (2.8) is continuous so that Proposition 2.10 guarantees the corresponding corrected scheme $\mathcal{S}^{\mathcal{D}}(u) = f_{\mathcal{D}}$ has at least one solution.

It has been proved in [91] that this correction gives a conservative and coercive scheme which has the LMP structure. This can also be shown by verifying this correction can be built following the guidelines 1–4 above. First, we consider the family $\gamma^{\mathcal{D}}$ given by (2.15) and then define, according to (2.24), correction $b^{\mathcal{D}}$ by:

$$\forall u \in \mathcal{H}_{\mathcal{M}}, \forall K \in \mathcal{M}, \forall Z \in V(K), \quad b_{K,Z}(u) = \frac{|\mathcal{A}_K(u)|}{\sum_{Y\in V(K)} |u_Y - u_K|}.$$

We then follow steps 2 and 3 taking $\tilde{\beta}_{K,L} = b_{K,L} + b_{L,K}$ in 3b and we augment $\tilde{\beta}^{\mathcal{D}}$ according to step 4. Equation (2.26) finally writes $\beta_{K,L} = \tilde{\beta}_{K,L} + |K|L|$.

Starting from a different choice for the family $\gamma^{\mathcal{D}}$, namely the one previously defined by (2.17), the steps 1–4 can lead to the correction defined, for all $u \in \mathcal{H}_{\mathcal{M}}$, all $K \in \mathcal{M}$ and all $Z \in V(K,u)^*$, by

$$\beta_{K,Z}(u) = \left\{ \max\left(\frac{|\mathcal{A}_K(u)|}{\mathrm{Card}\, V(K,u)^*}, \frac{|\mathcal{A}_Z(u)|}{\mathrm{Card}\, V(Z,u)^*} \right) + \sum_{\sigma \in K|Z} |\sigma|\, d_\sigma \right\} \frac{1}{|u_K - u_Z|} \tag{2.27}$$

where we set $\frac{|\mathcal{A}_Z(u)|}{\mathrm{Card}\, V(Z,u)^*} = 0$ if $Z = \sigma \in \mathcal{E}_{\mathrm{ext}}$. The corresponding conservative and coercive corrected scheme $\mathcal{S}^{\mathcal{D}}$ which has the LMP structure writes, for all $u \in \mathcal{H}_{\mathcal{M}}$ and all $K \in \mathcal{M}$,

$$\begin{aligned} \mathcal{S}_K(u) = &-\mathcal{A}_K(u) \\ &+ \sum_{Z \in V(K,u)^*} \left\{ \max\left(\frac{|\mathcal{A}_K(u)|}{\mathrm{Card}\, V(K,u)^*}, \frac{|\mathcal{A}_Z(u)|}{\mathrm{Card}\, V(Z,u)^*} \right) + \sum_{\sigma \in K|Z} |\sigma|\, d_\sigma \right\} \mathrm{sgn}(u_K - u_Z). \end{aligned} \tag{2.28}$$

Note that the use of the terms $\mathrm{sgn}(u_K - u_Z)$ in this last correction is reminiscent of the form of the non-linear stabilization term proposed in [25] to design a Galerkin approximation of the Laplacian operator guaranteeing a discrete maximum principle on arbitrary meshes. The main drawback of the scheme (2.28) is that the corrective term is not continuous so that the existence of solutions to the non-linear system $\mathcal{S}^{\mathcal{D}}(u) = f_{\mathcal{D}}$ is not ensured. To obtain continuity we present in section 2.3.2 a regularized version of this scheme.

2.2.3 Convergence preserving corrections

Let us consider a coercive and consistent original scheme in the sense of assumptions *(A2)-(A3)*. From section 2.2.2, we know how to correct it in order to obtain a scheme which has the LMP structure and is still coercive. This last property ensures that the solution of such a corrected scheme still converges, up to a subsequence, to a function $\bar{u} \in H_0^1(\Omega)$. Moreover, from the consistency of the original scheme with problem (2.1), the behavior of the original part of the corrected scheme is known. Therefore, a simple way to prove that the limit $\bar{u}$ is a weak solution to the problem (2.1) is to make sure the corrective term vanishes as the size of the mesh tends to 0.

In addition to the geometrical regularity of the mesh, measured by the quantity $\mathrm{regul}(\mathcal{D})$, we want to take into account its compatibility with the original discretized operator $\mathcal{A}^{\mathcal{D}}$. To this end we first define the sets $\tilde{V}(K)$ by adding to $V(K)$ all the cells crossed by some polygonal path coming from K *i.e.* of the form KL ($L \in V(K)$). The sets $\tilde{V}(K)$ are then completed so that they are still symmetric that is:

$$\forall (K,L) \in \mathcal{M}^2, \quad L \in \tilde{V}(K) \implies K \in \tilde{V}(L).$$

Then we define the following quantity

$$\mathrm{reg}_{\mathcal{A}}(\mathcal{D}) = \mathrm{regul}(\mathcal{D}) + \max_{K \in \mathcal{M}, L \in \tilde{V}(K)} \frac{\mathrm{diam}(L)}{\mathrm{diam}(K)} + \max_{K \in \mathcal{M}} \mathrm{Card}(\tilde{V}(K)).$$

Proposition 2.11 (Convergence of the corrected scheme)**.** *Let $(\mathcal{D}^n)_{n\geq 1}$ be a sequence of admissible meshes of Ω such that,* $\mathrm{size}(\mathcal{D}^n) \to 0$ *as* $n \to \infty$ *and* $(\mathrm{reg}_{\mathcal{A}}(\mathcal{D}^n))_{n\geq 1}$ *is bounded. Let $(\beta^n)_{n\geq 1}$ be a family of corrections associated with $(\mathcal{D}^n)_{n\geq 1}$ such that for all $n \geq 1$, β^n is symmetric and non-negative. For $n \geq 1$ we denote by $\mathcal{S}^n$ the corresponding corrected scheme.*

Assume that a family $(u^n)_{n\geq 1}$ satisfies:

- *For all $n \geq 1$, $u^n \in \mathcal{H}_\mathcal{M}$ is a solution to $\mathcal{S}^n$;*
- *As $n \to \infty$,*

$$\sum_{K\in\mathcal{M}^n} \mathrm{diam}(K) \sum_{Z\in V(K)} \beta^n_{K,Z}(u^n)\left|u^n_K - u^n_Z\right| \to 0. \tag{2.29}$$

Then, as $n \to \infty$, u^n converges in $L^2(\Omega)$ to the unique solution of (2.1).

Remark 2.12. In the case where $V(K) \cap \mathcal{M}$ reduces to the neighboring cells $\mathcal{N}_K$ and where the paths in $\mathcal{C}$ cross only one edge, the family of corrective fluxes $R = (R_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$ defined through (2.21) and (2.22) simply writes, for $\sigma \in K|L$,

$$R_{K,\sigma}(u) = \beta_{K,\sigma}(u)(u_L - u_K).$$

Let us define, for a family of fluxes $F = (F_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$ and for a finite $p \geq 1$, discrete norms $N_{p,\mathcal{D}}(F)$ by

$$N_{p,\mathcal{D}}(F)^p = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} |\sigma| \mathrm{diam}(K) \left|\frac{F_{K,\sigma}}{|\sigma|}\right|^p.$$

We also define

$$N_{\infty,\mathcal{D}}(F) = \max_{K\in\mathcal{M},\sigma\in\mathcal{E}_K} \left|\frac{F_{K,\sigma}}{|\sigma|}\right|.$$

Then condition (2.29) reads

$$N_{1,\mathcal{D}^n}(R(u^n)) \to 0 \text{ as } n \to \infty. \tag{2.30}$$

Notice that as a consequence of Hölder inequality, the following bound holds for any family of fluxes F and any $p \in [1, \infty]$:

$$N_{1,\mathcal{D}}(F) \leq (d\,|\Omega|\mathrm{regul}(\mathcal{D}))^{1-\frac{1}{p}} N_{p,\mathcal{D}}(F). \tag{2.31}$$

Thus, as $(\mathrm{regul}(\mathcal{D}^n))_{n\geq 1}$ is bounded, condition (2.29) holds if, for any $p > 1$, $N_{p,\mathcal{D}^n}(R(u^n)) \to 0$ as $n \to \infty$.

Remark 2.13. When choosing the augmentation in step 4 of the construction from section 2.2.2.4, one has to make sure this augmentation vanishes as $\mathrm{size}(\mathcal{D}) \to 0$ if not at the risk of jeopardizing the consistency of the scheme. In case this augmentation is conservative one may ensure that condition (2.30) holds. Note that this is the case for the above two corrections:

- Concerning (2.25)-(2.26) we know from Remark 2.5 that the additional numerical diffusion term writes, for all $u \in \mathcal{H}_\mathcal{M}$ and all $K \in \mathcal{M}$,

$$\mathrm{diam}(K)\Delta_K(u) = \sum_{\sigma\in\mathcal{E}_K} r_{K,\sigma}(u),$$

with $r_{K,\sigma}(u) = |\sigma|(u_Z - u_K)$. Now, remark that if $\theta \geq \mathrm{regul}(\mathcal{D})$ then we have

$$\forall u \in \mathcal{H}_\mathcal{M}, \quad N_{2,\mathcal{D}}(r(u)) \leq C_1\,\mathrm{size}(\mathcal{D})\,\|u\|_\mathcal{D},$$

with some $C_1 \in \mathbb{R}_+$ only depending on θ. Provided both $(\mathrm{regul}(\mathcal{D}^n))_{n\geq 1}$ and $(\|u^n\|_{\mathcal{D}^n})_{n\geq 1}$ are bounded, this entails that $N_{2,\mathcal{D}^n}(r(u^n)) \to 0$ as $n \to \infty$.

Replacing $|K|L|$ in (2.26) by the smaller quantity (used in section 2.3.1)

$$\min\left(|K|L|, \frac{|K|}{\sum_{Y\in V(K)}|u_Y - u_K|} + \frac{|L|}{\sum_{Y\in V(L)}|u_Y - u_L|}\right)$$

and denoting by $\tilde{r}_{K,\sigma}$ the corresponding fluxes, we have

$$N_{\infty,\mathcal{D}}(\tilde{r}) \leq 2 \max_{K\in\mathcal{M},\sigma\in\mathcal{E}_K} \frac{|K|}{|\sigma|}.$$

Assuming some reasonable regularity assumptions on the mesh, we can see that this last quantity scales as size($\mathcal{D}$) so that this augmentation vanishes more strongly than the previous one.

- The augmentation chosen in (2.27) is conservative with fluxes $\rho_{K,\sigma}$ defined by $\rho_{K,\sigma}(u) = |\sigma| d_\sigma \mathrm{sgn}(u_K - u_Z)$. In that case, (2.30) follows from the following estimate:

$$\forall u \in \mathcal{H}_\mathcal{M}, \quad N_{\infty,\mathcal{D}}(\rho(u)) \leq 2\,\mathrm{size}(\mathcal{D}).$$

Proof. We proceed as mentioned above: we use first coercivity to extract a convergent subsequence of $(u^n)_{n\geq 1}$, then the consistency of the original scheme together with assumption (2.29) allow to pass to the limit in the corrected scheme.

Given $n \geq 1$, Proposition 2.9 shows that $\mathcal{S}^n$ is coercive with constant ζ and thus that the *a priori* estimate (1.10) holds for u^n. Since $(\mathrm{regul}(\mathcal{D}^n))_{n\geq 1}$ is bounded and since ζ does not depend on n, this estimate proves that the sequence $(\|u^n\|_{\mathcal{D}^n})_{n\geq 1}$ is bounded. Thus, according to the discrete compactness results for bounded families in the discrete H_0^1 norm (see [65] lemmas 5.6 and 5.7 with $p = 2$), there exists $\bar{u} \in H_0^1(\Omega)$ such that, up to a subsequence, $u^n \to \bar{u}$ in $L^2(\Omega)$. Since (2.1) has a unique solution, if we prove that $\bar{u}$ is indeed this solution, then we get that the whole family $(u^n)_{n\geq 1}$ converges to $\bar{u}$ as $n \to \infty$.

To simplify the notations, we drop the index n and assume that $u = u^n$ converges to $\bar{u}$ as size($\mathcal{D}$) $\to 0$. Given $\varphi \in \mathcal{C}_c^\infty(\Omega)$ we set $\varphi_\mathcal{D} = (\varphi_K)_{K\in\mathcal{M}} \in \mathcal{H}_\mathcal{M}$ with $\varphi_K = \varphi(x_K)$. Multiplying the equation on K (2.7) by φ_K and summing over $K \in \mathcal{M}$ we get

$$-\sum_{K\in\mathcal{M}} \mathcal{A}_K(u)\varphi_K + \sum_{K\in\mathcal{M}} \mathcal{R}_K(u)\varphi_K = \int_\Omega f\varphi_\mathcal{D}. \tag{2.32}$$

The right-hand side tends to $\int_\Omega f\varphi$ as size($\mathcal{D}$) $\to 0$. Besides, since $\mathrm{reg}_\mathcal{A}(\mathcal{D})$ is bounded, assumption *(A3)* on the consistency of the original scheme ensures that, along the extracted subfamily, we have

$$-\sum_{K\in\mathcal{M}} \mathcal{A}_K(u)\varphi_K \to \int_\Omega D\nabla\bar{u}\nabla\varphi,$$

as size($\mathcal{D}$) $\to 0$.

Let us prove the corrected term in the left-hand side of (2.32) vanishes as size($\mathcal{D}$) $\to 0$. Gathering by polygonal paths, we can write

$$\sum_{K\in\mathcal{M}} \mathcal{R}_K(u)\varphi_K = \sum_{IJ\in\mathcal{C}} \beta_{I,J}(u)(u_I - u_J)(\varphi_I - \varphi_J) + \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{ext}} \beta_{K,\sigma}(u)u_K\varphi_K.$$

Hence

$$\left|\sum_{K\in\mathcal{M}} \mathcal{R}_K(u)\varphi_K\right| \leq \sum_{K\in\mathcal{M}} \sum_{Z\in V(K)} \beta_{K,Z}(u)|u_K - u_Z||\varphi_K - \varphi_Z|. \tag{2.33}$$

Now note that since φ is regular, compactly supported in Ω, and since $\mathrm{reg}_{\mathcal{A}}(\mathcal{D})$ is bounded, there exists C_2 not depending on $\mathcal{D}$ such that

$$|\varphi_K - \varphi_Z| \leq C_2 \operatorname{diam}(K)$$

for all $K \in \mathcal{M}$ and all $Z \in V(K)$. Using this last inequality in (2.33) proves, according to (2.29), that

$$\sum_{K\in\mathcal{M}} \mathcal{R}_K(u)\varphi_K \to 0$$

as $\mathrm{size}(\mathcal{D})$ goes to 0.

Sending $\mathrm{size}(\mathcal{D}) \to 0$ in (2.32) (along the extracted subfamily) we finally get, for any $\varphi \in \mathcal{C}_c^\infty(\Omega)$,

$$\int_\Omega D\nabla\bar{u}\nabla\varphi = \int_\Omega f\varphi,$$

which proves, as announced, that $\bar{u}$ is the weak solution to (2.1). □

2.3 Examples of corrections

In this section, we assume that the original scheme is coercive and consistent in the sense of *(A2)-(A3)*. Using the tools from the previous section we study two actual examples of corrections. In both cases, we provide a numerical condition under which the convergence of the scheme is ensured.

2.3.1 A first correction

Given some parameter $\eta > 0$, we consider first the following correction $\beta^{\mathcal{D}}$ defined, for all $u \in \mathcal{H}_{\mathcal{M}}$, all $K \in \mathcal{M}$ and all $Z \in V(K)$, by:

- If $Z = \sigma \in \mathcal{E}_{\mathrm{ext}}$, then

$$\beta_{K,\sigma}(u) = \frac{|\mathcal{A}_K(u)|}{\sum_{Y\in V(K)}|u_Y - u_K|} + \eta \min\left(|\sigma|, \frac{|K|}{\sum_{Y\in V(K)}|u_K - u_Y|}\right). \tag{2.34}$$

- If $Z = L \in \mathcal{M}$, then

$$\beta_{K,L}(u) = \frac{|\mathcal{A}_K(u)|}{\sum_{Y\in V(K)}|u_Y - u_K|} + \frac{|\mathcal{A}_L(u)|}{\sum_{Y\in V(L)}|u_Y - u_L|} + \eta \min\left(|K|L|, \frac{|K|}{\sum_{Y\in V(K)}|u_Y - u_K|} + \frac{|L|}{\sum_{Y\in V(L)}|u_Y - u_L|}\right). \tag{2.35}$$

This correction is slightly different from the one previously defined by (2.25)-(2.26). More precisely the difference lies in the last term that is in the augmentation chosen in step 4 of the guidelines from section 2.2.2.4. The modified augmentation chosen above still brings the LMP structure and takes better care of the convergence of the scheme since the stabilization term is smaller (see Remark 2.13).

Proposition 2.14. *Let $\eta > 0$ and let $(\mathcal{D}^n)_{n\geq 1}$ be a sequence of admissible meshes of Ω such that $\mathrm{size}(\mathcal{D}^n) \to 0$ as $n \to \infty$ and $(\mathrm{reg}_{\mathcal{A}}(\mathcal{D}^n))_{n\geq 1}$ is bounded. For all $n \geq 1$ we denote by $\mathcal{S}^n : \mathcal{H}_{\mathcal{M}^n} \to \mathcal{H}_{\mathcal{M}^n}$ the corrected scheme defined through (2.34)–(2.35). Let $(u^n)_{n\geq 1}$ be a sequence of discrete functions satisfying:*

- *For all $n \geq 1$, $u^n \in \mathcal{H}_{\mathcal{M}^n}$ is a solution to $\mathcal{S}^n$;*
- *As $n \to \infty$,*

$$\sup_{K \in \mathcal{M}^n} \left\{ \left| \mathcal{A}_K^{\mathcal{D}^n}(u^n) \right| \frac{\operatorname{diam}(K)}{|K|} \right\} \to 0. \tag{2.36}$$

Then, as $n \to \infty$, u^n converges in $L^2(\Omega)$ to the unique solution of (2.1).

Proof. We show that the family of solutions $(u^n)_{n \geq 1}$ matches condition (2.29). For simplicity, we drop the index n. For all $K \in \mathcal{M}$ and all $Z \in V(K)$ we have

$$\beta_{K,Z}(u)|u_K - u_Z| \leq |\mathcal{A}_K(u)| + |\mathcal{A}_Z(u)| + \eta\,|K|Z||u_K - u_Z|.$$

Thus, $\operatorname{reg}_{\mathcal{A}}(\mathcal{D})$ being bounded, there exists C_3 independent of $\mathcal{D}$ such that

$$\sum_{K \in \mathcal{M}} \operatorname{diam}(K) \sum_{Z \in V(K)} \beta_{K,Z}(u)|u_K - u_Z| \leq C_3 \sum_{K \in \mathcal{M}} \operatorname{diam}(K)|\mathcal{A}_K(u)| + \eta N_{1,\mathcal{D}}(r(u)), \tag{2.37}$$

with $N_{1,\mathcal{D}}(r(u)) = \sum_{K \in \mathcal{M}} \operatorname{diam}(K) \sum_{\sigma \in \mathcal{E}_K} |\sigma||u_K - u_L|$. Remark 2.13 together with inequality (2.31) yield

$$N_{1,\mathcal{D}}(r(u)) \xrightarrow[\operatorname{size}(\mathcal{D}) \to 0]{} 0. \tag{2.38}$$

Besides, the first term of the right hand side in (2.37) can be bounded above as follows

$$\sum_{K \in \mathcal{M}} \operatorname{diam}(K)|\mathcal{A}_K(u)| \leq |\Omega| \sup_{K \in \mathcal{M}} \left\{ |\mathcal{A}_K(u)| \frac{\operatorname{diam}(K)}{|K|} \right\},$$

which, thanks to (2.36), implies

$$\sum_{K \in \mathcal{M}} \operatorname{diam}(K)|\mathcal{A}_K(u)| \xrightarrow[\operatorname{size}(\mathcal{D}) \to 0]{} 0. \tag{2.39}$$

Substituting estimates (2.38) and (2.39) into (2.37) proves that, as $\operatorname{size}(\mathcal{D}) \to 0$,

$$\sum_{K \in \mathcal{M}} \operatorname{diam}(K) \sum_{Z \in V(K)} \beta_{K,Z}(u)|u_K - u_Z| \to 0,$$

which, according to Proposition 2.11, gives the desired result. □

2.3.2 A regularized correction

As we pointed out above, the main drawback of the correction defined by (2.27) is that the resulting scheme is not a continuous function of $u \in \mathcal{H}_{\mathcal{M}}$. Actually, discontinuity mainly comes from the family $\gamma^{\mathcal{D}}$ given by (2.17) which has been used to build the correction following the steps 1–4 from section 2.2.2.4. Given a positive parameter ε, let us replace $\gamma^{\mathcal{D}}$ by a smoothed family γ^{ε} which writes, for $u \in \mathcal{H}_{\mathcal{M}}$, $K \in \mathcal{M}$ and $Z \in V(K)$,

$$\gamma_{K,Z}^{\varepsilon}(u) = \frac{1}{\operatorname{Card}_{\varepsilon} V(K,u)^* (|u_K - u_Z| + \varepsilon)}, \tag{2.40}$$

in which the smoothed version $\operatorname{Card}_{\varepsilon} V(K,u)^*$ of $\operatorname{Card} V(K,u)^*$ is defined, for $u \in \mathcal{H}_{\mathcal{M}}$ and $K \in \mathcal{M}$, by

$$\operatorname{Card}_{\varepsilon} V(K,u)^* = \sum_{Z \in V(K)} \frac{|u_K - u_Z|}{|u_K - u_Z| + \varepsilon}.$$

Note that this smoothed version of $\gamma^{\mathcal{D}}$ still matches the condition (2.9) of Proposition 2.4 so that, following the steps given in section 2.2.2.4, we can start from γ^{ε} to build a smoothed correction β^{ε} defined, for $u \in \mathcal{H}_{\mathcal{M}}$, $K \in \mathcal{M}$ and $Z \in V(K)$, by

$$\beta^{\varepsilon}_{K,Z}(u) = \max\left(\frac{|\mathcal{A}_K(u)|}{\mathrm{Card}_{\varepsilon} V(K,u)^*}, \frac{|\mathcal{A}_Z(u)|}{\mathrm{Card}_{\varepsilon} V(Z,u)^*}\right)\frac{1}{|u_K - u_Z| + \varepsilon} + \frac{\sum_{\sigma \in K|Z} |\sigma| d_{\sigma}}{|u_K - u_Z| + \varepsilon} \tag{2.41}$$

with the convention $\frac{|\mathcal{A}_Z(u)|}{\mathrm{Card}_{\varepsilon} V(Z,u)^*} = 0$ if $Z = \sigma \in \mathcal{E}_{\mathrm{ext}}$.

The corresponding corrected scheme $\mathcal{S}^{\varepsilon}$ thus writes, for all $u \in \mathcal{H}_{\mathcal{M}}$ and all $K \in \mathcal{M}$,

$$\mathcal{S}^{\varepsilon}_K(u) = -\mathcal{A}_K(u) + \sum_{Z \in V(K)} \max\left(\frac{|\mathcal{A}_K(u)|}{\mathrm{Card}_{\varepsilon} V(K,u)^*}, \frac{|\mathcal{A}_Z(u)|}{\mathrm{Card}_{\varepsilon} V(Z,u)^*}\right)\mathrm{sgn}_{\varepsilon}(u_K - u_Z) + \sum_{\sigma \in \mathcal{E}_K} \rho^{\varepsilon}_{K,\sigma}(u), \tag{2.42}$$

where the real function $\mathrm{sgn}_{\varepsilon} : x \in \mathbb{R} \mapsto x/(|x| + \varepsilon)$ regularizes the function sgn and the additional corrective fluxes $\rho^{\varepsilon}_{K,\sigma}$ are defined, for $u \in \mathcal{H}_{\mathcal{M}}$, $K \in \mathcal{M}$, and $\sigma \in K|Z$ by $\rho^{\varepsilon}_{K,\sigma}(u) = |\sigma| d_{\sigma} \mathrm{sgn}_{\varepsilon}(u_K - u_Z)$. According to section 2.2.2.4 this scheme has the LMP structure, is coercive and Proposition 2.10 ensures it admits at least one solution. Moreover, if the original scheme is conservative, then this scheme is also.

Remark 2.15. Considering a sequence (u^{ε}) of solutions to the regularized schemes (2.42) and sending $\varepsilon \to 0$, one can expect to obtain a solution to the unregularized scheme defined by (2.28). Indeed, thanks to the *a priori* estimate (1.10), the sequence (u^{ε}) is bounded in the finite-dimensional space $\mathcal{H}_{\mathcal{M}}$ and then converges, up to a subsequence, to a discrete function $u \in \mathcal{H}_{\mathcal{M}}$. However, passing to the limit in (2.42) does not prove that u satisfies (2.28). Actually, since the function sgn is not continuous at the origin, we can only conclude that, up to a subsequence, as $\varepsilon \to 0$

$$\mathrm{sgn}_{\varepsilon}(u^{\varepsilon}_K - u^{\varepsilon}_Z) \to \begin{cases} \mathrm{sgn}(u_K - u_Z) & \text{if } u_Z \neq u_K \\ s_{K,Z} & \text{if } u_Z = u_K, \end{cases}$$

for some $s_{K,Z} \in [-1,1]$. Then, as $\varepsilon \to 0$, $\mathrm{Card}\, V(K,u^{\varepsilon})^* \to \Sigma(K)$ with

$$\Sigma(K) = \mathrm{Card}\, V(K,u)^* + \sum_{\substack{Z \in V(K) \\ u_Z = u_K}} \left|s_{K,Z}\right|.$$

Thus we can only conclude that u satisfies the limit scheme

$$-\mathcal{A}_K(u) + \sum_{Z \in V(K,u)^*} \left\{\max\left(\frac{|\mathcal{A}_K(u)|}{\Sigma(K)}, \frac{|\mathcal{A}_Z(u)|}{\Sigma(Z)}\right) + \sum_{\sigma \in K|Z} |\sigma| d_{\sigma}\right\}\mathrm{sgn}(u_K - u_Z) + \sum_{\substack{Z \in V(K) \\ u_Z = u_K}} \left\{\max\left(\frac{|\mathcal{A}_K(u)|}{\Sigma(K)}, \frac{|\mathcal{A}_Z(u)|}{\Sigma(K)}\right) + \sum_{\sigma \in K|Z} |\sigma| d_{\sigma}\right\} s_{K,Z} = |K| f_K,$$

which does not coincide with (2.28).

In order to address the question of convergence for the scheme $\mathcal{S}^\varepsilon$, the proposition bellow gives an estimate on $\mathcal{A}^{\mathcal{D}}(u)$ if u is a solution to (2.42).

The statement of this proposition uses the sets $V(K,u)^+$ and $V(K,u)^-$ defined, as said before, by:

$$V(K,u)^+ = \{Z \in V(K) \ ; \ \mathcal{A}_K(u)(u_Z - u_K) > 0\},$$
$$V(K,u)^- = \{Z \in V(K) \ ; \ \mathcal{A}_K(u)(u_Z - u_K) < 0\}.$$

Proposition 2.16. *Let $\mathcal{D}$ be an admissible mesh of Ω and let $\theta \geq \mathrm{regul}(\mathcal{D})$ and $\varepsilon > 0$. Let u be a solution to $\mathcal{S}^\varepsilon$ and let $K_0 \in \mathcal{M}$ be such that*

$$\frac{\left|\mathcal{A}_{K_0}(u)\right|}{\mathrm{Card}_\varepsilon V(K_0,u)^*} = \max_{K\in\mathcal{M}} \frac{|\mathcal{A}_K(u)|}{\mathrm{Card}_\varepsilon V(K,u)^*}. \tag{2.43}$$

Assume that u satisfies:

$$\text{there exists } Z \in V(K_0,u)^+ \text{ such that } \left|u_{K_0} - u_Z\right| \geq \varepsilon. \tag{2.44}$$

Then there exists C_4 only depending on d and θ such that, for all $K \in \mathcal{M}$,

$$\frac{|\mathcal{A}_K(u)|}{\mathrm{Card}_\varepsilon V(K,u)^*} \leq |K_0|\left|f_{K_0}\right| + C_4|K_0|. \tag{2.45}$$

Proof. It is sufficient to prove estimate (2.45) for $K = K_0$. Now the K_0 component of $\mathcal{S}^\varepsilon(u)$ reduces to

$$-\mathcal{A}_{K_0}(u) + \sum_{Z\in V(K_0)} \frac{\left|\mathcal{A}_{K_0}(u)\right|}{\mathrm{Card}_\varepsilon V(K_0,u)^*}\mathrm{sgn}_\varepsilon(u_{K_0} - u_Z) + \sum_{\sigma\in\mathcal{E}_{K_0}} \rho^\varepsilon_{K_0,\sigma}(u) = |K_0| f_{K_0}. \tag{2.46}$$

Summing separately on $V(K_0,u)^-$ and $V(K_0,u)^+$, we get

$$\begin{aligned}-\mathcal{A}_{K_0}(u) + \sum_{Z\in V(K_0)} &\frac{\left|\mathcal{A}_{K_0}(u)\right|}{\mathrm{Card}_\varepsilon V(K_0,u)^*}\mathrm{sgn}_\varepsilon(u_{K_0} - u_Z) \\ &= -\mathcal{A}_{K_0}(u)\left(1 - \sum_{Z\in V(K_0,u)^-} \frac{\left|\mathrm{sgn}_\varepsilon(u_{K_0} - u_Z)\right|}{\mathrm{Card}_\varepsilon V(K_0,u)^*} + \sum_{Z\in V(K_0,u)^+} \frac{\left|\mathrm{sgn}_\varepsilon(u_{K_0} - u_Z)\right|}{\mathrm{Card}_\varepsilon V(K_0,u)^*}\right).\end{aligned}$$

Since condition (2.9) for the family γ^ε can be written

$$\sum_{Z\in V(K_0,u)^-} \frac{\left|\mathrm{sgn}_\varepsilon(u_{K_0} - u_Z)\right|}{\mathrm{Card}_\varepsilon V(K_0,u)^*} + \sum_{Z\in V(K_0,u)^+} \frac{\left|\mathrm{sgn}_\varepsilon(u_{K_0} - u_Z)\right|}{\mathrm{Card}_\varepsilon V(K_0,u)^*} = 1,$$

we then have

$$\begin{aligned}-\mathcal{A}_{K_0}(u) + \sum_{Z\in V(K_0)} &\frac{\left|\mathcal{A}_{K_0}(u)\right|}{\mathrm{Card}_\varepsilon V(K_0,u)^*}\mathrm{sgn}_\varepsilon(u_{K_0} - u_Z) \\ &= \frac{-2\mathcal{A}_{K_0}(u)}{\mathrm{Card}_\varepsilon V(K_0,u)^*} \sum_{Z\in V(K_0,u)^+} \left|\mathrm{sgn}_\varepsilon(u_{K_0} - u_Z)\right|, \end{aligned} \tag{2.47}$$

Now since $\left|\mathrm{sgn}_\varepsilon(x)\right| \geq 1/2$ when $|x| \geq \varepsilon$, assumption (2.44) ensures that

$$\sum_{Z \in V(K_0,u)^+} \left|\mathrm{sgn}_\varepsilon(u_{K_0} - u_Z)\right| \geq 1/2.$$

Substituting (2.47) in (2.46), applying the triangle inequality and using this last bound lead to

$$\frac{\left|\mathcal{A}_{K_0}(u)\right|}{\mathrm{Card}_\varepsilon V(K_0,u)^*} \leq |K_0|\left|f_{K_0}\right| + \sum_{\sigma \in \mathcal{E}_{K_0}} \left|\rho^\varepsilon_{K_0,\sigma}(u)\right|. \tag{2.48}$$

Finally remark that, for all $K \in \mathcal{M}$,

$$\sum_{\sigma \in \mathcal{E}_K} \left|\rho^\varepsilon_{K,\sigma}(u)\right| \leq \sum_{\sigma \in \mathcal{E}_K} |\sigma| d_\sigma \leq d(1+\theta)|K|. \tag{2.49}$$

Plugging this last inequality with $K = K_0$ into (2.48) gives the desired estimates. □

Adding some regularity assumption on the mesh, the following result states the convergence of the solution to the scheme $\mathcal{S}^\varepsilon$ provided this solution fulfills condition (2.44) above. In the following, for $u \in \mathcal{H}_\mathcal{M}$, we say that $K \in \mathcal{M}$ is a maximal cell for u if

$$\frac{|\mathcal{A}_K(u)|}{\mathrm{Card}_\varepsilon V(K,u)^*} = \max_{L \in \mathcal{M}} \frac{|\mathcal{A}_L(u)|}{\mathrm{Card}_\varepsilon V(L,u)^*}. \tag{2.50}$$

Proposition 2.17. *Assume $f \in L^d(\Omega)$. Let $(\mathcal{D}^n)_{n\geq 1}$ be a sequence of admissible meshes of Ω such that $\mathrm{size}(\mathcal{D}^n) \to 0$ as $n \to \infty$ and $(\mathrm{reg}_\mathcal{A}(\mathcal{D}^n))_{n\geq 1}$ is bounded; assume that there exists $C_5 > 0$ verifying*

$$\forall n \geq 1, \forall K, L \in \mathcal{M}^n, \quad |K| \leq C_5 |L|, \tag{2.51}$$

$$\forall n \geq 1, \forall K \in \mathcal{M}^n, \quad \mathrm{diam}(K)^d \leq C_5 |K|. \tag{2.52}$$

Let $(\varepsilon_n)_{n\geq 1}$ be a sequence of positive real numbers and let $(u^n)_{n\geq 1}$ be a sequence of discrete functions satisfying:

- *For all $n \geq 1$, $u^n \in \mathcal{H}_{\mathcal{M}^n}$ is a solution to the scheme $\mathcal{S}^{\varepsilon_n}$.*
- *For all $n \geq 1$, there exists a maximal cell $K_0^n \in \mathcal{M}^n$ for u^n for which*

$$\text{there exists } Z \in V(K_0^n, u^n)^+ \text{ such that } \left|u^n_{K_0^n} - u^n_Z\right| \geq \varepsilon_n. \tag{2.53}$$

Then, as $n \to \infty$, u^n converges in $L^2(\Omega)$ to the unique solution of (2.1).

Proof. We show that, thanks to assumption (2.53) made on $(u^n)_{n\geq 1}$, condition (2.29) of Proposition 2.11 is satisfied. For simplicity we drop the index n. From Proposition 2.16 and the triangle inequality, we know since $\mathrm{reg}_\mathcal{A}(\mathcal{D})$ is bounded that there exists a constant C_6 independent of $\mathcal{D}$ and ε such that, for all $K \in \mathcal{M}$,

$$\sum_{Z \in V(K)} \beta^\varepsilon_{K,Z}(u)|u_K - u_Z| \leq C_6 \int_{K_0} (|f|+1) + \sum_{\sigma \in \mathcal{E}_K} \left|\rho^\varepsilon_{K,\sigma}(u)\right|.$$

From Hölder inequality and assumption (2.51) we get

$$\sum_{Z\in V(K)} \beta^{\varepsilon}_{K,Z}(u)|u_K - u_Z| \le C_7 |K|^{\frac{d-1}{d}} \left(\int_{K_0} (|f|+1)^d \right)^{\frac{1}{d}} + \sum_{\sigma\in\mathcal{E}_K} \left|\rho^{\varepsilon}_{K,\sigma}(u)\right|,$$

with $C_7 = \max\left(C_5^{\frac{d-1}{d}} C_6, C_6\right)$. Then, bounding $\mathrm{diam}(K)$ by $C_5^{\frac{1}{d}} |K|^{\frac{1}{d}}$, we get C_8 that does not depend on $\mathcal{D}$ or ε such that

$$\sum_{K\in\mathcal{M}} \mathrm{diam}(K) \sum_{Z\in V(K)} \beta^{\varepsilon}_{K,Z}(u)|u_K - u_Z| \le C_8 \left(\int_{K_0} (|f|+1)^d \right)^{\frac{1}{d}} + N_{1,\mathcal{D}}(\rho^{\varepsilon}(u)). \tag{2.54}$$

Since $|f|+1 \in L^d(\Omega)$, the first term of this right-hand side tends to 0 as $\mathrm{size}(\mathcal{D}) \to 0$. The norm comparison (2.31) shows that

$$N_{1,\mathcal{D}}(\rho^{\varepsilon}(u)) \le C_9 N_{\infty,\mathcal{D}}(\rho^{\varepsilon}(u)) \le C_{10}\,\mathrm{size}(\mathcal{D})$$

(with constants depending neither on $\mathrm{size}(\mathcal{D})$ nor on ε). Therefore, as $\mathrm{size}(\mathcal{D})$ tends to 0,

$$\sum_{K\in\mathcal{M}} \mathrm{diam}(K) \sum_{Z\in V(K)} \beta^{\varepsilon}_{K,Z}(u)|u_K - u_Z| \to 0.$$

This guarantees we can apply Proposition 2.11 and conclude that $u \to \bar{u}$ in $L^2(\Omega)$ as $\mathrm{size}(\mathcal{D}) \to 0$. □

2.4 Numerical results

To deal with the nonlinear terms, we perform an iterative algorithm. Let us denote u^i the value of the solution where i is a fixed point iteration. We fix $u = u^i$ in $\beta_{K,Z}(u)$ in (2.7) and the iterative scheme can be written :

$$\forall K \in \mathcal{M}, \quad -\mathcal{A}_K(u^{i+1}) + \sum_{Z\in V(K)} \beta_{K,Z}(u^i)(u_K^{i+1} - u_Z^{i+1}) = |K| f_K.$$

We stop the algorithm when the criterion $\dfrac{\|u^{i+1} - u^i\|}{\|u^i\|} \le 10^{-4}$ is satisfied. We start from the conservative and consistent original operator $\mathcal{A}^{\mathcal{D}}$ developed in [1]. Moreover, we use grids of squares of surface h^2 (h changing from $\frac{1}{8}$ to $\frac{1}{128}$) so that this scheme is also coercive (see Table 1.1). Some notations used to present the numerical results are given in Table 2.1.

2.4.1 Stationary analytical solution

In order to numerically estimate the convergence of the scheme, let us consider the following elliptic problem:

$$\begin{cases} -\nabla\cdot(D\nabla\bar{u}) = f & \text{in } \Omega = (0,0.5)\times(0,0.5), \\ \bar{u}(x,y) = \sin(\pi x)\sin(\pi y) & \text{for } (x,y)\in\partial\Omega \end{cases} \tag{2.55}$$

with

$$D = \frac{1}{x^2+y^2} \begin{pmatrix} y^2+\alpha x^2 & -(1-\alpha)xy \\ -(1-\alpha)xy & x^2+\alpha y^2 \end{pmatrix}$$

Table 2.1 – Notations.

h	size of the discretization
L^2 error	L^2 error of the computed solution with respect to the analytical solution
ratiol2	order of convergence, in L^2 norm, of the method
nit	number of iterations needed to compute the approximate solution of $\mathcal{S}$
Min. Val.	$\min\{u_K \ ; \ K \in \mathcal{M}\}$
Max. Val.	$\max\{u_K \ ; \ K \in \mathcal{M}\}$
$\left\|u_{K_0} - u_{Z^*}\right\|$	$\max\left\{\left\|u_{K_0} - u_Z\right\| \ ; \ Z \in V(K_0,u)^+\right\}$
$\frac{\|\mathcal{A}_{K^*}\|}{\|K^*\|}$	$\max\left\{\frac{\|\mathcal{A}_K\|}{\|K\|} \ ; \ K \in \mathcal{M}\right\}$

and

$$\begin{cases} u_{\text{ana}}(x,y) = \sin(\pi x)\sin(\pi y), \\ f = -\nabla \cdot D\nabla u_{\text{ana}}. \end{cases} \tag{2.56}$$

The parameter α is equal to 10^{-6} and the anisotropy ratio is equal to 10^6. We check that $f \geq 0$.
We show the results obtained in Table 2.2 with the scheme developed in [1] (S. 1), with the first correction (S. 2) and with the regularized correction (S. 3). For the scheme 2, we choose $\eta = 2$. For the scheme 3, we choose $\varepsilon = 4h^2$.

Table 2.2 – Numerical results for (2.55) with the original scheme, the first correction and the regularized correction as a function of the discretization step.

h	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{128}$
L^2 error (S. 1)	5.21×10^{-1}	1.96×10^{-1}	7.14×10^{-2}	1.65×10^{-2}	2.14×10^{-3}
ratiol2 (S. 1)		1.41	1.46	2.11	2.95
Undershoots (S. 1)	12.5 %	10 %	5 %	2 %	1 %
Min. Val. (S. 1)	-2.9×10^{-1}	-2.4×10^{-1}	-1.4×10^{-1}	-5.26×10^{-2}	-1.33×10^{-2}
L^2 error (S. 2)	1.59×10^{-1}	8.98×10^{-2}	4.73×10^{-2}	2.47×10^{-3}	1.30×10^{-2}
ratiol2 (S. 2)		0.82	0.93	0.94	0.93
nit	7	11	13	13	13
$\frac{\|\mathcal{A}_{K^*}\|}{\|K^*\|}$	13.26	15.80	16.60	17.25	18.09
L^2 error (S. 3)	9.03×10^{-2}	4.27×10^{-2}	2.12×10^{-2}	1.00×10^{-2}	4.75×10^{-3}
ratiol2 (S. 3)		1.08	1.01	1.07	1.08
nit	15	17	18	18	15
$\left\|u_{K_0} - u_{Z^*}\right\|$	1.43×10^{-1}	3.62×10^{-2}	9.10×10^{-3}	2.28×10^{-3}	5.70×10^{-4}
ε	6.25×10^{-2}	1.56×10^{-2}	3.90×10^{-3}	9.77×10^{-4}	2.44×10^{-4}

It is clear that the original scheme is at least second order in space but we observe large oscillations. Concerning the scheme 2 and 3, they become first order in space but all oscillations disappear. For the scheme 2, looking at the terms $\frac{\|\mathcal{A}_{K^*}\|}{\|K^*\|}$, the assumptions of Proposition 2.14 seem to hold in this case. For the scheme 3, we also check the assumptions of Proposition 2.17. As we use squares, the grids satisfy clearly the inequalities (2.51)-(2.52). Moreover, looking at the terms $\left|u_{K_0} - u_{Z^*}\right|$, the inequalities (2.53) are verified for all the grids considered so that we may expect this inequality to hold with further refinement of the grid.

2.4.2 Stationary non analytical solution

In order to evaluate the respect of the discrete maximum principle, we now consider the problem:

$$\begin{cases} -\nabla\cdot(D\nabla\bar{u}) = f & \text{in } \Omega = (0,0.5)\times(0,0.5), \\ \bar{u} = 0 & \text{on } \partial\Omega \end{cases} \tag{2.57}$$

and

$$f(x,y) = \begin{cases} 10 & \text{if } (x,y) \in (0.25,0.5)\times(0.25,0.5), \\ 0 & \text{otherwise,} \end{cases} \tag{2.58}$$

where D is as before (see (2.55)). We also choose $\eta = 2$ and $\varepsilon = 4h^2$.
The Table 2.3 shows the minimum and the maximum values for the original scheme, the first correction and the regularized correction. It is interesting to observe that the oscillations can be quite large unless the grid is thin. Figure 2.3 shows that they can be numerous even on the thin grid. On the other hand, as expected, no such oscillations appear with the modified schemes (Figure 2.2). For the two corrected schemes, the number of iterations seems to be bounded as a function of the discretization step when we refine the grid. Moreover, looking at the terms $\frac{|\mathcal{A}_{K^*}|}{|K^*|}$ and $\left|u_{K_0} - u_{Z^*}\right|$, the inequalities (2.36) and (2.53) are also satisfied for all the grids which signals a promising outlook for the convergence of the corrected schemes.

Table 2.3 – Numerical results for (2.57) with the original scheme, the first correction and the regularized correction as a function of the discretization step.

h	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{128}$
Undershoots (S. 1)	37 %	28%	21 %	19 %	20%
Min. Val. (S. 1)	-4.62×10^{-2}	-3.91×10^{-2}	-1.08×10^{-2}	-1.09×10^{-2}	-4.71×10^{-3}
Max. Val. (S. 1)	2.97×10^{-1}	3.3×10^{-1}	3.5×10^{-1}	3.8×10^{-1}	4.1×10^{-1}
Min. Val. (S. 2)	2.38×10^{-3}	1.16×10^{-4}	8.75×10^{-7}	3.30×10^{-10}	1.82×10^{-15}
Max. Val. (S. 2)	9.41×10^{-2}	1.13×10^{-1}	1.16×10^{-1}	2.12×10^{-1}	2.62×10^{-1}
nit	8	11	13	19	20
$\frac{\lvert\mathcal{A}_{K^*}\rvert}{\lvert K^*\rvert}$	7.06	11.81	14.43	16.94	17.81
Min. Val. (S. 3)	1.12×10^{-3}	5.90×10^{-5}	1.55×10^{-6}	3.53×10^{-8}	7.95×10^{-10}
Max. Val. (S. 3)	1.21×10^{-1}	1.41×10^{-1}	1.95×10^{-1}	2.48×10^{-1}	2.92×10^{-1}
nit	8	13	16	20	21
$\left\lvert u_{K_0} - u_{Z^*}\right\rvert$	6.88×10^{-2}	2.17×10^{-2}	5.14×10^{-3}	1.25×10^{-3}	3.07×10^{-4}
ε	6.25×10^{-2}	1.56×10^{-2}	3.90×10^{-3}	9.77×10^{-4}	2.44×10^{-4}

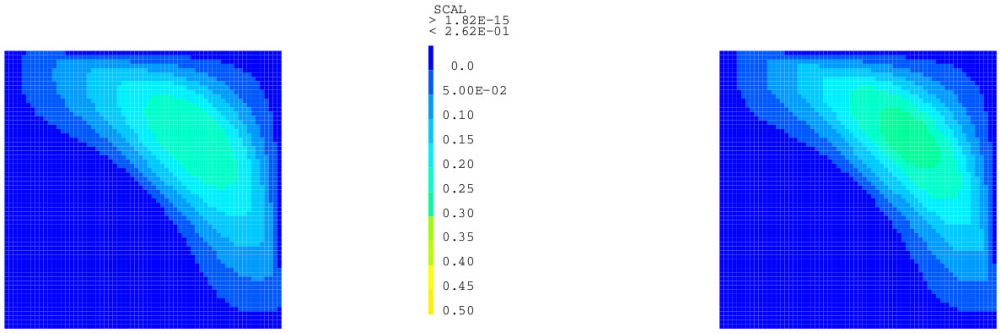

Figure 2.2 – Concentration on a grid made of 4096 squares for the first correction (maximum value 0.26, minimum value 1.82×10^{-15}) and the regularized correction (maximum value 0.29, minimum value 7.95×10^{-10}).

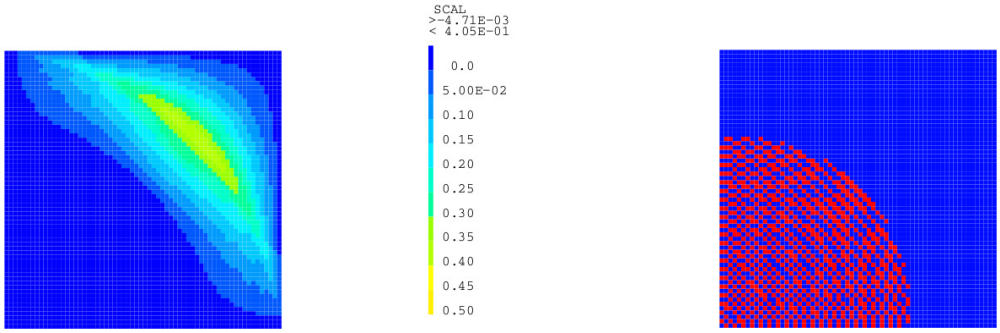

Figure 2.3 – Concentration and position of the undershoots (in red) for the original scheme on a grid made of 4096 cells (maximum value 0.41, minimum value -4.71×10^{-3}).

PARTIE

II

Prise en compte de topographies irrégulières dans des modèles shallow water

CHAPITRE

3 Shallow water waves over polygonal bottoms

The traditional shallow water model for waves propagating over varying bathymetry depends for its derivation on the asymptotic analysis of a Dirichlet-Neumann operator. This analysis however is restricted to smoothly varying topographies. In this chapter, we propose an adaptation to one dimensional polygonal bottoms using the conformal mapping idea of Hamilton and Nachbin. The asymptotic analysis of the Dirichlet-Neumann operator relies on an *ad hoc* transformation of the fluid domain into a flat bottom domain. We derive a new shallow water model which accounts for polygonal topographies. The proposed model is used in the next chapter as a reference model to validate the formal method developed therein.

Contents

3.1 Introduction

3.1.1 Water waves over polygonal topographies

Studies of surface water wave dynamics in the presence of variable topographies are of great interest from coastal engineering point of view. Despite this importance, there is no general agreement about how to describe shallow water flows over rough topographies. Actually, it has been known for quite some time that the presence of strongly varying topographies introduces special problems for the formal derivation of shallow water models. In [74], Hamilton raises the limitations of the long wave models derived by Mei and Le Méhauté [96] and Peregrine [108] when the bottom is strongly sloping. For two dimensional flows, he used a conformal mapping technique (inspired from Kreisel [83]) to derive a long wave model on a fluid of strongly varying depth. The restriction of this method is that it requires knowledge of the conformal mapping between the fluid domain and a flat strip. In case the topography has polygonal shape, Nachbin [97] used Schwarz-Christoffel theory to compute this conformal mapping (numerically) and derived a weakly nonlinear, weakly dispersive, terrain-following Boussinesq system.

The difficulties pointed out by Hamilton to derive shallow water models in the presence of non smooth topographies also occur if one wants to use the more recent method based on the Zakharov/Craig-Sulem formulation of the water waves problem. The main task of this method is the asymptotic analysis of the Dirichlet-Neumann operator involved in this particular formulation of the water waves problem (see *e.g.* [10] or [79]). Now this analysis depends upon the transformation of the fluid layer into a flat strip. Unfortunately, as noted by Lannes (see [85, section 2.5.3]), the classical diffeomorphism between the fluid layer and a flat strip cannot be used when the bottom parametrization is not regular.

On the basis of these considerations, we intend in this chapter to conduce the shallow water analysis of the Dirichlet-Neumann operator when the bottom has polygonal shape using the conformal mapping idea of Hamilton and Nachbin to straighten the fluid layer.

3.1.2 Formulation of the water waves problem

The water waves problem consists in describing the motion of the free surface, denoted by $\zeta(t,x)$, of an incompressible, homogeneous and inviscid fluid, under the influence of gravity. Thorough this chapter, we assume that the topography of the bottom is polygonal, with a finite number of edges (that is the bottom is flat at infinity). The fluid domain is given by

$$\Omega(\zeta,b)=\Big\{(x,z)\in\mathbb{R}^2\ \ ;\ \ -H_0+b(x)<z<\zeta(t,x)\Big\},$$

where H_0 is a reference depth and $b(x)$ denotes the polygonal variations of the bottom (see Figure 3.1). With the usual assumption of irrotational flow, the fluid velocity is represented by the gradient of a potential Φ.

The asymptotic analysis of the water waves problem requires the use of dimensionless quantities based on characteristics of the flow. More precisely, denoting by λ the typical wavelength of the waves, by $a_{\rm surf}$ their typical amplitude and by $a_{\rm bott}$ the typical amplitude of the bottom variations, we define dimensionless variables and unknowns as

$$x'=\frac{x}{\lambda},\quad z'=\frac{z}{H_0},\quad t'=\frac{\sqrt{gH_0}}{\lambda}t,$$

and

$$\zeta'=\frac{\zeta}{a_{\rm surf}},\quad b'=\frac{b}{a_{\rm bott}},\quad \Phi'=\frac{\Phi}{a_{\rm surf}\lambda\sqrt{g/H_0}}.$$

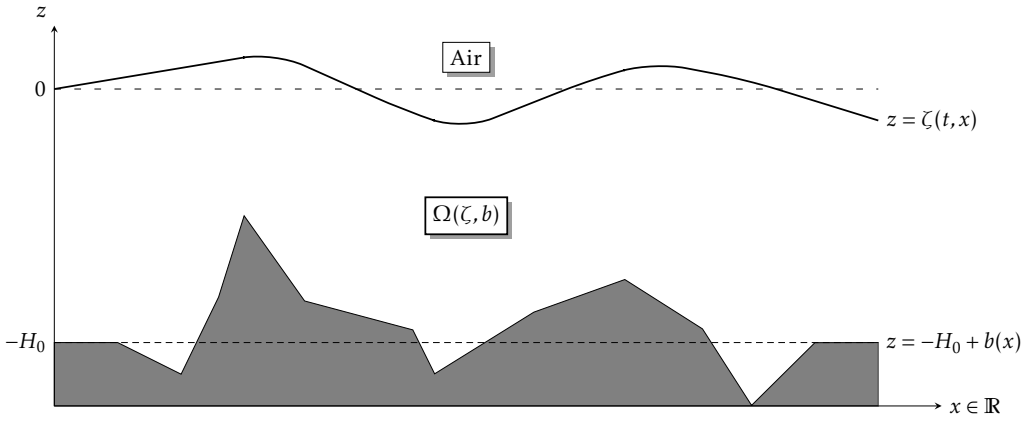

Figure 3.1 – Sketch of the fluid domain.

To simplify the notations we omit the prime symbol. From the previous physical scales we also define three independent parameters:

$$\mu = \frac{H_0^2}{\lambda^2}, \quad \varepsilon = \frac{a_{\text{surf}}}{H_0}, \quad \beta = \frac{a_{\text{bott}}}{H_0}.$$

Our analysis focuses on the shallow water regime $\mu \ll 1$. The parameters ε and β respectively account for the relative amplitude of the waves and of the bathymetry.

Zakharov [118] and Craig and Sulem [41, 40] remarked that the water waves equations can be written as a system of two scalar evolution equations on the surface elevation ζ and on the velocity potential at the surface $\psi = \Phi_{|z=\varepsilon\zeta}$. The key point is that at time t, given $\zeta(t,\cdot)$ and b, the knowledge of $\psi(t,\cdot)$ fully determines the velocity potential $\Phi(t,\cdot,\cdot)$ in the fluid domain as the solution of the following nondimensionalized elliptic problem

$$\begin{cases} \mu\partial_x^2\Phi + \partial_z^2\Phi = 0 & \text{in } \Omega(\varepsilon\zeta, \beta b), \\ \Phi = \psi & \text{on } \{z = \varepsilon\zeta\}, \\ \partial_{\mathbf{n}}\Phi = 0 & \text{on } \{z = -1 + \beta b\}, \end{cases} \tag{3.1}$$

where $\partial_{\mathbf{n}}$ stands for the outward conormal[1] derivative associated with the elliptic operator $\mu\partial_x^2 + \partial_z^2$. In particular, one may define the Dirichlet-Neumann operator as

$$\mathcal{G}_\mu[\varepsilon\zeta, \beta b] : \psi \mapsto \sqrt{1 + \varepsilon^2 |\partial_x\zeta|^2}\,\partial_{\mathbf{n}}\Phi_{|z=\varepsilon\zeta}. \tag{3.2}$$

The Zakharov/Craig-Sulem formulation of the water waves problem then reads, in dimensionless form,

$$\begin{cases} \partial_t\zeta - \dfrac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = 0, \\ \partial_t\psi + \zeta + \dfrac{\varepsilon}{2}\left|\partial_x\psi\right|^2 - \varepsilon\mu\dfrac{\left(\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi + \varepsilon\partial_x\zeta\partial_x\psi\right)^2}{2(1 + \varepsilon^2\mu|\partial_x\zeta|^2)} = 0. \end{cases} \tag{3.3}$$

[1] At the bottom boundary, the outward unit normal vector $\mathbf{n}$ is well defined everywhere except at the vertices.

Using the Zakharov/Craig-Sulem formulation (3.3) as a starting point, approximating the water waves equation in shallow water regime then amounts to understand the asymptotic behavior of the Dirichlet-Neumann operator when the shallowness parameter μ is small.

3.1.3 Asymptotic expansion of the Dirichlet-Neumann operator

Transforming the Laplace equation into an elliptic problem on a flat strip. Since the Dirichlet-Neumann operator is explicitly defined in terms of the velocity potential, a natural way to derive asymptotic properties of this operator is by studying the asymptotic behavior of the potential. The issue is then to study a Laplace equation on the unknown fluid domain $\Omega(\varepsilon\zeta, \beta b)$. An efficient approach to get around this issue is to transform the fluid domain to the flat strip $\mathcal{S} = (-1, 0) \times \mathbb{R}$ (see *e.g.* [20, 10, 79] or [101]). The main interest is that the resulting transformed potential on the flat strip then solves an elliptic boundary value problem with variable coefficients defined on the fixed domain $\mathcal{S}$. Since the Dirichlet-Neumann operator can be expressed in terms of the transformed potential, constructing a shallow water expansion of $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ reduces to finding an approximate solution to this new boundary value problem on $\mathcal{S}$.

Limitations of the classical approach. In the previous approach, the choice of the diffeomorphism between the flat strip and the fluid domain is important because it governs the form of the resulting elliptic problem on $\mathcal{S}$. More precisely, given a diffeomorphism Σ mapping $\mathcal{S}$ onto $\Omega(\varepsilon\zeta, \beta b)$, we know from Proposition 2.7 of [84] that the transformed velocity potential $\phi = \Phi \circ \Sigma$ satisfies

$$\nabla_{x,z} \cdot P[\Sigma]\nabla_{x,z}\phi = 0 \quad \text{in } \mathcal{S},$$

where $\nabla_{x,z} = [\partial_x, \partial_z]^T$ and where the matrix $P[\Sigma]$ is defined in terms of the Jacobian matrix J_Σ of Σ as

$$P[\Sigma] = |\det J_\Sigma| J_\Sigma^{-1} \begin{bmatrix} \mu & 0 \\ 0 & 1 \end{bmatrix} (J_\Sigma^{-1})^T.$$

To define this diffeomorphism, the simplest choice consists in transforming only the vertical coordinate:

$$\forall (x, z) \in \mathcal{S}, \quad \Sigma(x, z) = \Big(x, \varepsilon\zeta(x) + z(1 + \varepsilon\zeta(x) - \beta b(x))\Big).$$

Unfortunately, this choice requires some regularity on the bottom parametrization b since the coefficients of P involve, among others, the derivative of b.

Schwarz-Christoffel mappings as an adaptation to polygonal topographies. In this chapter, we intend to adapt the previous approach to the particular case of a polygonal topography. Therefore, the first task is to construct a diffeomorphism between the flat strip and the fluid domain with polygonal bottom. This task is undertaken in section 3.2, in which we transform the Laplace problem (3.1) on the polygonal bottom domain $\Omega(\varepsilon\zeta, \beta b)$ into a variable coefficients elliptic problem on the flat strip $\mathcal{S}$. Introducing complex canonical coordinates, we use a conformal mapping technique, namely Schwarz-Christoffel mapping theory (see [99] for example), to straighten the polygonal bottom. Section 3.3 is entirely devoted to the shallow water analysis of the Dirichlet-Neumann operator. Following the usual approach outlined above, we show that this operator can be expressed in terms of the solution of the boundary value problem on the flat strip. The asymptotic expansion of the Dirichlet-Neumann operator then hinges on the construction of an approximate solution to this boundary value problem. Using this asymptotic analysis, we finally derive in section 3.4 a shallow water model which accounts for polygonal topographies.

3.2 Reduction to a problem on the flat strip

In view of the shallow water analysis of the Dirichlet-Neumann operator, it is important to address the recovering of the velocity potential Φ from its trace ψ at the surface. For this reason, this section is devoted to the study of the Laplace equation on the physical domain $\Omega = \Omega(\varepsilon\zeta, \beta b)$ with polygonal topography:

$$\begin{cases} \mu\partial_x^2\Phi + \partial_z^2\Phi = 0 & \text{in } \Omega, \\ \Phi = \psi & \text{on } \{z = \varepsilon\zeta\}, \\ \partial_{\mathbf{n}}\Phi = 0 & \text{on } \{z = -1 + \beta b\}. \end{cases} \tag{3.4}$$

In what follows, we assume that the water depth remains positive:

$$\exists h_{\min} > 0, \quad 1 + \varepsilon\zeta - \beta b \geq h_{\min}. \tag{3.5}$$

In this section, we explain how to transform the problem (3.4) into a variable coefficients elliptic problem on the flat strip. The main issue is to straighten the polygonal bottom using a smooth mapping. Section 3.2.1 is devoted to the construction of such straightening mappings. The transformation of the Laplace problem on Ω into a boundary value problem on the flat strip then proceeds in two steps (see Figure 3.2):

1. Using the straightening of the bottom, we first transform in section 3.2.2 the Laplace problem (3.4) into a Laplace problem with flat bottom via a diffeomorphism, denoted by $\Sigma_{\text{bott}}^{-1}$.
2. Starting from this transformed Laplace problem with flat bottom, section 3.2.3 then addresses the flattening of the fluid boundary using the diffeomorphism $\Sigma_{\text{surf}}^{-1}$.

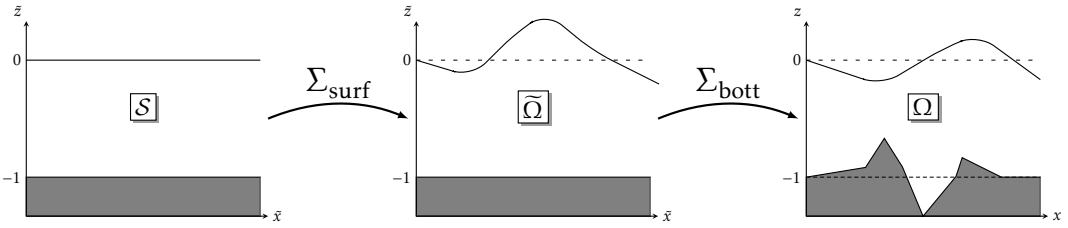

Figure 3.2 – Two step straightening of the physical domain: $\Sigma_{\text{bott}}^{-1}$ straightens the bottom, then $\Sigma_{\text{surf}}^{-1}$ straightens the fluid boundary.

3.2.1 Using conformal mappings to straighten the bottom

We first aim at mapping the flat strip $\mathcal{S}$ onto the domain at rest

$$\Omega_{\text{rest}} = \Big\{(x,z) \in \mathbb{R}^2 \ ; \ -1 + \beta b(x) < z < 0\Big\}.$$

To achieve this, we focus on *μ-conformal transformations* that is transformations that are conformal for the metric $\mu \mathrm{d}x^2 + \mathrm{d}z^2$. As we will see, working with such mappings is convenient because they leave invariant the nondimensionalized Laplace equation $\mu\partial_x^2\Phi + \partial_z^2\Phi = 0$.

3.2.1.1 Building μ-conformal mappings

Following Nachbin [97], a simple way to build μ-conformal maps from conformal maps is to use a vertical scaling by a factor $\sqrt{\mu}$, namely $T_\mu : (x,z) \mapsto (x, \sqrt{\mu} z)$. As a matter of fact, a transformation $\Sigma : \mathcal{S} \mapsto \Omega_{\text{rest}}$ is μ-conformal if and only if the transformation $\Sigma_\mu = T_\mu \circ \Sigma \circ T_\mu^{-1}$ is a conformal map. Building a μ-conformal transformation further amounts to building a conformal map Σ_μ between the scaled domains $\mathcal{S}^\mu = T_\mu(\mathcal{S})$ and $\Omega_{\text{rest}}^\mu = T_\mu(\Omega_{\text{rest}})$ (see Figure 3.3). The main interest of such a scaling is that conformal transformations between

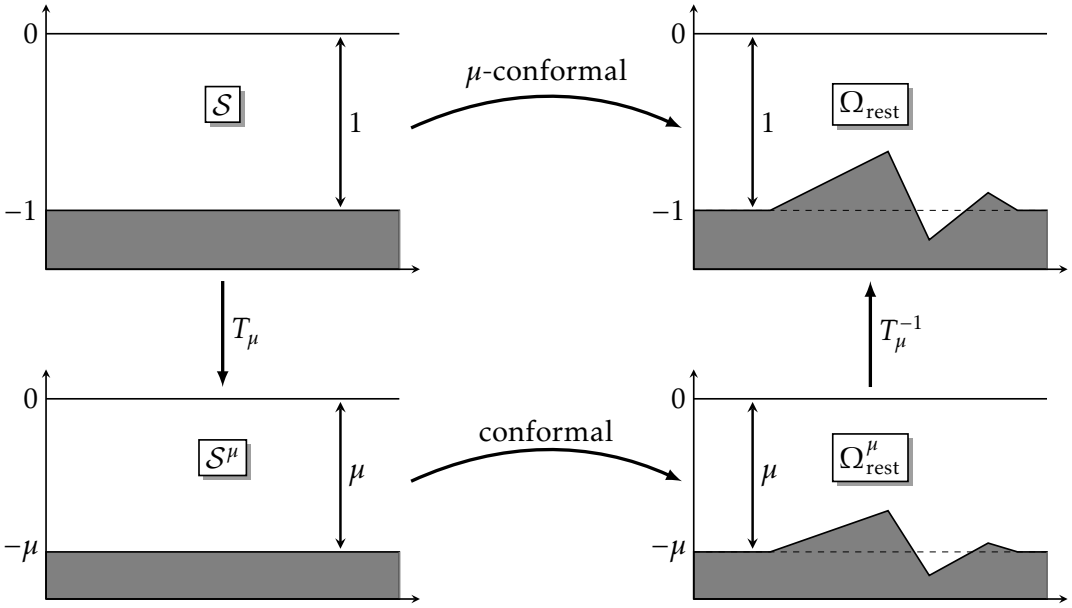

Figure 3.3 – Getting μ-conformal mappings from conformal ones.

a strip and a polygonal domain can be found from using Schwarz-Christoffel mapping theory. The latter ensures that the problem of seeking a conformal map from a strip to the interior of a polygonal region can be reduced to solving a nonlinear system of equations, whose unknowns are the pre-images of the vertices of the polygonal boundary. This problem, known as the Schwarz-Christoffel parameter problem, is nontrivial and generally analytically intractable but it can be solved numerically with very efficient methods (see for instance [50]). As an illustration for the use of the Schwarz-Christoffel formula, we construct in the following example an analytic function which maps $\mathcal{S}$ onto a step bottom domain.

Example 3.1. Let us consider the simple case of a step bottom, that is suppose that the bathymetry is parametrized by

$$b(x) = \begin{cases} 0 & \text{if } x < 0, \\ b_0 & \text{if } x > 0, \end{cases}$$

where b_0 is some positive constant. Identifying $\mathbb{R}^2$ with the complex plane $\mathbb{C}$, the rescaled strip $\mathcal{S}^\mu$ reads

$$\mathcal{S}^\mu = \left\{ \omega = x + iz \ ; \ (x,z) \in \mathbb{R}^2, \ -\sqrt{\mu} < z < 0 \right\}$$

and the rescaled domain at rest Ω^μ_{rest} is given as

$$\Omega^\mu_{\text{rest}} = \Big\{\omega = x + iz \ ; \ (x,z) \in \mathbb{R}^2, \ \sqrt{\mu}(-1 + \beta b(x)) < z < 0\Big\}.$$

We seek for an analytic function Σ_μ from $\mathcal{S}^\mu$ to Ω^μ_{rest} which maps the upper boundary $\{z = 0\}$

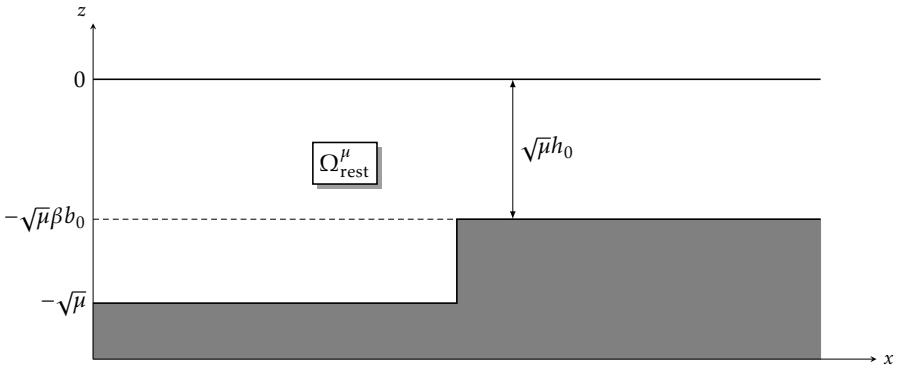

Figure 3.4 – Rescaled domain at rest in the case of a step

onto itself and the lower boundary $\{z = -\sqrt{\mu}\}$ onto the rescaled bottom boundary of Ω^μ_{rest}. Using the Schwarz-Christoffel formula (see *e.g* [50, Theorem 2.1]), the desired mapping can be found by integrating the following expression

$$\frac{\mathrm{d}\Sigma_\mu}{\mathrm{d}\omega}(\omega) = \left(\frac{1 + h_0^2 \exp\Big(\frac{\pi\omega}{\sqrt{\mu}}\Big)}{1 + \exp\Big(\frac{\pi\omega}{\sqrt{\mu}}\Big)}\right)^{1/2},$$

where $h_0 = (1 - \beta b_0)$ (see Figure 3.4). It follows that

$$\Sigma_\mu(\omega) = \sqrt{\mu}\frac{h_0}{\pi}\left(\log\Big(\frac{1+s}{1-s}\Big) - \frac{1}{h_0}\log\Big(\frac{s+h_0}{s-h_0}\Big)\right),$$

where $s = h_0\left(\frac{1+\exp\left(\frac{\pi\omega}{\sqrt{\mu}}\right)}{1+h_0^2\exp\left(\frac{\pi\omega}{\sqrt{\mu}}\right)}\right)^{\frac{1}{2}}$. See Figure 3.5 for a sketch of the behavior of the resulting conformal mapping Σ_μ.

3.2.1.2 Choice of a μ-conformal transformation

Let Σ_μ be the Schwarz-Christoffel mapping from the rescaled strip $\mathcal{S}^\mu$ to the rescaled physical domain at rest Ω^μ_{rest}. As explained above, a μ-conformal transformation Σ_{bott} from $\mathcal{S}$ to Ω_{rest} is given by $\Sigma_{\text{bott}} = T_\mu^{-1} \circ \Sigma_\mu \circ T_\mu$, where we recall that T_μ is the scaling defined as $T_\mu(x,z) = (x, \sqrt{\mu}z)$. Now, when solving the Schwarz-Christoffel parameter problem, the conformal map Σ_μ can be chosen so as to map the upper boundary $\{z = 0\}$ of $\mathcal{S}^\mu$ onto the corresponding boundary of Ω^μ_{rest}. Thus, Schwarz reflection principle ensures that Σ_μ can be analytically continued across $\{z = 0\}$ to the reflected strip $[0, \sqrt{\mu}) \times \mathbb{R}$. It follows that the associated μ-conformal transformation Σ_{bott} is actually a diffeomorphism between the strip $(-1, 1) \times \mathbb{R}$ and the augmented domain $\Omega_{\text{rest}} \cup \Omega^*_{\text{rest}}$ obtained as the union of Ω_{rest} with the

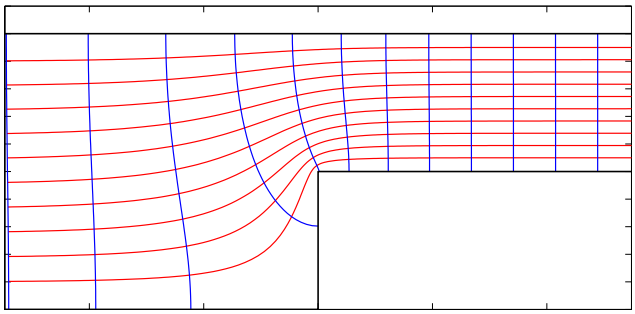

Figure 3.5 – Level lines for the conformal map of Example 3.1

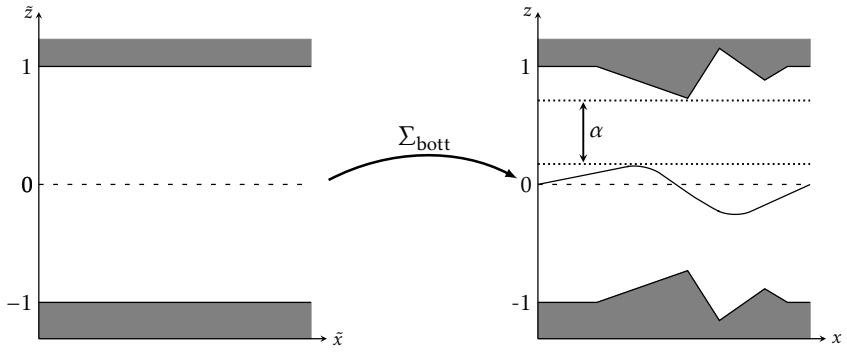

Figure 3.6 – Reflected domains and assumption (3.6).

reflected domain at rest $\Omega^*_{\text{rest}} = \left\{(x,z) \in \mathbb{R}^2 \ ; \ 0 < z < 1 - \beta b(x)\right\}$ (see Figure 3.6). We assume that the fluid domain Ω is contained within this augmented domain, more precisely

$$\exists \alpha > 0, \quad 1 - \beta b - \varepsilon\zeta \geq \alpha \quad \text{on } \mathbb{R}. \tag{3.6}$$

This allows us to set $\widetilde{\Omega} = \Sigma_{\text{bott}}^{-1}(\Omega)$ and to define functions σ_μ and ρ_μ through

$$\forall x \in \mathbb{R}, \quad (\sigma_\mu(x), \rho_\mu(x)) = \Sigma_{\text{bott}}^{-1}(x, \varepsilon\zeta(x)).$$

In other words, we get a transformed domain $\widetilde{\Omega}$ with flat bottom and whose free surface $\widetilde{\Gamma}$ is parametrized by $(\sigma_\mu(x), \rho_\mu(x))$[2]. Moreover the diffeomorphism Σ_{bott} between $\widetilde{\Omega}$ and Ω maps the bottom (resp. fluid) boundary of $\widetilde{\Omega}$ onto the bottom (resp. fluid) boundary of Ω (see Figure 3.7).

Remark 3.2. Even if the physical elevation is parametrized by a graph (namely $z = \varepsilon\zeta(x)$), the transformed free surface $\widetilde{\Gamma}$ of $\widetilde{\Omega}$ is not necessarily a graph too. However, for the sake of simplicity, we assume in what follows that $\widetilde{\Gamma}$ may be parametrized as a graph. More precisely, we make the following assumption

$$\exists \delta > 0, \quad \sigma'_\mu > \delta. \tag{3.7}$$

[2]The subscript μ on "σ" and "ρ" reminds one that the transformed free surface depends on μ. Indeed, by definition, the transformation Σ_{bott} is dependent on μ. Note that, as is clear from their definition, σ_μ and ρ_μ also depend on ε but this dependence is omitted.

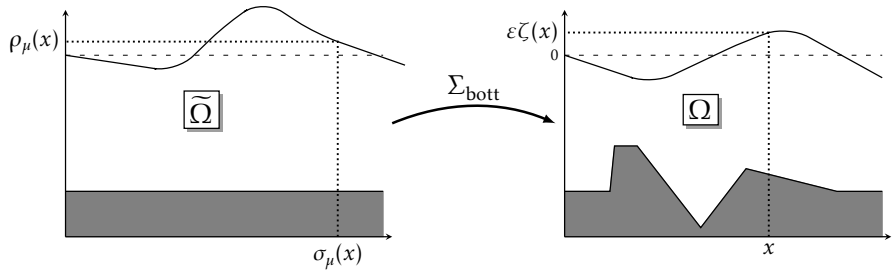

Figure 3.7 – Straightening the polygonal topography with a μ-conformal diffeomorphism.

This assumption implies that σ_μ is a diffeomorphism so that, setting $\widetilde{\zeta}_\mu = \rho_\mu \circ \sigma_\mu^{-1}$, the transformed surface may be parametrized as

$$\widetilde{\Gamma} = \Big\{(\tilde{x}, \widetilde{\zeta}_\mu(\tilde{x})) \ ; \ \tilde{x} \in \mathbb{R}\Big\}.$$

Note that, since $\Sigma_{\text{bott}}^{-1}$ is smooth near $\{z = 0\}$ and since this diffeomorphism maps this line onto itself, the previous assumption holds if ε is small enough.

3.2.2 Transformed Laplace equation with flat bottom

By carefully choosing the diffeomorphism Σ_{bott}, we have ensured that the transformed potential $\widetilde{\phi} = \Phi \circ \Sigma_{\text{bott}}$ still solves the (nondimensionalized) Laplace equation on the flat bottom domain $\widetilde{\Omega}$. This actually follows from the fact that Σ_{bott} is a μ-conformal transformation. More precisely, defining the inner product $\langle \cdot, \cdot \rangle_\mu$ as

$$\forall u, v \in \mathbb{R}^2, \quad \langle u, v \rangle_\mu = I_\mu u \cdot v,$$

where

$$I_\mu = \begin{bmatrix} \mu & 0 \\ 0 & 1 \end{bmatrix},$$

the Jacobian matrix $J_{\Sigma_{\text{bott}}}$ of Σ_{bott} enjoys the following property.

Lemma 3.3. *For all vectors $u, v \in \mathbb{R}^2$,*

$$\Big\langle J_{\Sigma_{\text{bott}}}^T u, J_{\Sigma_{\text{bott}}}^T v \Big\rangle_\mu = \Big|\det J_{\Sigma_{\text{bott}}}\Big| \langle u, v \rangle_\mu \tag{3.8}$$

Proof. The Jacobian matrix of the conformal transformation Σ_μ can be written as

$$J_{\Sigma_\mu} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

Differentiating the identity $\Sigma_{\text{bott}} = T_\mu^{-1} \circ \Sigma_\mu \circ T_\mu$ then gives

$$J_{\Sigma_{\text{bott}}} = \begin{bmatrix} \tilde{a} & -\sqrt{\mu}\tilde{b} \\ \frac{\tilde{b}}{\sqrt{\mu}} & \tilde{a} \end{bmatrix}, \tag{3.9}$$

where $\tilde{a} = a \circ T_\mu$ and $\tilde{b} = b \circ T_\mu$. Equality (3.8) then follows from direct computations using the latter expression for $J_{\Sigma_{\text{bott}}}$. □

Before stating the main proposition of this section let us recall that, with the notation above, the Laplace equation $\mu\partial_x^2\Phi + \partial_z^2\Phi = 0$ holds in $\mathcal{D}'(\Omega)$ if and only if

$$\forall \varphi \in \mathcal{D}(\Omega), \quad \int_\Omega \langle \nabla_{x,z}\Phi, \nabla_{x,z}\varphi \rangle_\mu = 0.$$

Proposition 3.4. *Assume that $\partial_x\psi \in H^{1/2}(\mathbb{R})$ and that $\zeta \in W^{1,\infty}(\mathbb{R})$. Then the transformed potential $\widetilde{\phi} = \Phi \circ \Sigma_{\text{bott}}$ is the solution of*

$$\begin{cases} \mu\partial_{\tilde{x}}^2\widetilde{\phi} + \partial_{\tilde{z}}^2\widetilde{\phi} = 0 & \text{in } \widetilde{\Omega}, \\ \widetilde{\phi} = \widetilde{\psi}_\mu & \text{on } \left\{\tilde{z} = \widetilde{\zeta}_\mu\right\}, \\ \partial_{\tilde{z}}\widetilde{\phi} = 0 & \text{on } \{\tilde{z} = -1\}, \end{cases} \tag{3.10}$$

where the velocity $\widetilde{\psi}_\mu$ at the transformed free surface is defined as $\widetilde{\psi}_\mu = \psi \circ \sigma_\mu^{-1}$.

Proof. Given the definitions of $\widetilde{\zeta}_\mu$ and $\widetilde{\psi}_\mu$, it is straightforward that $\widetilde{\phi}$ satisfies the Dirichlet boundary condition on the fluid boundary $\left\{\tilde{z} = \widetilde{\zeta}_\mu(\tilde{x})\right\}$. To prove that $\widetilde{\phi}$ solves (3.10), it therefore remains to show that we have

$$\int_{\widetilde{\Omega}} \left\langle \nabla_{\tilde{x},\tilde{z}}\widetilde{\phi}, \nabla_{\tilde{x},\tilde{z}}\tilde{v} \right\rangle_\mu = 0, \tag{3.11}$$

for any test function $\tilde{v}$ in the functional space

$$\widetilde{V} = \left\{\tilde{v} \in H^1(\widetilde{\Omega}) \ ; \ \tilde{v} = 0 \text{ on } \widetilde{\Gamma}\right\}.$$

Let us prove first that (3.11) holds for $\tilde{v}$ of the form $\tilde{v} = v \circ \Sigma_{\text{bott}}$, where v is any function in $\mathcal{D}(\Omega \cup \{z = -1 + \beta b\})$. Given such a function, we thus have to check that

$$\int_{\widetilde{\Omega}} \langle \nabla_{\tilde{x},\tilde{z}}(\Phi \circ \Sigma_{\text{bott}}), \nabla_{\tilde{x},\tilde{z}}(v \circ \Sigma_{\text{bott}}) \rangle_\mu = 0. \tag{3.12}$$

Using the chain rule and then applying Lemma 3.3 yields

$$\int_{\widetilde{\Omega}} \langle \nabla_{\tilde{x},\tilde{z}}(\Phi \circ \Sigma_{\text{bott}}), \nabla_{\tilde{x},\tilde{z}}(v \circ \Sigma_{\text{bott}}) \rangle_\mu = \int_{\Sigma_{\text{bott}}(\Omega)} \langle \nabla_{x,z}\Phi \circ \Sigma_{\text{bott}}, \nabla_{x,z} v \circ \Sigma_{\text{bott}} \rangle_\mu \left|\det J_{\Sigma_{\text{bott}}}\right|.$$

Using the mapping Σ_{bott} to perform a change of variable in the last integral, we get

$$\int_{\widetilde{\Omega}} \langle \nabla_{\tilde{x},\tilde{z}}(\Phi \circ \Sigma_{\text{bott}}), \nabla_{\tilde{x},\tilde{z}}(v \circ \Sigma_{\text{bott}}) \rangle_\mu = \int_\Omega \langle \nabla_{x,z}\Phi, \nabla_{x,z} v \rangle_\mu$$

and thus, since Φ solves (3.4),

$$\int_{\widetilde{\Omega}} \langle \nabla_{\tilde{x},\tilde{z}}(\Phi \circ \Sigma_{\text{bott}}), \nabla_{\tilde{x},\tilde{z}}(v \circ \Sigma_{\text{bott}}) \rangle_\mu = 0.$$

It is important to note that, due to the behavior of Σ_{bott} near the prevertices (in $\widetilde{\Omega}$) of the polygonal bottom, every function of the form $\tilde{v} = v \circ \Sigma_{\text{bott}}$ is not necessarily in $H^1(\widetilde{\Omega})$. Nonetheless, by construction of Σ_{bott}, we know that $v \circ \Sigma_{\text{bott}}$ is smooth on condition that v vanishes near the vertices of the polygonal bottom. Using such test function in (3.12), we deduce that (3.11) holds for all $\tilde{v} \in \mathcal{D}(-1 \leq \tilde{z} < 1 + \widetilde{\zeta}_\mu)$ supported away from the prevertices. Since this last set is dense in $\widetilde{V}$ (see [71, Lemma 2.1.2]) we finally conclude that (3.11) holds for any test function $\tilde{v} \in \widetilde{V}$. □

3.2.3 Transformed problem on the flat strip

We have seen in Proposition 3.4 that the Laplace problem with polygonal bottom (3.4) can be reduced to the same equation posed in a transformed fluid domain $\widetilde{\Omega}$ with flat bottom. Therefore, to reduce the problem to a boundary value problem on the flat strip, it simply remains to straighten the transformed fluid boundary $\left\{\tilde{z} = \widetilde{\zeta}_\mu(\tilde{x})\right\}$ of $\widetilde{\Omega}$. This can be done using the classical straightening diffeomorphism defined in section 3.1.3. Following this classical approach we define the diffeomorphism Σ_{surf} mapping $\mathcal{S}$ onto the flat bottom domain $\widetilde{\Omega}$ as

$$\begin{aligned}\Sigma_{\text{surf}}: \quad \mathcal{S} &\longrightarrow \widetilde{\Omega}\\ (\tilde{x},\tilde{z}) &\longmapsto \Big(\tilde{x}, \widetilde{\zeta}_\mu(\tilde{x}) + \tilde{z}(1+\widetilde{\zeta}_\mu(\tilde{x}))\Big).\end{aligned}$$

Then we know (see *e.g.* [85, Proposition 2.26]) that the Laplace problem on $\widetilde{\Omega}$, and thus the Laplace problem on Ω, are equivalent to the following elliptic equation on $\mathcal{S}$

$$\begin{cases}\nabla_{\tilde{x},\tilde{z}}\cdot\widetilde{P}\nabla_{\tilde{x},\tilde{z}}\phi = 0 & \text{in } \mathcal{S},\\ \phi = \widetilde{\psi}_\mu & \text{on } \{\tilde{z}=0\},\\ \partial_{\mathbf{n}}\phi = 0 & \text{on } \{\tilde{z}=-1\},\end{cases} \tag{3.13}$$

with $\phi = \widetilde{\phi}\circ\Sigma_{\text{surf}}$ and where the matrix $\widetilde{P}$ is given by

$$\widetilde{P} = \begin{bmatrix} \mu(1+\widetilde{\zeta}_\mu) & -\mu(\tilde{z}+1)\partial_{\tilde{x}}\widetilde{\zeta}_\mu \\ -\mu(\tilde{z}+1)\partial_{\tilde{x}}\widetilde{\zeta}_\mu & \frac{1+\mu(\tilde{z}+1)^2\left(\partial_{\tilde{x}}\widetilde{\zeta}_\mu\right)^2}{1+\widetilde{\zeta}_\mu}\end{bmatrix}, \tag{3.14}$$

and $\partial_{\mathbf{n}}\phi_{|\tilde{z}=-1} = \mathbf{e}_{\tilde{z}}\cdot\widetilde{P}\nabla_{\tilde{x},\tilde{z}}\phi_{|\tilde{z}=-1}$ denotes the upward conormal derivative.

3.3 Shallow-water analysis of the Dirichlet-Neumann operator

In this section we concentrate on the asymptotic analysis of the Dirichlet-Neumann operator in shallow water regime ($\mu \ll 1$). In the light of the previous section, the Laplace problem (3.4) with polygonal bottom boundary can be reduced to the elliptic problem (3.13) on the flat strip. This can then be used to build an asymptotic expansion of $\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi$ with respect to μ following the usual method for smooth topographies (see for instance [10] or [20, 87, 31] where the method has been used to derive long-wave models, see also [102] in which this change of variable approach is applied to address the analyticity of the Dirichlet-Neumann operator). As outlined in the introduction, this method consists of the following steps:

1. Express the Dirichlet-Neumann operator in terms of the solution ϕ to the problem (3.13) on the flat strip;
2. Approximate the transformed potential ϕ on $\mathcal{S}$ using a BKW procedure;
3. Plug this approximate solution back into the expression of (1) to compute an approximation of $\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi$.

The first step is undertaken in section 3.3.1, while the second and third steps are detailed in section 3.3.2.

3.3.1 The Dirichlet-Neumann operator on the flat strip

As seen in the introduction, the Dirichlet-Neumann operator is given in terms of the velocity potential as

$$\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = \partial_z \Phi_{|z=\varepsilon\zeta} - \mu\varepsilon \partial_x \zeta \partial_x \Phi_{|z=\varepsilon\zeta} \tag{3.15}$$

$$= \left\langle \nabla_{x,z} \Phi_{|z=\varepsilon\zeta}, \mathbf{n} \right\rangle_\mu, \tag{3.16}$$

where $\mathbf{n}$ is the (non-unit) normal vector to the free surface defined as $\mathbf{n} = [-\varepsilon\partial_x\zeta\,, 1]^T$. Using the μ-conformal transformation Σ_{bott}, this operator can be similarly expressed in terms of the transformed velocity potential $\widetilde{\phi} = \Phi \circ \Sigma_{\text{bott}}$ on the flat bottom domain $\widetilde{\Omega}$.

Proposition 3.5. *Assume that $\partial_x\psi \in H^{1/2}(\mathbb{R})$ and that $\zeta \in H^{t_0+2}(\mathbb{R})$ for some $t_0 > 1/2$. Then, the Dirichlet-Neumann operator can be written as*

$$\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = \left\langle \nabla_{\tilde{x},\tilde{z}} \widetilde{\phi} \circ (\sigma_\mu, \rho_\mu), \tilde{\mathbf{n}} \right\rangle_\mu, \tag{3.17}$$

where $\tilde{\mathbf{n}}$ is the (non-unit) normal vector to the transformed free surface $\widetilde{\Gamma}$ defined as $\tilde{\mathbf{n}} = [-\rho'_\mu\,, \sigma'_\mu]^T$.

Proof. The proof relies on the fact that the transformation Σ_{bott} is μ-conformal. Indeed, from the chain rule, we have

$$\left(J^T_{\Sigma_{\text{bott}}} \circ (\sigma_\mu, \rho_\mu)\right) \nabla_{x,z} \Phi_{|z=\varepsilon\zeta} = \nabla_{\tilde{x},\tilde{z}} \widetilde{\phi} \circ (\sigma_\mu, \rho_\mu).$$

Using Lemma 3.3 and the above equality in the definition (3.15) of $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]$, we obtain

$$\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = \left\langle \nabla_{\tilde{x},\tilde{z}} \widetilde{\phi} \circ (\sigma_\mu, \rho_\mu), \left(\frac{J^T_{\Sigma_{\text{bott}}}}{\left|\det J_{\Sigma_{\text{bott}}}\right|} \circ (\sigma_\mu, \rho_\mu) \right) \mathbf{n} \right\rangle_\mu,$$

and so it remains to show that

$$\left(\frac{J^T_{\Sigma_{\text{bott}}}}{\left|\det J_{\Sigma_{\text{bott}}}\right|} \circ (\sigma_\mu, \rho_\mu) \right) \mathbf{n} = \tilde{\mathbf{n}}. \tag{3.18}$$

To do this, we first note that a tangent vector to the transformed fluid boundary $\widetilde{\Gamma}$ is given by $\tilde{\mathbf{t}} = J^{-1}_{\Sigma_{\text{bott}}} \circ (\sigma_\mu, \rho_\mu)\mathbf{t}$, where $\mathbf{t} = [1\,, \varepsilon\partial_x\zeta]^T$ is a tangent vector to the (physical) free surface $\{z = \varepsilon\zeta(x)\}$ (see Figure 3.8). Now, using the expression (3.9) for $J^T_{\Sigma_{\text{bott}}}$, we get

$$\left(\frac{J^T_{\Sigma_{\text{bott}}}}{\left|\det J_{\Sigma_{\text{bott}}}\right|} \circ (\sigma_\mu, \rho_\mu) \right) I_\mu^{-1} \mathbf{t} = I_\mu^{-1} \tilde{\mathbf{t}}.$$

Equality (3.18) is then obtained from this last relation by noting that $\tilde{\mathbf{t}} = [\sigma'_\mu, \rho'_\mu]^T$ and using again the expression (3.9) for $J^T_{\Sigma_{\text{bott}}}$. □

Remark 3.6. A Dirichlet-Neumann operator $\widetilde{\mathcal{G}}_\mu[\tilde{\zeta}]\widetilde{\psi}_\mu$ associated with the transformed Laplace problem with flat bottom (3.10) can be defined as

$$\widetilde{\mathcal{G}}_\mu[\tilde{\zeta}]\widetilde{\psi}_\mu = \partial_{\tilde{z}} \widetilde{\phi}_{|\tilde{z}=\widetilde{\zeta}_\mu} - \mu \partial_{\tilde{x}} \widetilde{\zeta}_\mu \partial_{\tilde{x}} \widetilde{\phi}_{|\tilde{z}=\widetilde{\zeta}_\mu}. \tag{3.19}$$

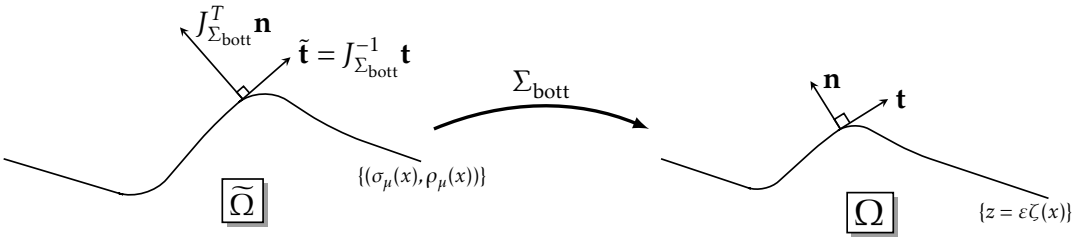

Figure 3.8 – Tangent and normal vectors to the free surface and the transformed free surface.

It is worth noticing that both quantities $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ and $\widetilde{\mathcal{G}}_\mu[\tilde{\zeta}]\widetilde{\psi}_\mu$ do not coincide since the μ-conformal diffeomorphism Σ_{bott} acts on the horizontal coordinate when straightening the bottom. More precisely, we can rewrite the above result as

$$\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = \sigma'_\mu \Big(\widetilde{\mathcal{G}}_\mu[\tilde{\zeta}]\widetilde{\psi}_\mu\Big) \circ \sigma_\mu, \tag{3.20}$$

where we recall that σ_μ is related to Σ_{bott} through

$$\forall x \in \mathbb{R}, \quad (\sigma_\mu(x), \rho_\mu(x)) = \Sigma^{-1}_{\text{bott}}(x, \varepsilon\zeta(x))$$

so that σ_μ can be considered as the horizontal deformation of the free surface $\{z = \varepsilon\zeta\}$ due to the straightening of the bottom.

Since we have transformed the Laplace problem with flat bottom (3.10) into the elliptic problem (3.13) on the flat strip using the trivial diffeomorphism Σ_{surf}, we know that the Dirichlet-Neumann operator $\widetilde{\mathcal{G}}_\mu[\tilde{\zeta}]\widetilde{\psi}_\mu$ coincides with the Dirichlet-Neumann operator that comes from this straightened elliptic problem on the flat strip (see *e.g.* [85, Remark 3.7]). Combining this expression with the previous relationship (3.20) between $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ and $\widetilde{\mathcal{G}}_\mu[\tilde{\zeta}]\widetilde{\psi}_\mu$ yields the desired expression of $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ in terms of the solution ϕ to the elliptic problem (3.13) on $\mathcal{S}$.

Proposition 3.7. *Assume that $\partial_x\psi \in H^{1/2}(\mathbb{R})$ and that $\zeta \in H^{t_0+2}(\mathbb{R})$ for some $t_0 > 1/2$. Then, the Dirichlet-Neumann operator can be written in terms of the transformed potential on the flat strip as*

$$\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = \sigma'_\mu\, \mathbf{e}_{\tilde{z}} \cdot \widetilde{P}\nabla_{\tilde{x},\tilde{z}}\phi \circ (\sigma_\mu, 0). \tag{3.21}$$

3.3.2 Asymptotic analysis of the Dirichlet-Neumann operator

Let us summarize the situation so far. Owing to an *ad hoc* straightening of the bottom, we have first reduced the Laplace problem with polygonal bottom (3.4) to the same Laplace equation on the flat bottom domain $\widetilde{\Omega}$ with transformed Dirichlet data $\widetilde{\psi}_\mu$ at the surface. Then we have seen that this transformed Laplace problem can in turn be reduced to a variable coefficients elliptic problem on the flat strip, namely equation (3.13), and that the Dirichlet-Neumann operator can be expressed in terms of the solution ϕ of this elliptic problem. Our strategy here is to use this expression to derive an asymptotic expansion of $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ in shallow water regime. To do so, we first construct an approximate solution ϕ_{app} to (3.13) and then replace the transformed potential ϕ in (3.21) with this approximate potential.

3.3.2.1 Asymptotic expansion of the transformed potential

We look for an asymptotic expansion of the transformed potential ϕ of the form

$$\phi_{\text{app}} = \phi_0 + \mu\phi_1. \tag{3.22}$$

This expansion is constructed as an approximate solution of the transformed elliptic problem on the flat strip, that is such that

$$\begin{cases} \nabla_{\tilde{x},\tilde{z}} \cdot \widetilde{P}\nabla_{\tilde{x},\tilde{z}}\phi_{\text{app}} = O(\mu^2) & \text{in } \mathcal{S}, \\ \phi = \widetilde{\psi}_\mu & \text{on } \{\tilde{z}=0\}, \\ \partial_{\mathbf{n}}\widetilde{\phi} = 0 & \text{on } \{\tilde{z}=-1\}. \end{cases} \tag{3.23}$$

To find this approximate solution, we plug the above expression for ϕ_{app} into the elliptic operator from (3.23), expand the resulting expression in powers of μ and then choose ϕ_0 and ϕ_1 so as to cancel the leading order terms. Mimicking the computations of [85, Lemma 3.42], the resulting approximate solution is given by

$$\phi_0(\tilde{x},\tilde{z}) = \widetilde{\psi}_\mu(\tilde{x}), \tag{3.24}$$

$$\phi_1(\tilde{x},\tilde{z}) = -\Big(\frac{\tilde{z}^2}{2}+\tilde{z}\Big)\Big(1+\widetilde{\zeta}_\mu(\tilde{x})\Big)^2\partial_{\tilde{x}}^2\widetilde{\psi}_\mu(\tilde{x}). \tag{3.25}$$

Remark 3.8. It is noteworthy to note that, in the expansion (3.22), both functions ϕ_0 and ϕ_1 also depend on μ.

3.3.2.2 Asymptotic expansion of the Dirichlet-Neumann operator

Given the previous approximate (transformed) potential ϕ_{app}, we may now compute a formal expansion of $\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi$. More precisely, using ϕ_{app} in the expression of the Dirichlet-Neumann operator on the flat strip given in Proposition 3.7, we get

$$\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi = \mathcal{G}_{\text{app}}\psi + O(\mu^2), \tag{3.26}$$

where $\mathcal{G}_{\text{app}}$ is defined, according to (3.21), as

$$\mathcal{G}_{\text{app}}\psi = \sigma_\mu'\,\mathbf{e}_{\tilde{z}} \cdot \widetilde{P}\nabla_{\tilde{x},\tilde{z}}\phi_{\text{app}} \circ (\sigma_\mu, 0). \tag{3.27}$$

Computing the right hand side with the help of the explicit expressions (3.24)-(3.25) for the functions ϕ_0 and ϕ_1 and then using that, from their definitions, $\widetilde{\psi}_\mu \circ \sigma_\mu = \psi$ and $\widetilde{\zeta}_\mu \circ \sigma_\mu = \rho_\mu$, we find

$$\mathcal{G}_{\text{app}}\psi = -\mu\partial_x\left(\frac{1+\rho_\mu}{\sigma_\mu'}\partial_x\psi\right) + O(\mu^2). \tag{3.28}$$

Due to the dependence on μ of the diffeomorphism Σ_{surf}, we need to make additional assumptions to establish an error estimate for the above approximate Dirichlet-Neumann operator. More precisely, we assume that the transformed free surface satisfies the following conditions uniformly with respect to μ:

(A1) There exists $\tilde{h}_{\min} > 0$, independent on μ, such that $1 + \rho_\mu \geq \tilde{h}_{\min}$.

(A2) There exists $\tilde{r} > 0$, independent on μ, such that $\left|\rho_\mu\right|_{W^{1,\infty}} \leq \tilde{r}$.

(A3) There exists $\delta > 0$, independent on μ, such that $\sigma'_\mu > \delta$.

As will be made clear in the proof of the following proposition, these extra assumptions give control on both the coercivity of the elliptic operator in (3.23) and on the $O(\mu^2)$ right hand side of this approximate equation.

Proposition 3.9. *Let $s \in \mathbb{N}^*$, $\zeta \in H^{s+5/2}(\mathbb{R})$ and $\partial_x\psi \in H^{s+7/2}(\mathbb{R})$ be such that (A1)-(A3) are satisfied and that $\sigma'_\mu \in W^{s,\infty}(\mathbb{R})$.Then the following estimate on the remainder holds*

$$\left|\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi + \mu\partial_x\left(\frac{1+\rho_\mu}{\sigma'_\mu}\partial_x\psi\right)\right|_{H^s} \leq \mu^2 C_0, \tag{3.29}$$

where C_0 is a constant of the form

$$C_0 = C\Big(\frac{1}{\delta}, \frac{1}{\tilde{h}_{\min}}, \tilde{r}, \big|\partial_{\tilde{x}}\widetilde{\psi}_\mu\big|_{H^{s+7/2}}, \big|\widetilde{\zeta}_\mu\big|_{H^{s+5/2}}, \big|\sigma'_\mu\big|_{W^{s,\infty}}\Big),$$

and C is a nondecreasing function of its arguments.

Proof. From the definition of $\mathcal{G}_{\text{app}}\psi$ and Proposition 3.7, we have

$$\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi - \mathcal{G}_{\text{app}}\psi = \sigma'_\mu\, \mathbf{e}_{\tilde{z}} \cdot \widetilde{P}\nabla_{\tilde{x},\tilde{z}}(\phi - \phi_{\text{app}}) \circ (\sigma_\mu, 0).$$

Now, from the construction of the approximate potential ϕ_{app}, one can check that $u = \phi - \phi_{\text{app}}$ solves

$$\begin{cases} \nabla_{\tilde{x},\tilde{z}} \cdot \widetilde{P}\nabla_{\tilde{x},\tilde{z}} u = \mu^2 R_\mu & \text{in } \mathcal{S}, \\ u = 0 & \text{on } \{\tilde{z} = 0\}, \\ \partial_{\mathbf{n}} u = 0 & \text{on } \{\tilde{z} = -1\}, \end{cases} \tag{3.30}$$

with R_μ satisfying

$$\big|R_\mu\big|_{L^2(-1,0\,;\,H^{s+1/2}(\mathbb{R}))} \leq C\Big(\big|\partial_{\tilde{x}}\widetilde{\psi}_\mu\big|_{H^{s+7/2}}, \big|\widetilde{\zeta}_\mu\big|_{H^{s+5/2}}\Big). \tag{3.31}$$

Then, since

$$\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi - \mathcal{G}_{\text{app}}\psi = \sigma'_\mu\, \mathbf{e}_{\tilde{z}} \cdot \widetilde{P}\nabla_{\tilde{x},\tilde{z}} u \circ (\sigma_\mu, 0), \tag{3.32}$$

one may feel inclined to deduce a control of the latter from (3.30) by resorting to elliptic estimates together with a trace inequality. Actually, since the coercivity constant of $\widetilde{P}$ depends on μ (it is of order $O(\mu)$), a straightforward application of such estimates does not directly yield (3.29). To face this difficulty, the idea is to consider the contribution of the shallowness parameter in the elliptic problem (3.30). More precisely, following for instance [10], the elliptic operator from (3.30) can be written as

$$\nabla_{\tilde{x},\tilde{z}} \cdot \widetilde{P}\nabla_{\tilde{x},\tilde{z}} = \nabla^\mu_{\tilde{x},\tilde{z}} \cdot \widetilde{P}_\mu \nabla^\mu_{\tilde{x},\tilde{z}},$$

where the twisted gradient operator $\nabla^\mu_{\tilde{x},\tilde{z}}$ is defined as $\nabla^\mu_{\tilde{x},\tilde{z}} = [\sqrt{\mu}\partial_{\tilde{x}}, \partial_{\tilde{z}}]^T$ and with

$$\widetilde{P}_\mu = \begin{bmatrix} 1 + \widetilde{\zeta}_\mu & -\sqrt{\mu}(\tilde{z}+1)\partial_{\tilde{x}}\widetilde{\zeta}_\mu \\ -\sqrt{\mu}(\tilde{z}+1)\partial_{\tilde{x}}\widetilde{\zeta}_\mu & \frac{1+\mu(\tilde{z}+1)^2\left(\partial_{\tilde{x}}\widetilde{\zeta}_\mu\right)^2}{1+\widetilde{\zeta}_\mu} \end{bmatrix}.$$

The advantage is that the matrix $\widetilde{P}_\mu$ at the core of this twisted formulation satisfies the following coercivity estimate

$$\forall \theta \in \mathbb{R}^2, \quad k_\mu \widetilde{P}_\mu \theta \cdot \theta \geq |\theta|^2,$$

with a constant k_μ of the form

$$k_\mu = 2\Big(1 + \big|\widetilde{\zeta}_\mu\big|_{L^\infty}\Big) + \frac{4}{\tilde{h}_{\min}}\Big(1 + \mu\big|\partial_{\tilde{x}}\widetilde{\zeta}_\mu\big|^2_{L^\infty}\Big).$$

Now, since $\mathbf{e}_{\tilde{z}} \cdot \widetilde{P}\nabla_{\tilde{x},\tilde{z}} u = \mathbf{e}_{\tilde{z}} \cdot \widetilde{P}_\mu \nabla^\mu_{\tilde{x},\tilde{z}} u$, we can apply the aforementioned elliptic estimate and trace inequality to obtain from (3.31)

$$\big|\mathbf{e}_{\tilde{z}} \cdot \widetilde{P}\nabla_{\tilde{x},\tilde{z}} u_{|\tilde{z}=0}\big|_{H^s} \leq \mu^2 C\Big(k_\mu, \big|\partial_{\tilde{x}}\widetilde{\psi}_\mu\big|_{H^{s+7/2}}, \big|\widetilde{\zeta}_\mu\big|_{H^{s+5/2}}\Big).$$

From (3.32), we deduce that

$$\big|\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi - \mathcal{G}_{\rm app}\psi\big|_{H^s} \leq \mu^2 C\Big(\frac{1}{\delta}, \big|\sigma'_\mu\big|_{W^{s,\infty}}, k_\mu, \big|\partial_{\tilde{x}}\widetilde{\psi}_\mu\big|_{H^{s+7/2}}, \big|\widetilde{\zeta}_\mu\big|_{H^{s+5/2}}\Big). \tag{3.33}$$

To conclude, we first note that (3.28) can be written as

$$\mathcal{G}_{\rm app}\psi = -\mu\partial_x\left(\frac{1+\rho_\mu}{\sigma'_\mu}\partial_x\psi\right) - \mu^2\sigma'_\mu\, r_\mu \circ \sigma_\mu,$$

where $r_\mu = (\partial_{\tilde{x}}\widetilde{\zeta}_\mu)^2(1+\widetilde{\zeta}_\mu)\partial^2_{\tilde{x}}\widetilde{\psi}_\mu$. From here, the desired estimate follows from (3.33) by applying the triangle inequality. □

Comment. The main drawback of the estimate furnished by the previous proposition is that the quantities that appear in the constant C_0 in (3.29) depend on the shallowness parameter. Actually, the μ-dependence is mainly due to the contribution of the parametrization (σ_μ, ρ_μ) of the transformed free surface in $\widetilde{\Omega}$. Therefore, to improve this estimate, one should more carefully focus on how the parametrization (σ_μ, ρ_μ) depends on μ. To do so, recall first that as we mentioned in section 3.2.1.1, the construction of the conformal mapping Σ_μ associated with the straightening diffeomorphism $\Sigma_{\rm bott}$ hinges on the resolution of a Schwarz-Christoffel parameter problem. Identifying $\mathbb{R}^2$ with the complex plane, denoting by a_1, a_2, ..., $a_n \in \mathbb{C}$ the vertices of the (rescaled) polygonal bottom and $\alpha_1\pi$, $\alpha_2\pi$,..., $\alpha_n\pi$ its interior angles (see Figure 3.9), this parameter problem actually consists in finding prevertices ω_1, ω_2, ..., $\omega_n \in \big\{z = -\sqrt{\mu}\big\}$ and a constant $\Lambda \in \mathbb{C}$ such that the desired conformal mapping $\Sigma_\mu : \mathbb{C} \to \mathbb{C}$ satisfies

$$\frac{\mathrm{d}\Sigma_\mu}{\mathrm{d}\omega}(\omega) = \Lambda \prod_{k=1}^{n}\Big((\exp\Big(\frac{\pi\omega_k}{\sqrt{\mu}}\Big) - \exp\Big(\frac{\pi\omega}{\sqrt{\mu}}\Big)\Big)^{\alpha_k - 1}.$$

Hence, the parametrization (σ_μ, ρ_μ) of the transformed free surface $\widetilde{\Gamma}$ satisfies the ordinary differential equation

$$\sigma'_\mu + i\sqrt{\mu}\rho'_\mu = \Lambda^{-1}(1 + i\varepsilon\sqrt{\mu}\zeta') \prod_{k=1}^{n}\Big((\exp\Big(\frac{\pi\omega_k}{\sqrt{\mu}}\Big) - \exp\Big(\frac{\pi\sigma_\mu}{\sqrt{\mu}}\Big)\exp(i\pi\rho_\mu)\Big)^{1-\alpha_k}.$$

In the case of a step this differential equation reads (see Example 3.1)

$$\sigma'_\mu + i\sqrt{\mu}\rho'_\mu = (1 + i\varepsilon\sqrt{\mu}\zeta')\left(\frac{1 + \exp\Big(\frac{\pi\sigma_\mu}{\sqrt{\mu}}\Big)\exp(i\pi\rho_\mu)}{1 + h_0^2\exp\Big(\frac{\pi\sigma_\mu}{\sqrt{\mu}}\Big)\exp(i\pi\rho_\mu)}\right)^{1/2}.$$

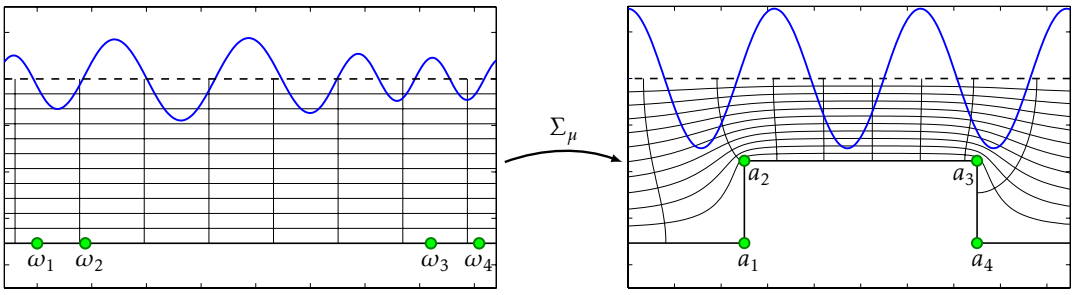

Figure 3.9 – Solution of the Schwarz-Christoffel parameter problem for a bottom with a rectangular hump.

A possible (but far from obvious), way to improve estimate (3.29) might be to perform an asymptotic analysis of the solution of the previous differential equation as $\mu \to 0$.

3.4 A shallow-water model for polygonal bottoms

In this last section, we focus on the study of shallow water waves over the polygonal bottom. Owing to the asymptotic analysis of the Dirichlet-Neumann operator conducted in the previous section, we derive a shallow water model that approximates, at order $O(\mu)$, the solutions of the water waves equations

$$\begin{cases} \partial_t \zeta - \dfrac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = 0, \\ \partial_t \psi + \zeta + \dfrac{\varepsilon}{2}\left|\partial_x \psi\right|^2 - \varepsilon\mu \dfrac{\left(\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi + \varepsilon\partial_x\zeta\partial_x\psi\right)^2}{2(1+\varepsilon^2\mu|\partial_x\zeta|^2)} = 0. \end{cases} \tag{3.34}$$

In the present case, since we consider a rough topography, we choose to work with variables located at the surface, that is away from the singularities of the bottom. Therefore the shallow water model is formulated in terms of the surface elevation ζ and the horizontal velocity at the surface $v_s = (\partial_x \Phi)_{|z=\varepsilon\zeta}$. From the definition of $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$, the vertical component of the velocity, on the other hand, can be written

$$(\partial_z \Phi)_{|z=\varepsilon\zeta} = \mu \frac{\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi + \varepsilon\partial_x\psi\partial_x\zeta}{1+\mu\varepsilon^2(\partial_x\zeta)^2}.$$

Since $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ gives first contributions at $O(\mu)$, we deduce that $(\partial_z\Phi)_{|z=\varepsilon\zeta}$ is of size $O(\mu)$. Since, by definition of $\psi = \Phi_{|z=\varepsilon\zeta}$, we have $\partial_x\psi = v_s + \varepsilon\partial_x\zeta(\partial_z\Phi)_{|z=\varepsilon\zeta}$, we finally get $\partial_x\psi = v_s + O(\mu)$. Plugging the latter in (3.28) yields

$$\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = -\partial_x\left(\frac{1+\rho_\mu}{\sigma_\mu'}v_s\right) + O(\mu), \tag{3.35}$$

where we recall that (σ_μ, ρ_μ) parametrizes the transformed free surface in $\widetilde{\Omega}$ and is defined from the straightening diffeomorphism Σ_{bott} by

$$\forall x \in \mathbb{R}, \quad (\sigma_\mu(x), \rho_\mu(x)) = \Sigma_{\text{bott}}^{-1}(x, \varepsilon\zeta(x)). \tag{3.36}$$

Defining the transformed (variable) free surface coefficient $M_\mu = M_\mu[\varepsilon\zeta, \beta b]$ as

$$M_\mu = \frac{1+\rho_\mu}{\sigma'_\mu}, \tag{3.37}$$

then substituting expansion (3.35) into the first equation of (3.34), we get the following approximate evolution equation for the elevation

$$\partial_t \zeta + \partial_x(M_\mu v_s) = O(\mu).$$

Then, differentiating the second equation of (3.34) with respect to x and using both $\partial_x \psi = v_s + O(\mu)$ and $\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = O(1)$, we are left with the following shallow water model with precision $O(\mu)$

$$\begin{cases} \partial_t \zeta + \partial_x(M_\mu v_s) = 0, \\ \partial_t v_s + \partial_x \zeta + \varepsilon v_s \partial_x v_s = 0. \end{cases} \tag{3.38}$$

Remark 3.10 (Flat bottom). In case the bottom is flat, the straightening diffeomorphism Σ_{bott} reduces to identity so that $\sigma_\mu(x) = x$ and $\rho_\mu(x) = \varepsilon\zeta(x)$. Consequently M_μ coincides with the water depth variable $h = 1 + \varepsilon\zeta$ and we recover the classical Saint-Venant system

$$\begin{cases} \partial_t \zeta + \partial_x(h v_s) = 0, \\ \partial_t v_s + \partial_x \zeta + \varepsilon v_s \partial_x v_s = 0. \end{cases} \tag{3.39}$$

3.5 Conclusion

The main limitation of the shallow water model (3.38) is that the transformed free surface coefficient M_μ depends on the variables (σ_μ, ρ_μ). This is problematic since, due to the analytic intractability of the underlying Schwarz-Christoffel parameter problem, we do not have analytical expression for the mapping $\Sigma^{-1}_{\text{bott}}$. Therefore $M_\mu = \frac{1+\rho_\mu}{\sigma'_\mu}$ cannot be explicitly written in terms of the variables ζ and v_s. As said above, a possible improvement could be provided by an asymptotic analysis of both functions σ_μ and ρ_μ. Unfortunately, we have been unable thus far to find explicit asymptotic expansions for these coefficients, even in the simple case of a step where the expression of the straightening diffeomorphism Σ_{bott} is known. It is notable that, for weakly nonlinear waves $\varepsilon \sim \mu \ll 1$, Nachbin [97] proposes an interesting approximation of a coefficient similar to M_μ by a time independent coefficient[3].

Given these limitations, we develop in the next chapter a different approach to study shallow water flows over rough bathymetries. The starting point of this approach is the shape analyticity of the Dirichlet-Neumann operator. More precisely, we know how to find explicit expressions for the shape derivatives of the Dirichlet-Neumann operator around $\zeta = 0$ and $b = 0$. We will see that these expressions only involve infinitely smooth contributions of the bottom. Therefore, we propose a more formal approach which consists in first approximating the Dirichlet-Neumann operator by its Taylor expansion around $\zeta = 0$ and $b = 0$ and then studying the shallow water asymptotic of the latter expansion.

Let us conclude by remarking that, even if the computation of the transformed free surface coefficient M_μ is not obvious (it requires to evaluate the inverse of the Schwarz-Christoffel

[3]The idea at the basis of this approximation is to evaluate $\Sigma^{-1}_{\text{bott}}$ in (3.36) not on the free surface points $(x, \varepsilon\zeta(x))$ but on the undisturbed surface points $(x, 0)$. Setting $(\sigma_0(x), 0) = \Sigma^{-1}_{\text{bott}}(x, 0)$, this amounts to replace M_μ in (3.38) by the time independent coefficient $M_0 = \frac{1}{\sigma'_0}$.

mapping Σ_μ at the free surface), it can be achieved numerically by using for instance the Schwarz-Christoffel Toolbox of Driscoll and Trefethen [50, Appendix], thus making possible the development of a numerical method for (3.38). More details about the numerical method for (3.38) are given in the appendix of the next chapter, in which we use equations (3.38) as a reference model to assess the performance of the formal approach developed therein. As an illustration, Figure 3.10 shows the time history of the surface elevation computed by (3.39) in the particular case of a rectangular bottom.

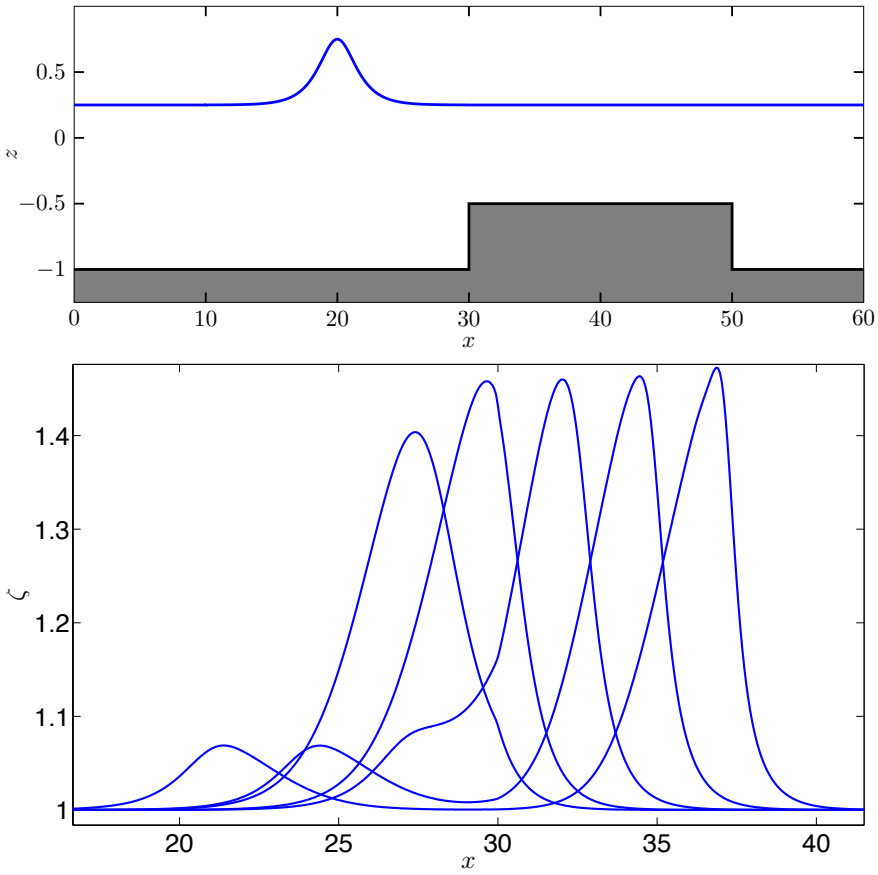

Figure 3.10 – Wave passing over a rectangular hump. Top: topography and initial condition. Bottom: time series of surface elevation.

CHAPITRE

4 Asymptotic shallow water models with non smooth topographies

In this chapter, we present new models to describe shallow water flows over non smooth topographies. The water waves problem is formulated as a system of two equations on surface quantities in which the topography is involved in a Dirichlet-Neumann operator. Starting from this formulation and using the joint analyticity of this operator with respect to the surface and the bottom parametrizations, we derive a nonlocal shallow water model which only includes smoothing contributions of the bottom. Under additional small amplitude assumptions, Boussinesq-type systems are also derived. Using these alternative shallow water models as references, we finally present numerical tests to assess the precision of the classical shallow water approximations over rough bottoms. In the case of a polygonal bottom, we show numerically that our new model is consistent with the approach developed in the previous chapter.

Contents

4.1 Introduction

4.1.1 Water waves over a rough bottom

Surface water waves propagation over a variable bottom has been widely studied over the past decades because of its importance in oceanography. Assuming the fluid is incompressible, homogeneous and inviscid, its motion is governed by the Euler equations with nonlinear boundary conditions at the surface. As the free surface boundary is part of the unknowns, the full problem, known as the water waves problem, is very difficult to solve both mathematically and numerically. Nonetheless, in some specific physical regimes it is possible to derive much simpler asymptotic models (see [86] for a recent review).

In shallow water conditions *i.e.* when the typical wavelength of the waves is much larger than the typical depth, the free surface problem is frequently approximated by the Saint-Venant equations. When the bottom parametrization is smooth, it is known [10, 106, 107, 80, 92, 79] that they provide a good approximation to the exact solution of the full water waves equations. However, in case the bottom is rough, there is no evidence that the Saint-Venant equations are still a relevant approximation of the water waves problem. As a matter of fact, the topography introduces singular terms in the Saint-Venant system if the bottom parametrization is not regular.

On that basis, some models have been proposed to handle rapidly varying periodic or random topographies. We cite in particular the papers of Rosales and Papanicolaou [109], Nachbin and Sølna [98], Craig *et al.* [36], Craig, Lannes and Sulem [37].

Concerning non smooth topographies, Hamilton [74] and Nachbin [97] used a conformal mapping technique to derive long wave models in the case of two-dimensional motions. In [97], a Boussinesq system is formulated to handle non smooth one-dimensional topographies. However, this technique only applies to polygonal (one-dimensional) bottom profiles.

The purpose of the present chapter is to derive alternatives to some classical shallow water models (namely Saint-Venant equations, Serre equations and Boussinesq system) which do not involve any singular term with non smooth topographies. The systems we obtain consist in modifying the topographical terms in the classical shallow water models. In case the bottom is smooth, these new systems are consistent with the former.

4.1.2 Formulation of the water waves problem

As in the previous chapter, we denote by $\zeta(t,x)$ the surface elevation and by $-H_0+b(x)$ a parametrization of the bottom, where H_0 is a reference depth (see Figure 4.1). The time-dependent fluid domain consists of the region

$$\Omega(\zeta,b)=\Big\{(x,z)\in\mathbb{R}^d\times\mathbb{R}\ ;\ -H_0+b(x)<z<\zeta(t,x)\Big\},$$

where $d=1,2$ denotes the spatial dimension of the free surface (and the bottom). Recall that the flow is assumed to be irrotational so that, from the incompressibility assumption, the velocity field is represented by the gradient of an harmonic potential Φ.

Dimensionless variables and unknowns are introduced in the conventional way (see Chapter 3), using the typical wavelength of the waves λ, their typical amplitude $a_{\rm surf}$ and the typical amplitude $a_{\rm bott}$ of the bottom variations. For the sake of simplicity, we use the same notations for the dimensionless quantities (*e.g.* Φ denotes the nondimensionalized velocity

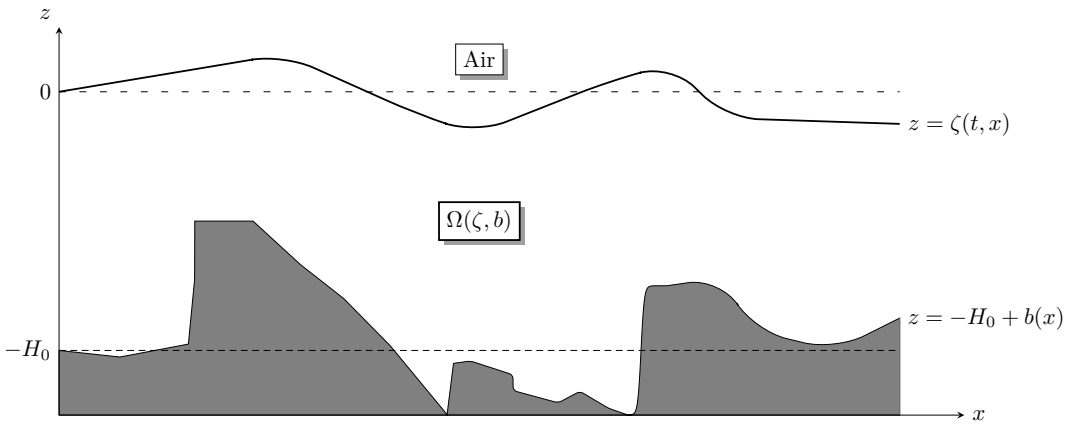

Figure 4.1 – Sketch of the domain

potential). We recall that three independent parameters can be defined from the previous physical scales:

$$\mu = \frac{H_0^2}{\lambda^2}, \quad \varepsilon = \frac{a_{\text{surf}}}{H_0}, \quad \beta = \frac{a_{\text{bott}}}{H_0},$$

where μ is the shallowness parameter, while ε and β are the amplitude parameters of the waves and of the bathymetry, respectively.

As remarked by Zakharov [118], the evolution of the flow is characterized by the evolution of only two quantities located at the surface, namely the surface elevation $\varepsilon\zeta$ and the trace of the velocity potential $\psi = \Phi_{|z=\varepsilon\zeta}$. Recall that the formulation of the water waves problem in the form of a system of two scalar evolution equations on (ζ, ψ), due to Craig and Sulem [41, 40], reads in dimensionless form,

$$\begin{cases} \partial_t \zeta - \dfrac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = 0, \\ \partial_t \psi + \zeta + \dfrac{\varepsilon}{2}\left|\nabla\psi\right|^2 - \varepsilon\mu \dfrac{\left(\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi + \varepsilon\nabla\zeta\cdot\nabla\psi\right)^2}{2(1+\varepsilon^2\mu|\nabla\zeta|^2)} = 0, \end{cases} \tag{4.1}$$

in which ∇ denotes the gradient operator in the horizontal variables. The key point in this formulation is the introduction of the Dirichlet-Neumann operator

$$\mathcal{G}_\mu[\varepsilon\zeta, \beta b] : \psi \mapsto \sqrt{1+\varepsilon^2|\nabla\zeta|^2}\,\partial_{\mathbf{n}}\Phi_{|z=\varepsilon\zeta}, \tag{4.2}$$

where the velocity potential Φ is the solution to the non-dimensionalized elliptic problem

$$\begin{cases} \mu\Delta\Phi + \partial_z^2\Phi = 0 & \text{in } \Omega(\varepsilon\zeta, \beta b), \\ \Phi = \psi(t,\cdot) & \text{on } \{z = \varepsilon\zeta(t,x)\}, \\ \partial_{\mathbf{n}}\Phi = 0 & \text{on } \{z = -1 + \beta b(x)\}; \end{cases} \tag{4.3}$$

in which Δ denotes the Laplace operator in the horizontal variables and $\partial_{\mathbf{n}}$ stands for the outward conormal derivative associated with the elliptic operator $\mu\Delta + \partial_z^2$.

4.1.3 Statement of results and outline of the chapter

As it appears from the above formulation, the derivation of shallow water models is governed by the asymptotic behavior of the Dirichlet-Neumann operator as $\mu \ll 1$. The main task consists in finding, in shallow water regime, an explicit relation between $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ and $\nabla\psi$ through expansion of the Dirichlet-Neumann operator with respect to μ. For smooth topographies, it is known (see Proposition 3.8 of [10]) that

$$\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = -\nabla\cdot((1+\varepsilon\zeta-\beta b)\nabla\psi) + O(\mu).$$

From the previous relation, one may deduce that, up to terms of order $O(\mu)$, the couple $(\zeta, \nabla\psi)$ satisfies the classical Saint-Venant equations

$$\begin{cases} \partial_t\zeta + \nabla\cdot((1+\varepsilon\zeta-\beta b)\nabla\psi) = 0, \\ \partial_t\nabla\psi + \nabla\zeta + \varepsilon(\nabla\psi\cdot\nabla)\nabla\psi = 0. \end{cases} \tag{4.4}$$

Now, in the presence of non smooth topographies, the contribution of the bottom to the first equation of (4.4) may be singular whereas, as regards the full Dirichlet-Neumann operator, the topographic contribution is still infinitely smooth from the ellipticity of the potential equation (4.3).

The main result of this chapter is the construction of an approximation that involves an infinitely smoothing contribution of the bottom, namely

$$\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = -\nabla\cdot\Big((1+\varepsilon\zeta-b_\mu[\beta b])\nabla\psi\Big) + O(\mu),$$

where $b_\mu[\beta b]$ is a regularization operator (defined below). This construction leads to the formal derivation of a nonlocal shallow water system allowing non smooth topographies. Under additional assumptions on ε, we also derive medium and small amplitude models including dispersive effects.

This chapter is organized as follows. Section 4.2 is devoted to the shallow water analysis of the Dirichlet-Neumann operator. Using the fact that this operator depends analytically on ζ and b, we show that the shallow water limit of $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ can be computed using explicit expressions for its shape derivatives with respect to the surface and the bottom parametrizations. This particular construction only involves smoothing contributions of the bottom. Using this asymptotic analysis, we address in section 4.3 the derivation of shallow water models under different sub-regimes, depending on the size of ε (that is the wave amplitude). All these alternative models account for non smooth topographies. The numerical results we present in section 4.4 confirm these alternative models are consistent with the classical shallow water systems in case the bottom parametrization is smooth. Moreover these new model can be used to asses the precision of the classical systems used with rough bottoms. In section 4.5, we present an additional numerical example with a polygonal bottom. In this particular case, the results obtained indicate that our new model is consistent with the approach developed in the previous chapter for polygonal topographies..

4.2 Asymptotic analysis of the Dirichlet-Neumann operator in shallow water regime

In this section, we focus on the asymptotic analysis of the Dirichlet-Neumann operator in shallow water regime. To handle rough bottoms, the strategy we adopt is to bring into play

the shape analyticity of the Dirichlet-Neumann operator, that is to say a Taylor expansion of $\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi$ with respect to the surface and the bottom parametrizations. The shape derivatives, *i.e.* the terms of this Taylor series, can be formally calculated and only give smooth contributions of the bottom. Thus we perform a shallow water limit of $\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi$ which allows rough bottoms by analyzing the asymptotic behavior of these derivatives as $\mu \ll 1$. In particular, attention is payed to check that, in shallow water regime, the higher the order of the shape derivative, the higher order in μ it contributes.

In the present section, the time variable does not play any role so that we drop the dependence on t to simplify notation.

4.2.1 Shape analyticity of the Dirichlet-Neumann operator

The analyticity of the Dirichlet-Neumann operator with respect to the surface elevation has been deeply investigated for the case of a flat bottom (see *e.g.* [26, 35, 39, 38, 78]). In case the topography is non-trivial, the shape analyticity of the Dirichlet-Neumann operator with respect to the surface and the bottom parametrizations has been more recently addressed by, among others, Nicholls and Taber [103] and Lannes [85, Theorem A.11].

4.2.1.1 Taylor expansion of the Dirichlet-Neumann operator in powers of ζ

From the analyticity with respect to the surface, if $\varepsilon\zeta$ lies in a small neighborhood of 0, the Dirichlet-Neumann operator can be expanded as

$$\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi = \sum_{n=0}^{+\infty} \mathcal{G}_\mu^n[\varepsilon\zeta,\beta b]\psi \tag{4.5}$$

where each mapping $\tilde\zeta \mapsto \mathcal{G}_\mu^n[\tilde\zeta,\beta b]\psi$ is homogeneous of degree n[1].

The description of the individual terms in this Taylor series expansion has been first addressed for flat bottoms by Craig and Sulem [40] in two dimensions and a generalization to three dimensions is given in [100]. Introducing an implicit operator $L_\mu[\beta b]$ to take into account the bottom variations, Craig *et al.* [36] showed that this description may be extended to the case of an uneven bottom.

In our non-dimensional framework, the first term of this expansion is given by

$$\mathcal{G}_\mu^0[\varepsilon\zeta,\beta b]\psi = \sqrt{\mu}|D|\tanh(\sqrt{\mu}|D|)\psi + \sqrt{\mu}|D|L_\mu[\beta b]\psi, \tag{4.6}$$

where $D = \frac{\nabla}{i}$ and where we used the Fourier multiplier notation $f(D)u$, defined in terms of Fourier transform by $\widehat{f(D)u} = f\hat{u}$. Then, adapting the computations of [36] to our dimensionless context, we get a similar recursive formula which reads, for the even terms,

$$\begin{aligned}
\mathcal{G}_\mu^{2n}[\varepsilon\zeta,\beta b]\psi =& \frac{\mu^n}{(2n)!} D\cdot\Big\{(\varepsilon\zeta)^{2n} D|D|^{2(n-1)}\mathcal{G}_\mu^0[\varepsilon\zeta,\beta b]\psi\Big\} \\
&- \sum_{p=0}^{n-1} \frac{\mu^{n-p}}{(2(n-p))!}\mathcal{G}_\mu^{2p}[\varepsilon\zeta,\beta b]\Big\{(\varepsilon\zeta)^{2(n-p)}|D|^{2(n-p)}\psi\Big\} \\
&- \sum_{p=0}^{n-1} \frac{\mu^{n-p-1}}{(2(n-p)-1)!}\mathcal{G}_\mu^{2p+1}[\varepsilon\zeta,\beta b]\Big\{(\varepsilon\zeta)^{2(n-p)-1}|D|^{2(n-p-1)}\mathcal{G}_\mu^0[\varepsilon\zeta,\beta b]\psi\Big\},
\end{aligned} \tag{4.7}$$

[1] Denoting by $d_{\tilde\zeta}^n\mathcal{G}_\mu[0,\beta b](\tilde\zeta)\psi$ the n-th derivative of $\zeta\mapsto\mathcal{G}_\mu[\zeta,\beta b]\psi$ at $\zeta=0$ in the direction $\tilde\zeta$, the n-th term in the Taylor expansion (4.5) is related to this shape derivative by $d_{\tilde\zeta}^n\mathcal{G}_\mu[0,\beta b](\tilde\zeta)\psi = n!\mathcal{G}_\mu^n[\tilde\zeta,\beta b]\psi$.

and for the odd terms,

$$\begin{aligned}\mathcal{G}_\mu^{2n+1}[\varepsilon\zeta,\beta b]\psi =& \frac{\mu^{n+1}}{(2n+1)!}D\cdot\Big\{(\varepsilon\zeta)^{2n+1}\,|D|^{2n}\,D\psi\Big\}\\ &-\sum_{p=0}^{n}\frac{\mu^{n-p}}{(2(n-p)+1)!}\mathcal{G}_\mu^{2p}[\varepsilon\zeta,\beta b]\Big\{(\varepsilon\zeta)^{2(n-p)+1}\,|D|^{2(n-p)}\,\mathcal{G}_\mu^0[\varepsilon\zeta,\beta b]\psi\Big\}\\ &-\sum_{p=0}^{n-1}\frac{\mu^{n-p}}{(2(n-p))!}\mathcal{G}_\mu^{2p+1}[\varepsilon\zeta,\beta b]\Big\{(\varepsilon\zeta)^{2(n-p)}\,|D|^{2(n-p)}\,\psi\Big\}.\end{aligned} \tag{4.8}$$

In particular, the linear operator $\mathcal{G}_\mu^1[\varepsilon\cdot,\beta\cdot]\psi$ is given by:

$$\mathcal{G}_\mu^1[\varepsilon\zeta,\beta b]\psi = -\mu\nabla\cdot(\varepsilon\zeta\nabla\psi) - \mathcal{G}_\mu^0[\varepsilon\zeta,\beta b](\varepsilon\zeta\mathcal{G}_\mu^0[\varepsilon\zeta,\beta b]\psi). \tag{4.9}$$

4.2.1.2 Taylor expansion of the Dirichlet-Neumann operator in powers of b

From the analyticity of the Dirichlet-Neumann operator with respect to the bottom parametrization we know that, if βb lies in a sufficiently small neighborhood of 0, the operator $|D|L_\mu[\beta b]\psi$ which stands for the contribution of the bottom in (4.6) can also be expressed as a convergent Taylor series expansion,

$$|D|L_\mu[\beta b]\psi = \sum_{n=0}^{+\infty}|D|L_\mu^n[\beta b]\psi, \tag{4.10}$$

where each mapping $\tilde{b}\mapsto L_\mu^n[\tilde{b}]\psi$ is homogeneous[2] of degree n. In [36], Craig *et al.* obtained a recursion formula for the $L_\mu^n[\beta b]\psi$. Guyenne and Nicholls remarked in [73] that this formula involves a smoothing operator, resulting in smooth contributions of the bottom in (4.10). More precisely, starting from the recursion formula of $|D|L_\mu^n[\beta b]\psi$ (see [36, Eq. (A 8)-(A 9)]), one can see that the individual terms in the Taylor expansion (4.10) take the form

$$\forall n\geq 1,\quad |D|L_\mu^n[\beta b]\psi = \sqrt{\mu}\nabla\cdot\Big\{B_\mu[\beta b]F_\mu^{n-1}[\beta b]\,\mathrm{sech}(\sqrt{\mu}|D|)\nabla\psi\Big\}, \tag{4.11}$$

where the smoothing operator $B_\mu[\beta b]$ is defined by

$$B_\mu[\beta b]\mathbf{v} = \mathrm{sech}(\sqrt{\mu}|D|)(\beta b\mathbf{v}). \tag{4.12}$$

Looking at (4.11), we see that any singular term introduced by the topography when computing $F_\mu^n[\beta b]$ (which is made explicit below) is then regularized using $B_\mu[\beta b]$ so much so that the topographic contribution given by (4.11) is infinitely smooth. Using (4.11), the description of the expansion (4.10) is computed from the following recursion formula for F_μ^n:

(i) for the even terms

$$\begin{aligned}F_\mu^{2n}[\beta b]\mathbf{v} =& \frac{\mu^n}{(2n+1)!}(\beta b)^{2n}\,|D|^{2n}\,\mathbf{v} - \sum_{p=1}^{n}\frac{\sqrt{\mu}^{2p}}{(2p)!}(\beta b)^{2p-1}\,|D|^{2(p-1)}\,D\Big(D\cdot\Big\{\beta b F_\mu^{2(n-p)}[\beta b]\mathbf{v}\Big\}\Big)\\ &+\sum_{p=0}^{n-1}\frac{\sqrt{\mu}^{2p+1}}{(2p+1)!}(\beta b)^{2p}\,|D|^{2p}\,T_\mu[\beta b]\Big\{F_\mu^{2(n-p)-1}[\beta b]\mathbf{v}\Big\},\end{aligned} \tag{4.13}$$

[2] Denoting by $d_b^n\mathcal{G}_\mu[0,0](\tilde{b})\psi$ the n-th derivative of $b\mapsto\mathcal{G}_\mu[0,b]\psi$ at $b=0$ in the direction $\tilde{b}$, the n-th term in the Taylor expansion (4.10) is related to this shape derivative by $d_b^n\mathcal{G}_\mu[0,0](\tilde{b})\psi = n!\sqrt{\mu}|D|L_\mu^n[\tilde{b}]\psi$.

(ii) for the odd terms

$$\begin{aligned} F_\mu^{2n+1}[\beta b]\mathbf{v} = &- \sum_{p=1}^{n} \frac{\sqrt{\mu}^{2p}}{(2p)!} (\beta b)^{2p-1} |D|^{2(p-1)} D\Big(D \cdot \Big\{\beta b F_\mu^{2(n-p)+1}[\beta b]\mathbf{v}\Big\}\Big) \\ &+ \sum_{p=0}^{n} \frac{\sqrt{\mu}^{2p+1}}{(2p+1)!} (\beta b)^{2p} |D|^{2p} T_\mu[\beta b]\Big\{F_\mu^{2(n-p)}[\beta b]\mathbf{v}\Big\}, \end{aligned} \tag{4.14}$$

where $T_\mu[\beta b]$ is defined as

$$T_\mu[\beta b]\mathbf{v} = D\Big(D \cdot \frac{\tanh(\sqrt{\mu}|D|)}{|D|}\{\beta b \mathbf{v}\}\Big). \tag{4.15}$$

Remark 4.1. It is important to note that, even if (4.11) eventually results in infinitely differentiable bottom contributions, it is by no means obvious how to define the operators $bF_\mu^j[\beta b]$ for general $b \in L^\infty(\mathbb{R}^d)$. For instance, computing the product $bF_\mu^1[\beta b]\mathbf{v} = \sqrt{\mu} b T_\mu[\beta b]\mathbf{v}$ requires to assign meaning to $b|D|(b\mathbf{v})$. The latter is of course well defined as soon as b belongs to $W^{1,\infty}(\mathbb{R}^d)$ but it seems much more technical to extend its definition to $L^\infty(\mathbb{R}^d)$ (see also Remark 4.8). More generally, it is known that the Dirichlet-Neumann operator is analytic with respect to Lipschitz deformations of the bottom (see [85, Theorem A.11]). Consequently, it is also possible to extend the definition of the higher order terms to $W^{1,\infty}$ parametrizations of the bottom. Such extensions could be due to non-obvious cancellations in (4.13)-(4.14). That being said, in the numerical simulations of section 4.4, we shall also consider more general topographies for which the present approach gives promising results[3].

4.2.1.3 Analyticity of the Dirichlet-Neumann operator in shallow water regime

As mentioned above, without any assumption on the shallowness parameter μ, both Taylor expansions (4.5) and (4.10) of the Dirichlet-Neumann operator can be written for small perturbations of the surface and the bottom, that is for ε and β small enough. In shallow water regime $\mu \ll 1$, one can roughly estimate from (4.6) and the recursion formulas (4.7) and (4.8) that $\mathcal{G}_\mu^n[\varepsilon\zeta, \beta b]\psi$ is at least of $O(\sqrt{\mu}^{n+1})$. Consequently the series in the right hand side of (4.5) converges (at least formally) without any further condition on ε. Similarly, because $F_\mu^n[\beta b]$ is of $O(\sqrt{\mu}^n)$, the right hand side of (4.10) also converges without any further condition on β. For these reasons, since we study the Dirichlet-Neumann in shallow water conditions, we still use expansion (4.5) and (4.10) to compute $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ whereas no assumption is made on ε and β (we only assume that they are of $O(1)$).

4.2.2 Shallow water expansion of the Dirichlet-Neumann operator

We adopt a formal procedure to derive an expansion of $\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ with respect to μ. The task consists in computing the relevant contributions of each term $\mathcal{G}_\mu^n[\varepsilon\zeta, \beta b]\psi$ from the Taylor series expansion (4.5) with respect to ζ. This is possible thanks to the fact that the order of the contribution in μ of $\mathcal{G}_\mu^n[\varepsilon\zeta, \beta b]\psi$ increases with n. Concerning the contribution of the topography note that, at the formal level, very little regularity is required on b to write the recursive formulation of $|D|L_\mu[\beta b]$. Indeed, as mentioned above, thanks to the

[3] The numerical experiment presented in section 4.5 also indicates that, for a step bottom, the present approach is consistent with the approach developed in chapter 3 for polygonal topographies.

smoothing operator $B_\mu[\beta b]$, each term $|D|L_\mu^n[\beta b]$ computed through (4.11) is well defined and gives smooth functions even for non smooth bottoms (see also remark 4.1). For this reason, allowing non smooth topographies, we use these formulas in order to estimate the shallow water contribution of the bottom *i.e.* the asymptotic behavior of $|D|L_\mu[\beta b]\psi$ as $\mu \to 0$.

To begin with, let us estimate the first contribution from $\mathcal{G}_\mu^0[\varepsilon\zeta, \beta b]\psi$. From (4.6), this term can be expanded as

$$\mathcal{G}_\mu^0[\varepsilon\zeta, \beta b]\psi = \sqrt{\mu}|D|L_\mu[\beta b]\psi + O(\mu).$$

Now, using the relation (4.11) and looking at the recursion formulas (4.13) and (4.14), one sees that $|D|L_\mu[\beta b]\psi$ gives first contributions at $O(\sqrt{\mu})$ so that $\mathcal{G}_\mu^0[\varepsilon\zeta, \beta b]\psi$ gives first contributions at $O(\mu)$. Starting from this and using (4.7) and (4.8), one readily proves by recursion that both $\mathcal{G}_\mu^{2n}[\varepsilon\zeta, \beta b]\psi$ and $\mathcal{G}_\mu^{2n+1}[\varepsilon\zeta, \beta b]\psi$ first contribute at $O(\mu^{n+1})$. Therefore, as we restrict our asymptotic analysis of the water waves equations to $O(\mu)$, we only need to compute the relevant contributions from the first two terms $\mathcal{G}_\mu^0[\varepsilon\zeta, \beta b]\psi$ and $\mathcal{G}_\mu^1[\varepsilon\zeta, \beta b]\psi$. Indeed, the operator $\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ in (4.1) can be formally expanded as

$$\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = \frac{1}{\mu}\Big(\mathcal{G}_\mu^0[\varepsilon\zeta, \beta b]\psi + \mathcal{G}_\mu^1[\varepsilon\zeta, \beta b]\psi\Big) + O(\mu), \tag{4.16}$$

so that one may approximate $\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ up to $O(\mu)$ by expanding both $\mathcal{G}_\mu^0[\varepsilon\zeta, \beta b]\psi$ and $\mathcal{G}_\mu^1[\varepsilon\zeta, \beta b]\psi$ up to $O(\mu^2)$.

In order to determine the contributions from the term involving $\sqrt{\mu}|D|L_\mu[\beta b]\psi$, we use the transformation (4.11) together with the recursion formulas (4.13) and (4.14). From these formulas, $\sqrt{\mu}|D|L_\mu[\beta b]\psi$ can be expanded as

$$\sqrt{\mu}|D|L_\mu[\beta b]\psi = \mu\nabla\cdot B_\mu[\beta b]\Big\{(F_\mu^0[\beta b] + F_\mu^1[\beta b])\,\mathrm{sech}(\sqrt{\mu}|D|)\nabla\psi\Big\} + O(\mu^2).$$

Setting $b_\mu[\beta b] = B_\mu[\beta b]\circ(F_\mu^0[\beta b] + F_\mu^1[\beta b])$ that is

$$b_\mu[\beta b]\mathbf{v} = B_\mu[\beta b]\Big\{\mathbf{v} + \sqrt{\mu}T_\mu[\beta b]\mathbf{v}\Big\},$$

the topographical term writes

$$\sqrt{\mu}|D|L_\mu[\beta b]\psi = \mu\nabla\cdot b_\mu[\beta b]\{\mathrm{sech}(\sqrt{\mu}|D|)\nabla\psi\} + O(\mu^2).$$

Considering the resulting approximation of $\mathcal{G}_\mu^0[\varepsilon\zeta, \beta b]\psi$ in (4.6) and performing a first order Taylor expansion of $\tanh(\sqrt{\mu}|D|)\psi$ and $\mathrm{sech}(\sqrt{\mu}|D|)\nabla\psi$ then lead to

$$\mathcal{G}_\mu^0[\varepsilon\zeta, \beta b]\psi = -\mu\nabla\cdot\Big((1 - b_\mu[\beta b])\nabla\psi\Big) + O(\mu^2). \tag{4.17}$$

The expansion of $\mathcal{G}_\mu^1[\varepsilon\zeta, \beta b]\psi$ in (4.9) follows from the fact that $\mathcal{G}_\mu^0[\varepsilon\zeta, \beta b]\psi$ is of size $O(\mu)$:

$$\mathcal{G}_\mu^1[\varepsilon\zeta, \beta b]\psi = -\mu\nabla\cdot(\varepsilon\zeta\nabla\psi) + O(\mu^2). \tag{4.18}$$

Gathering the last two approximations in (4.16), we finally deduce that

$$\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = -\nabla\cdot\Big((1 + \varepsilon\zeta - b_\mu[\beta b])\nabla\psi\Big) + O(\mu). \tag{4.19}$$

Remark 4.2. In (4.16), the residual is actually of order $O(\varepsilon^2\mu)$ so that if we consider moderate amplitude surface waves *i.e.* $\varepsilon = O(\sqrt{\mu})$, the resulting approximation in (4.16) is precise up to order $O(\mu^2)$. In that case, one may perform an asymptotic analysis up to $O(\mu^2)$ by expanding $\mathcal{G}_\mu^0[\varepsilon\zeta,\beta b]\psi$ and $\mathcal{G}_\mu^1[\varepsilon\zeta,\beta b]\psi$ up to $O(\mu^3)$. This can be achieved by first approximating the topographical term in (4.6) as

$$\sqrt{\mu}|D|L_\mu[\beta b]\psi = \mu\nabla\cdot B_\mu[\beta b]\Big\{(F_\mu^0[\beta b]+F_\mu^1[\beta b]+F_\mu^2[\beta b]+F_\mu^3[\beta b])\,\mathrm{sech}(\sqrt{\mu}|D|)\nabla\psi\Big\}+O(\mu^2).$$

Then, using the recursion formula (4.13)-(4.14) and following the same steps that led to (4.19), one finds that

$$\begin{aligned}\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi =& -\nabla\cdot\Big((1+\varepsilon\zeta-\tilde{b}_\mu[\beta b])\nabla\psi\Big)-\frac{\mu}{3}\nabla\cdot\Delta\nabla\psi+\frac{\mu}{2}\nabla\cdot(b_\mu[\beta b]\Delta\nabla\psi)\\ &+\mu\beta^2\nabla\cdot b_\mu[\beta b]\Big\{-\frac{b^2}{6}\Delta\nabla\psi+\frac{b}{2}\nabla\Big(\nabla\cdot(b\nabla\psi)\Big)\Big\}\\ &+\mu^{3/2}\beta^2\nabla\cdot B_\mu[\beta b]\Big\{-\frac{b^2}{6}\Delta T_\mu[\beta b]\nabla\psi+\frac{b}{2}\nabla\Big(\nabla\cdot(bT_\mu[\beta b]\nabla\psi)\Big)\Big\}\\ &-\mu\varepsilon\nabla\cdot(1-B_\mu[\beta b])\Big\{\nabla(\zeta\nabla\cdot(1-B_\mu[\beta b])\nabla\psi)\Big\}+O(\mu^2,\mu\varepsilon^2),\end{aligned}\tag{4.20}$$

where $\tilde{b}_\mu[\beta b]$ is defined as

$$\tilde{b}_\mu[\beta b]\mathbf{v} = B_\mu[\beta b]\Big\{(1+\sqrt{\mu}T_\mu[\beta b]+\mu T_\mu[\beta b]^2+\mu^{3/2}T_\mu[\beta b]^3)\mathbf{v}\Big\}.\tag{4.21}$$

When no assumption is made on ε, one also needs to compute the relevant contributions from $\mathcal{G}_\mu^2[\varepsilon\zeta,\beta b]\psi$ and $\mathcal{G}_\mu^3[\varepsilon\zeta,\beta b]\psi$ to perform an asymptotic analysis up to $O(\mu^2)$.

4.3 Derivation of shallow water models

This section is devoted to the study of shallow water waves without any regularity assumption on the bottom parametrization. Using the shallow water expansion of the Dirichlet-Neumann operator computed in section 4.2.2, we derive asymptotic models that approximate, in this particular regime, the solutions of the water waves equations

$$\begin{cases}\partial_t\zeta-\dfrac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi=0,\\[2ex] \partial_t\psi+\zeta+\dfrac{\varepsilon}{2}\left|\nabla\psi\right|^2-\varepsilon\mu\dfrac{\left(\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi+\varepsilon\nabla\zeta\cdot\nabla\psi\right)^2}{2(1+\varepsilon^2\mu|\nabla\zeta|^2)}=0.\end{cases}\tag{4.22}$$

In Section 4.3.1, a nonlinear shallow water model is obtained at first order (with respect to μ). Under additional assumptions on ε, asymptotic models with precision $O(\mu^2)$ are then derived in Section 4.3.2 and Section 4.3.3.

These approximate models are written in terms of the surface elevation ζ and the horizontal velocity at the surface $\mathbf{v}_\mathrm{s} = (\nabla\Phi)_{|z=\varepsilon\zeta}$, where we recall that Φ is the velocity potential given by (4.3). The link between $\nabla\psi$ and $\mathbf{v}_\mathrm{s}$ results from the application of the chain rule which yields $\mathbf{v}_\mathrm{s} = \nabla\psi-\varepsilon(\partial_z\Phi)_{|z=\varepsilon\zeta}\nabla\zeta$. Now, by definition of $\mathcal{G}_\mu[\varepsilon\zeta,\beta b]$,

$$(\partial_z\Phi)_{|z=\varepsilon\zeta} = \mu\frac{\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi+\varepsilon\nabla\psi\cdot\nabla\zeta}{1+\mu\varepsilon^2\left|\nabla\zeta\right|^2},$$

so that the horizontal velocity at the surface can be expressed as

$$\mathbf{v}_{\mathrm{s}} = \nabla\psi - \mu\varepsilon \frac{\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi + \varepsilon\nabla\psi\cdot\nabla\zeta}{1+\mu\varepsilon^2|\nabla\zeta|^2}\nabla\zeta. \tag{4.23}$$

To achieve the formal derivation of an approximate model with precision $O(\mu^k)$ ($k = 1$ or 2), the strategy we adopt in the next two sections is to take the following steps:

1. From (4.23), find an asymptotic expansion of order $O(\mu^k)$ of $\nabla\psi$ in terms of the velocity $\mathbf{v}_{\mathrm{s}}$;
2. Plug this expansion into (4.19) (or (4.20)) to get an asymptotic expansion of the Dirichlet-Neumann operator in terms of $\mathbf{v}_{\mathrm{s}}$;
3. In the first equation of (4.22), replace $-\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi$ by the approximation of step (2) and neglect the terms of order $O(\mu^k)$;
4. Take the gradient of the second equation of (4.22), insert the expansions from steps (1) and (2) and neglect the $O(\mu^k)$ terms.

4.3.1 The nonlinear shallow water equations for non smooth bottoms

Following the steps 1–4 above, we derive a nonlocal shallow water approximation of (4.22) at order $O(\mu)$. To this end, let us first remark that, at first order, $\mathbf{v}_{\mathrm{s}} = \nabla\psi + O(\mu)$. Plugging this last expansion in (4.19) we get

$$\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = -\nabla\cdot\Big((1+\varepsilon\zeta - b_\mu[\beta b])\mathbf{v}_{\mathrm{s}}\Big) + O(\mu), \tag{4.24}$$

where we recall that the smoothing operator $b_\mu[\beta b]$ is defined as

$$b_\mu[\beta b]\mathbf{v} = B_\mu[\beta b]\Big\{(1+\sqrt{\mu}T_\mu[\beta b])\mathbf{v}\Big\}, \tag{4.25}$$

with

$$B_\mu[\beta b]\mathbf{v} = \mathrm{sech}(\sqrt{\mu}|D|)(\beta b\mathbf{v}) \quad \text{and} \quad T_\mu[\beta b]\mathbf{v} = D\Big(D\cdot\frac{\tanh(\sqrt{\mu}|D|)}{|D|}\{\beta b\mathbf{v}\}\Big). \tag{4.26}$$

Therefore, substituting expansion (4.24) into the first equation of (4.22), then applying ∇ to the second equation and using both $\mathbf{v}_{\mathrm{s}} = \nabla\psi + O(\mu)$ and $\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta, \beta b]\psi = O(1)$, we obtain the following nonlocal approximate equations of motion up to terms of order $O(\mu)$

$$\begin{cases} \partial_t\zeta + \nabla\cdot((1+\varepsilon\zeta - b_\mu[\beta b])\mathbf{v}_{\mathrm{s}}) = 0, \\ \partial_t\mathbf{v}_{\mathrm{s}} + \nabla\zeta + \varepsilon(\mathbf{v}_{\mathrm{s}}\cdot\nabla)\mathbf{v}_{\mathrm{s}} = 0. \end{cases} \tag{4.27}$$

Remark 4.3. The classical shallow water approximation of (4.22) can be written

$$\begin{cases} \partial_t\zeta + \nabla\cdot((1+\varepsilon\zeta - \beta b)\mathbf{v}_{\mathrm{s}}) = 0, \\ \partial_t\mathbf{v}_{\mathrm{s}} + \nabla\zeta + \varepsilon(\mathbf{v}_{\mathrm{s}}\cdot\nabla)\mathbf{v}_{\mathrm{s}} = 0. \end{cases} \tag{4.28}$$

Hence in case the bottom parametrization is not regular, the alternative shallow water model (4.27) differs from the classical approximation by the presence of a regularized discharge,

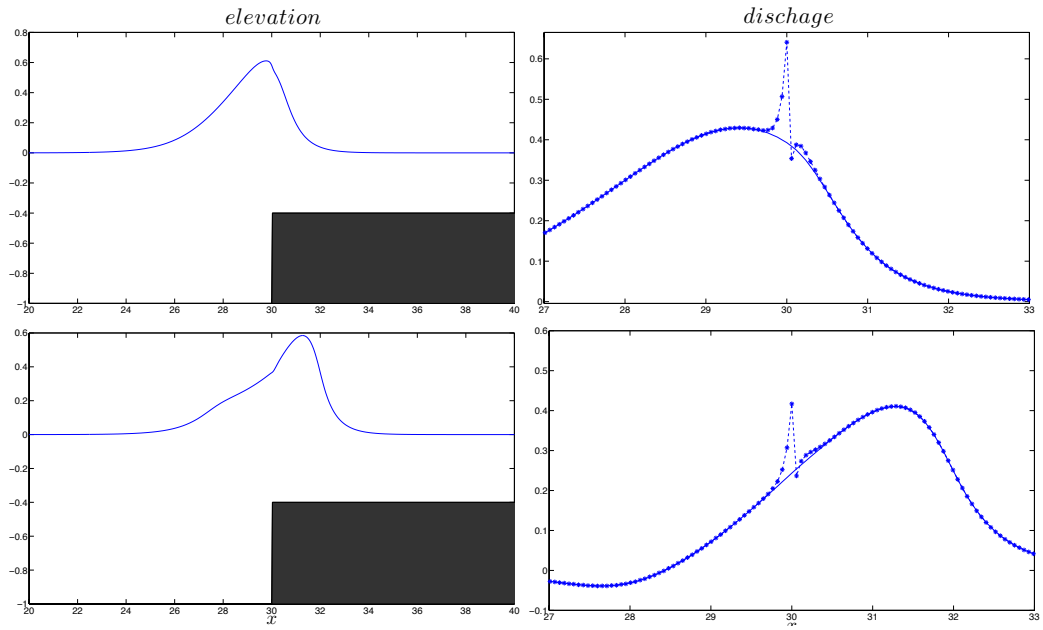

Figure 4.2 – Wave passing over a step. Left: elevation. Right: comparison of the regularized discharge (—) with the classical discharge (--*--).

namely $q_\mu = (1 + \varepsilon\zeta - b_\mu[\beta b])\mathbf{v}_s$, instead of the classical discharge $q = (1 + \varepsilon\zeta - \beta b)\mathbf{v}_s$. An illustration of this regularizing effect is given in Figure 4.2. It is also worth mentioning that the water depth variable $h = 1 + \varepsilon\zeta + \beta b$ has no regularized analogous in the present alternative shallow water model. Indeed, one may feel inclined to define a regularized water depth as $h_\mu = 1 + \varepsilon\zeta - b_\mu[\beta b]$. However, this last expression does not define a function but an operator (precisely because $b_\mu[\beta b]$ is an operator). As a particular consequence, unlike the Saint-Venant equations which can be formulated in (h, q) variables instead of (ζ, v), the alternative shallow water model can neither be formulated in terms of the water depth h (which may be singular for non smooth bottoms) nor in terms of the quantity h_μ (which is not a function).

Remark 4.4 (Smooth bottoms). In case the bottom parametrization is regular, a Taylor expansion of $T_\mu[\beta b]$ and $B_\mu[\beta b]$ in (4.25) ensures that

$$b_\mu[\beta b]\mathbf{v}_s = b\mathbf{v}_s + O(\mu).$$

Using this last approximation in (4.27), one recovers the classical Saint-Venant system (4.28) from the alternative equations (4.27).

Remark 4.5 (Dimensionalized equations). Going back to variables with dimensions, the nonlocal shallow water system reads

$$\begin{cases} \partial_t\zeta + \nabla\cdot((H_0 + \zeta - b_\mu[b])\mathbf{v}_s) = 0, \\ \partial_t\mathbf{v}_s + g\nabla\zeta + (\mathbf{v}_s\cdot\nabla)\mathbf{v}_s = 0, \end{cases}$$

where the dimensionalized smoothing operator $b_\mu[b]$ is now given by

$$b_\mu[b]\mathbf{v} = B[b]\Big\{(1 + T[b])\mathbf{v}\Big\},$$

with

$$B[b]\mathbf{v} = \mathrm{sech}(H_0|D|)(b\mathbf{v}) \quad \text{and} \quad T[b]\mathbf{v} = D\Big(D\cdot\frac{\tanh(H_0|D|)}{|D|}\{b\mathbf{v}\}\Big).$$

4.3.2 Medium amplitude models ($\varepsilon = O(\sqrt{\mu})$) for non smooth bottoms

In this section, besides the shallow water hypothesis, we assume that the amplitude parameter ε is of size $O(\sqrt{\mu})$. In case the bottom is smooth, this regime leads to the medium amplitude Green-Naghdi or Serre equations (see *e.g.* [85] for the derivation of these equations).

4.3.2.1 Derivation of an approximate model with precision $O(\mu^2)$

Under the previous assumption on ε, let us follow the steps (1)-(4) given above to derive an approximate model with precision $O(\mu^2)$. Since

$$\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi = -\nabla\cdot(1-B_\mu[\beta b])\mathbf{v}_\mathrm{s} + O(\sqrt{\mu}), \tag{4.29}$$

we can see from relation (4.23) that

$$\nabla\psi = \mathbf{v}_\mathrm{s} - \varepsilon\mu\nabla\cdot(1-B_\mu[\beta b])\mathbf{v}_\mathrm{s}\nabla\zeta + O(\mu^2). \tag{4.30}$$

Plugging this last approximation in (4.20), we obtain

$$\begin{aligned}\frac{1}{\mu}\mathcal{G}_\mu[\varepsilon\zeta,\beta b]\psi = -\nabla\cdot\Big((1+\varepsilon\zeta-\tilde{b}_\mu[\beta b])\mathbf{v}_\mathrm{s}\Big) - \frac{\mu}{3}\Delta\nabla\cdot\mathbf{v}_\mathrm{s} + \frac{\mu}{2}S_\mu[\beta b]\mathbf{v}_\mathrm{s}\\ - \mu\varepsilon\nabla\cdot(1-B_\mu[\beta b])\Big\{\zeta\nabla(\nabla\cdot(1-B_\mu[\beta b])\mathbf{v}_\mathrm{s})\Big\} + O(\mu^2),\end{aligned}$$

where we recall that $B_\mu[\beta b]$ and $\tilde{b}_\mu[\beta b]$ are respectively defined in (4.12) and (4.21), while the dispersive topographical term $S_\mu[\beta b]$ is defined as

$$\begin{aligned}S_\mu[\beta b]\mathbf{v} = \nabla\cdot(b_\mu[\beta b]\Delta\mathbf{v}) + \beta^2\nabla\cdot b_\mu[\beta b]\Big\{-\frac{b^2}{3}\Delta\mathbf{v} + b\nabla\Big(\nabla\cdot(b\mathbf{v})\Big)\Big\}\\ + \sqrt{\mu}\beta^2\nabla\cdot B_\mu[\beta b]\Big\{-\frac{b^2}{3}\Delta T_\mu[\beta b]\mathbf{v} + b\nabla\Big(\nabla\cdot(bT_\mu[\beta b]\mathbf{v})\Big)\Big\}.\end{aligned} \tag{4.31}$$

Inserting this last expansion in the first equation of (4.22) yields an evolution equation for the free surface up to $O(\mu^2)$ terms:

$$\begin{aligned}\partial_t\zeta + \nabla\cdot\Big((1+\varepsilon\zeta-\tilde{b}_\mu[\beta b])\mathbf{v}_\mathrm{s}\Big) + \frac{\mu}{3}\Delta\nabla\cdot\mathbf{v}_\mathrm{s} - \frac{\mu}{2}S_\mu[\beta b]\mathbf{v}_\mathrm{s}\\ + \mu\varepsilon\nabla\cdot(1-B_\mu[\beta b])\Big\{\zeta\nabla(\nabla\cdot(1-B_\mu[\beta b])\mathbf{v}_\mathrm{s})\Big\} = 0.\end{aligned} \tag{4.32}$$

Concerning the evolution of the velocity unknown we follow step (4) and take the gradient of the second equation (4.22). On using (4.29) and (4.30), the result is

$$\partial_t\Big((1-\mu\varepsilon A_\mu[\zeta,\beta b])\mathbf{v}_\mathrm{s}\Big) + \nabla\zeta + \varepsilon(\mathbf{v}_\mathrm{s}\cdot\nabla)\mathbf{v}_\mathrm{s} - \frac{\mu\varepsilon}{2}\nabla\Big(\nabla\cdot(1-B_\mu[\beta b])\mathbf{v}_\mathrm{s}\Big)^2 = O(\mu^2), \tag{4.33}$$

where $A_\mu[\zeta,\beta b]$ is defined as $A_\mu[\zeta,\beta b]\mathbf{v} = \nabla\zeta\nabla\cdot(1-B_\mu[\beta b])\mathbf{v}$. Now since, from (4.32), $\partial_t\zeta$ may be approximated as $\partial_t\zeta = -\nabla\cdot(1-B_\mu[\beta b])\mathbf{v}_\mathrm{s} + O(\sqrt{\mu})$, we get

$$\partial_t\Big((1-\mu\varepsilon A_\mu[\zeta,\beta b])\mathbf{v}_\mathrm{s}\Big) = \Big(1-\mu\varepsilon A_\mu[\zeta,\beta b]\Big)\partial_t\mathbf{v}_\mathrm{s} + \frac{\mu\varepsilon}{2}\nabla\Big(\nabla\cdot(1-B_\mu[\beta b])\mathbf{v}_\mathrm{s}\Big)^2 + O(\mu^2). \tag{4.34}$$

Gathering (4.34) and (4.33) leads to

$$(1-\mu\varepsilon A_\mu[\zeta,\beta b])\partial_t \mathbf{v}_\mathrm{s}+\nabla\zeta+\varepsilon(\mathbf{v}_\mathrm{s}\cdot\nabla)\mathbf{v}_\mathrm{s}=O(\mu^2),$$

from which we deduce the following approximate evolution equation for the velocity up to $O(\mu^2)$ terms

$$\partial_t \mathbf{v}_\mathrm{s}+\nabla\zeta+\varepsilon(\mathbf{v}_\mathrm{s}\cdot\nabla)\mathbf{v}_\mathrm{s}+\mu\varepsilon\nabla\zeta\nabla\cdot(1-B_\mu[\beta b])\nabla\zeta=0. \tag{4.35}$$

4.3.2.2 Improving the frequency dispersion of the model

Equations (4.32) and (4.35) only differ by nonlinear terms from the Boussinesq model formulated in terms of the velocity at the surface (see *e.g.* [108, Eq. (16) and (17)]). Now it is known that the latter is linearly ill-posed (see *e.g.* [19]), and so are the former. Indeed, the existence of non-trivial solutions (ζ,ψ) of the form $(\zeta_0,\psi_0)\mathrm{e}^{i(\mathbf{k}\cdot x-\omega t)}$ to the linearization of (4.32)-(4.35) around $\zeta=0$, $\nabla\psi=0$ and for flat bottom $b=0$ requires the dispersion relation

$$\omega_\alpha(\mathbf{k})^2=|\mathbf{k}|^2-\frac{\mu}{3}|\mathbf{k}|^4,$$

and this relation does not lead to real-valued frequencies $\omega_\alpha(\mathbf{k})$ for high wave numbers $|\mathbf{k}|$. To improve the linear dispersion frequencies of the model one can use the BBM "trick" [15]. The idea is to note that since $\partial_t\zeta+\nabla\cdot((1+\varepsilon\zeta-b_\mu[\beta b])\mathbf{v}_\mathrm{s})$ is of size $O(\mu)$, we can introduce a real parameter α by adding the quantity

$$-\mu\frac{\alpha}{3}(\Delta\partial_t\zeta+\Delta\nabla\cdot\mathbf{v}_\mathrm{s}+\varepsilon\Delta\nabla\cdot(\zeta\mathbf{v}_\mathrm{s})-\Delta\nabla\cdot b_\mu[\beta b]\mathbf{v}_\mathrm{s})=O(\mu^2)$$

to (4.32). The resulting approximate equation, together with (4.35), yields the following asymptotic model with precision $O(\mu^2)$ for shallow water medium amplitude waves

$$\begin{cases}(1-\dfrac{\mu\alpha}{3}\Delta)\partial_t\zeta+\nabla\cdot\Big((1+\varepsilon\zeta-\tilde{b}_\mu[\beta b])\mathbf{v}_\mathrm{s}\Big)+\dfrac{\mu}{3}(1-\alpha)\Delta\nabla\cdot\mathbf{v}_\mathrm{s}-\dfrac{\mu}{2}S_\mu[\beta b]\mathbf{v}_\mathrm{s}\\ \quad+\mu\varepsilon\nabla\cdot(1-B_\mu[\beta b])\Big\{\zeta\nabla(\nabla\cdot(1-B_\mu[\beta b])\mathbf{v}_\mathrm{s})\Big\}+\mu\dfrac{\alpha}{3}\Delta\nabla\cdot b_\mu[\beta b]\mathbf{v}_\mathrm{s}-\mu\varepsilon\dfrac{\alpha}{3}\Delta\nabla\cdot(\zeta\mathbf{v}_\mathrm{s})=0,\\ \partial_t\mathbf{v}_\mathrm{s}+\nabla\zeta+\varepsilon(\mathbf{v}_\mathrm{s}\cdot\nabla)\mathbf{v}_\mathrm{s}+\mu\varepsilon\nabla\zeta\nabla\cdot(1-B_\mu[\beta b])\nabla\zeta=0.\end{cases} \tag{4.36}$$

The dispersion relation associated to (4.36) now reads

$$\omega_\alpha(\mathbf{k})^2=|\mathbf{k}|^2\,\frac{1+\frac{\alpha-1}{3}\mu|\mathbf{k}|^2}{1+\frac{\alpha}{3}\mu|\mathbf{k}|^2}.$$

Consequently, the interest of the parameter α is that the corresponding system (4.36) is linearly well-posed as soon as $\alpha\geq 1$. Moreover this parameter can be adjusted (see for instance [34, 33]) to improve the dispersive characteristics embedded in the medium amplitude model (4.36). To this end, we set $\alpha=1.159$ in all the numerical tests of section 4.4.2.2. Following [34], this value has been chosen so that the phase and group velocities associated to (4.36) stay close to the reference velocities coming from the water waves equations (4.22). More precisely, this value minimizes the following joint error over the range $0\leq\sqrt{\mu}|k|\leq 3$:

$$\frac{\int_0^3(C_\alpha^p-C_{\mathrm{WW}}^p)^2\,\mathrm{d}\sqrt{\mu}|\mathbf{k}|}{\int_0^3(C_{\mathrm{WW}}^p)^2\,\mathrm{d}\sqrt{\mu}|\mathbf{k}|}+\frac{\int_0^3(C_\alpha^g-C_{\mathrm{WW}}^g)^2\,\mathrm{d}\sqrt{\mu}|\mathbf{k}|}{\int_0^3(C_{\mathrm{WW}}^g)^2\,\mathrm{d}\sqrt{\mu}|\mathbf{k}|}$$

where the phase and group velocities associated with (4.36) are defined as

$$C_\alpha^p(\sqrt{\mu}|\mathbf{k}|) = \frac{\omega_\alpha(\mathbf{k})}{|\mathbf{k}|} = \left(\frac{1+\frac{\alpha-1}{3}\mu|\mathbf{k}|^2}{1+\frac{\alpha}{3}\mu|\mathbf{k}|^2}\right)^{1/2},$$

and

$$C_\alpha^g(\sqrt{\mu}|\mathbf{k}|) = \frac{\mathrm{d}\omega_\alpha(\mathbf{k})}{\mathrm{d}|\mathbf{k}|} = \frac{1+\frac{2(\alpha-1)}{3}\mu|\mathbf{k}|^2+\frac{\alpha(\alpha-1)}{9}\mu^2|\mathbf{k}|^4}{\left(1+\frac{\alpha}{3}\mu|\mathbf{k}|^2\right)^{3/2}\left(1+\frac{\alpha-1}{3}\mu|\mathbf{k}|^2\right)^{1/2}},$$

while C_{WW}^p and C_{WW}^g represent respectively the phase and group velocities associated with the full water waves equations[4].

Remark 4.6. When the bottom is smooth, further improvements of the dispersive properties can be achieved by replacing the velocity variable $\mathbf{v}_\mathrm{s}$ at the surface with a different velocity variable linked to the velocity at an arbitrary elevation. The velocity at a certain depth is used in [105, 115] as dependent variable while a slightly different choice is made in [33] with the introduction of a new dependent variable (which is also related to the velocity at an arbitrary elevation as explained in [85]). In the present case, since we are dealing with rough bottoms, we decide to work with variables located at the surface where the irregularities of the bottom have the least effect.

4.3.3 A Boussinesq system for non smooth bottoms

The additional small amplitude assumption $\varepsilon = O(\mu)$ is the traditional assumption that leads to the usual Boussinesq models. In this particular regime, neglecting in (4.36) the terms of order $O(\varepsilon\mu) = O(\mu^2)$ yields the following Boussinesq-type approximation of the water waves equations (4.22) with precision $O(\mu^2)$

$$\begin{cases} (1-\dfrac{\mu\alpha}{3}\Delta)\partial_t\zeta + \nabla\cdot\Big((1+\varepsilon\zeta-\tilde{b}_\mu[\beta b])\mathbf{v}_\mathrm{s}\Big) \\ \qquad\qquad + \dfrac{\mu}{3}(1-\alpha)\Delta\nabla\cdot\mathbf{v}_\mathrm{s} - \dfrac{\mu}{2}S_\mu[\beta b]\mathbf{v}_\mathrm{s} + \mu\dfrac{\alpha}{3}\Delta\nabla\cdot b_\mu[\beta b]\mathbf{v}_\mathrm{s} = 0, \\ \partial_t\mathbf{v}_\mathrm{s} + \nabla\zeta + \varepsilon(\mathbf{v}_\mathrm{s}\cdot\nabla)\mathbf{v}_\mathrm{s} = 0, \end{cases} \tag{4.37}$$

where $S_\mu[\beta b]$ is defined as in (4.31).

Remark 4.7 (Smooth bottoms). Using the velocity at the surface as dependent variable, the usual Boussinesq model derived by Peregrine for smooth bottoms (see [108]) can be written

$$\begin{cases} (1-\dfrac{\mu\alpha}{3}\Delta)\partial_t\zeta + \nabla\cdot(h_b\mathbf{v}_\mathrm{s}) + \dfrac{\mu}{3}\nabla\cdot(h_b^3\nabla(\nabla\cdot\mathbf{v}_\mathrm{s})) - \mu\dfrac{\alpha}{3}\Delta\nabla\cdot\mathbf{v}_\mathrm{s} \\ \qquad - \dfrac{\mu\beta}{2}\nabla\cdot(h_b^2\nabla(\nabla b\cdot\mathbf{v}_\mathrm{s})) - \dfrac{\mu\beta}{2}\nabla\cdot(h_b^2\nabla\cdot\mathbf{v}_\mathrm{s}\nabla b) + \mu\beta\dfrac{\alpha}{3}\Delta\nabla\cdot(b\mathbf{v}_\mathrm{s}) = 0, \\ \partial_t\mathbf{v}_\mathrm{s} + \nabla\zeta + \varepsilon(\mathbf{v}_\mathrm{s}\cdot\nabla)\mathbf{v}_\mathrm{s} = 0, \end{cases} \tag{4.38}$$

where $h_b = 1-\beta b$ stands for the (nondimensional) still water depth. Assuming that the bottom is smooth, a Taylor expansion of both operators $T_\mu[\beta b]$ and $B_\mu[\beta b]$ in (4.21) ensures that

$$\tilde{b}_\mu[\beta b]\mathbf{v}_\mathrm{s} = \beta b\mathbf{v}_\mathrm{s} + \frac{\mu\beta}{2}\Delta(b\mathbf{v}_\mathrm{s}) - \mu\beta^2 b\nabla(\nabla\cdot(b\mathbf{v}_\mathrm{s})) + O(\mu^2).$$

[4] The dispersion relation associated to the (nondimensionalized) water waves equations (4.22) reads $\omega_{\mathrm{WW}}^2(\mathbf{k}) = |\mathbf{k}|^2\,\frac{\tanh(\sqrt{\mu}|\mathbf{k}|)}{\sqrt{\mu}|\mathbf{k}|}$.

Using this last expansion together with $b_\mu[\beta b] = \beta b \mathbf{v}_\mathrm{s} + O(\mu)$ in the first equation of (4.37) while keeping in mind that, as the bottom is smooth, $T_\mu[\beta b]$ gives first contributions at $O(\sqrt{\mu})$, one can check that this equation coincides with the first equation of (4.38) up to terms of order $O(\mu^2)$.

4.4 Numerical computations

In this section we describe spatial discretization and time integration of the nonlocal shallow water models derived in the previous section. Then we present some numerical simulations in order to illustrate the behavior of these asymptotic models.

4.4.1 Numerical scheme

The numerical simulations are made in the one dimensional case $d = 1$. In that case, the nonlocal operators $B_\mu[\beta b]$ and $T_\mu[\beta b]$ that occur in the definitions of $b_\mu[\beta b]$ and $\tilde{b}_\mu[\beta b]$ are given by

$$B_\mu[\beta b]v = \mathrm{sech}(\sqrt{\mu}|D|)(\beta b v) \quad \text{and} \quad T_\mu[\beta b]v = |D|\tanh(\sqrt{\mu}|D|)(\beta b v). \tag{4.39}$$

4.4.1.1 Numerical scheme for the nonlinear shallow water equations

In the one dimensional case, equations (4.27) reduce to

$$\begin{cases} \partial_t \zeta + \partial_x v_\mathrm{s} + \varepsilon \partial_x (v_\mathrm{s} \zeta) = \partial_x (b_\mu[\beta b] v_\mathrm{s}), \\ \partial_t v_\mathrm{s} + \partial_x \zeta + \varepsilon v_\mathrm{s} \partial_x v_\mathrm{s} = 0, \end{cases} \tag{4.40}$$

Time integration. Following the previous work of Besse and Bruneau, we use a Crank-Nicolson like scheme where the nonlinear part is avoided by doing a relaxation that is by writing the linear and the nonlinear parts to different times (see [18, 17] for a description of the method and *e.g.* [32, 58, 57] for applications to asymptotic models related to the water waves equations). More precisely, given a time step Δt, we consider functions (ζ^n, v^n) which approximate $\zeta(t^n, \cdot)$ and $v_\mathrm{s}(t^n, \cdot)$ at time $t^n = n\Delta t$ and $v^{n+\frac{1}{2}}$ which approximate $v_\mathrm{s}(t^{n+\frac{1}{2}}, \cdot)$ at $t^{n+\frac{1}{2}} = (n + \frac{1}{2})\Delta t$. Then the semi-discretized in time scheme for (4.27) reads, for all $n \geq 1$,

$$v^n = \frac{v^{n+\frac{1}{2}} + v^{n-\frac{1}{2}}}{2}$$

and

$$\begin{cases} \dfrac{\zeta^{n+1} - \zeta^n}{\Delta t} + \varepsilon \partial_x \left(v^{n+\frac{1}{2}} \dfrac{\zeta^{n+1} + \zeta^n}{2} \right) = \partial_x (b_\mu[\beta b] v^{n+\frac{1}{2}}) - \partial_x v^{n+\frac{1}{2}}, \\ \dfrac{v^{n+1} - v^n}{\Delta t} + \partial_x \left(\dfrac{\zeta^{n+1} + \zeta^n}{2} \right) + \varepsilon v^{n+\frac{1}{2}} \partial_x \left(\dfrac{v^{n+1} + v^n}{2} \right) = 0. \end{cases}$$

Spatial discretization. In all the test cases, the (one-dimensional) spatial domain is $(0, L)$. We assume periodic boundary conditions so that the nonlocal operator $b_\mu[\beta b]$ can be approximated using the discrete Fourier transform. This amounts to evaluating all the differential operators in (4.39) in Fourier space while performing nonlinear products in physical space. More precisely, if Δx is a spatial step (chosen such that $N = \frac{L}{\Delta x}$ is an integer), the spatial

domain is discretized by N equally spaced points $x_j = j\Delta x$, $j = 1,\dots,N$, and the corresponding discrete frequencies are given by $\mathbf{k} = \frac{2\pi}{L}\{-\frac{N}{2}+1,\dots,\frac{N}{2}\}$. Then, if we wish to evaluate the discrete analogue of $B_\mu[\beta b]$ applied to a discrete function $u = (u_j)_{1\le j\le N}$, we first multiply u by $(b(x_j))_{1\le j\le N}$, then transform to the Fourier space (using fast Fourier transform), multiply by the diagonal operator $\mathrm{sech}(\sqrt{\mu}\mathbf{k})$ and finally transform back to the physical space. Approximations of other such terms in (4.25) is achieved similarly which leads to a discrete approximation $b_\mu^{\Delta x}:\mathbb{R}^N\to\mathbb{R}^N$ of the nonlocal operator $b_\mu[\beta b]$.

Thus considering discrete unknowns $\zeta^n = (\zeta_j^n)_{1\le j\le N}$ and $v^n = (v_j^n)_{1\le j\le N}$ at time t^n and $v^{n+\frac12} = (v_j^{n+\frac12})_{1\le j\le N}$ at time $t^{n+\frac12}$, the fully discrete scheme reads, for all $n\ge 1$, $v^{n+\frac12} = 2v^n - v^{n-\frac12}$ and

$$\begin{cases} \dfrac{\zeta^{n+1}-\zeta^n}{\Delta t} + \varepsilon D_1\Big(v^{n+\frac12}\dfrac{\zeta^{n+1}+\zeta^n}{2}\Big) = D_1(b_\mu^{\Delta x}v^{n+\frac12}) - D_1 v^{n+\frac12}, \\ \dfrac{v^{n+1}-v^n}{\Delta t} + D_1\Big(\dfrac{\zeta^{n+1}+\zeta^n}{2}\Big) + \varepsilon v^{n+\frac12}D_1\Big(\dfrac{v^{n+1}+v^n}{2}\Big) = 0, \end{cases} \tag{4.41}$$

where D_1 stands for the classical centered discretization of ∂_x (with periodic boundary conditions). When comparing both asymptotic models (4.27) and (4.28), we use for the classical shallow water model (4.28) a finite difference scheme similar in principle to that described above for the alternative model.

4.4.1.2 Numerical scheme for the medium and small amplitude models

The one dimensional version of the medium amplitude model (4.36) reads

$$\begin{cases} (1-\dfrac{\mu\alpha}{3}\partial_x^2)\partial_t\zeta + \varepsilon(\partial_x - \dfrac{\mu\alpha}{3}\partial_x^3)(v_{\rm s}\zeta) + \dfrac{\mu}{3}(1-\alpha)\partial_x^3 v_{\rm s} \\ \qquad = \partial_x\tilde b_\mu[\beta b]v_{\rm s} - \partial_x v_{\rm s} + \dfrac{\mu}{2}S_\mu[\beta b]v_{\rm s} - \mu\dfrac{\alpha}{3}\partial_x^3 b_\mu[\beta b]v_{\rm s} \\ \qquad\quad - \mu\varepsilon\partial_x(1-B_\mu[\beta b])\Big\{\zeta\partial_x^2(1-B_\mu[\beta b])v_{\rm s}\Big\}, \\ \partial_t v_{\rm s} + \partial_x\zeta + \varepsilon v_{\rm s}\partial_x v_{\rm s} + \mu\varepsilon\partial_x((1-B_\mu[\beta b])\partial_x\zeta)\partial_x\zeta = 0, \end{cases} \tag{4.42}$$

and the dispersive topographical contribution $S_\mu[\beta b]$ is given by

$$\begin{aligned} S_\mu[\beta b]v = {} & \partial_x b_\mu[\beta b]\partial_x^2 v + \beta^2\partial_x b_\mu[\beta b]\Big\{-\frac{b^2}{3}\partial_x^2 v + b\partial_x^2(bv)\Big\} \\ & + \sqrt{\mu}\beta^2\partial_x B_\mu[\beta b]\Big\{-\frac{b^2}{3}\partial_x^2 T_\mu[\beta b]v + b\partial_x^2(bT_\mu[\beta b]v)\Big\}. \end{aligned}$$

Time integration is achieved using the aforementioned Crank-Nicolson like scheme. Concerning spatial discretization, we use discrete Fourier transform as described above to approximate each nonlocal operator that appears in (4.42). Thus the fully discrete scheme

reads, for all $n \geq 1$,

$$\begin{cases} (I - \dfrac{\mu\alpha}{3} D_2)\dfrac{\zeta^{n+1} - \zeta^n}{\Delta t} + \varepsilon(D_1 - \dfrac{\mu\alpha}{3} D_3)\Big(v^{n+\frac{1}{2}} \dfrac{\zeta^{n+1} + \zeta^n}{2}\Big) + \dfrac{\mu}{3}(1-\alpha) D_3\Big(\dfrac{v^{n+1} + v^n}{2}\Big) \\ \qquad = D_1(\tilde{b}_\mu^{\Delta x} v^{n+\frac{1}{2}}) - D_1 v^{n+\frac{1}{2}} + \dfrac{\mu}{2} S_\mu^{\Delta x} v^{n+\frac{1}{2}} - \mu \dfrac{\alpha}{3} D_3(b_\mu^{\Delta x} v^{n+\frac{1}{2}}) \\ \qquad - \mu\varepsilon D_1(1 - B_\mu^{\Delta x})\Big\{\zeta^{n+\frac{1}{2}} D_2(1 - B_\mu^{\Delta x}) v^{n+\frac{1}{2}}\Big\}, \\ \dfrac{v^{n+1} - v^n}{\Delta t} + D_1\Big(\dfrac{\zeta^{n+1} + \zeta^n}{2}\Big) + \varepsilon v^{n+\frac{1}{2}} D_1\Big(\dfrac{v^{n+1} + v^n}{2}\Big) \\ \qquad = \mu\varepsilon D_1((B_\mu^{\Delta x} - 1) D_1 \zeta^{n+\frac{1}{2}}) D_1\Big(\dfrac{\zeta^{n+1} + \zeta^n}{2}\Big), \end{cases} \tag{4.43}$$

where D_1, D_2 and D_3 stand for the classical centered discretizations of ∂_x, ∂_x^2 and ∂_x^3 while $B_\mu^{\Delta x}$, $b_\mu^{\Delta x}$, $\tilde{b}_\mu^{\Delta x}$ and $S_\mu^{\Delta x}$ are respectively the discrete approximations of the nonlocal operators $B_\mu[\beta b]$, $b_\mu[\beta b]$, $\tilde{b}_\mu[\beta b]$ and $S_\mu[\beta b]$. The one-dimensional version of the Boussinesq-like system (4.37) is similarly approximated.

4.4.2 Numerical results

Our goal in the computations presented in this chapter is to compare the results produced by the nonlocal shallow water systems for rough bottom derived in Section 4.3 with the ones obtained from the classical shallow water models.

All simulations have been performed using $N = 1024$ points and $\Delta t = 10^{-2}$. In all the test cases, the initial condition (ζ_0, v_0) consists of a unidirectional wave propagating to the right on a domain of length $L = 60$:

$$\zeta_0(x) = v_0(x) = a\,\mathrm{sech}^2\Big(\frac{x - 20}{2}\Big), \quad 0 < x < 60, \tag{4.44}$$

where a is an arbitrary parameter. The bathymetry can be parametrized as follows

$$b(x) = \frac{1}{2}\Big(\tanh\Big(2\Big(\frac{x - 30}{\delta}\Big)\Big) - \tanh(x - 49)\Big), \quad 0 < x < 60. \tag{4.45}$$

This parametrization is regular but it involves a slope of order $\frac{1}{\delta}$ around $x = 30$. Then as $\delta \ll 1$ this slope becomes steep and the corresponding bottom becomes rough (see Figure 4.3).

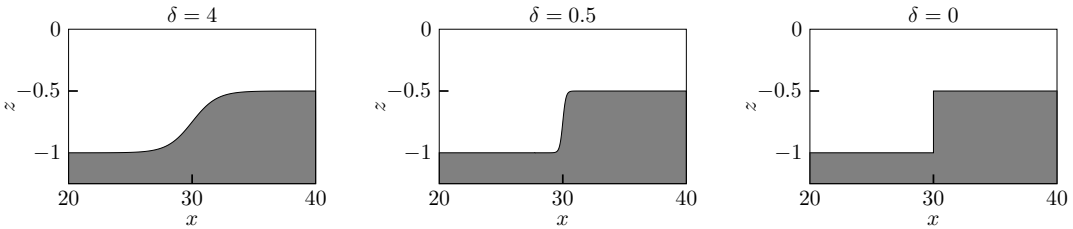

Figure 4.3 – Bathymetries at $x = 30$ for $\delta = 4$ (left), $\delta = 0.5$ (middle) and $\delta = 0$ (right).

4.4.2.1 Numerical results for the nonlinear shallow water equations

We aim at evaluating the difference between both classical Saint-Venant system (4.28) and the nonlocal alternative (4.27) in terms of the shallowness parameter μ. Since, for practical purposes, the classical Saint-Venant system is often used with non smooth topographies, the idea is to asses the price to pay when working with the classical model though the bottom is rough. In this section, the amplitude parameter is set to $\varepsilon = 0.1$.

Smooth bottom. In this test case, the topography parameter is $\beta = 0.6$. We set $\delta = 4$ so that the corresponding bathymetry is smooth (see Figure 4.3). In this situation, we know from Remark 4.3 that the nonlocal model (4.27) reduces to the classical shallow water approximation (4.28) up to $O(\mu)$ terms. In order to illustrate this precision, we computed the numerical solution given by (4.41) for several values of μ and we then compared them with the numerical solution of the classical shallow water equations. For each computation and each discrete time t^n, the L^∞-norm differences $E_\zeta^n = \|\zeta_{\rm NL}^n - \zeta_{\rm SV}^n\|_\infty$ and $E_v^n = \|v_{\rm NL}^n - v_{\rm SV}^n\|_\infty$ have been computed, where $(\zeta_{\rm NL}^n, v_{\rm NL}^n)_n$ is the numerical solution of the nonlocal alternative system (4.41) and $(\zeta_{\rm SV}^n, v_{\rm SV}^n)_n$ denotes the solution of the classical Saint-Venant scheme. Figure 4.4 depicts $e_\zeta = \max_n E_\zeta^n$ as a function of μ (the maximum is taken over a duration of 1500 time steps). The computed order of convergence is 0.95 which is consistent with the expected difference between both asymptotic models.

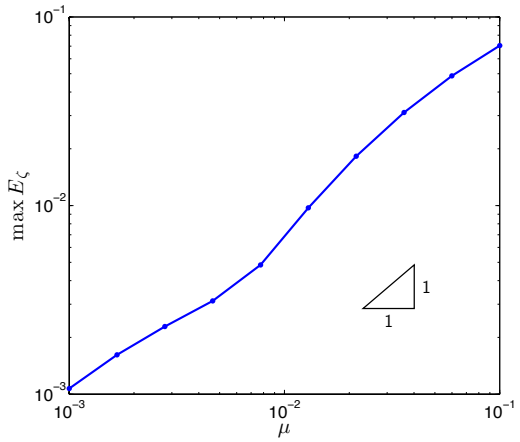

Figure 4.4 – Smooth bottom: convergence between the surface elevation computed by the classical Saint-Venant model and the alternative one, as functions of the parameter μ.

Rough bottom. In the following test cases, we focus on bottoms involving a steep slope. More precisely, the bathymetry is still given by (4.45) but the simulations have been performed for smaller values of δ, namely $\delta = 0.5$, $\delta = 0.1$ and the limit value $\delta = 0$. In the latter case, the bottom parametrization has a step at $x = 30$:

$$b(x) = \begin{cases} 0 & \text{if } 0 \le x < 30, \\ \frac{1}{2}\left(1 - \tanh(10(x-49))\right) & \text{if } 30 < x \le 60. \end{cases}$$

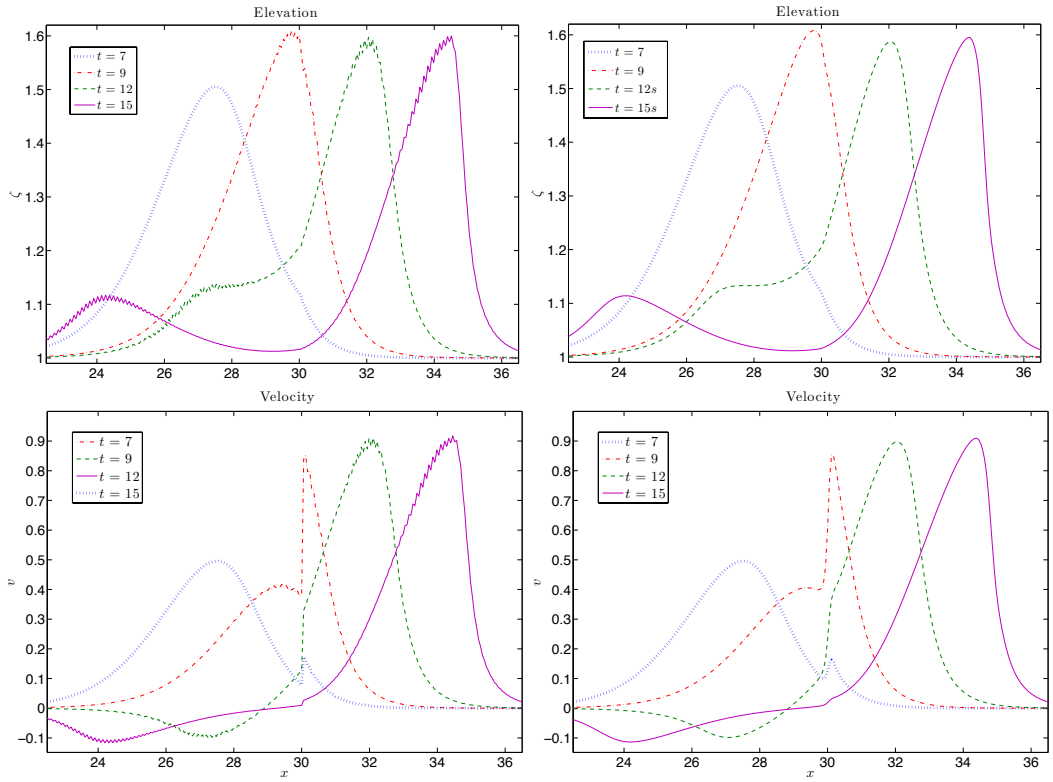

Figure 4.5 – Elevation and velocity for a wave passing over a step: classical Saint-Venant model (left) and nonlocal alternative (right).

Remark 4.8. Using the alternative shallow water model with the above step parametrization raises questions as to the meaning of the term $D\,\mathrm{sech}(\sqrt{\mu}|D|)bT_\mu[\beta b]v$ that appears in the definition of $b_\mu[\beta b]v$. As a matter of fact, defining this term amounts to defining the term $\mathcal{T} = D\,\mathrm{sech}(\sqrt{\mu}|D|)b|D|(bv)$, which is far from obvious for general $b \in L^\infty(\mathbb{R})$. Let us consider the case where b is the sign function and assume that v is smooth. In this case, we can actually define $\mathcal{T}$ as a smooth function. Indeed, note first that defining $b|D|(bv)$ as a tempered distribution is tantamount to defining its Fourier transform $\mathcal{F}(b|D|(bv))$. Since, up to a multiplicative constant, $\mathcal{F}b$ coincides with the principal value p.v.$\left(\frac{1}{\xi}\right)$, the Fourier transform $\mathcal{F}(b|D|(bv))$ can be formally written as a convolution product of the form p.v.$\left(\frac{1}{\xi}\right) * \mathcal{F}(|D|(bv)) = \mathcal{H}\mathcal{F}(|D|(bv))$, where $\mathcal{H}$ is the Hilbert transform. Now one can check that $\mathcal{F}(|D|(bv))$ takes the form $\mathcal{F}(|D|(bv)) = af_1 + f_2$ with $a \in \mathbb{C}$, $f_1 = v(0)\mathrm{sign}$ and $f_2 \in L^2(\mathbb{R})$ so that taking the Hilbert transform yields $\mathcal{F}(b|D|(bv)) = \tilde{a}\tilde{f}_1 + \tilde{f}_2$ where $\tilde{a} \in \mathbb{C}$, $\tilde{f}_1 = v(0)\log|\cdot|$ and $\tilde{f}_2 \in L^2(\mathbb{R})$. Since $\mathcal{F}\mathcal{T}(\xi) = \mathrm{sech}(\sqrt{\mu}|\xi|)\xi\mathcal{F}(b|D|(bv))$, the last logarithmic singularity gives rise to a term of the form $\mathrm{sech}(\sqrt{\mu}|\xi|)\xi\log|\xi|$, which is continuous and rapidly decreasing. This confirms that $\mathcal{T}$ makes sense and is a smooth function. The same conclusion holds for piecewise continuous parametrizations of the bottom.

Figure 4.5 shows the comparison between the wave profiles and the velocities determined from both the classical and the nonlocal shallow water models for a flow over such a step. The shallowness parameter is set to be $\mu = 0.01$ and the topography parameter is $\beta = 0.6$. As

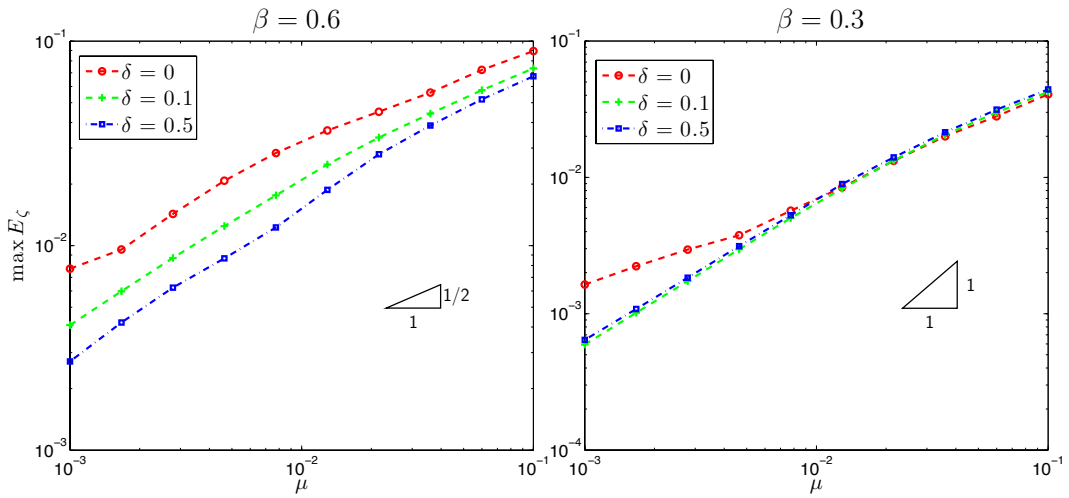

Figure 4.6 – Rough bottom: convergence between the surface elevation computed by the classical Saint-Venant model and the alternative one, as functions of the parameter μ.

δ	Convergence rate	
	$\beta = 0.6$	$\beta = 0.3$
4	0.96	0.94
0.5	0.71	0.94
0.1	0.63	0.94
0	0.54	0.71

Table 4.1 – Shallow water models: computed convergence rates with respect to μ.

the wave passes over the step (located at $x = 30$) the classical Saint-Venant model produces oscillations at the top of both main and reflected waves while the alternative model does not exhibit these oscillations. Moreover the velocity computed by the classical model has a jump discontinuity across the step. This discontinuity is "smoothed" by the nonlocal model. Note that the amplitude of the oscillations produced by the classical Saint-Venant equations decreases with decreasing the topography parameter β.

In order to estimate the error (in terms of the parameter μ) committed when using the classical Saint-Venant model, the quantity $e_\zeta = \max_n E_\zeta^n$ has again been computed and the results are plotted in Figure 4.6. To study the influence of the topography parameter, this figure also presents the results obtained with $\beta = 0.3$. As expected the error committed by the classical model increases with β. The computed orders of convergence with respect to μ for the surface elevation have been gathered in Table 4.1. For the case of small amplitude bottom ($\beta = 0.3$), the convergence rate remains close to 1 except for the step topography ($\delta = 0$). Now, for the case of large amplitude bottom ($\beta = 0.6$), the order of convergence decreases with the steepness of the topography. In the limit case $\delta = 0$, the classical model becomes a $O(\mu^{\frac{1}{2}})$ approximation (compared with $O(\mu)$ for smooth topographies). In other words, with this particular kind of rough bottoms, if one decides to use the classical shallow water model, the price to pay is at most one-half order of precision.

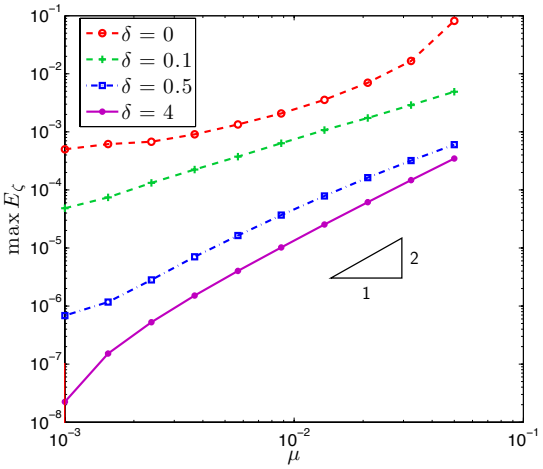

Figure 4.7 – Convergence between the surface elevation computed by the classical Boussinesq model and the alternative one, as functions of the parameter μ.

δ	Convergence rate
4	2.33
0.5	1.79
0.1	1.19
0	1.18

Table 4.2 – Boussinesq models: computed convergence rates with respect to μ for the wave amplitude.

4.4.2.2 Numerical results for the medium and small amplitude models

Convergence as functions of the shallowness parameter. We consider once again the L^∞-norm difference $E_\zeta^n = \|\zeta_\alpha^n - \zeta_{\text{Bouss}}^n\|_\infty$, where $(\zeta_\alpha^n)_n$ is the elevation computed by the nonlocal alternative Boussinesq model while $(\zeta_{\text{Bouss}}^n)_n$ denotes the numerical elevation given by the classical Boussinesq system. In the sake of evaluating the convergence between both classical Boussinesq system and the nonlocal alternative as functions of μ, the quantity $e_\zeta = \max_n E_\zeta^n$ has been computed over a duration of 1500 time steps and for several values of δ. The results are depicted in Figure 4.7 and the computed orders of convergence are given in Table 4.2. As we noticed in Remark 4.7, the alternative model (4.37) reduces to the standard Boussinesq system (4.38) up to $O(\mu^2)$ terms for smooth bottoms. The computed order of convergence of 2.33 obtained for the smooth step ($\delta = 4$) is thus consistent with the expected difference between both asymptotic models. In the limit case of the step bottom ($\delta = 0$) the convergence rate becomes 1.18. Consequently using the classical Boussinesq approximation costs at most about one order of precision.

The case of a step bottom ($\delta = 0$). In this limit case, two different behaviors emerge when comparing the classical shallow water medium amplitude model with the nonlocal alternative:

i) For the small values of the shallowness, say $\mu < 0.01$, the elevation and velocity computed by the classical model are close to those given by the nonlocal alternative (see Figure 4.8

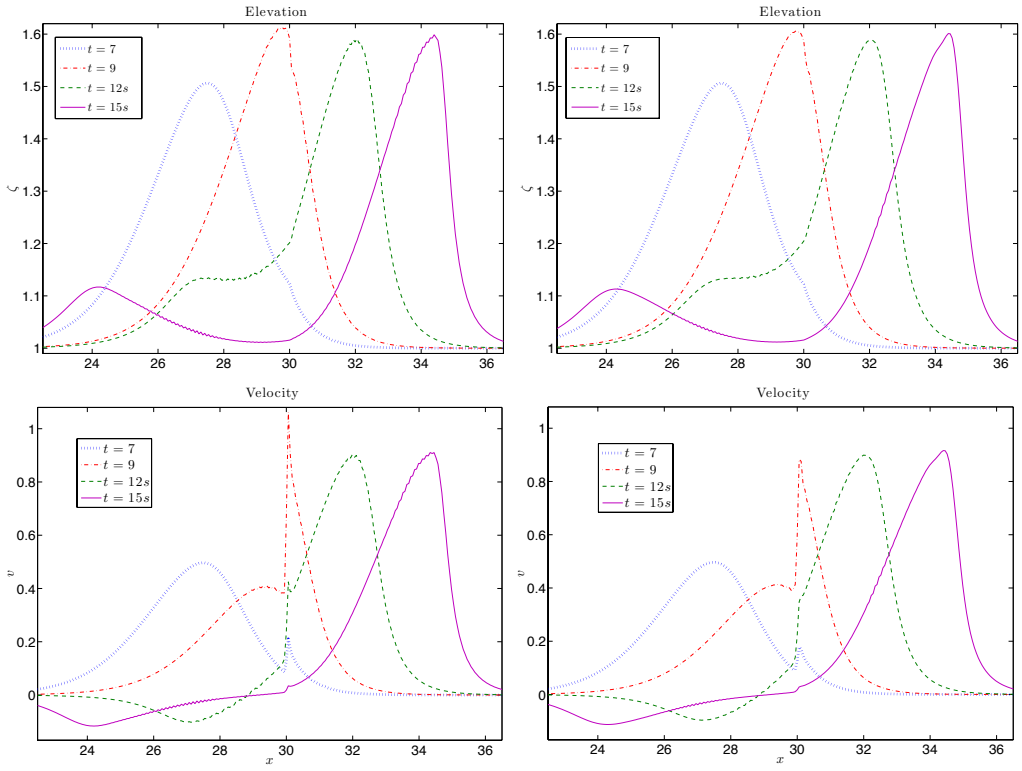

Figure 4.8 – Elevation and velocity for a wave passing over a step ($\mu = 0.01$, $\varepsilon = 0.1$). Classical medium amplitude model (left) and nonlocal alternative (right).

obtained for $\mu = 0.01$). In particular, as seen when comparing Figure 4.8 to Figure 4.5 both obtained for $\mu = 0.01$ and $\varepsilon = 0.1$, the amplitude of the oscillations produced by the classical medium amplitude model are lower than those obtained with the classical Saint-Venant model.

ii) For values of the shallowness parameter in the range $0.01 < \mu < 0.05$, some instabilities arise when using the classical medium and small amplitude models with a step bottom. This behavior is illustrated in Figure 4.9, which shows a comparison between the wave profiles and the velocities determined from both the classical and the nonlocal shallow water medium amplitude models. Note that these instabilities do not vanish for small values of the time step.

4.5 Appendix: the case of polygonal topographies

In the case of two dimensional motions and when the bottom has polygonal shape, the conformal mapping approach developed in the previous chapter yielded a different shallow water model. This method requires knowledge of a diffeomorphism Σ_{bott} which takes a uniform strip to the fluid domain at rest (see Section 3.2.1). Setting $(\sigma(x), \rho(x)) = \Sigma_{\text{bott}}^{-1}(x, \varepsilon\zeta(x))$ (the transformed free surface), recall that the resulting shallow water system with polygonal

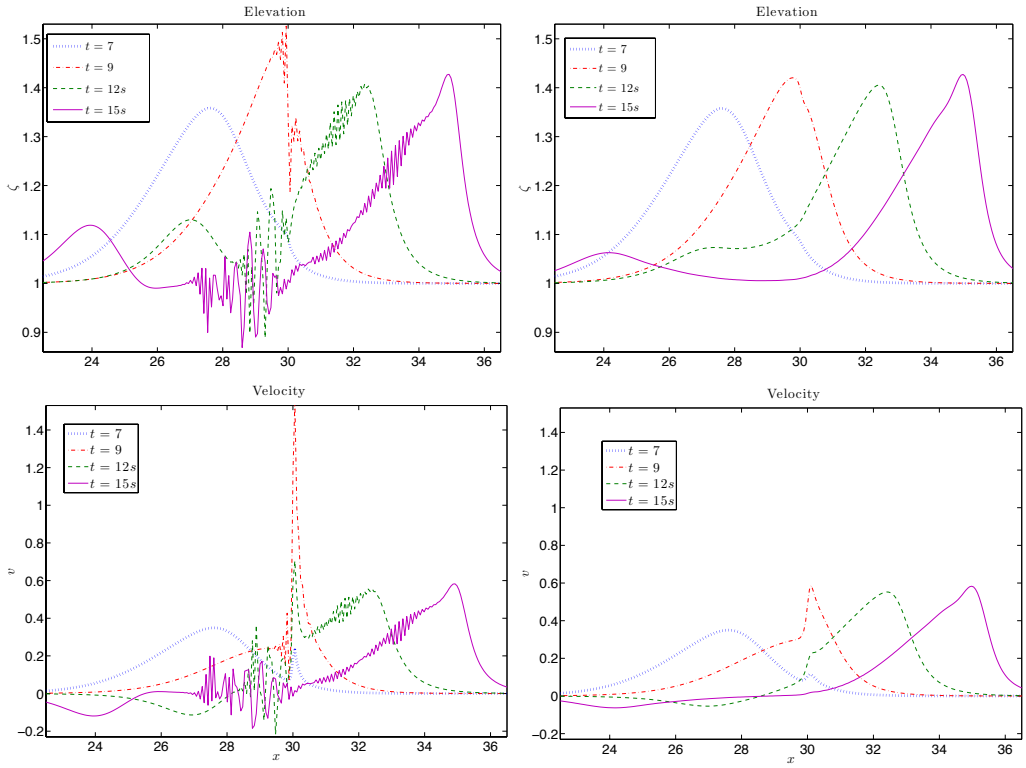

Figure 4.9 – Elevation and velocity for a wave passing over a step ($\mu = 0.04$, $\varepsilon = 0.2$). Classical medium amplitude model (left) and nonlocal alternative (right).

topography reads

$$\begin{cases} \partial_t \zeta + \partial_x (M_\mu v_{\rm s}) = 0, \\ \partial_t v_{\rm s} + \partial_x \zeta + \varepsilon v_{\rm s} \partial_x v_{\rm s} = 0, \end{cases} \tag{4.46}$$

where the transformed free surface coefficient M_μ is defined as

$$M_\mu = \frac{1 + \rho_\mu}{\sigma'_\mu}. \tag{4.47}$$

For polygonal topographies, the surface elevation computed by this model can be used as a reference solution to validate the formal approach developed in the present chapter.

4.5.1 Numerical computation of the transformed free surface coefficient

Computing the coefficient M_μ requires to evaluate the inverse of the diffeomorphism $\Sigma_{\rm bott}$ at the free surface mesh points. Now, as we have seen in Section 3.2.1.2, computing $\Sigma_{\rm bott}$ reduces to the computation of a Schwarz-Christoffel map. To compute such a map, we use the Schwarz-Christoffel Toolbox of Driscoll and Trefethen (see [50, Appendix] for more details on the use of this Toolbox, see also [68, Appendix A] for an application to the conformal mapping of a fluid domain with polygonal bottom). Then, given any family $\zeta = (\zeta_j)$ of discrete elevations at the grid points $\mathbf{x} = (x_j)$, the computation of M_μ results from evaluating

the transformed free surface $(\sigma_\mu, \rho_\mu) = \Sigma_{\text{bott}}^{-1}(\mathbf{x}, \zeta)$. Using a finite difference scheme similar to that given in Section 4.4, the numerical scheme for (4.46) reads

$$\begin{cases} \dfrac{\zeta^{n+1} - \zeta^n}{\Delta t} + \left(M_\mu^{n+1/2} \dfrac{v^{n+1} + v^n}{2} \right) = 0, \\ \dfrac{v^{n+1} - v^n}{\Delta t} + D_1\left(\dfrac{\zeta^{n+1} + \zeta^n}{2} \right) + \varepsilon v^{n+\frac{1}{2}} D_1\left(\dfrac{v^{n+1} + v^n}{2} \right) = 0, \end{cases} \tag{4.48}$$

where the predictive terms $M_\mu^{n+1/2}$ and $v^{n+\frac{1}{2}}$ are defined as

$$M_\mu^{n+1/2} = 2\frac{1+\rho_\mu^n}{D_1 \sigma_\mu^n} - M_\mu^{n-1/2} \quad \text{and} \quad v^{n+1/2} = 2v^n - v^{n-1/2}.$$

4.5.2 Numerical comparison of the shallow water models

To evaluate the behavior of the nonlocal shallow water model (4.27) when the bottom has polygonal shape, we compare the solutions produced by both systems (4.41) and (4.48) in the particular case of a rectangular bottom. The bathymetry is given by

$$b(x) = \begin{cases} 0 & \text{if } x < 30 \text{ or } x > 50, \\ 1 & \text{if } 30 \le x \le 50, \end{cases}$$

as illustrated in Figure 4.10. The initial condition is the unidirectional wave defined in (4.44)

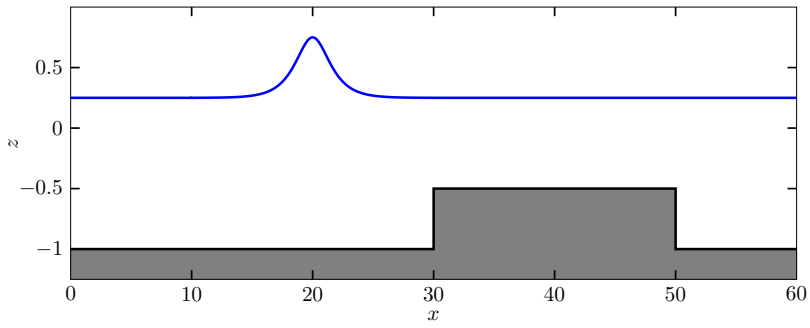

Figure 4.10 – Initial condition and rectangular bottom.

and the amplitude parameters are set to $\varepsilon = 0.1$ and $\beta = 0.6$. Time histories of the surface elevation computed by both models are shown in Figure 4.11. The simulation was performed using $N = 1024$ points and $\Delta t = 10^{-1}$. As the wave passes over the step, both models produce similar results.

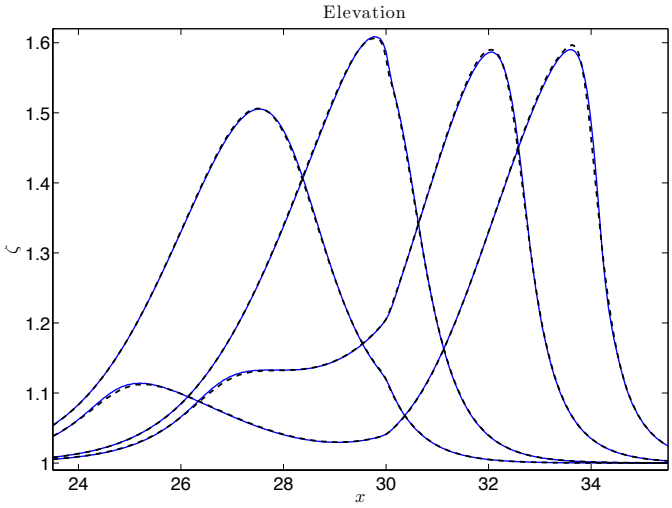

Figure 4.11 – Elevation at times $t = 7$, 9, 12 and 15 for a wave passing over a rectangular hump. Conformal mapping approach (dashed line) and nonlocal model (solid line).

Bibliographie

[1] I. Aavatsmark, T. Barkve, Ø. Bøe, and T. Mannseth. Discretization on unstructured grids for inhomogeneous, anisotropic media. I. Derivation of the methods. *SIAM J. Sci. Comput.*, 19(5) :1700–1716 (electronic), 1998. 3, 22, 30, 32, 49, 50

[2] I. Aavatsmark, T. Barkve, Ø. Bøe, and T. Mannseth. Discretization on unstructured grids for inhomogeneous, anisotropic media. II. Discussion and numerical results. *SIAM J. Sci. Comput.*, 19(5) :1717–1736, 1998. 22

[3] I. Aavatsmark, G. T. Eigestad, B. Heimsund, B. T. Mallison, J. M. Nordbotten, and E. Øian. A new finite volume approach to efficient discretization on challenging grids. In *Proc. SPE 106435*, Houston, USA, 2007. 22

[4] I. Aavatsmark, G. T. Eigestad, B. T. Mallison, and J. M. Nordbotten. A compact multipoint flux approximation method with improved robustness. *Numer. Methods Partial Differential Equations*, 24(5) :1329–1360, 2008. 22

[5] L. Agélas, D. A. Di Pietro, and J. Droniou. The G method for heterogeneous anisotropic diffusion on general meshes. *M2AN Math. Model. Numer. Anal.*, 44(4) :597–625, 2010. 3, 22

[6] L. Agelas, R. Eymard, and R. Herbin. A nine-point finite volume scheme for the simulation of diffusion in heterogeneous media. *C. R. Math. Acad. Sci. Paris*, 347(11-12) :673–676, 2009. 30, 32

[7] L. Agelas, C. Guichard, and R. Masson. Convergence of finite volume MPFA O type schemes for heterogeneous anisotropic diffusion problems on general meshes. *Int. J. Finite Vol.*, 7(2) :33, 2010. 3, 30

[8] L. Agelas and R. Masson. Convergence of the finite volume MPFA O scheme for heterogeneous anisotropic diffusion problems on general meshes. *C. R. Math. Acad. Sci. Paris*, 346(17-18) :1007–1012, 2008. 32

[9] T. Alazard, N. Burq, and C. Zuily. On the water-wave equations with surface tension. *Duke Math. J.*, 158(3) :413–499, 2011. 12

[10] B. Alvarez-Samaniego and D. Lannes. Large time existence for 3D water-waves and asymptotics. *Invent. Math.*, 171(3) :485–541, 2008. 11, 13, 14, 56, 58, 65, 69, 76, 78

[11] B. Andreianov, F. Boyer, and F. Hubert. Discrete duality finite volume schemes for Leray-Lions-type elliptic problems on general 2D meshes. *Numer. Methods Partial Differential Equations*, 23(1) :145–195, 2007. 23

[12] D. N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.*, 19(4) :742–760, 1982. 23

[13] G. A. Baker. Finite element methods for elliptic equations using nonconforming elements. *Math. Comp.*, 31(137) :45–59, 1977. 23

[14] P. Bastian, C. Engwer, J. Fahlke, and O. Ippisch. A unfitted discontinuous Galerkin method for pore-scale simulations of solute transport. *Math. Comput. Simulation*, 81(10) :2051–2061, 2011. 23

[15] T. B. Benjamin, J. L. Bona, and J. J. Mahony. Model equations for long waves in nonlinear dispersive systems. *Philos. Trans. Roy. Soc. London Ser. A*, 272(1220) :47–78, 1972. 87

[16] E. Bertolazzi and G. Manzini. A second-order maximum principle preserving finite volume method for steady convection-diffusion problems. *SIAM J. Numer. Anal.*, 43(5) :2172–2199, 2005. 4

[17] C. Besse. Schéma de relaxation pour l'équation de Schrödinger non linéaire et les systèmes de Davey et Stewartson. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 326(12) :1427 – 1432, 1998. 89

[18] C. Besse and C. H. Bruneau. Numerical study of elliptic-hyperbolic Davey-Stewartson system : dromions simulation and blow-up. *Math. Models Methods Appl. Sci.*, 8(8) :1363–1386, 1998. 89

[19] J. L. Bona, M. Chen, and J.-C. Saut. Boussinesq equations and other systems for small-amplitude long waves in nonlinear dispersive media. I. Derivation and linear theory. *J. Nonlinear Sci.*, 12(4) :283–318, 2002. 87

[20] J. L. Bona, T. Colin, and D. Lannes. Long wave approximations for water waves. *Arch. Ration. Mech. Anal.*, 178(3) :373–410, 2005. 13, 14, 58, 65

[21] F. Boyer and F. Hubert. Finite volume method for 2D linear and nonlinear elliptic problems with discontinuities. *SIAM J. Numer. Anal.*, 46(6) :3032–3070, 2008. 3, 23

[22] F. Brezzi, K. Lipnikov, and M. Shashkov. Convergence of the mimetic finite difference method for diffusion problems on polyhedral meshes. *SIAM J. Numer. Anal.*, 43(5) :1872–1896, 2005. 3, 22

[23] F. Brezzi, K. Lipnikov, and V. Simoncini. A family of mimetic finite difference methods on polygonal and polyhedral meshes. *Math. Models Methods Appl. Sci.*, 15(10) :1533–1551, 2005. 3, 22

[24] C. Buet and S. Cordier. On the non existence of monotone linear schema for some linear parabolic equations. *C. R. Math. Acad. Sci. Paris*, 340(5) :399–404, 2005. 4, 23

[25] E. Burman and A. Ern. Discrete maximum principle for Galerkin approximations of the Laplace operator on arbitrary meshes. *C. R. Math. Acad. Sci. Paris*, 338(8) :641–646, 2004. 4, 32, 41

[26] A.-P. Calderón. Cauchy integrals on Lipschitz curves and related operators. *Proc. Nat. Acad. Sci. U.S.A.*, 74(4) :1324–1327, 1977. 15, 79

[27] C. Cancès, M. Cathala, and C. Le Potier. Monotone corrections for generic cell-centered finite volume approximations of anisotropic diffusion equations. *Numerische Mathematik*, 2013. ix

[28] M. Cathala. Asymptotic shallow water models with non smooth topographies. Preprint hal-00804047, Mar. 2013. ix

[29] M. Cathala. Shallow water waves over polygonal bttoms. Preprint hal-00847843, July 2013. ix

[30] G. Chavent and J. Jaffré. *Mathematical Models and Finite Elements for Reservoir Simulation : Single Phase, Multiphase and Multicomponent Flows through Porous Media*. Studies in Mathematics and its Applications. Elsevier Science, 1986. 3

[31] F. Chazel. Influence of bottom topography on long water waves. *M2AN Math. Model. Numer. Anal.*, 41(4) :771–799, 2007. 14, 65

[32] F. Chazel. On the Korteweg-de Vries approximation for uneven bottoms. *Eur. J. Mech. B Fluids*, 28(2) :234–252, 2009. 89

[33] F. Chazel, D. Lannes, and F. Marche. Numerical simulation of strongly nonlinear and dispersive waves using a Green-Naghdi model. *J. Sci. Comput.*, 48(1-3) :105–116, 2011. 87, 88

[34] R. Cienfuegos, E. Barthélemy, and P. Bonneton. A fourth-order compact finite volume scheme for fully nonlinear and weakly dispersive Boussinesq-type equations. I. Model development and analysis. *Internat. J. Numer. Methods Fluids*, 51(11) :1217–1253, 2006. 87

[35] R. Coifman and Y. Meyer. Nonlinear harmonic analysis and analytic dependence. In *Pseudodifferential operators and applications (Notre Dame, Ind., 1984)*, volume 43 of *Proc. Sympos. Pure Math.*, pages 71–78. Amer. Math. Soc., Providence, RI, 1985. 15, 79

[36] W. Craig, P. Guyenne, D. P. Nicholls, and C. Sulem. Hamiltonian long-wave expansions for water waves over a rough bottom. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 461(2055) :839–873, 2005. 16, 76, 79, 80

[37] W. Craig, D. Lannes, and C. Sulem. Water waves over a rough bottom in the shallow water regime. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 29(2) :233–259, 2012. 76

[38] W. Craig and D. P. Nicholls. Travelling two and three dimensional capillary gravity water waves. *SIAM J. Math. Anal.*, 32(2) :323–359, 2000. 15, 79

[39] W. Craig, U. Schanz, and C. Sulem. The modulational regime of three-dimensional water waves and the Davey-Stewartson system. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 14(5) :615–667, 1997. 15, 79

[40] W. Craig and C. Sulem. Numerical simulation of gravity waves. *J. Comput. Phys.*, 108(1) :73–83, 1993. 9, 16, 57, 77, 79

[41] W. Craig, C. Sulem, and P.-L. Sulem. Nonlinear modulation of gravity waves : a rigorous approach. *Nonlinearity*, 5(2) :497–522, 1992. 57, 77

[42] G. de Marsily. *Hydrogéologie quantitative*. Collection Sciences de la terre. Masson, 1981. 2

[43] B. Després. Non linear finite volume schemes for the heat equation in 1D. Technical report, Laboratoire Jacques-Louis Lions, 2012. 4, 32

[44] D. A. Di Pietro. Cell centered Galerkin methods. *C. R. Math. Acad. Sci. Paris*, 348(1-2) :31–34, 2010. 22

[45] D. A. Di Pietro. Cell centered Galerkin methods for diffusive problems. *ESAIM Math. Model. Numer. Anal.*, 46(1) :111–144, 2012. 22

[46] D. A. Di Pietro and A. Ern. *Mathematical aspects of discontinuous Galerkin methods*, volume 69 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Heidelberg, 2012. 3

[47] D. A. Di Pietro, A. Ern, and J.-L. Guermond. Discontinuous Galerkin methods for anisotropic semidefinite diffusion with advection. *SIAM J. Numer. Anal.*, 46(2) :805–831, 2008. 23

[48] D. A. Di Pietro and M. Vohralík. A review of recent advances in discretization methods, a posteriori error analysis, and adaptive algorithms for numerical modeling in geosciences. *Oil and Gas Science and Technology*, page To appear, June 2013. 22

[49] K. Domelevo and P. Omnes. A finite volume method for the Laplace equation on almost arbitrary two-dimensional grids. *M2AN Math. Model. Numer. Anal.*, 39(6) :1203–1249, 2005. 3, 22

[50] T. A. Driscoll and L. N. Trefethen. *Schwarz-Christoffel mapping*, volume 8 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2002. 12, 15, 60, 61, 73, 97

[51] J. Droniou. Finite volume schemes for diffusion equations : Introduction to and review of modern methods. Preprint hal-00813613, 2013. 23

[52] J. Droniou and R. Eymard. A mixed finite volume scheme for anisotropic diffusion problems on any grid. *Numer. Math.*, 105(1) :35–71, 2006. 3, 22

[53] J. Droniou and R. Eymard. Study of the mixed finite volume method for Stokes and Navier-Stokes equations. *Numer. Methods Partial Differential Equations*, 25(1) :137–171, 2009. 3, 22

[54] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin. A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods. *Math. Models Methods Appl. Sci.*, 20(2) :265–295, 2010. 22

[55] J. Droniou, T. Eymard, R.and Gallouët, and R. Herbin. Gradient schemes : A generic framework for the discretisation of linear, nonlinear and nonlocal elliptic and parabolic equations. *Mathematical Models and Methods in Applied Sciences*, 0(0) :1–38, 0. 22

[56] J. Droniou and C. Le Potier. Construction and convergence study of schemes preserving the elliptic local maximum principle. *SIAM J. Numer. Anal.*, 49(2) :459–490, 2011. 4, 5, 23, 26, 32, 33

[57] V. Duchêne. Boussinesq/Boussinesq systems for internal waves with a free surface, and the KdV approximation. *ESAIM Math. Model. Numer. Anal.*, 46(1) :145–185, 2012. 89

[58] M. Duruflé and S. Israwi. A numerical study of variable depth KdV equations and generalizations of Camassa-Holm-like equations. *J. Comput. Appl. Math.*, 236(17) :4149–4165, 2012. 89

[59] M. G. Edwards and C. F. Rogers. Finite volume discretization with imposed flux continuity for the general tensor pressure equation. *Computational Geosciences*, 2(4) :259–290, 1998. 3, 22

[60] R. E. Ewing, editor. *The mathematics of reservoir simulation*, volume 1 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1983. 22

[61] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of numerical analysis, Vol. VII*, Handb. Numer. Anal., VII, pages 713–1020. North-Holland, Amsterdam, 2000. 4, 22, 24, 26

[62] R. Eymard, T. Gallouët, and R. Herbin. A cell-centered finite-volume approximation for anisotropic diffusion operators on unstructured meshes in any space dimension. *IMA J. Numer. Anal.*, 26(2) :326–353, 2006. 30, 32

[63] R. Eymard, T. Gallouët, and R. Herbin. A new finite volume scheme for anisotropic diffusion problems on general grids : convergence analysis. *C. R. Math. Acad. Sci. Paris*, 344(6) :403–406, 2007. 3, 22

[64] R. Eymard, T. Gallouët, and R. Herbin. Discretization schemes for linear diffusion operators on general non-conforming meshes. In *Finite volumes for complex applications V*, pages 375–382. ISTE, London, 2008. 3, 22

[65] R. Eymard, T. Gallouët, and R. Herbin. Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes SUSHI : a scheme using stabilization and hybrid interfaces. *IMA J. Numer. Anal.*, 30(4) :1009–1043, 2010. 3, 22, 28, 30, 43

[66] R. Eymard, G. Henry, R. Herbin, F. Hubert, R. Klöfkorn, and G. Manzini. 3d benchmark on discretization schemes for anisotropic diffusion problems on general grids. In J. Fořt, J. Fürst, J. Halama, R. Herbin, and F. Hubert, editors, *Finite Volumes for Complex Applications VI Problems & Perspectives*, volume 4 of *Springer Proceedings in Mathematics*, pages 895–930. Springer Berlin Heidelberg, 2011. 4, 23

[67] R. Eymard and R. Herbin. A new colocated finite volume scheme for the incompressible Navier-Stokes equations on general non matching grids. *C. R. Math. Acad. Sci. Paris*, 344(10) :659–662, 2007. 30

[68] A. S. Fokas and A. Nachbin. Water waves over a variable bottom : a non-local formulation and conformal mappings. *J. Fluid Mech.*, 695 :288–309, 2012. 7, 15, 97

[69] A. Genty and C. Le Potier. Positivity problem and reactive transport simulation in porous media. In *XVI International Conference on Computational Methods in Water Resources*, 2006. 3

[70] A. Genty and C. Le Potier. Maximum and minimum principles for radionuclide transport calculations in geological radioactive waste repository : comparison between a mixed hybrid finite element method and finite volume element discretizations. *Transp. Porous Media*, 88(1) :65–85, 2011. 3, 32

[71] P. Grisvard. *Singularities in boundary value problems*, volume 22 of *Recherches en Mathématiques Appliquées [Research in Applied Mathematics]*. Masson, Paris, 1992. 64

[72] P. Guyenne and D. P. Nicholls. Numerical simulation of solitary waves on plane slopes. *Math. Comput. Simulation*, 69(3-4) :269–281, 2005. 16

[73] P. Guyenne and D. P. Nicholls. A high-order spectral method for nonlinear water waves over moving bottom topography. *SIAM J. Sci. Comput.*, 30(1) :81–101, 2007/08. 16, 80

[74] J. Hamilton. Differential equations for long-period gravity waves on fluid of rapidly varying depth. *J. Fluid Mech.*, 83(2) :289–310, 1977. 12, 56, 76

[75] R. Herbin and F. Hubert. Benchmark on discretization schemes for anisotropic diffusion problems on general grids. In *Finite volumes for complex applications V*, pages 659–692. ISTE, London, 2008. 4, 23, 32

[76] F. Hermeline. A finite volume method for the approximation of diffusion operators on distorted meshes. *J. Comput. Phys.*, 160(2) :481–499, 2000. 3, 22

[77] P. Houston, C. Schwab, and E. Süli. Discontinuous *hp*-finite element methods for advection-diffusion-reaction problems. *SIAM J. Numer. Anal.*, 39(6) :2133–2163, 2002. 23

[78] B. Hu and D. P. Nicholls. Analyticity of Dirichlet-Neumann operators on Hölder and Lipschitz domains. *SIAM J. Math. Anal.*, 37(1) :302–320 (electronic), 2005. 16, 79

[79] T. Iguchi. A shallow water approximation for water waves. *J. Math. Kyoto Univ.*, 49(1) :13–55, 2009. 11, 13, 56, 58, 76

[80] T. Kano and T. Nishida. Sur les ondes de surface de l'eau avec une justification mathématique des équations des ondes en eau peu profonde. *J. Math. Kyoto Univ.*, 19(2) :335–370, 1979. 11, 76

[81] I. V. Kapyrin. A family of monotone methods for the numerical solution of three-dimensional diffusion problems on unstructured tetrahedral meshes. *Dokl. Akad. Nauk*, 416(5) :588–593, 2007. 4, 23, 32

[82] E. Keilegavlen, J. M. Nordbotten, and I. Aavatsmark. Sufficient criteria are necessary for monotone control volume methods. *Appl. Math. Lett.*, 22(8) :1178–1180, 2009. 4, 23

[83] G. Kreisel. Surface waves. *Quarterly of Applied Mathematics*, 7 :21–44, 1949. 12, 56

[84] D. Lannes. Well-posedness of the water-waves equations. *J. Amer. Math. Soc.*, 18(3) :605–654 (electronic), 2005. 13, 58

[85] D. Lannes. *The water waves problem : mathematical analysis and asymptotics*. AMS, to appear, 2013. 10, 13, 16, 56, 65, 67, 68, 79, 81, 86, 88

[86] D. Lannes and P. Bonneton. Derivation of asymptotic two-dimensional time-dependent equations for surface water wave propagation. *Physics of Fluids*, 21(1) :016601, 2009. 76

[87] D. Lannes and J.-C. Saut. Weakly transverse Boussinesq systems and the Kadomtsev-Petviashvili approximation. *Nonlinearity*, 19(12) :2853–2875, 2006. 14, 65

[88] C. Le Potier. Schéma volumes finis monotone pour des opérateurs de diffusion fortement anisotropes sur des maillages de triangles non structurés. *C. R. Math. Acad. Sci. Paris*, 341(12) :787–792, 2005. 4, 23

[89] C. Le Potier. Schéma volumes finis pour des opérateurs de diffusion fortement anisotropes sur des maillages non structurés. *C. R. Math. Acad. Sci. Paris*, 340(12) :921–926, 2005. 30, 32

[90] C. Le Potier. A nonlinear finite volume scheme satisfying maximum and minimum principles for diffusion operators. *Int. J. Finite Vol.*, 6(2) :20, 2009. 4, 23, 32

[91] C. Le Potier. Correction non linéaire et principe du maximum pour la discrétisation d'opérateurs de diffusion avec des schémas volumes finis centrés sur les mailles. *C. R. Math. Acad. Sci. Paris*, 348(11-12) :691–695, 2010. 4, 32, 37, 38, 40

[92] Y. A. Li. A shallow-water approximation to the full water wave problem. *Comm. Pure Appl. Math.*, 59(9) :1225–1285, 2006. 11, 76

[93] K. Lipnikov, M. Shashkov, D. Svyatskiy, and Y. Vassilevski. Monotone finite volume schemes for diffusion equations on unstructured triangular and shape-regular polygonal meshes. *J. Comput. Phys.*, 227(1) :492–512, 2007. 4, 23, 32

[94] K. Lipnikov, M. Shashkov, and I. Yotov. Local flux mimetic finite difference methods. *Numer. Math.*, 112(1) :115–152, 2009. 30, 32

[95] K. Lipnikov, D. Svyatskiy, and Y. Vassilevski. Interpolation-free monotone finite volume method for diffusion equations on polygonal meshes. *J. Comput. Phys.*, 228(3) :703–716, 2009. 4, 23, 32

[96] C. C. Mei and B. Le Méhauté. Note on the equations of long waves over an uneven bottom. *Journal of Geophysical Research*, 71(2) :393–400, 1966. 12, 56

[97] A. Nachbin. A terrain-following Boussinesq system. *SIAM J. Appl. Math.*, 63(3) :905–922 (electronic), 2003. 12, 14, 15, 56, 60, 72, 76

[98] A. Nachbin and K. Sølna. Apparent diffusion due to topographic microstructure in shallow waters. *Phys. Fluids*, 15(1) :66–77, 2003. 76

[99] Z. Nehari. *Conformal mapping*. McGraw-Hill Book Co., Inc., New York, Toronto, London, 1952. 12, 58

[100] D. P. Nicholls. Traveling water waves : spectral continuation methods with parallel implementation. *J. Comput. Phys.*, 143(1) :224–240, 1998. 79

[101] D. P. Nicholls and F. Reitich. A new approach to analyticity of Dirichlet-Neumann operators. *Proc. Roy. Soc. Edinburgh Sect. A*, 131(6) :1411–1433, 2001. 13, 58

[102] D. P. Nicholls and F. Reitich. A new approach to analyticity of Dirichlet-Neumann operators. *Proceedings of the Royal Society of Edinburgh : Section A Mathematics*, 131 :1411–1433, 11 2001. 65

[103] D. P. Nicholls and M. Taber. Joint analyticity and analytic continuation of Dirichlet-Neumann operators on doubly perturbed domains. *J. Math. Fluid Mech.*, 10(2) :238–271, 2008. 16, 79

[104] J. M. Nordbotten, I. Aavatsmark, and G. T. Eigestad. Monotonicity of control volume methods. *Numer. Math.*, 106(2) :255–288, 2007. 3, 23, 32

[105] O. Nwogu. Alternative Form of Boussinesq Equations for Nearshore Wave Propagation. *J. Waterway., Port, Coast., and Ocean Eng.*, 119(6) :618–638, 1993. 88

[106] L. V. Ovsjannikov. To the shallow water theory foundation. *Arch. Mech. (Arch. Mech. Stos.)*, 26 :407–422, 1974. Papers presented at the Eleventh Symposium on Advanced Problems and Methods in Fluid Mechanics, Kamienny Potok, 1973. 11, 76

[107] L. V. Ovsjannikov. Cauchy problem in a scale of Banach spaces and its application to the shallow water theory justification. In *Applications of methods of functional analysis to problems in mechanics (Joint Sympos., IUTAM/IMU, Marseille, 1975)*, pages 426–437. Lecture Notes in Math., 503. Springer, Berlin, 1976. 11, 76

[108] D. H. Peregrine. Long waves on a beach. *Journal of Fluid Mechanics*, 27 :815–827, 1967. 12, 56, 87, 88

[109] R. Rosales and G. Papanicolaou. Gravity waves in a channel with a rough bottom. *Stud. Appl. Math.*, 68(2) :89–102, 1983. 76

[110] Z. Sheng and G. Yuan. The finite volume scheme preserving extremum principle for diffusion equations on polygonal meshes. *J. Comput. Phys.*, 230(7) :2588–2604, 2011. 4, 23, 32

[111] S. Sun and M. F. Wheeler. Discontinuous Galerkin methods for coupled flow and reactive transport problems. *Appl. Numer. Math.*, 52(2-3) :273–298, 2005. 23

[112] S. Sun and M. F. Wheeler. $L^2(H^1)$ norm a posteriori error estimation for discontinuous Galerkin approximations of reactive transport problems. *J. Sci. Comput.*, 22/23 :501–530, 2005. 23

[113] P. K. Sweby. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM J. Numer. Anal.*, 21(5) :995–1011, 1984. 4

[114] A. Uzureau. *Modélisation et calculs pour la cicatrisation osseuse. Application à la modélisation d'un bioréacteur*. PhD thesis, Université de Nantes, 2013. 2

[115] G. Wei, J. T. Kirby, S. T. Grilli, and R. Subramanya. A fully nonlinear Boussinesq model for surface waves. I. Highly nonlinear unsteady waves. *J. Fluid Mech.*, 294 :71–92, 1995. 88

[116] M. F. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM J. Numer. Anal.*, 15(1) :152–161, 1978. 23

[117] Z. Yuan, G.and Sheng. Monotone finite volume schemes for diffusion equations on polygonal meshes. *J. Comput. Phys.*, 227(12) :6288–6312, 2008. 4, 23, 32

[118] V. Zakharov. Stability of periodic waves of finite amplitude on the surface of a deep fluid. *Journal of Applied Mechanics and Technical Physics*, 9 :190–194, 1968. 9, 57, 77

Problématiques d'analyse numérique et de modélisation pour écoulements de fluides environnementaux

Ce travail s'inscrit dans l'étude mathématique d'écoulements de fluides environnementaux. Nous en abordons deux aspects, à travers deux contextes distincts d'application.

En lien avec la simulation des écoulements en milieux poreux, on s'intéresse dans une première partie à la discrétisation d'opérateurs de diffusion anisotropes hétérogènes par des méthodes de volumes finis sur des maillages généraux. Dans le but d'obtenir des solutions approchées qui respectent les bornes physiques des modèles, notre attention se porte sur la conservation du principe du maximum pour les opérateurs elliptiques. Nous présentons des mécanismes généraux permettant de corriger tout schéma volumes finis afin de garantir un principe du maximum discret tout en préservant certaines de ses propriétés principales. On étudie en particulier les propriétés de coercivité et de convergence des schémas corrigés.

La deuxième partie est consacrée à la construction de modèles approchés pour la propagation des vagues en eaux peu profondes et sur des topographies irrégulières. A cet effet, nous proposons tout d'abord une adaptation de la démarche d'étude classique à des écoulements bidimensionnels sur des topographies polygonales. Dans un cadre plus général, nous développons ensuite une démarche formelle qui débouche sur des alternatives non locales à quelques modèles classiques (équations de Saint-Venant, équations de Serre, système de Boussinesq). Ces nouveaux modèles contiennent des termes régularisants pour les contributions du fond.

Mots clés : Equations elliptiques – Schémas volumes finis – Anisotropie – Principe du maximum – Corrections non linéaires – Equations d'Euler à surface libre – Modèles shallow water – Topographies irrégulières – Opérateur de Dirichlet-Neumann.

Mathematical modeling and numerical analysis of environmental flows

This work investigates two research questions associated with environmental flows and their mathematical modeling.

The first part is devoted to the development of finite volume methods for anisotropic and heterogeneous diffusion operators arising in models of porous media flows. To ensure that the approximate solutions lie within physical bounds, we aim at maintaining a discrete analogous of the maximum principle for elliptic operators. Starting from any given cell-centered finite volume scheme, we present a general approach to devise non-linear corrections providing a discrete maximum principle while retaining some main properties of the scheme. In particular, we study the coercivity and convergence properties of the modified schemes.

The second part of this work focuses on the derivation of approximate models for shallow water wave propagation over rough topographies. In the particular case of one-dimensional polygonal bottom profiles, we first propose an adaptation of the usual derivation method using complex analysis tools. We then develop a formal approach to account for more general topographies. We propose nonlocal alternatives to some classical models (namely Saint-Venant equations, Serre equations and Boussinesq system). All these alternative models only involve smoothing contributions of the bottom.

Keywords: Elliptic equations – Finite volume schemes – Anisotropy – Maximum principle – Nonlinear corrections – Water waves – Shallow water models – non smooth topographies – Dirichlet-Neumann operator.

AMS subject classification: 65N08, 65N12, 35J05, 76B15, 35Q35.