

Master Analyse, modélisation, simulation

Modélisation et simulation  
des écoulements de fluides  
dans la géosphère  
Aspects numériques

Michel KERN \*

2024–2025  
Version de décembre 2024

---

\*. INRIA, CENTRE DE PARIS, 75012 PARIS, [Michel.Kern@inria.fr](mailto:Michel.Kern@inria.fr)



# Table des matières

<b>1</b>	<b>Écoulement stationnaires : volumes finis pour l'équation de Darcy</b>	<b>6</b>
1.1	Rappel du modèle, inconvénients de méthodes usuelles . . . . .	6
1.1.1	Modèle physique . . . . .	6
1.1.2	Lien avec la formulation primale . . . . .	7
1.2	Volumes finis en dimension 1 . . . . .	9
1.2.1	Construction du schéma dans le cas homogène . . . . .	11
1.2.2	Extension au cas hétérogène . . . . .	12
1.2.3	Interprétation algébrique – propriétés de la matrice . . . . .	13
1.2.4	Convergence du schéma . . . . .	17
1.2.5	Quelques éléments de mise en oeuvre . . . . .	19
1.3	Volumes finis en dimension 2 ou 3 . . . . .	20
1.3.1	Maillage – inconnues discrètes . . . . .	20
1.3.2	Construction du schéma dans le cas homogène . . . . .	21
1.3.3	Extension au cas hétérogène . . . . .	23
1.3.4	Propriétés du schéma . . . . .	24
1.3.5	Convergence et estimation d'erreur . . . . .	25
1.3.6	Au delà du schéma à deux points . . . . .	27
<b>2</b>	<b>Volumes finis pour l'équation de diffusion transitoire</b>	<b>29</b>
2.1	Un (tout petit) peu de physique . . . . .	29
2.1.1	Lois de conservation . . . . .	29
2.1.2	De la physique aux mathématiques . . . . .	30
2.2	Volumes finis pour une équation de diffusion transitoire . . . . .	31
2.2.1	Stabilité du $\theta$ -schéma . . . . .	34
<b>3</b>	<b>Lois de conservations – méthode de Godunov</b>	<b>36</b>
3.1	Introduction, exemples . . . . .	36
3.2	Analyse mathématique . . . . .	38
3.2.1	Solution classiques – méthode des caractéristiques . . . . .	38
3.2.2	Solutions faibles . . . . .	41
3.2.3	Solutions faibles entropiques . . . . .	46
3.2.4	Solution du problème de Riemann . . . . .	47
3.3	Le schéma de Godunov . . . . .	49
3.3.1	Méthodes de volumes finis pour les lois de conservation . . . . .	49
3.3.2	Le schéma de Godunov . . . . .	51
3.4	Volumes finis en dimension 2 . . . . .	53

<b>A</b>	<b>Analyse des schémas par différences finies</b>	<b>54</b>
A.1	Différences finies . . . . .	54
A.2	Équation d'advection . . . . .	55
A.3	Équation de la chaleur . . . . .	61
<b>B</b>	<b>Équations d'advection–diffusion</b>	<b>64</b>
B.1	Équation d'advection–diffusion . . . . .	64
	<b>Bibliographie</b>	<b>67</b>

## Avertissement

Cette édition du cours est encore un travail non-achevé. Il est possible, et malheureusement probable, qu'il reste des erreurs, des fautes de frappe, des explications peu claires, et même peut-être des fautes de mathématiques. Par conséquent, lisez ce cours avec attention et un esprit critique.

Dans l'espoir d'améliorer ce document pour les promotions à venir, je vous serais gré de me communiquer les erreurs que vous pourriez découvrir, par un message à `michel.kern@inria.fr`

# Chapitre 1

## Écoulement stationnaires : volumes finis pour l'équation de Darcy

Dans ce chapitre, nous étudions la simulation numérique d'un écoulement stationnaire, régi par la loi de Darcy.

La principale méthode étudiée est la méthodes des volumes finis « à deux points ». Après un rappel des équations à résoudre, nous introduisons la méthode dans le cas d'un écoulement 1D, puis nous montrons comment généraliser à 2 (ou 3) dimensions. Nous présenterons des résultats de convergence, et une estimation d'erreur de la méthode, sans entrer dans le détails des démonstrations.

Ce chapitre est basé sur le cours de Master de R. Herbin [Her18], et sur des notes de cours de F. Boyer[Boy10]. La lectrice ou le lecteur souhaitant plus de détails pourra consulter [EGH00]. Signalons également la référence [AGL07], dans un esprit différent des précédentes, mais d'une grande qualité pédagogique.

### 1.1 Rappel du modèle, inconvénients de méthodes usuelles

#### 1.1.1 Modèle physique

Détails dans le cours d'E. Mouche.

$\Omega \subset \mathbf{R}^d$  ( $d = 1, 2, 3$ ) ouvert borné, connexe, régulier (coins autorisés, pas de fissure)

#### Écoulement stationnaire, incompressible, loi de Darcy

**Grandeurs physiques (unités)**

**Charge piézométrique**  $H$  (m) ;

**Vitesse de Darcy**  $\vec{u}$  (m/s) ;

**Perméabilité**  $K$  (m/s) ;

**Source**  $f$  ( $s^{-1}$ ).

$$(1.1) \quad \operatorname{div} \vec{u} = f, \text{ dans } \Omega \quad \text{incompressibilité}$$

$$(1.2) \quad \vec{u} = -K \nabla H, \text{ dans } \Omega \quad \text{loi de Darcy}$$

**Hypothèses** sur les données :

—  $K$  est une fonction scalaire (tenseurs exclus ici!),

$$(1.3) \quad K \in L^\infty(\Omega), \exists K_*, K^* \text{ tq : } 0 < K_* \leq K(x) \leq K^* < \infty, \text{ pp } x \in \Omega$$

Hypothèse physique, renforcée pour obtenir un problème mathématique *bien posé* ;

—  $f \in L^2(\Omega)$  (peut être affaibli, pour prendre en compte des sources ponctuelles).

**Conditions aux limites** Partition de la frontière de  $\Omega$  :  $\partial\Omega := \Gamma, \bar{\Gamma} = \bar{\Gamma}_D \cup \bar{\Gamma}_N, \Gamma_D \cap \Gamma_N = \emptyset$ .

**Dirichlet** charge imposée :

$$(1.4) \quad H = H_D \text{ sur } \Gamma_D$$

**Neumann** flux imposé,  $\vec{n}$  est la normale unitaire sortante à  $\Omega$  :

$$(1.5) \quad \vec{u} \cdot \vec{n} = -g \text{ sur } \Gamma_N$$

*Remarque 1.1.* D'après Darcy,  $\vec{u} \cdot \vec{n} = -K \frac{\partial H}{\partial n}$  (dérivée normale)

*Remarque 1.2* (Problème de Neumann « pur »). Si il n'y a pas de condition de Dirichlet ( $\Gamma = \Gamma_N$ ), le problème (1.1), (1.2), (1.5) est *mal posé*.

**Pas d'existence** Pour qu'il existe une solution, les données  $f$  et  $g$  doivent vérifier une condition nécessaire de compatibilité.

Intégrer (1.1) sur  $\Omega$ , la formule de Green donne :

$$\int_{\Omega} f = \int_{\Omega} \operatorname{div} \vec{u} = \int_{\Gamma} \vec{u} \cdot \vec{n} d\gamma = - \int_{\Gamma} g d\gamma,$$

soit  $\int_{\Omega} f + \int_{\Gamma} g d\gamma = 0$ , qui ne fait pas intervenir les inconnues.

Physiquement, cela signifie que la contribution algébrique de tous les termes sources (intérieurs pour  $f$  et de bord pour  $g$ ) doivent s'équilibrer.

**Pas d'unicité** En effet, il est clair que si  $H$  est solution du problème (1.1), (1.2), (1.5), alors pour toute constante  $C \in \mathbf{R}$ ,  $H + C$  est également solution. Il n'existe aucune condition pour « fixer » la constante.

En conséquence, nous ferons toujours l'hypothèse dans la suite que  $\Gamma_D \neq \emptyset$ . Pour obtenir un résultat mathématique, on doit même faire l'hypothèse plus forte que la mesure de  $\Gamma_D$  (dans  $\Gamma$ !) est strictement positive. En théorie, on ne peut pas seulement fixer la pression en un point!

### 1.1.2 Lien avec la formulation primale

Formulation primale = élimination de  $\vec{u}$ , équation du second ordre en  $H$  :

$$(1.6) \quad \begin{cases} -\operatorname{div}(K \operatorname{grad} H) = f & \text{dans } \Omega \\ H = H_D & \text{sur } \Gamma_D \\ K \frac{\partial H}{\partial n} = g & \text{sur } \Gamma_N. \end{cases}$$

Avec les hypothèses de la section précédente, on a existence et unicité de  $H \in H^1(\Omega)$  (espace de Sobolev) via le théorème de Lax–Milgram. On calcule ensuite  $\vec{u} = -K \text{grad } H$ . On peut alors utiliser une méthode d’éléments finis de Lagrange.

En pratique, ce n’est pas ce qui est fait pour les problèmes d’écoulement (c’est OK en mécanique ou en thermique). Raison :  $\vec{u}$  est souvent la quantité la plus importante (c’est le « moteur » pour le transport par advection), on veut la calculer directement.

**Exemple 1.1** (Écoulement en 1D).

Voir plus de détails dans le cours d’E. Mouche.

$\Omega = ]0, L[$  ( $L > 0$ ),  $f = 0$ , et un milieu « bi-couche » avec

$$K(x) = \begin{cases} K_1 & 0 < x < L/2; \\ K_2 & L/2 < x < L, \end{cases}$$

et les conditions aux limites (charge imposée) :  $H(0) = H_g$ ,  $H(L) = H_d$ .

D’après la conservation de la masse (en 1D) :  $u'(x) = 0$ , pour tout  $x \in ]0, L[$  et donc  $u(x) = \bar{u}$ , où  $\bar{u}$  est une constante dont la valeur est encore à déterminer.

D’après la loi de Darcy, on a alors :

$$H(x) = \begin{cases} H_g - \frac{\bar{u}}{K_1}x & \text{pour } 0 < x \leq L/2, \\ H_d + \frac{\bar{u}}{K_2}(L - x) & \text{pour } L/2 \leq x < L. \end{cases}$$

Pour déterminer la valeur de  $\bar{u}$ , on écrit la continuité de la charge  $H(x)$  en  $x = L/2$ , ce qui donne  $\bar{u} = \frac{\bar{K}}{L}(H_g - H_d)$ , où  $\bar{K}$  est la *moyenne harmonique* de  $K_1$  et  $K_2$ , définie par :

$$\frac{2}{\bar{K}} = \frac{1}{K_1} + \frac{1}{K_2}.$$

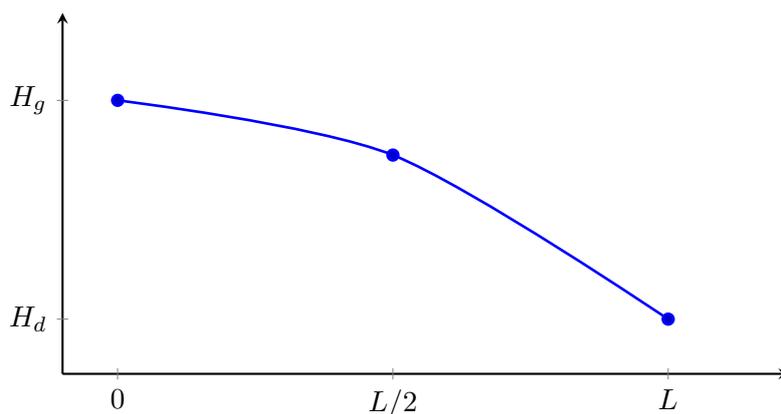


FIGURE 1.1 – Écoulement en dimension 1, avec perméabilité discontinue (cas où  $H_g > H_d$ )

Sur cet exemple 1D,  $H$  est continue, mais n’est manifestement pas de classe  $C^1$ , alors que la vitesse  $u$  est plus régulière (ici constante).

L'exemple suivant montre pourquoi il peut ne pas être recommandé de calculer  $H$  par une méthode d'éléments finis puis d'en déduire la vitesse.

**Exemple 1.2** (Éléments finis en 2D).

On discrétise les système (1.1)– (1.5) par une méthode d'éléments finis de Lagrange de degré 1. Pour cela, on se donne une triangulation  $\mathcal{T}_h$  du domaine  $\Omega \subset \mathbf{R}^2$  (pour fixer les idées), supposé polygonal. On rappelle que la méthode des éléments finis de Lagrange cherche une approximation  $H_h$  de la charge  $H$  dans l'espace (de dimension finie)

$$V_h = \{v_h \in C^0(\Omega), v_h|_T \text{ est un polynôme de degré 1 sur chaque élément } T \text{ du maillage}\},$$

et que  $V_h$  est un sous-ensemble de  $H^1(\Omega)$  (approximation conforme).

La charge approchée  $H_h$  est donc affine sur chaque triangle  $T \in \mathcal{T}_h$ , et par conséquent son gradient est *constant* sur chaque élément. Dans le cas le plus simple d'un milieu homogène (la perméabilité  $K$  est constante), la vitesse de Darcy  $\vec{u}$  est donc également constante sur chaque élément.

Étant donné deux éléments  $T_1$  et  $T_2$  possédant une arête commune notée  $e$  (voir figure 1.2, il n'y a aucune raison pour que les flux calculés de part et d'autre :

$$Q_1 = \int_e \vec{u}|_{T_1} \cdot \vec{n}_1 \, ds, \quad Q_2 = \int_e \vec{u}|_{T_2} \cdot \vec{n}_2 \, ds,$$

soient égaux. Mais pourtant, la condition d'incompressibilité  $\text{div } \vec{u} = 0$  impose que, sur tout arête, les flux d'un coté et de l'autre doivent être égaux (conséquence de la formule de Green).

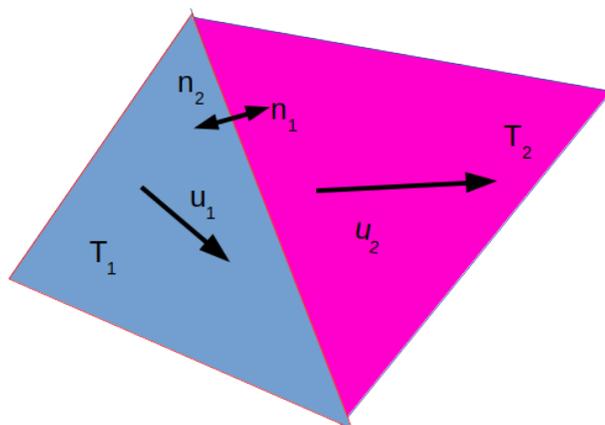


FIGURE 1.2 – Discontinuité du flux à travers une arête du maillage

Cet exemple n'est naturellement pas en contradiction avec les résultats de convergence obtenus sur la méthode des éléments finis! La convergence du gradient se produit dans  $L^2(\Omega)$ , ce qui n'entraîne pas la continuité des dérivées normales.

## 1.2 Volumes finis en dimension 1

Nous exposerons d'abord la méthode des volumes finis en dimension 1. Dans tout ce paragraphe, on prendra donc  $\Omega = ]0, L[$ , où  $L > 0$  est fixé. Le système d'équations à résoudre s'écrit

donc

$$(1.7) \quad u(x) = -K(x)H'(x) \quad 0 < x < L,$$

$$(1.8) \quad u'(x) = f(x) \quad 0 < x < L$$

où  $f \in L^2(0, L)$  est un terme source donné et  $K \in L^\infty(0, L)$  est la perméabilité. On ajoute les conditions aux limites, qui peuvent être :

**Dirichlet**  $H(0) = H_g, H(L) = H_d,$

**Neumann**  $u(0) = g_g, u(L) = -g_d,$

où  $H_g, H_d$  et  $g_g, g_d$  sont des réels fixés (on doit avoir une et une seule condition à la limite à chaque extrémité, chacune peut être soit de type Dirichlet, soit de type Neumann).

### Maillage et inconnues discrètes

On se donne

— une famille d'intervalles (les *volumes de contrôle*) recouvrant  $]0, L[$  :

$$(1.9) \quad T_i = ]x_{i-1/2}, x_{i+1/2}[, \quad \text{pour } i = 1, \dots, N$$

vérifiant

$$0 = x_{1/2} < x_1 < \dots < x_{i-1/2} < x_{i+1/2} < \dots < x_{N+1/2} = L;$$

On notera  $\mathcal{T} = \{T_i, i = 1, \dots, N\}$  l'ensemble des éléments (on dit aussi « mailles ») du maillage, et nous adopterons la convention de dénoter avec un symbole «  $\mathcal{T}$  » en indice un objet défini sur le maillage.

— une famille de points  $x_i \in T_i, i = 1, \dots, N, x_0 = 0, x_{N+1} = L$ . Le plus souvent (mais ce n'est pas nécessaire),  $x_i$  sera le milieu du  $T_i$  :  $x_i = \frac{x_{i-1/2} + x_{i+1/2}}{2}$ .

On note :  $h_i = x_{i+1/2} - x_{i-1/2}$  la longueur de  $T_i$ , et également  $h_{i+1/2} = x_{i+1} - x_i$ . Dans le cas d'un maillage régulier, on a  $h_i = h, \forall i$ , et aussi  $h_{i+1/2} = h$ .

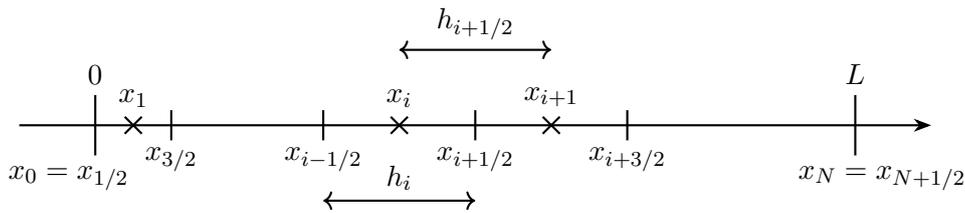


FIGURE 1.3 – Maillage volumes finis en 1D

**Définition 1.1** (Inconnues discrètes). Les inconnues discrètes du problème approché sont les nombres

$$(1.10) \quad H_i \approx \frac{1}{|T_i|} \int_{T_i} H(x) dx$$

Si la solution  $H$  est une fonction régulière, on a bien entendu  $H_i \approx H(x_i)$ . Mais la définition précédente est préférable pour conserver « l'esprit » des volumes finis.

### 1.2.1 Construction du schéma dans le cas homogène

Dans ce premier paragraphe, nous faisons donc l'hypothèse que la perméabilité  $K$  est une constante (strictement positive).

La construction du schéma procède en deux étapes :

**Définition des flux discrets** en intégrant l'équation (1.8) sur chaque volume de contrôle ;

**Approximation des flux** par différences finies.

Dans la phase de construction, nous procéderons formellement, en supposant que les fonctions sont suffisamment régulières pour justifier les calculs. L'analyse numérique viendra dans un deuxième temps.

#### Définition des flux discrets

Comme indiqué, nous intégrons l'équation de conservation (1.8) sur chaque volume de contrôle  $T_i$ , ce qui donne

$$u(x_{i+1/2}) - u(x_{i-1/2}) = h_i f_i, \quad i = 1, \dots, N,$$

en notant  $f_i = \frac{1}{h_i} \int_{T_i} f(x) dx$  la valeur moyenne de la source. Notons que cette première étape ne comporte aucune approximation.

La loi de Darcy (1.7) nous permet de remplacer la vitesse  $u$  par la charge  $H$  ( $K$  est ici constante) :

$$-KH'(x_{i+1/2}) + KH'(x_{i-1/2}) = h_i f_i, \quad i = 1, \dots, N,$$

#### Approximation des flux

Il faut maintenant pouvoir relier les flux  $-KH'(x_{i+1/2})$  aux inconnues discrètes, qui sont les moyennes de  $H$  sur  $T_i$ . En considérant que la charge  $H$  est *constante* sur chaque volume de contrôle  $T_i$  (qui sont bien entendu destinés à devenir « petits »), il est naturel d'approcher cette dérivée par le quotient aux différences :

$$H'(x_{i+1/2}) \approx = \frac{H_{i+1} - H_i}{x_{i+1} - x_i} = \frac{H_{i+1} - H_i}{h_{i+1/2}},$$

ce qui conduit à la définition du schéma (à l'intérieur)

$$(1.11) \quad \boxed{-K \frac{H_{i+1} - H_i}{h_{i+1/2}} + K \frac{H_i - H_{i-1}}{h_{i-1/2}} = h_i f_i, \quad i = 1, \dots, N.}$$

Ce schéma fait intervenir les valeurs des inconnues discrètes, mais notons que les valeurs des inconnues « au bord »  $H_0$  et  $H_{N+1}$  ne sont pas définies. Il manque deux équations.

#### Conditions aux limites

Nous indiquerons ce qui se passe sur le bord « gauche » ( $x = 0$ ), la transposition en  $x = L$  étant immédiate.

**Condition de Dirichlet** Sur la première maille, le terme  $\frac{H_i - H_{i-1}}{h_{i-1/2}}$  n'est pas défini. Une solution possible (ce n'est pas la seule) est de considérer qu'il s'agit d'une demi-maille, de

remplacer la valeur inconnue  $H_0$  par la donnée  $H_g$ , et de prendre par définition  $h_{1/2} = x_{1/2} - x_0$  (dans le cas d'un maillage uniforme, on a donc  $h_{1/2} = h/2$ , ce qui correspond bien à une demi-maille).

**Condition de Neumann** Dans ce cas, on revient à la première étape, et on remplace *le flux* par sa valeur imposée :  $u(0) = -KH'(0) = g_g$  (noter que le flux est par rapport à la « normale » sortante, ce qui explique l'absence de signe « - ». Sur le bord droit, on a  $u(L) = KH'(L) = -g_d$ . Dans tous les cas, faire attention aux conventions de signe.

On voit ainsi que dans tous les cas, le nombre d'inconnues reste égal aux nombre de volumes de contrôle. Il y a bien un traitement séparé des conditions aux limites, mais sans avoir besoin d'éliminer des inconnues.

### 1.2.2 Extension au cas hétérogène

Dans ce paragraphe, et dans la suite,  $K$  est maintenant une fonction. En vue des applications en hydrogéologie, il est important de prendre en compte des perméabilités discontinues, et même fortement discontinues. Ce n'est pas une restriction en pratique de faire l'hypothèse que le maillage respecte les discontinuités, et donc que la perméabilité varie de manière régulière dans les mailles. En première approximation, nous ferons l'hypothèse que la perméabilité est constante sur chaque maille, et nous noterons  $K_i$  sa valeur sur la maille  $T_i$ .

La première étape du paragraphe précédent est toujours valable, il s'agit maintenant d'approcher la quantité

$$u(x_{i+1/2}) = -(KH')(x_{i+1/2})$$

sachant que tant  $K$  que  $H'$  ne sont pas définies de manière unique au point  $x_{i+1/2}$ . La résolution de ce problème passe par l'introduction d'une inconnue fictive en  $x_{i+1/2}$ , qui permet de calculer un flux de chaque côté du point en question, puis qui sera éliminée en imposant l'égalité des deux flux calculés. Le schéma obtenu sera conservatif par construction.

Notons donc temporairement  $H_{i+1/2} \approx H(x_{i+1/2})$ , et notons également  $F_{i+1/2} \approx u(x_{i+1/2})$  l'approximation du flux (qui reste à définir). Le schéma volume fini s'écrira donc

$$(1.12) \quad F_{i+1/2} - F_{i-1/2} = h_i f_i, \quad i = 1, \dots, N$$

et notre but est de définir  $F_{i+1/2}$ .

On introduit donc :

$$\text{Sur } (x_i, x_{i+1/2}) \quad F_{i+1/2}^- = -K_i \frac{H_{i+1/2} - H_i}{x_{i+1/2} - x_i},$$

$$\text{Sur } (x_{i+1/2}, x_{i+1}) \quad F_{i+1/2}^+ = -K_{i+1} \frac{H_{i+1} - H_{i+1/2}}{x_{i+1} - x_{i+1/2}}.$$

Un calcul simple mais fastidieux montre que la condition  $F_{i+1/2}^+ = F_{i+1/2}^-$  permet d'éliminer l'inconnue fictive  $H_{i+1/2}$  entre les deux équations précédentes et donne pour valeur commune du flux

$$(1.13) \quad F_{i+1/2} = -\bar{K}_{i+1/2} \frac{H_{i+1} - H_i}{h_{i+1/2}}, \quad i = 1, \dots, N-1,$$

où la perméabilité moyenne entre les mailles  $T_i$  et  $T_{i+1}$  est donnée par

$$(1.14) \quad \boxed{\bar{K}_{i+1/2} = \frac{K_i K_{i+1} h_{i+1/2}}{K_i h_{i+1/2}^+ + K_{i+1} h_{i+1/2}^-}},$$

en notant  $h_{i+1/2}^+ = x_{i+1} - x_{i+1/2}$ , et  $h_{i+1/2}^- = x_{i+1/2} - x_i$  les longueurs des « demi-maillages » de chaque côté de  $x_{i+1/2}$  ( $h_{i+1/2}^+ + h_{i+1/2}^- = h_{i+1/2}$ ).

Dans le cas d'un maillage régulier ( $h_{i+1/2} = h$  et  $h_{i+1/2}^+ = h_{i+1/2}^- = h/2$ , pour tout  $i$ ), la formule précédente se simplifie

$$\bar{K}_{i+1/2} = \frac{2K_i K_{i+1}}{K_i + K_{i+1}},$$

ce qui montre que  $\bar{K}_{i+1/2}$  est la *moyenne harmonique* de  $K_i$  et  $K_{i+1}$ , définie par

$$\frac{2}{\bar{K}_{i+1/2}} = \frac{1}{K_i} + \frac{1}{K_{i+1}}.$$

Dans le cas général, on peut vérifier que  $\frac{K_{i+1/2}}{h_{i+1/2}}$  est la moyenne harmonique de  $\frac{K_i}{h_i}$  et de  $\frac{K_{i+1}}{h_{i+1}}$ .

*Remarque 1.3* (Justification physique de la moyenne harmonique). Considérons deux mailles voisines  $T_1$  et  $T_2$  dont les perméabilités sont  $K_1 = 1$  et  $K_2 = \varepsilon$ , avec naturellement  $\varepsilon \ll 1$ . On s'attend à ce que la perméabilité intermaille  $\bar{K}_{i+1/2}$  soit très faible, de façon à ne pas laisser passer de fluide vers la maille la plus imperméable.

C'est bien ce qui se produit avec la moyenne harmonique :

$$\bar{K}_{\text{harm}} = \frac{2\varepsilon}{1 + \varepsilon} \approx 2\varepsilon.$$

Si l'on avait, au contraire, utilisé la moyenne arithmétique pour calculer la perméabilité intermaille, on aurait obtenu :

$$\bar{K}_{\text{arith}} = \frac{1 + \varepsilon}{2} \approx \frac{1}{2},$$

ce qui conduit à un comportement non physique.

En résumé, dans le cas général, le schéma est défini par les équations (1.12), (1.13), (1.14). La définition des flux numériques doit être complétée pour les mailles voisines du bord par la prise en compte des conditions aux limites.

### Conditions de Dirichlet

$$(1.15) \quad F_{1/2} = -K_1 \frac{H_1 - H_g}{h_1}, \quad F_{N+1/2} = -K_N \frac{H_d - H_N}{h_N},$$

Noter que  $h_{1/2}$  et  $h_{N+1/2}$  correspondent à des demi-maillages, et que dans le cas d'un maillage uniforme, on a donc  $h_{1/2} = h_{N+1/2} = h/2$ .

### Conditions de Neumann

$$(1.16) \quad F_{1/2} = g_g, \quad F_{N+1/2} = -g_d$$

### 1.2.3 Interprétation algébrique – propriétés de la matrice

Le schéma défini au paragraphe précédent peut s'écrire comme un système linéaire.

Nous considérons d'abord le cas des conditions de Dirichlet. Dans le cas d'un maillage avec  $N$  mailles, les  $N$  inconnues sont les composantes du vecteur  $H_{\mathcal{T}} = (H_1, \dots, H_N)^T$ , et les  $N$  équations fournies par (1.12) conduisent un système linéaire

$$(1.17) \quad \tau_{i-1/2}(H_i - H_{i-1}) + \tau_{i+1/2}(H_i - H_{i+1}) = h_i f_i, \quad i = 1, \dots, N,$$

où les *transmissivités*  $\tau_{i+1/2}$  sont définies par

$$\tau_{i+1/2} = \frac{\bar{K}_{i+1/2}}{h_{i+1/2}},$$

et où nous avons défini les inconnues additionnelles  $H_0 = H_g$  et  $H_{N+1} = H_d$  (il s'agit uniquement d'un artifice mathématique pour conserver une écriture compacte).

Dans le cas des conditions de Neumann, l'équation (1.12) reste valable, ainsi que la définition des flux pour les points intérieurs (1.13), mais la définition des flux sur les deux interfaces du bord est différente et donnée par (1.16).

Dans tous les cas, on obtient un système linéaire

$$(1.18) \quad A_{\mathcal{T}} H_{\mathcal{T}} = f_{\mathcal{T}}$$

pour le vecteur inconnu  $H_{\mathcal{T}}$ . La matrice  $A_{\mathcal{T}}$  est tridiagonale. Pour une ligne « intérieure », c'est-à-dire correspondant à un indice  $i$  entre 2 et  $N - 1$ , on a

$$(1.19) \quad (A_{\mathcal{T}})_{i,i-1} = -\tau_{i-1/2}, \quad (A_{\mathcal{T}})_{ii} = \tau_{i-1/2} + \tau_{i+1/2}, \quad (A_{\mathcal{T}})_{i,i+1} = -\tau_{i+1/2}.$$

Pour ces mêmes indices, le second membre est donné par  $(f_{\mathcal{T}})_i = h_i f_i$ .

La première et la dernière ligne de la matrice, et le premier et le dernier élément du second membre tiennent compte des conditions aux limites :

**CL Dirichlet** En définissant les transmissivités pour tenir compte de (1.15), on a

$$\text{Bord gauche } (A_{\mathcal{T}})_{i,1} = \tau_{1/2} + \tau_{3/2}, \quad (A_{\mathcal{T}})_{12} = -\tau_{3/2}, \quad (f_{\mathcal{T}})_1 = h_1 f_1 + \tau_{1/2} H_g;$$

$$\text{Bord droit } (A_{\mathcal{T}})_{i,N-1} = -\tau_{N-1/2}, \quad (A_{\mathcal{T}})_{NN} = \tau_{N-1/2} + \tau_{N+1/2}, \quad (f_{\mathcal{T}})_N = h_N f_N + \tau_{N+1/2} H_d;$$

**CL Neumann** Nous utilisons cette fois (1.16).

$$\text{Bord gauche } (A_{\mathcal{T}})_{i,1} = \tau_{3/2}, \quad (A_{\mathcal{T}})_{12} = -\tau_{3/2}, \quad (f_{\mathcal{T}})_1 = h_1 f_1 + g_g;$$

$$\text{Bord droit } (A_{\mathcal{T}})_{i,N-1} = -\tau_{N-1/2}, \quad (A_{\mathcal{T}})_{NN} = \tau_{N-1/2}, \quad (f_{\mathcal{T}})_N = h_N f_N + g_d;$$

*Remarque 1.4.* Il est naturellement possible d'avoir des type de conditions à la limite différents aux deux extrémités.

*Remarque 1.5* (Problème de Neumann pur). Comme on peut s'y attendre, le problème de Neumann pur est *mal posé*, la matrice correspondante n'est pas inversible. Le noyau de  $A_{\mathcal{T}}$  est de dimension 1, engendré par le vecteur  $e = (1, \dots, 1)^T$ , qui correspond à la discrétisation sur la grille d'une fonction constante (cf remarque 1.2).

## Comparaison avec les différences finies

Il est instructif de comparer le système obtenu avec celui que donnerait une méthode de différences finies. Nous illustrons ce point dans le cas d'un milieu homogène (la perméabilité  $K$  est constante) et d'un maillage uniforme de pas  $h$ , avec des conditions de Dirichlet à chaque extrémité.

Les équations à l'intérieur se réduisent au schéma à trois points bien connu :

$$\frac{K}{h} (H_i - H_{i+1}) + \frac{K}{h} (H_i - H_{i-1}) = h f_i, \quad i = 2, \dots, N - 1$$

ou de manière équivalente

$$K \frac{-H_{i-1} + 2H_i - H_{i+1}}{h^2} = f_i,$$

ce qui est le même schéma que celui qui aurait été obtenu par différences finies. Mais la prise en compte des conditions aux limites est différente. Puisque le bord  $x = 0$  est à une demi-maille du centre de la première maille, la première équation s'écrit

$$\frac{K}{h/2}(H_1 - H_g) + \frac{K}{h}(H_1 - H_2) = hf_1$$

soit

$$\frac{3K}{h}H_1 - \frac{K}{h}H_2 = hf_1 \frac{2K}{h}H_g,$$

et l'expression analogue sur la dernière maille.

Dans le cas des différences finies, le coefficient « 3 » serait remplacé par un « 2 ». Les deux systèmes ne sont donc pas identiques !

### Propriétés de la matrice – formulation variationnelle

Dans ce paragraphe nous utilisons temporairement des notations plus proches des mathématiques. Nous ne considérons que le cas des conditions de Dirichlet aux deux extrémités.

Nous noterons  $u_\tau, v_\tau$  des vecteurs de  $\mathbf{R}^N$ , et il sera commode d'ajouter des valeurs fictives (nulles) correspondant aux conditions aux limites. Ainsi

$$u_\tau = (u_0, u_1, \dots, u_N, u_{N+1})^T, \quad \text{avec } u_0 = u_{N+1} = 0.$$

Nous munirons  $\mathbf{R}^N$  du produit scalaire usuel

$$(u_\tau, v_\tau) = \sum_{i=1}^N u_i v_i = \sum_{i=0}^{N+1} u_i v_i,$$

la seconde égalité étant vraie d'après notre convention.

**Lemme 1.1** (Intégration par parties discrète). *Étant donné  $(u_\tau, v_\tau) \in \mathbf{R}^N \times \mathbf{R}^N$ , on a*

$$(1.20) \quad (A_{\mathcal{T}} u_\tau, v_\tau) = \sum_{i=1}^N \bar{K}_{i+1/2} \frac{u_{i+1} - u_i}{h_{i+1/2}} \cdot \frac{v_{i+1} - v_i}{h_{i+1/2}} h_{i+1/2}.$$

*Démonstration.* La preuve est un simple calcul.

$$\begin{aligned} (A_{\mathcal{T}} u_\tau, v_\tau) &= \sum_{i=1}^N (F_{i+1/2} - F_{i-1/2}) v_i = \sum_{i=1}^{N-1} F_{i+1/2} v_i + F_{N+1/2} v_N - \sum_{i=2}^N F_{i-1/2} v_i - F_{1/2} v_1 \\ &= \sum_{i=1}^{N-1} F_{i+1/2} (v_i - v_{i+1}) + F_{N+1/2} v_N - F_{1/2} v_1 \\ &= \sum_{i=0}^N F_{i+1/2} (v_i - v_{i+1}) = \sum_{i=1}^N \bar{K}_{i+1/2} \frac{u_{i+1} - u_i}{h_{i+1/2}} \cdot \frac{v_{i+1} - v_i}{h_{i+1/2}} h_{i+1/2}. \end{aligned}$$

Le passage de la première à la deuxième ligne est un changement d'indice, le passage de la deuxième à la troisième utilise les conditions aux limites, puis l'expression du flux (1.13). ■

On en déduit facilement que le système approché (1.18) a une solution unique.

**Théorème 1.1.** *La matrice  $A_{\mathcal{T}}$  est symétrique et défini positive*

*Démonstration.* La symétrie est claire d'après l'expression des coefficients hors-diagonaux dans (1.17). Pour montrer que la matrice est définie positive, nous utilisons le lemme précédent, avec  $v_{\mathcal{T}} = u_{\mathcal{T}}$ .

$$(Au_{\mathcal{T}}, u_{\mathcal{T}}) = \sum_{i=1}^{N-1} \bar{K}_{i+1/2} \left( \frac{u_{I+1} - u_i}{h_{i+1/2}} \right)^2 h_{i+1/2}$$

et cette expression est manifestement positive ou nulle. De plus, si le produit scalaire  $(Au_{\mathcal{T}}, u_{\mathcal{T}})$  est nul, chacun des termes de la somme est nul, ce qui veut dire que le vecteur  $u_{\mathcal{T}}$  est constant, mais comme (d'après les conditions de Dirichlet)  $u_0 = u_{N+1} = 0$ , le vecteur  $u_{\mathcal{T}}$  est nul. ■

On voit donc que  $A_{\mathcal{T}}$  est inversible, ce qui justifie l'existence et l'unicité de la solution discrète (pour un maillage fixé!) annoncé avant le théorème.

Une conséquence utile de la formule d'intégration par parties discrète est une interprétation *variationnelle* du schéma. La démonstration du résultat suivant est essentiellement contenue dans celle du Lemme 1.1.

**Proposition 1.1** (Formulation variationnelle discrète). *La solution  $H_{\mathcal{T}}$  du schéma volume fini (1.18) vérifie la formulation variationnelle suivante :*

$$(1.21) \quad \forall G_{\mathcal{T}} \in \mathbf{R}^N, \sum_{i=1}^{N-1} \bar{K}_{i+1/2} \left( \frac{H_{i+1} - H_i}{h_{i+1/2}} \right) \left( \frac{G_{i+1} - G_i}{h_{i+1/2}} \right) h_{i+1/2} = \sum_{I=1}^N f_I G_I h_I.$$

Bien entendu, ce résultat est à rapprocher de la formulation variationnelle « continue » vérifiée par la solution du problème continu.  $H \in H_0^1(0, L)$  vérifie

$$\forall G \in H_0^1(0, L), \int_0^L K(x) H'(x) G'(x) dx = \int_0^L f(x) G(x) dx.$$

Le résultat suivant généralise également une propriété de la solution du problème continu. Il est fort utile, dans la mesure où il montre que le schéma numérique préserve cette propriété de *monotonie*, ce qui n'est pas toujours le cas.

**Proposition 1.2** (Principe du maximum). *Si le second membre  $f$  vérifie  $f_i \geq 0, \forall i$ , alors la solution du schéma volume fini (1.18) vérifie également  $H_i \geq 0, \forall i$ .*

*Démonstration.* La démonstration est par l'absurde. Supposons donc qu'il existe un indice  $i$  pour lequel  $H_i < 0$ . Soit maintenant  $i_0$  le plus petit indice auquel  $H_{\mathcal{T}}$  atteint son minimum. Nécessairement,  $H_{i_0} < 0$  et on a  $H_{i_0} < H_{i_0-1}$  (inégalité stricte par définition de  $i_0$ ) et  $H_{i_0} \leq H_{i_0+1}$ . Écrivons alors l'équation discrète correspondant à  $i_0$  :

$$-\bar{K}_{i_0+1} \frac{H_{i_0+1} - H_{i_0}}{h_{i_0+1/2}} + \bar{K}_{i_0} \frac{H_{i_0} - H_{i_0-1}}{h_{i_0-1/2}} = h_{i_0} f_{i_0}.$$

Par hypothèse, le second membre est positif, mais chacun des deux termes du premier membre est négatif, et même strictement négatif pour le second, ce qui est bien la contradiction attendue. ■

### 1.2.4 Convergence du schéma

Dans cette section, nous énonçons et démontrons un résultat de convergence pour le schéma défini par (1.12) et (1.13), dans le cas d'un milieu homogène ( $K = 1$ ), avec des conditions de Dirichlet homogènes ( $H(0) = H(L) = 0$ ).

Ce résultat sera démontré sous des hypothèses assez restrictives, qui permettent de donner une estimation de l'erreur entre la solution exacte  $H$  et la solution approchée  $H_{\mathcal{T}}$ . Ce résultat peut être largement amélioré, nous y reviendrons quand nous étudierons la convergence du schéma en 2D. Il reste intéressant de démontrer le résultat ci-dessous, dans la mesure où cela permet de mettre en place des idées qui resteront utiles dans un cadre plus général.

Avant d'énoncer le théorème correspondant, revenons sur la comparaison avec la méthodes des différences finies, pour donner cette fois un résultat négatif.

*Remarque 1.6* (Non-consistance du schéma volumes finis). Nous allons effectivement vérifier que le schéma volumes finis (1.12), (1.13) n'est *pas* consistant au sens des différences finies.

Rappelons que la consistance au sens des différences finies revient à montrer que l'erreur de troncature tend vers 0 avec le pas du maillage.

Nous considérons donc la solution du problème continu :

$$(1.22) \quad -H''(x) = f(x), \quad 0 < x < L, \quad H(0) = H(L) = 0,$$

et nous faisons l'hypothèse que  $H \in C^2(0, L)$  (ce qui sera vrai si  $f$  est assez régulière). Nous calculons alors l'erreur de troncature, qui dans notre cas est (d'après un développement de Taylor) :

$$\varepsilon_i = H''(x_i) - \frac{1}{h_i} \left( \frac{H(x_{i+1}) - H(x_i)}{h_{i+1/2}} - \frac{H(x_i) - H(x_{i-1})}{h_{i-1/2}} \right) = r_i H''(x_i) + O(h),$$

où  $r_i$  est défini par

$$r_i = 1 - \frac{h_{i+1/2} + h_{i-1/2}}{h_i}.$$

Si nous prenons, par exemple, un maillage non-régulier avec  $h_{2i} = h$  et  $h_{2i+1} = h/2$ , pour  $h > 0$  donné, et avec  $x_i$  au milieu du segment  $x_{i-1/2}, x_{i+1/2}$ , on voit que

$$r_{2i} = \frac{1}{4} \text{ alors que } r_{2i+1} = -\frac{1}{2},$$

de sorte que l'erreur de troncature ne peut pas tendre vers 0 avec  $h$ .

Nous allons néanmoins démontrer le théorème suivant, cf. [EGH00, Thm 2.1]!

**Théorème 1.2** (Convergence du schéma volumes finis). *Considérons une famille de maillage de paramètre  $h = \max_{i=1, \dots, N} h_i$ , et notons  $H_{\mathcal{T}}$  la solution du schéma défini par (1.12), (1.13) correspondant (avec la convention que  $H_0 = H_{N+1} = 0$ ).*

*L'erreur au point  $x_i$  est notée*

$$e_i^{\mathcal{T}} = H(x_i) - H_i, \quad i = 0, \dots, N+1, \quad (e_0 = e_{N+1} = 0).$$

*Sous les hypothèses suivantes :*

- le second membre  $f \in C^1(0, L)$  ;
- la solution de (1.22)  $H \in C^2(0, L)$ ,

il existe une constante  $C$  indépendante de  $h$  (mais qui peut dépendre de  $L$  et de  $f$ ) telle que

- Norme  $H^1$  discrète

$$(1.23) \quad \sum_{i=0}^N \left( \frac{e_{i+1}^{\mathcal{T}} - e_i^{\mathcal{T}}}{h_{i+1/2}} \right)^2 h_{i+1/2} \leq Ch^2,$$

- Norme du sup

$$(1.24) \quad |e_i^{\mathcal{T}}| \leq Ch, \forall 1 \leq i \leq N.$$

Le schéma est donc d'ordre 1 (la première estimation porte sur le carré de la norme).

*Démonstration.* La notion de *flux* joue un rôle fondamental dans la preuve, qui consiste à comparer différentes approximations des flux pour obtenir l'estimation cherchée. Nous définissons donc :

$$\begin{aligned} \bar{F}_{i+1/2} &= -H'(x_{i+1/2}) && \text{(flux exact) ;} \\ F_{i+1/2} &= -\frac{H_{i+1} - H_i}{h_{i+1/2}} && \text{(flux numérique) ;} \\ F_{i+1/2}^* &= -\frac{H(x_{i+1/2}) - H(x_i)}{h_{i+1/2}} && \text{(flux numérique de la solution exacte) ;} \end{aligned}$$

- Par définition du schéma :

$$(1.25) \quad F_{i+1/2} - F_{i-1/2} = h_i f_i;$$

- En intégrant l'équation continue, on obtient :

$$(1.26) \quad \bar{F}_{i+1/2} - \bar{F}_{i-1/2} = h_i f_i.$$

La régularité de  $H$  permet d'obtenir un résultat de *consistance des flux*. En notant  $R_{i+1/2} = F_{i+1/2}^* - \bar{F}_{i+1/2}$ , la formule de Taylor (puisque  $H$  est supposée de classe  $C^2$ ) montre qu'il existe  $C > 0$  (indépendante de  $h$ ) telle que

$$(1.27) \quad |R_{i+1/2}| \leq Ch,$$

en posant  $h = \max_i h_i$ .

En soustrayant (1.26) de (1.25) on obtient

$$(F_{i+1/2} - \bar{F}_{i+1/2}) - (F_{i-1/2} - \bar{F}_{i-1/2}) = 0..$$

ce qui permet alors de relier  $F_{i+1/2}$  et  $F_{i+1/2}^*$  :

$$(1.28) \quad (F_{i+1/2}^* - F_{i+1/2}) - (F_{i-1/2}^* - F_{i-1/2}) = R_{i+1/2} - R_{i-1/2}.$$

Nous pouvons maintenant relier l'erreur  $e_i$  à  $R$ . L'égalité précédente nous dit simplement que

$$-\frac{e_{i+1} - e_i}{h_{i+1/2}} + \frac{e_i - e_{i-1}}{h_{i-1/2}} = R_{i+1/2} - R_{i-1/2}, \quad i = 1, \dots, N$$

avec  $e_0 = e_{N+1} = 0$ .

On multiplie l'équation précédente par  $e_i$ , et on somme par rapport à  $i$ . Il vient :

$$-\sum_{i=1}^N \frac{(e_{i+1} - e_i)e_i}{h_{i+1/2}} + \sum_{i=1}^N \frac{(e_i - e_{i-1})e_i}{h_{i-1/2}} = \sum_{i=1}^N R_{i+1/2}e_i - \sum_{i=1}^N R_{i-1/2}e_i.$$

On peut réécrire cette égalité sous la forme (en tenant compte des conditions aux limites)

$$-\sum_{i=1}^N \frac{(e_{i+1} - e_i)e_i}{h_{i+1/2}} + \sum_{i=0}^N \frac{(e_{i+1} - e_i)e_{i+1}}{h_{i+1/2}} = \sum_{i=0}^N R_{i+1/2}e_i - \sum_{i=0}^N R_{i+1/2}e_{i+1} = \sum_{i=0}^N R_{i+1/2}(e_i - e_{i+1}),$$

et la majoration (1.27) donne alors

$$\sum_{i=0}^N \frac{(e_{i+1} - e_i)^2}{h_{i+1/2}} \leq Ch \sum_{i=0}^N |e_{i+1} - e_i|.$$

On peut majorer le second membre par l'inégalité de Cauchy-Schwarz :

$$\sum_{i=0}^N |e_{i+1} - e_i| \leq \left( \sum_{i=0}^N \frac{(e_{i+1} - e_i)^2}{h_{i+1/2}} \right)^{1/2} \cdot \left( \sum_{i=0}^N h_{i+1/2} \right)^{1/2},$$

et on remarque que  $\sum_{i=0}^N h_{i+1/2} = L$ . Donc, en désignant toujours par  $C$  une constante différente (toujours indépendante de  $h$ ),

$$\sum_{i=0}^N \frac{(e_{i+1} - e_i)^2}{h_{i+1/2}} \leq Ch^2,$$

qui est bien la première estimation cherchée.

La seconde estimation s'en déduit en écrivant :

$$e_i = \sum_{j=1}^i (e_j - e_{j-1}),$$

et en appliquant une nouvelle fois l'inégalité de Cauchy-Schwarz. ■

## 1.2.5 Quelques éléments de mise en oeuvre

### Calcul de la matrice et du second membre

**Données** Sommets du maillage (extrémités des mailles), centres des mailles, conditions aux limites, perméabilité (une valeur par maille), terme source (une valeur par maille).

**Assemblage** Tableau  $\text{TAU}(N+1)$ ,  $\text{TAU}(i) = \tau_{i+1/2} = \frac{\bar{K}_{i+1/2}}{h_{i+1/2}}$ .

Pour calculer  $A_\tau$ , technique de *l'assemblage*. Boucle sur les éléments, pour chaque élément, on ajoute la contribution aux deux arêtes. L'élément  $T_i$  contribue aux arêtes en position  $i - 1$  et  $i + 1$ . Dans un premier temps, on ne tient pas compte des conditions aux limites

**pour  $i = 2, \dots, N$  faire**

$$\begin{aligned}
A_\tau(i-1, i-1) &= A_\tau(i-1, i-1) + \text{TAU}(i) \\
A_\tau(i, i) &= A_\tau(i, i) + \text{TAU}(i) \\
A_\tau(i-1, i) &= A_\tau(i-1, i) - \text{TAU}(i) \\
A_\tau(i, i-1) &= A_\tau(i, i-1) - \text{TAU}(i)
\end{aligned}$$

**fin pour**  
**pour**  $i = 1, \dots, N$  **faire**  
 $b_\tau(i) = h_i f_i$ .  
**fin pour**

### Conditions aux limites

**Si CL gauche est Dirichlet**  $A_\tau(1, 1) = A_\tau(1, 1) + \frac{K_{1/2}}{h_{1/2}}$ ,  $b_\tau(1) = b_\tau(1) + \frac{K_{1/2}}{h_{1/2}} H_g$  ;

**Si CL gauche est Neumann** Pas de modification de la matrice,  $b_\tau(1) = b_\tau(1) + g_g$  ;

**Si CL droite est Dirichlet**  $A_\tau(N, N) = A_\tau(N, N) + \frac{K_{N+1/2}}{h_{N+1/2}}$ ,  $b_\tau(N) = b_\tau(N) + \frac{K_{N+1/2}}{h_{N+1/2}} H_d$  ;

**Si CL droite est Neumann** Pas de modification de la matrice,  $b_\tau(N) = b_\tau(N) + g_d$ .

La matrice obtenue est tridiagonale, symétrique et définie positive, le système  $A_\tau H_\tau = b_\tau$  peut-être résolu par la variante tridiagonale de l'algorithme d'élimination de Gauss, qui est dans ce cas de complexité  $O(N)$ .

## 1.3 Volumes finis en dimension 2 ou 3

Dans cette partie  $\Omega$  est un ouvert borné de  $\mathbf{R}^d$  (avec  $d = 2$  ou  $3$ ), qui sera supposé polygonal en 2D, ou polyédral en 3D. Pour simplifier dans la suite nous écrirons « polygonal » en considérant les deux situations. Cette régularité est suffisante pour analyser les méthodes (on peut définir une normale presque partout), et permet un maillage exact du domaine par des éléments de forme simple.

### 1.3.1 Maillage – inconnues discrètes

**Définition 1.2** (Maillage admissible orthogonal). On se donne

- une famille  $T \in \mathcal{T}$  de sous-domaines compacts polygonaux convexes (non vides) tels que :

$$\begin{aligned}
\overset{\circ}{T} \cap \overset{\circ}{T}' &= \emptyset, \quad \text{si } T \neq T' \\
\Omega &= \bigcup_{T \in \mathcal{T}} T.
\end{aligned}$$

- Une famille de segments (faces)  $\mathcal{E}$  tels que pour  $T \in \mathcal{T}$  on ait  $\partial T = \cup_{\sigma \in \mathcal{E}_K} \bar{\sigma}$ , avec  $\mathcal{E}_K \subset \mathcal{E}$ .
- Une famille de points  $(x_T)_{T \in \mathcal{T}} \in T$  tels que si  $(T, T') \in \mathcal{T}^2$  avec  $T \neq T'$ , soit  $\bar{T} \cap \bar{T}'$  est vide ou réduit à un point, soit  $\bar{T} \cap \bar{T}' \sigma \in \mathcal{E}$ . Dans ce cas on notera  $\sigma = T|T'$ .

On suppose de plus que les points  $x_T$  peuvent être choisis de sorte que le segment  $[x_T, x_{T'}]$  est *orthogonal* à  $T|T'$ .

*Remarque 1.7* (Orthogonalité). La dernière hypothèse constitue une restriction importante sur le maillage.

Elle est vérifiée en 2D si les éléments sont tous des triangles aigus, ou proviennent d'une triangulation de Delaunay (voir [EGH00, exemple 3.2] pour des détails), ou si les éléments sont des rectangles droits.

En 3D, elle est vérifiée pour des hexaèdres droits, mais pas nécessairement pour des tétraèdres.

## Notations

- Ensembles d'arêtes (faces)
  - $\mathcal{E}$  toutes les arêtes du maillage ;
  - $\mathcal{E}_{\text{int}}$  arêtes intérieures (communes à deux éléments) ;
  - $\mathcal{E}_{\text{ext}}$  arêtes du bord,  $\sigma \subset \partial\Omega$  ;
  - $\mathcal{E}_T$  arêtes de l'élément  $T \in \mathcal{T}$ .
- Normale :  $\nu_T$  normale *sortante* à  $T$ ,  $\nu_{T,\sigma}$  normale sortante à  $T$  le long de  $\sigma$ ,  $\nu_{T,\sigma} \perp \sigma$  ;
- Mesures, distances
  - $|T|$  aire de  $T \in \mathcal{T}$  ;
  - $|\sigma|$  longueur de  $\sigma \in \mathcal{E}$  ;
  - $d_{T,T'} = d(x_T, x_{T'})$ ,  $d_\sigma = d_{TT'}$  si  $\sigma = T|T'$  ;
  - $d_{T\sigma} = d(x_T, \sigma)$
- Taille du maillage  $\text{size}(\mathcal{T}) = \max_{T \in \mathcal{T}}(\text{diam}(T))$ , où  $\text{diam}(T)$  est la plus grande distance entre deux points de  $T$ .

**Exemple 1.3** (Cas des triangles).

Voir figure 1.3.1

à faire

FIGURE 1.4 – Quantités associées au maillage

## Choix des inconnues discrètes

Une inconnue par maille

$$(1.29) \quad H_{\mathcal{T}} = (H_T)_{T \in \text{cal}\mathcal{T}}, \quad H_T \approx \frac{1}{|T|} \int_T H \, dx$$

### 1.3.2 Construction du schéma dans le cas homogène

On résout le problème d'écoulement stationnaire (1.1)– (1.5). Même méthode qu'en dimension 1.

## Flux discrets

Intégration sur la maille

On intègre la loi de conservation (1.1) sur une maille  $T \in \mathcal{T}$ , en utilisant la *formule de Green*

$$(1.30) \quad \int_T \operatorname{div} \vec{u} = \int_{\partial T} \vec{u} \cdot \nu_T d\gamma = \sum_{\sigma \in \mathcal{E}_T} \int_{\sigma} \vec{u} \cdot \nu_{T\sigma}.$$

Notations :

—  $\bar{F}_{T,\sigma} = \int_{\sigma} \vec{u} \cdot \nu_{T\sigma}$ , flux exact ;

—  $f_T = \frac{1}{|T|} \int_T f$

Loi de conservation (1.1) donne :

$$(1.31) \quad \sum_{\sigma \in \mathcal{E}_T} \bar{F}_{T,\sigma} = |T| f_T, \quad \forall T \in \mathcal{T}.$$

## Définition des flux approchés

$F_{T,\sigma} \approx \int_{\sigma} \vec{u} \cdot \nu_{T\sigma} = - \int_{\sigma} K \nabla H \cdot \nu_{T,\sigma}$

Propriété fondamentale :  $F_{T,\sigma} + F_{T',\sigma} = 0$  si  $\sigma = T|T'$  (vrai pour le flux exact  $\bar{F}_{T,\sigma}$ ).  
*Conservativité* du schéma.

**Arête intérieure**  $\sigma = T|T'$ . Condition d'orthogonale :  $x_{T'} - x_T = d_{T,T'} \nu_{T\sigma}$

Pour  $x \in \sigma$ ,  $\nabla H(x) \cdot \nu_{T\sigma} \frac{H_{T'} - H_T}{d_{T,T'}} + O(\text{size}(\mathcal{T}))$ .

$$(1.32) \quad \text{Flux discret : } \boxed{F_{T,\sigma} = -K |\sigma| \frac{H_{T'} - H_T}{d_{T,T'}}$$

Vérifie bien la conservativité  $F_{T,\sigma} + F_{T',\sigma} = 0$  si  $\sigma = T|T'$

## Conditions aux limites

**Arêtes Dirichlet** Dans ce cas soit  $x_{\sigma}$  la projection orthogonale de  $x_T$  sur  $\sigma$ ,  $x_{\sigma} - x_T = d_{T\sigma} \nu_{T\sigma}$ ,

$$\nabla H(x) \cdot \nu_{T\sigma} = \frac{H(x_{\sigma}) - H_T}{d_{T,T'}} = \frac{H_D(x_{\sigma}) - H_T}{d_{T\sigma}}$$

$$(1.33) \quad \boxed{F_{T\sigma} = -K |\sigma| \frac{H_D(x_{\sigma}) - H_T}{d_{T\sigma}}}$$

## Arête Neumann

$$(1.34) \quad \boxed{F_{T\sigma} = -|\sigma| g_N(x_{\sigma})}.$$

### Définition du schéma

$$(1.35) \quad \boxed{\sum_{\sigma \in \mathcal{E}_T} F_{T,\sigma} = |T| f_T \quad \forall T \in \mathcal{T}.}$$

Système de  $N$  équations à  $N$  inconnues, avec  $N = \#\mathcal{T}$ , nombre d'éléments du maillage  
Schéma appelé « à deux points » (TPFA, Two point flux approximation en Anglais), puisque le flux est approché par une différence finie à deux points.

### 1.3.3 Extension au cas hétérogène

Hypothèse de modélisation : le maillage respecte les discontinuités de la perméabilité.  
On prend  $K$  **constant par maille** :

$$K_{\mathcal{T}} = (K_T)_{T \in \mathcal{T}}.$$

Comme en 1D, on cherche un flux approché sous la forme

$$(1.36) \quad F_{T,\sigma} = -|\sigma| K_{\sigma} \frac{H_{T'} - H_T}{d_{TT'}}, \quad \sigma = T|T'$$

où  $K_{\sigma}$  est à définir pour assurer la conservativité.

Introduction d'une inconnue fictive  $H_{\sigma}$ , flux local sur chaque élément  $F_{T,\sigma} = -|\sigma| K_{\sigma} \frac{H_{\sigma} - H_T}{d_{TT'}}$   
et conservativité :  $F_{T,\sigma} + F_{T',\sigma} = 0$ .

On trouve

$$(1.37) \quad K_{\sigma} = \frac{d_{TT'}}{\frac{d_{T\sigma}}{K_T} + \frac{d_{T'\sigma}}{K_{T'}}}$$

et on retrouve une *moyenne harmonique* :

$$\frac{d_{TT'}}{K_{\sigma}} = \frac{d_{T\sigma}}{K_T} + \frac{d_{T'\sigma}}{K_{T'}}, \quad d_{TT'} = d_{T\sigma} + d_{T'\sigma}.$$

*Remarque 1.8* (Moyenne harmonique). On démontre (voir [Boy10]) que ce choix est celui qui permet d'obtenir une méthode *d'ordre 1*. Tout autre choix de moyenne provoque la perte d'un demi-ordre de convergence.

*Remarque 1.9* (Conditions aux limites). Il n'y a pas de changement par rapport au cas homogène, à condition de définir  $K_{\sigma} = K_T$  pour les arêtes du bord Dirichlet.

Il est instructif de détailler le système 1.35 en séparant les contributions des arêtes internes et des arêtes du bord, en fonction du type de condition aux limites, et en ne faisant intervenir que les inconnues discrètes  $(H_T)_{T \in \mathcal{T}}$ . Il est habituel de définir la où la *transmissivité*  $\tau_{\sigma}$  par

$$(1.38) \quad \tau_{\sigma} = |\sigma| K_{\sigma} \frac{1}{d_{\sigma}}.$$

Le schéma s'écrit alors

$$(1.39) \quad \sum_{\substack{\sigma \in \mathcal{E}_T \cap \mathcal{E}_{\text{int}} \\ \sigma = T|T'}} \tau_\sigma (H_T - H_{T'}) + \sum_{\sigma \in \mathcal{E}_T \cap \mathcal{E}_D} \tau_\sigma (H_T - H_{D\sigma}) - \sum_{\sigma \in \mathcal{E}_T \cap \mathcal{E}_N} \sigma g_\sigma = |T| f_T, \quad \forall T \in \mathcal{T}_h,$$

ou, en séparant les quantités connues et inconnues :

$$\sum_{\substack{\sigma \in \mathcal{E}_T \cap \mathcal{E}_{\text{int}} \\ \sigma = T|T'}} \tau_\sigma (H_T - H_{T'}) + \left( \sum_{\sigma \in \mathcal{E}_T \cap \mathcal{E}_D} \tau_\sigma \right) H_T = |T| f_T + \sum_{\sigma \in \mathcal{E}_T \cap \mathcal{E}_D} \tau_\sigma H_{D\sigma} + \sum_{\sigma \in \mathcal{E}_T \cap \mathcal{E}_N} \sigma g_\sigma, \quad \forall T \in \mathcal{T}_h,$$

### 1.3.4 Propriétés du schéma

Dans ce paragraphe nous nous restreignons aux conditions de Dirichlet homogènes sur tout le bord.

**Théorème 1.3** (Intégration par parties discrètes). *Soit  $H_{\mathcal{T}}$  une éventuelle solution du problème discret (1.35). Pour tout  $G_{\mathcal{T}} \in \mathbf{R}^N$ ,*

$$(1.40) \quad \sum_{\substack{\sigma \in \mathcal{E} \\ \sigma = T|T'}} |\sigma| K_\sigma \frac{H_{T'} - H_T}{d_\sigma} \frac{G_{T'} - G_T}{d_\sigma} d_\sigma = \sum_{T \in \mathcal{T}} |T| f_T G_T.$$

avec la convention  $H_{T'} = G_{T'} = 0$  si  $\sigma \subset T$  est une arête du bord.

*Démonstration.* On multiplie l'équation discrète (1.35) par  $G_T$  et l'on somme sur  $T \in \mathcal{T}$ , pour obtenir :

$$\sum_{T \in \mathcal{T}} \left( \sum_{\sigma \in \mathcal{E}_T} F_{T,\sigma} \right) G_T = \sum_{T \in \mathcal{T}} |T| f_T G_T.$$

Pour transformer le premier membre, nous utilisons une technique qui revient fréquemment dans l'analyse des méthodes de volumes finis. On réordonne la double somme comme une somme sur l'ensemble des arêtes. Les arêtes du bord sont elles comptées une seule fois. On obtient donc

$$\sum_{T \in \mathcal{T}} \left( \sum_{\sigma \in \mathcal{E}_T} F_{T,\sigma} \right) G_T = \sum_{\substack{\sigma \in \mathcal{E} \\ \sigma = T|T'}} (F_{T,\sigma} G_T + F_{T',\sigma} G_{T'}) = \sum_{\substack{\sigma \in \mathcal{E} \\ \sigma = T|T'}} F_{T,\sigma} (G_T - G_{T'}),$$

où la deuxième égalité résulte de la conservativité des flux. Cela donne bien la relation attendue, étant donné la définition du flux discret (1.36).  $\blacksquare$

La relation (1.40) est une *formulation variationnelle discrète*. De plus, on vérifie que la quantité

$$[H_{\mathcal{T}}, G_{\mathcal{T}}] \stackrel{\text{def}}{=} \sum_{\substack{\sigma \in \mathcal{E} \\ \sigma = T|T'}} |\sigma| K_\sigma \frac{H_{T'} - H_T}{d_\sigma} \frac{G_{T'} - G_T}{d_\sigma} d_\sigma$$

définit un produit scalaire sur  $\mathbf{R}^{|\mathcal{T}|}$  (toujours sous la convention d'extension des vecteurs par 0 sur les arêtes de Dirichlet).

**Corollaire 1.1.** *Le schéma discret (1.39) admet une solution unique.*

*Démonstration.* Le problème discret s'écrit comme un système linéaire à matrice carrée, et le théorème 1.3 montre qu'une éventuelle solution est unique, et donne également l'existence d'une solution. ■

**Proposition 1.3** (Principe du maximum). *Si le second membre vérifie  $f_T \geq 0, \forall T \in \mathcal{T}$ , alors  $H_T \geq 0, \forall T \in \mathcal{T}$ .*

La démonstration est analogue à celle du résultat en dimension 1, proposition 1.2, en utilisant la positivité des transmissivités.

### Propriétés de la matrice du système

Nous notons  $A_{\mathcal{T}}$  la matrice du système linéaire (1.39). La matrice  $A_{\mathcal{T}}$  possède les propriétés suivantes :

- $A_{\mathcal{T}}$  est symétrique et définie positive (d'après le théorème 1.3) ;
- La matrice  $A_{\mathcal{T}}$  est une matrice creuse : l'élément  $A_{TT'}$  est non nul si et seulement si  $T$  et  $T'$  sont voisins (avec une arête commune). Pour un maillage triangulaire (en 2D), chaque triangle a au plus 3 voisins, ce qui veut dire qu'une ligne de la matrice a, au plus, quatre éléments non-nuls.
- $A_{\mathcal{T}}$  est une M-matrice, ce qui veut dire qu'elle vérifie

$$(A_{\mathcal{T}}^{-1})_{TT'} \geq 0, \quad \forall T, T'.$$

C'est une conséquence immédiate du principe du maximum 1.3.

### 1.3.5 Convergence et estimation d'erreur

Pour pouvoir parler de convergence, on doit considérer des fonctions. Nous associerons donc à chaque vecteur  $H_{\mathcal{T}} \in \mathbf{R}^{|\mathcal{T}|}$  la fonction constante sur chaque élément (que nous noterons également  $H_{\mathcal{T}}$ ) dont la valeur sur  $T$  est  $H_T$ .

**Théorème 1.4** (Convergence). *Soit  $f \in L^2(\Omega)$ ,  $K \in L^\infty(\Omega)$  vérifiant les hypothèses (1.3) et  $H \in H_0^1(\Omega)$  l'unique solution du problème continu.*

*On considère une famille de maillage admissibles  $(\mathcal{T}_n)_{n \in \mathbf{N}}$ , dont la taille tend vers 0 :  $\text{size}(\mathcal{T}_n) \xrightarrow{n \rightarrow \infty} 0$ , et on note  $H_n$  la solution correspondante.*

*On suppose que la perméabilité  $K_\sigma$  est définie par la moyenne harmonique (1.37).*

*Alors la suite  $(H_n)_{n \in \mathbf{N}} \in \mathbf{N}$  converge fortement vers  $H$  dans  $L^2(\Omega)$ .*

La preuve dépasse le niveau de ce cours. Elle peut être trouvée dans [Boy10], ou dans un cadre plus général dans [EGH00, Thm 3.1]. Elle repose sur un résultat de compacité, qui permet d'extraire une sous-suite convergente. On montre que la limite est solution faible du problème continu, et l'unicité permet de conclure à la convergence de toute la suite vers la solution du problème continu.

*Remarque 1.10.* — Ce résultat est général : il montre la convergence du schéma sans hypothèses supplémentaires par rapport au théorème d'existence pour le problème continu ;

- Ce théorème ne donne par contre *pas* de taux de convergence. Pour cela, il faut faire une hypothèse supplémentaire de *régularité* de la solution continue (cf. le théorème 1.5 ci-dessous). Cette situation est analogue à ce qui se produit dans l'étude des méthodes d'éléments finis : une estimation d'erreur demande une hypothèse de régularité ;
- Il n'y a *pas* de convergence dans  $H^1(\Omega)$  : les gradients discrets ne convergent pas ( voir la discussion dans [Boy10]) ;

**Théorème 1.5** (Estimation d'erreur pour les solutions régulières). *On prend les hypothèses du théorème 1.4. On suppose de plus que  $H \in C^2(\bar{\Omega})$ .*

*Il existe une constante  $C > 0$  indépendante de  $\text{size}(\mathcal{T})$  telle que*

$$(1.41) \quad \|H - H_{\mathcal{T}}\|_{L^2(\Omega)} \leq C \text{size}(\mathcal{T}) \|D^2H\|_{L^\infty(\Omega)},$$

où  $D^2H$  désigne la matrice des dérivées secondes de  $H$ .

Nous ne donnerons pas la démonstration, qui se trouve dans [EGH00, Thm 3.3]. Elle est analogue à la démonstration du théorème 1.2.

L'hypothèse  $H \in C^2(\bar{\Omega})$  est *très forte*, en particulier si  $K$  est discontinue. Elle peut être affaiblie en  $H \in H^2(\Omega)$ , cf [EGH00, Thm 3.4].

### Quelques éléments de mise en oeuvre

La formulation (1.39) montre que le calcul de la matrice du système discret peut s'effectuer *par arête*. Pour chaque arête, on ajoute la contribution aux deux éléments contenant cette arête (ou un seul élément si il s'agit d'une arête du bord).

Les données nécessaires sont :

- une liste des sommets, avec leurs coordonnées ;
- Pour chaque élément  $T$ 
  - les coordonnées de son centre  $x_T$  ;
  - sa mesure  $|T|$  ;
  - la perméabilité  $K_T$ , le second membre  $f_T$ .
- Pour chaque arête :
  - Sa mesure  $\sigma$  ;
  - les deux éléments voisins  $T$  et  $T'$  ;
  - les deux distances de l'arête aux centres des voisins  $d_{T,\sigma}$  et  $d_{T',\sigma}$ . On en déduit  $d_{T,T'} = d_{T,\sigma} + d_{T',\sigma}$ .

### Assemblage pour les arêtes internes

**pour  $\sigma = 1, \dots, \#(\mathcal{E}_{\text{int}})$  faire**

Déterminer les deux voisins  $T$  et  $T'$  de  $\sigma$

Calculer la transmissivité  $\text{TAU}(\sigma)$  par l'équation (1.38)

$$\begin{aligned} A_\tau(T, T) &= A_\tau(T, T) + \text{TAU}(\sigma) \\ A_\tau(T', T') &= A_\tau(T', T') + \text{TAU}(\sigma) \\ A_\tau(T, T') &= A_\tau(T, T') - \text{TAU}(\sigma) \\ A_\tau(T', T) &= A_\tau(T', T) - \text{TAU}(\sigma) \end{aligned}$$

**fin pour**

### 1.3.6 Au delà du schéma à deux points

Malgré sa simplicité, le schéma « à deux points » vu au paragraphe précédent a d'importantes limitations. Nous en indiquons certaines ci-dessous, et nous donnerons quelques indications sur des méthodes plus sophistiquées qui permettent de s'affranchir de ces restrictions. Notons qu'il s'agit encore d'un sujet de recherche en pleine activité!

*Remarque 1.11* (Limitations du schéma à deux points). Les principales limitations de ce schéma sont les suivantes :

**Diffusion anisotrope** Le schéma du paragraphe précédent ne permet de prendre en compte de manière satisfaisante le cas où la diffusion est un tenseur non-diagonal. Une telle extension existe de manière théorique sous la condition que le maillage est  $K$ -orthogonal (cf [EGH00, Defn 3.7]). Cela conduit à des maillages dépendant de la perméabilité, et n'est pas réalisable en pratique ;

**Géométrie des éléments** La condition de maillage admissible 1.2 n'est pas réalisable pour toutes les géométries d'éléments. Cette limitation est particulièrement sensible en dimension 3. Dans certaines applications, les maillages rencontrés sont souvent déformés, et ne vérifient pas les hypothèses de la définition 1.2.

**Convergence des gradients discrets** Nous avons déjà noté cette limitation après le théorème 1.4. On n'obtient qu'une approximation de la dérivée dans la direction normale à chaque arête.

Diverses généralisations de la méthode des volumes finis à deux points ont été proposées dans la littérature ces dernières années. Elles permettent de s'affranchir des limitations évoquées ci-dessus. Malheureusement, elles perdent la simplicité de la méthode à deux points, et sont également plus coûteuses à mettre en oeuvre. En restant dans le cadre des volumes finis, on peut soit ajouter des inconnues sur les arêtes (méthode DDFV, ou SUSHI), soit calculer le gradient en prenant en compte plus d'éléments (méthode MPFA). La lectrice ou le lecteur désireux d'en savoir plus pourra consulter l'article de revue [Dro14], ou les notes de cours [Boy10]. Une introduction aux méthodes multi-points se trouve dans [AGL07], et un schéma basé sur la reconstruction d'un gradient discret est présenté dans [EGH09].

Bien entendu, il est également possible d'approcher l'équation de Darcy de manière conservative par d'autres méthodes que les volumes finis.

Les méthodes *d'éléments finis mixtes* permettent d'approcher en même temps une fonction et son gradient. Elles reposent sur une formulation variationnelle dite « formulation mixte », et se déclinent selon des familles paramétrées par le degré de l'approximation. Des présentations accessibles (sans analyse) de la plus simple de ces méthodes se trouvent dans [AGL07] ou [CR91]. Le livre [Gat14] contient une introduction mathématique abordable, et une référence complète est [BBF13].

Les méthodes d'éléments finis mixtes sont utilisables sur des éléments de forme simple (triangles ou quadrangles en deux dimensions, tétraèdres ou hexaèdres réguliers en 3D). Elles prennent naturellement en compte les diffusions anisotropes. Leur principal inconvénient est leur coût élevé, lié au nombre important d'inconnues. Dans le cas le plus simple, on obtient un système linéaire avec une inconnue par arête en 2D, ou par face en 3D. Or un maillage général contient plus de faces que d'éléments.

Parmi les autres méthodes plus récentes, citons les méthodes de Galerkin Discontinu, voir [DE12 ; Riv08]. Encore plus récemment, des méthodes permettant d'utiliser des mailles basées sur des

polyèdres ou des polygones généraux ont été développées, voir par exemple [BLM14; DD20] ou [Bei+16]. Plusieurs de ces méthodes sont introduites dans [Bar+16].

## Chapitre 2

# Volumes finis pour l'équation de diffusion transitoire

### 2.1 Un (tout petit) peu de physique

#### 2.1.1 Lois de conservation

Transport d'une substance (*concentration*, température, énergie)  $c$ , flux à travers  $\partial dV$   $j$ .  
Dans un *domaine élémentaire*  $dV$ , temps  $dt$

Variation de la concentration + variation du flux à travers  $dV = 0$

$$\frac{\partial \rho c}{\partial t} + \text{div } j = 0$$

**Lois de comportement : relier  $j$  et  $c$**

**Advection**  $j = vc$ ,  $v$  vitesse d'écoulement

**Diffusion**  $j = -\mathbf{D} \text{ grad } c$  (loi de Fick)

#### Advection

Transport par un écoulement, à vitesse  $v$ .  
Pas de déformation du profil, voir Figure 2.1.

#### Diffusion

Mouvement dû à l'agitation moléculaire, des régions de forte concentration vers les régions de faible concentration, voir Figure 2.2.

#### Un exemple : transport de polluant

- $c$  concentration
- $\rho$  masse volumique
- $\phi$  porosité
- $\mathbf{j}$  flux de masse

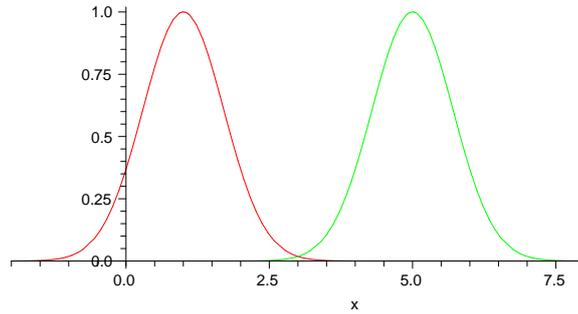


FIGURE 2.1 – Transport par advection

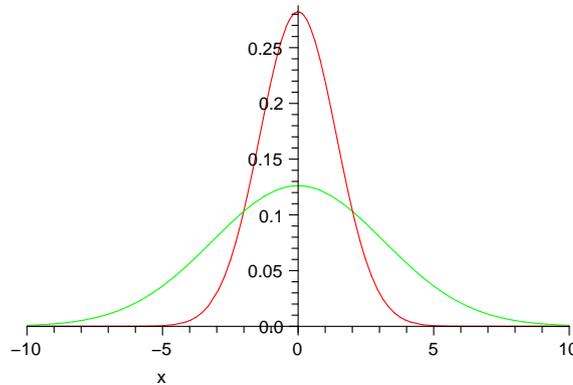


FIGURE 2.2 – Transport par diffusion

—  $f$  source de concentration

**Conservation de la masse**  $\rho\phi \frac{\partial c}{\partial t} + \nabla \cdot (\rho \mathbf{j}) = f$

**Définition du flux**  $\mathbf{j} = \mathbf{j}_a + \mathbf{j}_d$ , avec

**Flux d'advection**  $\mathbf{j}_a = vc$ ,  $v$  vitesse de l'écoulement

**Flux de diffusion**  $\mathbf{j}_d = -\mathbf{D}\nabla c$ , (loi de Fick)  $\mathbf{D}$  : coefficient de diffusion moléculaire

### 2.1.2 De la physique aux mathématiques

Équation de convection–diffusion  $x \in \Omega$  (domaine de  $\mathbf{R}^d$ ,  $d = 1, 2$  ou  $3$ ),  $0 < t < T_f$ .

$$\rho \left( \frac{\partial c}{\partial t} + v \cdot \nabla c \right) - \nabla \cdot (\mathbf{D}\nabla c) = f, \quad x \in \Omega, 0 < t < T_f$$

- Conditions aux limites  $x \in \partial\Omega = \Gamma_D \cup \Gamma_N$ ,
  - Concentration imposée  $c(x, t) = T_d(x)$ ,  $x \in \Gamma_D$
  - Flux de matière imposé  $\mathbf{q} \cdot \mathbf{n} = q_d(x)$ ,  $x \in \Gamma_N$
- Condition initiale  $c$  donné en  $t = 0$ ,  $x \in \Omega$

## Modèles simplifiés

### Une dimension

$$\frac{\partial c}{\partial t} - \frac{\partial}{\partial x} \left( \mathbf{D} \frac{\partial c}{\partial x} - vc \right) = 0, \quad 0 < x < L$$

### Pas de convection

$$\frac{\partial c}{\partial t} - \mathbf{D} \Delta c = g, \quad x \in \Omega, 0 < t < T_f$$

### Stationnaire

$$v \cdot \nabla c - \nabla \cdot (\mathbf{D} \nabla c) = f, \quad x \in \Omega,$$

## 2.2 Volumes finis pour une équation de diffusion transitoire

Dans ce chapitre, on étudie l'équation de diffusion transitoire (l'advection, linéaire et non-linéaire sera traitée au chapitre suivant).

Le modèle est le suivant :

$$(2.1) \quad \begin{aligned} \phi \frac{\partial c}{\partial t} - \operatorname{div}(D \nabla c) &= f && \text{dans } \Omega \\ c(x, t) &= c_d(x, t) && \text{sur } \Gamma_D \\ D \nabla c \cdot n &= g(x, t) && \text{sur } \Gamma_N \\ c(x, 0) &= c_0(x) && \text{dans } \Omega. \end{aligned}$$

La description de la géométrie du domaine, du maillage et les notations sont les mêmes qu'au paragraphe 1.3. Les notations sont différentes : l'inconnue est maintenant  $c$ , et la perméabilité  $K$  est remplacée par la diffusion  $D$ .

Nous procéderons en deux étapes :

**Discretisation en espace** nous utiliserons la méthode de volumes finis du paragraphe 1.3, ce qui nous conduira à un système d'équations différentielles ordinaires ;

**Discretisation en temps** nous utiliserons une méthode de résolution d'équation différentielle pour obtenir un système entièrement discret.

On intègre donc la première équation de (2.1) sur un volume de contrôle  $T \in \mathcal{T}_h$ , pour obtenir :

$$(2.2) \quad \int_T \phi \frac{\partial c}{\partial t}(x, t) dx + \sum_{\sigma \in \mathcal{E}_T} \int_{\sigma} D \nabla c(x, t) \cdot \nu = \int_T f(x, t), \quad \forall T \in \text{cal} \mathcal{T}_h.$$

On introduit maintenant une inconnue semi-discrète, en notant

$$c_T(t) = \frac{1}{T} \int_T c(x, t) dx$$

ainsi que le second membre

$$f_T(t) = \frac{1}{T} \int_T f(x, t) dx,$$

et la condition initiale

$$c_{0T} = \frac{1}{T} \int_T c_0(x) dx.$$

On suppose également que les paramètres du milieu (porosité  $\phi$  et diffusion  $D$  sont constants sur chaque élément, et on note  $\phi_T$  et  $D_T$  leur valeur sur l'élément  $T$ .

On peut donc réécrire l'équation (2.2) sous la forme :

$$(2.3) \quad \phi_T |T| \frac{dc_T}{dt} + \sum_{\sigma \in \mathcal{E}_T} F_{T,\sigma}(t) = |T| f_T, \quad \forall T \in \mathcal{T}_h,$$

avec la définition du flux numérique obtenue en (1.36) (avec les modifications habituelles pour prendre en compte les conditions aux limites) :

$$F_{T,\sigma}(t) = -|\sigma| D_\sigma \frac{c_{T'}(t) - c_T(t)}{d_{TT'}}, \quad \sigma = T|T'$$

où la diffusion  $D_\sigma$  est obtenue par une moyenne harmonique comme en (1.37). On ajoute la condition initiale sous la forme

$$(2.4) \quad c_T(0) = c_{0T}, \forall T \in \mathcal{T}_h.$$

De manière plus explicite, en distinguant les arêtes (ou faces en 3D) internes, celles sur un bord portant une condition de Dirichlet, et celles sur un bord portant une condition de Neumann, on obtien le système semi-discrétisé en espace ( $\tau_\sigma$  a été défini à l'équation (1.38) :

$$(2.5) \quad \phi_T |T| \frac{dc_T}{dt} + \sum_{\substack{\sigma \in \mathcal{E}_T, \\ \sigma \in \mathcal{E}_{\text{int}}, \\ \sigma = T|T'}} \tau_\sigma (c_T(t) - c_{T'}(t)) + \sum_{\substack{\sigma \in \mathcal{E}_T \\ \sigma \in \mathcal{E}_D}} \tau_\sigma (c_T(t) - c_{D\sigma}(t)) - \sum_{\substack{\sigma \in \mathcal{E}_T \\ |\sigma| \in \mathcal{E}_N}} \sigma g_\sigma(t) = |T| f_T, \quad \forall T \in \mathcal{T}_h,$$

Le système (2.5)–(2.4) consitue un système d'équations différentielles ordinaires. Il est intéressant d'écrire ce système sous forme vectorielle. Par analogie avec le Chapitre (2), nous noterons  $c_{\text{cal}T} = (c_{T_1}, \dots, c_{T_N})^T$  le vecteur des concentration (avec une notation analogue pour les autres quantités). Le système différentiel s'écrit alors sous la forme

$$(2.6) \quad \Phi_{\mathcal{T}} \frac{dc_{\mathcal{T}}}{dt} + A_{\mathcal{T}} c_{\mathcal{T}} = b_{\mathcal{T}}(t), \quad 0 < t < T_f,$$

$$c_{\mathcal{T}}(0) = c_{0\mathcal{T}},$$

où la matrice  $A_{\mathcal{T}}$  est celle déjà rencontrée dans le cas stationnaire,  $\Phi_{\mathcal{T}}$  est la matrice diagonale dont l'élément de la ligne  $T$  est

$$(\Phi_{\mathcal{T}})_{T,T} = \phi_T |T|,$$

et où le vecteur  $b_{\mathcal{T}}(t)$  est défini par :

$$(b_{\mathcal{T}}(t))_T = |T| f_T(t) + \sum_{\sigma \in \mathcal{E}_T \cap \mathcal{E}_D} \tau_\sigma c_{D\sigma}(t) + \sum_{\sigma \in \mathcal{E}_T \cap \mathcal{E}_N} |\sigma| g_\sigma(t).$$

Nous allons maintenant discrétiser ce système par rapport au temps. Nous nous restreindrons à l'étude d'une des méthodes les plus simples, mais fréquemment utilisée en pratique : le *theta*-schéma.

Nous nous donnons donc un pas de temps  $\Delta t$ , choisi constant pour simplifier l'exposé, et nous notons  $N_T = T_f/\Delta t$  (on suppose également que  $\Delta T$  est choisi de manière à ce que ce rapport soit

entier). Nous noterons  $t^n = n\Delta t$ ,  $n = 0, \dots, N_T$  les instants correspondants. Par convention, pour toute fonction (continue)  $t \rightarrow X(t)$  dépendant du temps, nous noterons  $X^n = X(t^n)$ , ou éventuellement une approximation de cette quantité. Les inconnues du schéma discret sont donc les quantités  $c_{\mathcal{T}}^n$ , avec  $T \in \mathcal{T}_h$  et  $n \in [0, N_T]$ ,

Nous nous donnons également un paramètre  $\theta \in [0, 1]$ . Pour toute suite  $(X^n)_{n \in [0, N_T]}$  dépendant du temps, nous noterons

$$X^{n+\theta} = (1 - \theta)X^n + \theta X^{n+1}.$$

Le  $\theta$ -schéma est alors défini de la manière suivante.

$$(2.7) \quad \begin{aligned} \Phi_{\mathcal{T}} \frac{c_{\mathcal{T}}^{n+1} - c_{\mathcal{T}}^n}{\Delta t} + A_{\mathcal{T}} c_{\mathcal{T}}^{n+\theta} &= b_{\mathcal{T}}^{n+\theta}, \quad 0 < t < T_f, \\ c_{\mathcal{T}}^0 &= c_{0\mathcal{T}}. \end{aligned}$$

Rappelons brièvement quelques propriétés immédiates du  $\theta$ -schéma (voir [DL88, Chap 20, §2], [All12, §8.6.2]).

— Pour  $\theta = 0$ , on retrouve la méthode d'Euler explicite, pour  $\theta = 1$ , on a la méthode d'Euler implicite. Notons aussi que pour  $\theta = 1/2$ , la méthode porte le nom de Crank-Nicolson.

— Pour  $\theta = 0$  la méthode est explicite. Elle est implicite pour  $0 < \theta \leq 1$ .

— La méthode est d'ordre 1 pour  $\theta \neq 0$ . Elle devient d'ordre 2 pour  $\theta = 1/2$ .

Précisons la forme que prend le schéma pour  $\theta = 0$ , puis dans le cas général.

Pour  $\theta = 0$ , on a simplement

$$\begin{aligned} \Phi_{\mathcal{T}} \frac{c_{\mathcal{T}}^{n+1} - c_{\mathcal{T}}^n}{\Delta t} + A_{\mathcal{T}} c_{\mathcal{T}}^n &= b_{\mathcal{T}}^n, \quad 0 < t < T_f, \\ c_{\mathcal{T}}^0 &= c_{0\mathcal{T}}. \end{aligned}$$

c'est-à-dire, à chaque pas de temps :

$$\Phi c_{\mathcal{T}}^{n+1} = \Phi c_{\mathcal{T}}^n - \Delta t A_{\mathcal{T}} c_{\mathcal{T}}^n + \Delta t b_{\mathcal{T}}^n,$$

Et ce système d'équations linéaires est simple à résoudre parce que sa matrice est diagonale. Plus précisément, en revenant à la forme de  $A_{\mathcal{T}}$  vue au Chapitre 2, on obtient l'équation sur chaque élément (dans le cas où il n'y a que des conditions aux limites de type Dirichlet, pour simplifier) :

$$(2.8) \quad \phi_T |T| c_T^{n+1} = \phi_T |T| c_T^n - \Delta t \left( \sum_{\substack{\sigma \in \mathcal{E}_T \cap \mathcal{E}_{\text{int}} \\ \sigma = T|T'}} \tau_{\sigma} (c_T^n - c_{T'}^n) + \sum_{\sigma \in \mathcal{E}_T \cap \mathcal{E}_D} \tau_{\sigma} (c_T^n - c_{D\sigma}^n) + |T| f_T^n \right), \quad \forall T \in \mathcal{T}$$

Pour  $\theta > 0$ , on doit résoudre un « vrai » système linéaire. La méthode s'écrit

$$(\Phi_{\mathcal{T}} + \theta \Delta t A_{\mathcal{T}}) c_{\mathcal{T}}^{n+1} = (\Phi_{\mathcal{T}}(1 - \theta) \Delta t A_{\mathcal{T}}) c_{\mathcal{T}}^n + \Delta f^{n+\theta}.$$

La matrice de ce système est symétrique et définie positive : c'est déjà le cas de la matrice  $A_{\mathcal{T}}$ , et  $\Phi_{\mathcal{T}}$  est diagonale à éléments positifs.

### 2.2.1 Stabilité du $\theta$ -schéma

Nous allons étudier la stabilité de la méthode dans deux situations :

- Pour  $\theta = 0$ , nous établirons un résultat de stabilité conditionnel, pour la norme du sup ;
- Pour  $1/2 \leq \theta \leq 1$ , la méthode est inconditionnellement stable pour le norme  $L^2$ .

Dans les deux cas, pour étudier la stabilité, nous prendrons  $f_{\mathcal{T}} = 0$ , et des conditions de Dirichlet homogènes,

Pour  $\theta = 0$ , nous repartons de l'expression (2.8), qui donne (à cause des conditions aux limites, la somme porte sur toutes les arêtes)

$$c_T^{n+1} = \left( 1 - \frac{\Delta t}{\phi_T |T|} \sum_{\sigma \in \mathcal{E}_T} \tau_\sigma \right) c_T^n + \frac{\Delta t}{\phi_T |T|} \sum_{\substack{\sigma \in \mathcal{E}_T \\ \sigma = T|T'}} \tau_\sigma c_{T'}^n, \quad \forall T \in \mathcal{T}.$$

On constate que la somme des termes pondérant  $c_T^n$  et les différents  $c_{T'}^n$  vaut 1. Par conséquent, si tous sont positifs, l'expression précédente sera une combinaison convexe, et on obtiendra facilement la stabilité. Comme les transmissivités sont positives, une condition suffisante pour obtenir une combinaison convexe est donc que

$$(2.9) \quad \frac{\Delta t}{\phi_T |T|} \sum_{\sigma \in \mathcal{E}_T} \tau_\sigma \leq 1.$$

En revenant à la définition des transmissivités, on a donc le résultat suivant

**Théorème 2.1** (Stabilité du schéma d'Euler explicite). *Sous la condition de stabilité*

$$(2.10) \quad \Delta t \leq \min_{T \in \mathcal{T}} \frac{\phi_T |T|}{\sum_{\sigma \in \mathcal{E}_T} |\sigma| K_\sigma / d_\sigma}$$

le schéma d'Euler explicite est stable pour la norme  $\infty(\Omega)$ .

*Démonstration.* Supposons que la concentration initiale  $c_T^0$  soit comprise entre deux bornes  $c_m$  et  $c_M$  :

$$c_m \leq c_T^0 \leq c_M, \quad \forall T \in \mathcal{T}.$$

Une récurrence immédiate basée sur le calcul qui précède l'énoncé du Théorème montre que, sous l'hypothèse (2.10), cette inégalité reste vraie pour tout  $n$  :

$$c_m \leq c_T^n \leq c_M, \quad \forall T \in \mathcal{T}, \quad \forall n = 0, \dots, N_T.$$

Cette inégalité correspond bien à la stabilité dans la norme  $L^\infty(\Omega)$ , pour la suite de fonctions discrètes  $(c_T^n)_{n \in [0, \dots, N]}$ , définies comme d'habitude en considérant la fonction constante par morceaux dont la restriction à  $T$  vaut  $c_T^n$ . ■

Il est utile d'ajouter quelques commentaires à ce résultat

*Remarque 2.1.* On a en fait démontré plus que la stabilité. La double inégalité démontrée dans la preuve donne le *principe du maximum* discret. La solution du schéma d'Euler explicite (sous la condition de stabilité) respecte les mêmes bornes que la condition initiale. Il s'agit là d'une propriété forte, pas nécessairement vérifiée pour les autres valeurs de  $\theta$  (nous verrons qu'elle est également vraie pour  $\theta = 1$ ).

*Remarque 2.2* (Interprétation de la condition de stabilité). Il n'est pas facile d'interpréter la condition (2.10), en particulier vis-à-vis de la condition de stabilité « usuelle », qui prend la forme  $\frac{K\Delta t}{h^2} \leq C$  (où  $C$  est une constante, qui vaut  $1/2$  en dimension 1).

Plaçons nous dans le cas où  $\Omega$  est le carré  $[0, 1] \times [0, 1]$ , avec des propriétés physiques constantes, et d'un maillage régulier du carré, de pas  $h$ . Dans ce cas, les deux quantités  $|\sigma|$  et  $d_\sigma$  valent  $h$ , alors que  $|T| = h^2$ . La somme porte sur quatre termes, ce qui veut dire que la condition (2.10) se réduit à

$$\frac{K}{\phi} \Delta t \leq \frac{h^2}{4}.$$

Plus généralement, on démontre que si le maillage possède certaines propriétés de « régularité » (que nous ne détaillerons pas), l'expression (2.10) se réécrit également sous la forme

$$\Delta t \leq C \min_{T \in \mathcal{T}} \frac{\phi_T |T|}{K_T}$$

où  $C$  est une constante (indépendante de  $\mathcal{T}$ ).

Passons maintenant au second résultat de stabilité, qui est valable pour  $1/2 \leq \theta \leq 1$ . Pour l'exprimer nous définissons une norme sur l'espace des concentrations, qui est une norme  $L^2$  discrète pondérée :

$$(2.11) \quad \|c_{\mathcal{T}}\|_{\mathcal{T}} = \left( \sum_{T \in \mathcal{T}} \phi_T |c_T|^2 |T| \right)^{1/2}.$$

Cette quantité définit bien une norme sous l'hypothèse que la porosité est uniformément positive ( $\exists \phi_0 < \phi_o, \phi_0 \leq \phi(x), \forall x \in \Omega$ ). On alors

**Théorème 2.2.** *Sous l'hypothèse  $1/2 \leq \theta \leq 1$ , le  $\theta$ -schéma est inconditionnellement stable (sans restriction sur le pas de temps), et la solution  $(c_{\mathcal{T}}^n)_{n \in [0, N]}$  vérifie l'inégalité suivante ;*

$$(2.12) \quad \|c_{\mathcal{T}}^{n+1}\|_{\mathcal{T}} \leq \|c_{\mathcal{T}}^n\|_{\mathcal{T}}.$$

*Démonstration.* Remarquons que la norme  $\|\cdot\|_{\mathcal{T}}$  peut aussi se calculer comme

$$\|c_{\mathcal{T}}\|_{\mathcal{T}}^2 = c_{\mathcal{T}}^T \Phi_{\mathcal{T}} c_{\mathcal{T}}.$$

Pour démontrer le théorème nous prenons le produit scalaire de (2.7) avec  $c_{\mathcal{T}}^{n+\theta}$ . En utilisant l'identité

$$c^{n+\theta} = \frac{1}{2} (c^{n+1} + c^n) + \left(\theta - \frac{1}{2}\right) (c^{n+1} - c^n),$$

il vient

$$\frac{1}{2} \left( \|c_{\mathcal{T}}^{n+1}\|_{\mathcal{T}}^2 - \|c_{\mathcal{T}}^n\|_{\mathcal{T}}^2 \right) + \left(\theta - \frac{1}{2}\right) \|c_{\mathcal{T}}^{n+1} - c_{\mathcal{T}}^n\|_{\mathcal{T}}^2 + (c_{\mathcal{T}}^{n+\theta})^T A_{\mathcal{T}} c_{\mathcal{T}}^{n+\theta} = 0.$$

■

La matrice  $A_{\mathcal{T}}$  est définie positive, et par hypothèse  $\theta \geq 1/2$ . et donc les deux derniers termes sont positifs, ce qui permet de conclure.

## Chapitre 3

# Lois de conservations – méthode de Godunov

Ce chapitre contient une introduction à l’analyse numérique des lois de conservation, dans le but de présenter la méthode de Godunov. Il ne fait qu’effleurer la surface de ce sujet, en se bornant aux lois de conservation scalaires, et en ne considérant qu’un schéma d’ordre 1. Le sujet fait l’objet d’un (très) grand nombre de travaux. Les références qui suivent sont donc nécessairement incomplètes, et reflètent les connaissances (et les goûts) de l’auteur.

On trouvera des chapitres consacrés aux schémas numériques pour les lois de conservation dans les ouvrages [Asc08, chap10] et [LL16, Chap. 10]. Pour une approche par volumes finis (et avec un niveau mathématique élevé) on pourra consulter les chapitres V et VI de [EGH00]. Plusieurs livres de R. LeVeque sont consacrés à ce sujet. Le plus récent [KLd20] présente une approche originale en combinant un livre (dont le texte est également disponible en ligne) avec des *notebooks* interactifs Jupyter, permettant de voir les méthodes en action, et d’expérimenter avec les paramètres. Il est fortement recommandé de consulter le site [http://www.clawpack.org/riemann\\_book/](http://www.clawpack.org/riemann_book/). Les autres livres [LeV92; LeV02] sont plus complets, et bien sûr plus difficiles d’accès. Le second décrit la bibliothèque CLAWPACK [Cla20], qui a depuis été réécrite entièrement en Python et est disponible à l’adresse <http://www.clawpack.org>. D’autres ouvrages sur la construction et l’analyse des schémas numériques sont, par ordre de difficulté approximativement croissante [Gui03], avec une approche plus pratique, [Des10; Hes18] (avec des exemples en Matlab, voir <https://www.epfl.ch/labs/mcss/books/>) ou encore [GR21]. Le livre [Tor09], est consacré au système des équations d’Euler. Finalement, parmi les références sur l’étude mathématique (solutions faibles, conditions de choc) citons le livre (court, mais dense) de P. Lax [Lax73], ainsi que les notes de cours de S. Kruzhkov [CG09].

### 3.1 Introduction, exemples

Une *loi de conservation scalaire* (hyperbolique) en une dimension d’espace est une EDP de la forme :

$$(3.1) \quad \begin{cases} \partial_t u + \partial_x (f(u)) = 0 & x \in \mathbf{R}, t > 0 \\ u(x, 0) = u_0(x), \end{cases}$$

où  $f \in C^1(\mathbf{R})$ ,  $u_0 \in L^\infty(\mathbf{R})$  et la fonction inconnue est  $u : \mathbf{R} \rightarrow \mathbf{R}$ .

En dimension  $d$  ( $d = 2$  ou  $3$ ) d'espace, le modèle s'écrit sous la forme

$$(3.2) \quad \begin{cases} \partial_t u + \operatorname{div}(F(u)) = 0 & x \in \mathbf{R}^d, t > 0 \\ u(x, 0) = u_0(x), \end{cases}$$

où  $F \in C^1(\mathbf{R})$  est à valeurs dans  $\mathbf{R}^d$  et la fonction inconnue est  $u : \mathbf{R} \times \mathbf{R}_+ \rightarrow \mathbf{R}$ .

Au paragraphe 3.4 nous étudierons le modèle de Buckley-Leverett, qui s'écrit sous la forme

$$F(x, u) = \vec{v}(x) f(u)$$

où  $\vec{v}$  est un champ de vecteurs donné  $\vec{v} : \mathbf{R}^d \rightarrow \mathbf{R}^d$  et  $f \in C^1(\mathbf{R})$ .

*Exemple 1* (Quelques exemples). **Advection linéaire** L'exemple de base est celui que nous avons rencontré au Chapitre B. L'équation d'advection linéaire (avec  $a \in \mathbf{R}$  une constante donnée)

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = 0,$$

entre bien dans le cadre général avec  $f(u) = au$ .

**Équation de Burgers** C'est le cas  $f(u) = \frac{1}{2}u^2$ . Cet exemple est le plus simple des modèles non-linéaires. Il est à l'origine de l'étude des lois de conservation scalaires, en tant que modèle très simplifié de la dynamique des gaz.

**Traffic routier** Voir [Des10, Par. 2.1.1] ou [Kld20, Chap. 5]. On modélise le trafic sur une autoroute (à une seule voie, sans rampe d'accès). L'inconnue  $u$  représente la densité de véhicule, et on suppose que la vitesse  $v$  dépend de la densité par une formule de la forme

$$v = v_{\max} \left( 1 - \frac{u}{u_{\max}} \right).$$

Si la densité est très faible, les véhicules peuvent atteindre leur vitesse maximale, si la densité est trop élevée, la vitesse devient nulle. Après adimensionnalisation, on obtient (3.1) avec  $f(u) = u(1 - u)$ . Noter que la fonction  $f$  est encore quadratique, comme à l'exemple précédent, mais est concave alors que la fonction pour Burgers est convexe. Nous verrons que cette question joue un rôle important.

**Buckley-Leverett** Il s'agit d'un modèle simplifié d'écoulement diphasique en milieu souterrain (voir le cours d'E. Mouche, ou [LeV02, Chap. 16]). Il correspond au déplacement d'huile par injection d'eau et  $u$  représente maintenant la proportion d'eau au voisinage d'un point (par définition  $0 \leq u \leq 1$ ). L'interface entre la zone contenant de l'eau et celle contenant de l'huile est diffuse, et on montre que  $u$  satisfait encore une équation du type (3.1) avec cette fois

$$(3.3) \quad f(u) = \frac{u^2}{u^2 + a(1 - u)^2},$$

où  $a$  est une constante positive, qui représente le rapport des mobilités des deux fluides. Cet exemple présente un changement de concavité, qui rend son analyse plus difficile que les deux exemples précédents. La figure 3.1 présente la fonction  $f$  et sa dérivée.

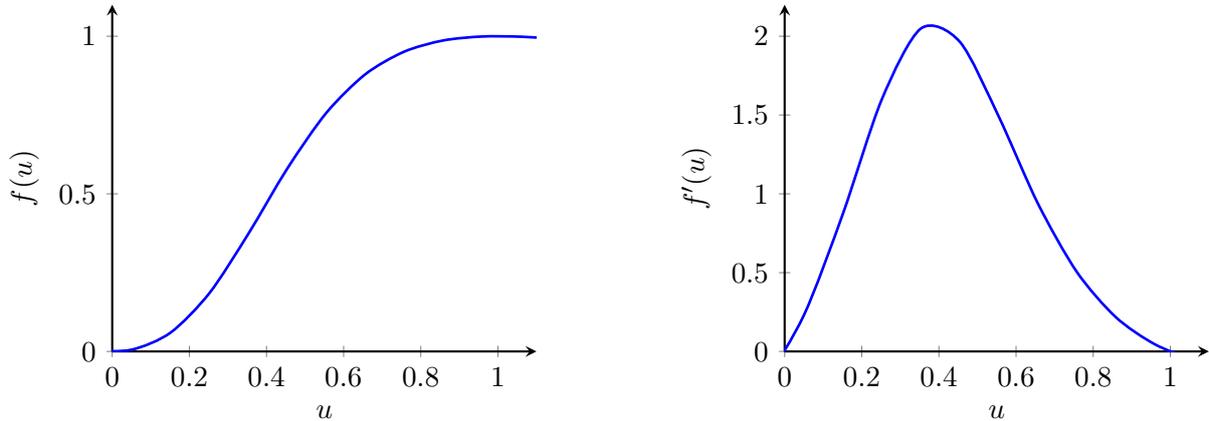


FIGURE 3.1 – Fonction de flux et sa dérivée pour le modèle de Buckley-Leverett

## 3.2 Analyse mathématique

### 3.2.1 Solution classiques – méthode des caractéristiques

Dans cette section, nous cherchons des solutions régulières (de classe  $C^1$  de (3.1)). Nous verrons qu'en général, ces solutions ne peuvent exister que sur un temps fini, et qu'il faut alors élargir la notion de solution. Mais la définition naturelle de solution faible entraîne une perte d'unicité, et nous verrons comment retrouver parmi les différentes solutions faible la « bonne » solution (celle qui correspond au phénomène physique sous-jacent, grâce à la notion d'entropie).

Remarquons tout d'abord que si  $u$  est une solution régulière de (3.1), nous pouvons utiliser la dérivation composée, pour obtenir l'équation

$$\frac{\partial u}{\partial t} + f'(u) \frac{\partial u}{\partial x} = 0, \quad x \in \mathbf{R}, t > 0.$$

Il est important d'insister sur le fait que cette manipulation n'est licite que si  $u$  est régulière!

On voit donc que (3.1) modélise un phénomène d'advection, dont la vitesse dépend de la solution elle-même.

Une **courbe caractéristique** de l'équation (3.1) est une solution  $t \rightarrow X(t)$  de l'équation différentielle ordinaire

$$(3.4) \quad \frac{dX}{dt} = f'(u(X(t), t)).$$

La propriété remarquable des caractéristiques est la suivante :

**Proposition 3.1.** *Les solutions de la loi de conservation (3.1) sont constantes le long des caractéristiques.*

*Autrement dit, si  $u$  est une solution régulière de (3.1) et si  $X$  est une caractéristique alors la fonction  $v : t \rightarrow u(X(t), t)$  est constante.*

*Démonstration.* La preuve est un simple calcul de dérivée composée :

$$\frac{dv}{dt} = \frac{d}{dx} u(X(t), t) = \frac{\partial u}{\partial x}(X(t), t) \frac{dX}{dt} + \frac{\partial u}{\partial t}(X(t), t) = 0,$$

d'après la définition des caractéristiques. ■

La remarque suivante est simple, mais fondamentale : d'après la proposition 3.1, le second membre de (3.4) est constant. Il en résulte que les solutions de (3.4) sont des fonctions affines de  $t$ , ou en termes plus géométriques que *les caractéristiques sont des droites* ! En notant  $X(0) = x_0$  la valeur initiale de la courbe, on en déduit l'équation de la caractéristique :

$$(3.5) \quad X(t) = x_0 + tf'(u_0(x_0)).$$

Dans le cas de l'équation d'advection linéaire, les caractéristiques sont des droites de pente  $a$ , puisque l'équation (3.4) devient dans ce cas

$$\frac{dX}{dt} = a.$$

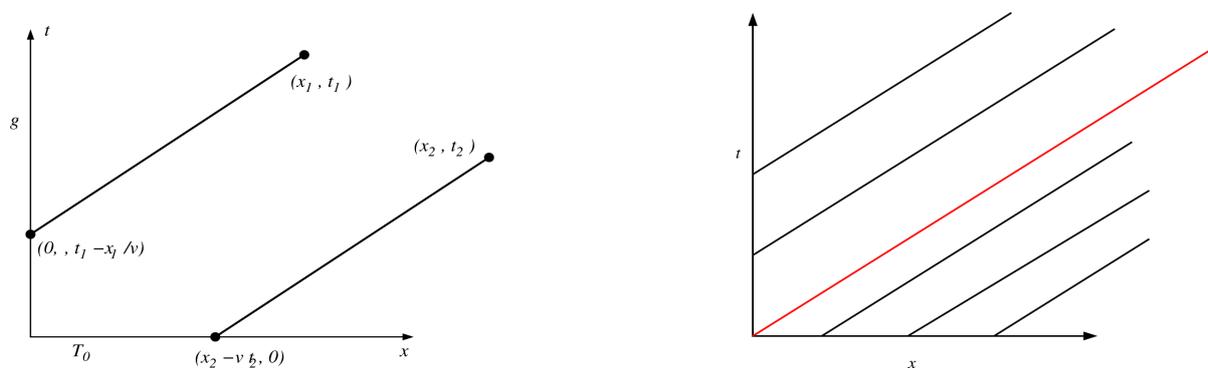


FIGURE 3.2 – Caractéristiques, vitesse constante

Il y a bien une caractéristique unique en chaque point  $(x, t)$  (voir figure A.2, et on en déduit facilement que la solution du problème de Cauchy s'écrit dans ce cas

$$(3.6) \quad u(x, t) = u_0(x - at), x \in \mathbf{R}, t > 0.$$

En suivant la figure, on peut d'ailleurs généraliser cette solution au problème sur la demi-droite  $\mathbf{R}^{+*}$  avec un condition à la limite en  $x = 0$  :

$$(3.7) \quad \begin{aligned} \frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} &= 0, & 0 < x, 0 < t \\ u(0, t) &= g(t) & 0 < t \\ u(x, 0) &= u_0(x), & 0 < x \end{aligned}$$

dont la solution (voir la figure A.2) s'écrit

$$u(x, t) = \begin{cases} g(t - x/a), & x < vt, \\ u_0(x - at), & vt < x. \end{cases}$$

Dans le cas général d'une loi non-linéaire, on a vu que les caractéristiques restent des droites, mais celles-ci ne sont plus parallèles. Dans le cas général la pente de ces droites dépend du point initial. Il en résulte que les caractéristiques peuvent se croiser, nous y reviendrons longuement !

On peut maintenant construire géométriquement les solutions régulières de (3.1). Pour cela, fixons  $x \in \mathbf{R}$  et  $t \in \mathbf{R}_+$ , et cherchons la valeur de la solution  $u(x, t)$ . D'après ce qui précède, cette solution est constante le long d'une caractéristique passant par  $x$  à l'instant  $t$ . À  $t = 0$ , une telle caractéristique sera au point  $x_0$  si

$$(3.8) \quad x = x_0 + t f'(u_0(x_0)),$$

ce qui constitue une équation implicite pour  $x_0$ . Si cette équation peut être résolue, on aura nécessairement

$$u(x, t) = u_0(x_0).$$

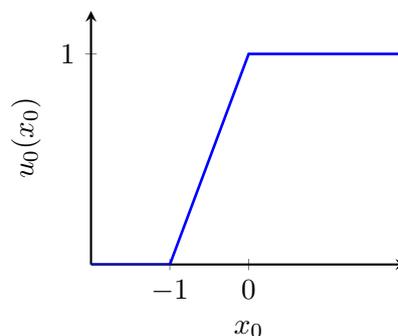
Nous allons examiner deux exemples dans le cas de l'équation de Burgers, on pourra consulter les références pour une discussion du cas général (existence en temps fini, voir [CG09]).

*Exemple 2* (Onde de détente). Dans le cas de l'équation de Burgers,  $f(u) = u$ , et l'équation des caractéristiques est

$$X(t) = x_0 + t u_0(x_0).$$

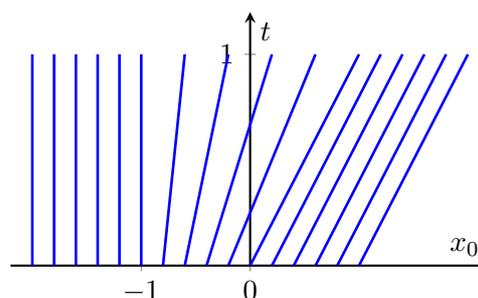
Considérons le cas

$$u_0(x_0) = \begin{cases} 0 & x_0 \leq -1 \\ 1 + x_0 & -1 \leq x_0 \leq 0 \\ 1 & 0 \leq x_0. \end{cases}$$



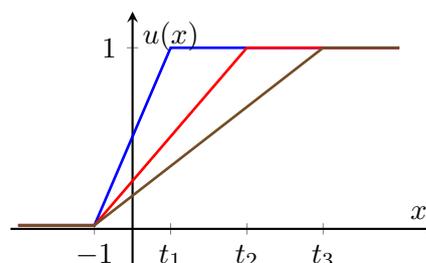
Les caractéristiques sont donc

$$X(t, x_0) = \begin{cases} x_0 & x_0 \leq -1 \\ x_0 + t(1 + x_0) & -1 \leq x_0 \leq 0 \\ x_0 + t & 0 \leq x_0. \end{cases}$$



et la solution est alors, pour  $t > 0$

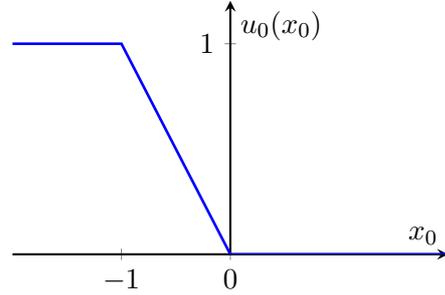
$$u(x, t) = \begin{cases} 0 & x_0 \leq -1 \\ 1 + \frac{x-t}{1+t} & -1 \leq x_0 \leq t \\ 1 & t \leq x_0. \end{cases}$$



Ces solutions sont représentées pour différents instants sur la figure à droite. Pour cet exemple, la méthode des caractéristiques permet de construire une solution pour tout  $t$ .

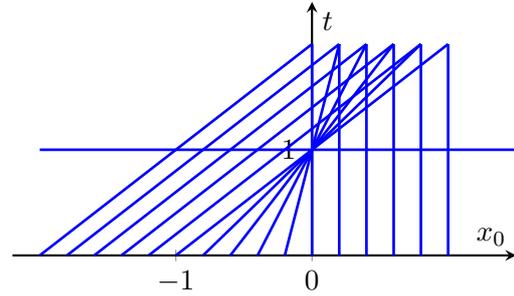
*Exemple 3* (Apparition d'un choc). On prend maintenant

$$u_0(x_0) = \begin{cases} 1 & x_0 \leq -1 \\ -x_0 & -1 \leq x_0 \leq 0 \\ 0 & 0 \leq x_0. \end{cases}$$



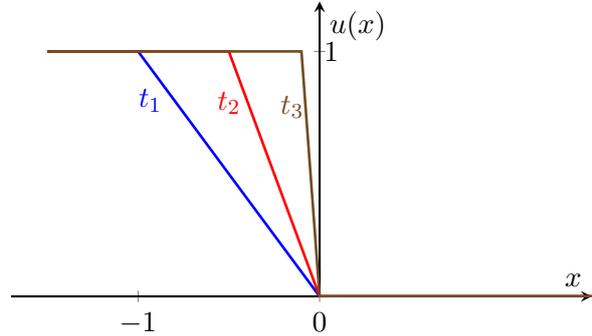
Dans ce cas les caractéristiques sont

$$X(t, x_0) = \begin{cases} x_0 + t & x_0 \leq -1 \\ x_0 - tx_0 & -1 \leq x_0 \leq 0 \\ x_0 & 0 \leq x_0. \end{cases}$$



Comme on peut le voir sur la figure, les caractéristiques se croisent au point  $(0, 0)$ . On ne peut donc construire la solution que jusqu'au temps  $t = 1$  (au-delà, l'équation non-linéaire (3.8) n'a pas une solution unique). Pour  $t < 1$  on a

$$u(x, t) = \begin{cases} 1 & x_0 \leq -1 \\ -\frac{x}{1-t} & -1 \leq x_0 \leq t \\ 0 & t \leq x_0. \end{cases}$$



À  $t = 1$ , une discontinuité apparaît, et la solution classique n'existe plus.

### 3.2.2 Solutions faibles

Les solutions discontinues comme celle de l'exemple 3 correspondent à des situations physiques, et il est important de pouvoir les prendre en compte. Nous allons donc étendre la notion de solution pour le problème (3.1), et définir les solutions faibles.

**Définition 3.1** (Solution faible d'une loi de conservation). On dit qu'une fonction  $u$  localement bornée est *solution faible* du problème (3.1) si et seulement si elle vérifie :

$$(3.9) \quad \int_{\mathbf{R}_+} \int_{\mathbf{R}} (u \partial_t \varphi + f(u) \partial_x \varphi) dx dt + \int_{\mathbf{R}} u_0(x) \varphi(x, 0) dx = 0, \quad \forall \varphi \in C_0^1(\mathbf{R} \times \mathbf{R}_+).$$

Rappelons que  $C_0^1(\mathbf{R} \times \mathbf{R}_+)$  désigne les fonctions de classe  $C^1$  à support compact, c'est-à-dire qui sont nulles pour  $t$  ou  $|x|$  assez grand (mais ces fonctions ne sont en général pas nulles en  $t = 0$ ).

On vérifie facilement (par intégration par parties) qu'une solution classique est une solution faible, et inversement qu'une solution faible régulière est une solution classique. Pour ce deuxième point, on procède en deux étapes, en choisissant d'abord  $\varphi$  nulle ne 0, ce qui donne l'EDP, puis  $\varphi$  quelconque pour obtenir la condition initiale.

Les solutions faibles peuvent présenter des discontinuités (on parle de *choc*), mais la loi de conservation contraint la vitesse de propagation de ces dernières.

**Proposition 3.2** (Relations de Rankine-Hugoniot). *Soit  $u : \mathbf{R} \times \mathbf{R}_+ \rightarrow \mathbf{R}$  une fonction localement bornée, supposée de classe  $C^1$  de part et d'autre d'une courbe régulière  $\Gamma : t \rightarrow x(t)$ . La fonction  $u$  est une solution faible de (3.1) si et seulement si*

- i)  $u$  est une solution régulière de chaque côté de  $\Gamma$  ;
- ii) la relation suivante est vérifiée :

$$(3.10) \quad -x'(t) [u(x(t)^+, t) - u(x(t)^-, t)] + [f(u(x(t)^+, t)) - f(u(x(t)^-, t))] = 0.$$

La courbe  $\Gamma$  étant régulière, elle sépare l'ouvert  $\mathbf{R} \times \mathbf{R}_+$  en deux parties. Nous notons donc  $\Omega^+ = \{(x, t), x > x(t)\}$ ,  $\Omega^- = \{(x, t), x < x(t)\}$ , et nous ajoutons un exposant + ou - aux quantités « de chaque côté de  $\Gamma$  ».

*Démonstration.* Nous considérons des fonctions test  $\varphi$  qui s'annulent en  $t = 0$ , de manière à ce que le terme contenant la condition initiale dans (3.9) disparaisse. En coupant l'intégrale en espace en deux parties, on obtient alors

$$\int \int_{\Omega^-} (u \partial_t \varphi + f(u) \partial_x \varphi) dx dt + \int \int_{\Omega^+} (u \partial_t \varphi + f(u) \partial_x \varphi) dx dt = 0.$$

Comme  $u$  est supposée régulière dans  $\Omega^-$  et  $\Omega^+$ , nous pouvons appliquer la formule de Green dans chaque domaine. Nous noterons  $\mathbf{n}^\pm$  la normale unitaire *sortante* à  $\Omega^\pm$  le long de  $\Gamma$ . Le champ de vecteur dans le plan  $(x, t)$  auquel nous appliquons la formule de Green est  $\begin{pmatrix} f(u) \\ u \end{pmatrix}$ .

Nous obtenons alors :

$$\begin{aligned} \int_{\Omega^-} -(\partial_t u + \partial_x f(u)) \varphi + \int_{\Omega^+} -(\partial_t u + \partial_x f(u)) \varphi \\ + \int_{\Gamma} \begin{pmatrix} f(u^-) \\ u^- \end{pmatrix} \cdot \mathbf{n}^- \varphi d\gamma + \int_{\Gamma} \begin{pmatrix} f(u^+) \\ u^+ \end{pmatrix} \cdot \mathbf{n}^+ \varphi d\gamma = 0 \end{aligned}$$

En prenant une fonction test  $\varphi$  à support contenu dans  $\Omega^-$  (resp.  $\Omega^+$ ), on voit que  $u$  est une solution régulière de (3.1) dans  $\Omega^-$  (resp.  $\Omega^+$ ).

La première ligne de l'équation précédent est donc nulle *pour toute fonction test*  $\varphi$ , et il ne reste alors que les deux intégrales le long de  $\Gamma$ . Il nous reste à expliciter l'expression des normales  $\mathbf{n}^\pm$ . À une normalisation près (pour les rendre unitaires), on a

$$\mathbf{n}^\pm = \pm \begin{pmatrix} 1 \\ -x'(t) \end{pmatrix},$$

et on en déduit, pour tout  $\varphi \in C_0^1(\mathbf{R} \times \mathbf{R}_+)$  :

$$\int_{\Gamma} [f(u(x(t)^+, t)) - f(u(x(t)^-, t))] \varphi d\gamma - \int_{\Gamma} x'(t) [u(x(t)^+, t) - u(x(t)^-, t)] \varphi(x(t), t) d\gamma = 0,$$

ce qui donne la relation (3.10). ■

Si la fonction  $u$  est continue à travers  $\Gamma$ , on retrouve le fait que c'est une solution classique. Si elle est discontinue, la relation (3.10) donne la vitesse de déplacement du choc. Cette proposition est presque toujours utilisée de la manière suivante :

**Corollaire 3.1** (Relation de Rankine-Hugoniot). *On considère la fonction  $u$  définie par (on suppose  $u_L \neq u_R$ )*

$$u(x, t) = \begin{cases} u_L & x < \sigma t \\ u_R & x > \sigma t. \end{cases}$$

$u$  est solution faible de (3.1) si et seulement si :

$$(3.11) \quad \sigma = \frac{f(u_R) - f(u_L)}{u_R - u_L} = \frac{[[f(u)]]}{[[u]]}.$$

Cette relation exprime que les discontinuités des solutions faibles ne peuvent se propager qu'aux vitesses autorisées par la relation (3.10), ou si c'est à vitesse constante, par la relation (3.11).

Dans le cas de l'équation de Burgers, on obtient par exemple

$$\sigma = \frac{u_R + u_L}{2}.$$

*Remarque 3.1* (Ne pas manipuler des lois de conservation). Cette remarque insiste sur le danger de raisonner sur les lois de conservation comme si les solutions étaient régulières (cf. [LeV92, Sec. 3.7]). Prenons ainsi l'exemple de l'équation de Burgers :

$$(3.12) \quad \partial_t u + \frac{1}{2} \partial_x (u^2) = 0.$$

Comme nous venons de le voir, les chocs possibles se déplacent à la vitesse  $\sigma = \frac{u_R + u_L}{2}$ . Si on multiplie « naïvement » cette équation par  $2u$ , on obtient la loi de conservation

$$\partial_t (u^2) + \frac{2}{3} \partial_x (u^3) = 0.$$

Les solutions régulières de ces deux équations sont les mêmes, mais *pas les solutions faibles !* D'après la relation de Rankine-Hugoniot, les chocs de cette dernière équation se déplacent à la vitesse

$$\sigma' = \frac{2 u_R^3 - u_L^3}{3 u_R^2 - u_L^2} \neq \sigma.$$

Ainsi, la transformation de l'équation a changé la vitesse des chocs !

**Exemple 3.1** (Application à l'équation de Buckley-Leverett).

Nous allons construire une solution faible pour l'équation de Buckley-Leverett (3.3) dans le cas du problème de Riemann, c'est-à-dire lorsque la condition initiale est de la forme

$$u_o(x) = \begin{cases} 1 & x < 0 \\ 0 & x > 0, \end{cases}$$

ce qui correspond à une injection d'eau dans un milieu contenant uniquement de l'huile initialement.

Comme on l'a vu en (3.8), la solution peut s'écrire

$$u(x + tf'(u_0(x)), t) = u_0(x),$$

ce qui peut s'interpréter en disant que l'on « décale » la condition initiale d'une quantité  $tf'(u_0(x))$  pour obtenir la solution au point  $x$  à l'instant  $t$ . Cela revient à regarder la graphe de  $f'$  après une rotation de  $90^\circ$ , comme sur la figure 3.3. Mais la solution devient multi-valuée, et on doit donc « introduire » un choc, à une position qui est déterminée par la relation de Rankine-Hugoniot. Notons  $u^*$  la valeur de la solution juste avant le choc, et  $x(t)$  sa position à

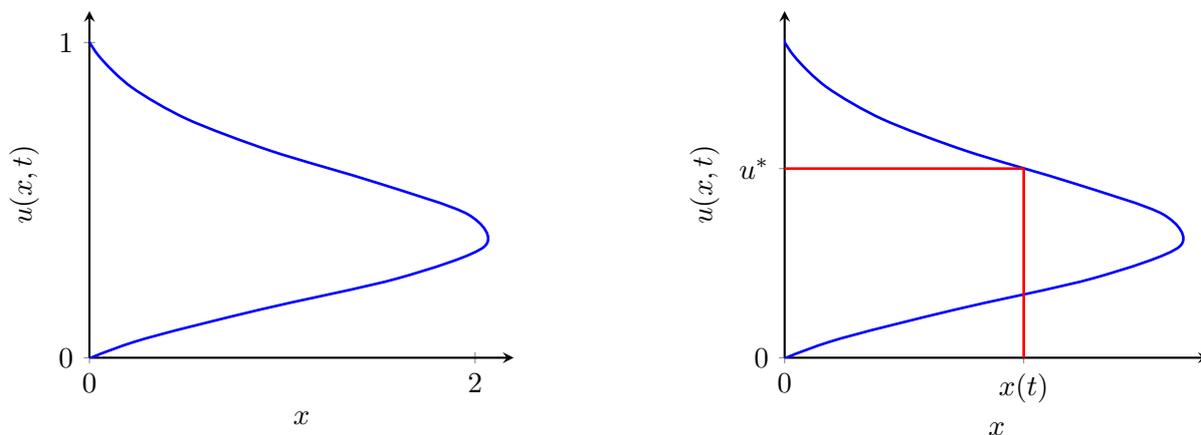


FIGURE 3.3 – solution de l'équation de Buckley-Leverett. À gauche solution par les caractéristiques, la solution devient multi-valuée. À droite, introduction d'un choc d'après la relation de Rankine-Hugoniot.

l'instant  $t$ .

- D'après les caractéristiques, le point  $(x(t), u^*)$  se trouve sur la courbe, donc  $x(t) = tf'(u^*)$ ;
- La relation de Rankine Hugoniot donne la vitesse de déplacement du choc :

$$x'(t) = \frac{0 - f(u^*)}{0 - u^*} = \frac{f(u^*)}{u^*}.$$

En comparant ces deux égalités (dériver la première!), on trouve que la valeur de  $u$  en amont du choc doit vérifier

$$f(u^*) = u^* f'(u^*),$$

ce qui exprime que le point  $u^*$  est celui où la tangente à la courbe de flux est égal à la corde entre l'origine et  $(u^*, f(u^*))$ .

Noter que l'on retrouve ici la construction de la tangente de Welge (voir le cours d'E. Mouche). Noter également que ce choix revient à remplacer la courbe de flux par *son enveloppe convexe*. Cette remarque est générale, voir [CG09] ou [LeV92, pp. 49-50].

La définition des solutions faibles permet donc bien de prendre en compte des solutions discontinues. Malheureusement, cette définition s'accompagne d'une perte d'unicité comme le montre l'exemple suivant.

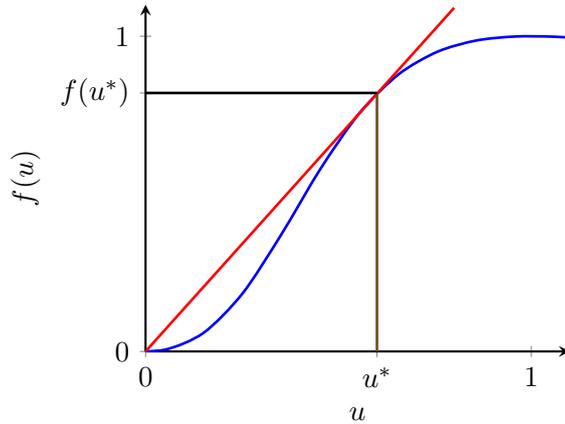


FIGURE 3.4 – Construction de la tangente de Welge et de l’enveloppe convexe

**Exemple 3.2** (Non-unicité des solutions faibles).

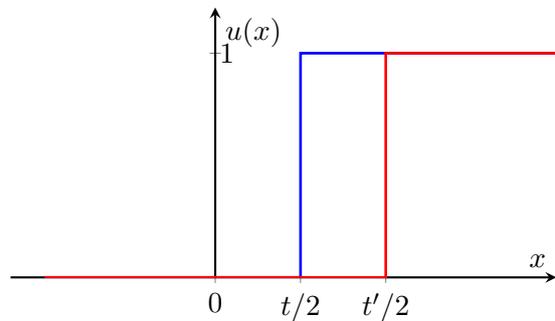
On considère l’équation de Burgers, avec la conditions initiale :

$$u_0(x) = \begin{cases} 0, & x < 0 \\ 1, & x > 0. \end{cases}$$

Nous allons construire deux solutions distinctes :

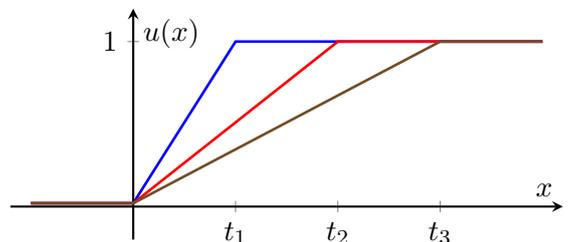
- i) Un choc, se propageant vers la droite à vitesse  $\frac{1}{2}$ , conformément à la relation de Rankine-Hugoniot :

$$u(x, t) = \begin{cases} 0 & x < t/2, \\ 1 & x > t/2. \end{cases}$$



- ii) Une onde de détente, qui est une solution continue.

$$u(x, t) = \begin{cases} 0, & x \leq 0 \\ \frac{x}{t}, & 0 \leq x \leq t \\ 1, & t \leq x. \end{cases}$$



On peut se convaincre que, pour des raisons physiques, la « bonne » solution est la détente. En effet, la solution du choc n’est pas stable par perturbations. Une autre raison est que la détente est une solution régulière, c’est celle qui serait obtenue par les caractéristiques. La solution du choc est donc un artefact de la définition de solution faible, et il nous faut donc un critère permettant de l’éliminer.

### 3.2.3 Solutions faibles entropiques

Le critère qui permet de discriminer parmi les solutions faibles celle qui correspond à une solution physique est la notion d'entropie, et de solution entropique. Une remarque : il s'agit ici de l'entropie mathématique, qui n'est pas la même que celle des physiciens !

La motivation pour introduire l'entropie vient de la considération d'un modèle qui est proche de (3.1) mais contient de la dissipation. On considère le problème parabolique non-linéaire :

$$(3.13) \quad \begin{cases} \partial_t u_\varepsilon + \partial_x (f(u_\varepsilon)) - \varepsilon \partial_{xx} u_\varepsilon = 0 & x \in \mathbf{R}, t > 0 \\ u_\varepsilon(x, 0) = u_0(x), \end{cases}$$

où  $\varepsilon > 0$  est bien entendu un petit paramètre qui représente une viscosité, et le modèle (3.1) est en fait une idéalisation de (3.13) dans la limite d'une viscosité tendant vers 0. On cherche donc une solution faible qui soit la limite de  $u_\varepsilon$  pour  $\varepsilon \rightarrow 0$ . On peut démontrer (voir [Des10] ou [GR21], il s'agit d'un résultat assez difficile) que le modèle parabolique admet bien une solution unique, et que cette solution converge (en un sens qu'il faudrait préciser) vers une solution faible de (3.1). Naturellement, cette caractérisation de la « bonne » solution faible de (3.1) n'est pas très maniable. On cherche donc une caractérisation intrinsèque.

Une *fonction d'entropie* est une fonction  $\eta : \mathbf{R} \rightarrow \mathbf{R}$  de classe  $C^2$  et convexe.

Le flux d'entropie associé est la fonction  $\xi$  vérifiant ( $f$  est le flux de la loi de conservation).

$$\xi(u) = \int \eta'(u) f'(u) du.$$

On montre également que la limite  $u$  des solutions  $u_\varepsilon$  vérifie l'inégalité d'entropie suivante, pour toute entropie  $(\eta, \xi)$  :

$$(3.14) \quad - \int_0^{+\infty} \int_{\mathbf{R}} (\eta(u) \partial_t \varphi + \xi(u) \partial_x \varphi) dx dt - \int_{\mathbf{R}} \eta(u_0(x)) \varphi(x, 0) dx \leq 0, \quad \forall \varphi \in C_0^1, \varphi \geq 0.$$

**Définition 3.2** (Solution faible entropique). Une solution faible entropique de la loi de conservation (3.1) est une solution faible qui vérifie l'inégalité d'entropie (3.14) pour tout couple d'entropies  $(\eta, \xi)$ .

On peut alors compléter la proposition 3.2 : parmi les solutions faibles qui vérifient la condition de saut (3.10), une solution est entropique si et seulement si elle vérifie l'inégalité

$$-x'(t)[[\eta(u)]] + [[\xi(u)]] \leq 0,$$

pour tout couple d'entropies  $(\eta, \xi)$ .

Le critère précédent reste encore difficile à appliquer en pratique. Pour déterminer si un choc est entropique, on dispose heureusement de conditions suffisantes qui ne font intervenir que la fonction de flux  $f$  et les états gauche et droit du choc. Nous résumons la discussion dans le résultat suivant

**Théorème 3.1** (Caractérisation des chocs entropiques). *On considère un choc défini par les états  $u_L$  et  $u_R$  et la quantité  $\sigma$  définie par la relation de Rankine-Hugoniot*

$$\sigma = \frac{f(u_R) - f(u_L)}{u_R - u_L}.$$

On suppose que la fonction de flux  $f$  est dérivable.

Le choc se propageant à la vitesse  $\sigma$  est entropique si l'inégalité suivante (appelée critère de Lax) est vérifiée

$$(3.15) \quad f'(u_L) \geq \sigma \geq f'(u_R)$$

Un cas particulier est celui des fonctions de flux convexes. Dans ce cas la dérivée est croissante, et le critère de Lax se réduit à

$$u_L \geq u_R.$$

On retrouve bien que la solution discontinue du paragraphe précédent n'est pas entropique.

Un critère plus général que le critère de Lax est le critère d'Oleinik : le choc est entropique si et seulement si

$$\frac{f(u) - f(u_L)}{u - u_L} \geq \sigma \geq \frac{f(u) - f(u_R)}{u - u_R},$$

pour tout  $u$  entre  $u_L$  et  $u_R$  (on ne dit rien sur l'ordre entre  $u_L$  et  $u_R$ ). Ce critère donne la même condition que précédemment dans le cas convexe.

### 3.2.4 Solution du problème de Riemann

Dans ce paragraphe nous allons construire la solution entropique du problème de Riemann, qui correspond à une condition initiale discontinue (on supposera  $u_L \neq u_R$ ) :

$$(3.16) \quad u_0(x) = \begin{cases} u_L & x < 0 \\ u_R & 0 < x. \end{cases}$$

Sans perte de généralité, nous avons arbitrairement fixé la discontinuité initiale en 0, puisque le problème est invariant par translation.

En plus de son intérêt propre, la solution du problème de Riemann sert à construire le schéma de Godunov, que nous étudierons au paragraphe 3.3.

La situation est particulièrement simple dans le cas convexe.

**Proposition 3.3** (Solution du problème de Riemann, cas convexe). *On suppose que  $f''(u) > 0$ . La solution du problème de Riemann est donnée par :*

**Cas**  $u_L < u_R$  Une détente :

$$(3.17) \quad u(x, t) = \begin{cases} u_L, & x \leq tf'(u_L) \\ (f')^{-1}\left(\frac{x}{t}\right), & tf'(u_L) \leq x \leq tf'(u_R) \\ u_R, & tf'(u_R) \leq x, \end{cases}$$

en remarquant que  $f'$ , qui est strictement croissante, est inversible sur  $f'(\mathbf{R})$ .

**Cas**  $u_L > u_R$  Un choc (entropique) à vitesse

$$\sigma = \frac{f(u_R) - f(u_L)}{u_R - u_L}.$$

*Démonstration.* On cherche une solution auto-similaire, c'est-à-dire de la forme

$$u(x, t) = v\left(\frac{x}{t}\right),$$

où  $v$  est une fonction à déterminer. Calculons les dérivées partielles :

$$\begin{aligned} - \partial_x u(x, t) &= \frac{1}{t} v' \left( \frac{x}{t} \right); \\ - \partial_t u(x, t) &= -\frac{x}{t^2} v' \left( \frac{x}{t} \right). \end{aligned}$$

Et par conséquent, par dérivation composée

$$\partial_t u + \partial_x f(u) = - \left( \frac{x}{t^2} + f'(v) \frac{1}{t} \right) v' \left( \frac{x}{t} \right).$$

En notant  $\xi = \frac{x}{t}$ , on voit que  $v$  doit vérifier l'équation

$$(\xi - f'(v(\xi)))v'(\xi) = 0.$$

Si  $\frac{x}{t} < f'(u_L)$  ou  $\frac{x}{t} > f'(u_R)$ , la solution est constante. Sinon,  $\xi$  vérifie  $f'(v(\xi)) = \xi$ . ■

Le cas général fait appel à l'enveloppe convexe ou concave de  $f$ .

On remarque que la solution du problème de Riemann peut s'écrire dans tous les cas sous la forme d'une fonction auto-similaire :

$$(3.18) \quad u(x, t) = w_R(x/t, ; u_L, u_R).$$

La valeur en  $x = 0$  de la solution du problème de Riemann sera utile pour définir le schéma de Godunov. Le résultat suivant est établi d'abord dans le cas convexe, mais il est vrai en général.

Dans le cas convexe, l'examen de la proposition 3.3 montre que la valeur en  $x = 0$  de la solution du problème de Riemann est donnée par :

$$(3.19) \quad u(0, t) = \begin{cases} u_L & \text{si } u_L > u_R \text{ et } \sigma > 0 \\ u_R & \text{si } u_L > u_R \text{ et } \sigma < 0 \\ u_L & \text{si } u_L \leq u_R \text{ et } f'(u_L) \geq 0 \\ u_R & \text{si } u_L \leq u_R \text{ et } f'(u_R) \geq 0 \\ (f')^{-1}(0) & \text{si } u_L \leq u_R \text{ et } f'(u_L) < 0 < f'(u_R). \end{cases}$$

Le dernier cas est une détente transsonique, par analogie avec la dynamique des gaz. C'est le cas où le choc n'est pas entropique.

Nous aurons besoin au paragraphe suivant de la valeur de  $f(u(0, t))$ , où  $u$  est la solution du problème de Riemann. La réponse est donnée par le résultat suivant :

**Proposition 3.4** (Flux de la solution en  $x = 0$  du problème de Riemann). *Pour un flux quelconque, la valeur de  $f(u(0, t))$ , où  $u$  est la solution du problème de Riemann avec données  $u_L, u_R$ , est*

$$(3.20) \quad f(u(0, t)) = \begin{cases} \min_{u_L \leq \xi \leq u_R} f(\xi) & \text{si } u_L < u_R \\ \max_{u_R \leq \xi \leq u_L} f(\xi) & \text{si } u_R < u_L \end{cases}$$

La preuve dans le cas convexe consiste à examiner tous les cas. L'extension au cas général vient d'un article de S. Osher [Osh84]. Il est curieux de noter que ce résultat est cité dans plusieurs des références citées dans l'introduction de ce chapitre, mais que sa démonstration n'est jamais donnée.

### 3.3 Le schéma de Godunov

#### 3.3.1 Méthodes de volumes finis pour les lois de conservation

On se donne un maillage de  $\mathbf{R}$ , de pas constant  $\Delta x$  (pour simplifier), et on se donne également un pas de temps  $\Delta t$  (voir Figure 3.5), avec  $N\Delta t = t_f$ .

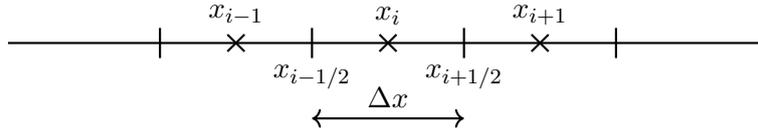


FIGURE 3.5 – Maillage volumes finis en 1D

On note  $x_i = i\Delta x, i \in \mathbf{Z}, t^n = n\Delta t, n = 0, \dots, N$ . Comme pour le cas des problèmes elliptiques au chapitre 2, nous construirons le schéma en intégrant l'équation sur un volume de contrôle, qui sera l'intervalle  $]x_{i-1/2}, x_{i+1/2}[$ . En ce qui concerne la discrétisation en temps, nous utiliserons un schéma explicite d'ordre 1. Les inconnues discrètes sont des approximations des moyennes de la solution sur chaque volume de contrôle :

$$u_i^n = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t^n) dx.$$

En intégrant (3.1) sur  $]x_{i-1/2}, x_{i+1/2}[$  nous obtenons

$$\Delta x \frac{u_i^{n+1} - u_i^n}{\Delta t} + f(u(x_{i+1/2}, t^n)) - f(u(x_{i-1/2}, t^n)) = 0, \quad i \in \mathbf{Z}, t = 0, \dots, N-1,$$

mais comme  $u(x_{i\pm 1/2}, t^n)$  n'est pas connu (notre approximation est discontinue au passage des volumes de contrôle), nous devons approcher le flux  $f(u(x_{i-1/2}, t^n))$ .

Pour cela nous considérons un flux approché  $F_{i+1/2}$ , utilisant les valeurs de la fonction inconnue sur les deux volumes de contrôle entourant l'interface  $x_{i+1/2}$  :

$$(3.21) \quad F_{i+1/2} = F(u_i^n, u_{i+1}^n),$$

où  $F$  est une *fonction de flux numérique* qui définit le schéma. En divisant par  $\Delta x$ , l'approximation est alors définie récursivement par

$$(3.22) \quad \frac{u_i^{n+1} - u_i^n}{\Delta t} + \frac{F_{i+1/2} - F_{i-1/2}}{\Delta x} = 0, \quad i \in \mathbf{Z}, t = 0, \dots, N-1,$$

les valeurs initiales étant définies par

$$(3.23) \quad u_i^0 = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} u_0(x) dx.$$

Quelque soit le choix de la fonction  $F$ , le schéma est *conservatif* par construction. En faisant l'hypothèse que la condition initiale est à support compact ( $u_0(x) = 0$  si  $|x| > A$  pour un certain  $A$ , cette propriété reste vraie à chaque pas de temps (le support peut augmenter avec le temps, mais reste compact), et

$$\sum_{i \in \mathbf{Z}} u_i^n = \sum_{i \in \mathbf{Z}} u_i^{n+1}, \quad \text{pour tout } n.$$

Pour obtenir un schéma *consistant*, on impose la condition

$$(3.24) \quad F(u, u) = f(u).$$

*Remarque 3.2* (De l'importance des schémas conservatifs). Il est dangereux d'utiliser un schéma qui n'est pas sous forme conservative. Prenons l'exemple de l'équation de Burgers, avec la condition initiale

$$u_0(x) = \begin{cases} 1 & x < 0, \\ 0 & x > 0. \end{cases}$$

L'unique solution faible entropique est un choc se déplaçant à vitesse  $1/2$ .

En écrivant l'équation de Burgers sous la forme non-conservative

$$\partial_t u + u \partial_x u = 0,$$

il peut sembler naturel de considérer le schéma explicite suivant :

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + u_i^n \frac{u_i^n - u_{i-1}^n}{\Delta x} = 0,$$

où on a introduit un décentrage pour obtenir un schéma stable (il faudrait le démontrer). La discrétisation (naturelle) de la condition initiale

$$u_i^n = \begin{cases} 1 & i < 0 \\ 0 & i \geq 0, \end{cases}$$

conduit à la solution

$$u_i^n = u_i^0, \quad i \in \mathbf{Z}, n > 0,$$

soit un choc stationnaire, qui ne peut pas être une approximation convergente de la solution exacte.

*Exemple 4* (Schéma de Lax-Friedrichs). Il s'agit d'une généralisation au cas non-linéaire . La fonction de flux est

$$F(u, v) = \frac{1}{2} (f(u) + f(v)) - \frac{\Delta x}{2\Delta t} (v - u),$$

ce qui correspond au schéma numérique :

$$(3.25) \quad \frac{u_i^{n+1} - \frac{u_{i-1}^n + u_{i+1}^n}{2}}{\Delta t} + \frac{f(u_{i+1}^n) - f(u_{i-1}^n)}{2\Delta x} = .$$

Ce schéma est consistant, et stable.

Revenons sur le cas de l'advection linéaire, avec  $f(u) = au$  (prenons  $a > 0$  pour fixer les idées). Les schémas vus au Chapitre B rentrent dans le cadre général de ce paragraphe :

**Centré** On prend  $F(u, v) = \frac{a}{2}(u + v)$ . Rappelons que ce schéma est *instable* ;

**Décentré amont** Ici  $F(u, v) = au$  (avec  $a > 0$ ).

### 3.3.2 Le schéma de Godunov

Il est difficile de généraliser le schéma décentré amont aux problèmes non-linéaires (voir le contre-exemple du paragraphe précédent) tout en restant dans le cadre des schémas conservatifs. La solution a été proposée par S. Godunov à la fin des années 1950, qui a qui a proposé d'utiliser la solution des problèmes de Riemann locaux pour construire une solution approchée entropique.

Pour définir la méthode, il sera commode d'introduire la fonction constante par morceaux  $u_h^n$  définie par

$$u_h^n(x) = u_i^n \quad \text{pour } x \text{ in } ]x_{i-1/2}, x_{i+1/2}[.$$

L'idée du schéma de Godunov consiste

- 1) Résoudre les problèmes de Riemann locaux

$$(3.26) \quad \begin{cases} \partial_t u + \partial_x (f(u)) = 0 & x \in \mathbf{R}, t \in [t^n, t^{n+1}[ \\ u(x, 0) = u_h^n(x). \end{cases}$$

On note  $\tilde{u}_h^{n+1}(x) = u_h^n(x, t^{n+1})$ .

- 2) Définir une nouvelle solution approchée comme la moyenne de la fonction  $\tilde{u}_h^{n+1}$  sur la maille  $]x_{i-1/2}, x_{i+1/2}[$ .

Au passage, on montre que le schéma est bien sous forme conservative, avec un flux numérique défini par

$$(3.27) \quad F_{i+1/2}^n = \frac{1}{\Delta} \int_{t^n}^{t^{n+1}} f(u_h^{n+1}(x_{i+1/2}, t)) dt = f(\tilde{u}_h^{n+1}(x_{i+1/2})) \cdot$$

Nous allons détailler chacune des étapes, pour montrer que le schéma de Godunov rentre bien dans le cadre introduit au paragraphe précédent, et donne l'expression du flux numérique.

#### Solution des problèmes de Riemann locaux

On résout un problème de Riemann local autour de chaque interface  $x = x_{i+1/2}$ , avec une donnée  $u_L = u_i^n, u_R = u_{i+1}^n$ . Il faut donc s'assurer que les solutions des différents problèmes n'interfèrent pas sur la durée du pas de temps. Ce sera le cas si le pas de temps est restreint par

$$(3.28) \quad \frac{\Delta t}{\Delta x} \max_i |f'(u_i^n)| \leq \frac{1}{2}.$$

Sous cette hypothèse, la solution  $u_h^n$  peut s'écrire comme une superposition de problèmes de Riemann indépendants, de sorte que

$$\tilde{u}_h^{n+1}(x) = w_R \left( \frac{x - x_{i+1/2}}{\Delta t}; u_i, u_{i+1} \right), \quad x_i < x < x_{i+1}, i \in \mathbf{Z}.$$

où  $w_R$  a été définie en (3.18).

## Calcul du flux numérique

Pour trouver une expression du flux numérique, on intègre l'équation (3.26) sur le domaine  $[x_{i-1/2}, x_{i+1/2}[\times]t^n, t^{n+1}[$ , pour obtenir

$$(3.29) \quad \int_{x_{i-1/2}}^{x_{i+1/2}} \tilde{u}_h^{n+1}(x) dx - \int_{x_{i-1/2}}^{x_{i+1/2}} \tilde{u}_h^n(x) dx \\ + \int_{t^n}^{t^{n+1}} f(u_h^n(x_{i+1/2}^-, t)) dt - \int_{t^n}^{t^{n+1}} f(u_h^n(x_{i-1/2}^+, t)) dt = 0$$

Le premier et le second terme vont simplement constituer la définition de  $u_i^n$  et  $u_i^{n+1}$  (au facteur  $\Delta x$  près).

Considérons le troisième terme (l'analyse pour le dernier est similaire).

$$\int_{t^n}^{t^{n+1}} f(u_h^n(x_{i+1/2}^-, t)) dt = \Delta t f(\tilde{u}_h^{n+1}(x_{i+1/2}^-)) = \Delta t f(w_R(0^-, u_i, u_{i+1})),$$

puisque la solution à l'origine du problème de Riemann est constante en temps d'après (3.19). Pour retrouver l'expression annoncée en (3.27), il faut encore remarquer que la fonction :  $\xi \rightarrow f(w_R(\xi; u_L, u_R))$  est continue en 0 (alors que  $w_R(\xi; u_L, u_R)$  ne l'est pas forcément). Si la fonction  $\xi \rightarrow w_R(\xi; u_L, u_R)$  est continue en 0, le résultat est vrai par composition. Et si cette fonction est discontinue en 0, on a un choc stationnaire, et la condition de Rankine-Hugoniot montre qu'alors  $f(w_R(0+; u_L, u_R)) = f(w_R(0-; u_L, u_R))$ . On peut donc noter sans ambiguïté  $f(w_R(0, u_i, u_{i+1}))$  cette valeur.

Finalement, (3.29) se réécrit :

$$(3.30) \quad \Delta x (u_i^{n+1} - u_i^n) = \Delta t \left( f(w_R(0, u_i^n, u_{i+1}^n)) - f(w_R(0, u_{i-1}^n, u_i^n)) \right).$$

Comme nous l'avons vu au paragraphe 3.2.4, cette expression ne dépend que des valeurs de  $u_i^n$  et  $u_{i+1}^n$ . De plus, nous avons obtenu une expression explicite, qui donne finalement une définition exploitable du flux numérique :

$$(3.31) \quad F(u_i, u_{i+1}) = \begin{cases} \min_{u_i \leq \xi \leq u_{i+1}} f(\xi) & \text{si } u_i < u_{i+1} \\ \max_{u_{i+1} \leq \xi \leq u_i} f(\xi) & \text{si } u_{i+1} < u_i \end{cases}$$

Ce flux est bien consistant avec l'équation : si  $u_i = u_{i+1}$ , la solution du problème de Riemann est une constante.

Le schéma de Godunov est conservatif, et on démontre qu'il converge vers la solution faible entropique du problème (3.1) (c'est une conséquence de ce qu'on utilise la solution entropique du problème de Riemann).

La principale limitation du schéma est qu'il est d'ordre 1, et donc induit une diffusion numérique importante. Un théorème de Godunov indique qu'un schéma numérique linéaire et monotone (nous n'avons pas défini ce terme, il veut dire que le schéma n'a pas d'oscillations) est nécessairement d'ordre 1.

Un schéma plus précis ne sera formellement d'ordre 2 que dans les régions où la solution est régulière, donc « loin » des chocs. Et un tel schéma présentera nécessairement des oscillations.

On entre dans le vaste sujet des méthodes de « haute résolution », qui tentent de conserver une meilleure précision quand la solution est régulière, tout en contrôlant la taille des oscillations. Pour cela il faut faire appel à des limiteurs de pente, ou des limiteurs de flux. La lectrice, ou même le lecteur, qui voudrait en savoir plus est invitée à consulter les références citées dans l'introduction.

### **3.4 Volumes finis en dimension 2**

## Annexe A

# Analyse des schémas par différences finies

Références pour ce chapitre : [Ise09 ; LeV07 ; LeV02 ; GDS15 ; LL16]

### A.1 Différences finies

#### Différences finies (1)

Approcher  $c'(x), c''(x), \dots$  pour une fonction  $u$  *inconnue*.

Fonction évaluée sur une grille  $x_j = jh, j \in \mathbf{Z}$

Formule de Taylor

$$c(x \pm h) = c(x) \pm hc'(x) + \frac{h^2}{2}c''(x) + O(h^3)$$

#### Approximations de $c'$

Différence avant  $c'(x) = \frac{c(x+h) - c(x)}{h} + O(h)$     Ordre 1

Différence arrière  $c'(x) = \frac{c(x) - c(x-h)}{h} + O(h)$     Ordre 1

Différence centrée  $c'(x) = \frac{c(x+h) - c(x-h)}{2h} + O(h^2)$     Ordre 2

#### Différences finies (2)

#### Approximations de $c''$

Différence centrée  $c''(x) = \frac{c(x+h) + 2c(x) - c(x-h)}{h^2} + O(h^2)$     Ordre 2

Différence arrière  $c''(x) = \frac{-c(x-3h) + 4c(x-2h) - 5c(x-h) + 2c(x)}{h^2}$     Ordre 2

#### Généralisation à plusieurs variables

$c(x, y) \in C^2(\mathbf{R}^2)$ , points  $(x_i, y_j) = (i\Delta x, j\Delta y)$ ,  $(i, j) \in \mathbf{Z}^2$

$$\Delta c(x, y) = \frac{c(x - \Delta x, y) - 2c(x, y) + c(x + \Delta x, y)}{\Delta x^2} + \frac{c(x, y - \Delta y) - 2c(x, y) + c(x, y + \Delta y)}{\Delta y^2} + O(\Delta x^2 + \Delta y^2)$$

## A.2 Équation d'advection

L'équation d'advection (1)

Équation en dimension 1

$$\begin{aligned} \frac{\partial c}{\partial t} + v \frac{\partial c}{\partial x} &= 0, \quad x \in \mathbf{R}, 0 < t \\ c(x, 0) &= c_0(x), \quad 0 < x \end{aligned}$$

Solution exacte (caractéristiques)

$$c(x, t) = c_0(x - vt), \quad x \in \mathbf{R}$$

La solution est *constante* le long des caractéristiques  $x - vt = \text{cte.}$  voir Figure A.1

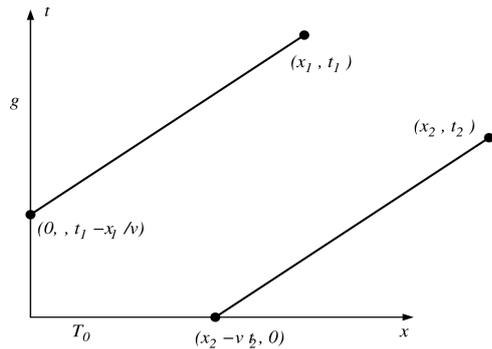


FIGURE A.1 – Caractéristiques, vitesse constante

L'équation d'advection (2)

Équation en dimension 1

$$\begin{aligned} \frac{\partial c}{\partial t} + v \frac{\partial c}{\partial x} &= 0, \quad 0 < x, 0 < t \\ c(0, t) &= g(t) \quad 0 < t \\ c(x, 0) &= c_0(x), \quad 0 < x \end{aligned}$$

Solution exacte (caractéristiques) voir Figure A.2

$$c(x, t) = \begin{cases} g(t - x/v), & x < vt \\ c_0(x - vt), & vt < x \end{cases}$$

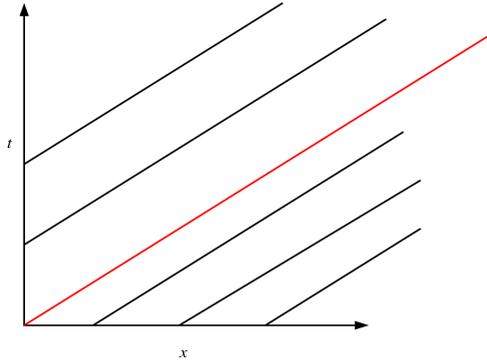


FIGURE A.2 – Caractéristiques, vitesse constante

### Schémas aux différences

Grille  $(x_j, t^n) = (j\Delta x, n\Delta t)$ ,  $j \in \mathbf{Z}, t \in \mathbf{N}$ ,  $c_j^n \approx c(x_j, t^n)$ .  $v > 0$ .

### Quelques schémas

**Centré en espace**  $\frac{c_j^{n+1} - c_j^n}{\Delta t} + v \frac{c_{j+1}^n - c_{j-1}^n}{2\Delta x} = 0,$

**Décentré amont**  $\frac{c_j^{n+1} - c_j^n}{\Delta t} + v \frac{c_j^n - c_{j-1}^n}{\Delta x} = 0,$

**Lax–Friedrichs**  $\frac{c_j^{n+1} - \frac{c_{j+1}^n + c_{j-1}^n}{2}}{\Delta t} + v \frac{c_{j+1}^n - c_{j-1}^n}{2\Delta x} = 0,$

Schémas *explicites* :  $c_j^{n+1}$  se calcule en fonctions de quantités *connues*.

### Schéma centré

$$c_j^{n+1} = c_j^n + \frac{v\Delta t}{2\Delta x} (c_{j+1}^n - c_{j-1}^n)$$

### Précision : erreur de troncature

$$\begin{aligned} \frac{c(x_j, t^{n+1}) - c(x_j, t^n)}{\Delta t} &= \frac{\partial c}{\partial t}(x_j, t^n) + O(\Delta t) \\ \frac{c(x_{j+1}, t^n) - c(x_{j-1}, t^n)}{2\Delta x} &= \frac{\partial c}{\partial x}(x_j, t^n) + O(\Delta x^2) \end{aligned}$$

Si  $c$  est solution de l'EDP,

$$\frac{c(x_j, t^{n+1}) - c(x_j, t^n)}{\Delta t} + v \frac{c(x_{j+1}, t^n) - c(x_{j-1}, t^n)}{2\Delta x} = O(\Delta x^2 + \Delta t)$$

*Schéma d'ordre 2 en espace, 1 en temps*

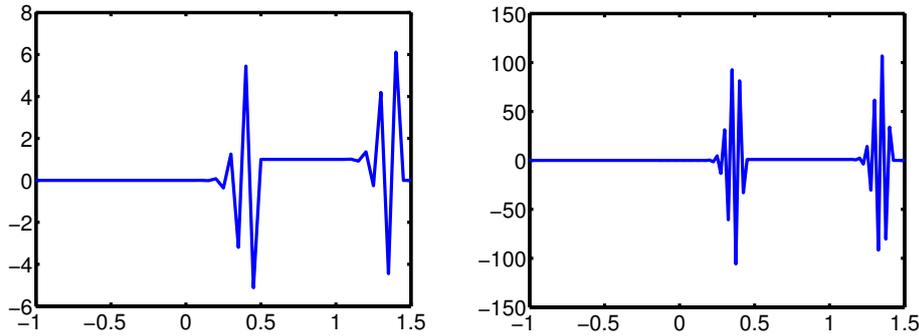


FIGURE A.3 – Instabilité du schéma centré

### Schéma centré : exemple

Calcul avec  $\frac{v\Delta t}{\Delta x} = 1.2$ . À gauche  $h = 0.05$ , à droite  $h = 0.025$ .

*Instable*

### Le schéma décentré amont

$$c_j^{n+1} = (1 - \alpha)c_j^n + \alpha c_{j-1}^n$$

avec  $\alpha = \frac{v\Delta t}{\Delta x}$  le nombre de Courant–Friedrichs–Lewy (CFL).

### Précision : erreur de troncature

$$\frac{c(x_j, t^{n+1}) - c(x_j, t^n)}{\Delta t} = \frac{\partial c}{\partial t}(x_j, t^n) + O(\Delta t)$$

$$\frac{c(x_j, t^n) - c(x_{j-1}, t^n)}{\Delta x} = \frac{\partial c}{\partial x}(x_j, t^n) + O(\Delta x)$$

Si  $c$  est solution de l'EDP,

$$\frac{c(x_j, t^{n+1}) - c(x_j, t^n)}{\Delta t} + v \frac{c(x_j, t^n) - c(x_{j-1}, t^n)}{\Delta x} = O(\Delta x + \Delta t)$$

*Schéma d'ordre 1*

### Stabilité

$\alpha \leq 1$   $c_j^{n+1}$  combinaison *convexe* de  $c_j^n$  et  $c_{j-1}^n$ . Donc  $|c_j^{n+1}| \leq \max_i |c_i^n|$ . [2ex]

$\alpha > 1$  Plus de combinaison convexe.

$$c^n = \begin{cases} 1, & j = 0, \dots, i-1 \\ 0, & j = i, i+1, \dots \end{cases} \Rightarrow c^{n+1} = \begin{cases} 1, & j = 0, \dots, i-1 \\ \alpha, & j = i \\ 0, & j = i+1, i+2, \dots \end{cases}$$

### Stabilité du schéma décentré amont

Schéma stable sous la condition CFL  $\frac{v\Delta t}{\Delta x} < 1$ .

### Diffusion numérique

Terme d'erreur principal du schéma amont :

$$\frac{c_j^{n+1} - c_j^n}{\Delta t} + v \frac{c_j^n - c_{j-1}^n}{\Delta x} = -\frac{v\Delta x}{2} \frac{\partial^2 c}{\partial x^2} + \frac{\Delta t}{2} \frac{\partial^2 c}{\partial t^2} + O(\Delta x^2 + \Delta t^2)$$

mais (pour la solution exacte)  $\frac{\partial c}{\partial t} = -v \frac{\partial c}{\partial x}$ , donc  $\frac{\partial^2 c}{\partial t^2} = v^2 \frac{\partial^2 c}{\partial x^2}$

Finalement, le schéma approche à l'ordre 2

$$\frac{\partial c}{\partial t} + v \frac{\partial c}{\partial x} - \frac{v\Delta x}{2} (1 - \alpha) \frac{\partial^2 c}{\partial x^2} = 0.$$

Équation modifiée : *dépend du maillage*

Si  $0 \leq \alpha \leq 1$  (coeff positif) : diffusion *numérique* (provient du schéma).

### Lax-Wendroff

Taylor :  $c(x_j, t^{n+1}) = c(x_j, t^n) + \Delta t \frac{\partial c}{\partial t}(x_j, t^n) + \frac{\Delta t^2}{2} \frac{\partial^2 c}{\partial t^2}(x_j, t^n)$

Équation :  $\frac{\partial^2 c}{\partial t^2} = v^2 \frac{\partial^2 c}{\partial x^2}$

### Schéma de Lax-Wendroff

$$\frac{c_j^{n+1} - c_j^n}{\Delta t} + v \frac{c_{j+1}^n - c_{j-1}^n}{2\Delta x} - \left(\frac{v\Delta t}{\Delta x}\right)^2 \frac{c_{j+1}^n - 2c_j^n + c_{j-1}^n}{\Delta x^2} = 0$$

Ordre 2 en espace

Comme schéma centré (instable), mais *stabilisé* par la diffusion numérique.

### Stabilité et condition CFL

Schéma borné, indépendamment de  $(\delta x, \Delta t)$

$$\sum_{j \in \mathbf{Z}} |c_j^n|^2 \leq C_T \sum_{j \in \mathbf{Z}} |c_{0j}|^2, \quad \text{pour } n\Delta t \leq T$$

Pas d'amplification des erreurs aux pas de temps précédents.

Condition *nécessaire de stabilité* :  $\mathcal{D}(x_j, t^n) \subset$  domaine de dépendance numérique.

*Condition CFL* (Courant, Friedrichs, Lewy) :  $\left| \frac{u\Delta t}{\Delta x} \right| \leq 1$  voir Figure A.4

Outils techniques : analyse de Fourier, von Neumann

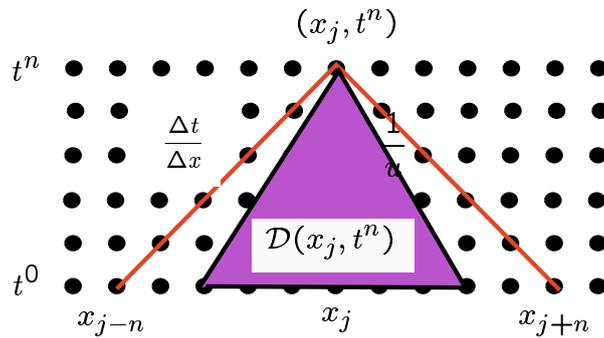
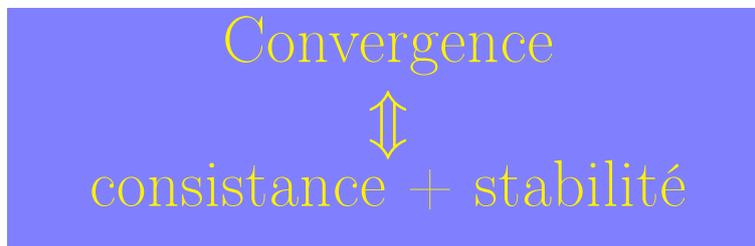


FIGURE A.4 – Cones de dépendance discret et continu

### Théorème de Lax–Richtmyer



- Schéma *instable*  $\implies$  ne peut pas être convergent [2ex]
- Montrer la convergence : Consistance (facile, Taylor) et stabilité (Fourier, von Neumann, énergie)

### Exemple : solution régulière

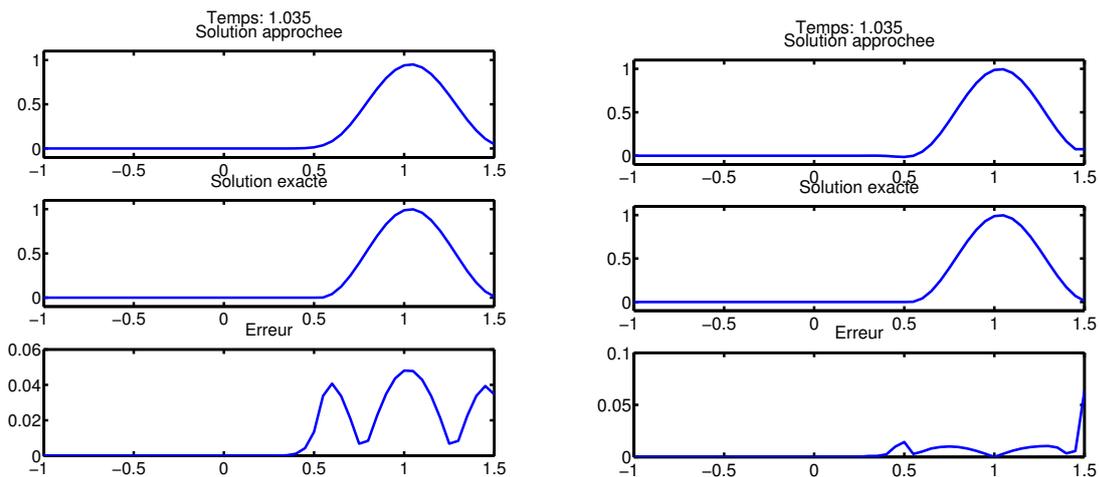


FIGURE A.5 – Comparaison des schémas décentré amont (gauche) et Lax–Wendroff (droite)

### Exemple : solution discontinue

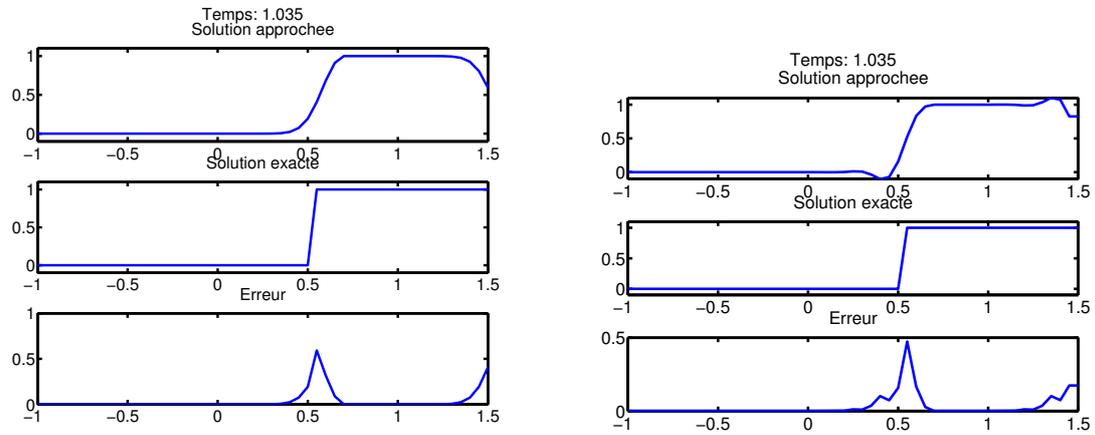


FIGURE A.6 – Comparaison des schémas décentré amont (gauche) et Lax–Wendroff (droite)

### Comparaison de schémas

Voir Figure A.7

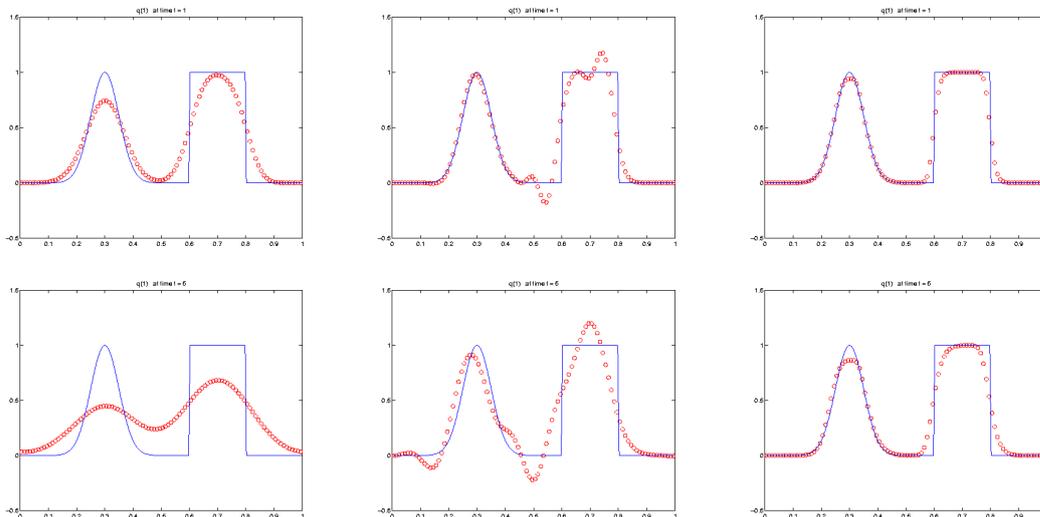


FIGURE A.7 – Comparaison des schémas : Décentré (gauche), Lax–Wendroff (milieu), Lw + limiteur (droite), figure du haut  $t = 1$ , figure du bas  $t = 5$

### Cas général : vitesse variable

Loi de conservation : forme intégrée sur  $[x_{j-1/2}, x_{j+1/2}] \times [t^n, t^{n+1}]$

$$\frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} c(x, t^{n+1}) dx - \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} c(x, t^n) dx = - \frac{1}{\Delta x} \left[ \int_{t^n}^{t^{n+1}} v(x_{j-1/2}) c(x_{j-1/2}, t) dt - \int_{t^n}^{t^{n+1}} v(x_{j+1/2}) c(x_{j+1/2}, t) dt \right]$$

Inconnue  $c_j^n \approx \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} c(x, t^n) dx$

Flux numérique  $F_{j-1/2}^n(c_{j-1}^n, c_j^n) \approx \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} v(x_{j-1/2}) c(x_{j-1/2}, t) dt$

Discrétisation conservative :  $\frac{c_j^{n+1} - c_j^n}{\Delta t} + \frac{F_{j+1/2}^n - F_{j-1/2}^n}{\Delta x} = 0$

### Exemples

Notation :  $v_{j+1/2} = v(x_{j+1/2})$

Schéma décentré amont

$$F_{j+1/2}^n = \begin{cases} v_{j+1/2} c_j^n & \text{si } v_{j+1/2} \leq 0 \\ v_{j+1/2} c_{j+1}^n & \text{si } v_{j+1/2} \geq 0 \end{cases}$$

Schéma de Lax – Wendroff deux étapes :

$$\frac{c_{j+1/2}^{n+1/2} - \frac{c_{j+1}^n + c_{j-1}^n}{2}}{\Delta t/2} + \frac{v_{j+1} c_{j+1} - v_j c_j}{\Delta x} = 0$$

$$\frac{c_j^{n+1} - c_j^n}{\Delta t} + \frac{v_{j+1/2} c_{j+1/2}^{n+1/2} - v_{j-1/2} c_{j-1/2}^{n+1/2}}{\Delta x} = 0$$

$c^n$  linéaire sur  $[x_{j-1}, x_j]$

Supprimer les oscillations : *limiteur de pente* (Godunov, van Leer)

## A.3 Équation de la chaleur

L'équation de la chaleur

Équation en dimension 1

$$\frac{\partial c}{\partial t} - D \frac{\partial^2 c}{\partial x^2} = 0, \quad 0 < x < L, \quad 0 < t$$

$$c(0, t) = 0, \quad c(L, t) = 0, \quad t > 0$$

$$c(x, 0) = c_0(x), \quad 0 < x < L.$$

Solution sur  $\mathbf{R}$

$$c(x, t) = \int_{\mathbf{R}} \frac{1}{2\sqrt{\pi t}} c_0(x - y) e^{-y^2/(4t)} dy$$

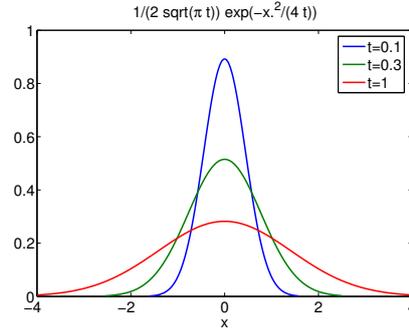


FIGURE A.8 – solution élémentaire de l'équation de la chaleur

### Un schéma explicite

Grilles en espace :  $x_j = j\Delta x, j = 0, \dots, J, J\Delta x = L$ , en temps  $t^n = n\Delta t, n = 0, \dots, N, N\Delta t = T$ .

$$\frac{c_j^{n+1} - c_j^n}{\Delta t} - D \frac{c_{j+1}^n - 2c_j^n + c_{j-1}^n}{\Delta x^2} = 0, \quad j = 1, \dots, J-1$$

$$c_0^{n+1} = 0, \quad c_J^{n+1} = 0$$

$$c_j^0 = c_0(j\Delta x)$$

- Schéma explicite,
- Ordre 1 en temps, 2 en espace
- $F = \frac{D\Delta t}{\Delta x^2}$  : nombre de Fourier (sans dimension)
- Stable si  $F \leq \frac{1}{2}$ , très restrictif.

### Un schéma implicite

$$\frac{c_j^{n+1} - c_j^n}{\Delta t} - D \frac{c_{j+1}^{n+1} - 2c_j^{n+1} + c_{j-1}^{n+1}}{\Delta x^2} = 0, \quad j = 1, \dots, J-1$$

$$c_0^{n+1} = 0, \quad c_J^{n+1} = 0$$

$$c_j^0 = c_0(j\Delta x)$$

- Schéma implicite : Système linéaire à chaque pas de temps,  $Ac^{n+1} = c^n$
- Ordre 1 en temps, 2 en espace
- Stable pour *tout*  $\Delta t$

### Système linéaire

$$A = \begin{bmatrix} 1+F & -F/2 & & & \\ -F/2 & 1+F & -F/2 & & \\ & \ddots & \ddots & \ddots & \\ & & -F/2 & 1+F & -F/2 \\ & & & -F/2 & 1+F \end{bmatrix} = I + F \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}$$

### Propriétés de $A$

- Tridiagonale
- Symétrique
- Définie positive

### $\theta$ -schéma

$$\begin{aligned} \frac{c_j^{n+1} - c_j^n}{\Delta t} - (1 - \theta)D \left( \frac{c_{j+1}^n - 2c_j^n + c_{j-1}^n}{\Delta x^2} \right) \\ - \theta D \left( \frac{c_{j+1}^{n+1} - 2c_j^{n+1} + c_{j-1}^{n+1}}{\Delta x^2} \right) = 0, \quad j = 1, \dots, J-1 \\ c_0^{n+1} = 0, \quad c_J^{n+1} = 0 \\ c_j^0 = c_0(j\Delta x) \end{aligned}$$

- En temps : ordre 1 pour  $\theta \neq 1/2$ , ordre 2 pour  $\theta = 1/2$ .
- En espace : ordre 2
- Stabilité

$\theta \geq 1/2$  Stable pour *tout*  $\Delta t$ ,

$\theta < 1/2$  Stable sous condition  $\Delta t \leq \frac{\Delta x^2}{2(1 - \theta)D}$

## Annexe B

# Équations d'advection–diffusion

### B.1 Équation d'advection–diffusion

#### Équation d'advection–diffusion

##### Modèle en dimension 1

$$(B.1) \quad \begin{cases} \frac{\partial u}{\partial t} - \frac{\partial}{\partial x} \left( D \frac{\partial u}{\partial x} - au \right) = f(x, t), & (x, t) \in [0, L] \times [0, T_f], \\ u(x, 0) = u_0(x) & \text{à } t = 0, \\ u(0, t) = g_G(t) & \text{en } x = 0, \\ D \frac{\partial u}{\partial x}(L, t) = 0 & \text{en } x = L \end{cases}$$

#### Le nombre de Péclet

Compare la « force » de la diffusion et de l'advection. Longueur  $L$

— Temps d'advection :  $t_a = L/a$  ;

— Temps de diffusion :  $t_d = L^2/D$

$$Pe = \frac{t_a}{t_d} = \frac{aL}{D}$$

#### Équation sans dimension

$x \rightarrow x/L, t \rightarrow t/t_a$  (choix du temps).

$$\frac{\partial u}{\partial t} - \frac{1}{Pe} \frac{\partial^2 u}{\partial x^2} + \frac{\partial u}{\partial x} = f(Lx, t_a t), \quad (x, t) \in [0, 1] \times [0, 1],$$

Nombre de Péclet de maille : longueur = taille de la maille,  $Pe = \frac{a\Delta x}{D}$ .

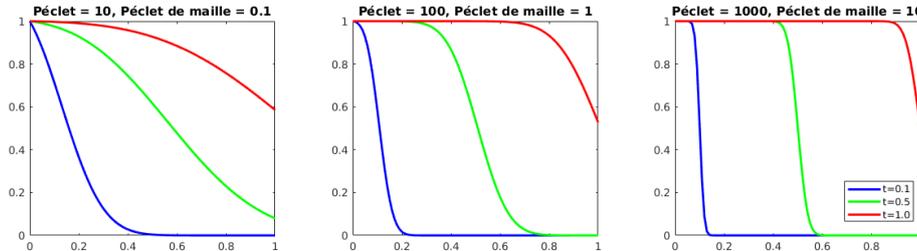
#### *Heuristique*

Résoudre le transport  $\Rightarrow Pe < 2$

Contrainte sur le pas d'espace

## Solution analytique (sur $\mathbf{R}^+$ )

$$u(x, t) = \frac{gG}{2} \exp\left(\frac{ax}{2D}\right) \left\{ \exp\left(-\frac{ax}{2D}\right) \operatorname{erfc}\left(\frac{x-at}{2\sqrt{Dt}}\right) + \exp\left(\frac{ax}{2D}\right) \operatorname{erfc}\left(\frac{x+at}{2\sqrt{Dt}}\right) \right\},$$



## Advection–diffusion : schémas numériques

### Propriétés d'un schéma

- Stabilité
- Précision
- Monotonie
- Coût

### Les choix

- Explicite : contrainte de stabilité forte ;
- Implicite : système linéaire à résoudre. [1.2ex]
- Premier ordre (décentré) : diffusion numérique
- Second ordre (centré) : oscillations

### Choix des paramètres de discrétisation

Si le nombre de Péclet est grand :

- Pas d'oscillation  $Pe = a\Delta x/D \leq 2 \Rightarrow \Delta x$  petit ;
- Précision CFL  $= a\Delta t/\Delta x \leq 1 \Rightarrow \Delta t$  petit ;

### Un exemple de schéma numérique

$$\frac{c_j^{n+1} - c_j^n}{Dt} - D \frac{c_{j+1}^{n+\theta} - 2c_j^{n+\theta} + c_{j-1}^{n+\theta}}{\mathcal{D}x^2} + a \frac{c_{j+1}^{n+\theta} - c_{j-1}^{n+\theta}}{2\mathcal{D}x} - \frac{\gamma \mathcal{D}x}{2} |a| \frac{c_{j+1}^{n+\theta} - 2c_j^{n+\theta} + c_{j-1}^{n+\theta}}{\mathcal{D}x^2} = f_j^{n+\theta}$$

$$c^{n+\theta} = (1 - \theta)c^n + \theta c^{n+1}$$

- $\gamma$  contrôle le décentrage ( $\gamma = 0$  : centré,  $\gamma = 1$  : décentré)
- Schéma implicite
- Inconditionnellement stable
- Ordre 1 ou 2 ( $\theta = 1/2$  et  $\gamma = 0$ )
- Conditions aux limites ...

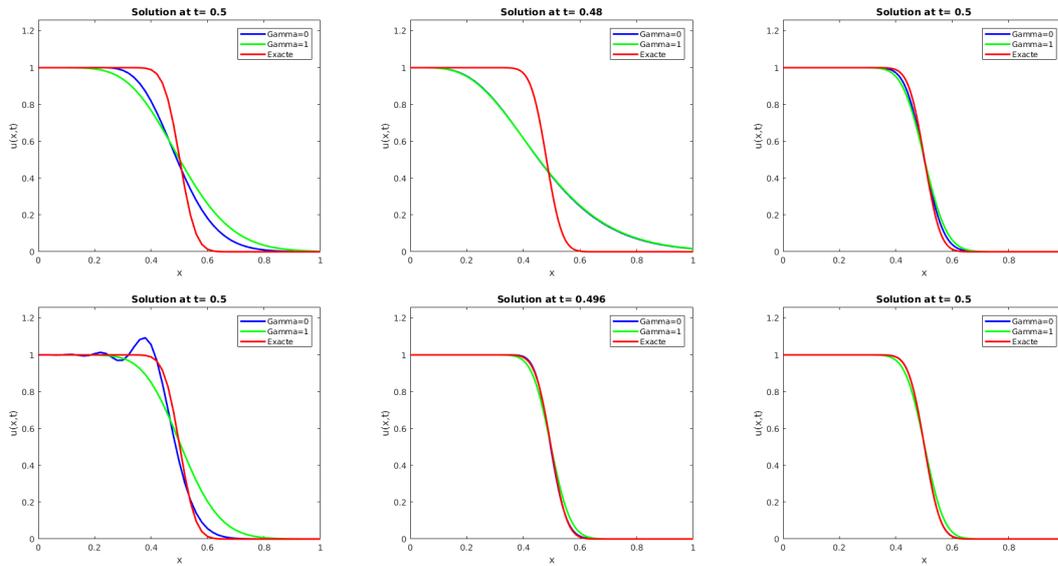


FIGURE B.1 – Équation d’advection–diffusion :  $Pe=20$ ,  $CFL=1$  (gauche),  $Pe=2$ ,  $CFL=4$  (milieu),  $Pe=2$ ,  $CFL=1$  (droite). Figure du haut  $\theta = 1$ , figure du bas  $\theta = 0.5$

### Exemples numériques ( $L = 1$ , $a = 1$ , solutions à $t = 0.5$ )

#### Compléments

##### Problèmes multi-dimensionnels

- Discrétisation dans chaque direction
- – Géométrie complexe, conditions aux limites
- + Facilité de mise en oeuvre, matrices structurées

##### Lois de conservations non–linéaires

- Fréquent en pratique : équations d’Euler, acoustique, ...
- Apparition de *chocs* (discontinuités). Notion de solution faible, entropique
- Numérique : nécessité d’introduire des *limiteurs de pentes* pour éviter les oscillations.

#### Méthodes de volumes finis

# Bibliographie

## Références de base

- [AGL07] Jørg E. AARNES, Tore GIMSE et Knut-Andreas LIE. “An introduction to the numerics of flow in porous media using Matlab”. In : *Geometric modelling, numerical simulation, and optimization : applied mathematics at SINTEF*. disponible à <https://folk.ntnu.no/andreas/papers/ResSimMatlab.pdf>. Springer, Berlin, 2007, p. 265-306. DOI : 10.1007/978-3-540-68783-2\_9.
- [All12] Grégoire ALLAIRE. *Analyse numérique et optimisation*. 2ème. Éditions de l'École Polytechnique, 2012. ISBN : 2-7302-1255-8. URL : <http://www.cmap.polytechnique.fr/~allaire/livre2.html>.
- [Asc08] Uri M. ASCHER. *Numerical methods for evolutionary differential equations*. T. 5. Computational Science & Engineering. Society for Industrial et Applied Mathematics (SIAM), Philadelphia, PA, 2008, p. xiv+395. ISBN : 978-0-898716-52-8. DOI : 10.1137/1.9780898718911.
- [Boy10] Franck BOYER. *Méthodes de volumes finis pour les écoulements en milieux poreux*. Disponible en ligne. 2010. URL : [https://www.math.univ-toulouse.fr/~fboyer/\\_media/exposes/tipaza2010.pdf](https://www.math.univ-toulouse.fr/~fboyer/_media/exposes/tipaza2010.pdf).
- [Des10] Bruno DESPRÉS. *Lois de conservations eulériennes, lagrangiennes et méthodes numériques*. T. 68. Mathématiques & Applications (Berlin) [Mathematics & Applications]. Springer-Verlag, Berlin, 2010, p. x+284. ISBN : 978-3-642-11656-8. DOI : 10.1007/978-3-642-11657-5.
- [DL88] Robert DAUTRAY et Jacques-Louis LIONS. *Analyse mathématique et calcul numérique pour les sciences et les techniques*. Vol. 9. INSTN : Collection Enseignement. Évolution : numérique, transport. Masson, Paris, 1988, i-xliv and 855-1303. ISBN : 2-225-81303-5.
- [EGH00] Robert EYMARD, Thierry GALLOUËT et Raphaèle HERBIN. “Finite volume methods”. In : *Handbook of numerical analysis, Vol. VII*. Handb. Numer. Anal., VII. disponible à <http://www.i2m.univ-amu.fr/~herbin/PUBLI/bookevol.pdf>. North-Holland, Amsterdam, 2000, p. 713-1020. DOI : 10.1086/phos.67.4.188705.
- [GDS15] David F. GRIFFITHS, John W. DOLD et David J. SILVESTER. *Essential partial differential equations*. Springer Undergraduate Mathematics Series. Analytical and computational aspects. Springer, Cham, 2015, p. xi+368. ISBN : 978-3-319-22568-5. DOI : 10.1007/978-3-319-22569-2.

- [Gui03] Vincent GUINOT. *Godunov-type Schemes. An Introduction for Engineers*. 1<sup>re</sup> éd. Elsevier, 2003. DOI : 10.1016/B978-0-444-51155-3.X5000-2.
- [Her18] Raphael HERBIN. “EDP : aspects numériques”. <https://www.i2m.univ-amu.fr/perso/raphaele.herbin/PUBLI/anedp.pdf>. 2018.
- [Ise09] Arieh ISERLES. *A first course in the numerical analysis of differential equations*. Second. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2009, p. xx+459. ISBN : 978-0-521-73490-5. DOI : 10.1017/CB09780511995569.
- [KLd20] David I. KETCHESON, Randall J. LEVEQUE et Mauricio J. DEL RAZO. *Riemann problems and Jupyter solutions*. English. T. 16. Philadelphia, PA : Society for Industrial et Applied Mathematics (SIAM), 2020, p. xii + 166. ISBN : 978-1-61197-620-5/pbk ; 978-1-61197-621-2/ebook. DOI : 10.1137/1.9781611976212.
- [LeV02] Randall J. LEVEQUE. *Finite volume methods for hyperbolic problems*. English. Cambridge : Cambridge University Press, 2002, p. xix + 558. ISBN : 0-521-00924-3/pbk ; 0-521-81087-6/hbk.
- [LeV07] Randall J. LEVEQUE. *Finite difference methods for ordinary and partial differential equations. Steady-state and time-dependent problems*. Society for Industrial et Applied Mathematics (SIAM), Philadelphia, PA, 2007, p. xvi+341. ISBN : 978-0-898716-29-0. DOI : 10.1137/1.9780898717839.
- [LeV92] Randall J. LEVEQUE. *Numerical methods for conservation laws. 2nd ed*. English. 2nd ed. Basel : Birkhäuser, 1992, p. x + 214. ISBN : 3-7643-2464-3.
- [LL16] Hervé LE DRET et Brigitte LUCQUIN. *Partial differential equations : Modeling, analysis and numerical approximation*. English. T. 168. Cham : Birkhäuser/Springer, 2016, p. xi + 395. ISBN : 978-3-319-27065-4/hbk ; 978-3-319-27067-8/ebook. DOI : 10.1007/978-3-319-27067-8.

## Pour en savoir plus

- [Bar+16] Gabriel R. BARRENECHEA et al., éd. *Building bridges : connections and challenges in modern approaches to numerical partial differential equations*. T. 114. Lecture Notes in Computational Science and Engineering. Selected papers from the 101st LMS-EPSRC Symposium held at Durham University, Durham, July 8–16, 2014. Springer, [Cham], 2016, p. viii+431. ISBN : 978-3-319-41638-0 ; 978-3-319-41640-3. DOI : 10.1007/978-3-319-41640-3.
- [BBF13] Daniele BOFFI, Franco BREZZI et Michel FORTIN. *Mixed finite element methods and applications*. T. 44. Springer Series in Computational Mathematics. Springer, Heidelberg, 2013, p. xiv+685. ISBN : 978-3-642-36518-8 ; 978-3-642-36519-5. DOI : 10.1007/978-3-642-36519-5.
- [Bei+16] Lourenco BEIRÃO DA VEIGA et al. “Virtual element implementation for general elliptic equations”. In : *Building bridges : connections and challenges in modern approaches to numerical partial differential equations*. T. 114. Lect. Notes Comput. Sci. Eng. Springer, [Cham], 2016, p. 39-71. DOI : 10.1007/978-3-319-41640-3\_2.

- [BLM14] Lourenço BEIRÃO DA VEIGA, Konstantin LIPNIKOV et Gianmarco MANZINI. *The mimetic finite difference method for elliptic problems*. T. 11. MS&A. Modeling, Simulation and Applications. Springer, Cham, 2014, p. xvi+392. ISBN : 978-3-319-02662-6 ; 978-3-319-02663-3. DOI : 10.1007/978-3-319-02663-3.
- [CG09] Gregory A. CHECHKIN et Andrei Yu. GORITSKY. “S.N. Kruzhkov’s lectures on first-order quasilinear PDEs”. In : *De Gruyter Proceedings in Mathematics*. Sous la dir. d’E. EMMRICH et P. WITTBOLD. Analytical and Numerical Aspects of Partial Differential Equations. traduit du russe par B. Andreianov. De Gruyter, juill. 2009, pp. 1-68. DOI : 10.1515/9783110212105.1. URL : <https://hal.archives-ouvertes.fr/hal-00363287>.
- [Cla20] CLAWPACK DEVELOPMENT TEAM. *Clawpack software*. Version 5.7.1. 2020. DOI : <https://doi.org/10.5281/zenodo.4025432>. URL : <http://www.clawpack.org>.
- [CR91] G. CHAVENT et J.E. ROBERTS. “A unified physical presentation of mixed, mixed-hybrid finite elements and standard finite difference approximations for the determination of velocities in waterflow problems”. In : *Advances in Water Resources* 14.6 (1991), p. 329-348. ISSN : 0309-1708. DOI : [https://doi.org/10.1016/0309-1708\(91\)90020-0](https://doi.org/10.1016/0309-1708(91)90020-0).
- [DD20] Daniele Antonio DI PIETRO et Jérôme DRONIOU. *The hybrid high-order method for polytopal meshes*. T. 19. MS&A. Modeling, Simulation and Applications. Design, analysis, and applications. Springer, Cham, [2020] ©2020, p. xxxi+525. ISBN : 978-3-030-37202-6 ; 978-3-030-37203-3. DOI : 10.1007/978-3-030-37203-3.
- [DE12] Daniele Antonio DI PIETRO et Alexandre ERN. *Mathematical aspects of discontinuous Galerkin methods*. T. 69. Mathématiques & Applications (Berlin) [Mathematics & Applications]. Springer, Heidelberg, 2012, p. xviii+384. ISBN : 978-3-642-22979-4. DOI : 10.1007/978-3-642-22980-0.
- [Dro14] Jerome DRONIOU. “Finite volume schemes for diffusion equations : Introduction to and review of modern methods”. In : *Mathematical Models and Methods in Applied Sciences* 24.08 (2014), p. 1575-1619. DOI : 10.1142/S0218202514400041.
- [EGH09] R. EYMARD, T. GALLOUËT et R. HERBIN. “Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes SUSHI : a scheme using stabilization and hybrid interfaces”. In : *IMA Journal of Numerical Analysis* 30.4 (juin 2009), p. 1009-1043. ISSN : 0272-4979. DOI : 10.1093/imanum/drn084.
- [Gat14] G.N. GATICA. *A simple introduction to the mixed finite element method. Theory and applications*. SpringerBriefs in Mathematics. Springer, 2014, p. xii+132. ISBN : 978-3-319-03694-6 ; 978-3-319-03695-3. DOI : 10.1007/978-3-319-03695-3.
- [GR21] Edwige GODLEWSKI et Pierre-Arnaud RAVIART. *Numerical approximation of hyperbolic systems of conservation laws*. T. 118. Applied Mathematical Sciences. Second edition [of 1410987]. Springer-Verlag, New York, [2021] ©2021, p. xiii+840. ISBN : 978-1-0716-1342-9 ; 978-1-0716-1344-3. DOI : 10.1007/978-1-0716-1344-3.
- [Hes18] Jan S. HESTHAVEN. *Numerical methods for conservation laws*. T. 18. Computational Science & Engineering. From analysis to algorithms. Society for Industrial et Applied Mathematics (SIAM), Philadelphia, PA, 2018, p. xvi+570. ISBN : 978-1-611975-09-3. DOI : 10.1137/1.9781611975109.

- [Lax73] Peter D. LAX. *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 11. Society for Industrial et Applied Mathematics, Philadelphia, Pa., 1973, p. v+48.
- [Osh84] Stanley OSHER. “Riemann Solvers, the Entropy Condition, and Difference”. In : *SIAM Journal on Numerical Analysis* 21.2 (1984), p. 217-235. DOI : 10.1137/0721016. eprint : <https://doi.org/10.1137/0721016>.
- [Riv08] Béatrice RIVIÈRE. *Discontinuous Galerkin methods for solving elliptic and parabolic equations*. T. 35. Frontiers in Applied Mathematics. Theory and implementation. Society for Industrial et Applied Mathematics (SIAM), Philadelphia, PA, 2008, p. xxii+190. ISBN : 978-0-898716-56-6. DOI : 10.1137/1.9780898717440.
- [Tor09] Eleuterio F. TORO. *Riemann solvers and numerical methods for fluid dynamics*. Third. A practical introduction. Springer-Verlag, Berlin, 2009, p. xxiv+724. ISBN : 978-3-540-25202-3. DOI : 10.1007/b79761.