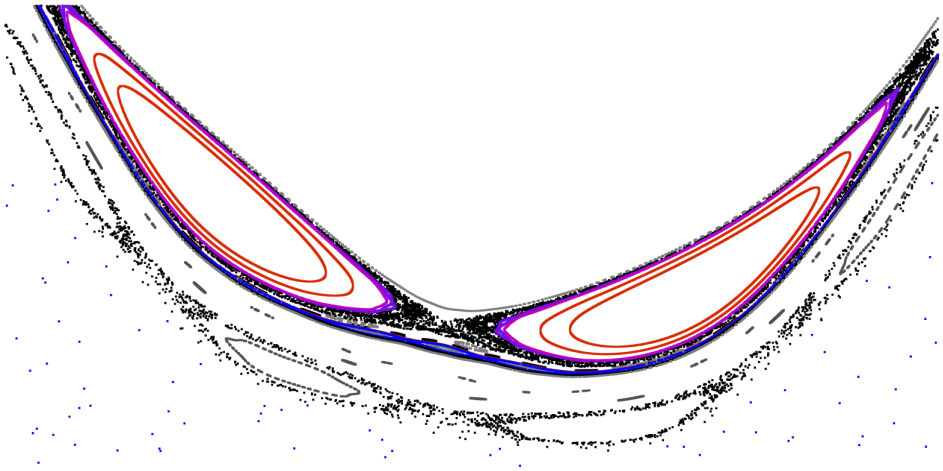
PHYSIQUE NUMÉRIQUE

Laurent Villard

Swiss Plasma Center
Faculté des Sciences de Base
Section de Physique
Ecole Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland
SPC - EPFL



Swiss Plasma Center

Version 25.2

3 mars 2025

Remerciements

Je remercie chaleureusement mon collègue Andreas Läuchli pour son aide précieuse lors de la première édition de ce cours, tout particulièrement à la création des exercices, qui sont une part essentielle de la pédagogie de ce cours.

Année après année, j'ai bénéficié du support enthousiaste de plusieurs équipes d'assistants, qui ont apporté une contribution primordiale au succès de ce cours. Je leur exprime ici toute ma gratitude.

Image de couverture : sections de Poincaré d'un pendule simple exclité non-amorti avec différentes conditions initiales. Schéma de Verlet.

Table des matières

Chapitre 1

Introduction

1.1 Présentation du cours

Ce cours est destiné aux étudiants de la Section de Physique de l'EPFL de deuxième année. Il suppose avoir acquis les notions des cours de mathématiques (analyse et algèbre linéaire), d'analyse numérique, de physique et d'informatique de l'année propédeutique.

Ce cours n'est ni un cours de programmation scientifique avancée, ni un cours de mathématiques discrètes. C'est un cours de physique.

Il n'a pas pour objectif de former des théoriciens spécialistes pointus des algorithmes numériques. Cependant, il est clair aujourd'hui que tout physicien sera confronté un jour ou l'autre à un problème de nature numérique, que ce soit seulement en tant qu'utilisateur d'un "package", ou que ce soit un expérimentateur confronté à des problèmes d'échantillonage et d'analyse du signal. D'autre part, il serait dommage, vu la puissance de calcul et l'aisance d'utilisation des ordinateurs, de se passer d'un outil qui, comme l'outil analytique, et en complément de celui-ci, permet de résoudre des problèmes de physique et ainsi aider à la compréhension de nombreux phénomènes.

Les problèmes que l'on peut résoudre "exactement" (c.a.d. par des méthodes analytiques) sont très restreints, dans le sens qu'ils sont souvent basés sur une idéalisation, une simplification de la réalité : par exemple, on néglige la résistance de l'air, ou on néglige la présence d'un troisième corps céleste pour calculer les trajectoire d'une planète, etc. La réalité est bien plus complexe, et la confrontation théorie - expérience doit faire face au dilemne suivant : les différences entre prédictions et mesures sont-elles dues à des effets négligés (pour pouvoir résoudre *exactement* des équations représentant une *approximation* de la réalité), ou à des imprécisions de mesure, ou encore indiquent-elles une défaillance fondamentale de la théorie utilisée ?

L'approche numérique permet de tenir compte de plusieurs effets traditionnellement négligés. Mais elle a elle-même ses limites : la solution numérique est *fragmentaire* et *approximative*. C'est pourquoi il est **absolument crucial de pouvoir évaluer la qualité de la solution numérique**. C'est un des objectifs essentiels de ce cours que de développer ce type d'attitude face à la solution numérique.

Objectifs

- Aborder, formuler et résoudre des problèmes de physique pouvant être décrits par des équations différentielles ordinaires ou aux dérivées partielles, en utilisant des méthodes numériques.
- Comprendre les avantages et les limites de ces méthodes.
- Etendre les applications aux problèmes difficilement traitables par les méthodes analytiques.
- Apprendre à utiliser les concepts physiques pour vérifier et valider les résultats numériques.
- Contrôler la précision en estimant les erreurs, en examinant la stabilité et la convergence.
- Compléter et illustrer différents sujets de physique traités dans d'autres cours.

Organisation

Le cours est organisé avec une partie d'enseignement "frontal", où les problèmes physiques et leur modélisation, ainsi que les méthodes numériques sont présentées. Il est suivi d'une partie de "travaux pratiques", où il s'agit de résoudre des exercices. Ces exercices feront l'objet de rapports à rendre, qui seront ensuite évalués et notés.

Contenu

Après une introduction à la discrétisation et aux concepts de la convergence et de la stabilité numériques, le cours aborde les problèmes d'évolution temporelle à valeur initiale, en partant du cas le plus simple.

On s'intéresse ensuite aux problèmes, essentiellement tirés de la mécanique Newtonienne, pour lesquelles divers intégrateurs numériques sont développés et analysés, et qui permettent d'aller au-delà, dans les applications, des exemples traditionnellement choisis. On verra, par exemple, l'apparition de mouvement chaotique dans des systèmes simples, ce qui aura des conséquences sur les notions de prédictabilité et de déterminisme.

On aborde ensuite les problèmes d'intégration d'équations différentielles dans l'espace, à une, puis deux dimensions. Les applications physiques seront tirées de la thermodynamique et de l'électromagnétisme.

Physique	**Numérique**
1. Introduction	1. Discrétisation, erreurs, convergence, stabilité
2. Evolution temporelle. Problèmes à valeur initiale décrits par des équations différentielles ordinaires. Oscillations. Chaos. Gravitation à 1,2 et 3 corps. Particules dans champ EM. Problèmes 1D à valeurs aux bords traités comme à valeur initiale.	2. Schémas explicites : Euler explicite, Euler symplectique, Verlet, Leapfrog, Runge-Kutta. Schéma d'Euler implicite. Schéma semi-implicite : Boris-Buneman. Stabilité et convergence. Pas de temps adaptatif. Traitement de la singularité des équations.
3. Intégration spatiale. Problèmes à valeurs aux bords. Electrostatique, magnétostatique, chaleur stationnaire.	3. Différences finies. Méthodes accélératrices : Gauss-Seidel, surrrelaxation. Elements finis. Grille non-uniforme.
4. Intégration spatio-temporelle : problèmes décrits par des équations aux dérivées partielles. Advection-Diffusion. Ondes : propagation, réflexion, superposition, milieux inhomogènes. Mécanique quantique : Schrödinger dépendante du temps, principe d'incertitude, puits et barrières de potentiel, effet tunnel, oscillateur harmonique, potentiel périodique. Schrödinger stationnaire, états propres.	4. Différences finies. Schémas explicites à 2 et 3 niveaux. Application de diverses conditions initiales et conditions aux bords. Analyse de stabilité. Monte Carlo Langevin (marche aléatoire). Schéma semi-implicite de Crank-Nicholson. Propriétés de conservation.
5. Physique statistique. Transitions de phase.	5. Monte Carlo. Algorithme de Metropolis.

TABLE 1.1 – *Correspondance entre les problèmes de physique abordés (colonne de gauche) et les méthodes numériques introduites (colonne de droite). Les numéros correspondent aux chapitres du cours.*

Les problèmes d'évolution spatio-temporelles seront évoqués, avec applications possibles aux problèmes de l'advection-diffusion, de la propagation d'ondes et de la mécanique quantique.

Un exemple simple d'application de la méthode de Monte Carlo, permettant de simuler le comportement statistique des systèmes à plusieurs degrés de liberté, sera présenté.

Enfin, le lecteur pourra à profit consulter les quelques ouvrages de référence, articles scientifiques et liens sur la toile mentionnés dans la Bibliographie, p. 206.

Structure

La présentation de ce cours suit une double logique. D'une part, la motivation est basée sur la physique, où des problèmes de complexité croissante sont abordés. Ces problèmes servent de motivation à l'introduction de méthodes numériques. La table 1.1 indique cette correspondance.

1.2 Discrétisation

On appelle **discrétisation** le processus de remplacer un système d'équations sur un espace continu, généralement des équations différentielles, en le représentant de façon **approximative** en termes d'un ensemble discret (dénombrable et fini) de quantités.

La figure 1.1 illustre ce propos, avec, sur la partie gauche, la représentation continue, et à droite, la représentation discrète.

Les quantités discrètes peuvent par exemple être les valeurs des fonctions en des points d'un réseau (appelé aussi *maillage*). Pour obtenir une estimation de la dérivée de ces fonctions, on peut, par exemple, utiliser alors des différences finies, qui seront une *approximation*, plus ou moins bonne, selon l'ordre utilisé. L'annexe A décrit plus en détail comment obtenir les formules de différences finies.

Une autre possibilité est de représenter les fonctions en termes d'une somme de fonctions de base ayant un support fini, également défini sur un réseau (maillage). Les méthodes d'éléments finis sont un exemple d'une telle approche. Les méthodes dites spectrales, avec des fonctions de base globales, typiquement harmoniques ou polynomiales, sont un autre exemple.

On obtient ainsi, à partir du système originel d'équations différentielles, un système d'équations algébriques, qui peut être résolu par des opérations arithmétiques. Moyen-

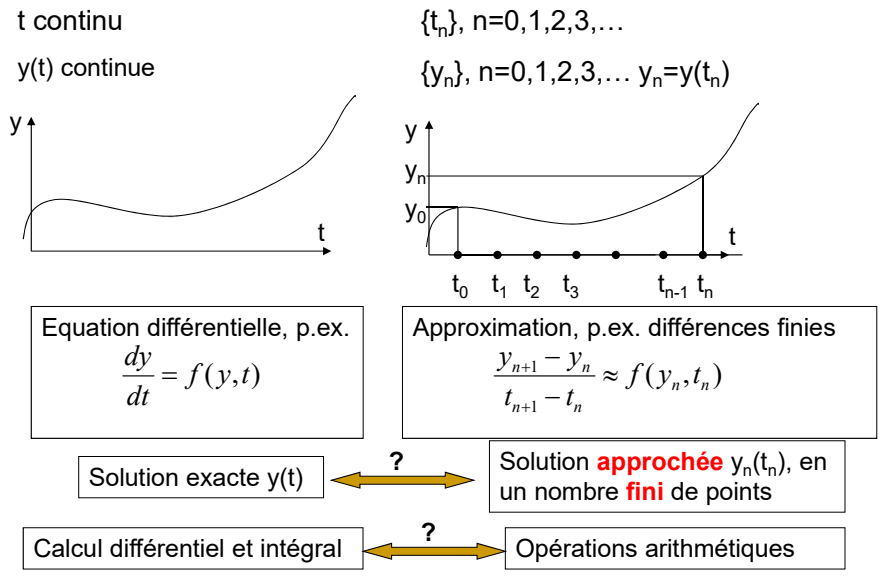


FIGURE 1.1 – *Correspondance entre la représentation continue, à gauche, et la représentation discrète, à droite.*

nant traduction par un langage de programmation, ces opérations peuvent s'exécuter par le processeur arithmétique d'un ordinateur.

Il est crucial de comprendre à quel point la solution discrétisée de notre problème repésente fidèlement, ou non, la solution du problème continu initial. Il s'agit ici de pouvoir *quantifier* cette *"fidélité"*. Un des objectifs principaux de ce cours est donc la compréhension des **erreurs**, des proprétés de **convergence** et de **stabilité** numériques. Ces concepts sont introduits brièvement dans les prochaines sections.

1.3 Erreurs de troncature et d'arrondi

Tout processus de discrétisation s'accompagne généralement d'erreurs. On parle d'erreurs de troncature et d'arrondi. Les erreurs de *troncature* sont directement liées à la façon dont on a approximé le problème continu. L'effet de ces erreurs dépend du type d'équations et peut être bénin. Un "bon" schéma numérique est tel que plus on utilise un réseau (maillage) fin, plus l'erreur de discrétisation diminue, pour tendre vers zéro. Cette propriété s'appelle la **convergence** numérique.

On verra que les conséquences des erreurs de discrétisation peuvent parfois être dramatiques et conduire à des *instabilités* totalement non-physiques : l'erreur croît exponentiellement dans le temps.

Les erreurs d'*arrondi* sont dues à la représentation des nombres réels par un nombre fini de bits : c'est le cas de toute arithmétique exécutée par un processeur.

Nous nous bornerons ici à évoquer brièvement le problème des erreurs et de leur accumulation. Les formules de différences finies sont introduites, pour certaines d'entre elles, ''a la section suivante et à l' annexe A. Des formules d'intégration numérique sont démontrées à l'annexe B.

La plupart des problèmes de physique que nous allons aborder aboutissent à une (ou plusieurs) équation(s) différentielle(s). Au coeur du problème de leur résolution numérique se trouve donc la question de la représentation et du calcul d'une dérivée. Par exemple, la formule d'ordre le plus bas en $h = \Delta x$ de la dérivée de premier ordre pour une fonction $f(x)$, Eq.(A.22), est :

$$\frac{\mathrm{d}f}{\mathrm{d}x}(x) = \frac{f(x+\Delta x) - f(x)}{\Delta x} + \mathcal{O}(\Delta x) \ . \tag{1.1}$$

Le symbole $\mathcal{O}(\Delta x)$ signifie que l'erreur commise dans l'approximation de la dérivée sera linéairement proportionnelle à Δx, dans la limite $\Delta x \to 0$. C'est **l'erreur de troncature** (on a tronqué ici le développement limité de la fonction f pour obtenir cvette formule de différence finie). On retrouve donc l'erreur de troncature lors de l'évaluation numérique de la dérivée.

Mais il y a un autre type d'erreur : non seulement les fonctions ne sont connues qu'en un ensemble *discret* de points, la représentation numérique d'un nombre réel dans un processeur utilise *un nombre fini de bits*, autrement dit la représentation de chaque valeur f_j et chaque x_j est elle-même *discrète*. On appelle ce type d'erreur **l'erreur d'arrondi** (le nombre réel est "arrondi" à sa plus proche représentation binaire sur le processeur).

Ces deux types d'erreurs sont illustrées à la FIG. 1.2 pour le cas de l'évaluation numérique de $\mathrm{d}\sin(x)/\mathrm{d}x$ en $x = \pi/4$. On a représenté, avec des axes logarithmiques, l'erreur par rapport à la solution exacte en fonction du choix de Δx. pour des Δx pas trop petits, l'erreur diminue effectivement linéairement avec Δx (la pente est $+1$ sur ce graphique logarithmique). Mais pour Δx tendant vers zéro, l'erreur *diverge* et l'évaluation de la dérivée **diverge** : ce comportement est dû aux erreurs d'arrondi, exemplifiées par la formule ci-dessus qui, au numérateur, évalue la différence de deux nombres voisins : il y a perte d'information. On constate que si le comportement de l'erreur de troncature est régulier, celui de l'erreur d'arrondi est erratique. Cela tient à la nature discrète de la représentation binaire d'un nombre réel. En conséquence de ces deux sources d'erreur, la précision sur le résultat est en pratique limitée à $\sim 10^{-7} - 10^{-8}$, qui est à peu près la moitié de la précision de la représentation à 64 bits.

Dans la suite du cours, nous allons intégrer des équations différentielles en prenant un grand nombre de points de la grille, ou de pas temporels, et la question de *l'accumulation des erreurs de troncature et d'arrondi* doit être surveillée.

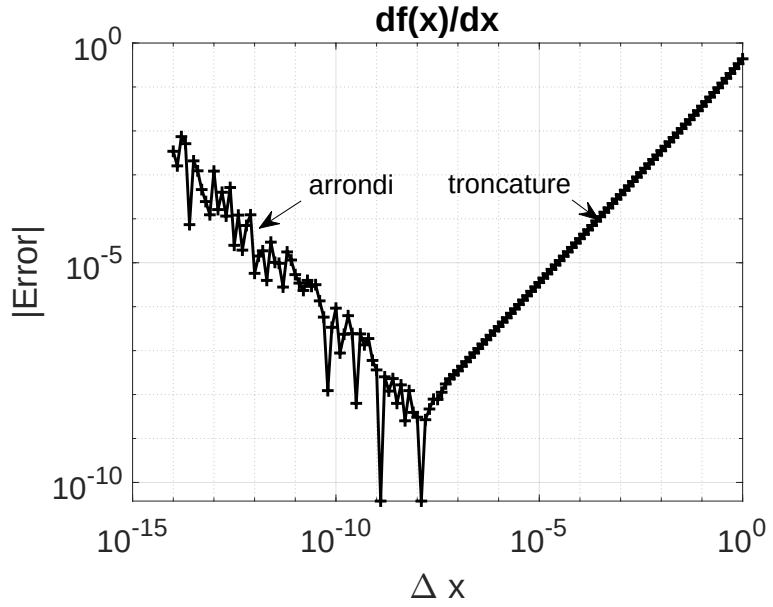


FIGURE 1.2 – *Erreur sur l'évaluation numérique de la dérivée* $\mathrm{d}\sin(x)/\mathrm{d}x$ *en* $x = \pi/4$*, en fonction de* Δx*. On a utilisé la formule de différence finies "forward", Eq.(1.1).*

1.4 Différences finies et développements limités

Dans la plupart des situations, des considérations d'ordre physique nous permettent de supposer que les fonctions décrivant le système et son évolution sont continues et n fois (si ce n'est indéfiniment) différentiables. On peut donc se baser sur le développement limité de Taylor de ces fonctions au voisinage des points de discrétisation (appelés "points du réseau" ou "points du maillage").

Soit une fonction $f \in \mathcal{C}^\infty(\mathbb{R})$. Soit une discrétisation x_j avec des points de maillage équidistants, $h_j = x_{j+1} - x_j = h, \forall j$. Soit $f_j = f(x_j)$ et $f'_j = \mathrm{d}f(x_j)/\mathrm{d}x$. On écrit les développements limités de la fonction f au voisinage du point de maillage x_j, que l'on exprime aux points de maillage $x_{j\pm1}, x_{j\pm2}$:

$$f_{j-2} = f_j - 2hf'_j + 2h^2 f''_j - \frac{8}{6}h^3 f^{(3)}_j + \frac{16}{24}h^4 f^{(4)}_j - \frac{32}{120}h^5 f^{(5)}_j + \mathcal{O}(h^6) \quad (1.2)$$

$$f_{j-1} = f_j - hf'_j + \frac{1}{2}h^2 f''_j - \frac{1}{6}h^3 f^{(3)}_j + \frac{1}{24}h^4 f^{(4)}_j - \frac{1}{120}h^5 f^{(5)}_j + \mathcal{O}(h^6) \quad (1.3)$$

$$f_{j+1} = f_j + hf'_j + \frac{1}{2}h^2 f''_j + \frac{1}{6}h^3 f^{(3)}_j + \frac{1}{24}h^4 f^{(4)}_j + \frac{1}{120}h^5 f^{(5)}_j + \mathcal{O}(h^6) \quad (1.4)$$

$$f_{j+2} = f_j + 2hf'_j + 2h^2 f''_j + \frac{8}{6}h^3 f^{(3)}_j + \frac{16}{24}h^4 f^{(4)}_j + \frac{32}{120}h^5 f^{(5)}_j + \mathcal{O}(h^6) \quad (1.5)$$

De l'Eq.(1.3), on obtient :

$$f_{j-1} - f_j = -hf'_j\mathcal{O}(h^2)\ .$$

En divisant par $-h$, on obtient :

$$\boxed{f'_j = \frac{f_j - f_{j-1}}{h} + \mathcal{O}(h)}\,. \tag{1.6}$$

C'est la formule de la première dérivée "en arrière" ("backward") ou "rétrograde". Elle est *d'ordre 1* en h (autrement dit, l'erreur de troncature est proportionnelle à la taille du maillage h).

On peut procéder de même à partir de l'Eq.(1.4) :

$$f_{j+1} - f_j = h f'_j \mathcal{O}(h^2)\ .$$

En divisant par h, on obtient :

$$\boxed{f'_j = \frac{f_{j+1} - f_j}{h} + \mathcal{O}(h)}\,. \tag{1.7}$$

C'est la formule de la première dérivée "en avant" ("forward") ou "progressive". Elle a une erreur *d'ordre 1* en h (autrement dit, l'erreur de troncature est proportionnelle à la taille du maillage h).

On peut faire mieux. En faisant la différence de l'Eq.(1.3) et de l'Eq.(1.4), on élimine les termes de dérivées d'ordre pair et on obtient :

$$f_{j+1} - f_{j-1} = 2h f'_j + \mathcal{O}(h^3)$$

En divisant cette expression par $2h$, on a :

$$\boxed{f'_j = \frac{f_{j+1} - f_{j-1}}{2h} + \mathcal{O}(h^2)}\,. \tag{1.8}$$

C'est la formule de la première dérivée ***centrée***. Elle a une erreur *d'ordre 2* en h (autrement dit, l'erreur de troncature est proportionnelle au carré de la taille du maillage, h^2).

On constate que **prendre un schéma centré augmente l'ordre, donc la précision obtenue, par rapport aux schémas décentrés.** Ceci est illustré à la Fig.1.3, où on a représenté l'erreur sur l'évaluation de $\mathrm{d}\sin(x)/\mathrm{d}x$ en $x = \pi/4$ en utilisant le schéma centré, Eq.(1.8) (en rouge). On constate que l'erreur de troncature varie bien en h^2 (la pente est 2 sue le diagramme log-log). Par rapport au schéma décentré "forward", Eq.(1.7) (en noir), on constate bien l'avantage du schḿa centré. Pour une taille de maillage donnée, la précision est bien meilleure, tant qu'on n'est pas dominé par les erreurs d'arrondi. Le minimum de l'erreur est $\sim 10^{-11}$, bien meilleur que les 10^{-8} du schéma forward.

Mais quelle que soit la qualité du schéma et son ordre de convergence, les erreurs d'arrondi sont inévitables. On trouve une divergence de l'erreur en $\sim 1/h$, indépendemment de l'ordre du schéma utilisé.

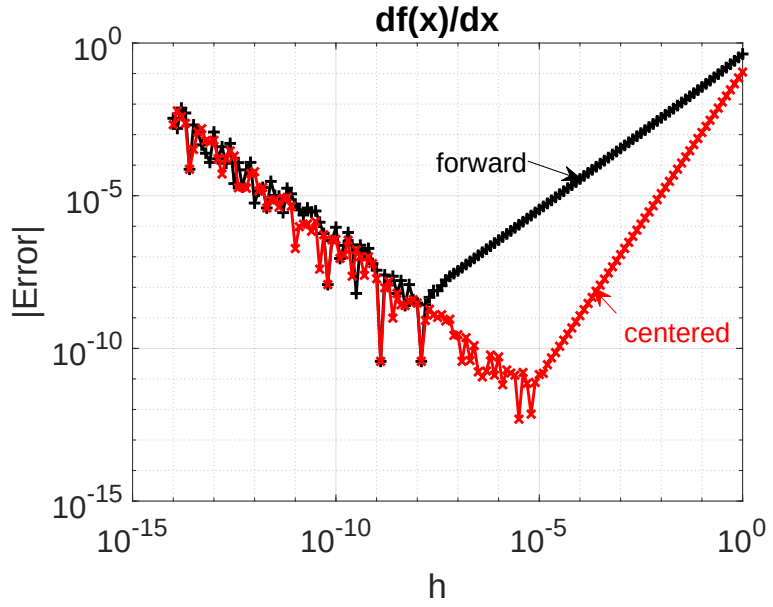


FIGURE 1.3 – *Erreur sur l'évaluation numérique de la dérivée* $\mathrm{d}\sin(x)/\mathrm{d}x$ *en* $x = \pi/4$, *en fonction de* Δx, *avec la formule de différences finies centrées, Eq.(1.8), (rouge) et avec la formule de différences finies "forward", Eq.(1.7) (noir).*

Une formule de différences finies que nous allons utiliser plusieurs fois dans la suite du cours est pour la deuxième dérivée. Elle s'obtient en additionnant les Eqs.(1.3) et Eq.(1.4), ce qui élimine les termes de dérivées d'ordre impair :

$$f_{j-1} + f_{j+1} = 2f_j + h^2 f''_j + \mathcal{O}(h^4)\ . \tag{1.9}$$

En divisant par h^2, on obtient

$$\boxed{f''_j = \frac{f_{j-1} - 2f_j + f_{j+1}}{h^2} + \mathcal{O}(h^2)}\ . \tag{1.10}$$

C'est également une expression *centrée*. Elle a une erreur d'ordre 2 en h. Pour augmenter l'ordre de l'erreur - et donc la précision - des formules de différences finies, il faut inclure non seulement les points immédiatement voisins du point j, $j \pm 1$, mais aussi les points au-delà : $j \pm 2, \pm 3, ...$ Par exemple, on obtient la deuxième dérivée avec une erreur d'ordre 4 en faisant la somme des Eqs(1.3) et(1.4), puis de Eqs(1.2) et (1.5)

$$f_{j-1} + f_{j+1} = 2f_j + h^2 f''_j + \frac{1}{12}h^4 f^{(4)}_j + \mathcal{O}(h^6) \tag{1.11}$$

$$f_{j-2} + f_{j+2} = 2f_j + 4h^2 f''_j + \frac{4}{3}h^4 f^{(4)}_j + \mathcal{O}(h^6) \tag{1.12}$$

En prenant $16\times$ Eq.(1.11) - Eq.(1.12), on élimine $f^{(4)}_j$:

$$-f_{j-2} + 16f_{j-1} + 16f_{j+1} - f_{j+2} = 30f_j + 12h^2 f''_j + \mathcal{O}(h^6)$$

et on obtient :

$$\boxed{f''_j = \frac{1}{12h^2}\left(-f_{j-2} + 16f_{j-1} - 30f_j + 16f_{j+1} - f_{j+2}\right) + \mathcal{O}(h^4)}\ . \tag{1.13}$$

On trouvera en annexe A la dérivation d'autres formules de différences finies.

1.5 Convergence numérique

Dans cette section, nous abordons la question de savoir comment les erreurs de troncature des schémas numériques peuvent être quantifiés. En particulier, il est important de déterminer comment la solution numérique se comporte avec une discrétisation de plus en plus fine. Par exemple, pour les problèmes d'évolution temporelle avec des pas de temps Δt :

1. La solution numérique, pour $\Delta t \to 0$, tend-elle vers une solution finie ?
2. Cette solution coïncide-t-elle avec la solution exacte (analytique) du problème ?
3. De quelle façon la précision du résultat numérique augmente-t-elle lorsque Δt diminue ?

En d'autres termes : (1) le schéma numérique **converge**-t-il, (2) converge-t-il vers la **bonne solution**, et (3) à quelle **ordre** le schéma converge-t-il ?

Ces questions sont typiquement abordées dans un cours d'analyse numérique, avec des démonstrations mathématiques rigoureuses. Nous nous bornerons ici à donner la définition de l'ordre de convergence, dans le cas d'une équation différentielle pour une fonction $y(t)$, du type

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(y,t)$$

avec une fonction connue $f(y,t)$ et une condition initiale connue $y(0) = y_0$. On s'intéresse à la solution obtenue au temps final $t = t_f$. On supposera que la solution exacte du problème est analytique, c'est-à-dire qu'elle possède un développement en série entière au voisinage de tout point. La plupart des méthodes numériques ne s'appliquent correctement que si une telle hypothèse est vérifiée. La solution numérique en $t = t_f$ sera généralement différente de la solution exacte en ce point. En discrétisant l'intervalle $[0, t_f]$ en N points de maillage équidistants, $\Delta t = t_f/N$, on dit que **la solution numérique converge à l'ordre** n si on peut écrire :

$$y_{\mathrm{num}}(t_f) = y_{\mathrm{exact}}(t_f) + c_n(\Delta t)^n + c_{n+1}(\Delta t)^{n+1} + ... \tag{1.14}$$

Autrement dit, tous les termes en $c_m(\Delta t)^m$, avec $m < n$, doivent être tels que c_m=0.

Dans ce qui suit, nous nous intéresserons à la façon d'exécuter une étude de convergence et de représenter ("montrer") les résultats.

Deux cas de figure peuvent se présenter. Dans le premier cas, si on dispose d'une solution exacte du problème, alors on peut calculer une **erreur**, qui est la différence entre la solution numérique et la solution exacte. En effectuant une série de simulations avec des

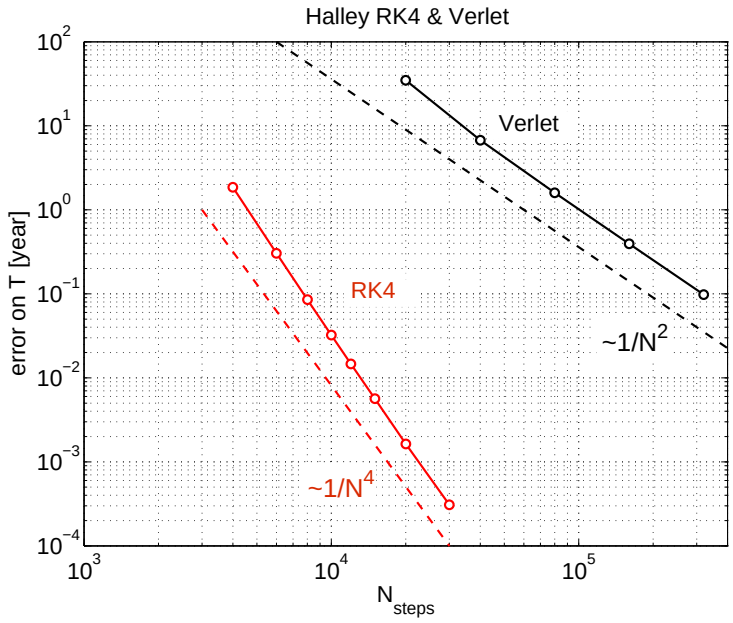


FIGURE 1.4 – *Convergence numérique de la période de révolution de la comète de Halley, obtenue avec le schéma de Verlet (noir) et le schéma de Runge-Kutta d'ordre 4. Ici, on connaît la solution exacte et donc on peut calculer une ereur = différence entre solutions numérique et exacte. On a représenté l'erreur en fonction du nombre de pas de temps sur une échelle log-log. Les lignes traitillées sont de pente* -2 *(noir), respectivement* -4 *(rouge), indiquant que l'ordre de convergence de ces schémas est de* 2 *(Verlet), respectivement 4 (RK4).*

discrétisations de plus en plus fines, on peut alors représenter la valeur absolue de cette erreur en fonction de Δt. En choisissant des échelles log-log, on obtient la réponse aux trois questions ci-dessus. En particulier, **la pente du graphe de l'erreur en fonction de Δt sur un diagramme log-log, dans la limite $\Delta t \to 0$, est l'ordre de convergence numérique**. En effet, à partir de l'Eq.(1.14), on obtient :

$$\log(|y_{\rm num}(t_f) - y_{\rm exact}(t_f)|) = \log(c_n) + n\log(\Delta t) + ... \tag{1.15}$$

On peut aussi représenter le logarithme de l'erreur en fonction du logarithme du nombre de pas de temps N :

$$\log(|y_{\rm num}(t_f) - y_{\rm exact}(t_f)|) = \log(c_n t_{fin}^n) - n\log(N) + ... \tag{1.16}$$

Au signe près, la pente est l'ordre de convergence. Un exemple est montré à la Fig. 1.4. Il s'agit de la période de révolution d'une comète, obtenue avec divers schémas numériques qui seront examinés au Chapitre 2.]

Dans le deuxième cas, si on ne dispose pas d'une solution exacte, on ne peut alors pas calculer une erreur, et la question (2) restera ouverte. Par contre, on peut néanmoins répondre aux questions (1) et (3). **Il n'est pas opportun de représenter la quantité calculée en fonction de Δt sur un diagramme log-log :** ce genre de diagramme de

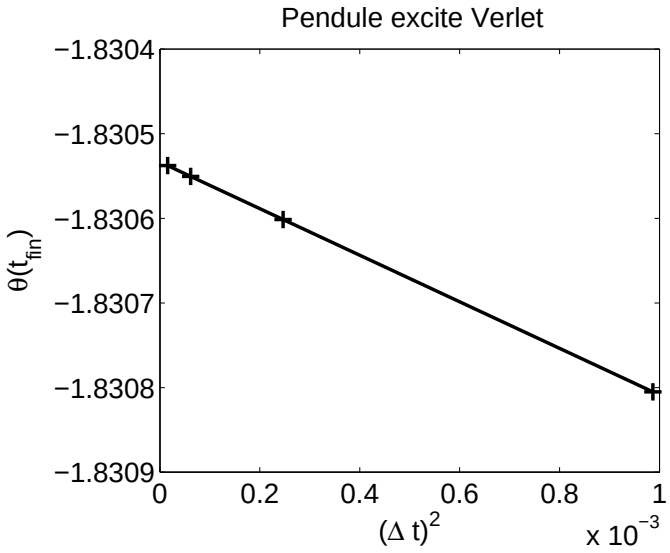


FIGURE 1.5 – *Convergence numérique de la position finale d'un pendule simple soumis à une excitation verticale et un amortissement, obtenue avec le schéma de Verlet. Ici, on ne connaît pas la solution exacte, et donc on ne peut pas a priori calculer l'erreur. On a représenté la valeur de la position finale du pendule en fonction de* $(\Delta t)^2$ *; le fait que les points s'alignent sur ce graphique indique que l'ordre de convergence de ce schéma est de* 2.

nous apporte pas l'information recherchée. En effet, en prenant le log de l'Eq.(1.14), on obtient :

$$\log(y_{\text{num}}(t_f)) = \log(y_{\text{exact}}(t_f)) + (c_n/y_{\text{exact}}(t_f))(\Delta t)^n + ...) \tag{1.17}$$

dont la représentation graphique en fonction de $\log(\Delta t)$ ne donne pas d'information évidente sur n. **La bonne méthode est de représenter la quantité calculée en fonction de** $(\Delta t)^n$**, sur une échelle linéaire-linéaire**. On utilise alors directement l'expression de l'Eq.(1.14). Si les points de mesure s'alignent sur une droite (dans la limite des petits Δt), alors on illustre ainsi que l'ordre de convergence est n. Un exemple est montré à la Fig. 1.5. Il s'agit ici de la position finale d'un pendule simple avec amortissement et excitation extérieure, pour lequel il n'existe pas de solution exacte. On peut alors, en supposant que le comportement en $(\Delta t)^n$ se prolonge jusque dans la limite $\Delta t \to 0$, extrapoler les données pour prendre $\lim \Delta t \to 0$: cette valeur sera la **valeur convergée**.

On peut alors définir une "erreur" comme la différence entre un résultat pour un Δt donné et cette valeur convergée. Ensuite, on peut représenter cette "erreur" en fonction de Δt sur un diagramme log-log. On devrait alors confirmer que la pente sur se diagramme est bien n, l'ordre de convergence.

Pratiquement tous les schémas numériques sont basés sur des développements limités des fonctions jusqu'à un certain ordre. Par exemple, si l'erreur d'un schéma, pour **un** pas de temps, est $\mathcal{O}(\Delta t)^2$, diminuer le pas de temps d'un facteur 2 résulte un une erreur 4 fois plus petite. Cependant, la solution à un temps $t = t_{fin}$ donné, sera entachée d'une

erreur $\mathcal{O}(\Delta t)^1$: ceci parce que les erreurs faites à chaque pas de temps s'additionnent, et le nombre de pas de temps est $t_{fin}/(\Delta t)$. Ainsi, dans cet exemple, diminuer le pas de temps d'un facteur 2 résulte en une erreur diminuée d'un facteur 2 au temps $t = t_{fin}$, et on parle de schéma d'ordre 1. Notons cependant qu'il existe des situations où l'ordre de convergence observé est supérieur à celui attendu : il peut en effet arriver que les erreurs d'ordre le plus bas s'annullent.

1.6 Stabilité numérique

Dans les problèmes d'évolution temporelle, on s'intéresse ici à la façon dont l'erreur numérique commise à un pas de temps $t = t_n$ se "propage" aux pas de temps ultérieurs $t = t_{n+1}, t = t_{n+2}, ...$

Il arrive malheureusement, pour certains schémas numériques et pour certaines équations, que la norme de l'erreur numérique soit amplifiée à chaque pas de temps par un facteur supérieur à 1. L'erreur numérique augmente ainsi exponentiellement au cours du temps. Le schéma numérique est alors dit **instable** pour l'équation considérée.

Il ne faut pas confondre les notions de stabilité et de convergence numériques. Un schéma peut très bien converger et être instable : en effet, on peut avoir une solution numérique qui tend vers la solution exacte dans la limite $\Delta t \to 0$, à un instant $t = t_{fin}$ donné, mais dont l'erreur numérique augmente exponentiellement en fonction de t_{fin}.

Réciproquement, un schéma numérique qui est stable peut ne pas converger. L'erreur numérique, dans ce cas, n'augmente pas exponentiellement au cours du temps, mais ne tend pas vers zéro lorsque Δt tend vers zéro.

Notons encore qu'un schéma numérique peut être stable pour une certaine équation mais instable pour une autre équation. Par exemple, nous verrons que le schéma d'Euler explicite est stable pour le problème de la désintégration, mais instable pour le problème de l'oscillateur harmonique.

Nous reviendrons plus en détail sur ces questions dans le chapitre suivant.

Chapitre 2

Evolution Temporelle - Problèmes à valeurs initiales

Dans ce chapitre, nous allons présenter quelques-unes des méthodes numériques utilisées pour la résolution de problèmes donnés par un système d'équations différentielles ordinaires (EDO) couplées, dont la solution unique requiert la connaissance des conditions initiales. Nous commencerons par le cas le plus simple d'une EDO du premier ordre pour une seule fonction inconnue du temps, $y(t)$, avec une condition initiale y_0 donnée :

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(y,t) \quad y(0) = y_0 \tag{2.1}$$

Nous commencerons avec le schéma le plus simple : Euler explicite. Puis nous généralserons aux systèmes d'équations couplées et introduirons progressivement des schémas numériques plus sophistiqués.

2.1 Schéma d'Euler explicite

Nous introduisons dans cette section un des schémas numériques les plus simples : la méthode d'Euler explicite (appelée parfois *Euler progressive*). Nous partirons d'un exemple physique simple.

2.1.1 Exemple : force de visosité

Soit un corps, point matériel de masse m, soumis à une force de viscosité $\vec{F}_{\text{visc}} = -\kappa\vec{v}$, avec κ un coefficient constant. En restreignant le mouvement à une dimension d'espace,

on a donc, de la deuxième loi de Newton, en posant $\gamma = \kappa/m$:

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -\gamma v \tag{2.2}$$

avec la condition initiale

$$v(0) = v_0 \ . \tag{2.3}$$

On trouve facilement la soution exacte de (2.2)-(2.1.1) :

$$v(t) = v_0 \, e^{-\gamma t} \ . \tag{2.4}$$

La solution numérique s'obtient sur un ensemble discret de valeurs du temps, $\{t_n\}_{i=0}^{\mathrm{nsteps}}$. Nous posons, pour simplifier, que ces valeurs discrètes sont équidistantes, avec $t_{n+1} - t_n = \Delta t, \forall n$.

Notre point de départ est le développement en série de Taylor de la fonction $v(t)$, Eq.(1.4) avec $f = v$ et $h = \Delta t$:

$$v(t + \Delta t) = v(t) + \frac{\mathrm{d}v}{\mathrm{d}t}(t)\Delta t + \frac{1}{2}\frac{\mathrm{d}^2 v}{\mathrm{d}t^2}(t)(\Delta t)^2 + \mathcal{O}(\Delta t)^3 \ . \tag{2.5}$$

Le schéma numérique le plus simple s'obtient en *négligeant les termes d'ordre supérieur à 1 en Δt :*

$$v(t + \Delta t) \approx v(t) + \frac{\mathrm{d}v}{\mathrm{d}t}(t)\Delta t \ . \tag{2.6}$$

Autrement dit, en écrivant cette relation pour les temps discrétisés $\{t_n\}$, on a

$$\boxed{v(t_{n+1}) \approx v(t_n) + \frac{\mathrm{d}v}{\mathrm{d}t}(t_n)\Delta t \ .} \tag{2.7}$$

En substituant avec l'équation différentielle (2.2), on obtient :

$$\boxed{v(t_{n+1}) \approx v(t_n) - \gamma v(t_n)\Delta t} \ . \tag{2.8}$$

Cette approche pour calculer $v(t)$ s'appelle **la méthode d'Euler explicite**. Le qualificatif "explicite" vient du fait que l'on obtient une expression pour l'état du système au temps t_{n+1} explicitement en fonction de l'état au pas de temps précédent t_n, supposé connu. Nous verrons plus loin, à la section 2.5 une méthode *implicite*, et à la section 2.11.2 une méthode *semi-implicite*.

Pour que le schéma fonctionne, il faut l'initialiser avec la valeur $v(t_0) = v_0$ donnée. En détaillant, le schéma d'Euler explicite consiste donc :

1. déclaration des variables nécessaires, tableaux, etc
2. initialisation : lire la valeur de γ, celle de v_0, celles du temps de début et de fin de la simulation et de la taille du pas temporel Δt
3. calcul de valeurs auxiliaires : nombre de pas temporels nsteps

4. boucle temporelle : obtenir la valeur de $v_{n+1} = v_n - \gamma * v_n * \Delta t$
5. $n \to n+1$
6. répéter la boucle temporelle tant que $n <$ nsteps
7. imprimer et faire un graphique du résultat
8. diagnostic de la solution : comparer si possible avec la solution exacte

On peut utiliser le schéma d'Euler explicite pour intégrer d'autres équations différentielles que celles décrivant le mouvement d'un corps soumis à une force visqueuse. On peut généraliser à toute équation de la forme

$$\boxed{\frac{\mathrm{d}y}{\mathrm{d}t} = f(y,t)}, \tag{2.9}$$

avec f une fonction donnée de deux variables. Il suffit, dans le schéma d'Euler explicite, de substituer le point 4 ci-dessus par

$$\boxed{y_{n+1} = y_n + f(y_n, t_n)\Delta t} \tag{2.10}$$

et d'écrire la fonction $f(y,t)$.

La solution numérique obtenue avec (2.10) n'est qu'une **approximation** de la solution exacte. Un des problèmes auxquels nous serons régulièrement confrontés est de déterminer à quel point la solution numérique est précise, voire même si elle a un sens physique : la solution numérique ne va-t-elle pas se "noyer" par une accumulation d'erreurs faites à chaque pas temporel ?

Un bon moyen est d'effectuer des tests de **convergence numérique** : choisissant des pas temporels Δt de plus en plus petits, on examine quelle est la valeur de la solution numérique, à un instant t donné, en fonction de Δt. On devrait pouvoir ensuite effectuer une extrapolation des résultats dans la limite $\Delta t \to 0$.

Dans les cas où on dispose d'une solution exacte, on peut calculer l'erreur de façon précise, et vérifier si l'erreur converge bien vers zéro. On note cependant qu' à cause des erreurs d'arrondi dues à la représentation avec un nombre fini de bits des nombres réels, on ne peut pas excéder la *précision machine*; et, dans certains cas, ces erreurs peuvent faire **diverger** la solution !

Le schéma d'Euler explicite (2.10) est dit *d'ordre 1*, parce qu'il provient d'un développement limité d'ordre 1 en Δt, les termes en $(\Delta t)^2$ ayant été négligés. A chaque pas de temps, on fait une erreur proportionnelle à $(\Delta t)^2$. Pour simuler jusqu'à un temps final donné t_{fin}, on doit faire un nombre de pas de temps inversément proportionnel à Δt, $N_{\text{steps}} = t_{\text{fin}}/\Delta t$. L'erreur au temps final est le résultat de l'accumulation des erreurs faite à chaque pas de temps et est donc proportionnelle à $N_{\text{steps}}(\Delta t)^2 \sim (\Delta t)^1$.

L'application du schéma d'Euler explicite au problème du corps soumis à une force de viscosité, Eq.(2.2), donne les résultats de la Fig. 2.1. On voit clairement la différence entre

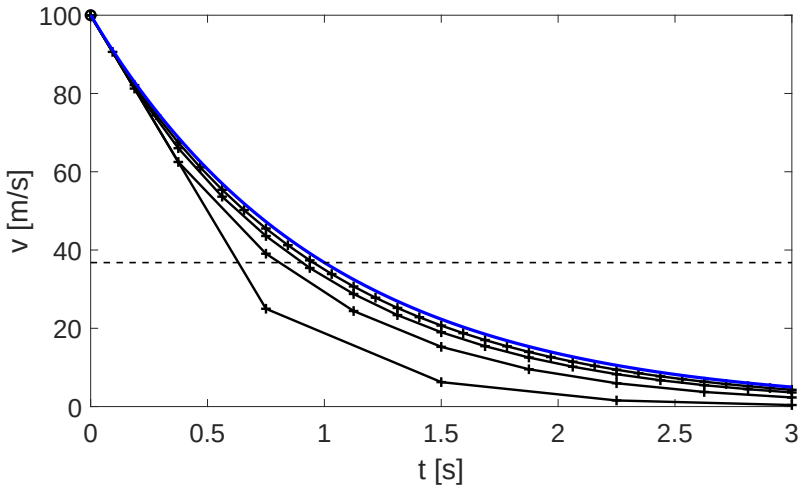


FIGURE 2.1 – *Vitesse d'un corps soumis à une force de viscosité ($\gamma = 1$), calculée avec le schéma d'Euler explicite et pour différentes valeurs du pas temporel (lignes avec croix). La solution exacte est la ligne bleue.*

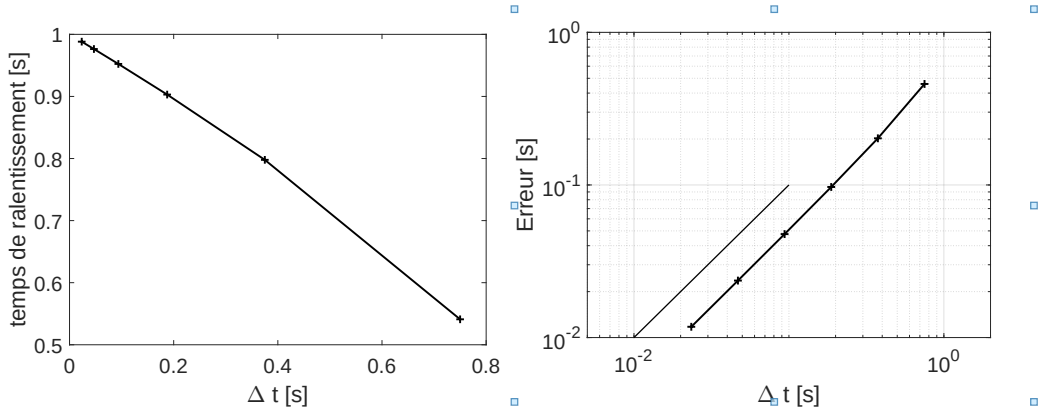


FIGURE 2.2 – *Etude de convergence du temps caractéristique de ralentissement pour le cas de la Fig. 2.1. A gauche, résultats en fonction de Δt sur des échelles linéaires. A droite, valeur absolue de l'erreur en fonction de Δt sur des échelles logarithmiques. La ligne mince est de pente 1. La solution numérique converge bien vers la valeur analytique exacte, et l'ordre de convergence est 1.*

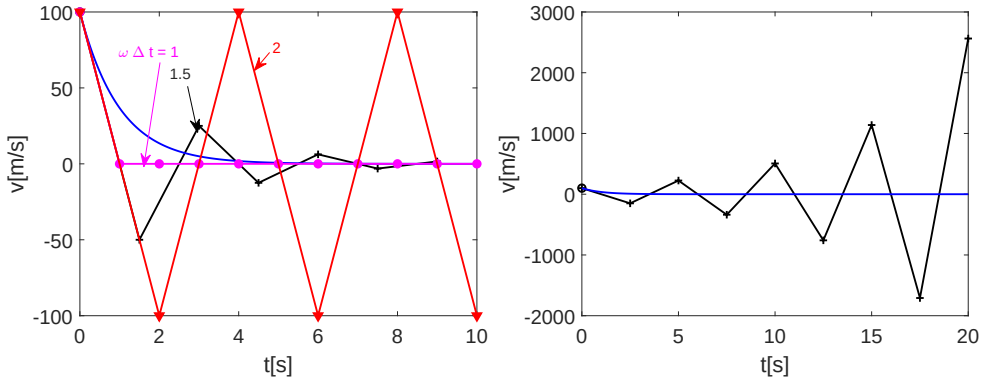


FIGURE 2.3 – *Schéma d'Euler explicite pour un corps soumis à une force de viscosité pour de grandes valeurs de* $\gamma\Delta t$. *A gauche : pour* $\gamma\Delta t = 1.0, 1.5, 20.0$. *A droite pour* $\gamma\Delta t = 2.5$.

la solution exacte et la solution numérique. Fort heureusement, la solution numérique **converge** vers la solution exacte lorsque le pas temporel Δt est choisi de plus en plus petit. On représente à la Fig. 2.2 deux façons de faire une étude de convergence. Comme quantité pour laquelle nous vérifions la convergence, nous avons ici choisi le temps caractéristique de ralentissement, défini comme le temps pour que la vitesse décroisse d'un facteur $1/e$ (v_0/e est représenté par la ligne horizontale traitillée de la Fig. 2.1). La première méthode consiste à reporter la quantité voulue en fonction de Δt (Fig. 2.2, à gauche), sur des échelles linéaires. Dans la limite $\Delta t \to 0$, les résultats numériques s'alignent bien, ce qui indique une convergence d'ordre 1 (parce que l'axe des x est $(\Delta t)^1$). La deuxième méthode, qui s'applique ici car on connaît la solution exacte, consiste à représenter la valeur absolue de l'erreur sur le résultat en fonction de Δt, sur des échelles logarithmiques (Fig. 2.2, à droite). La pente du graphe est 1, ce qui indique une convergence d'ordre 1. Ainsi, l'ordre de convergence de 1, tel que prédit par la théorie, est bien vérifié par nos simulations numériques.

Après avoir examiné le comportement du schéma numérique dans la limite des petits Δt (convergence), explorons ce qui se passe pour les grandes valeurs de Δt. Les résultats sont montrés à la Fig. 2.3. Pour $\Delta t = 1/\gamma$, le schéma d'Euler explicite donne la solution nulle après un pas temporel. Pour $1/\gamma < \Delta t < 2/\gamma$, la solution numérique oscille autour de zéro ; elle tend bien vers zéro pour $t \to \infty$, mais comme elle a $v < 0$ à certains pas de temps, on rejette cete solution comme étant **non physique** : une force de viscosité ne peut jamais, à elle seule, inverser le sens de la vitesse. Pour $\Delta t > 2/\gamma$, la solution numérique oscille avec une amplitude qui croît exponentiellement : il y a **instabilité numérique**.

En résumé :

- $\lim \gamma\Delta t \to 0$, la solution numérique converge vers la solution exacte.
- $\gamma\Delta t = 1$, solution nulle après 1 pas temporel.

• $1 < \gamma\Delta t$, solution non physique car présentant des inversions du sens de la vitesse.
• $2 < \gamma\Delta t$, solution numériquement instable : erreur d'amplitude croissant exponentiellement dans le temps.

Ainsi, il est nécessaire de choisir le pas temporel Δt plus petit que le temps caractéristique de l'équation considérée $1/\gamma$: $\boxed{\gamma\Delta t << 1}$.

2.1.2 Généralisation à un système d'équations couplées

Le schéma d'Euler explicite se généralise aisément aux systèmes d'équations différentielles ordinaires couplées. Notant l'ensemble de N_f fonctions du temps $\{y^{(j)}(t)\}_{j=1}^{N_f}$ par un vecteur de fonctions $\mathbf{y}(t)$, et l'ensemble de N_f fonctions à $N_f + 1$ variables (fonctions des $y^{(j)}$ et du temps t) par un vecteur $\mathbf{f}(\mathbf{y}, t)$, on écrit le système d'équations différentielles ordinaires couplées comme

$$\boxed{\frac{\mathrm{d}y^{(j)}}{\mathrm{d}t} = f^{(j)}(y^{(1)}, ..., y^{(N_\mathrm{f})}, t),\ j = 1..N_\mathrm{f}} \Leftrightarrow \boxed{\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = \mathbf{f}(\mathbf{y}, t)}\,. \tag{2.11}$$

Le schéma numérique d'Euler explicite s'écrit

$$\boxed{y_{n+1}^{(j)} = y_n^{(j)} + f^{(j)}(y_n^{(1)}, ..., y_n^{(N_\mathrm{f})}, t_n)\,\Delta t} \Leftrightarrow \boxed{\mathbf{y}_{n+1} = \mathbf{y}_n + \mathbf{f}(\mathbf{y}_n, t_n)\,\Delta t}\,. \tag{2.12}$$

On a distingué, dans les notations, les indices (subscripts), qui indiquent le numéro du pas temporel, des exposants (superscipts), qui indiquent le numéro de la fonction.

Application : système à trois niveaux

Nous allons voir un des éléments constituant le LASER. Soit un ensemble d'atomes identiques, dont on considère trois niveaux d'énergie. Depuis le niveau fondamental no.1, les atomes ont une probabilité de transiter vers un état excité no.2, par exemple parce qu'ils sont illuminés par des photons. Voir la FIG. 2.4.

Le niveau excité no.2 se désintègre spontanément en un niveau intermédiaire no.3. Celui-ci de désintègre dans le niveau fondamental. On note la probabilité par unité de temps de chacune de ces transitions par $\gamma_1, \gamma_2, \gamma_3$. La population d'atomes dans chacun de ces niveaux, notée N_1, N_2, N_3, est donc décrite par :

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = \gamma_3 N_3 - \gamma_1 N_1 \tag{2.13}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = \gamma_1 N_1 - \gamma_2 N_2 \tag{2.14}$$

$$\frac{\mathrm{d}N_3}{\mathrm{d}t} = \gamma_2 N_2 - \gamma_3 N_3 \tag{2.15}$$

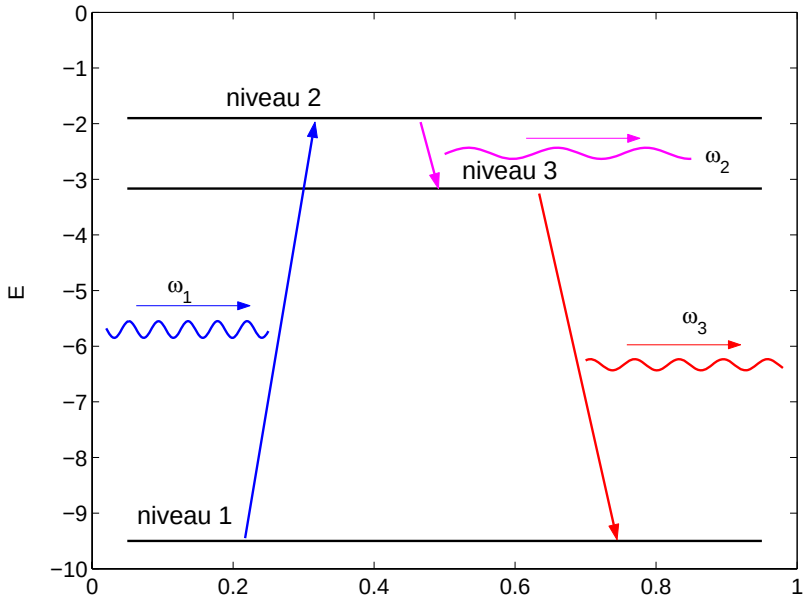


FIGURE 2.4 – *Laser à trois niveaux. Schéma de principe. Absorption ("pompe") de photons d'énergie $\hbar\omega_1 = E_2 - E_1$, avec probabilité γ_1 ; émission spontanée de photons d'énergie $\hbar\omega_2 = E_2 - E_3$, avec probabilité γ_2 ; émission induite de photons d'énergie $\hbar\omega_3 = E_3 - E_1$ avec probabilité γ_3. Inversion de population : il y a plus d'atomes dans le niveau 3 que dans le niveau 1 si la "pompe" est d'intensité suffisante, si le niveau E_2 est instable (donc temps de vie court), et si le niveau E_3 est métastable.*

Suggestion d'exercice Résoudre ce problème. La figure 2.5 montre un résultat pour $\gamma_1 = 1$, $\gamma_2 = 10/3$, $\gamma_3 = 1/4$, avec $\Delta t = 0.05$, à partir d'une condition initiale où tous les atomes sont dans l'état fondamental. Le niveau 1 se dépeuple au profit du no.2, qui atteint un peuplement transitoire maximal, puis se dépeuple au profit du no.3. On analysera le comportement asymptotique ($t \to \infty$) et on comparera avec le calcul analytique. On vérifiera également que le nombre total d'atomes est conservé.

Un résultat physique intéressant est qu'on observe une *inversion de population* : pour des temps longs, le nombre d'atomes dans l'état excité no.3 est supérieur à celui dans l'état fondamental (no.1). Un tel processus joue un rôle important dans les lasers.

2.2 Désintégration. Modélisation statistique (Monte Carlo).

On observe un processus de *désintégration* dans de nombreux systèmes physiques. Par exemple, de nombreux noyaux atomiques sont instables. Ou encore, les niveaux d'énergie supérieurs des atomes sont généralement instables, à cause du couplage avec le champ électromagnétique : un atome, dans un niveau d'énergie dit excité, se relaxe spontanément

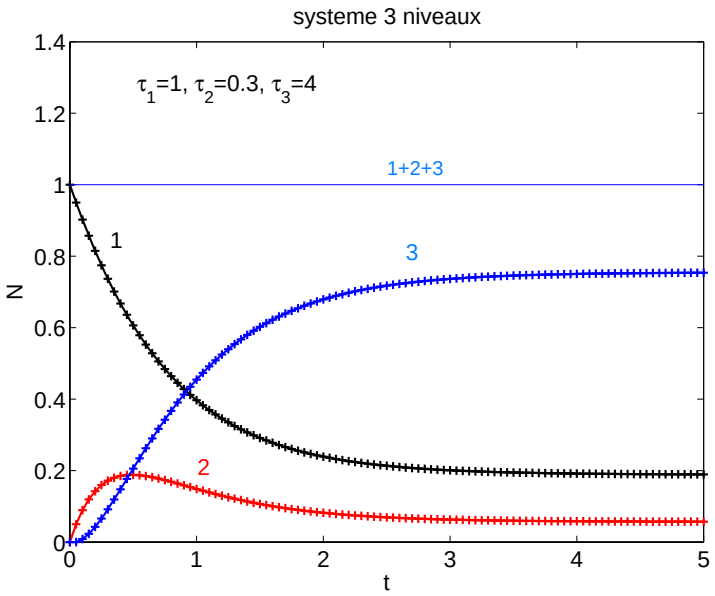


FIGURE 2.5 – *Populations dans 3 niveaux en fonction du temps, calculées avec le schéma d'Euler explicite pour* $\Delta t = 0.05$.

en un niveau d'énergie inférieur, tout en émettant un photon dont l'énergie est égale à la différence des niveaux d'énergie (initial - final).

Le processus semble aléatoire, dans le sens qu'il est impossible, en ne considérant qu'**un** noyau atomique, de savoir exactement quand il va se désintégrer. On ne peut prédire que la *probabilité* qu'un noyau donné se désintègre pendant un intervalle de temps donné. Cette probabilité est *constante* au cours du temps.

Donc, en considérant un grand nombre de noyaux instables (ou d'atomes excités), le nombre de désintégrations par unité de temps est proportionnel au nombre de noyaux instables (ou atomes excités) non encore désintégrés. La constante de proportionnalité est le *taux* de désintégration γ. On exprime ceci mathématiquement (en faisant la limite statistique d'un très grand nombre) par :

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\gamma N \tag{2.16}$$

avec la condition initiale

$$N(0) = N_0 \tag{2.17}$$

donnée. On définit un "temps de vie", ou "constante de temps" $\tau = 1/\gamma$. C'est la mécanique quantique, et la physique atomique ou nucléaire, qui permet, en principe, de calculer la valeur de γ pour un niveau d'énergie atomique ou un noyau donné. Nous ne ferons pas ce calcul ici, mais supposerons γ donné. Le but est de résoudre l'équation ci-dessus donnant l'évolution temporelle du nombre d'atomes ou de noyaux. On trouve facilement la solution de (2.16)-(2.17) :

$$N(t) = N_0 e^{-\gamma t} \ . \tag{2.18}$$

On a donc une équation de même nature que celle pour le corps soumis à une force de viscosité, et on peut donc résoudre le problème avec le schéma d'Euler explicite, comme à la section précédente.

On présente ci-dessous une autre approche. Le processus de désintégration étant intrinsèquement *aléatoire*, on est tenté d'essayer de reproduire cette caractéristique numériquement. Ci-dessus, on a construit, à partir d'une *nature discrète et aléatoire* (les désintégrations ont lieu de façon aléatoire, soudaine et spontanée), un *modèle continu et déterministe* (voir l'Eq.2.16) : connaissant le nombre de particules à un instant donné, Eq.2.17, on connaît exactement, pour tous les temps ultérieurs, le nombre restant de particules : la solution du problème mathématique existe et est unique. Or, dans la réalité, deux échantillons identiques (par example de matière radioactive) ne vont jamais donner exactement le même $N(t)$. Il y a une certaine dispersion statistique des résultats. C'est ce que nous aimerions obtenir par un calcul numérique.

On notera au passage que la résolution numérique de ce modèle continu que nous avons faite à la Section 2.1 faisait appel à des équations *discrètes* (et déterministes). Cependant, la nature discrète de ce type de méthodes numériques n'a rien à voir avec la nature discrète du phénomène physique de la désintégration.

L'idée est de *simuler* la réalité. En outre, cela nous permettra de comprendre comment le modèle continu et déterministe peut être obtenu comme une limite du modèle intrinsèquement discret et aléatoire. En d'autres termes, on aura obtenu une autre méthode pour résoudre l'Eq.(2.16). Dans le processus de désintégration, on ne peut connaître que la *probabilité* par unité de temps qu'une particule se désintègre. Cette probabilité est :

- constante au cours du temps,
- identique pour chaque particule du même type,
- indépendante de la désintégration ou non des autres particules.

On construit le modèle numérique directement à partir de là.

1. Initialisation : nombre de particules en $t = 0$ (N_0), probabilité par unité de temps (γ), choix d'un pas temporel (Δt), probabilité par intervalle temporel (P).
2. Boucle temporelle
3. Boucle sur les particules non encore désintégrées
4. Pour chaque particule non encore désintégrée et pour chaque intervalle de temps, on choisit un nombre aléatoire entre 0 et 1 selon une distribution uniforme.
5. Si ce nombre est inférieur à P, on diminue d'une unité le nombre de particules.
6. Fin de la boucle sur les particules
7. Fin de la boucle temporelle
8. Impression des résultats et comparaison avec le modèle continu

Ce type de méthode nécessite un générateur de nombres aléatoires. En fait, comme les algorithmes sont de nature intrinsèquement déterministe, on parle en fait de générateur *pseudo-aléatoire*. Il n'est en fait pas si facile qu'il n'y paraît de prime abord de construire

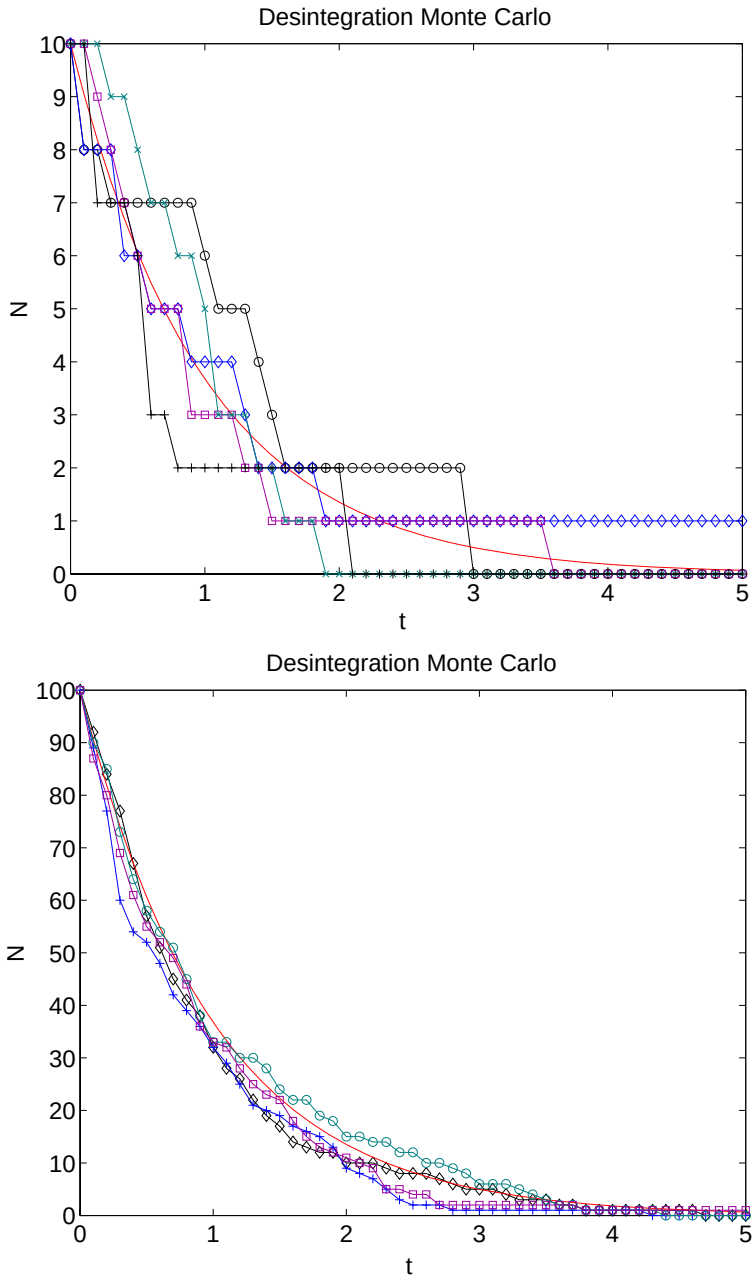


FIGURE 2.6 – *Désintégration d'une population de particules instables, en fonction du temps, calculé avec le modèle numérique probabiliste présenté à la Section 2.2.* $N_0 = 10$ *(haut),* $N_0 = 100$ *(bas),* $\Delta t = 0.1$, $\gamma = 1$. *Pour chaque* N_0, *quatre éxécutions du code sont représentées.*

un "bon" générateur pseudo-aléatoire. Nous laissons ce sujet en dehors du champ de ce cours, et nous bornerons à utiliser de tels générateurs issus de librairies.

Les résultats de ce modèle sont présentés à la FIG. 2.6, pour $\gamma = 1$ et différentes valeurs de N_0. Les résultats de chaque simulation sont différents : ceci est dû au tirage du nombre aléatoire dans l'algorithme. Ceci reproduit bien l'expérience : 4 échantillons de 10 particules instables ne se désintègrent jamais exactement de la même façon.

Pour une taille de l'échantillon initial de 10 (FIG. 2.6, haut), on note en particulier une grande disparité des résultats pour N au temps de vie caractéristique $\tau = 1/\gamma$ (=1 dans ce cas) : entre 2 et 6 particules. La moyenne de ces prédictions est 4, ce qui s'approche du résultat de la solution exacte du modèle continu.

Ceci suggère la façon dont les prédictions du modèle discret aléatoire probabiliste vont tendre vers le résultat du modèle continu déterministe. Augmentant la taille de l'échantillon initial à $N_0 = 100$, on obtient les résultats de la FIG. 2.6 (bas). On constate que les écarts entre les 4 échantillons pour le nombre de particules *relatif*, $N(t)/N_0$, sont nettement plus faibles que pour les simulations avec une taille de $N_0 = 10$ de la FIG. 2.6.

On peut aussi considérer que les simulations avec un N_0 donné correspondent à un échantillonage d'un système physique réel contenant un grand nombre de particules. Plus N_0 est élevé, meilleur est l'échantillonage, et plus petite est la dispersion statistique des résultats. On peut montrer que cette dispersion statistique σ tend vers zéro comme

$$\sigma \sim \frac{1}{\sqrt{N}} \, . \tag{2.19}$$

Ces simulations sont aussi discrétisées dans le temps, et on peut vérifier (suggestion d'exercice) que les résultats des simulations convergent vers la solution exacte du modèle continu lorsque $N_0 \to \infty$ **et** $\Delta t \to 0$. Il faut effectuer une double convergence.

Ainsi, on peut considérer cet algorithme comme une façon de résoudre numériquement l'Eq.(2.16). Ce type de méthode, faisant appel à un échantillonage statistique, est souvent appelé **Monte Carlo**. On y aura recours à la Section 4.1.4 et au Chapitre 5.

2.3 Applications du schéma d'Euler explicite

Le schéma numérique très simple (Euler explicite) présenté à la Section 2.1.2 permet déjà de résoudre de nombreux problèmes physiques que l'on ne peut pas résoudre analytiquement, ou du moins très difficilement.

Un exemple typique est l'étude de la dynamique Newtonienne (mécanique classique),

où de nombreux problèmes commencent par la phrase "on négligera l'effet des forces de frottement". Ou alors, il s'agit de problèmes que l' on ne peut résoudre que dans certaines limites (du style chute d'un corps dans la limite $t \to \infty$, en régime stationnaire, etc).

Mais la dynamique Newtonienne est bien plus riche que ne le laisseraient supposer les quelques problèmes que l'on sait résoudre analytiquement. De plus, toute théorie devant être confrontée à l'expérience, on aimerait disposer d'un outil permettant cette confrontation dans des situations réalistes.

On observera toujours un certain écart entre valeurs théoriques et mesures expérimentales. Il est important de déterminer quelle est la part de cet écart qui est due aux imprécisions de mesure de la part qui est due aux effets que l'on a négligés dans la résolution des équations... C'est une étape indispensable pour la validation d'une théorie ou d'un modèle.

2.3.1 Véhicule avec force de traînée aérodynamaique

Une voiture de masse m a un moteur de puissance maximale $P_{\max}$ et un couple maximal donnant une force de poussée maximale $F_{\max}$. [Dans la réalité cette puissance et cette force sont fonction du nombre de tours/minute du moteur et du rapport de transmission. Ici, pour simplifier, on supposera $F_{\max}$ et $P_{\max}$ constantes.] On aimerait calculer la vitesse au cours du temps pour un départ arrêté, sur une route horizontale. On tiendra compte de la force de traînée aérodynamique, avec un coefficient C_x (que l'on suposera constant, pour simplifier)

$$\vec{F}_t = -\frac{1}{2}\rho S C_x v^2 \vec{e}_v \tag{2.20}$$

avec S une surface effective de la voiture, ρ la densité de l'air, et $\vec{e}_v = \vec{v}/|\vec{v}|$.

L'équation du mouvement pour $v(t)$ est donnée par le théorème de l'énergie cinétique

$$\frac{\mathrm{d}E_{\mathrm{cin}}}{\mathrm{d}t} = P_{\max} - F_t v \tag{2.21}$$

$$\Rightarrow mv\frac{\mathrm{d}v}{\mathrm{d}t} = P_{\max} - F_t v \tag{2.22}$$

$$\Rightarrow \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{P_{\max}}{mv} - \frac{F_t}{m}\,. \tag{2.23}$$

Sans force de traînée, et sans tenir compte de la limite de la force de poussée $F_{\max}$, la solution exacte donne

$$v(t) = \sqrt{v_0^2 + 2P_{\max}t/m}\,, \tag{2.24}$$

où v_0 est la vitesse en $t = 0$. Cette solution, pour un départ arrêté ($v_0 = 0$), donne une accélération infinie en $t = 0$. Ce n'est pas physique. Il faut tenir compte de la limite de

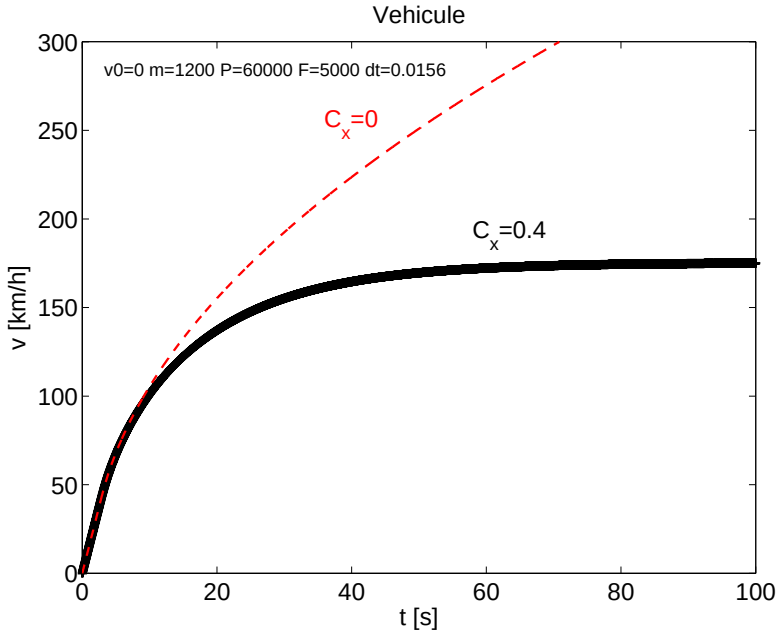


FIGURE 2.7 – *Véhicule en accélération départ arrêté, en tenant compte de la force de traînée aérodynamique. Le cas sans force de traînée est en traitillés. Paramètres :* $m = 1200\text{kg}$, $P_{\max} = 60\text{kW}$, $F_{\max} = 5000\text{N}$, $C_x = 0.4$, $S = 2\text{m}^2$, *schéma d'Euler explicite avec* $\Delta t = 0.015625\text{s}$.

la force de poussée $F_{\max}$. L'équation à résoudre est finalement :

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{1}{m}\left(\min\{\frac{P_{\max}}{v}, F_{\max}\} - F_t\right) \ . \tag{2.25}$$

On applique le schéma d'Euler explicite pour résoudre ce problème.

Un exemple est donné à la FIG. 2.7, avec pour paramètres $m = 1200\text{kg}$, $P_{\max} = 60\text{kW}$, $F_{\max} = 5000\text{N}$, $C_x = 0.4$. Pour comparaison, le cas sans effet de traînée aérodynamique est représenté en traitillés.

Suggestion d'exercice. Calculer le temps pour une accélération de 0 à 100 km/h et tester la convergence numérique. On peut obtenir aussi la limite asymptotique $\lim_{t\to\infty} v(t)$ analytiquement et la comparer aux résultats numériques. On calculera le même problème, mais pour une route en pente. On peut aussi regarder ce qui se passe lorsque le conducteur coupe le moteur à partir d'une vitesse initiale $v_0 \neq 0$, quelle est la vitesse limite en fonction de la pente, etc.

2.3.2 Rentrée dans l'atmosphère

La chute des corps au voisinage de la surface terrestre est fortement influencée par la présence de l'atmosphère. On verra dans cet exemple à quel point l'atmosphère joue un

rôle de "bouclier protecteur" contre les objets célestes (météorites, etc) attirés par la gravitation terrestre.

On considère une météorite de masse m, densité $\rho_M = 5000\text{kg/m}^3$, arrivant " de l'infini" à proximité de la terre. Elle a une vitesse v_0 de 11 km/s verticale lorsque son altitude est $z_0 = 200\text{km}$. On aimerait connaître quelle sera la vitesse d'impact au sol. On tiendra compte de l'atmosphère, avec une densité $\rho(x) = \rho_0\, e^{-x/\lambda}$, $\rho_0 = 1.3\text{kg/m}^3$ et une épaisseur caractéristique $\lambda = 20\text{km}$. On supposera la météorite de forme sphérique et un $C_x = 0.3$ constant donné.

[N.B. : d'où vient cette dépendance de la densité exponentiellement décroissante avec l'altitude ? La valeur de $\lambda = 20\text{km}$ est/elle réaliste pour l'atmosphère terrestre ?].

Suggestion d'exercice. On appliquera le schéma d'Euler explicite à ce problème. On remarque que les équations de base sont semblables à celles de la voiture, mais avec la gravitation en plus et la puissance du moteur en moins. Un résultat est montré à la FIG. 2.8.

Suggestion d'exercice. On étudiera la convergence numérique des résultats avec Δt, et on fera une étude de la vitesse d'impact en fonction de la masse de la météorite. On calculera la puissance de la force de traînée en fonction du temps.

Dans la réalité, la puissance de cette force de traînée est convertie en chaleur. Une partie de cette chaleur chauffe l'atmosphère, à des températures telles que l'air devient partiellement ionisé (plasma) [Dans un tel plasma les ondes RF utilisées pour la communication avec les astronautes ne se propagent plus : c'est la raison du "blackout" observé lors des rentrées dans l'atmosphère des astronautes]. Une autre partie de cette chaleur chauffe le météorite [ou l'engin spatial... cause de la catastrophe d'Atlantis] et le sublime : c'est une bonne nouvelle pour nous s'il s'agit d'une météorite, mais une difficulté pour le design des vaisseaux spatiaux [bouclier thermique, tuiles céramiques, etc]. Une simulation plus réaliste tiendrait compte de cet effet *d'ablation*, la masse de la météorite se réduisant lors de sa chute, avec émission de gaz très chauds, qui forment la traînée lumineuse des "étoiles filantes".

On peut aussi étudier ce qui se passe avec un projectile lancé depuis le sol, avec une vitesse initiale vers le haut ; on comparera avec ce qui se passerait s'il n'y avait pas d'atmosphère.

Il est aisé de généraliser ces problèmes au cas où la vitesse n'est pas purement verticale

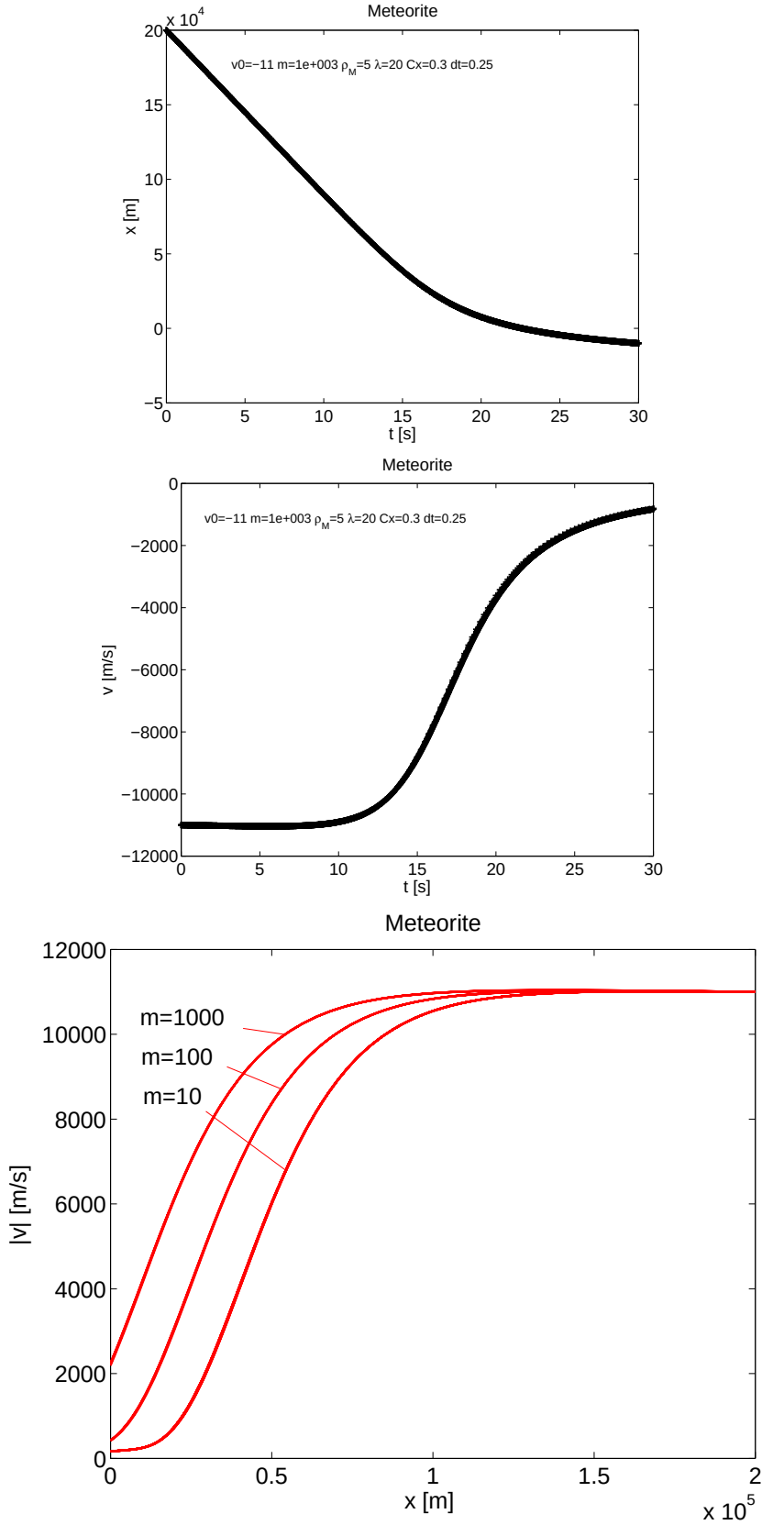


FIGURE 2.8 – *Impact d'une météorite de densité* $5 \times 10^3 \mathrm{kg/m^3}$*, tenant compte de la force de frottement de l'air de l'atmosphère terrestre. Méthode d'Euler explicite. Position (haut) et vitesse (milieu) en fonction du temps, pour une masse* $m = 1000\mathrm{kg}$*. Vitesse en fonction de l'altitude (bas), pour 3 météorites de masses* $m = 10, 100$ *et* $1000\mathrm{kg}$*.*

(v_z), mais a aussi des composantes horizontales v_x, v_y. On définit le "vecteur" de fonctions

$$\mathbf{y} = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \\ v_x(t) \\ v_y(t) \\ v_z(t) \end{pmatrix} . \tag{2.26}$$

On a

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = \begin{pmatrix} v_x \\ v_y \\ v_z \\ -\frac{1}{2m}\rho S C_x v v_x \\ -\frac{1}{2m}\rho S C_x v v_y \\ -\frac{1}{2m}\rho S C_x v v_z - g \end{pmatrix} \tag{2.27}$$

avec $v = |\vec{v}|$.

2.3.3 Balistique avec rotation : portance, effet Magnus

Il s'agit de tenir compte d'une force de portance aérodynamique qui s'exerce sur les corps en mouvement combiné de translation et de rotation dans les fluides. Un corps en rotation dans un fluide va entraîner l'air dans son voisinage de telle sorte que, combinées à la vitesse de translation, les vitesses résultantes du fluide soient différentes d'un côté et de l'autre du corps. L'application de l'équation de Bernoulli donne une force résultante perpendiculaire au vecteur vitesse de rotation et perpendiculaire à la vitesse de translation. C'est *l'effet Magnus*. Voir Cours de Physique III. On obtient une résultante

$$\vec{F}_p = \rho v_t L C \vec{e}_\omega \times \vec{e}_v \tag{2.28}$$

avec ρ la densité du fluide, v la vitesse de translation, L la dimension transversale du corps, $\vec{e}_\omega$ la direction de l'axe de rotation du corps et $C = \oint_\Gamma \vec{v} \cdot \vec{dl}$ la circulation de la vitesse autour du corps. La circulation de la vitesse est proportionnelle à la vitesse angulaire de rotation ω, au carré des dimensions linéaires du corps et à la vitesse de translation v. On a donc une force $|\vec{F}_p| \propto \rho v^2 S$, avec $S \sim l^2$, que l'on écrit habituellement sous la forme

$$F_p = \frac{1}{2}\rho S C_y v^2 \tag{2.29}$$

C'est une **force de portance**, avec C_y le *coefficient de portance* et S une surface de référence du corps considéré. Dans le cas du corps en rotation, le C_y est proportionnel à la vitesse de rotation. Le calcul détaillé du C_y n'est pas possible analytiquement, sauf pour des cas très simples, et moyennant un certain nombre d'hypothèses simplificatrices : par exemple un écoulement fluide stationnaire, incompressible, 2D, autour d'un cylindre de rayon R. On trouve $C_y \sim 2\pi\omega R/v$ (Evt. en exercice). Pour un ballon sphérique, il est

extrêmement difficile de le calculer analytiquement. On supposera dans la suite la force de portance de l'effet Magnus donnée par

$$\vec{F}_p = \mu R^3 \rho \vec{\omega} \times \vec{v} \tag{2.30}$$

avec un coefficient (sans dimensions) μ donné.

Cette force de portance est responsable des effets de courbure de la trajectoire de la balle dans de nombreux sports (football, ping-pong, baseball, etc). On se propose ici de calculer numériquement de telles trajectoires.

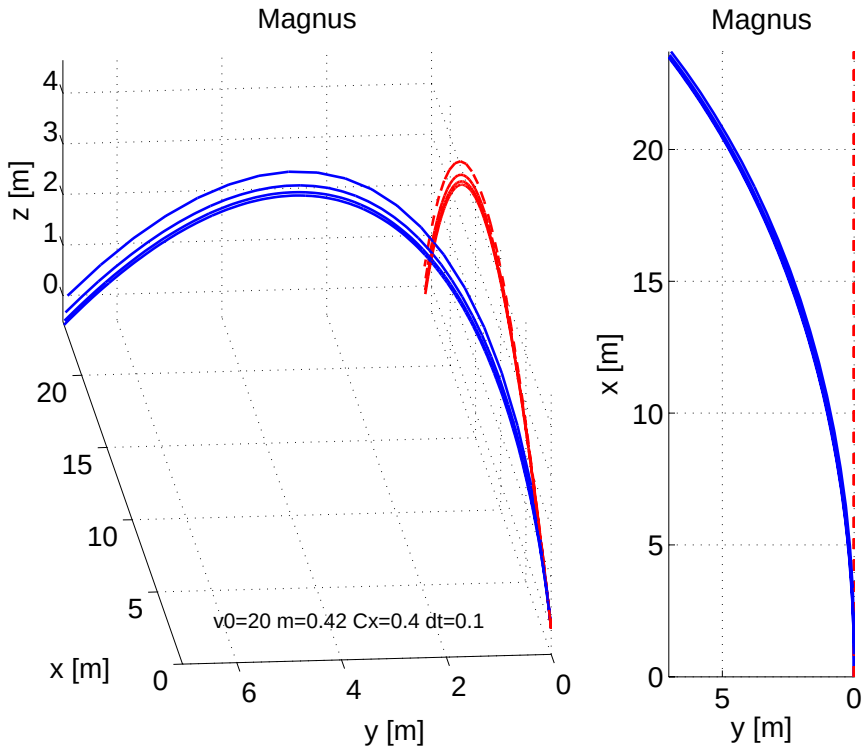


FIGURE 2.9 – *Tir d'un coup franc, avec rotation de la balle, incluant la force de traînée et la force de portance due à l'effet Magnus. Vue en 3D (à gauche), et vue d'en haut (à droite). Méthode d'Euler explicite, 4 exécutions avec* $\Delta t = 0.1, 0.05, 0.025, 0.0125$s. *Pour comparaison, on a représenté en rouge traitillés la trajectoire sans rotation de la balle.*

Pour simplifier, on supposera $\vec{\omega} =$ constant, négligeant ainsi le ralentissement de la vitesse de rotation par l'effet des forces de viscosité. Le vecteur $\vec{\omega}$ fait un angle γ avec la verticale, dans le plan (y, z). On a donc, dans les notations de l'Eq.(2.27),

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = \begin{pmatrix} v_x \\ v_y \\ v_z \\ -\frac{1}{2m}\rho S C_x v v_x + \frac{1}{m}\mu R^3 \rho\omega(\sin\gamma\, v_z - \cos\gamma\, v_y) \\ -\frac{1}{2m}\rho S C_x v v_y + \frac{1}{m}\mu R^3 \rho\omega\cos\gamma\, v_x \\ -\frac{1}{2m}\rho S C_x v v_z - g - \frac{1}{m}\mu R^3 \rho\omega\sin\gamma\, v_x \end{pmatrix}. \tag{2.31}$$

Ceci est implémenté en utilisant l'algorithme d'Euler explicite. Un résultat est montré à la FIG. 2.9. On tire un coup franc au football, avec $|\vec{v}_0| = 20\text{m/s}$, $\vec{v}_0$ faisant un angle $\alpha = 30^o$ avec l'horizontale, et le ballon faisant 2 tours/s autour de l'axe vertical ($\gamma = 0$). Les paramètres sont : $m = 0.42\text{kg}$, $C_x = 0.4$, $\mu = 6.28$, $\rho = 1.3\text{kg/m}^3$, $R = 0.11\text{m}$. Le tir est initialement dans le plan (x, z), mais on voit clairement la déviation du ballon selon y, qui atteint environ 7m à son point de chute.

Suggestion d'exercice. Considérer le tir d'une balle de tennis, avec vecteur vitesse de rotation dans le plan horizontal. Etudier les effets (lift, slice) sur les trajectoires.

2.4 Instabilité numérique - schéma d'Euler explicite - mouvements oscillatoires

Dans cette section, nous allons résoudre des problèmes oscillatoires. Ils sont donnés par une équation différentielle du 2e ordre du type

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = \frac{F(x, v, t)}{m} \ . \tag{2.32}$$

2.4.1 Desciption de l'instabilité numérique - oscillateur harmonique

Dans l'exemple le plus simple du ressort linéaire, on a $F(x, v, t) = F(x) = -kx$, où k est une constante. On obtient alors la solution générale de cette équation :

$$x(t) = A\cos(\omega t + \varphi) \tag{2.33}$$

avec $\omega = \sqrt{k/m}$. A et φ sont des constantes réelles, déterminées par les conditions initiales (position et vitesse en $t = 0$).

Pour résoudre numériquement cette équation, (et ultérieurement avec des forces $F(x, v, t)$ plus compliquées que le ressort linéaire), on commence par la réécrire en définissant un vecteur de fonctions $\mathbf{y}(t)$ dont les composantes sont

$$\mathbf{y}^{(1)}(t) = x(t), \quad \mathbf{y}^{(2)}(t) = \frac{\mathrm{d}x}{\mathrm{d}t}(t) \ . \tag{2.34}$$

Ceci permet d'écrire l'équation du 2e ordre Eq.(2.32) comme un système d'équations différentielles couplées du 1er ordre

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{y} = \mathbf{f}(\mathbf{y}) \tag{2.35}$$

avec

$$\mathbf{f}^{(1)} = \mathbf{y}^{(2)} \tag{2.36}$$
$$\mathbf{f}^{(2)} = F(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, t)/m \tag{2.37}$$

Ceci est donc équivalent à l'Eq.(2.11), écrit ici sous une forme vectorielle. Avec $v(t) = dx(t)/dt$, l'Eq. (2.32) est équivalente au système

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ v \end{pmatrix} = \begin{pmatrix} v \\ F(x,v,t)/m \end{pmatrix} . \tag{2.38}$$

Pour le problème du ressort linéaire, cela revient à

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ v \end{pmatrix} = \begin{pmatrix} v \\ -(k/m)x \end{pmatrix} . \tag{2.39}$$

En utilisant le schéma d'Euler explicite, Eq.(2.12), on obtient le résultat de la FIG. 2.10. Il y a manifestement un problème. La solution est bien proche de la solution exacte pour les temps courts, mais pour les temps longs elle s'en écarte avec une amplitude des oscillations qui croît exponentiellement.

On peut donc imaginer qu'en choisissant un Δt plus petit on va converger vers le bon résultat. Mais le problème subsiste, il est simplement repoussé à des temps ultérieurs : l'amplitude des oscillations finit toujours par croître exponentiellement. Aussi petit soit $\Delta t \neq 0$, il existe donc un temps au delà duquel le calcul numérique s'écarte complètement de la solution physique correcte.

C'est un problème **d'instabilité numérique**.

2.4.2 Analyse de stabilité du schéma d'Euler explicite : propagation de l'erreur

Soit $\mathbf{y}(t)$ la solution exacte de l'équation différentielle (2.35). Soit $\mathbf{y}_n$ la valeur de $\mathbf{y}$ au temps t_n produite par le schéma numérique. Soit $\mathbf{e}_n$ l'erreur, telle que $\mathbf{y}_n = \mathbf{y}(t_n) + \mathbf{e}_n$. Le but du calcul ci-dessous est de déterminer l'erreur au temps $n+1$, en d'autres termes de déterminer comment l'erreur va se "propager".

Le schéma d'Euler explicite est

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \mathbf{f}(\mathbf{y}_n)\Delta t \quad \Rightarrow \tag{2.40}$$
$$\mathbf{y}_{n+1} = \mathbf{y}(t_n) + \mathbf{e}_n + \mathbf{f}(\mathbf{y}(t_n) + \mathbf{e}_n)\Delta t \quad \Rightarrow \tag{2.41}$$
$$\mathbf{y}_{n+1} \approx \mathbf{y}(t_n) + \mathbf{e}_n + \left[\mathbf{f}(\mathbf{y}(t_n)) + \frac{\partial \mathbf{f}}{\partial \mathbf{y}}\mathbf{e}_n\right]\Delta t\,. \tag{2.42}$$

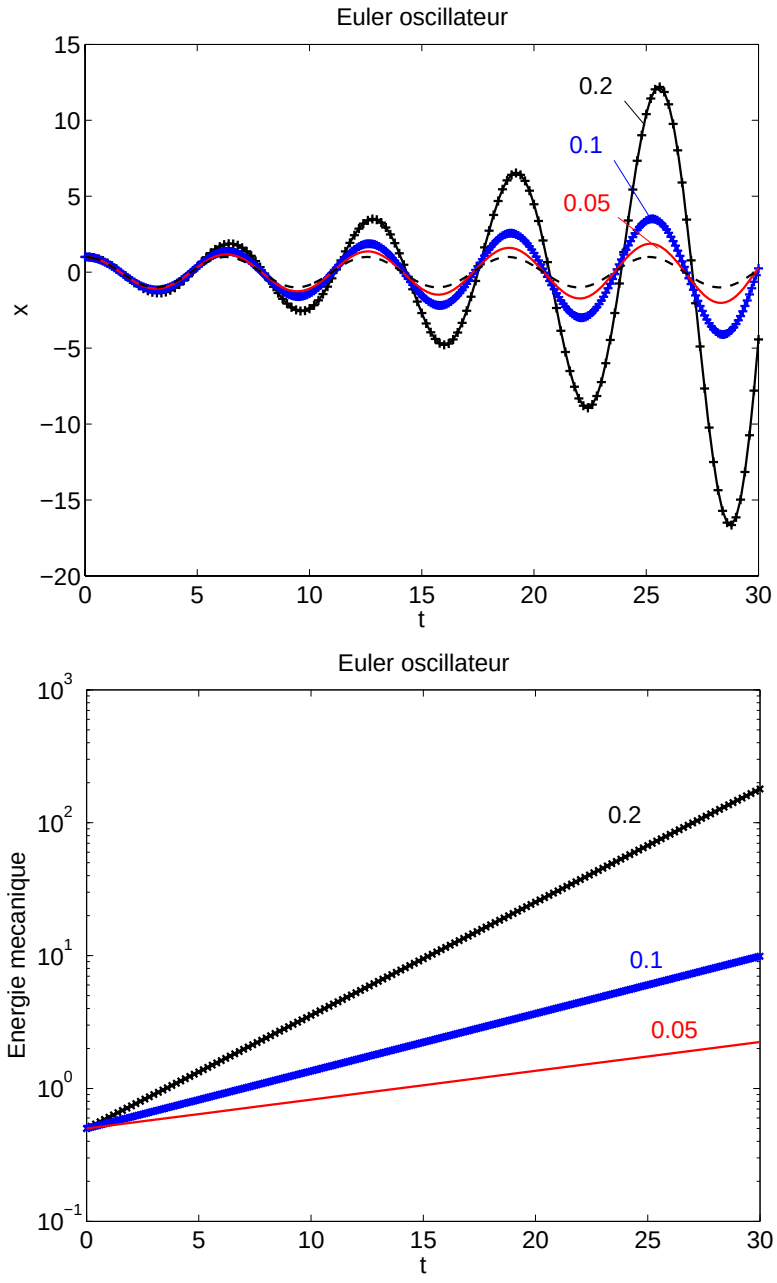


FIGURE 2.10 – *Oscillateur harmonique avec la méthode d'Euler explicite, $k = 1$, $m = 1$. Trois exécutions avec $\Delta t = 0.2, 0.1, 0.05$. Le schéma est instable, avec une croissance exponentielle de l'amplitude des oscillations (haut) et de l'énergie mécanique (bas) (qui devrait être conservée). Le taux de croissance est proportionnel à Δt.*

Avec $\mathbf{y}_{n+1} = \mathbf{y}(t_{n+1}) + \mathbf{e}_{n+1}$ et $\mathbf{y}(t_{n+1}) \approx \mathbf{y}(t_n) + (d\mathbf{y}/dt)\Delta t = \mathbf{y}(t_n) + \mathbf{f}(\mathbf{y}(t_n))\Delta t$, on a

$$\boxed{\mathbf{e}_{n+1} = \left(\mathbf{I} + \Delta t \frac{\partial \mathbf{f}}{\partial \mathbf{y}}\right) \mathbf{e}_n}\,, \tag{2.43}$$

où $\mathbf{I}$ est la matrice identité et $(\partial \mathbf{f}/\partial \mathbf{y})_{ij} = \partial \mathbf{f}_i/\partial \mathbf{y}_j$. En définissant la *matrice de gain* $\mathbf{G}$ telle que

$$\mathbf{e}_{n+1} = \mathbf{G}\mathbf{e}_n \tag{2.44}$$

on a

$$\boxed{\mathbf{G} = \left(\mathbf{I} + \Delta t \frac{\partial \mathbf{f}}{\partial \mathbf{y}}\right)}\,. \tag{2.45}$$

La norme de l'erreur va s'amplifier, et donc le schéma numérique sera **instable** s'il existe une valeur propre de λ_i de $\mathbf{G}$ avec

$$|\lambda_i| > 1\,. \tag{2.46}$$

Réciproquement, le schéma sera **stable** si les valeurs propres λ_i de $\mathbf{G}$ sont telles que

$$|\lambda_i| \leq 1\,,\ \forall i\,. \tag{2.47}$$

Appliquons cette analyse de stabilité au schéma d'Euler explicite dans le cas de l'oscillateur harmonique. Avec $\mathbf{f}_1 = v$ et $\mathbf{f}_2 = -(k/m)x$, on a

$$\mathbf{G} = \begin{pmatrix} 1 & \Delta t \\ -(k/m)\Delta t & 1 \end{pmatrix}\,. \tag{2.48}$$

On a l'équation caractéristique pour les valeurs propres λ_i de $\mathbf{G}$:

$$(1-\lambda)^2 + (k/m)(\Delta t)^2 = 0 \tag{2.49}$$

dont les solutions sont

$$\lambda_{1,2} = 1 \pm i\sqrt{k/m}\Delta t\,. \tag{2.50}$$

On a

$$|\lambda_{1,2}| = \sqrt{1 + (k/m)(\Delta t)^2} \ \Rightarrow\ |\lambda_{1,2}| > 1\,,\ \forall \Delta t\,, \tag{2.51}$$

ce qui veut dire que **le schéma d'Euler explicite est toujours instable pour le problème de l'oscillateur harmonique**.

2.4.3 Analyse de stabilité du schéma d'Euler explicite : solution analytique des équations discrétisées

Dans cette section, on donne une autre analyse du phénomène d'instabilité numérique. Pour des équations linéaires, on peut écrire le systèmes d'équations différentielles de la forme

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{y} = \mathbf{M}\mathbf{y} \tag{2.52}$$

où $\mathbf{M}$ est une matrice $n_f \times n_f$ (n_f est le nombre de fonctions dans le vecteur $\mathbf{y}$). Pour illustration, dans le cas de l'oscillateur harmonique, on a

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ v \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -k/m & 0 \end{pmatrix}\begin{pmatrix} x \\ v \end{pmatrix} . \tag{2.53}$$

Le schéma d'Euler explicite s'écrit alors

$$\mathbf{y}_{n+1} = (\mathbf{I} + \mathbf{M}\Delta t)\mathbf{y}_n . \tag{2.54}$$

Cherchons une solution à ces équations discrétisées du type $\mathbf{y}(t) = \mathbf{A}e^{i\omega t}$, avec $\mathbf{A} \in C^{n_f}$ et $\omega \in C$.[1] [On rappelle que si ω est réel, cela correspond à une oscillation non amortie ; si ω est purement imaginaire, cela correspond à une solution exponentiellement décroissante ou croissante, selon le signe de $\Im(\omega)$; si ω a une partie réelle et une partie imaginaire, cela correspond à une solution oscillante dont l'amplitude croît ou décroît selon le signe de $\Im(\omega)$]. Introduisant cet Ansatz dans le schéma d'Euler explicite ci-dessus, on a

$$\mathbf{A}e^{i\omega t}e^{i\omega\Delta t} = (\mathbf{I} + \mathbf{M}\Delta t)\mathbf{A}e^{i\omega t} \tag{2.55}$$

En supposant

$$\boxed{\omega\Delta t << 1}, \tag{2.56}$$

on fait un développement limité à l'ordre 2 en $\omega\Delta t$:

$$e^{i\omega\Delta t} = 1 + i\omega\Delta t - \frac{1}{2}\omega^2\Delta t^2 + \mathcal{O}((\omega\Delta t)^3) . \tag{2.57}$$

Insérant dans (2.55), on obtient le système d'équations linéaire homogène

$$\left[(i\omega - \frac{1}{2}\omega^2\Delta t)\mathbf{I} - \mathbf{M}\right]\mathbf{A} = 0 \tag{2.58}$$

qui a une solution non triviale si et seulement si son déterminant est nul. Pour le problème de l'oscillateur harmonique, on a

$$\begin{pmatrix} i\omega - \frac{1}{2}\omega^2\Delta t & -1 \\ k/m & i\omega - \frac{1}{2}\omega^2\Delta t \end{pmatrix}\mathbf{A} = 0 . \tag{2.59}$$

A l'ordre 1 en $\omega\Delta t$, on obtient $-\omega^2(1 + i\omega\Delta t) + k/m = 0 \Rightarrow \omega^2 = (k/m)(1 - i\omega\Delta t) \Rightarrow$

$$\boxed{\omega = -i\frac{k}{m}\frac{\Delta t}{2} \pm \sqrt{\frac{k}{m}}}. \tag{2.60}$$

Ainsi

$$\boxed{\begin{pmatrix} x \\ v \end{pmatrix} = \mathbf{A}e^{i\left(\sqrt{k/m}\right)t}e^{(k/m)(\Delta t/2)\,t}}. \tag{2.61}$$

On retrouve bien le résultat numérique de la FIG. 2.10 : une oscillation sinusoïdale de fréquence $\sqrt{k/m}$, avec une amplitude croissant exponentiellement dans le temps, avec un taux de croissance $\gamma = (k/m)(\Delta t/2)$. Cette croissance exponentielle, non-physique, est la signature d'une instabilité numérique.

1. La solution "physique" est la partie réele de cete fonction complexe.

2.4.4 Vérification de la conservation de l'énergie, schéma d'Euler explicite

Dans le cas de l'oscillateur harmonique, on sait que l'énergie mécanique est conservée :

$$E_{\text{mec}}(t) = \frac{1}{2}mv^2(t) + \frac{1}{2}kx^2 = \text{const.} \tag{2.62}$$

Nous allons vérifier cette propriété pour le schéma d'Euler explicite. La méthode est simple : il s'agit d'écrire l'énergie mécanique au temps t_{n+1} et au temps t_n, puis comparer les deux expressions. On a :

$$E_{\text{mec},n+1} = \frac{1}{2}mv_{n+1}^2 + \frac{1}{2}kx_{n+1}^2 \ .$$

On substitue le schéma d'Euler pour v_{n+1} et x_{n+1} :

$$E_{\text{mec},n+1} = \frac{1}{2}m\left(v_n - \frac{k}{m}x_n\Delta t\right)^2 + \frac{1}{2}k\left(x_n + v_n\Delta t\right)^2$$

$$E_{\text{mec},n+1} = \frac{1}{2}mv_n^2 + \frac{1}{2}kx_n^2 - v_nkx_n\Delta t + v_nkx_n\Delta t + \frac{1}{2}\frac{k^2}{m}x_n^2\Delta t^2 + \frac{1}{2}kv_n^2\Delta t^2 \ .$$

On identifie $E_{\text{mec},n}$ et on obtient :

$$E_{\text{mec},n+1} = E_{\text{mec},n} + \frac{k}{m}E_{\text{mec},n}\Delta t^2 \ . \tag{2.63}$$

Comme $k > 0$, $m > 0$ et $E_{\text{mec}} > 0$ (sauf pour le cas trivial $x(t) = 0, v(t) = 0$), on a que

$$\boxed{E_{\text{mec},n+1} > E_{\text{mec},n}}\ , \forall n. \tag{2.64}$$

Ainsi, **l'énergie mécanique, au lieu de rester constante, croît à chaque pas de temps.** On peut même écrire l'Eq.(2.63), soustrayant $E_{\text{mec},n}$ puis en divisant par Δt :

$$\frac{E_{\text{mec},n+1} - E_{\text{mec},n}}{\Delta t} = \left(\frac{k}{m}\Delta t\right)E_{\text{mec},n}$$

Cette équation n'est autre que l'approximation par différences finies de l'équation différentielle

$$\frac{\mathrm{d}E_{\text{mec}}}{\mathrm{d}t} = \left(\frac{k}{m}\Delta t\right)E_{\text{mec}}$$

Dont la solution est

$$\boxed{E_{\text{mec}}(t) = E_{\text{mec}}(0)\exp(\tilde{\gamma}t)}\ , \tag{2.65}$$

L'énergie mécanique du schéma d'Euler explicite, au lieu de rester constante, croît exponentiellement dans le temps, avec un taux de croissance $\tilde{\gamma} = (k/m)\Delta t$.

On note que ce taux de croissance est le double de celui trouvé pour la solution analytique du schéma d'Euler, voir Eq.(2.61).

Bien que cette instabilité numérique ait un taux de croissance qui tende vers zéro avec Δt, on aimerait avoir un algorithme qui évite complètement l'instabilité. C'est le but des trois sections suivantes.

2.5 Schéma d'Euler implicite

Le schéma d'Euler explicite est basé sur l'approximation de la première dérivée des fonctions inconnues $\mathbf{y}$ au temps t_n, c'est-à-dire au début de l'intervalle temporel $[t_n, t_{n+1}]$:

$$\frac{\mathbf{y}_{n+1} - \mathbf{y}_n}{\Delta t} = \mathbf{f}(\mathbf{y}_n, t_n) + \mathcal{O}(\Delta t) \tag{2.66}$$

On en avait obtenu le schéma d'Euler explicite, que l'on réécrit ici :

$$\boxed{\mathbf{y}_{n+1} = \mathbf{y}_n + \mathbf{f}(\mathbf{y}_n, t_n)\Delta t}\,, \tag{2.67}$$

qui est directement utilisable pour implémenter l'algorithme.

L'idée de la méthode implicite se base sur l'approximation de la première dérivée de la fonction $\mathbf{y}$ au temps t_{n+1}, c'est-à-dire *à la fin de l'intervalle temporel* $[t_n, t_{n+1}]$:

$$\frac{\mathbf{y}_{n+1} - \mathbf{y}_n}{\Delta t} = \mathbf{f}(\mathbf{y}_{n+1}, t_{n+1}) + \mathcal{O}(\Delta t) \tag{2.68}$$

On obtient, en multipliant par Δt et négligeant les termes d'ordre 2 en Δt, une équation pour $\mathbf{y}_{n+1}$:

$$\boxed{\mathbf{y}_{n+1} = \mathbf{y}_n + \mathbf{f}(\mathbf{y}_{n+1}, t_{n+1})\Delta t}\,. \tag{2.69}$$

La résolultion de cette équation n'est pas toujours triviale, selon la complexité des fonctions $\mathbf{f}$. Il y a plusieurs méthodes possibles, dont des méthodes itératives. Nous présentons ici la plus simple de celles-ci : la méthode du point fixe. L'idée est de choisir comme première estimation ($k = 0$, k sera un compteur des itérations) :

$$\mathbf{y}_{n+1}^{(k=1)} = \mathbf{y}_n \tag{2.70}$$

Ensuite, on effecture une boucle itérative, $k \to k + 1$:

$$\boxed{\mathbf{y}_{n+1}^{(k+1)} = \mathbf{y}_n + \mathbf{f}(\mathbf{y}_{n+1}^{(k)}, t_{n+1})\Delta t} \tag{2.71}$$

Pour arrêter la boucle itérative, on mesure l'erreur que l'on fait sur la résolution de l'Eq.(2.69) :

$$d = ||\mathbf{y}_{n+1}^{(k+1)} - \mathbf{y}_n - \mathbf{f}(\mathbf{y}_{n+1}^{(k+1)}, t_{n+1})\Delta t||\,, \tag{2.72}$$

et on indique à l'algorithme d'arrêter les itérations lorsque cette erreur d est plus petite qu'une tolérance spécifiée ϵ. Cette méthode fonctionne bien dans le cas de l'oscillateur harmonique.

Ainsi, l'algorithme d'Euler implicite consiste en **deux itérations imbriquées** :

- Une boucle itérative sur le temps ($t_n \to t_{n+1}$),
- ... et, à chaque pas de temps, une itération du point fixe ($k \to k + 1$).

On suggère de l'implémenter et de le tester en exercice.

En résumé :

1. Le schéma d'Euler implicite **converge** : à un intant donné t, la solution numérique tend vers la solution exacte pour $\Delta t \to 0$.
2. La convergence est **d'ordre un** en Δt : l'erreur (i.e. la différence entre solution numérique et solution convergée) est proportionnelle à $(\Delta t)^1$.
3. Le schéma d'Euler implicite est inconditionnellement **stable** pour le problème de l'oscillateur harmonique. Cela signifie que, quelle que soit la valeur de Δt, l'erreur numérique ne croît pas exponentiellement dans le temps. En effet, en appliquant l'analyse de stabilité de propagation de l'erreur, comme à la section 2.4.2, on obtient cette fois :

$$\boxed{\mathbf{e}_{n+1} = \left(\mathbf{I} - \Delta t \frac{\partial \mathbf{f}}{\partial \mathbf{y}}\right)^{-1} \mathbf{e}_n}, \tag{2.73}$$

et pour l'oscillateur harmonique la matrice de gain est :

$$\mathbf{G} = \begin{pmatrix} 1 & -\Delta t \\ (k/m)\Delta t & 1 \end{pmatrix}^{-1} = \frac{1}{\left(1 + \frac{k}{m}\Delta t^2\right)} \begin{pmatrix} 1 & \Delta t \\ -(k/m)\Delta t & 1 \end{pmatrix} \tag{2.74}$$

Les valeurs propres de $\mathbf{G}$ sont

$$\lambda_{1,2} = \frac{1}{\left(1 + \frac{k}{m}\Delta t^2\right)} \left(1 \pm i\sqrt{\frac{k}{m}}\Delta t\right) \tag{2.75}$$

et leur norme est

$$|\lambda_{1,2}| = \frac{1}{\sqrt{1 + \frac{k}{m}\Delta t^2}} < 1, \ \forall \Delta t \ . \tag{2.76}$$

4. La solution analytique du schéma d'Euler implicite pour le problème de l'oscillateur harmonique est une sinusoïdale avec une amplitude exponentiellement décroissante dans le temps. **Cet amortissement est d'origine numérique**, le taux de décroissance est $\gamma = -(k/m)(\Delta t/2)$. En appliquant la même démarche et les mêmes définitions qu'à la section 2.4.3, le schéma d'Euler implicite s'écrit :

$$(\mathbf{I} - \mathbf{M}\Delta t)\mathbf{y}_{n+1} = \mathbf{y}_n \ . \tag{2.77}$$

Avec l'Ansatz

$$\mathbf{y}_n = \mathbf{A} e^{i\omega t_n} \tag{2.78}$$

et le développement linité de $\exp(i\omega\Delta t)$ au 2e ordre, on obtient :

$$\left[\left(1 + i\omega\Delta t - \frac{1}{2}(\omega\Delta t)^2\right)(\mathbf{I} - \Delta t\mathbf{M}) - \mathbf{I}\right]\mathbf{y}_n = 0 \ . \tag{2.79}$$

Ce système algébrique d'équations linéaire homogène pour $\mathbf{y}_n$ n'admet de solution non triviale que si son déterminant est nul, ce qui donne :

$$\left(\omega^2 - \frac{k}{m}\right) + i\left(2\frac{k}{m} - \omega^2\right)\ \omega\Delta t = 0 \ . \tag{2.80}$$

En rsolvant perrturbativement ordre par ordre en $(\omega\Delta t)$, on a, à l'ordre 0 :

$$\omega^{(0)} = \pm\sqrt{\frac{k}{m}} \tag{2.81}$$

et à l'ordre 1 :

$$\omega^{(1)} = i\frac{k}{m}\frac{\Delta t}{2} \tag{2.82}$$

Donc

$$\boxed{\omega = +i\frac{k}{m}\frac{\Delta t}{2} \pm \sqrt{\frac{k}{m}}}\,. \tag{2.83}$$

Ainsi

$$\boxed{\begin{pmatrix} x \\ v \end{pmatrix} = \mathbf{A}e^{i\left(\sqrt{k/m}\right)t}e^{-(k/m)(\Delta t/2)\,t}}\,. \tag{2.84}$$

C'est une oscillation sinusoïdale de fréquence $\sqrt{k/m}$, d'amplitude décroissant exponentiellement avec le taux $\gamma = -(k/m)\Delta t/2$. Cela décrit bien un mouvement amorti. Cependant, cet amortissement est d'origine purement numérique. Comparez avec le cas du schéma d'Euler explicite, Eq.(2.61).

5. Le schéma d'Euler implicite **ne conserve pas** l'énergie mécanique. En fait, quelle que soit la valeur de Δt, il **dissipe** l'énergie mécanique. Cette dissipation, d'origine numérique et non physique, implique que le mouvement tend toujours asymptotiquement ($\lim_{t\to\infty}$) vers la position d'équilibre. En faisant la même démarche qu'à la section 2.4.4, on obtient :

$$E_{\mathrm{mec},n+1} = E_{\mathrm{mec},n} - \frac{k}{m}E_{\mathrm{mec},n}\Delta t^2\ . \tag{2.85}$$

Comme $k > 0$, $m > 0$ et $E_{\mathrm{mec}} > 0$ (sauf pour le cas trivial $x(t) = 0, v(t) = 0$), on a que

$$\boxed{E_{\mathrm{mec},n+1} < E_{\mathrm{mec},n}}\,, \forall n. \tag{2.86}$$

Ainsi, **l'énergie mécanique, au lieu de rester constante, dimunue à chaque pas de temps.** On peut même écrire l'Eq.(2.63), soustrayant $E_{\mathrm{mec},n}$ puis en divisant par Δt :

$$\frac{E_{\mathrm{mec},n+1} - E_{\mathrm{mec},n}}{\Delta t} = -\left(\frac{k}{m}\Delta t\right)E_{\mathrm{mec},n}$$

Cette équation n'est autre que l'approximation par différences finies de l'équation différentielle

$$\frac{\mathrm{d}E_{\mathrm{mec}}}{\mathrm{d}t} = -\left(\frac{k}{m}\Delta t\right)E_{\mathrm{mec}}$$

Dont la solution est

$$\boxed{E_{\mathrm{mec}}(t) = E_{\mathrm{mec}}(0)\exp(-\tilde{\gamma}t)}\,, \tag{2.87}$$

avec $\tilde{\gamma} = (k/m)\Delta t$. Le taux d'amortissement est proportionnel à Δt. Il est donc bien d'origine numérique. Ce taux tend vers zéro lorsque $\Delta t \Rightarrow 0$: la méthode converge.

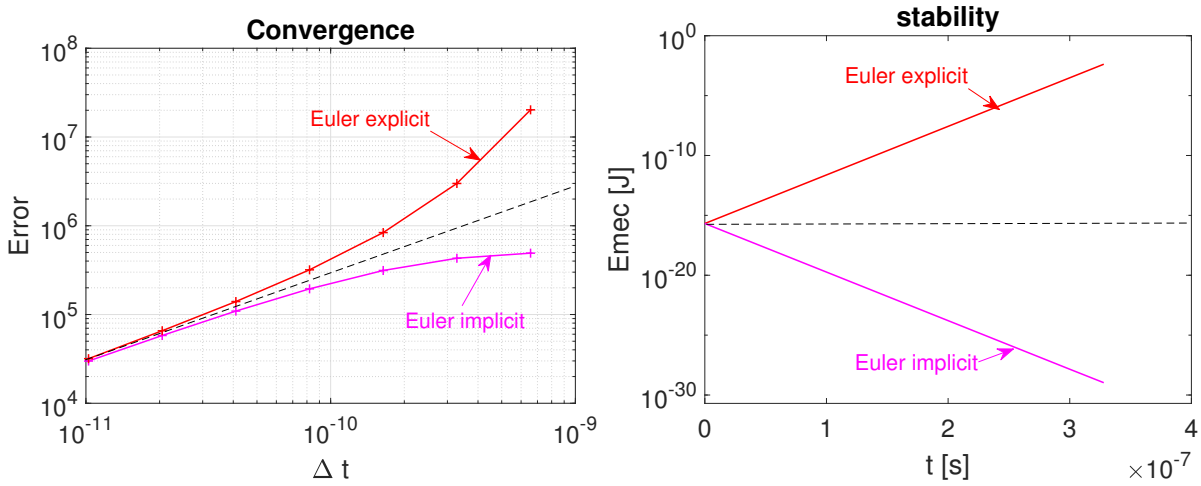


FIGURE 2.11 – A gauche : convergence de l'erreur sur la vitesse finale des schémas d'Euler explicite et implicite. La ligne traitillée est une droite de pente 1. A droite : énergie en fonction du temps. La ligne horizontale en traitillés est la solution analytique : l'énergie est constante. Proton dans un champ magnétique.

En résumé, le schéma d'Euler implicite, appliqué à l'oscillateur harmonique :

- converge à l'ordre 1 en Δt,
- est stable,
- n'est pas conservatif,
- est dissipatif.

Illustration : application au mouvement d'une particule chargée dans un champ magnétique uniforme et constant.

On montre facilement (exercice) que les équations du mouvement pour la vitesse sont mathématiquement semblables à celles de l'oscillateur harmonique. On a appliqué les schémas d'Euler explicite et implicite pour les comparer. Le problème est résoluble analytiquement, on a la solution exacte et donc on peut obtenir l'erreur numérique.

Le cas physique est celui d'un proton, de masse $m = 1.6726 \times 10^{-27}$kg, de charge $q = 1.6022 \times 10^{-19}$C, dans un champ magnétique $B = 4$T, avec une vitesse initiale $v_0 = 5 \times 10^5$m/s. L'étude de convergence de l'erreur sur la vitesse finale après 5 périodes de rotation est montrée à la Fig.2.11. On observe bien une convergence d'ordre 1 pour les deux schémas. Le schéma explicite tend systématiquement à produire une erreur supérieure à cele du schéma implicite.

Pour ce qui est de la stabilité numérique et des propriétés de conservation de l'énergie cinétique, on montre à la Fig.2.11 l'évolution temporelle de l'énergie. L'énergie du schéma explicite croît exponentiellement dans le temps, ce qui est bien le signe d'une instabilité, alors que celle du schéma implicite décroît exponentiellement dans le temps, ce qui est signe de stabilité, mais malheureusement aucun des deux schéma ne conserve l'énergie.

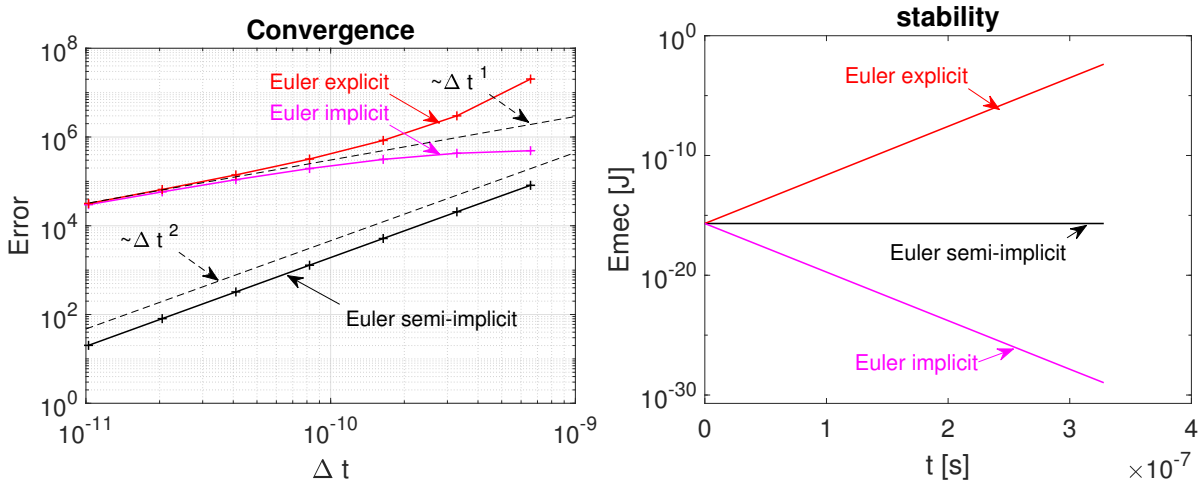


FIGURE 2.12 – A gauche : convergence de l'erreur sur la vitesse finale des schémas d'Euler semi-implicite, explicite et implicite. Les lignes traitillées sont des droites de pente 1 et 2, respectivement. A droite : énergie en fonction du temps. Le schéma semi-implicite satisfait la conservation de l'énergie. Proton dans un champ magnétique.

2.6 Schéma d'Euler semi-implicite

L'inspection de la Fig.2.11 suggère qu'un schéma qui mélangerait les aspects explicite et implicite serait meilleur que chacun des deux schémas pris séparément. D'où l'idée d'un schéma ***semi-implicite***.

On peut unifier la présentation de ces trois schémas d'Euler (explicite, implicite, semi-implicite) en faisant la moyenne pondérée des Eqs.(2.67) et (2.69) :

$$\mathbf{y}_{n+1} = \mathbf{y}_n + (\alpha \mathbf{f}(\mathbf{y}_n, t_n) + (1-\alpha)\mathbf{f}(\mathbf{y}_{n+1}, t_{n+1}))\Delta t \qquad (2.88)$$

avec $\alpha = 1$ pour le schéma explicite, $\alpha = 0$ pour le schéma implicite et $\alpha = 1/2$ pour le schéma semi-implicite.

Le schéma semi-implicite implique la résolution de la partie implicite. On peut utiliser, en l'adaptant, la méthode du point fixe présentée à la section précédente, Eq.(2.71). Comme précédemment, on choisit comme première estimation ($k = 0$, k sera un compteur des itérations) :

$$\mathbf{y}_{n+1}^{(k=1)} = \mathbf{y}_n \qquad (2.89)$$

Ensuite, on effectue une boucle itérative, $k \to k+1$:

$$\boxed{\mathbf{y}_{n+1}^{(k+1)} = \mathbf{y}_n + (\alpha \mathbf{f}(\mathbf{y}_n) + (1-\alpha)\mathbf{f}(\mathbf{y}_{n+1}^{(k)}, t_{n+1}))\Delta t}\,. \qquad (2.90)$$

Pour arrêter la boucle itérative, on mesure l'erreur que l'on fait sur la résolution de l'Eq. (2.88) :

$$d = ||\mathbf{y}_{n+1}^{(k+1)} - \mathbf{y}_n - (\alpha \mathbf{f}(\mathbf{y}_n) + (1-\alpha)\mathbf{f}(\mathbf{y}_{n+1}^{(k+1)}, t_{n+1}))\Delta t||\,, \qquad (2.91)$$

et on indique à l'algorithme d'arrêter les itérations lorsque cette erreur d est plus petite qu'une tolérance spécifiée ϵ.

On illustre à la Fig.2.12 les résultats obtenus pour le problème de la particule chargée dans un champ magnétique, en reportant les trois schémas pour les comparer.

Pour ce qui est de la convergence, les résultats du schéma semi-implicite ne se situent pas entre ceux des schémas explicite et implicite. On observe un ordre de convergence d'ordre 2, et non d'ordre 1. La précision du résultat numérique est plusieurs ordres de grandeur meilleure. Nous avions déjà observé qu'un schéma de différences finies *centré* offre un ordre de convergence supérieur pour l'évaluation de la dérivée, voir Fig.1.3.

Pour ce qui est de l'évolution temporelle de l'énergie mécanique, on observe que, à une tolérance ϵ près sur les itérations du point fixe, le schéma semi-implicite **conserve l'énergie exactement**. C'est une propriété très intéressante, qui permet de faire de longues simulations, en évitant l'instabilité du schéma explicite tout en évitant l'érosion de l'énergie du schéma implicite.

2.7 Schémas symplectiques : Euler-Cromer, Verlet et variantes

Voir cours de mécanique analytique : la dynamique Hamiltonienne est une reformulation de la dynamique Newtonienne. Les équations du mouvement sont écrites pour un système classique à M degrés de liberté, en utilisant les coordonnées généralisées $\mathbf{q}$ et les moments conjugués $\mathbf{p}$, et l'Hamiltonien $H(\mathbf{p},\mathbf{q})$. On a utilisé la notation vectorielle $\mathbf{q} = (q_1, q_2, ..., q_M)$, $\mathbf{p} = (p_1, p_2, ..., p_M)$.

$$\frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t} = \frac{\partial H}{\partial \mathbf{p}}\,, \quad \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = -\frac{\partial H}{\partial \mathbf{q}}\,. \tag{2.92}$$

Par exemple, pour un système de particules de masses m, d'énergie cinétique $K = \sum_i p_i^2/2m$, soumises à des forces dérivant d'un potentiel $V(q_1, q_2, .., q_M)$, les équations s'écrivent

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = -\frac{\partial V}{\partial \mathbf{q}} = \mathbf{F}\,, \tag{2.93}$$

$$\frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t} = \mathbf{p}/m\,. \tag{2.94}$$

On sait bien que dans de tels systèmes l'énergie mécanique (qui n'est autre que la valeur numérique de l'Hamiltonien) est une constante du mouvement. On vérifie cette propriété pour contrôler la qualité de la solution numérique. Mais il y a bien d'autres quantités qui

sont conservées. En reliant les deux vecteurs $\mathbf{q}$ et $\mathbf{p}$ pour former un vecteur $\mathbf{z}$ (appelé vecteur de l'espace de phase) :

$$\mathbf{z} = \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} , \tag{2.95}$$

on peut écrire les équations du mouvement Hamiltoniennes, Eq.(2.92), comme une équation d'évolution temporelle pour $\mathbf{z}$

$$\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t} = \mathbf{J} \cdot \frac{\partial H}{\partial \mathbf{z}} \tag{2.96}$$

avec la matrice $\mathbf{J}$

$$\mathbf{J} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{pmatrix} \tag{2.97}$$

appelée "matrice symplectique". Une propriété des transformations canoniques est de conserver la "forme symplectique" définie par

$$s(\mathbf{z}_1, \mathbf{z}_2) = \mathbf{z}_1^T \cdot \mathbf{J}\mathbf{z}_2 \ . \tag{2.98}$$

Par exemple, si $\mathbf{z}_1(0)$ et $\mathbf{z}_2(0)$ représentent deux conditions initiales différentes, la quantité $s(\mathbf{z}_1(t), \mathbf{z}_2(t))$ est une constante du mouvement. L'idée est d'utiliser cette quantité conservée comme indicateur de qualité des schémas numériques.

2.7.1 Algorithme d'Euler-Cromer

Le schéma d'Euler explicite, Section 2.1.2, appliqué au problème Eqs. (2.93-2.94) donnerait

$$\mathbf{p}(t + \Delta t) = \mathbf{p}(t) + \Delta t\, \mathbf{F}(\mathbf{q}(t)) \tag{2.99}$$
$$\mathbf{q}(t + \Delta t) = \mathbf{q}(t) + \Delta t\, \mathbf{p}(t)/m \tag{2.100}$$

L'idée est, au lieu d'utiliser le moment $\mathbf{p}(t)$ pour avancer les coordonnées dans (2.100), d'utiliser le moment à l'instant $t + \Delta t$ obtenu de (2.99). On a ainsi l'algorithme Euler-Cromer "A" :

$$\mathbf{p}(t + \Delta t) = \mathbf{p}(t) + \Delta t\, \mathbf{F}(\mathbf{q}(t)) \tag{2.101}$$
$$\mathbf{q}(t + \Delta t) = \mathbf{q}(t) + \Delta t\, \mathbf{p}(t + \Delta t)/m \tag{2.102}$$

On peut aussi inverser l'ordre dans lequel on évalue $\mathbf{p}$ et $\mathbf{q}$, ce qui donne l'algorithme Euler-Cromer "B" :

$$\mathbf{q}(t + \Delta t) = \mathbf{q}(t) + \Delta t\, \mathbf{p}(t)/m \tag{2.103}$$
$$\mathbf{p}(t + \Delta t) = \mathbf{p}(t) + \Delta t\, \mathbf{F}(\mathbf{q}(t + \Delta t)) \tag{2.104}$$

Ces algorithmes ont des propriétés remarquables lorsqu'ils sont appliqués au problème de l'oscillateur harmonique, par exemple. La FIG. 2.13 montre que l'instabilité numérique a disparu. La convergence numérique avec Δt est bien supérieure : les trois résultats, pour

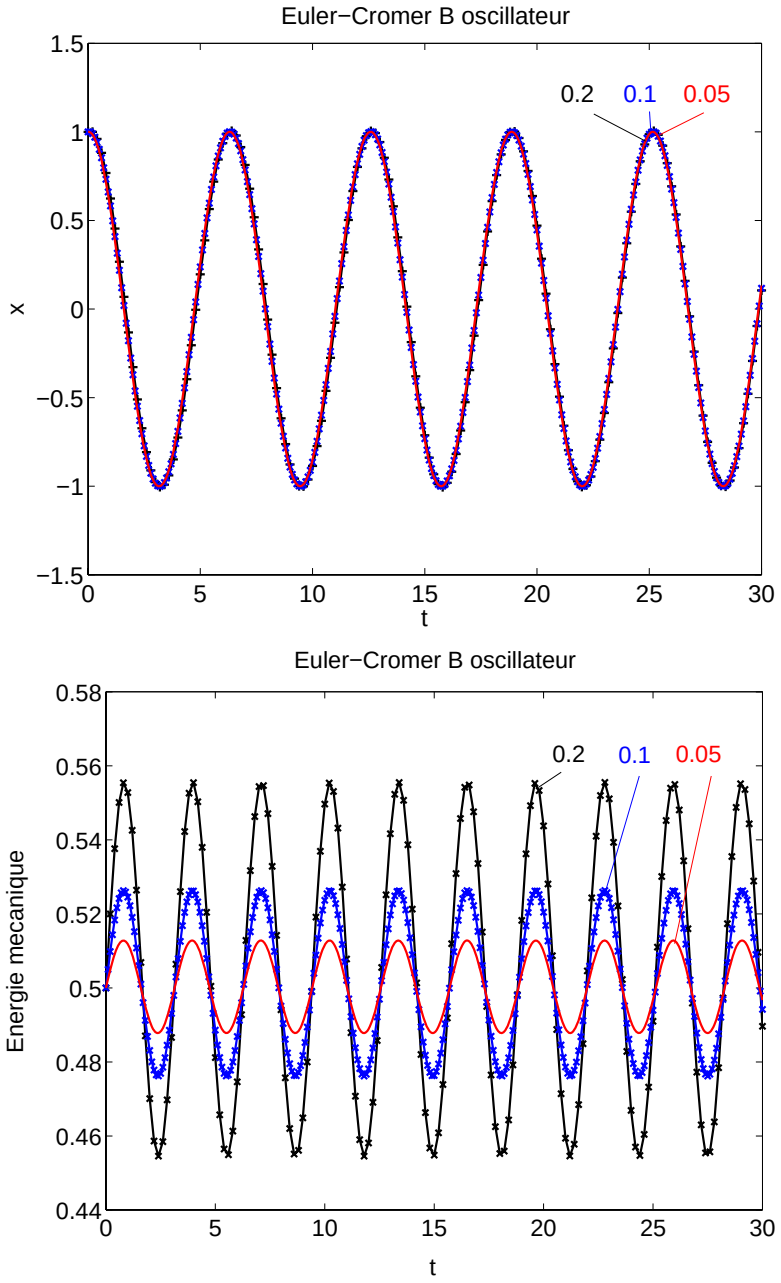


FIGURE 2.13 – *Oscillateur harmonique avec la méthode d'Euler-Cromer, $k = 1$, $m = 1$. Trois exécutions avec $\Delta t = 0.2, 0.1, 0.05$. Le schéma est stable, l'amplitude des oscillations (haut) est constante et l'énergie mécanique (bas) est conservée en moyenne, avec une erreur instantanée qui tend vers zéro avec Δt.*

$\Delta t = 0.2, 0.1, 0.05$ sont pratiquement indiscernables à l'échelle de la figure (comparer avec le schéma d'Euler explicite, FIG. 2.10). On note que l'énergie mécanique, bien que pas exactement conservée à tous les temps, est conservée exactement en moyenne temporelle, la non-conservation instantanée tendant vers zéro proportionellement à Δt. C'est une propriété intéressante, surtout lorsque l'on veut faire de longues simulations.

Les algorithmes "A" et "B" donnent des résultats très similaires (FIG. 2.14). On remarque que les erreurs sur l'énergie mécanique des deux schémas sont opposées. Cela suggère de combiner ces deux algorithmes.

2.7.2 Algorithme de Verlet et ses variantes

On obtient cet algorithme en divisant le pas temporel Δt en deux. Pour la première moitié, on utilise l'algorithme Euler-Cromer "A", et pour la deuxième moitié l'algorithme Euler-Cromer "B". On obtient

$$\begin{array}{rcl} \mathbf{p}(t+\Delta t/2) &=& \mathbf{p}(t) + (\Delta t/2)\mathbf{F}(\mathbf{q}(t)) \\ \mathbf{q}(t+\Delta t/2) &=& \mathbf{q}(t) + (\Delta t/2m)\mathbf{p}(t+\Delta t/2) \\ \mathbf{q}(t+\Delta t) &=& \mathbf{q}(t+\Delta t/2) + (\Delta t/2m)\mathbf{p}(t+\Delta t/2) \\ \mathbf{p}(t+\Delta t) &=& \mathbf{p}(t+\Delta t/2) + (\Delta t/2)\mathbf{F}(\mathbf{q}(t+\Delta t)) \end{array} \tag{2.105}$$

En éliminant les quantités évaluées au milieu de l'intervalle temporel $(t + \Delta t/2)$, on obtient

$$\boxed{\begin{array}{rcl} \mathbf{q}(t+\Delta t) &=& \mathbf{q}(t) + (\Delta t/m)\,\mathbf{p}(t) + ((\Delta t)^2/2m)\,\mathbf{F}(\mathbf{q}(t)) + \mathcal{O}\left((\Delta t)^4\right) \\ \mathbf{p}(t+\Delta t) &=& \mathbf{p}(t) + (\Delta t/2)\,[\mathbf{F}(\mathbf{q}(t+\Delta t)) + \mathbf{F}(\mathbf{q}(t))] \end{array}} \tag{2.106}$$

L'algorithme de Stormer-Verlet existe en plusieurs formulations. Celle que nous avons présentée ici est due à Swope en 1982 ; elle est parfois appelée *velocity Verlet*. Une autre formulation est le *Verlet leapfrog*, (ou *saute-mouton*), due à Vineyard en 1962 :

$$\boxed{\begin{array}{rcl} \mathbf{p}(t+\Delta t/2) &=& \mathbf{p}(t-\Delta t/2) + \Delta t\,\mathbf{F}(\mathbf{q}(t)) \\ \mathbf{q}(t+\Delta t) &=& \mathbf{q}(t) + (\Delta t/m)\,\mathbf{p}(t+\Delta t/2) + \mathcal{O}\left((\Delta t)^4\right) \end{array}} \tag{2.107}$$

Dans cette formulation, lorsque les conditions initiales sont connues en $t = 0$, l'algorithme doit être initialisé par un "demi-pas temporel" pour $\mathbf{p}$. Plus précisément, pour le premier pas temporel, on remplace la première ligne de l'Eq.(2.107) par $\mathbf{p}(\Delta t/2) = \mathbf{p}(0) + (\Delta t/2)\,\mathbf{F}(\mathbf{q}(0))$. L'algorithme peut ensuite se poursuivre normalement.

Une troisième formulation (celle de Verlet en 1967) [2] s'obtient de l'équation

$$\frac{\mathrm{d}^2\mathbf{q}}{\mathrm{d}t^2} = \frac{1}{m}\mathbf{F}(\mathbf{q}(t)) \tag{2.108}$$

2. Stormer avait déjà utilisé ce schéma en ... 1907 pour le calcul des trajectoires des particules piégées dans le champ magnétique terretre.

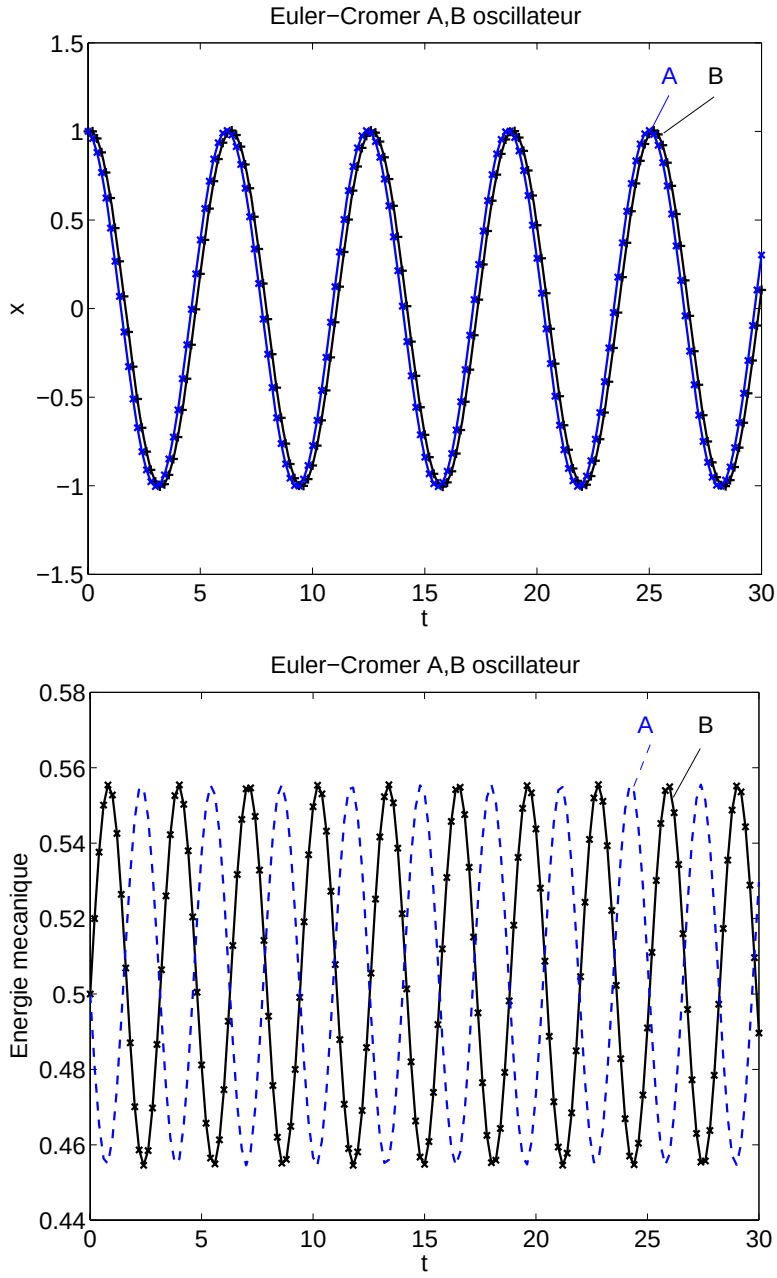


FIGURE 2.14 – *Oscillateur harmonique avec la méthode d'Euler-Cromer,* $k = 1$, $m = 1$, $\Delta t = 0.2$. *Les algorithmes "A" et "B" donnent des résultats très similaires.*

avec l'approximation par différences finies

$$\frac{\mathrm{d}^2\mathbf{q}}{\mathrm{d}t^2}(t) = \frac{1}{(\Delta t)^2}\left(\mathbf{q}(t+\Delta t) - 2\mathbf{q}(t) + \mathbf{q}(t-\Delta t)\right) + \mathcal{O}\left((\Delta t)^2\right) \tag{2.109}$$

ce qui conduit à

$$\boxed{\mathbf{q}(t+\Delta t) = 2\mathbf{q}(t) - \mathbf{q}(t-\Delta t) + \frac{(\Delta t)^2}{m}\,\mathbf{F}(\mathbf{q}(t)) + \mathcal{O}\left((\Delta t)^4\right)}\,. \tag{2.110}$$

Cet algorithme requiert une estimation pour $\mathbf{q}(-\Delta t)$ pour être initialisé.

On remarque que les moments $\mathbf{p}$ n'interviennent pas explicitement dans cet algorithme. On peut les obtenir *a posteriori* à partir de la deuxième des équations Eq.(2.106), ou avec une estimation de différences finies centrées :

$$\mathbf{p}(t) = \frac{m}{2\Delta t}(\mathbf{q}(t+\Delta t) - \mathbf{q}(t-\Delta t)) + \mathcal{O}\left((\Delta t)^2\right) \tag{2.111}$$

Les formulations de "velocity-Verlet", Eq.(2.106) et "Verlet", Eq.(2.110) sont strictement équivalentes. En effet, en notant, pour simplifier, $\mathbf{q}_j = \mathbf{q}(t_j)$, $\mathbf{p}_j = \mathbf{p}(t_j)$, l'Eq.(2.106) s'écrit, pour le pas t_{j+1} et le pas t_j :

$$\begin{aligned}
\mathbf{q}_{j+1} &= \mathbf{q}_j + (\Delta t/m)\mathbf{p}_j + ((\Delta t)^2/2m)\mathbf{F}(\mathbf{q}_j) && (2.112)\\
\mathbf{q}_j &= \mathbf{q}_{j-1} + (\Delta t/m)\mathbf{p}_{j-1} + ((\Delta t)^2/2m)\mathbf{F}(\mathbf{q}_{j-1}) && (2.113)\\
\mathbf{p}_{j+1} &= \mathbf{p}_j + (\Delta t/2)(\mathbf{F}(\mathbf{q}_{j+1} + \mathbf{F}(\mathbf{q}_j))) && (2.114)\\
\mathbf{p}_j &= \mathbf{p}_{j-1} + (\Delta t/2)(\mathbf{F}(\mathbf{q}_j + \mathbf{F}(\mathbf{q}_{j-1}))) && (2.115)
\end{aligned}$$

Soustrayant (2.112)-(2.113), on a :

$$\mathbf{q}_{j+1} - \mathbf{q}_j = \mathbf{q}_j - \mathbf{q}_{j-1} + (\Delta t/m)(\mathbf{p}_j - \mathbf{p}_{j-1}) + ((\Delta t)^2/2m)(\mathbf{F}(\mathbf{q}_j) - \mathbf{F}(\mathbf{q}_{j-1})) \tag{2.116}$$

Substituant $\mathbf{p}_j - \mathbf{p}_{j-1}$ à partir de l'Eq.(2.115), on obtient

$$\mathbf{q}_{j+1} = 2\mathbf{q}_j - \mathbf{q}_{j-1} + ((\Delta t)^2/m)\mathbf{F}(\mathbf{q}_j) \tag{2.117}$$

qui est bien l'expression de l'Eq.(2.110).

2.7.3 Analyse de la stabilité du schéma de Verlet

On fait une analyse des erreurs comme à la Section 2.4.2. Pour le problème de l'oscillateur harmonique, en utilisant l'algorithme Eq.(2.110), les erreurs $\mathbf{e}(t)$ obéissent à :

$$\mathbf{e}(t+\Delta t) = (2 - (k/m)(\Delta t)^2)\mathbf{e}(t) - \mathbf{e}(t-\Delta t)\ . \tag{2.118}$$

Définissons une matrice d'amplification d'erreurs **G** par

$$\begin{pmatrix} \mathbf{e}(t+\Delta t) \\ \mathbf{e}(t) \end{pmatrix} = \mathbf{G} \begin{pmatrix} \mathbf{e}(t) \\ \mathbf{e}(t-\Delta t) \end{pmatrix} . \tag{2.119}$$

On a donc dans notre cas

$$\mathbf{G} = \begin{pmatrix} (2-(k/m)(\Delta t)^2) & -1 \\ 1 & 0 \end{pmatrix} . \tag{2.120}$$

La condition de stabilité est que toutes les valeurs propres λ_i de G soient de module inférieur ou égal à 1. L'équation caractéristique pour ces valeurs propres est

$$\det(\mathbf{G} - \lambda \mathbf{I}) = 0 \,, \tag{2.121}$$

ce qui donne

$$\lambda^2 - (2-(k/m)(\Delta t)^2)\lambda + 1 = 0 \;\Rightarrow\; \lambda_{1,2} = \left(1 - \frac{k}{m}\frac{(\Delta t)^2)}{2}\right) \pm \sqrt{\frac{k^2}{m^2}\frac{(\Delta t)^4}{4} - \frac{k}{m}(\Delta t)^2} \,. \tag{2.122}$$

Pour $(k/m)(\Delta t)^2 < 4$, on a $|\lambda_{1,2}| = 1$ et l'algorithme de Verlet est stable.

Suggestion d'exercice. On vérifiera les propriétés de stabilité et de convergence de cet algorithme.

On montre un example à la FIG. 2.15 d'application du schéma, sous la forme de l'Eq.(2.106) ; la précision est bien meilleure que pour l'algorithme d' Euler-Cromer (voir FIG. 2.13) pour un même pas temporel Δt. **Cependant, l'intérêt majeur de cet algorithme est qu'il peut être utilisé pour de longues simulations, sans qu'il y ait accumulation systématique d'erreurs**. Par exemple, l'énergie mécanique reste conservée en moyenne sur une période. De plus, l'erreur instantanée sur E_{mec} converge vers zéro en $(\Delta t)^2$ (voir exercice).

La propriété "symplectique" de l'algorithme est illustrée à la FIG. 2.16. La forme symplectique, Eq.(2.98), peut être interprétée géométriquement dans l'espace de phase (q,p) comme l'aire du quadrilatère construit sur $\mathbf{z}_1$ et $\mathbf{z}_2$. Ainsi, les systèmes Hamiltoniens **conservent les aires dans l'espace de phase**. Dans la FIG. 2.16 on a choisi 4 conditions initiales voisines de $(x = 1, v = 0)$, formant un quadrilatère. On a représenté des instatntanés de ce quadrilatère au cours de son évolution temporelle. S'il y a très légère déformation, l'aire est par contre exactement conservée.

Suggestion d'exercice. Pendule simple (force en $\sin\theta$), étudier la période des oscillations en fonction de l'amplitude. A partir de la conservation de l'énergie mécanique, dériver une expression pour la période, que l'on intégrera numériquement (avec, p.ex. la méthode des trapèzes, voir Annexe B). Comparer avec les résultats des simulations.

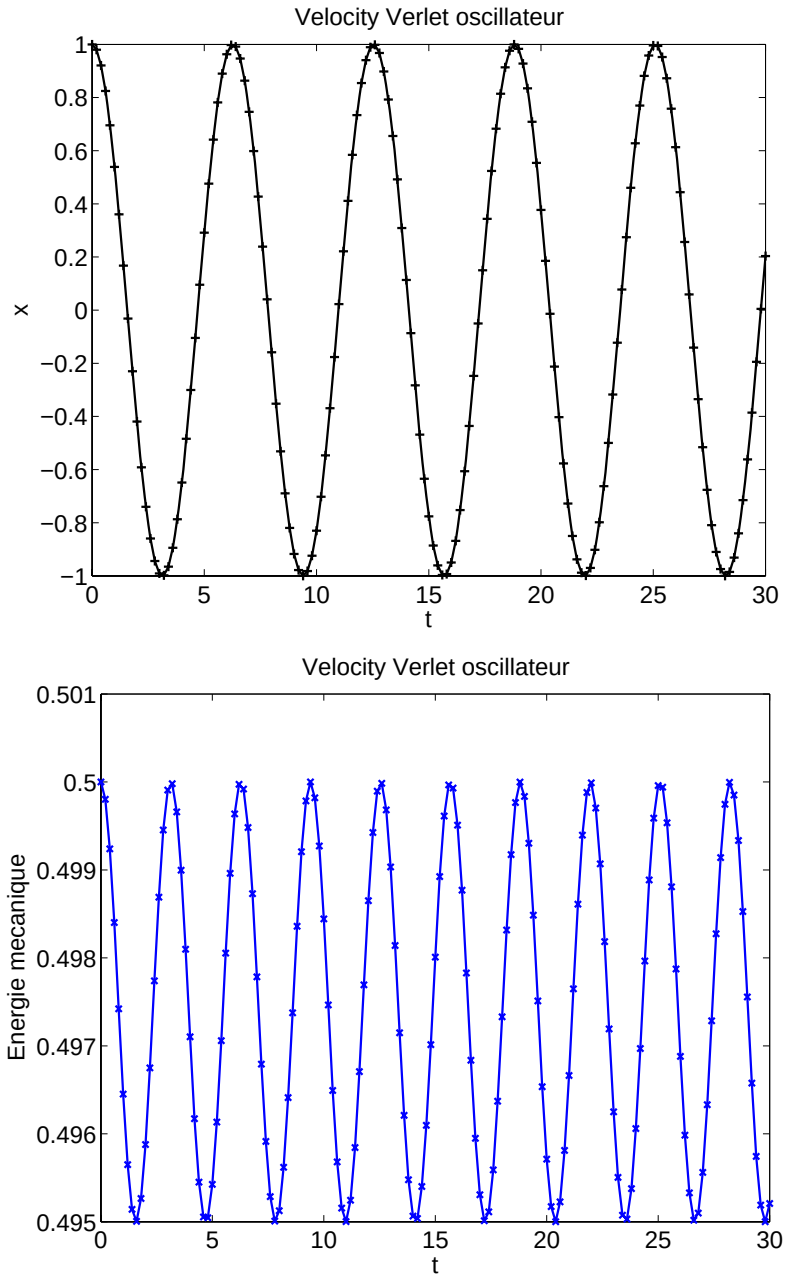


FIGURE 2.15 – *Oscillateur harmonique avec la méthode de Verlet ("velocity Verlet", Eq.(2.106)), $k = 1$, $m = 1$, $\Delta t = 0.2$. Le schéma est stable, l'amplitude des oscillations (haut) est constante et l'énergie mécanique (bas) est conservée en moyenne, avec une erreur instantanée qui tend vers zéro avec $(\Delta t)^2$.*

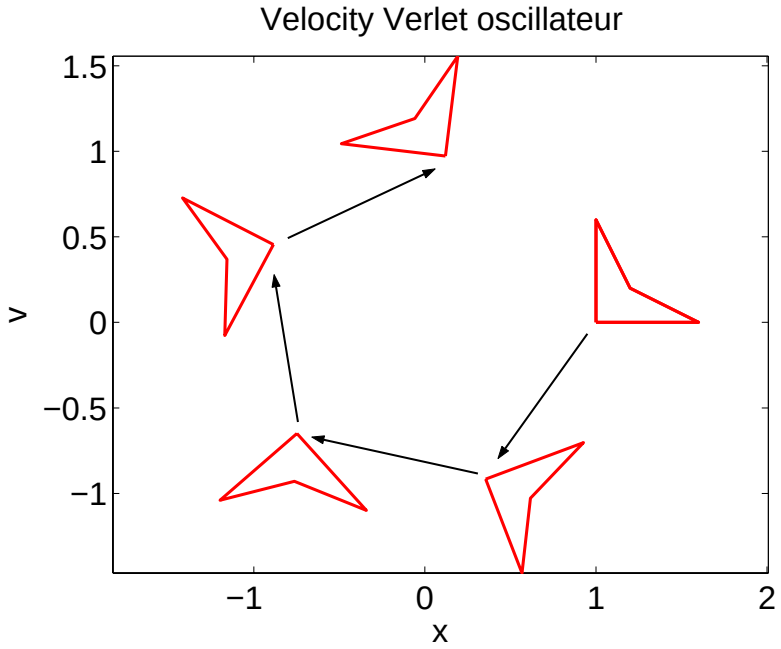


FIGURE 2.16 – *Oscillateur harmonique avec la méthode de Verlet ("velocity Verlet", Eq.(2.106)), $k = 1$, $m = 1$, $\Delta t = 0.2$, 4 conditions initiales au voisinage de $(x = 0, v = 0)$ formant un quadrilatère dont l'évolution temporelle est illustrée par des instantanés. L'algorithme reproduit fidèlement la propriété fondamentale de conservation de l'aire de l'espace de phase.*

2.7.4 Extension de Verlet à des forces dépendant explicitement du temps et de la vitesse

La section précédente a présenté des algorithmes qui s'appliquent directement à des systèmes dynamiques où les forces ne dépendent explicitement que de la position. Ainsi, l'algorithme de Verlet *"velocity-Verlet"*, Eq.(2.106), pour les équations

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\vec{x}\\ \vec{v}\end{pmatrix} = \begin{pmatrix}\vec{v}\\ \vec{F}(\vec{x})/m\end{pmatrix} \tag{2.123}$$

s'écrit, en posant $\vec{a}(\vec{x}) = \vec{F}(\vec{x})/m$,

$$\begin{aligned}\vec{x}_{j+1} &= \vec{x}_j + \vec{v}_j\Delta t + \tfrac{1}{2}\vec{a}(\vec{x}_j)(\Delta t)^2\\ \vec{v}_{j+1} &= \vec{v}_j + \tfrac{1}{2}\left(\vec{a}(\vec{x}_j) + \vec{a}(\vec{x}_{j+1})\right)\Delta t\end{aligned} \tag{2.124}$$

Une première généralisation, à des forces dépendant explicitement du temps, est immédiate, en posant $\vec{a}(\vec{x}, t) = \vec{F}(\vec{x}, t)/m$:

$$\begin{aligned}\vec{x}_{j+1} &= \vec{x}_j + \vec{v}_j\Delta t + \tfrac{1}{2}\vec{a}(\vec{x}_j, t_j)(\Delta t)^2\\ \vec{v}_{j+1} &= \vec{v}_j + \tfrac{1}{2}\left(\vec{a}(\vec{x}_j, t_j) + \vec{a}(\vec{x}_{j+1}, t_{j+1})\right)\Delta t\end{aligned} \tag{2.125}$$

Si on a des forces qui dépendent explicitement de la vitesse, alors il n'est souvent pas possible de trouver un algorithme symplectique, tout simplement parce que le système

d'équations ne satisfait alors plus la condition symplectique. C'est le cas par exemple des forces de friction ou de traînée aérodynamique. On peut néanmoins proposer un algorithme qui sera encore d'ordre 2 et qui sera symplectique dans la limite des forces de friction ou de traînée tendant vers zéro.

On se restreindra au cas où la partie de la force dépendant explicitement de la vitesse est additive :

$$\vec{F}(\vec{x},\vec{v},t) = \vec{F}_1(\vec{x},t) + \vec{F}_2(\vec{v},t) \tag{2.126}$$

On définit alors, en divisant par la masse m :

$$\vec{a}(\vec{x},\vec{v},t) = \vec{a}_1(\vec{x},t) + \vec{a}_2(\vec{v},t)\ . \tag{2.127}$$

Dans la boucle temporelle, on commence par faire la mise à jour de la position, comme dans l'algorithme de base :

$$\boxed{\vec{x}_{j+1} = \vec{x}_j + \vec{v}_j\Delta t + \frac{1}{2}\vec{a}(\vec{x}_j,\vec{v}_j,t_j)(\Delta t)^2} \tag{2.128}$$

On fait alors un ***demi-pas*** pour la vitesse :

$$\boxed{\vec{v}_{j+1/2} = \vec{v}_j + \frac{1}{2}\vec{a}(\vec{x}_j,\vec{v}_j,t_j)\Delta t} \tag{2.129}$$

La mise à jour de la vitesse se fait avec la même expression que dans l'algorithme de base pour la partie $\vec{a}_1(\vec{x},t)$, et avec un pas centré en $j+1/2$ pour la partie $\vec{a}_2(\vec{v})$:

$$\boxed{\vec{v}_{j+1} = \vec{v}_j + \frac{1}{2}\left(\vec{a}_1(\vec{x}_j,t_j) + \vec{a}_1(\vec{x}_{j+1},t_{j+1})\right)\Delta t + \vec{a}_2(\vec{v}_{j+1/2},t_{j+1/2})\Delta t} \tag{2.130}$$

L'algorithme est constitué des expressions (2.128)(2.129) (2.130). Il implique l'écriture de deux fonctions distinctes $a_1(\vec{x},t)$ et $a_2(\vec{v},t)$. A chaque pas de temps, ces fonctions sont appelées deux fois.

On peut formuler l'expression (2.130) autrement, en écrivant le dernier terme comme $(1/2)\vec{a}_2(\vec{v}_{j+1/2},t_{j+1/2})\Delta t + (1/2)\vec{a}_2(\vec{v}_{j+1/2},t_{j+1/2})\Delta t$ et en l'insérant dans la parenthèse du 2e terme, pour obtenir :

$$\vec{v}_{j+1} = \vec{v}_j + \frac{1}{2}\left(\vec{a}_1(\vec{x}_j,t_j) + \vec{a}_2(\vec{v}_{j+1/2},t_{j+1/2}) + \vec{a}_1(\vec{x}_{j+1},t_{j+1}) + \vec{a}_2(\vec{v}_{j+1/2},t_{j+1/2})\right)\Delta t \tag{2.131}$$

Si $\vec{a}_1$ et $\vec{a}_2$ ne dépendent pas tous deux explicitement du temps, alors en regroupant et se rappelant la définition de la fonction $\vec{a} = \vec{a}_1 + \vec{a}_2$, on a :

$$\boxed{\vec{v}_{j+1} = \vec{v}_j + \frac{1}{2}\left(\vec{a}(\vec{x}_j,\vec{v}_{j+1/2},t_j) + \vec{a}(\vec{x}_{j+1},\vec{v}_{j+1/2},t_{j+1})\right)\Delta t} \tag{2.132}$$

L'algorithme est alors composé des expressions (2.128)(2.129) (2.132). Il implique à chaque pas de temps 3 appels à la fonction accéleration, $\vec{a}(\vec{x},\vec{v},t)$, avec **3 combinaisons différentes d'arguments**.

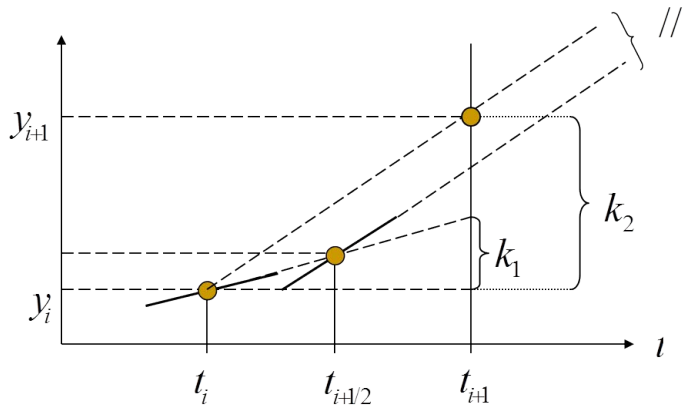


FIGURE 2.17 – *Schéma de Runge-Kutta d'ordre 2.*

2.8 Schémas de Runge-Kutta

Ces schémas s'appliquent à des équations différentielles du 1er ordre

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = \mathbf{f}(\mathbf{y}, t) \ . \tag{2.133}$$

On se limitera ici à énoncer l'idée de base de tels schémas. Si on revient au schéma d'Euler, Eq.(2.10), il approxime la valeur de $\mathbf{f}$ sur l'intervalle $[t_i \ t_{i+1}]$ par une constante, évaluée au début de l'intervalle temporel. Une meilleure approche est de calculer d'abord un "prédicteur" pour $\mathbf{y}$ à la mi-temps $t_{i+1/2}$, évaluer $\mathbf{f}_{i+1/2} = \mathbf{f}(y_{i+1/2}, t_{i+1/2})$, puis refaire un pas complet en utilisant $\mathbf{f}_{i+1/2}$ au lieu de $\mathbf{f}_i$. Cela donne l'algorithme de **Runge-Kutta d'ordre 2**, voir Fig. 2.17 :

$$\boxed{\begin{array}{rcl} \mathbf{k}_1 &=& \Delta t\, \mathbf{f}(\mathbf{y}_i, t_i) \\ \mathbf{k}_2 &=& \Delta t\, \mathbf{f}(\mathbf{y}_i + \frac{1}{2}\mathbf{k}_1, t_{i+1/2}) \\ \mathbf{y}_{i+1} &=& \mathbf{y}_i + \mathbf{k}_2 \end{array}} \tag{2.134}$$

Cet algorithme est d'ordre 2, autrement dit l'erreur sur **un** pas de temps est d'ordre $\mathcal{O}(\Delta t)^3$, et l'erreur numérique globale jusqu'à un temps donné $t = t_{fin}$ est $\mathcal{O}(\Delta t)^2$, ceci pour autant que la fonction $\mathbf{f}$ et la solution $\mathbf{y}$ soient infiniment différentiables. La preuve mathématique est présentée ci-dessous. De plus, on peut se poser la question si l'algorithme ci-dessus, Eq.(2.134), est le seul algorithme d'ordre 2 possible : par exemple, peut-on choisir un autre point que le milieu de l'intervalle, $t_{i+1/2}$, pour évaluer la fonction $\mathbf{f}$? La réponse est qu'il y a une infinité d'autres choix possibles.

Essayons de généraliser l'algorithme Eq.(2.134). Soit 3 nombres a, b, λ entre 0 et 1. On pose alors un schéma généralisé :

$$\begin{array}{rcl} \mathbf{k}_1 &=& \Delta t\, \mathbf{f}(\mathbf{y}_i, t_i) \\ \mathbf{k}_2 &=& \Delta t\, \mathbf{f}(\mathbf{y}_i + \lambda\mathbf{k}_1, t + \lambda\Delta t) \\ \mathbf{y}_{i+1} &=& \mathbf{y}_i + a\mathbf{k}_1 + b\mathbf{k}_2 \end{array} \tag{2.135}$$

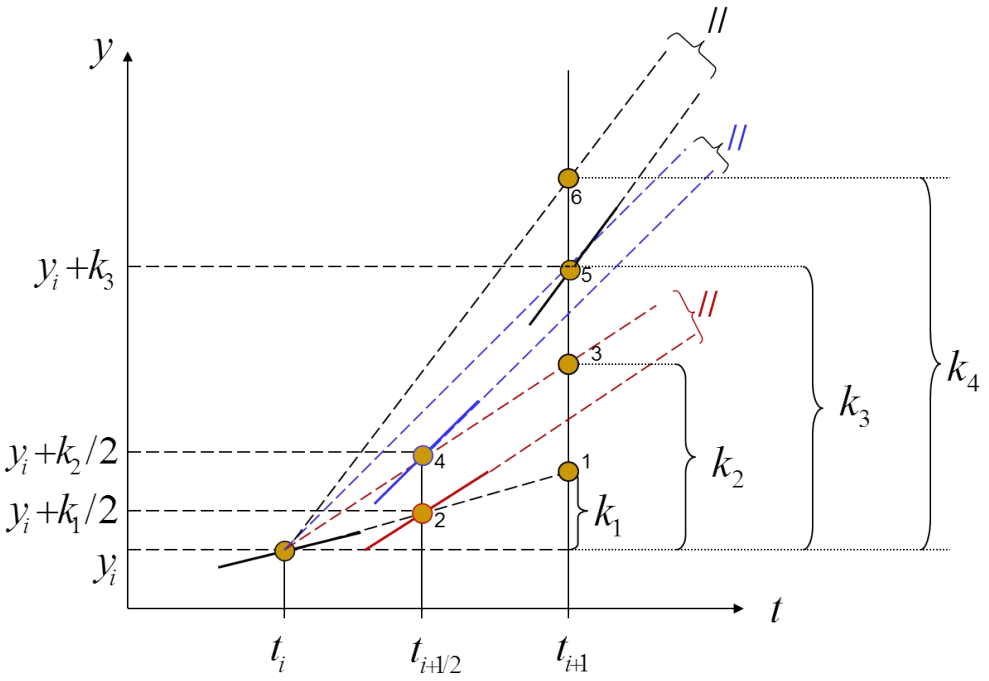


FIGURE 2.18 – *Schéma de Runge-Kutta d'ordre 4.*

Ainsi, on obtient la solution au pas suivant en prenant comme estimation de $\mathrm{d}\mathbf{y}/dt$ (donc de $\mathbf{f}$) une somme pondérée d'une évaluation en début d'intervalle ($\mathbf{k}_1$) et d'une évaluation quelque part entre le début et la fin de l'intervalle ($\mathbf{k}_2$). Le but est de trouver un ensemble d'équations qui nous donnera les valeurs de a, b, λ telles que l'algorithme soit d'ordre 2. Pour que l'algorithme soit d'ordre 2, il faut que la solution numérique coïncide avec le développement limité de la solution exacte au moins jusqu'aux termes d'ordre 2. A partir du schéma Eq.(2.135), on obtient :

$$\mathbf{y}_{i+1} = \mathbf{y}_i + a\Delta t f(y_i, t_i) + b\Delta t f(\mathbf{y}_i + \lambda \mathbf{k}_1, t_i + \lambda \Delta t) \tag{2.136}$$

En développant la fonction $\mathbf{f}$ au voisinage de $(\mathbf{y}_i, t_i)$, on obtient :

$$\mathbf{y}_{i+1} = \mathbf{y}_i + \Delta t(a+b)f + (\Delta t)^2 \left(\lambda b \frac{\partial \mathbf{f}}{\partial \mathbf{y}} \mathbf{f} + \lambda b \frac{\partial \mathbf{f}}{\partial t} \right) + \mathcal{O}(\Delta t)^3 \tag{2.137}$$

D'autre part, nous obtenons le développement limité de la solution exacte :

$$y(t_i + \Delta t) = y(t_i) + \Delta t \mathbf{f} + (\Delta t)^2 \frac{1}{2} \left(\frac{\partial \mathbf{f}}{\partial t} + \frac{\partial \mathbf{f}}{\partial \mathbf{y}} \mathbf{f} \right) + \mathcal{O}(\Delta t)^3 \, . \tag{2.138}$$

(NB : pour obtenir cette relation, on a substitué $\mathrm{d}\mathbf{y}/\mathrm{d}t$ par $\mathbf{f}$, conformément à l'équation différentielle). En comparant ces deux dernières expressions, nous obtenons les conditions suivantes pour que le schéma soit d'ordre 2 :

$$a + b = 1 \, , \quad \lambda b = \frac{1}{2}; \, . \tag{2.139}$$

L'algorithme donné précédemment, Eq.(2.134), correspond ainsi à $a = 0, b = 1, \lambda = 1/2$.

On peut aussi choisir, par exemple, $a = b = 1/2, \lambda = 1$, ce qui donne :

$$\begin{aligned} \mathbf{k}_1 &= \Delta t\, \mathbf{f}(\mathbf{y}_i, t_i) \\ \mathbf{k}_2 &= \Delta t\, \mathbf{f}(\mathbf{y}_i + \mathbf{k}_1, t + \Delta t) \\ \mathbf{y}_{i+1} &= \mathbf{y}_i + \tfrac{1}{2}(\mathbf{k}_1 + \mathbf{k}_2) \end{aligned} \tag{2.140}$$

Dans cet algorithme, on estime ainsi la "pente" comme la moyenne des estimations en début et fin d'intervalle.

On peut faire mieux encore avec l'algorithme de **Runge-Kutta d'ordre 4**, voir Fig. 2.18 :

$$\boxed{\begin{aligned} \mathbf{k}_1 &= \Delta t\, \mathbf{f}(\mathbf{y}_i, t_i) \\ \mathbf{k}_2 &= \Delta t\, \mathbf{f}(\mathbf{y}_i + \tfrac{1}{2}\mathbf{k}_1, t_{i+1/2}) \\ \mathbf{k}_3 &= \Delta t\, \mathbf{f}(\mathbf{y}_i + \tfrac{1}{2}\mathbf{k}_2, t_{i+1/2}) \\ \mathbf{k}_4 &= \Delta t\, \mathbf{f}(\mathbf{y}_i + \mathbf{k}_3, t_{i+1}) \\ \mathbf{y}_{i+1} &= \mathbf{y}_i + \tfrac{1}{6}\left[\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4\right] \end{aligned}} \tag{2.141}$$

Les schémas de Runge-Kutta du 4e ordre sont très utilisés dans toutes sortes d'applications de la physique et des sciences de l'ingénieur. Ce sont des schémas qui ont montré leur "robustesse", dans le sens qu'ils donnent de bons résultats dans la plupart des cas. Leur avantage principal réside dans la précision élevée obtenue avec relativement peu de pas temporels : la convergence est très rapide, à cause de l'ordre élevé du schéma. Mais il faut malgré tout faire attention : il y a des situations pour lesquelles ces schémas Runge-Kutta ne convergent pas du tout ou sont instables. Nous ne ferons pas l'analyse numérique de la convergence et de la stabilité des schémas Runge-Kutta dans ce cours, mais nous en ferons des applications dans la suite. De même que pour le schéma Runge-Kutta d'ordre 2, il existe une infinité de schémas de type Runge-Kutta d'ordre 4. On trouve dans la littérature plusieurs variantes de ces schémas.

2.9 Applications à divers systèmes oscillants

On considère un pendule à ressort amorti, puis un pendule simple excité et amorti, et enfin un pendule articulé. On observe dans certains cas l'apparition du chaos.

2.9.1 Pendule amorti

Um corps de masse m est attaché à un ressort de constante k. Il subit également une force de frottement visqueux $F_v = -\nu\, v$.

Equation du mouvement : $ma = F$, avec $F = -kx - \nu\, v \Rightarrow$

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + (\nu/m)\frac{\mathrm{d}x}{\mathrm{d}t} + (k/m)x = 0\,. \tag{2.142}$$

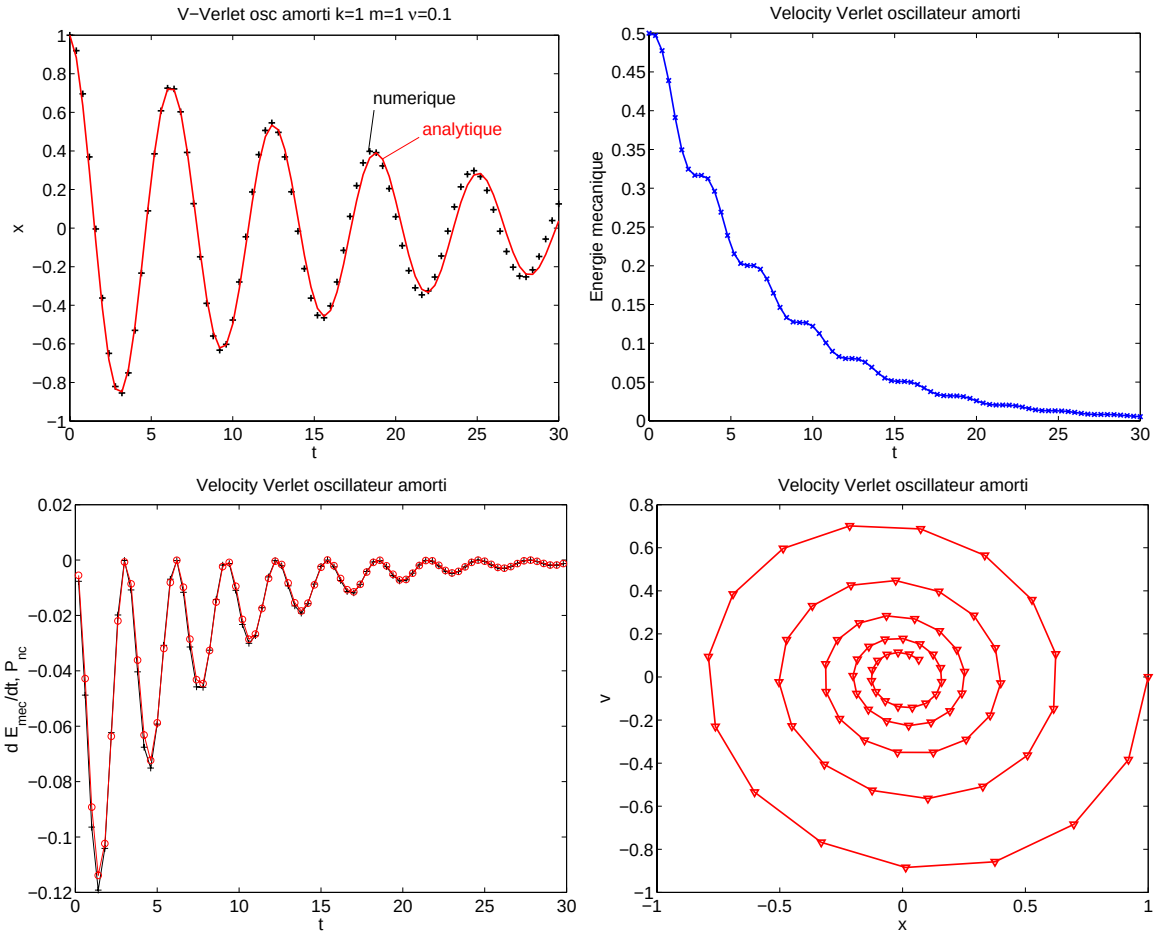


FIGURE 2.19 – *Oscillateur harmonique amorti, Eq.(2.142), avec la méthode de Verlet ("velocity Verlet", Eq.(2.106)),* $k = 1$, $m = 1$, $\nu = 0.1$, $\Delta t = 0.4$. *Position* $x(t)$ *(en haut à gauche), énergie mécanique* $E_{\rm mec}(t)$ *(en haut à droite), vérification du théorème de l'énergie mécanique* $\mathrm{d}E_{\rm mec}/\mathrm{d}t$ *(courbe avec +) et puissance des forces non conservatives* $P_{\rm nc} = -\nu\, v^2$ *(courbe avec o) (en bas à gauche), et orbite dans l'espace de phase* (x, v) *(en bas à droite).*

Nous utiliserons la méthode de "velocity Verlet", Eq.(2.106). La force dépend de la position *et de la vitesse*. Cela requiert une modification de l'algorithme (suggestion d'exercice). Un exemple de résultat est donné à la FIG. 2.19.

2.9.2 Pendule avec excitation extérieure. Résonance. Régime chaotique.

Soit un point matériel de masse m attaché à une tige mince rigide de masse négligeable, de longueur l, à un point O fixe. En plus de la gravitation, il est soumis à une force de frottement visqueux $F_v = -\kappa v$ et à un couple de force $M_E = M\sin(\Omega t)$. Et bien sûr, il y a la force de liaison (= force de soutien = "tension du fil"). En partant de

$$\frac{\mathrm{d}\vec{L}_O}{\mathrm{d}t} = \vec{M}_{\text{ext}} \ , \tag{2.143}$$

écrite en utilisant les coordonnées polaires (r, θ), on obtient

$$ml^2\ddot{\theta} = -lmg\sin\theta - l^2\kappa\dot{\theta} + M\sin(\Omega t) \ . \tag{2.144}$$

En posant $\nu = \kappa/m$, $A = M/(ml^2)$, on a

$$\boxed{\ddot{\theta} + \nu\dot{\theta} + \frac{g}{l}\sin\theta = A\sin(\Omega t)} \ . \tag{2.145}$$

Pour de petits mouvements ($\sin\theta \approx \theta$), $\omega_0 = \sqrt{g/l}$ est la fréquence propre du système libre ($A = 0$) non amorti ($\nu = 0$). Lorsqu'on excite le pendule avec un couple extérieur de fréquence $\Omega = \omega_0$, apparaît le phénomène de *résonance*, illustré à la FIG. 2.20. L'amplitude des oscillations est bien plus élevée que lorsque la fréquence d'excitation ne correspond pas à la fréquence propre. Le système accumule ainsi une plus grande quantité d'énergie mécanique.

Le mouvement devient vraiment intéressant pour des amplitudes d'excitation plus importantes. A la FIG. 2.21, on compare le mouvement périodique, régulier, obtenu avec $A = 1$ au mouvement irrégulier, "capricieux", **chaotique** avec une amplitude légèrement plus élevée, $A = 1.25$. Il ne s'agit pas d'une phase transitoire juste un peu plus longue : ce comportement chaotique se poursuit indéfiniment (image du bas de la FIG. 2.21).

2.9.3 Section de Poincaré. Attracteurs étranges. Divergence des orbites.

L'orbite dans l'espace de phase (θ, ω), où on a défini $\omega \equiv \dot{\theta} \equiv \mathrm{d}\theta/\mathrm{d}t$, FIG. 2.22, montre clairement que le pendule ne suit pas un mouvement régulier. On a l'impression que si

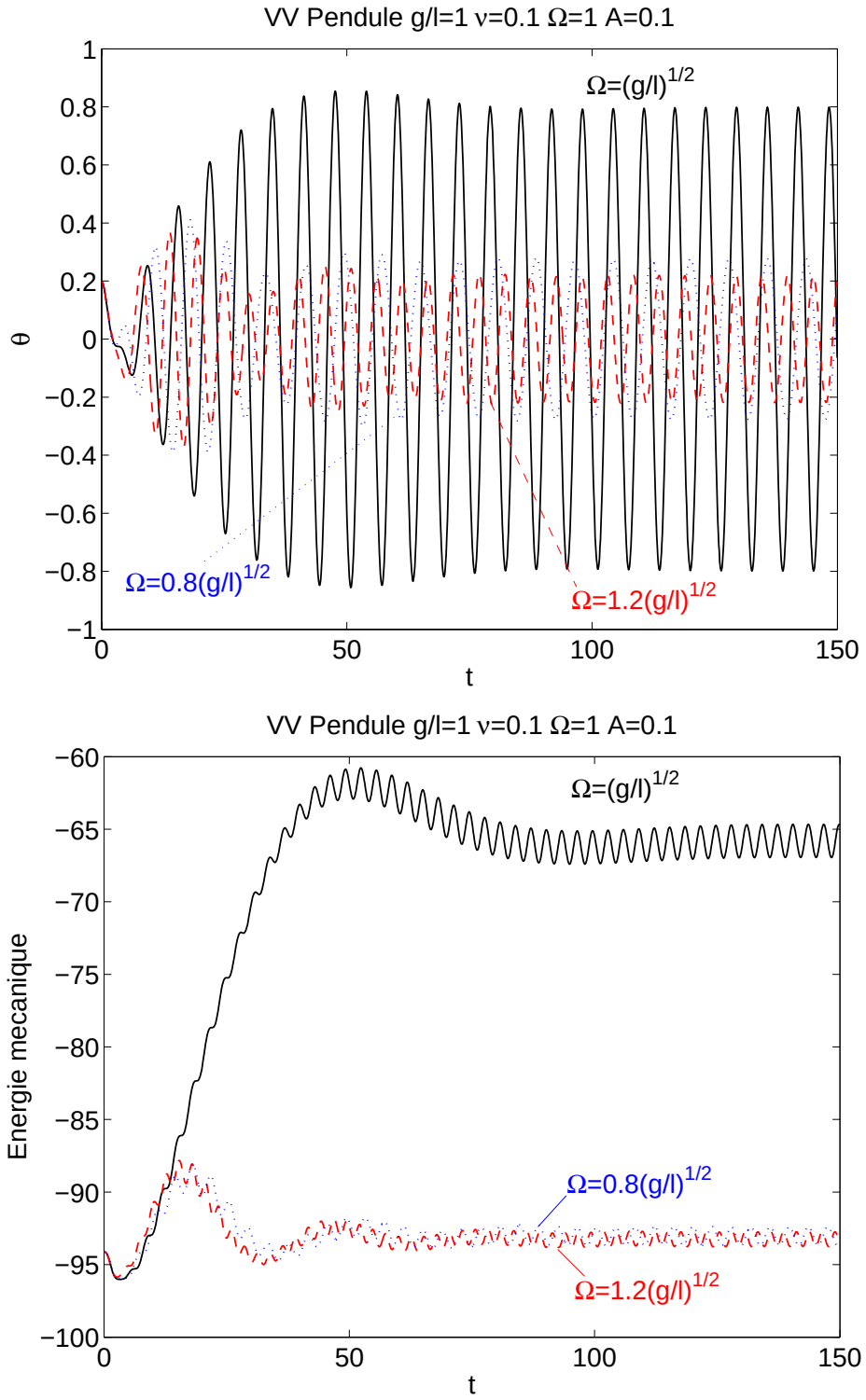


FIGURE 2.20 – *Résonance d'un pendule amorti excité , Eq.(2.145), avec la méthode de Verlet ("velocity Verlet", Eq.(2.106)), $g/l = 1$, $\nu = 0.1$, $A = 0.1$. Position $\theta(t)$ (en haut), énergie mécanique $E_{\rm mec}(t)$ (en bas), pour trois valeurs de Ω.*

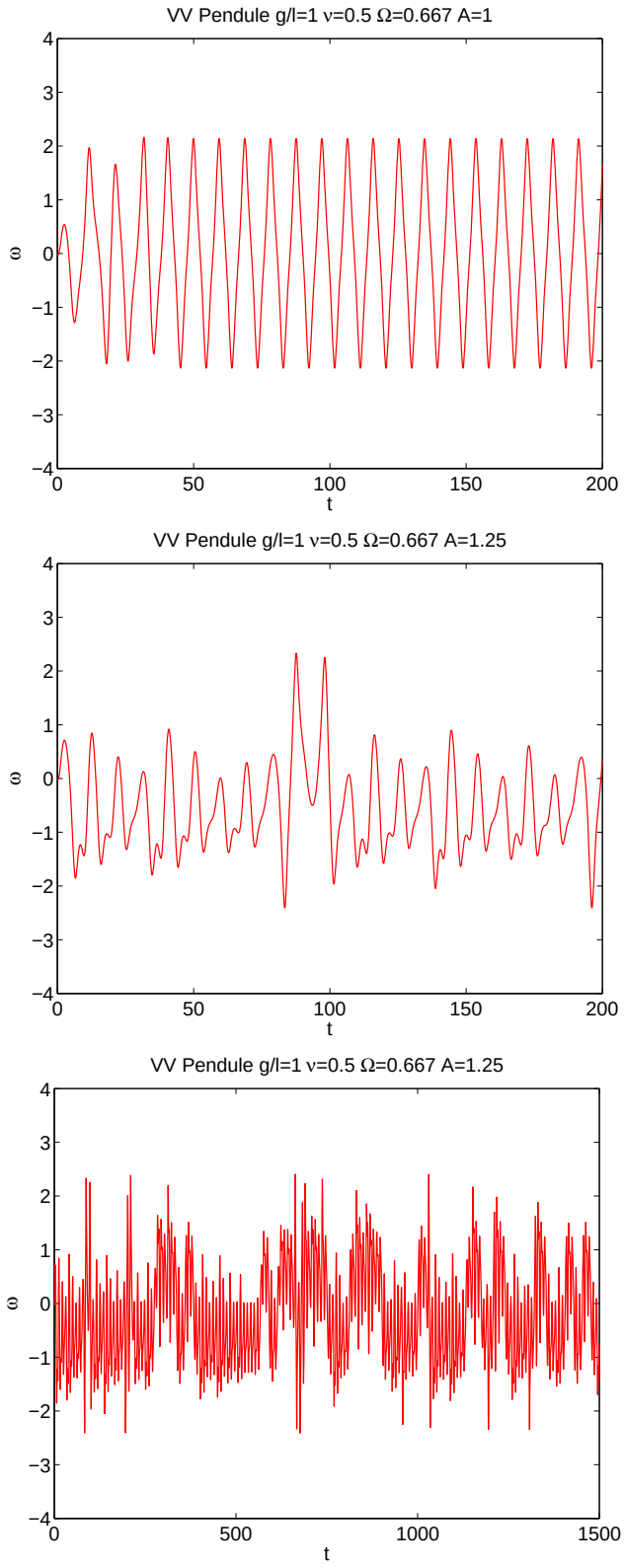


FIGURE 2.21 – *Transition vers le chaos d'un pendule amorti excité , Eq.(2.145), avec la méthode de Verlet ("velocity Verlet", Eq.(2.106)), $g/l = 1$, $\nu = 0.5$, $\Omega = 2/3$. On a représenté la vitesse angulaire $\omega(t) \equiv \dot{\theta}(t)$. Avec $A = 1.0$ (haut), le mouvement est régulier, périodique. Avec $A = 1.25$, le mouvement est* ***chaotique****.*

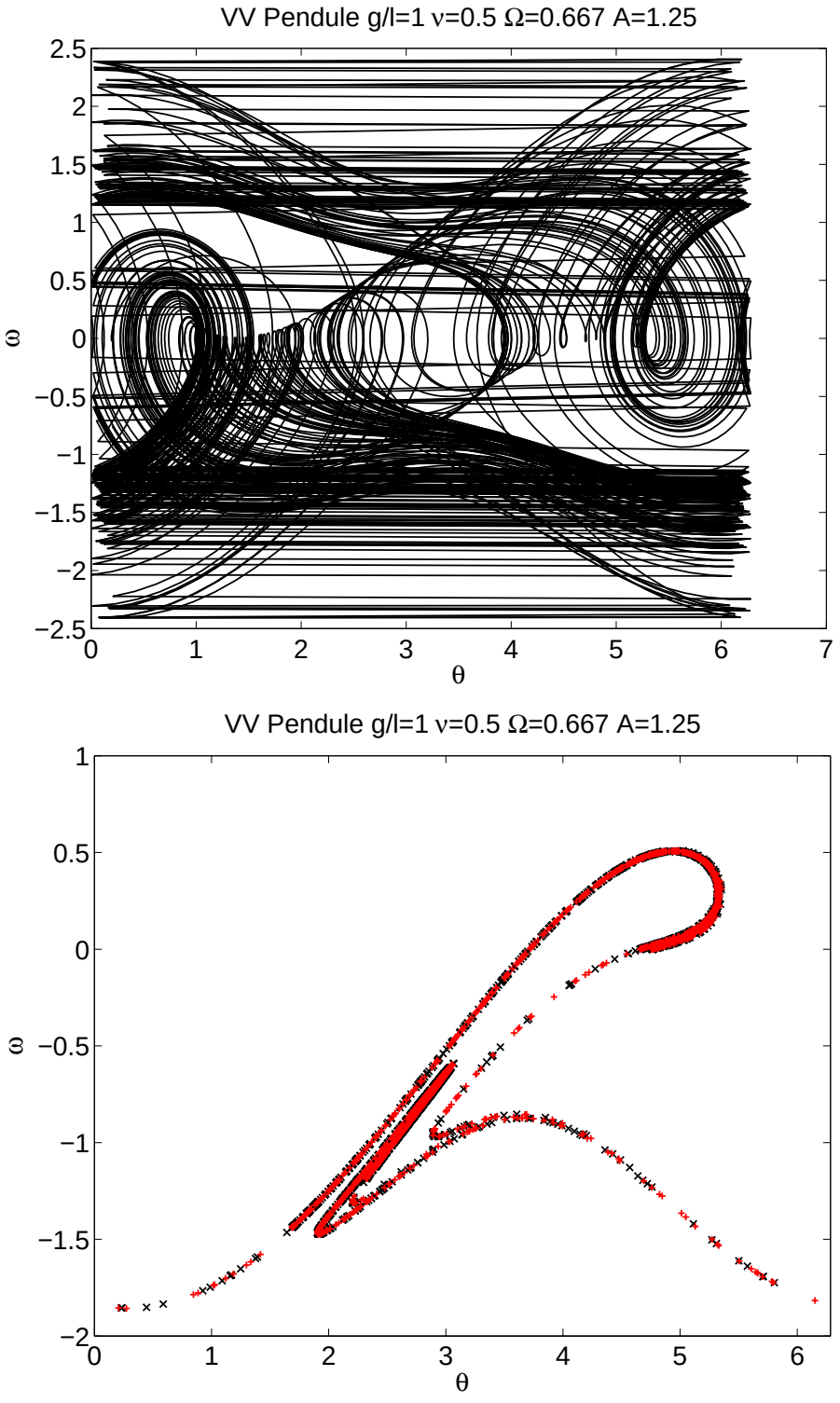


FIGURE 2.22 – *Mouvement d'un pendule amorti excité , Eq.(2.145), en régime chaotique, avec la méthode de Verlet ("velocity Verlet", Eq.(2.106)),* $g/l = 1$, $\nu = 0.5$, $\Omega = 2/3$, $A = 1.25$. *On a représenté l'orbite dans l'espace de phase* (θ, ω) *(haut) et la section de Poincaré (bas) pour deux conditions initiales très différentes :* $(\theta(0) = 1, \omega(0) = 0)$ *(points noirs) et* $(\theta(0) = 0, \omega(0) = 1)$ *(points rouges) .*

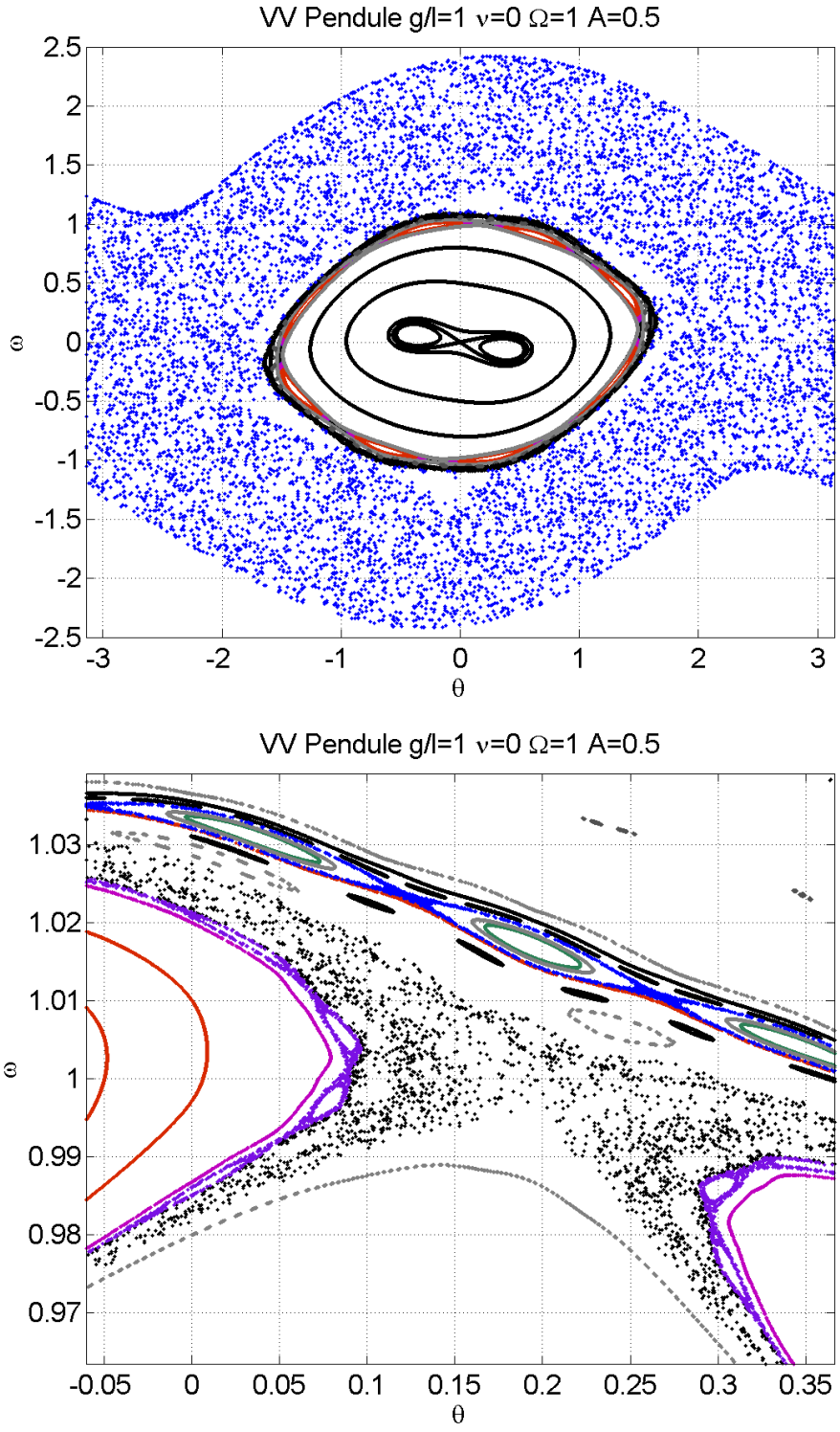


FIGURE 2.23 – *Sections de Poincaré pour diverses conditions initiales d'un pendule simple avec excitation verticale et sans amortissement, $g = l = 1$, $\nu = 0$, $\Omega = 1$, $A = 0.5$. L'image du bas est un zoom d'une région de l'image du haut.*

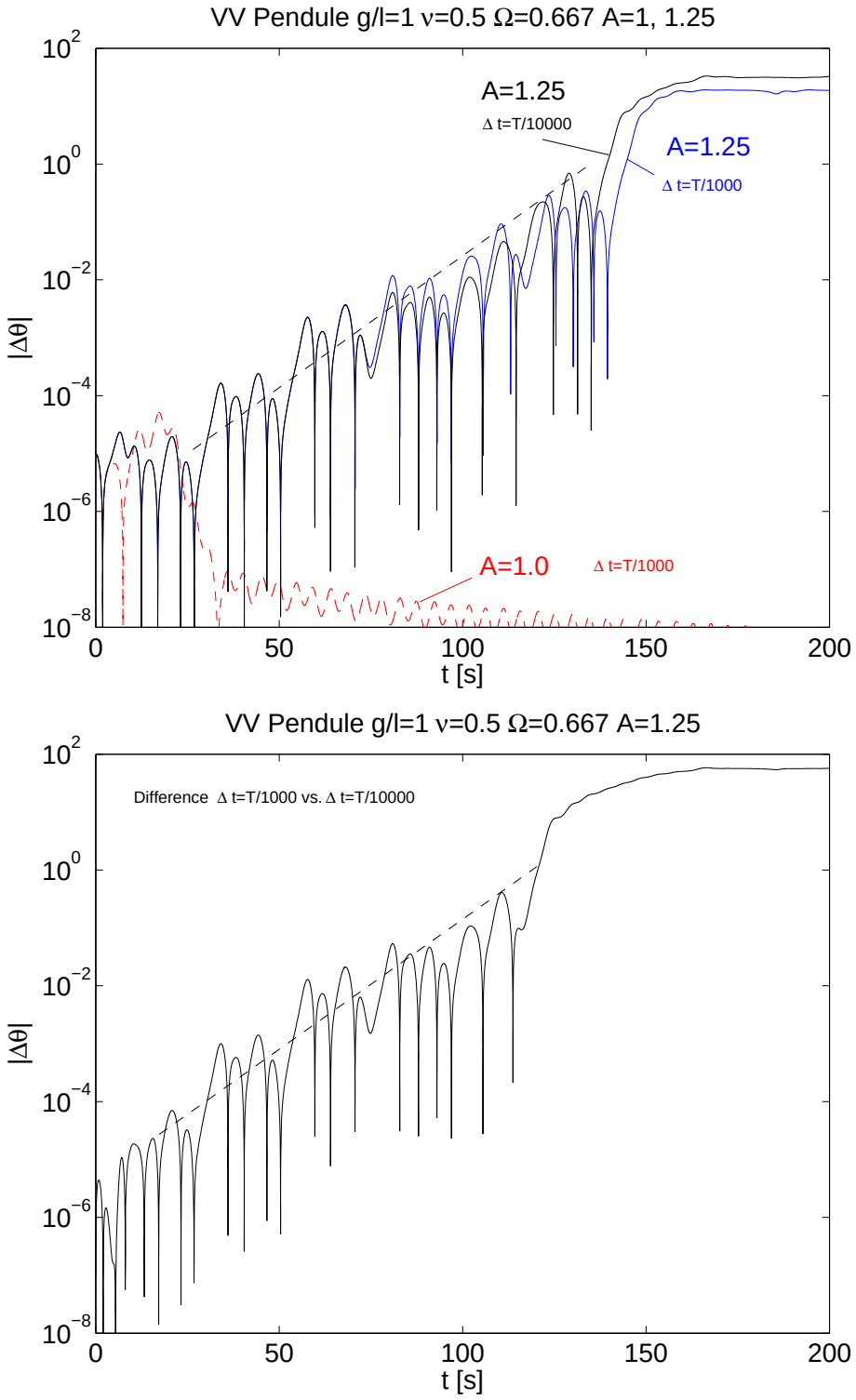


FIGURE 2.24 – *Caractérisation du mouvement chaotique d'un pendule amorti excité, Eq.(2.145), par la sensitivité aux conditions initiales. Haut : écart $|\Delta\theta|$ entre deux trajectoires obtenues à partir de conditions initiales qui diffèrent de 10^{-5} Pour comparaison, un régime non chaotique, avec $A = 1.0$, est représenté en traitillés : dans ce cas l'écart entre les trajectoires reste du même ordre que l'écart initial. Bas : écart $|\Delta\theta|$ entre 2 trajectoires obtenues à partir de la même condition initiale, mais deux valeurs différentes de Δt. Les droites traitillées soulignent le caractère exponentiellement divergent, dans le cas chaotique.*

on continue la simulation assez longtemps, l'orbite va finir par "remplir" uniformément l'espace de phase, tout point de cet espace pouvant être atteint par le pendule à un moment donné ou un autre, et ceci avec une densité de probabilité uniforme (c.a.d. la même probabilité quel que soit le point de l'espace de phase), ce qui serait le cas si le mouvement du pendule était parfaitement **aléatoire**. Mais la réalité est bien différente. Le mouvement du pendule n'est pas aléatoire ; bien qu'irrégulier, il y a une certaine "structure" au mouvement.

On peut s'en rendre compte graphiquement en ne représentant pas tous les points de l'orbite, mais seulement ceux aux temps multiples de la période d'excitation, donc l'ensemble des points $(\theta(t), \omega(t))$ tels que $t = n\,2\pi/\Omega$, avec n entier. On appelle ce type de représentation une **section de Poincaré**. On voit sur la FIG. 2.22 (en bas) qu'une structure apparaît : les points sont arrangés selon un ensemble de lignes de formes compliquées, que l'on appelle *attracteur étrange*. La notion *d'attracteur* se comprend à partir d'un exemple plus simple : celui du pendule amorti mais non excité. Danc ce cas, toute condition initiale va résulter en une solution $(\theta = 0, \omega = 0)$ pour des temps suffisamments longs. Le point $(\theta = 0, \omega = 0)$ "attire" toutes les conditions initiales, c'est l'attracteur, trivial, dans ce cas. Dans le cas qui nous intéresse, avec amortissment ET excitation, toute condition initiale va résulter en une section de Poincaré ayant la même structure de lignes complexes sur lesquelles les points vont se trouver. La FIG. 2.22 (en bas) montre en fait deux sections de Poincaré, l'une avec les points noirs, l'autre avec les points rouges, correspondants à deux conditions initiales très différentes. (Note technique : on a ignoré dans cette figure les 50 premiers points, i.e. les 50 premières périodes d'oscillation). Aucun point rouge ne coïncide exactement avec un point noir, et cependant les points noirs et rouges se rassemblent selon de la même structure de lignes. Cette structure "attire" toutes les conditions initiales, d'où le nom d'*attracteur*.

Pour compléter la discussion, lorsqu'il n'y pas *pas d'amortissement, il n'y a généralement pas d'attracteur*. Par exemple, pour un pendule simple sans excitation ni amortissement, le mouvement oscillatoire va continuer indéfiniment, et le point $(\theta = 0, \omega = 0)$ n'est plus un attracteur. Si on considère un pendule simple, avec excitation verticale [3] et sans amortissement, chaque condition initiale produit sa section de Poincaré généralement distincte des autres, voir FIG.2.23. L'espace de phase se sépare en régions avec des sections imbriquées les unes dans les autres, des régions avec des chaînes d'ilôts et des régions stochastiques où le mouvement est chaotique.

Une des "signatures" du mouvement chaotique est la sensibilité aux conditions initiales. Dans le régime chaotique, deux trajectoires obtenues à partir de conditions initiales *infinitésimalement voisines* finissent toujours par *diverger exponentiellement*. On l'illustre à la FIG. 2.24. Dans le cas non chaotique ($A = 1.0$, traitillés), les deux trajectoires restent proches l'une de l'autre tout au long du mouvement. Alors que dans le cas chaotique ($A = 1.25$), il y a toujours un moment où les trajectoires s'écartent

3. L'équation différentielle est alors quelque peu différente de l'Eq.(2.145) (suggestion d'exercice).

l'une de l'autre de plusieurs ordres de grandeur, aussi petit que soit l'écart entre les deux conditions initiales. Une conséquence de ce comportement est d'empêcher la convergence numérique avec Δt au dela de quelques secondes (image du bas).

On sait de la théorie des équations différentielles ordinaires qu'une fois que les conditions initiales sont spécifiées, la solution du mouvement est unique. Le système est dit *déterministe*. Mais comment se fait-il que le mouvement en régime chaotique soit *à la fois déterministe et pratiquement imprédictible ?*

Suggestion d'exercice. Avec les paramètres des FIGS.2.22-2.24, choisir différentes conditions initiales et montrer que l'attracteur étrange (section de Poincaré) est toujours le même. Choisir des Δt différents, pour une même condition initiale, et montrer que les trajectoires simulées finissent par s'écarter l'une de l'autre. Changer l'amplitude et essayez d'obtenir d'autres attracteurs. Essayez de trouver A pour que le mouvement devienne périodique avec une période multiple de la période d'excitation. (Indication : essayez $A \in [1.4\ 1.5]$ ou $A \in [1.6\ 1.8]$).

Suggestion d'exercice. Considérer un pendule de longueur l, masse m, avec amortissement, dont le point d'attache est un point O' mobile animé d'un mouvement oscillatoire vertical $y_{O'}(t) = d\sin(\Omega t)$.

a) Mettre le pendule "à l'envers" (condition initiale θ_0 proche de π). Etudier ce qui se passe lorsque le point d'attache est immobile ($d = 0$). Prendre ensuite les paramètres $l = 1\text{m}$, $\nu = 0.1$, $d = 0.3\text{m}$, $\Omega/2\pi = 3$. C'est le phénomène de *stabilisation non linéaire*.

b) Avec $l = 1\text{m}$, $\nu = 0.1$, $d = 0.07\text{m}$, choisissez une condition initiale autour de $\theta_0 = 0.15$ et variez la fréquence entre $\Omega/2\pi = 0.7$ et $\Omega/2\pi = 1.2$. Observez ce qui se passe autour de la fréquence $\Omega/2\pi = 1.0$: on obtient des oscillations de grande amplitude, dont la fréquence est la moitié de la fréquence d'excitation. Alors que pour des autres valeurs de la fréquence, les oscillations restent de petite amplitude. C'est le phénomène de *résonance paramétrique*.

2.9.4 Pendule articulé. Chaos dans un système conservatif.

On considère un système formé de deux tiges minces de longueurs L_1, L_2, masses m_1, m_2, attachées l'une à l'autre par une extrémité. Une des tiges est attachée à un point fixe O. On négligera les forces de frottement. Voir FIG. 2.25.

C'est un système conservatif à deux degrés de liberté. On choisira θ_1 et θ_2, les angles de chacune des deux tiges par rapport à la verticale.

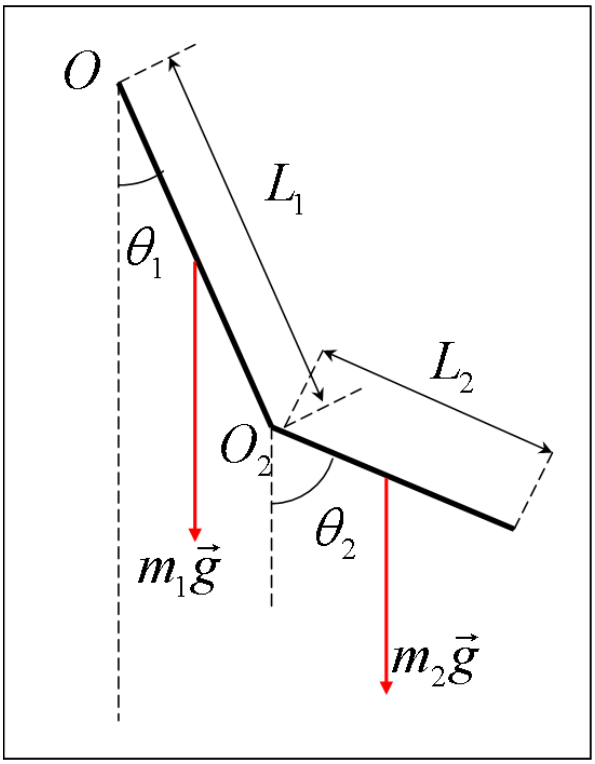


FIGURE 2.25 – *Pendule articulé.*

Les équations du mouvement peuvent s'obtenir soit directement à partir des équations de Newton (2e et 3e loi) et le théorème du moment cinétique. Ou encore par le Lagrangien, ou par l'Hamiltonien (voir cours de Mécanique Analytique). On préfèrera la méthode des équations de Lagrange (exercice).

[N.B. : L'Hamiltonien est en effet non séparable. Ceci a pour conséquence qu'il n'est pas aisé de trouver un algorithme symplectique. En particulier, l'application de l'algorithme de Verlet, Eq.(2.106) à ce problème (en exercice), montre que l'énergie mécanique n'est pas bien conservée, sauf pour les petits mouvements.]

On comparera la méthode de Verlet avec une méthode Runge-Kutta du 4e ordre. On montre aux FIGS.2.26, 2.27 et 2.28 le cas de 2 tiges de densité uniforme, de masses $m_1 = m_2 = 0.2$kg, et longueurs $L_1 = L_2 = 0.2$m. Pour des conditions initiales voisines de la position d'équilibre $\theta_1 = \theta_2 = 0$, le mouvement reste dans le voisinage, c'est un point d'équilibre stable. On observe un mouvement quasi-périodique de plus en plus complexe à mesure que l'on augmente les amplitudes des conditions initiales. A partir d'une certaine amplitude, le mouvement devient chaotique.

Suggestion d'exercice. On vérifiera une des caractéristiques du chaos, à savoir la sensibilité aux conditions initiales. On contrôlera la qualité de la simulation numérique en mesurant la conservation de l'énergie mécanique et en variant Δt.

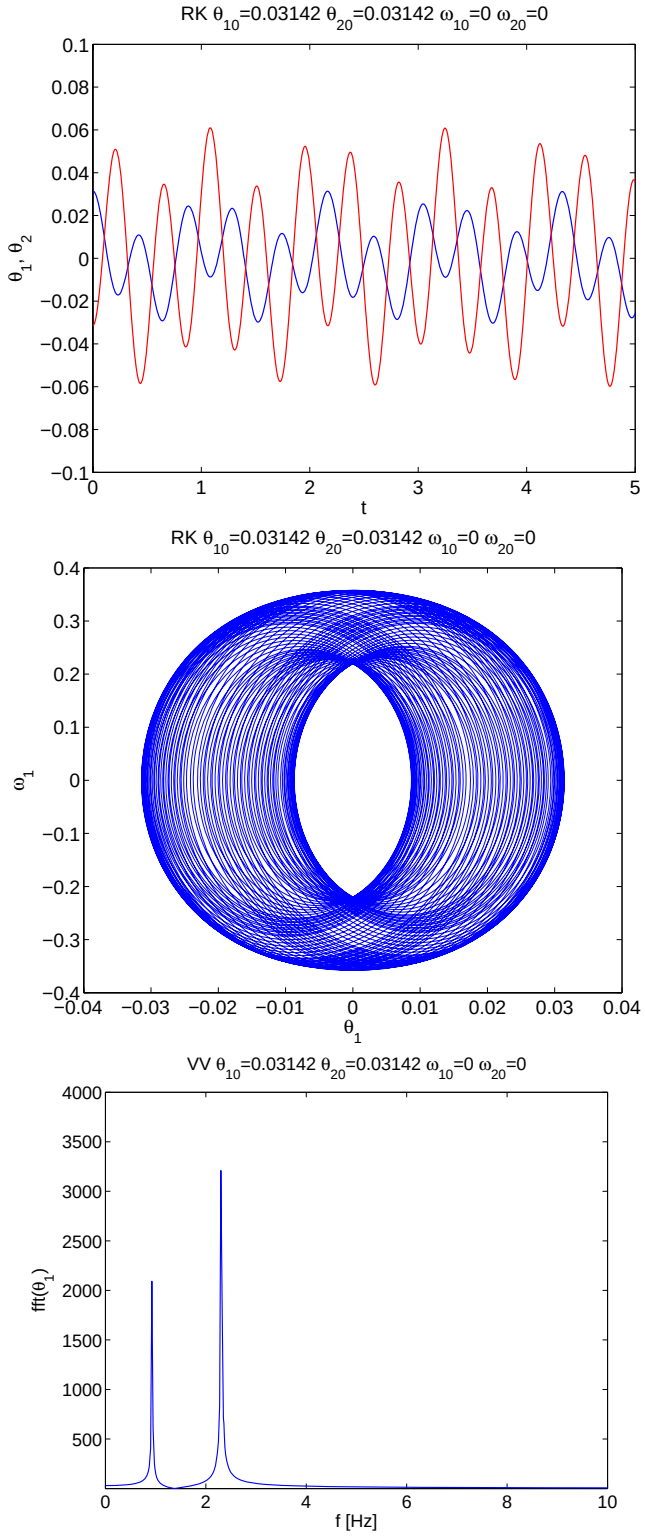


FIGURE 2.26 – *Mouvement du pendule articulé pour de faibles amplitudes, pour une condition initiale $\theta_{10} = -\theta_{20} = \pi/100$. : $\theta_1(t)$, $\theta_2(t)$ (en haut), (θ_1, ω_1) (au milieu), et analyse de la fréquence du signal (en bas). Le mouvement est constitué d'une superposition des 2 modes propres linéaires.*

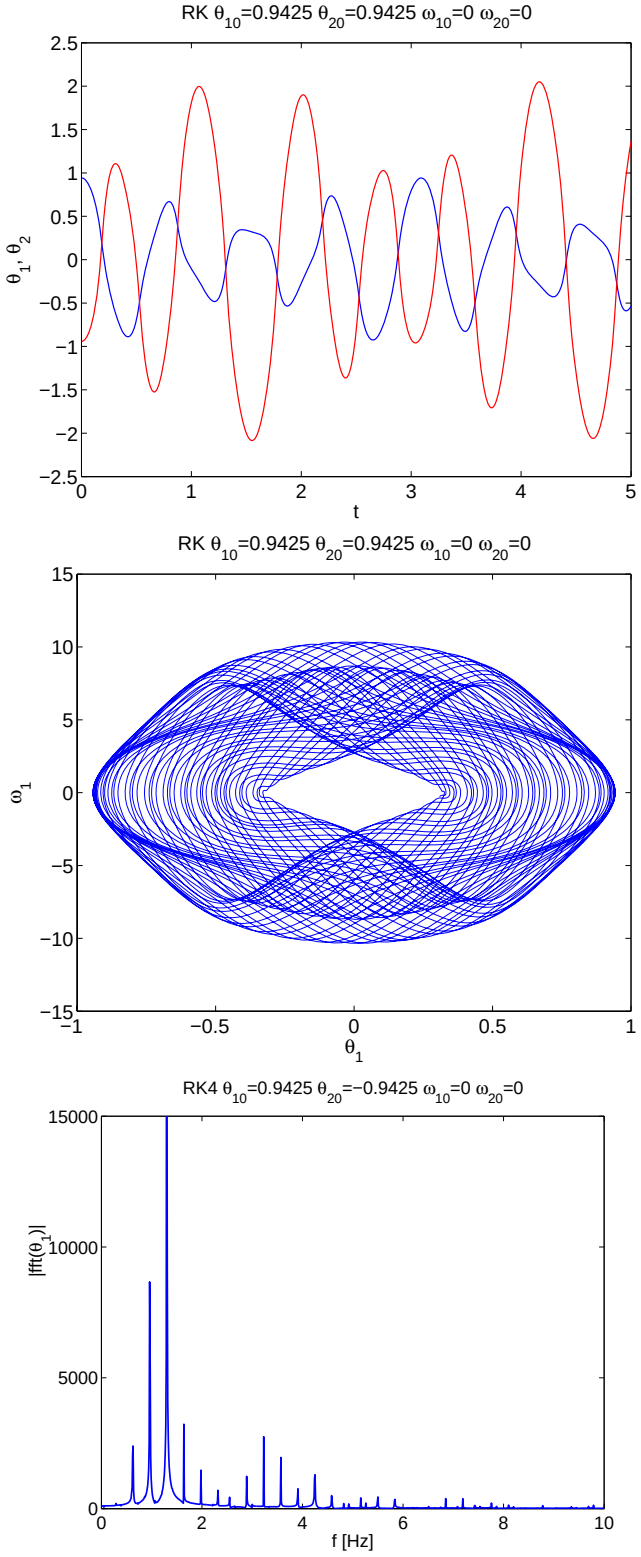


FIGURE 2.27 – *Mouvement du pendule articulé pour une condition initiale* $\theta_{10} = -\theta_{20} = 0.3\pi$. $\theta_1(t)$, $\theta_2(t)$ *(en haut),* (θ_1, ω_1) *(au milieu), et analyse de la fréquence du signal (en bas). Le mouvement apparaît comme une superposition de plusieurs fréquences dues aux effets de couplage non linéaire.*

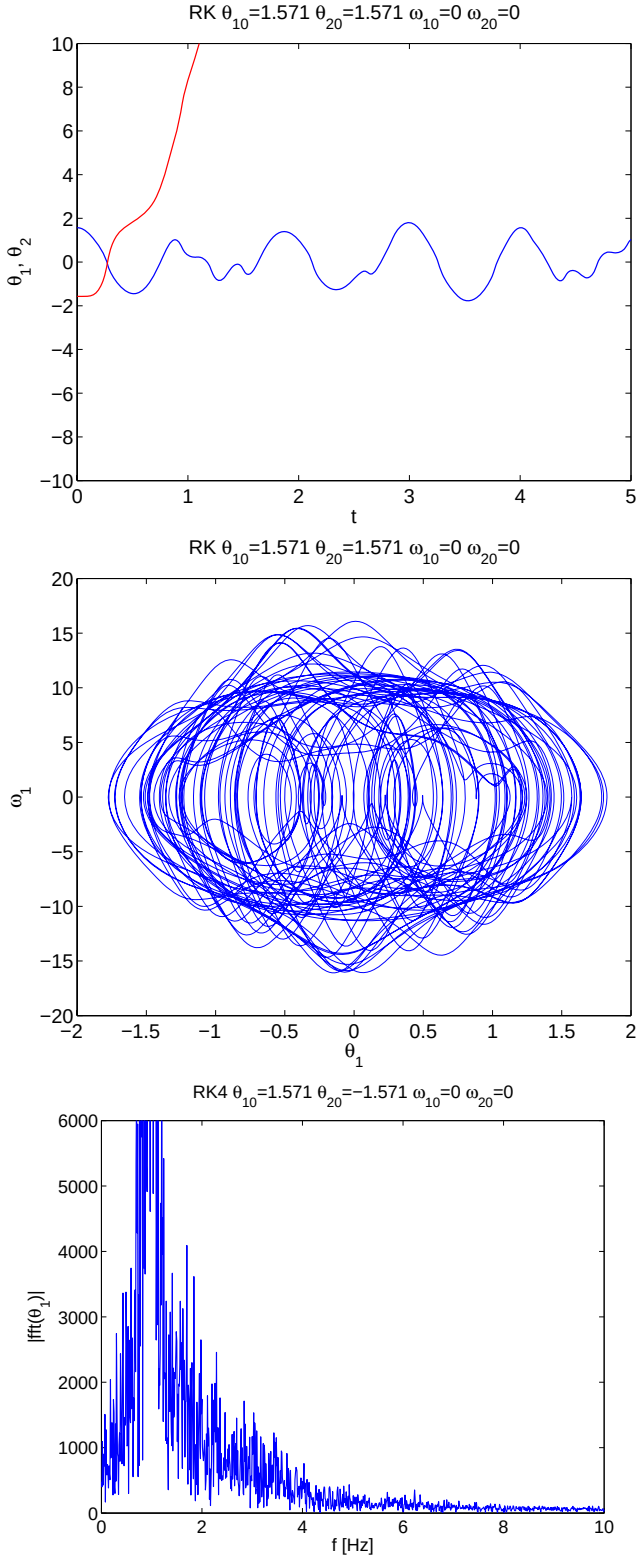


FIGURE 2.28 – *Mouvement du pendule articulé pour une condition initiale* $\theta_{10} = -\theta_{20} = \pi/2$. $\theta_1(t)$, $\theta_2(t)$ *(en haut),* (θ_1, ω_1) *(au milieu), et analyse de la fréquence du signal (en bas). Le mouvement est chaotique.*

Pour de faibles amplitudes, l'algorithme de Verlet conserve bien l'énergie mécanique en moyenne temporelle, mais ce n'est plus le cas lorsque l'amplitude devient plus importante. L'algorithme de Runge Kutta du 4e ordre s'avère alors plus performant.

2.10 Gravitation. Schémas adaptatifs

2.10.1 Généralités : 1 ou 2 corps - mais pas plus

Les problèmes de mouvements gravitationnels (force en $1/r^2$) sont abordables par des méthodes analytiques pour 1 ou 2 corps. Dès que le système considéré comporte 3 corps ou plus, les choses deviennent extrêmement complexes et il n'y a pas de solution exacte.

Les méthodes numériques, cependant, peuvent assez aisément se généraliser à un nombre de corps quelconque. Nous allons en montrer quelques exemples. Dans cette section, on vérifiera la précision des simulations pour des cas à 1 ou 2 corps en comparant les résultats avec les solutions analytiques exactes.

Soit un système de N corps de masses $m_i, i = 1..N$. Les équations du mouvement s'obtiennent de la 2e loi de Newton pour chacun des corps :

$$m_i \frac{\mathrm{d}^2 \vec{x}_i}{\mathrm{d}t^2} = \sum_{j \neq i}^{N} -\frac{G m_i m_j}{r_{ij}^3} \vec{r}_{ij} \,, \qquad \vec{r}_{ij} = \vec{r}_i - \vec{r}_j \,. \tag{2.146}$$

Le système est conservatif, avec des forces ne dépendant pas de la vitesse et dérivant d'un potentiel. Les algorithmes d'Euler-Cromer et de Verlet sont appropriés à ce genre de situation. On peut aussi utiliser l'algorithme de Runge-Kutta.

On simule l'orbite terrestre, sachant que la distance terre-soleil est au minimum $r_{\min} = 147098074$km, au maximum $r_{\max} = 152097701$km, et que la vitesse de la terre est au minimum $v_{\min} = 29.291$km/s, au maximum $v_{\max} = 30.287$km/s. On rappelle que l'orbite est une ellipse. Le moment cinétique est conservé, ce qui veut dire que $r_{\max} v_{\min} = r_{\min} v_{\max}$. Avec la méthode Euler-Cromer, on montre les résultats à la FIG. 2.29. L'orbite est presque circulaire, l'ellipticité étant à peine visible sur la trajectoire dans le plan (x, y). En représentant $r(t)$, la non-circularité est bien visible. Par une étude de convergence avec Δt, on peut obtenir des valeurs précises pour la période, les distances minimales et maximales, etc.

Suggestion d'exercice. Trouver la masse du soleil, en utilisant l'intégration numérique d' Euler-Cromer (ou Verlet ou Runge Kutta), connaissant $r_{\min} = 147098074$km et $v_{\max} = 30.287$km/s. On ajustera la masse du soleil jusqu'à trouver la bonne période d'un an $= 365.2564$ jours et le bon $r_{\max} = 152097701$km.

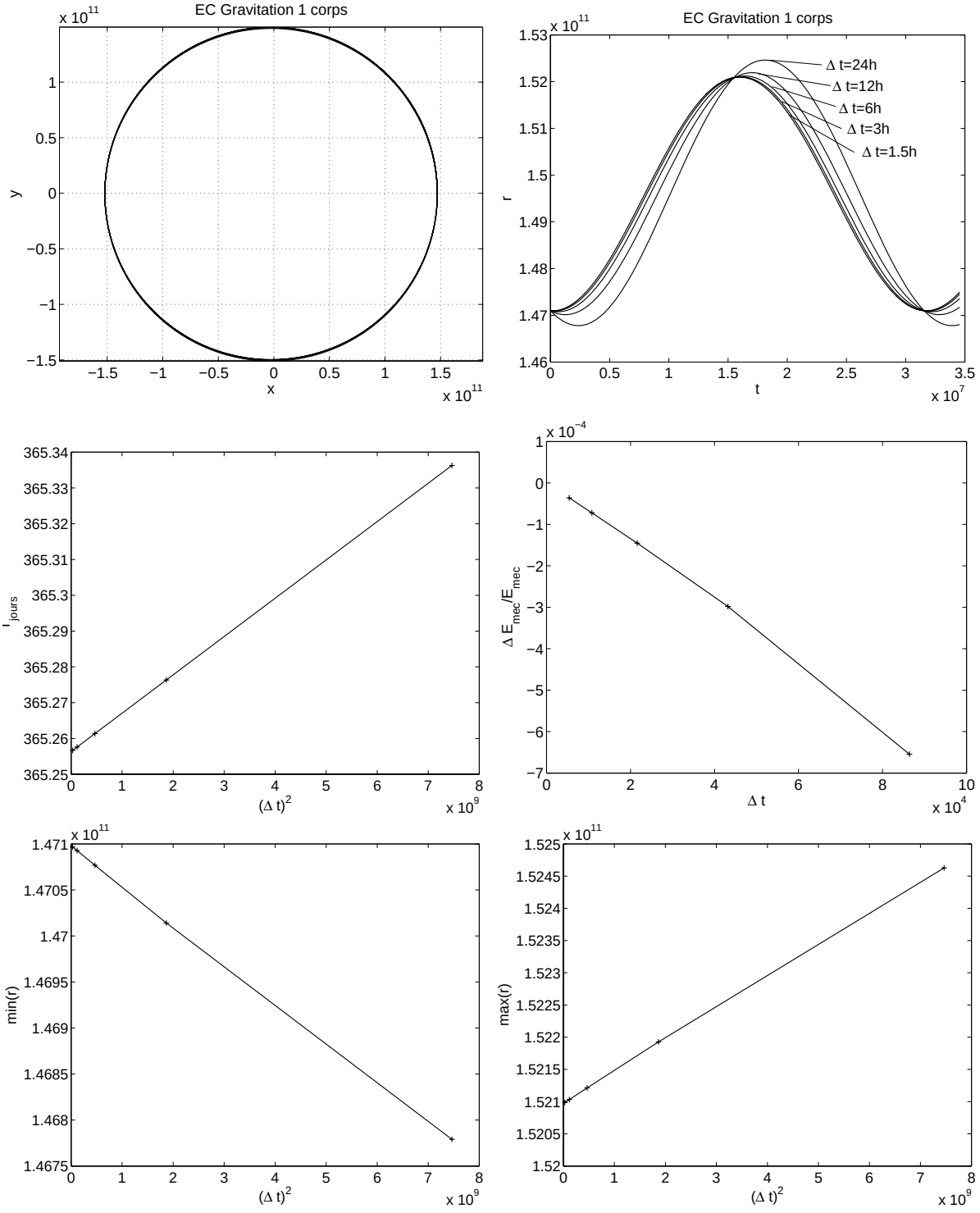


FIGURE 2.29 – *Simulation de l'orbite terrestre avec l'algorithme d' Euler-Cromer. En haut à gauche : vue dans le plan (x, y). En haut à droite : distance terre-soleil en fonction du temps, 5 simulations à des valeurs de Δt différentes. Au milieu, convergence de la période (gauche) et de la conservation de l'énergie mécanique (droite). En bas, convergence de la distance terre-soleil minimale (gauche) et maximale (droite).*

Suggestion d'exercice. Vérifier la 3e loi de Kepler pour diverses planètes et comètes du système solaire. Rappel : $T^2/a^3 = \text{const}$, où T est la période de révolution et a le demi grand-axe.

Suggestion d'exercice. Considérer un système gravitationnel à 2 corps. Vérifier que le mouvement relatif $\vec{r}_2 - \vec{r}_1$ est équivalent au mouvement à 1 corps, mais avec la *masse réduite* $\mu = m_1 m_2/(m_1+m_2)$. Vérifier que le centre de masse du système a un mouvement rectiligne uniforme. Représenter les trajectoires dans le référentiel du centre de masse.

Suggestion d'exercice. Etudier ce qui se passerait si la force de gravitation était en $1/r^\beta$, et examiner 3 cas : $\beta < 1$, $\beta = 2.5$ et $\beta = 3$.

Suggestion d'exercice. Précession de l'orbite de Mercure (effet de relativité générale) : on modélise cet effet par une force

$$F \approx -G\frac{m_1 m_2}{r_{12}^2}\left(1 + \frac{\alpha}{r^2}\right) \ . \tag{2.147}$$

Considérer d'abord le cas $\alpha = 0$, et trouver les conditions initiales pour Mercure, sachant que le demi grand axe est $a = 0.39$ AU, et l'excentricité $e = 0.206$. Indications : 1 AU= distance moyenne terre-soleil = 149 597 870 691 m. $r_{\max} = (1+e)a$, $r_{\min} = (1-e)a$. Trouver $v_{\max}$ et $v_{\min}$.

Prendre ensuite $\alpha = 0.005\text{AU}^{-2}$ et calculer l'angle θ au périhélie de plusieurs révolutions successives, puis en déduire la vitesse de précession $\mathrm{d}\theta/\mathrm{d}t$. Prendre des valeurs décroissantes de α et extrapoler la valeur de $\mathrm{d}\theta/\mathrm{d}t$ pour la valeur physique de $\alpha = 1.1 \times 10^{-8}\text{AU}^{-2}$. Comparer avec la valeur mesurée de 43 secondes d'arc par siècle.

Suggestion d'exercice. Engin spatial avec poussée, transitions d'orbites.

Suggestion d'exercice. Comparer les algorithmes Runge-Kutta d'ordre 2, Runge-Kutta d'ordre 4, Euler-Cromer et Verlet pour longues simulations d'une orbite gravitationnelle. Contrôler la précision de la conservation de l'énergie, et celle de la conservation du moment cinétique.

2.10.2 Problème à 3 corps

Les systèmes gravitationnels que l'on peut résoudre exactement avec des méthodes analytiques se limitent aux problèmes à 1 ou 2 corps. Pour 1 corps, on suppose un des deux objets célestes de masse beaucoup plus élevée que l'autre, et ainsi on le suppose fixe. On obtient un mouvement central en $1/r^2$, avec les lois de Kepler : I- trajectoires côniques (ellipse, parabole ou hyperbole) avec le corps central à l'un des foyers ; II- loi des aires

(conservation du moment cinétique) ; III- le rapport des carrés des périodes de révolution est égal au rapport des cubes des demi-grands axes dans le cas d'orbites elliptiques. Voir cours de Physique de 1ere année.

Si les 2 corps célestes ont des masses comparables (étoile double, par exemple), le mouvement relatif, $\vec{r}_{12} = \vec{r}_2 - \vec{r}_1$, obéit formellement aux mêmes équations que le problème à un corps, mais avec la *masse réduite* $\mu = m_1 m_2/(m_1 + m_2)$ au lieu de la masse d'un des 2 corps. Une autre façon d'aborder le problème à 2 corps est de se placer dans le référentiel du centre de masse ; chacun des 2 objets célestes obéira au lois de Kepler, mais où le centre de masse est un des foyers de chacune des côniques.

Mais le mode réel a bien plus de 2 corps célestes. Nous allons examiner ce qui se passe avec 3 corps célestes.

Orbite terrestre dans un système d'étoiles doubles

Examinons d'abord l'effet de Jupiter sur l'orbite de la Terre. On considère donc 3 corps : la terre, Jupiter, et le soleil, que l'on considèrera comme des points matériels (particules). Les équations du mouvement s'obtiennent de la 2e loi de Newton :

$$m_\alpha \frac{\mathrm{d}\vec{v}_\alpha}{\mathrm{d}t} = \sum_{\beta\neq\alpha} \vec{F}_{\alpha\beta}, \quad \vec{F}_{\alpha\beta} = -\frac{G m_\alpha m_\beta}{r^3}\left(\vec{r}_\alpha - \vec{r}_\beta\right) \ , \quad \alpha,\beta = 1,2,3, \tag{2.148}$$

où $\vec{F}_{\alpha\beta}$ est la force exercée sur la particule α par la particule β.

Comme schéma numérique, comme il s'agit d'un système conservatif, on utilisera l'algorithme de "Velocity Verlet", Eq.(2.106). On peut en principe utiliser un algorithme encore plus simple, Euler-Cromer par exemple, mais la précision n'est pas suffisante et contraint à prendre des Δt trop petits.

Plutôt que d'utiliser les unités S.I., on écrit les équations dans le système d'unités normalisées suivant. L'unité de longueur est l' Unité Astronomique, [UA], définie comme le demi grand-axe de l'orbite terrestre autour du soleil, qui vaut 149.597871 millions de km, et 1 année comme unité de temps, soit 365.256898 jours.

Si on prend les valeurs réelles des masses de la terre, de Jupiter et du soleil, on constate que l'effet de Jupiter sur l'orbite terrestre est extrêmement petit. L'orbite terrestre est stable, ce qui n'est pas vraiment une surprise, étant donné les 4.6 millards d'années que cela dure.

Supposons maintenant que Jupiter soit de masse plus élevée. En multipliant la masse réelle de Jupiter par un facteur de 300, on obtient les résultats de la FIG. 2.30. Les orbites du soleil et de "Jupiter" sont des ellipses pratiquement circulaires avec un des

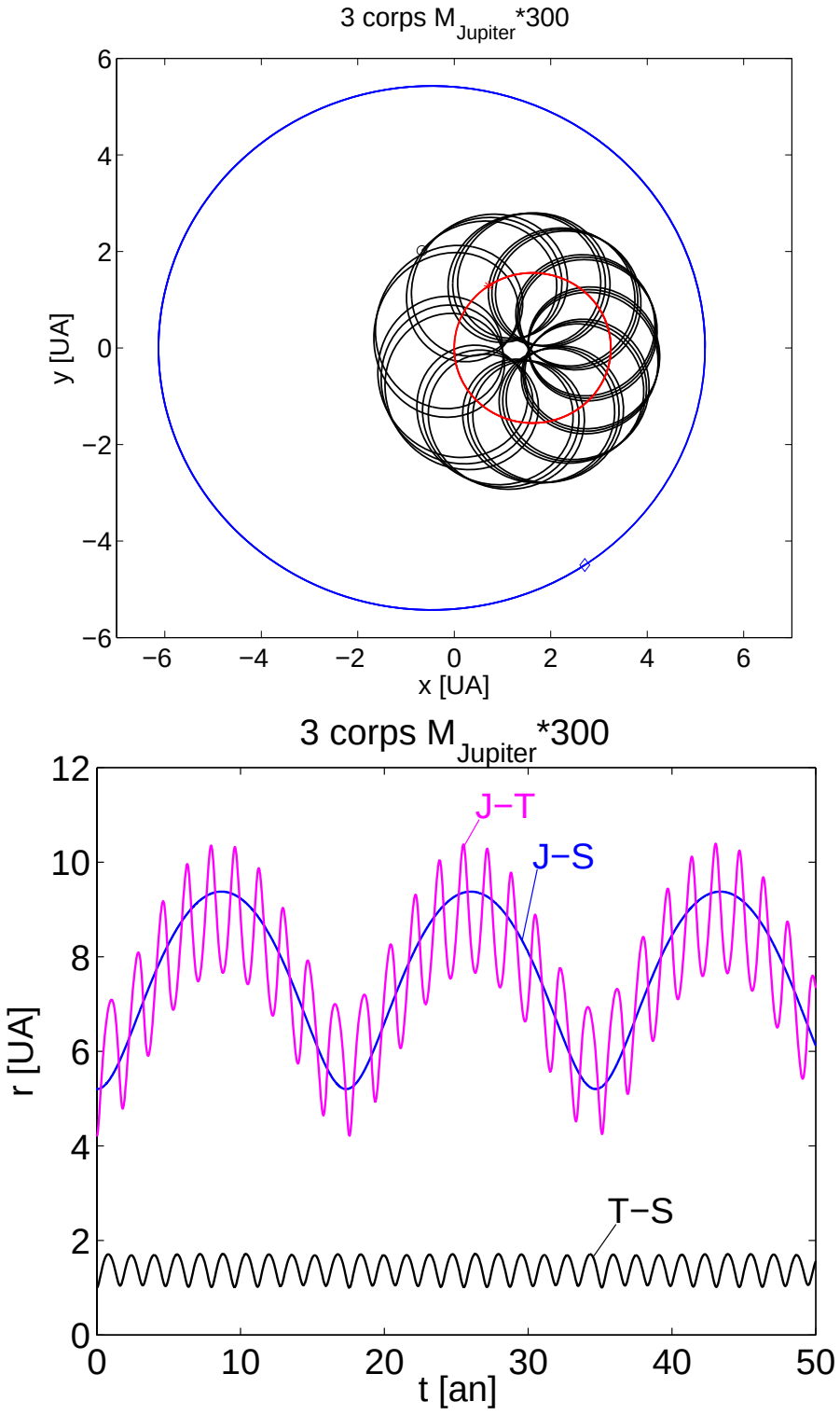


FIGURE 2.30 – *Trajectoires du soleil (rouge, étoile), de "Jupiter" s'il avait 300 fois sa masse réelle (bleu, diamant) et de la terre (noir, cercle). Algorithme de Störmer-Verlet (Velocity Verlet), Eq.(2.106), avec* $\Delta t = 0.01$an. *En bas, l'évolution temporelle des distances entre les 3 corps est représentée.*

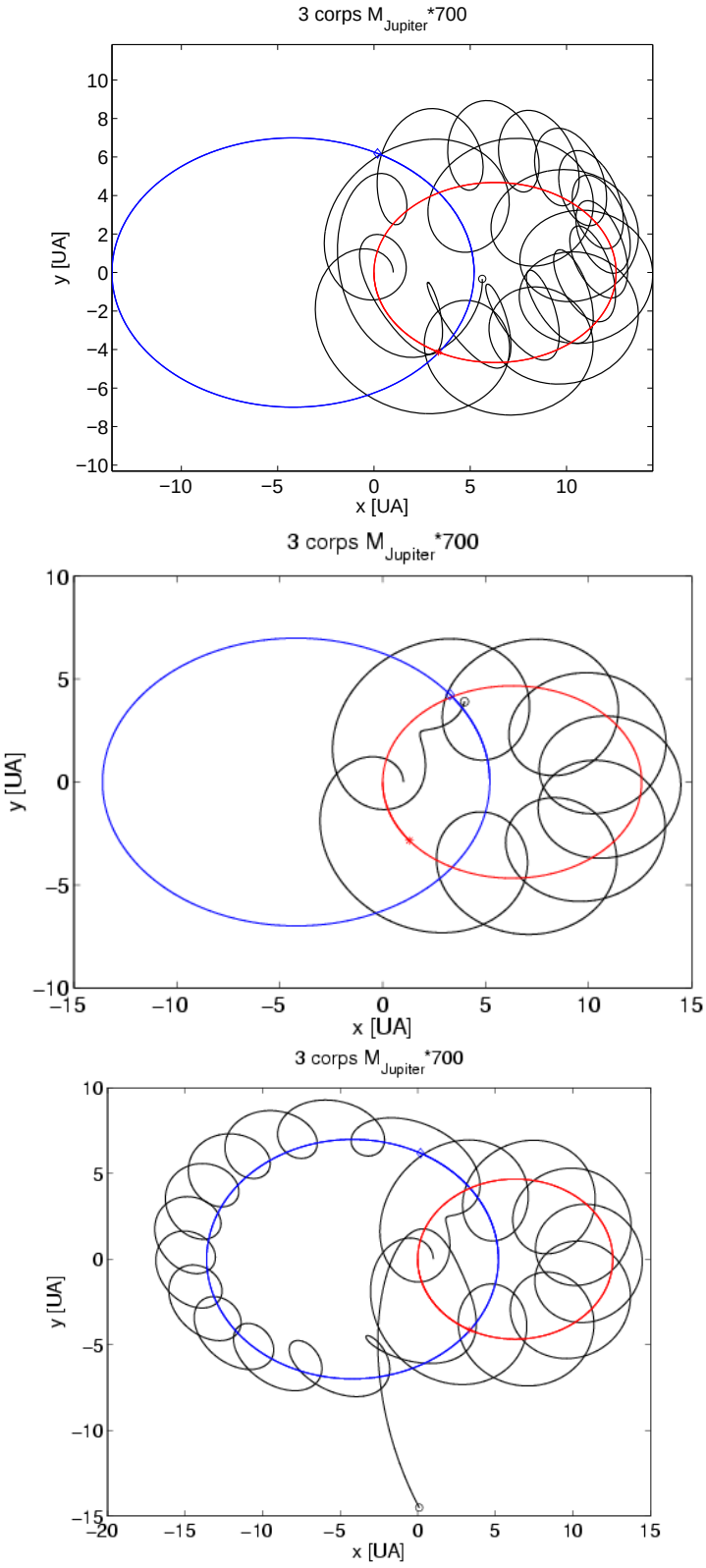


FIGURE 2.31 – *Trajectoires du soleil (rouge, étoile), de "Jupiter" s'il avait 700 fois sa masse réelle (bleu, diamant) et de la terre (noir, cercle). Algorithme de Störmer-Verlet (Velocity Verlet), Eq.(2.106), avec* $\Delta t = 0.01$an *(en haut) et* $\Delta t = 0.001$an *(au milieu et en bas). Il y a capture de la terre par Jupiter. Durée de 50 ans (en haut et au milieu) et de 100 ans (en bas).*

foyers au centre de masse du sous-système {soleil,"Jupiter"}. On a en effet, pour cette simulation, choisi les conditions initiales du soleil et de "Jupiter" de telle sorte que la vitesse du centre de masse du sous-système soit nulle. Pour ces paramètres, la masse de la terre est beaucoup plus petite que les deux autres corps, et ainsi le mouvement du sous-système {soleil,"Jupiter"} est pratiquement un mouvement à deux corps. Le mouvement de la terre est par contre fortement différent ! Il serait très difficile de vivre sur cette terre-là : la distance terre-soleil varierait beaucoup au cours de l'année.

Si Jupiter était encore plus massique, la situation serait vraiment catastrophique pour nous. On montre à la FIG. 2.31 un résultat avec la masse réelle de Jupiter multipliée par 700. La trajectoire de la terre devient chaotique. Elle est par moments "capturée" par "Jupiter". Elle entre presque en collision avec le soleil ou avec "Jupiter". Il est intéressant de constater la grande sensibilité de la simulation : l'image du haut a été obtenue avec $\Delta t = 0.01$an, celle du milieu avec $\Delta t = 0.001$an. La durée physique (50 ans) simulée, et les conditions initiales sont les mêmes. Cependant, dans la première simulation, la terre se trouve toujours proche du soleil, alors que la deuxième simulation, plus précise, prédit presque une collision avec "Jupiter".

2.10.3 Schémas adaptatifs : pas d'intégration variable

Une analyse plus fine des résultats numériques de la section précédente montre que les erreurs s'accumulent surtout lorsque 2 des corps sont proches l'une de l'autre. C'est à ce moment-là que leur vitesse et leur accélération sont les plus élevées.

C'est la raison pour laquelle il faut un pas temporel Δt suffisamment petit. Mais prendre un Δt petit est très coûteux en temps de calcul. De plus, la plupart du temps, les corps célestes sont assez éloignés et un Δt assez grand suffit à garantir une certaine précision.

L'idée est donc de **choisir le pas temporel en l'ajustant dynamiquement au cours de la simulation, afin de garantir un niveau de précision donné.** Les algorithmes "single full timestep", comme le Velocity Verlet, Eq.(2.106), et les algorithmes de Runge-Kutta, Eqs.(2.134-2.141), se prêtent relativement facilement à cette modification.

La stratégie est, au pas temporel t_i, étant donné la solution $\mathbf{y}_i$, de faire 2 estimations de la solution au pas temporel suivant $t = t_i + \Delta t$. Voir Fig.2.32. La première estimation, $\mathbf{y}_{i+1}^{(1)}$, s'obtient en faisant un pas temporel entier Δt. La deuxième estimation, $\mathbf{y}_{i+1}^{(2)}$, s'obtient en faisant un premier demi-pas pour aller de t_i à $t_i + \Delta t/2$, suivi d'un deuxième demi-pas pour aller de $t_i + \Delta t/2$ à $t_i + \Delta t$. On mesure la valeur absolue de la différence obtenue, d, et on la compare à une précision requise donnée, ϵ.

— Si $d < \epsilon$, on passe au pas de temps suivant, en essayant de l'augmenter par rapport au pas de temps actuel.

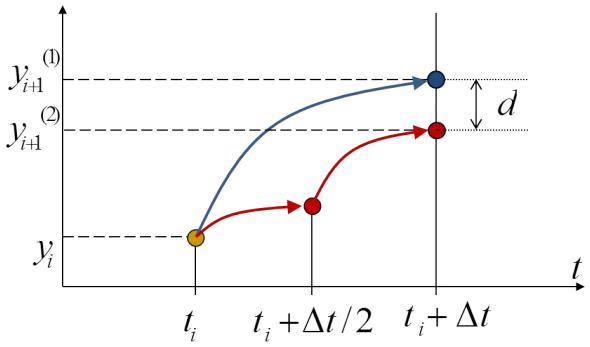


FIGURE 2.32 – *Principe du schéma adaptatif. On obtient une estimation de l'erreur en mesurant la différence d entre deux évaluations de la solution en $t + \Delta t$.*

— Si $d > \epsilon$, il faut recommencer, c'est-à-dire *revenir* au temps t_i, choisir un pas de temps plus court, et refaire les deux évaluations comme décrit au paragraphe ci-dessus pour obtenir une nouvelle évaluation de la différence d. Tant que $d > \epsilon$, on doit recommencer.

La question est donc : comment choisir le pas de temps (plus long ou plus court), étant donné une mesure de l'écart d ? La réponse à cette question dépend de *l'ordre du schéma numérique considéré*. Supposons que le schéma soit convergent au voisinage de $(\mathbf{y}_i, t_i)$, d'ordre n. Cela implique donc :

$$\mathbf{y}_i^{(1)} = \mathbf{y}_{\text{exact}} + \mathbf{C}_1 (\Delta t)^{n+1} + \mathcal{O}(\Delta t)^{n+2} \quad (2.149)$$

$$\mathbf{y}_i^{(2)} = \mathbf{y}_{\text{exact}} + \mathbf{C}_{2a} (\Delta t/2)^{n+1} + \mathbf{C}_{2b} (\Delta t/2)^{n+1} + \mathcal{O}(\Delta t)^{n+2} \quad (2.150)$$

avec $\mathbf{C}_1, \mathbf{C}_{2a}, \mathbf{C}_{2b}$ des vecteurs de constantes. Pour Δt suffisamment petit, on peut faire l'hypothèse simplificatrice $\mathbf{C}_1 = \mathbf{C}_{2a} = \mathbf{C}_{2b} = \mathbf{C}$, qui veut dire que la vitesse de convergence est la même dans le voisinage considéré de $(\mathbf{y}_i, t_i)$. Négligeant $\mathcal{O}(\Delta t)^{n+2}$, on obtient donc

$$d = |\mathbf{y}_i^{(1)} - \mathbf{y}_i^{(2)}| = |C| (\Delta t)^{n+1} \left(1 - \frac{1}{2^n}\right) \quad (2.151)$$

On veut choisir une nouvelle valeur de Δt, Δt_{new}, telle que l'écart d_{new} calculé avec cette nouvelle valeur soit inférieur à la précision demandée :

$$|C| (\Delta t_{\text{new}})^{n+1} \left(1 - \frac{1}{2^n}\right) \leq \epsilon \quad (2.152)$$

En comparant ces deux dernières expressions, on a donc :

$$\Delta t_{\text{new}} = \Delta t \left(\frac{\epsilon}{d}\right)^{\frac{1}{n+1}} . \quad (2.153)$$

Pour éviter une éventuelle boucle infinie lorsqu'on raccourcit le pas de temps, il est judicieux de multiplier par un facteur $f < 1$, c'est-à-dire *tant que $d > \epsilon$, on refait le pas avec*

$$\Delta t_{\text{refaire}} = f\ \Delta t \left(\frac{\epsilon}{d}\right)^{\frac{1}{n+1}} . \quad (2.154)$$

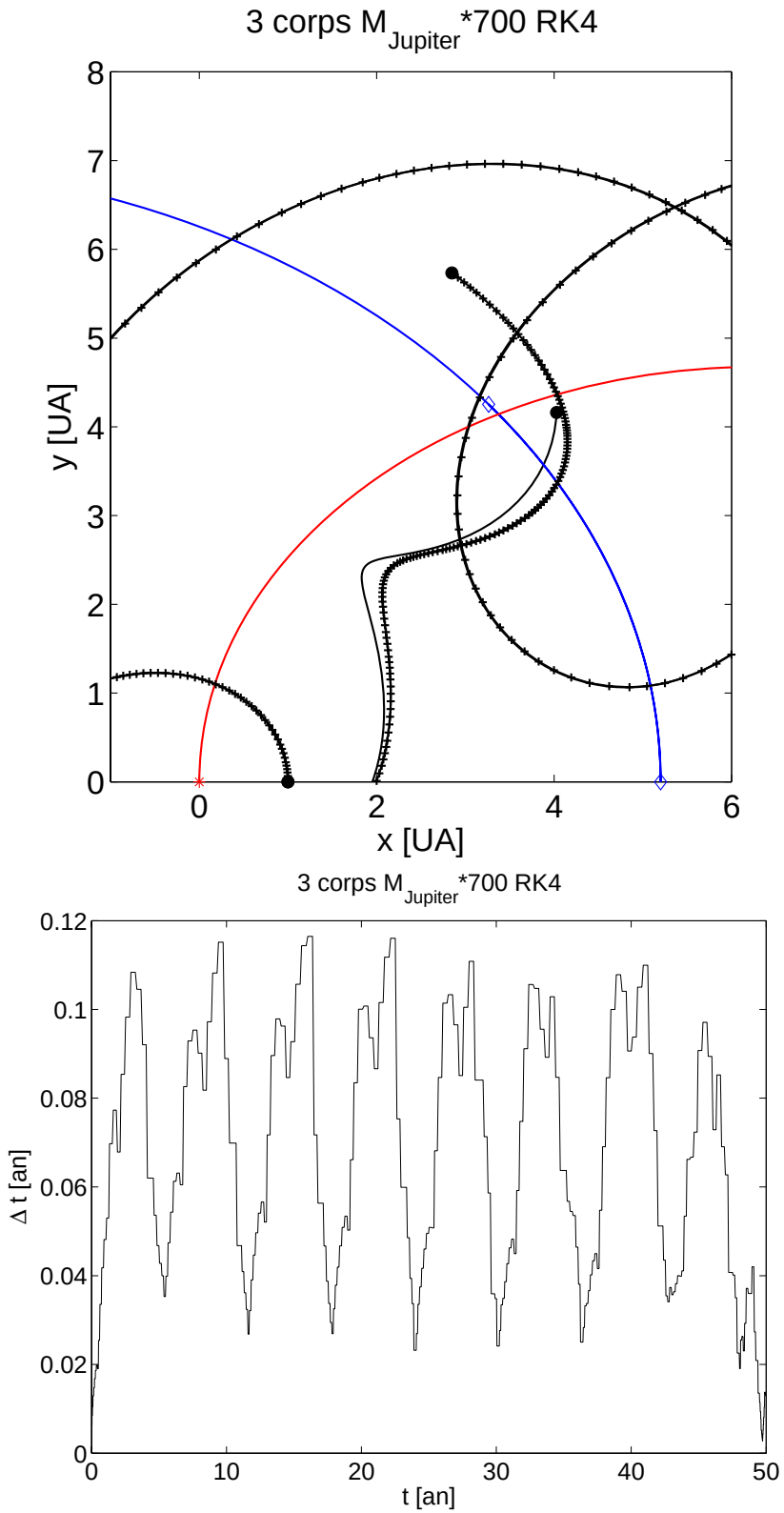


FIGURE 2.33 – *Trajectoires du soleil (rouge, étoile), de "Jupiter" s'il avait 700 fois sa masse réelle (bleu, diamant) et de la terre (noir, cercle). Mêmes paramètres physiques que la FIG. 2.31. Algorithme de Runge-Kutta du 4e ordre, Eq.(2.141), avec pas temporel variable, deux éxécutions avec deux précisions différentes : 965 pas (ligne avec +) et 1469 pas (ligne). En bas, l'évolution temporelle de Δt est représentée pour le cas à 965 pas.*

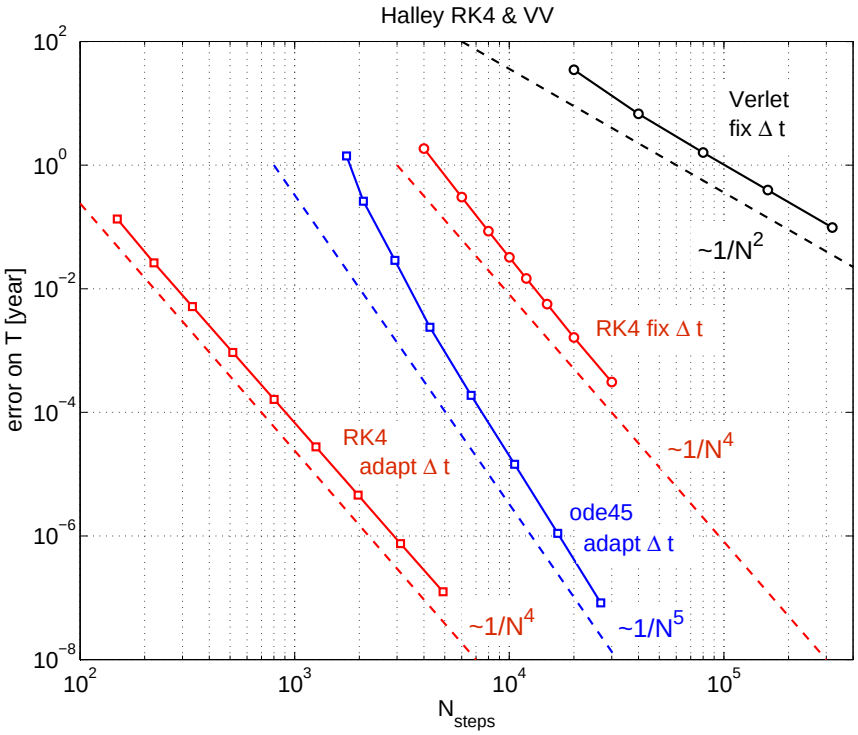


FIGURE 2.34 – *Convergence de l'erreur sur la période de révolution de la comète de Halley, en fonction du nombre de pas de temps. Différents schémas sont utilisés : Verlet (noir), Runge-Kutta (rouge) avec Δt fixe (cercles) ou adaptatif (carrés). Pour comparaison, on a également utilisé la fonction Matlab* `ode45`*, qui est aussi à pas de temps adaptatif.*

On prend typiquement $f \approx 0.95 - 0.99$.

On montre un exemple, le problème gravitationnel à trois corps avec un schéma Runge-Kutta d'ordre 4 à pas adaptatif, à la FIG. 2.33. Avec 965 pas temporels, la méthode est aussi précise que la méthode de Verlet à pas fixe avec 50000 pas temporels. On a représenté aussi comment le pas temporel Δt varie au cours du temps.

Un autre exemple, le problème gravitationnel à un corps, pour la trajectoire de la comète de Halley, est illustré à la Fig. 2.34, où l'erreur sur la période de révolution est représentée en fonction du nombre de pas temporels effectués. On varie le nombre de pas en variant la précision requise ϵ. Cette comète a une trajectoire très elliptique, et la distance avec le soleil varie fortement au cours de son orbite. Dans de telles situations, un schéma à pas temporel adaptatif est particulièrement efficace : dans cet exemple, avec la méthode Runge-Kutta d'ordre 4, pour 5000 pas de temps, le schéma adaptatif est 10 millions de fois plus précis qu'avec un Δt fixe !

Suggestion d'exercice. Prendre 3 corps célestes de masses dans le rapport 3 :4 :5. Condition initiale : positions aux sommets d'un triangle rectangle de longueurs de côtés 3 :4 :5 (faisant face à la masse correspondante), et vitesses initiales toutes nulles. Etudier et comparer divers schémas numériques à pas temporel fixe, leur comportement avec Δt, puis considérer un schéma à pas variable.

Suggestion d'exercice : Points de Lagrange. Problème à 3 corps où un des corps est de masse bien plus petite que les deux autres (appelé problème réduit). Se placer dans le référentiel tournant dans lequel les 2 corps massiques sont fixes. Calculer l'énergie potentielle pour le 3e corps et observer qu'il y a des positions d'équilibre, soit en forme d'extrema soit en forme de points selle. Prendre pour le mouvement du 3e corps des conditions initiales voisines d'une des positions d'équilibre et étudier la stabilité des orbites obtenues.

2.10.4 Solide en rotation chaotique dans un champ gravitationnel

On considère une planète de masse M autour de laquelle gravite un satellite, modélisé par un ensemble de 2 points matériels de masses m_1, m_2 reliés par une tige rigide de masse négligeable, de longueur L. On étudiera le cas où la tige est dans le plan de l'orbite du satellite. On supposera $M >> m_1, m_2$, de telle sorte que la planète puisse être considérée comme immobile. La longueur L est beaucoup plus petite que la distance r_G du centre de gravité du satellite au centre de la planète. Mais L n'est pas nul. Donc, puisque les positions $\vec{r}_1$ et $\vec{r}_2$ des deux points matériels au centre de la planète sont différents, les forces gravitationnelles exercent un couple non nul sur le satellite. Il y aura donc accélération

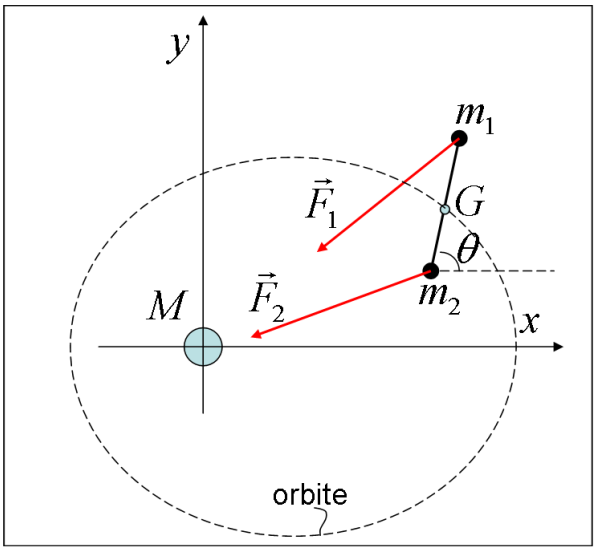


FIGURE 2.35 – *Modèle d'un satellite solide rigide en orbite autour d'une planète de masse M. Pour des raisons de clarté de la figure, la taille du satellite a été fortement exagérée.*

angulaire du satellite. Voir FIG. 2.35. Soit θ l'angle de la tige avec l'axe des x. Soit $\vec{r}_G$ la position du centre de masse. Le moment des forces de gravitation par rapport à G est

$$\vec{M}_G^{\text{ext}} = (\vec{r}_1 - \vec{r}_G) \times \vec{F}_1 + (\vec{r}_2 - \vec{r}_G) \times \vec{F}_2 \,, \tag{2.155}$$

avec

$$\vec{F}_i = -\frac{GMm_i}{r_i^3}\vec{r}_i \,, \quad i = 1, 2. \tag{2.156}$$

Le moment cinétique relatif à G est

$$\vec{L}_G = I_G\vec{\omega} \,, \quad \omega = \mathrm{d}\theta/\mathrm{d}t \,, \quad I_G = \sum_{i=1}^{2} m_i(\vec{r}_i - \vec{r}_G)^2 \,. \tag{2.157}$$

L'équation du moment cinétique dans le référentiel du centre de masse (rappel : c'est un référentiel en translation avec G)

$$\frac{\mathrm{d}\vec{L}_G}{\mathrm{d}t} = \vec{M}_G^{\text{ext}} \tag{2.158}$$

nous donne, à l'ordre le plus bas en L/r_G (exercice) :

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = -\frac{3GM}{r_G^5}(x_G \sin\theta - y_G \cos\theta)(x_G \cos\theta + y_G \sin\theta) \,. \tag{2.159}$$

En utilisant le programme écrit pour le mouvement gravitationnel à 1 corps et y rajoutant l'intégration de $\mathrm{d}\theta/\mathrm{d}t = \omega$ et $\mathrm{d}\omega/\mathrm{d}t$ de l'Eq.(2.159) ci-dessus, on obtient les résultats de la FIG. 2.36. L'algorithme de Runge-Kutta d'ordre 4 avec un pas de temps fixe, $\Delta t = 10^{-3}$ an, a été utilisé.

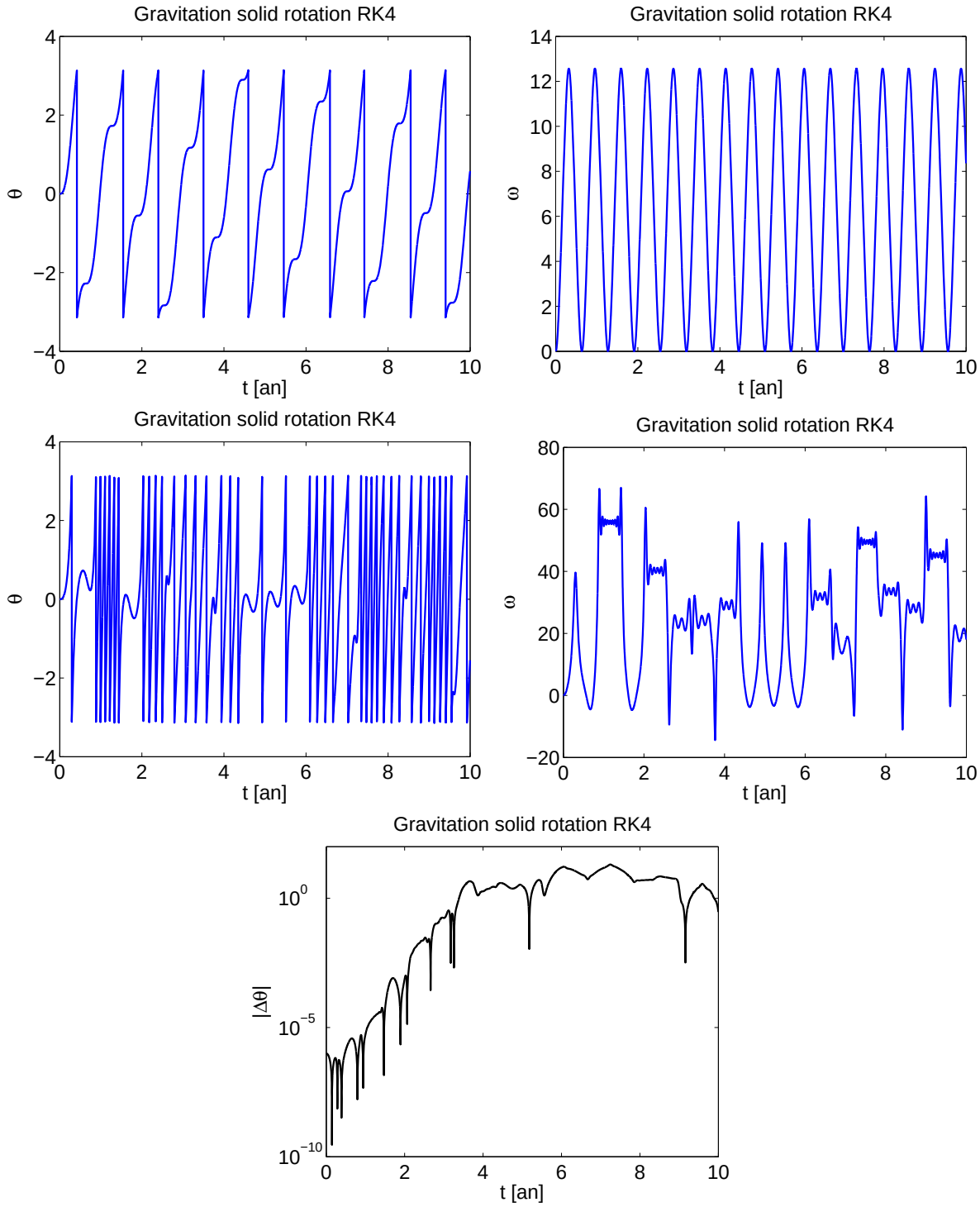


FIGURE 2.36 – *Rotation d'un satellite de longueur finie en orbite autour d'une planète. Cas d'une orbite circulaire (haut) : mouvement régulier, périodique. Cas d'une orbite elliptique (milieu) : mouvement irrégulier, chaotique. Sensibilité aux conditions initiales dans le cas chaotique : écart entre 2 trajectoires $|\Delta\theta(t)|$ pour deux conditions initiales voisines, $|\Delta\theta(0)| = 10^{-6}$ (bas). Schéma de Runge-Kutta d'ordre 4 à pas de temps fixe $\Delta t = 10^{-3}$ an.*

Tout d'abord, pour une orbite circulaire, le mouvement de rotation du satellite est régulier, périodique. Pour une orbite elliptique, cependant, le mouvement de rotation devient chaotique. Il devient effectivement imprédictible, ce que l'on peut vérifier en prenant deux conditions initiales très proches et en mesurant l'écart $\Delta\theta$ entre les deux mouvements : il y a croissance exponentielle au cours du temps.

Il existe un exemple spectaculaire de rotation chaotique dans le système solaire : Hyperion, une des lunes de Saturne. Elle a une orbite excentrique, qui lui cause ce mouvement de rotation chaotique. Pour plus de détails, voir par exemple sous
http://solarviews.com/eng/hyperion.htm.

Suggestion d'exercice. Solide en rotation libre. Ecrire, puis résoudre les équations pour la vitesse angulaire $\vec{\omega}$ d'un corps solide dont les moments principaux d'inertie sont $I_1 < I_2 < I_3$. Choisir des conditions initiales avec $\vec{\omega}$ proche des axes principaux. Analyser la stabilité ou l'instabilité de la rotation au voisinage des axes principaux.

Suggestion d'exercice. Solide soumis à des couples de forces. Précession de l'axe des pôles due au couple de forces exercé par la lune sur la terre (aplatie par l'effet de sa rotation propre).

2.11 Particules dans un champ magnétique

2.11.1 Dérive des particules dans des champs inhomogènes

Soit une particule chargée, de masse m, charge q, en mouvement dans un champ magnétique statique $\vec{B}(\vec{x})$ et un champ électrique statique $\vec{E}(\vec{x})$. Les équations du mouvement s'obtiennent directement de la force de Lorentz et de la deuxième loi de Newton :

$$\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = \frac{q}{m}\left(\vec{E}(\vec{x}) + \vec{v} \times \vec{B}(\vec{x})\right) \ . \tag{2.160}$$

Dans le cas $\vec{B}$ uniforme, et $\vec{E} = 0$, il est facile d'obtenir la solution analytique exacte du mouvement : il est uniforme dans la direction de $\vec{B}$, et circulaire uniforme dans le plan perpendiculaire à $\vec{B}$. La fréquence angulaire du mouvement circulaire est la *fréquence cyclotronique* $\omega_c = qB/m$, le rayon du cercle est appelé *rayon de Larmor* $\rho_L = v_\perp/\omega_c$.

Les choses se compliquent quand $\vec{B}$ n'est pas uniforme. Mais il est facile d'intégrer numériquement les équations du mouvement. Pour fixer les idées, plaçons un système de coordonnées cartésiennes avec $z \parallel \vec{B}$, et supposons

$$\vec{B}(\vec{x}) = B_0(1 + \alpha x)\vec{e}_z \ . \tag{2.161}$$

Le mouvement parallèle, selon z, est toujours uniforme. Mais, dans le plan perpendiculaire, (x, y), on observe qu'au mouvement circulaire se superpose une **dérive** dans la direction y. La FIG. 2.37 montre les résultats de l'intégration numérique avec un schéma Runge-Kutta d'ordre 4. La dérive a lieu dans la direction $\perp \vec{B}$ et $\perp \nabla B$. La direction de cette dérive dépend du signe de la charge : on montre les résultats pour un ion positif et pour un électron.

Cette constatation est à la base de travaux théoriques analysant le mouvement en le séparant en une composante rapide (le mouvement de gyration cyclotronique) et une composante lente (la dérive). La dérivation des équations de dérive sort du cadre de ce cours : elle sera abordée au cours de Physique des Plasmas. Mentionnons juste le résultat : la vitesse de dérive due au gradient de champ magnétique s'écrit

$$\vec{v}_{\nabla B} = \frac{v_\parallel^2 + v_\perp^2/2}{\omega_c B^2} \vec{B} \times \nabla B \; . \tag{2.162}$$

Ce mouvement de dérive est représenté en traitillés sur la FIG. 2.37. Le calcul numérique vérifie donc bien la théorie. On peut aussi vérifier les points suivants : le rayon de gyration est bien $\rho_L = v_\perp/\omega_c$, la fréquence du mouvement de gyration est bien $\omega_c = qB/m$, et l'énergie cinétique de la particule est bien conservée. On montre à la FIG. 2.37 comment la conservation de l'énergie converge en prenant des Δt de plus en plus petits.

Suggestion d'exercice. Superposer au champ magnétique $\vec{B}(\vec{x}) = B_0(1 + \alpha x)\vec{e}_z$ un champ électrique uniforme dans la direction y. Montrer qu'à la vitesse de dérive due au gradient de champ magnétique se superpose une dérive $\vec{v}_E = \vec{E} \times \vec{B}/B^2$, indépendante de la charge, de la masse, et de la vitesse de la particule.

Suggestion d'exercice. Considérer un champ magnétique curviligne et non uniforme avec $B_z(z) = B_0 + B_1 \cos(2\pi z/L)$, B_0, B_1 et L étant des constantes données. Utiliser l'équation $\nabla \cdot \vec{B} = 0$ en coordonnées cylindriques pour trouver $B_r(r, z)$. On projettera ensuite sur les coordonnées cartésiennes. Etudier le mouvement d'une particule chargée de charge q, masse m dans ce champ magnétique. Choisir différentes conditions initiales, et observer qu'à partir d'un certain rapport entre $v_\perp$ et $v_\parallel$ la particule est réfléchie dans la direction z : c'est *l'effet miroir*, observé par exemple pour les particules du vent solaire dans le champ magnétique terrestre. On montre à la FIG. 2.38 quelques résultats typiques. On a choisi la postition initiale à l'endroit B minimum. Avec $v_\perp/v_\parallel$ au temps $t = 0$ suffisamment petit, la particule arrive à passer les maxima de B , elle est dite "passante" (trajectoire noire sur la FIG. 2.38). Avec $v_\perp/v_\parallel(t = 0)$ suffisamment large, la particule est réfléchie là où le champ magnétique est plus intense, elle est dite "piégée" (trajectoire bleue sur la FIG. 2.38. Dans la théorie des dérives, on montre que *le moment magnétique de la particule*

$$\mu = \frac{m v_\perp^2}{2B} \tag{2.163}$$

est *conservé en moyenne* : mis à part des oscillations lors du mouvement cyclotronique rapide, le moment est conservé aux échelles de temps du mouvement de dérive lent. Avec

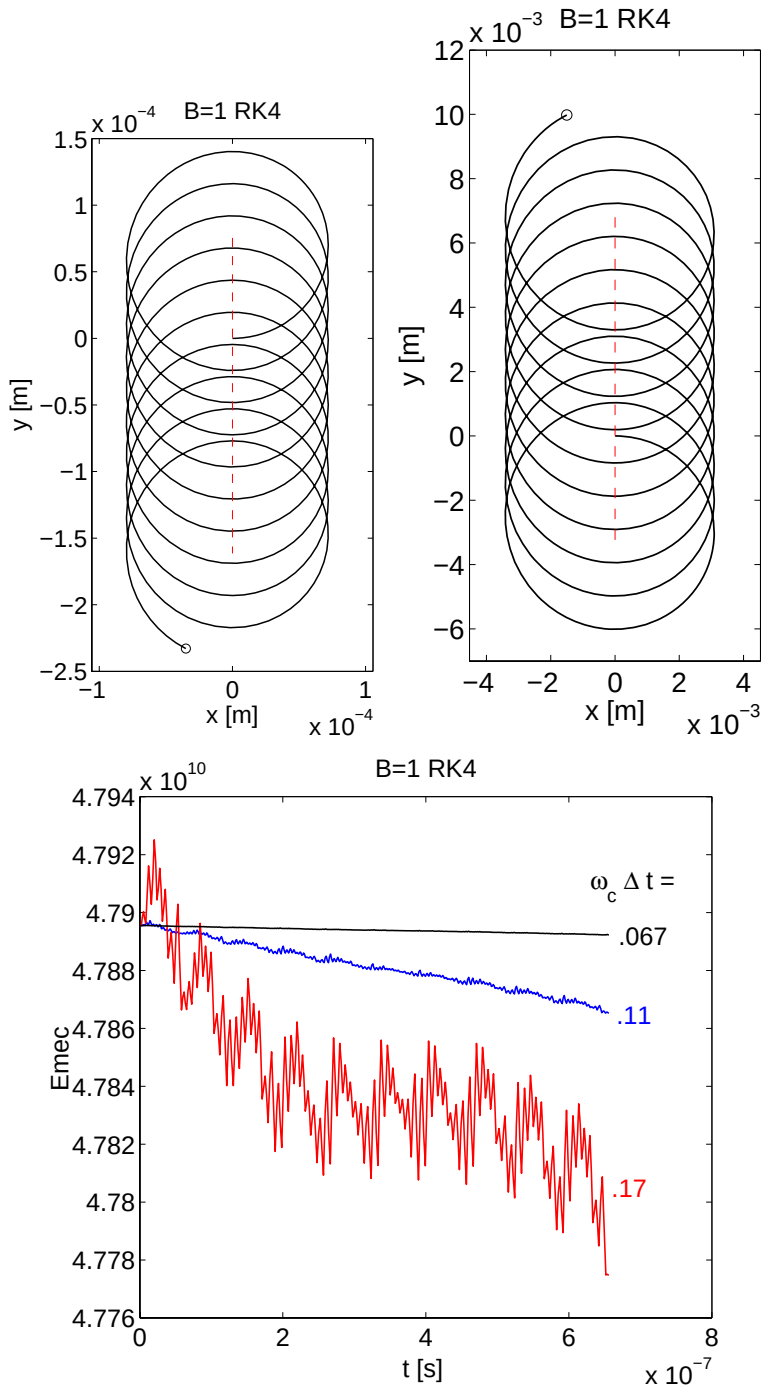


FIGURE 2.37 – *Trajectoire d'un électron (en haut à gauche), d'un ion positif (en haut à droite) dans un champ magnétique rectiligne parallèle à z et d'intensité variable selon x. Une dérive lente dans la direction $\perp \vec{B}$ et $\perp \nabla B$ se superpose au mouvement rapide de gyration cyclotronique. Schéma de Runge-Kutta d'ordre 4. On vérifie (en bas) la conservation de l'énergie de la particule.*

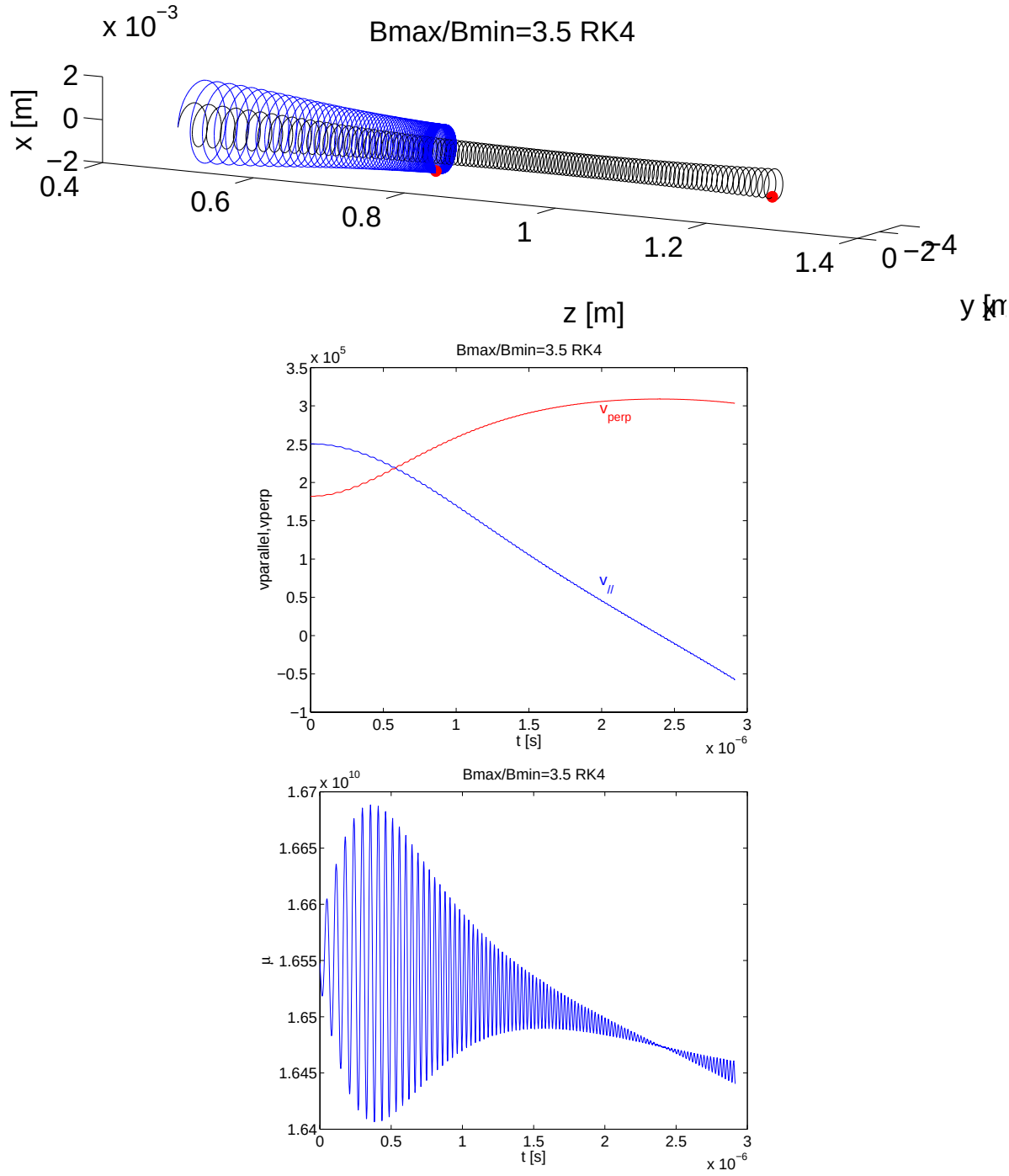


FIGURE 2.38 – *Effet miroir dans un champ magnétique curviligne. En haut : particule passante (noir) et particule piégée (bleu). Au milieu : $v_\perp(t)$ et $v_\parallel(t)$ pour la particule piégée (on remarque que le signe de $v_\parallel$ s'inverse !). En bas : conservation approximative, et en moyenne, du moment magnétique de la particule piégée. Schéma de Runge-Kutta d'ordre 4.*

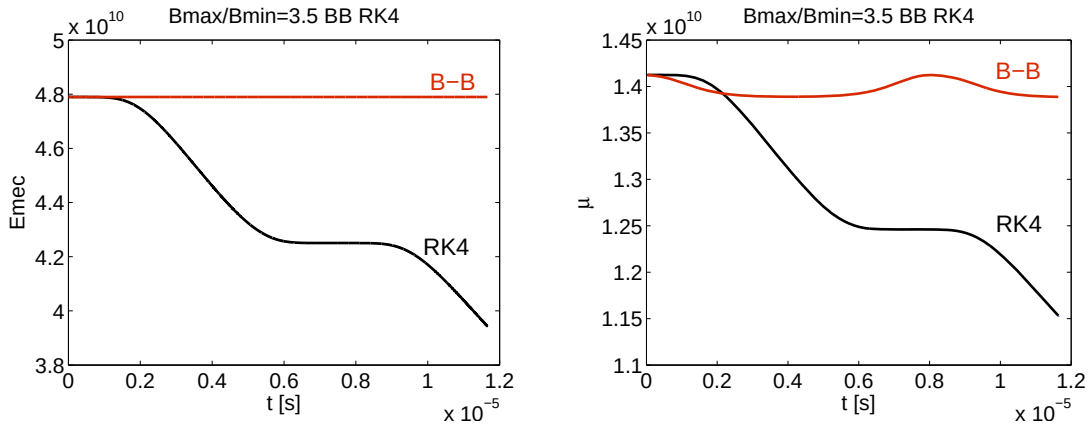


FIGURE 2.39 – *Energie (à gauche) et moment magnétique (à droite) de la prticule au cours du temps. Normalement, l'énergie devrait ête consrevée exactement, et le moment magnétique devrait être conservé en moyenne temporelle. Le schéma de Runge-Kutta ne conserve pas bien, alors que le schéma de Boris-Buneman est bien meilleur.*

le fait que l'énergie cinétique de la particule est conservé, on peut obtenir la condition pour laquelle une particule sera réfléchie (piégée) :

$$\frac{v_\parallel(t=0)}{v_\perp(t=0)} \quad < \quad \sqrt{\frac{B_{\max}}{B(\vec{r}(t=0))} - 1} \quad \Rightarrow \quad \text{piégée.} \tag{2.164}$$

2.11.2 Schéma de Boris-Buneman

Le schéma de Runge-Kutta, même s'il est d'ordre 4, ne conserve pas bien l'énergie et le moment magnétique pour de longues simulations. Il y a systématiquement une accumulation d'erreurs qui fait que l'énergie de la particule, ainsi que son moment magnétique, "s'érodent" au cours du temps. La Fig. 2.39 illustre ceci, où on a prolongé la simulation de la Fig. 2.38.

On peut faire mieux, avec le schéma de **Boris-Buneman** [9, 10, 11]. Pour simplifier les notations, nous écrivons $\vec{v}_- = \vec{v}_n$, la vitesse de la particule à l'instant $t = t_n$, et $\vec{v}_+ = \vec{v}_{n+1}$ sa vitesse à l'instant $t = t_{n+1}$. Pour une particule dans un champ magnétique seulement ($\vec{E} = 0$), on écrit l'équation différentielle du mouvement, Eq.(2.160), avec des différences finies centrées pour la dérivée temporelle, et la moyenne des vitesses en début et fin d'intervalle pour le membre de droite :

$$\frac{\vec{v}_+ - \vec{v}_-}{\Delta t} = \frac{q}{m}\left(\frac{\vec{v}_+ + \vec{v}_-}{2}\right) \times \vec{B}\,. \tag{2.165}$$

En rappelant la définition $\omega_c = qB/m$, et en posant $\vec{e_\parallel} = \vec{B}/B$, on a :

$$\vec{v}_+ = \vec{v}_- + \frac{\omega_c \Delta t}{2}(\vec{v}_+ + \vec{v}_-) \times \vec{e_\parallel} \tag{2.166}$$

C'est un schéma dit **semi-implicite** : la solution au pas de temps ultérieur dépend de la solution au pas précédent, $\vec{v}_-$, (partie dite explicite) et de la solution au temps ultérieur, $\vec{v}_+$, (partie dite implicite). Dans le cas précis, on peut en fait résoudre la partie semi-implicite analytiquement. On obtient, après quelques calculs :

$$\vec{v}_+ = \vec{v}_- + \frac{\omega_c \Delta t}{1 + (\omega_c \Delta t/2)^2} \left(\vec{v}_- \times \vec{e_\parallel} + \frac{\omega_c \Delta t}{2} (\vec{v}_- \times \vec{e_\parallel}) \times \vec{e_\parallel} \right) . \tag{2.167}$$

On peut montrer (exercice) que le schéma de Boris-Buneman conserve l'énergie mécanique exactement. La conservation du moment magnétique est également bien meilleure : voir Fig.2.39.

Dans le cas d'une présence simultanée d'un champ électrique $\vec{E}(\vec{x})$ et d'un champ magnétique $\vec{B}(\vec{x})$, le schéma de Boris-Buneman s'écrit :

$$\boxed{\begin{array}{rcl} \vec{x}_- & = & \vec{x}_n + \vec{v}_n \Delta t/2 \\ \vec{v}_- & = & \vec{v}_n + (q/m)\vec{E}(\vec{x}_-)\Delta t/2 \\ \vec{v}_+ & = & \vec{v}_- + \frac{\omega_c(\vec{x}_-)\Delta t}{1+(\omega_c(\vec{x}_-)\Delta t/2)^2} \left(\vec{v}_- \times \vec{e_\parallel}(\vec{x}_-) + \frac{\omega_c(\vec{x}_-)\Delta t}{2} (\vec{v}_- \times \vec{e_\parallel}(\vec{x}_-)) \times \vec{e_\parallel}(\vec{x}_-) \right) \\ \vec{v}_{n+1} & = & \vec{v}_+ + (q/m)\vec{E}(\vec{x}_-)\Delta t/2 \\ \vec{x}_{n+1} & = & \vec{x}_- + \vec{v}_{n+1}\Delta t/2 \end{array}} \tag{2.168}$$

Chapitre 3

Intégration Spatiale : Problèmes aux limites

On s'intéresse ici à l'évolution de systèmes physiques variant dans l'espace et le temps. De tels systèmes sont souvent décrit par des Equations aux Dérivées Partielles (EDP) opérant sur des champs scalaires et/ou vectoriels. Dans ce cours, nous nous limiterons aux champs scalaires, c'est-à-dire des fonctions $f(\vec{x}, t)$ à valeurs réelles ou complexes.

Nous considèrerons trois équations fondamentales très importantes de la physique : l'équation d'advection-diffusion, l'équation d'onde et l'équation de Schrödinger. Trois schémas numériques seront introduits et utilisés pour résoudre ces trois équations : les différences finies explicites à deux et trois niveaux et le schéma semi-implicite de Crank-Nicolson.

3.1 Cas 1-D : méthode de tir

Dans le cas unidimensionnel (1-D), il est souvent possible de se ramener à un problème aux valeurs initiales. On peut ainsi utiliser les méthodes numériques vues au chapitre précédent, en remplaçant formellement $t \to x$.

Dans cette section, nous allons présenter deux exemples : (1) calculer la distribution de pression, densité et température dans l'atmosphère terrestre, connaissant la température et la densité au sol ; (2) calculer la distribution de pression, densité et température au coeur du soleil.

Ces deux études nous permettront d'aborder le problème des singularités des équations. Elles sont de deux origines différentes : la première est d'origine physique, lorsque la

densité est nulle ; la deuxième est d'ordre géométrique, il s'agit de la singularité du système de coordonnées utilisées (sphériques dans le cas du soleil). La présence de ces singularités nécessite des adaptations des schémas numériques et peuvent altérer leurs propriétés de convergence.

3.1.1 Modèles fluides d'atmosphère planétaire. Singularité de l'équation

On supposera l'épaisseur de l'atmosphère négligeable par rapport au rayon de la planète. On supposera la masse de l'atmosphère négligeable par rapport à celle de la planète. Ceci nous permet d'approximer l'accélération de la pesanteur par une constante $\vec{g}$. On néglige le mouvement de l'atmosphère et la rotation de la terre. Les équations de base sont celles de la mécanique des fluides, avec un champ de vitesse fluide $\vec{v}(\vec{x},t) = 0, \forall t, \forall \vec{x}$ (statique), un champ de densité $\rho(\vec{x},t)$, un champ de pression $P(\vec{x},t)$ et un champ de température $T(\vec{x},t)$. De Navier-Stokes (ou Euler), on a

$$0 = -\nabla P + \rho\vec{g}\,. \tag{3.1}$$

De l'équation d'état de la thermodynamique des gaz parfaits, on a

$$P = (\rho/m)k_B T\,, \tag{3.2}$$

où m est la masse d'une molécule, $k_B = 1.3807 \times 10^{-23}$ est la constante de Boltzmann. Comme tout système fluide, les équations doivent être complétées par une hypothèse supplémentaire (on parle de *fermeture* du système d'équations).

Dans **le modèle isotherme**, on suppose $T = T_0 = \text{const}$. Il vient donc

$$0 = -\nabla\rho\frac{k_B T_0}{m} + \rho\vec{g}\,. \tag{3.3}$$

Avec les hypothèses d'atmosphère mince et statique, on a des champs qui ne dépendent que de l'altitude z (axe cartésien vertical). Il vient donc

$$\frac{\mathrm{d}\rho}{\mathrm{d}z} + \frac{mg}{k_B T_0}\rho = 0\,, \tag{3.4}$$

qui s'intègre facilement, à partir de la condition initiale $\rho_0 = \rho(0) = mP_0/(k_B T_0)$, comme

$$\rho = \rho_0 e^{-z/\lambda}\,, \quad \lambda = \frac{k_B T_0}{mg}\,. \tag{3.5}$$

Le modèle isotherme n'est certainement pas très réaliste : on sait bien que la température varie avec l'altitude. Un autre modèle est basé sur l'hypothèse que les échanges de chaleur (transport) sont négligeables ($\delta Q \approx 0$). Dans le **modèle adiabatique**, appelé aussi

polytropique, on a $P\rho^{-\gamma} = \text{const}$, avec γ l'indice d'adiabaticité ou rapport des chaleurs spécifiques. On rappelle que $\boxed{1 \leq \gamma < 2}$. On pose $P = C\rho^{\gamma}$, où C est une constante déterminée par les conditions au bord $z = 0$:

$$C = \left(\frac{k_B T_0}{m}\right)^{\gamma} \frac{1}{P_0^{\gamma-1}} \tag{3.6}$$

et on trouve, en substituant dans l'Eq.(3.1) (exercice) :

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\rho^{\gamma-1}\right) = -\frac{g(\gamma-1)}{C\gamma} \tag{3.7}$$

dont la solution est, avec la condition au bord $\rho(0) = \rho_0$,

$$\rho = \left(\rho_0^{\gamma-1} - \frac{g(\gamma-1)}{C\gamma} z\right)^{1/(\gamma-1)} \quad . \tag{3.8}$$

Nous allons intégrer numériquement ces équations. Cela nous permettra d'illustrer le problème de la **singularité**. On réécrit l'équation différentielle ci-dessus, Eq.(3.7), comme

$$\boxed{\frac{\mathrm{d}\rho}{\mathrm{d}z} = -\frac{g}{C\gamma}\rho^{2-\gamma}} \, . \tag{3.9}$$

[N.B. : Pour $\gamma = 1$, cette équation, avec C tiré de (3.6), conduit à l'Eq.(3.4) pour le modèle de l'atmosphère isotherme.] C'est une équation différentielle du 1er ordre, et on peut utiliser le schéma d'Euler, Eq.(2.10), avec la variable d'intégration z remplaçant la variable temporelle. On trouve le résultat de la FIG. 3.1. L'intégration se passe sans problème jusqu'au moment où la densité devient nulle, en

$$z_0 = \frac{\gamma}{\gamma-1}\frac{k_B T_0}{mg} \, . \tag{3.10}$$

Mathématiquement, notre équation a des problèmes à ce point-là ; elle est *singulière*, avec le comportement suivant :

$$\lim_{z\to z_0} \rho(z) = 0 \, ; \quad \lim_{z\to z_0} \frac{\mathrm{d}^m \rho}{\mathrm{d}z^m}(z) = 0, \forall m < \frac{1}{\gamma-1}; \quad \lim_{z\to z_0} \frac{\mathrm{d}^n \rho}{\mathrm{d}z^n}(z) = \infty \, , \forall n \geq \frac{1}{\gamma-1} \, . \tag{3.11}$$

Pour $z > z_0$, la solution mathématique est complexe, et n'a pas de signification physique : z_0 représente, dans ce modèle adiabatique, le sommet de l'atmosphère.

Il est intéressant d'essayer d'intégrer le problème inverse : supposons la position du sommet de l'atmosphère, z_0, connue, et calculons quelle est la densité au sol. Intégrer "en marche arrière", c.a.d. à partir de la condition initiale $z = z_0$, $\rho = 0$, avec un pas $\Delta z < 0$ donne la solution numérique ρ nulle partout ! Le problème vient de la singularité.

La solution est d'examiner analytiquement le comportement au voisinage de la singularité, et prendre une condition initiale en $z = z_{\text{init}} = z_0 - \epsilon$ avec $\rho(z_0 - \epsilon)$ *consistant avec le*

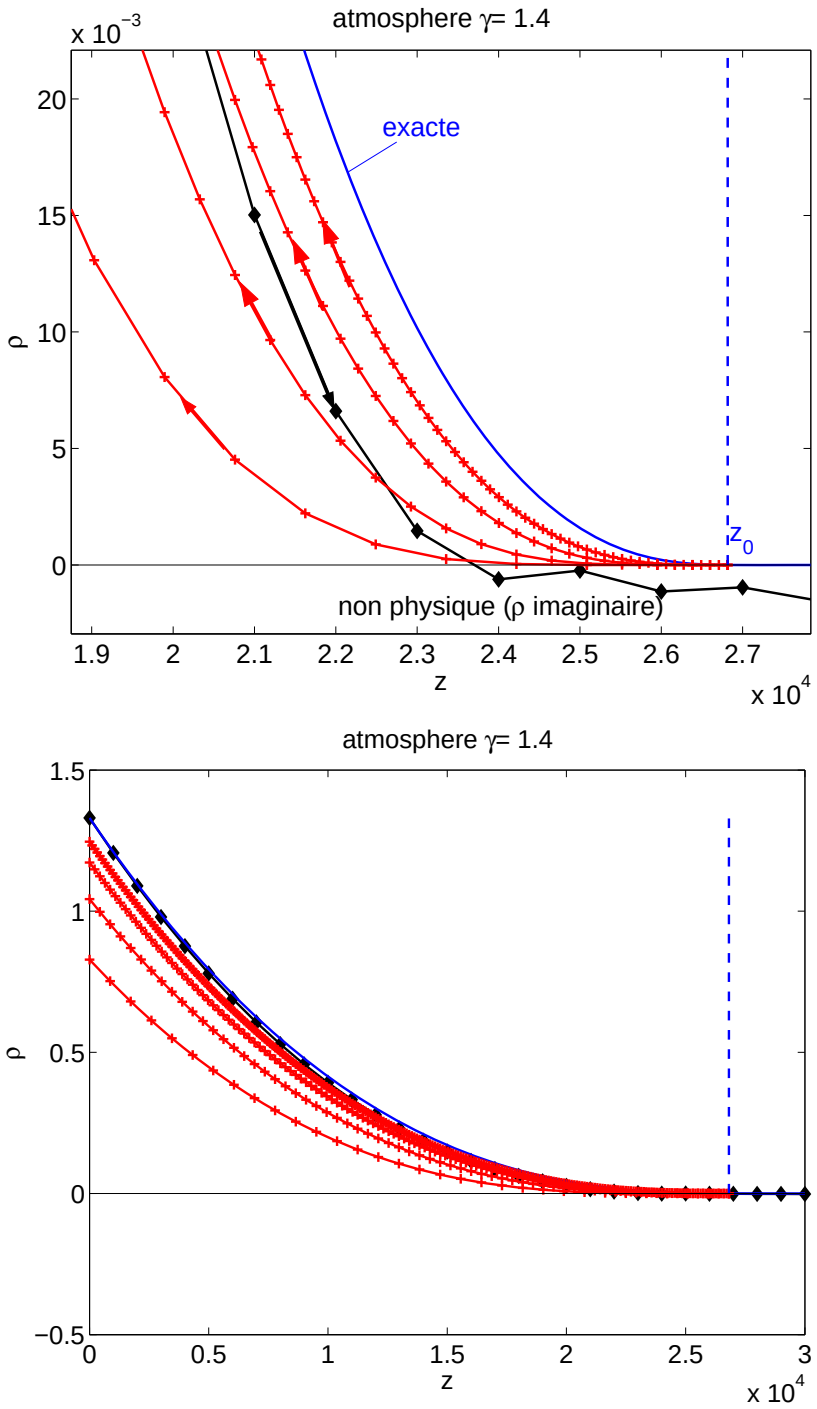


FIGURE 3.1 – *Densité de l'atmosphère terrestre calculée avec le schéma d'Euler explicite. Modèle adiabatique, $\gamma = 7/5$. En $z = z_0 = \gamma/(\gamma-1)(k_B T_0/mg)$ (ligne traitillée verticale), la densité s'annulle. L'intégration numérique "en avant" (ligne noire avec losanges) devient non physique dès que la solution numérique trouve $\rho < 0$. L'intégration numérique "en arrière" (lignes rouges avec croix), à partir de $z_0 - \epsilon$, doit se faire en tenant compte du caractère singulier de l'équation au voisinage de $z = z_0$. On peut montrer que la densité au sol, calculée à partir du sommet de l'atmosphère, converge bien avec $\Delta z \to 0$.*

comportement singulier de la solution au voisinage de z_0. On obtient le comportement au voisinage de z_0 avec l'Ansatz

$$\rho = A(z_0 - z)^\alpha \ = A\epsilon^\alpha \tag{3.12}$$

que l'on substitue dans l'équation différentielle (3.9) pour obtenir

$$\alpha = \frac{1}{\gamma - 1}\ , \quad A = \left(\frac{C\gamma}{g(\gamma - 1)}\right)^{1/(1-\gamma)} \tag{3.13}$$

et donc

$$\rho(z_{\text{init}}) = \left(\frac{C\gamma}{g(\gamma - 1)}\right)^{1/(1-\gamma)} \epsilon^{1/(\gamma-1)}\ . \tag{3.14}$$

On effectue ensuite l'intégration numérique, dont on étudie les propriétés de convergence avec Δt, voir à la FIG. 3.1, et la dépendance en ϵ. Le résultat convergé en Δt ne devrait pas dépendre du choix de ϵ. Plus exactement, on devrait faire $\lim_{\epsilon \to 0+}$. Cependant, ceci n'est pas faisable numériquement, car plus ϵ est petit, plus on se rapproche de la singularité de l'équation différentielle, et pour ϵ trop petit, les erreurs numériques dues à cette proximité l'emportent sur l'approximation ϵ fini. En fait, si on utilise un schéma numérique d'ordre élevé, la proximité de la singularité peut faire *perdre l'ordre de convergence du schéma* : en effet, l'ordre de convergence n'est effectif que pour une régularité suffisante de la solution, ce qui n'est pas le cas au voisinage d'une singularité.

Suggestion d'exercice. Calculer quelle serait la densité au sol si l'atmosphère avait une hauteur de 50km. Faire les études de convergence avec Δt et de comportement au voisinage de la singularité.

3.1.2 Distribution de pression, densité et température au coeur du soleil

Singularité du système de coordonnées

Quelle est la densité au centre du soleil ? L'impossibilité de mesures expérimentales *in situ* implique la nécessité de développer des modèles théoriques, basés sur un certain nombre d'hypothèses. Le problème est un peu plus compliqué que le cas d'une atmosphère planétaire mince, où on négligeait la masse de l'atmosphère par rapport à celle de la terre (solide). Dans une étoile, la masse *est* celle du gaz, et cette masse dépend du rayon.

Nous allons faire les hypothèses suivantes :

- fluide au repos
- équilibre des forces de pression et des forces gravitationnelles
- équation d'état *polytropique* :

$$P\rho^{-\gamma} = \text{const} \tag{3.15}$$

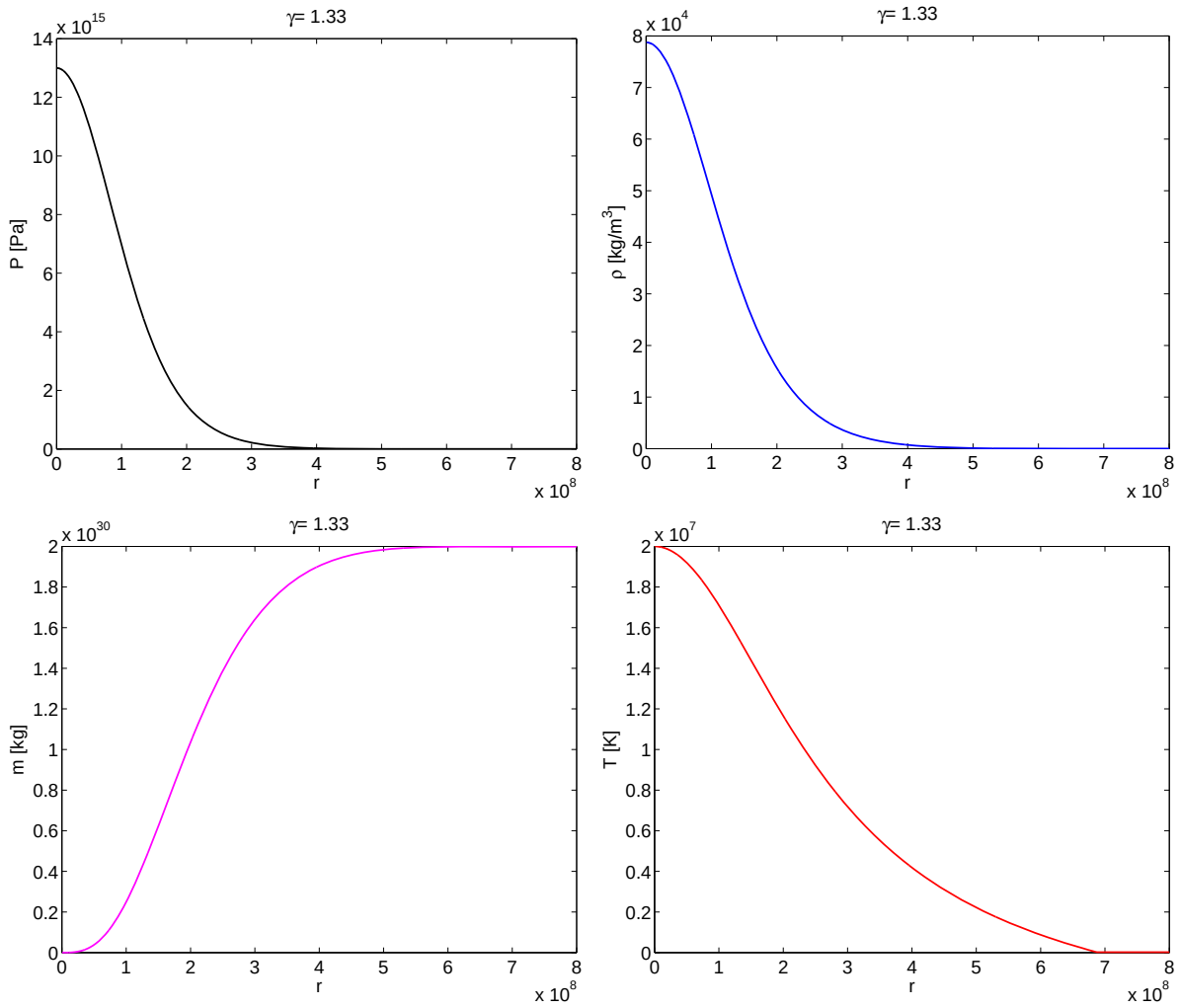


FIGURE 3.2 – *Pression, densité, masse et température à l'intérieur du soleil, calculés avec le modèle polytropique, $\gamma = 4/3$, et utilisant un schéma Runge-Kutta d'ordre 4.*

avec γ une constante donnée

— symétrie sphérique, on néglige la rotation du soleil

Le soleil est composé majoritairement d'hydrogène complètement ionisé, pour lequel $\gamma = 5/3$, et de photons, pour lesquels $\gamma = 4/3$. On simplifiera en ne considérant qu'une seule valeur de γ pour tout l'intérieur du soleil.

Soit (r, θ, φ) les coordonnées sphériques centrées au centre de masse du soleil. Soient $P(r)$ la pression, $m(r)$ la masse contenue à l'intérieur d'une sphère de rayon r, et $\rho(r)$ la densité. Les équations de base sont ainsi :

1) Statique des fluides $\Rightarrow$

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{\rho G m}{r^2} \,. \tag{3.16}$$

2) Masse dm contenue dans une sphère de rayon r, d'épaisseur dr $\Rightarrow$

$$\frac{\mathrm{d}m}{\mathrm{d}r} = 4\pi r^2 \rho \; . \tag{3.17}$$

3) Equation d'état polytropique $\Rightarrow$

$$P\rho^{-\gamma} = K \; . \tag{3.18}$$

A partir de ces équations, on peut montrer (exercice) que l'on obtient l'équation suivante pour la densité :

$$\boxed{\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 K\gamma\rho^{\gamma-2}\frac{\mathrm{d}\rho}{\mathrm{d}r}\right) = -4\pi\rho G} \; . \tag{3.19}$$

Méthode de tir. Elle consiste à intégrer numériquement, avec une des méthodes pour les valeurs initiales développées précédemment (avec r au lieu de t comme variable d'intégration) : Euler, Runge-Kutta, etc. Il faut préciser 2 conditions initiales, puisque c'est une équation du 2e ordre :

$$\left\{ \begin{array}{rcl} \rho(0) & = & \rho_0 \\ \frac{\mathrm{d}\rho}{\mathrm{d}r}(0) & = & 0 \end{array} \right. \tag{3.20}$$

On ajustera la valeur initiale ρ_0 jusqu'à obtenir le rayon du soleil $R = 7 \times 10^8$m et la masse du soleil $M = m(R) = 2 \times 10^{30}$kg.

Singularité en $r = 0$. L'équation à résoudre est singulière en $r = 0$. Ce type de singularité est lié au choix des coordonnées sphériques. Ce n'est donc pas une singularité d'origine physique : physiquement parlant, tout est régulier en $r = 0$. Mais ceci impose de prendre la condition initiale non pas en $r = 0$, mais en $r = \epsilon$. On choisira $\epsilon << R$. Pour démarrer l'intégration correctement, il faut choisir $\mathrm{d}\rho/\mathrm{d}r(\epsilon)$. En partant des équations de base (3.16)-(3.18), on obtient

$$\frac{\mathrm{d}\rho}{\mathrm{d}r}(\epsilon) \approx -\frac{1}{\gamma K}\rho_0^{2-\gamma}\frac{Gm(\epsilon)}{\epsilon^2} \tag{3.21}$$

avec $m(\epsilon) \approx (4/3)\pi\rho_0\epsilon^3$ la masse contenue à l'intérieur de la petite sphère de rayon ϵ. (N.B. L'approximation vient du fait que l'on a considéré $\rho \approx \mathrm{const} = \rho_0$ à l'intérieur de cette petite sphère). Un exemple, avec $\gamma = 4/3$, est montré à la FIG. 3.2.

Suggestion d'exercice. Calculer la température, la densité et la pression au centre du soleil. Comparer les cas $\gamma = 5/3$ et $\gamma = 4/3$. Indication : prendre la densité à la surface du soleil $\rho(R) = 10^{-1}$kg/m^3 ou $\rho(R) = 10^{-4}$kg/m^3 comme critère de détermination de R.

3.2 Diifférences finies. Equation de Poisson

3.2.1 Electrodynamique et limite statique

On se bornera dans cette section à rappeler l'essentiel des équations de base. Pour plus de détails, voir les cours de Physique 3 et 4.

On décrit l'interaction entre particules chargées électriquement par l'intermédiaire du concept de **champs électromagnétiques, abrégé EM** : $\vec{E}(\vec{x},t), \vec{B}(\vec{x},t)$, champs vectoriels.

Un ensemble de charges q_i sera décrit par un champ scalaire **densité de charge** $\rho(\vec{x},t)$.

Un ensemble de charges en mouvement sera décrit par un courant électrique I, ou un champ vectoriel **densité de courant** $\vec{j}(\vec{x},t)$.

Les champs $\vec{E}, \vec{B}$ sont créés par les champs $\rho, \vec{j}$. Ces champs obéissent aux **équations de Maxwell** :

$$\boxed{\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0}} \quad (a) \qquad \boxed{\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}} \quad (b)$$

$$\boxed{\nabla \cdot \vec{B} = 0} \quad (c) \qquad \boxed{\nabla \times \vec{B} = \mu_0 \vec{j} + \frac{1}{c^2}\frac{\partial \vec{E}}{\partial t}} \quad (d) \ . \tag{3.22}$$

Les champs $\vec{E}$ et $\vec{B}$ sont tels qu'une charge q dans un champ EM subit une force, la **force de Lorentz** :

$$\boxed{\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right)} \, . \tag{3.23}$$

Avec les lois de la dynamique de Newton, $\boxed{m\vec{a} = \vec{F}}$, on a une théorie décrivant **l'ensemble des phénomènes EM** (électrodynamique classique).

Dans la limite de champs statiques, $\partial/\partial t = 0$ et les équations de pour les champs $\vec{E}$ et $\vec{B}$ se *découplent*. Pour l'électrostatique,

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0} \, , \qquad \nabla \times \vec{E} = 0 \, . \tag{3.24}$$

De la 2e équation, on tire l'existence d'un potentiel scalaire $\phi(\vec{x})$ tel que

$$\vec{E} = -\nabla\phi \, , \quad \nabla^2\phi = -\frac{\rho}{\varepsilon_0} \, . \tag{3.25}$$

Pour la magnétostatique [1]

$$\nabla \cdot \vec{B} = 0 \,, \qquad \nabla \times \vec{B} = \mu_0 \vec{j} \,. \tag{3.26}$$

De la 1e équation, on tire l'existence d'un potentiel vecteur $\vec{A}(\vec{x})$ tel que

$$\vec{B} = \nabla \times \vec{A} \,, \qquad \nabla^2 \vec{A} = -\mu_0 \vec{j} \,. \tag{3.27}$$

Pour la dernière équation, on a utilisé la *jauge de Coulomb :* $\nabla \cdot \vec{A} = 0$.

Les équations pour $\vec{E}(\vec{x})$ et pour $\vec{B}(\vec{x})$ dans le vide (cas $\rho = 0, \vec{j} = 0$) sont identiques. Cela ne signifie pas pour autant qu'elles ont des solutions identiques : la différence tient aux *sources de ces champs et de leur topologie* : des charges ponctelles pour $\vec{E}$, des boucles de courant pour $\vec{B}$. Cela implique des topologies fondamentalement différentes pour les lignes de champ $\vec{E}$ et $\vec{B}$.

Les équations pour ϕ et pour $\vec{A}$ sont également de même nature (du moins si on les écrit en coordonnées cartésiennes). Cela veut dire que les méthodes pour résoudre les problèmes d'électrostatique sont en principe applicables aux problèmes de magnétostatique. Il faut cependant faire attention aux conditions aux bords, qui pourront être différentes d'un problème à l'autre.

3.2.2 Equations aux différences finies. Formulation matricielle

On s'intéresse ici aux problèmes à valeurs aux bords, décrits par un système linéaire d'équations aux dérivées partielles (EDP). Comme exemple type, on considèrera le problème électrostatique, décrit par un potentiel scalaire $\phi(\vec{x})$, satisfaisant l'équation de Poisson :

$$\nabla^2 \phi(\vec{x}) = -\rho(\vec{x})/\varepsilon_0 \,, \quad \forall \vec{x} \in \Omega \,, \tag{3.28}$$

où $\rho(\vec{x})$ est le champ densité de charge, ε_0 la permittivité du vide et Ω le domaine spatial considéré. Le problème a une solution unique si on pose des conditions aux limites, ou conditions aux bords

$$\phi(\vec{x}) = V(\vec{x}) \,, \quad \forall \vec{x} \in \partial\Omega \,, \tag{3.29}$$

avec $V(\vec{x})$ une fonction connue sur le bord $\partial\Omega$ du domaine Ω.

La structure générale de telles équations peut s'écrire

$$\boxed{\mathcal{L}(\phi(\vec{x})) = b(\vec{x})}, \;\; \forall \vec{x} \in \Omega; \;\; \phi(\vec{x}) = V(\vec{x}) \,, \;\; \forall \vec{x} \in \partial\Omega \,, \tag{3.30}$$

avec $\mathcal{L}$ un opérateur différentiel linéaire.

1. En fait on devrait parler de "magnétostationnaire", puisqu'il s'agit du champ magnétique créé par écoulement stationnaire de charges.

La discrétisation consiste à définir un maillage $\{(x_i, y_j, z_k)\}$, équidistant dans chaque direction, avec $h_x = \Delta x = x_{i+1} - x_i$, $h_y = \Delta y = y_{j+1} - y_j$, $h_z = \Delta z = z_{k+1} - z_k$. On approxime ensuite les opérateurs différentiels apparaissant dans le système d'EDP par des différences finies. On obtient ainsi un système d'équations linéaires **algébrique** pour les inconnues discrétisées $\phi_{i,j,k} = \phi(x_i, y_j, z_k)$ qui approxime le problème exact. Par exemple, les différences finies d'ordre le plus bas, Eq.(A.7), donnent

$$\left.\frac{\partial^2\phi}{\partial x^2}\right|_{i,j,k} \approx \frac{1}{h_x^2}\left(\phi_{i-1,j,k} - 2\phi_{i,j,k} + \phi_{i+1,j,k}\right) \ . \tag{3.31}$$

On fait de même pour $\partial^2\phi/\partial y^2$ et $\partial^2\phi/\partial z^2$ pour obtenir le système linéaire d'équations algébriques approximant l'équation de Poisson :

$$\begin{array}{l} \frac{-2\left(h_y^2h_z^2+h_x^2h_z^2+h_x^2h_y^2\right)}{h_x^2h_y^2h_z^2}\,\phi_{i,j,k} + \frac{1}{h_x^2}\left(\phi_{i-1,j,k} + \phi_{i+1,j,k}\right) \\ +\frac{1}{h_y^2}\left(\phi_{i,j-1,k} + \phi_{i,j+1,k}\right) + \frac{1}{h_z^2}\left(\phi_{i,j,k-1} + \phi_{i,j,k+1}\right) \quad = \quad b_{i,j,k} \ . \end{array} \tag{3.32}$$

Il est intéressant d'écrire ces équations dans le cas particulier $h_x = h_y = h_z = h$:

$$\frac{1}{h^2}\left(-6\phi_{i,j,k} + \phi_{i-1,j,k} + \phi_{i+1,j,k} + \phi_{i,j-1,k} + \phi_{i,j+1,k} + \phi_{i,j,k-1} + \phi_{i,j,k+1}\right) = b_{i,j,k} \ . \tag{3.33}$$

Dans le cas $\rho(\vec{x}) = 0, \forall\vec{x} \in \Omega - \partial\Omega$, (équation de Laplace : potentiel électrostatique dans le vide), on a $b_{i,j,k} = 0$ pour tous les points intérieurs au domaine Ω. On a alors

$$\phi_{i,j,k} = \frac{1}{6}\left(\phi_{i-1,j,k} + \phi_{i+1,j,k} + \phi_{i,j-1,k} + \phi_{i,j+1,k} + \phi_{i,j,k-1} + \phi_{i,j,k+1}\right) \ , \forall(x_i, y_j, z_k) \notin \partial\Omega \tag{3.34}$$

ce qui veut dire que le potentiel aux points intérieurs du maillage est la moyenne arithmétique des valeurs du potentiel aux points les plus proches voisins du maillage. Les valeurs du potentiel aux points du maillage situés sur le bord $\partial\Omega$ sont données directement par les conditions aux bords.

Dans tous les cas, on peut écrire le problème EDP linéaire discrétisé comme

$$\boxed{\mathbf{A}\Phi = \mathbf{b}} \tag{3.35}$$

avec $\mathbf{A}$ une matrice $N \times N$ avec $N = N_xN_yN_z$ le nombre total de points du maillage, Φ et $\mathbf{b}$ des vecteurs de N élements. Pour ce faire, il faut définir une **numérotation** des points de maillage. En 1-D, c'est trivial. En 2-D (et 3-D), il faut décider si on numérote d'abord en suivant x, puis y, (puis z), ou dans un autre ordre. Voir un exemple 2-D en FIG. 3.3. On obtient la matrice $\mathbf{A}$ avec une structure de bande ("multi-diagonale"). Dans

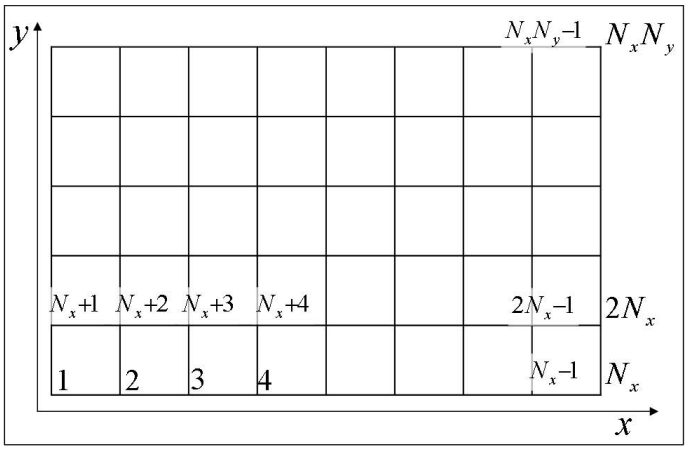


FIGURE 3.3 – *Numérotation des noeuds d'un maillage 2-D. Ici, on a numéroté en suivant d'abord selon x, puis selon y.*

le cas 2-D de l'équation de Poisson, avec $h_x = h_y = h_z = h$, on obtient :

$$\mathbf{A} = \frac{1}{h^2}\begin{pmatrix} -4 & 1 & . & . & 1 & . & . & . & . & . & . & . \\ 1 & -4 & 1 & . & . & 1 & . & . & . & . & . & . \\ . & 1 & -4 & 1 & . & . & 1 & . & . & . & . & . \\ . & . & 1 & -4 & . & . & . & 1 & . & . & . & . \\ 1 & . & . & . & -4 & 1 & . & . & 1 & . & . & . \\ . & 1 & . & . & 1 & -4 & 1 & . & . & 1 & . & . \\ . & . & 1 & . & . & 1 & -4 & 1 & . & . & 1 & . \\ . & . & . & 1 & . & . & 1 & -4 & . & . & . & 1 \\ . & . & . & . & 1 & . & . & . & -4 & 1 & . & . \\ . & . & . & . & . & 1 & . & . & 1 & -4 & 1 & . \\ . & . & . & . & . & . & 1 & . & . & 1 & -4 & 1 \\ . & . & . & . & . & . & . & 1 & . & . & 1 & -4 \end{pmatrix} \tag{3.36}$$

La matrice $\mathbf{A}$ ci-dessus doit encore être modifiée pour inclure les conditions aux bords. Il faut remplacer, pour tous les indices de ligne et colonne qui correspondent à un point du bord, l'équation par la condition au bord correspondante : on met 1 sur la diagonale de A et on met la valeur au bord $V(x_i, y_j, z_j)$ à l'élément correspondant du vecteur $\mathbf{b}$.

Si on n'inclut pas ces conditions aux limites, la matrice $\mathbf{A}$ est singulière et il est impossible de résoudre le système linéaire $\mathbf{A}\Phi = \mathbf{b}$.

3.2.3 Résolution du système linéaire. Méthodes directes (Gauss) et itératives (Jacobi, Gauss-Seidel, SOR)

Il y a deux groupes de méthodes pour résoudre le système linéaire $\mathbf{A}\Phi = \mathbf{b}$. Les méthodes **directes** et les méthodes **itératives**. On se limitera ici à en rappeler et en décrire quelques-unes.

Méthodes directes : élimination de Gauss, décomposition $A = LL^T$, décomposition $A = LDU$.

Méthodes itératives : La plupart sont applicables seulement aux matrices symétriques définies positives. C'est notamment le cas de l'opérateur Laplacien discrétisé que nous étudions, (mais ce n'est pas le cas de toute équation différentielle !). Ecrivons la matrice $\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{R}$, avec $\mathbf{L}$ une matrice triangulaire inférieure, $\mathbf{D}$ une matrice diagonale et $\mathbf{R}$ une une matrice triangulaire supérieure.

La *méthode de Jacobi* consiste à résoudre, à la $(n+1)$-ième itération,

$$\mathbf{D}\Phi^{(n+1)} = \mathbf{b} - \mathbf{L}\Phi^{(n)} - \mathbf{R}\Phi^{(n)} \ . \tag{3.37}$$

On notera que dans le cas du problème de Poisson dans le vide, on a un système homogène ($\mathbf{b} = 0$), et la méthode de Jacobi revient à prendre, à chaque itération, la moyenne des plus proches voisins, voir Eq.(3.34).

La *méthode de Gauss-Seidel* consiste à résoudre, à la $(n+1)$-ième itération :

$$(\mathbf{L} + \mathbf{D})\Phi^{(n+1)} = \mathbf{b} - \mathbf{R}\Phi^{(n)} \ . \tag{3.38}$$

Ce système est facile à résoudre en commençant à un bout de la matrice et en résolvant une ligne après l'autre, séquentiellement (substitution "forward").

La *méthode SOR*, dite aussi de *sur-relaxation*, est une façon d'accélérer la convergence de la méthode de Gauss-Seidel. Soit α un nombre réel. En multipliant l'Eq.(3.38) par α et en soustrayant formellement $(\alpha - 1)\mathbf{D}\Phi$ de part et d'autre, on résout, à la $(n+1)$-ième l'itération :

$$(\alpha\mathbf{L} + \mathbf{D})\Phi^{(n+1)} = \alpha\mathbf{b} - (\alpha\mathbf{R} + (\alpha - 1)\mathbf{D})\,\Phi^{(n)} \ . \tag{3.39}$$

On peut remarquer que la matrice $((\alpha\mathbf{L} + \mathbf{D})$ reste triangulaire, et on résout, comme Gauss-Seidel, par substitution. Notons qu'il n'y a pas besoin de stocker explicitement les matrices $\mathbf{L}, \mathbf{D}, \mathbf{R}$, etc., en mémoire. L'implémentation de cet algorithme revient, *à l'intérieur* de la boucle de substitution (ligne no.i), à faire l'étape Gauss-Seidel

$$\phi^{(*)} = \left(b_i - L\Phi^{(n+1)} - R\Phi^{(n)}\right)/D_i \ , \tag{3.40}$$

puis la surrelaxation proprement dite :

$$\Phi_i^{(n+1)} = \Phi_i^{(n)} + \alpha\left(\phi^{(*)} - \Phi_i^{(n)}\right) \ . \tag{3.41}$$

Avec $\alpha < 1$ on a une méthode de sous-relaxation, avec $\alpha = 1$ on retrouve la méthode de Gauss-Seidel, avec $1 < \alpha < 2$ on parle de sur-relaxation, alors qu'avec $\alpha \geq 2$ l'algorithme diverge.

On définit le *résidu* à l'itération n

$$r^{(n)} = ||\mathbf{b} - \mathbf{A}\Phi^{(n)}|| \tag{3.42}$$

et on choisit un critère d'arrêt pour les itérations $r < \epsilon$, où ϵ est une précision requise.

Il y a bien d'autres méthodes itératives. Par exemple, celles basées sur les *gradients conjugués*, qui consistent à choisir judicieusement les directions des relaxations successives. Elles nécessitent généralement un *préconditionnement* de la matrice pour être efficaces.

Mentionnons que de nombreuses librairies numériques sont disponibles pour résoudre les systèmes algébriques linéaires.

3.2.4 Electrostatique en 2-D, différences finies, GS-SOR. Convergence des itérations

On applique la discrétisation par différences finies au problème d'un condensateur rectangulaire, de taille finie. On montre des exemples aux FIGS.3.4 et 3.5. Les méthodes de Gauss-Seidel sans et avec sur-relaxation (SOR) convergent vers la même solution. Pour un maillage $N_x = 101,\ N_y = 61$, après quelques dizaines d'itérations SOR avec $\alpha = 1.9$, la solution est convergée "à l'oeil nu", c.à.d. qu'on ne distingue plus de différence, à l'échelle de la figure, sur les lignes de niveau du potentiel. Pour ces figures, on a utilisé le critère d'arrêt des itérations : résidu $r < \epsilon = 10^{-6}$.

On remarque, au niveau de la physique du résultat, le champ électrique plus intense vers les angles du conducteur intérieur. (Le champ électrique étant le gradient du potentiel, des équipotentielles serrées indiquent un champ électrique intense). La densité de charge à la surface d'un conducteur, et donc l'intensité du champ électrique dans son voisinage, est inversément proportionnelle au rayon de courbure de la surface (voir cours de Physique III-IV). C'est *l'effet de pointe*. Sur l' image du bas de la FIG. 3.5, l'effet de renforcement de l'intensité du champ électrique est encore plus manifeste lorsque l'électrode intérieure est mince et est placée à proximité du conducteur extérieur. S'il y a claquage, c'est vers la pointe que cela se produira : c'est sur ce principe que sont basés les paratonnerres.

On peut utiliser la loi de Gauss pour le champ électrique

$$\oiint_S \vec{E} \cdot d\vec{\sigma} = Q_{\text{enf}}/\varepsilon_0 \tag{3.43}$$

pour calculer la charge sur le conducteur interne. Pour ce faire, il faut d'abord calculer le champ $\vec{E}$ à partir de la solution numérique $\phi(x_i, y_j)$. On peut le faire par exemple en utilisant les différences finies centrées du premier ordre, Eq.(A.16),

$$\begin{array}{rcl} E_x|_{i+1/2} & = & -\partial\phi/\partial x|_{i+1/2} \approx -(\phi(i+1,j) - \phi(i,j))/\Delta x \\ E_y|_{j+1/2} & = & -\partial\phi/\partial y|_{j+1/2} \approx -(\phi(i,j+1) - \phi(i,j))/\Delta y \end{array} \tag{3.44}$$

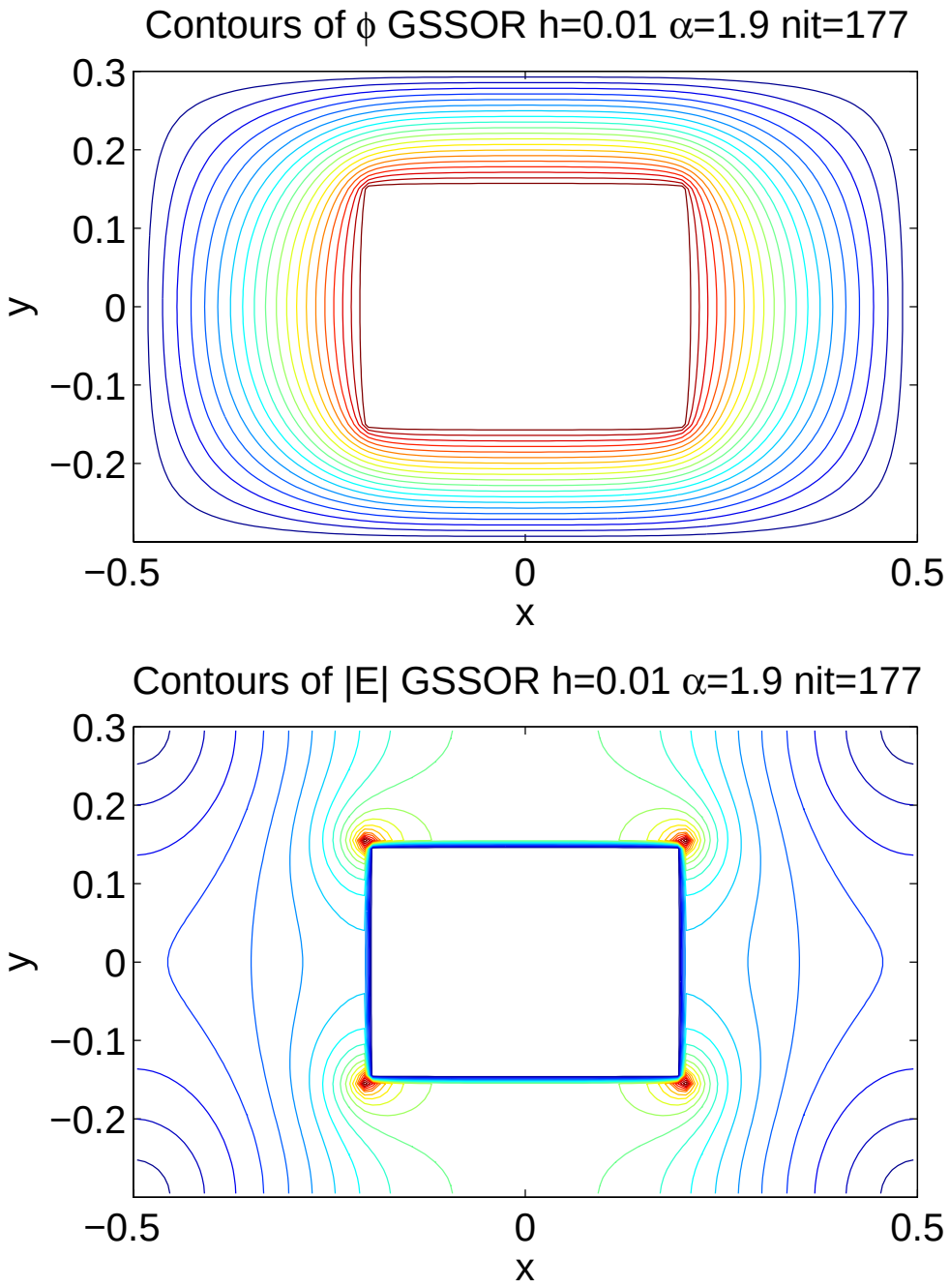


FIGURE 3.4 – *Condensateur rectangulaire. Conditions aux bords* $\phi = 1$V *sur le conducteur intérieur,* $\phi = 0$ *sur le conducteur extérieur. Méthode de différences finies, maillage* $N_x = 101$, $N_y = 61$ *. Problème matriciel résolu avec Gauss-Seidel et SOR, paramètre de sur-relaxation* $\alpha = 1.9$*, précision requise : résidu* $r < \epsilon = 10^{-6}$*. En haut : lignes de niveau du potentiel (de 0 à 1V). En bas : lignes de niveau de* $|\vec{E}|$ *(de 0 à 13.6 V/m).*

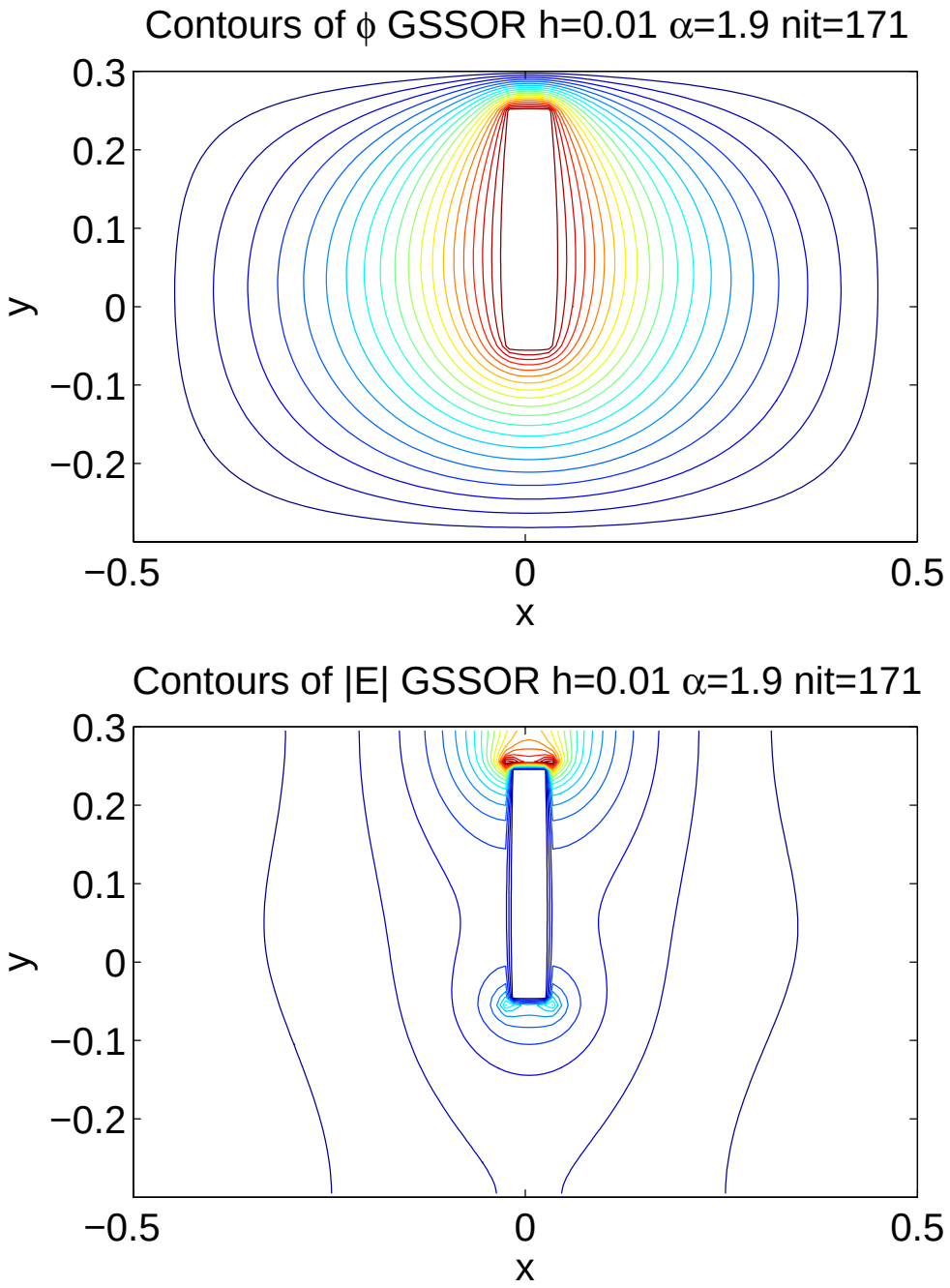


FIGURE 3.5 – *Condensateur rectangulaire. Conditions aux bords* $\phi = 1\text{V}$ *sur le conducteur intérieur,* $\phi = 0$ *sur le conducteur extérieur. Méthode de différences finies, maillage* $N_x = 101,\ N_y = 61$. *Problème matriciel résolu avec Gauss-Seidel et SOR, paramètre de sur-relaxation* $\alpha = 1.9$, *précision requise : résidu* $r < \epsilon = 10^{-6}$. *En haut : lignes de niveau du potentiel (de 0 à 1V). En bas : lignes de niveau de* $|\vec{E}|$ *(de 0 à 26.4 V/m).*

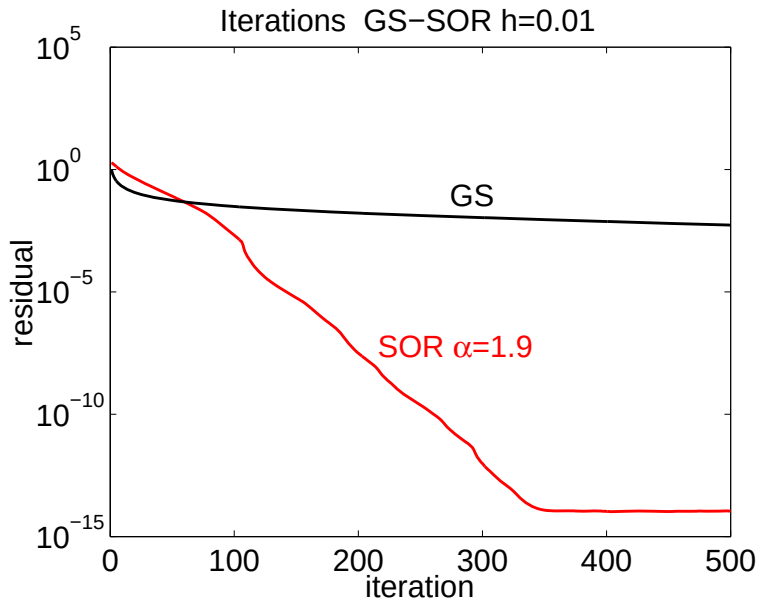


FIGURE 3.6 – *Convergence du résidu avec les itérations des méthodes de Gauss-Seidel et de sur-relaxation (SOR), pour le cas de la FIG. 3.5.*

Il faut noter que le champ $\vec{E}$ est défini aux milieux des segments du maillage, et que E_x et E_y ne sont ainsi pas définis aux mêmes points. On choisit ensuite une surface fermée S entourant le conducteur interne. On prendra pour surface S un rectangle aligné avec les milieux des cellules du maillage. Enfin, on doit effectuer l'intégrale. On choisira la méthode d'ordre le plus bas, étant donné que l'approximation par différences finies que nous avons choisie est d'ordre le plus bas. On prendra garde à l'orientation de $\vec{d\sigma}$ sur les 4 faces du rectangle.

Au niveau de la numérique, on étudie la convergence des itérations pour résoudre le problème matriciel (FIG. 3.6). Avec sur-relaxation (SOR) et un paramètre de sur-relaxation $\alpha = 1.9$, la convergence est fortement accélérée par rapport à Gauss-Seidel ($\alpha = 1$). On remarque clairement aussi qu'on ne peut pas aller en dessous d'un résidu de 10^{-14} : on a alors atteint la précision machine (64 bits dans ce cas).

Un autre test de validation de la solution numérique consiste à vérifier la loi de Gauss pour des surfaces fermées S différentes. Pour toute surface fermée S entourant le conducteur interne et entièrement contenue à l'intérieur du conducteur externe, l'intégrale de Gauss (flux du champ électrique à travers S) devrait donner le *même résultat*. Les différences sont ainsi une mesure de l'erreur numérique. *La valeur du résultat permet de calculer, avec la relation $Q = C\Delta V$, la capacité du système.* Pour toute surface fermée n'entourant aucune partie de conducteur, la charge enfermée est nulle, et le flux du champ électrique devrait être nul. Dans ce cas, la valeur du résultat fournit une autre mesure de l'erreur numérique.

L'erreur numérique a en principe trois origines distinctes : premièrement, l'erreur sur la convergence de la méthode itérative pour résoudre le système matriciel (Gauss-Seidel ou SOR) ; deuxièmement, l'erreur de discrétisation venant du fait que la taille du maillage est finie (h), ou erreur de troncature ; troisièmement, les erreurs d'arrondi. Pour le cas de la FIG. 3.5, l'intégrale de Gauss donne, après 171 itérations de Gauss-Seidel SOR ($\alpha = 1.9$), 6.1511173 pour S tout près du conducteur interne et 6.1511178 pour S tout près du conducteur externe, soit une erreur relative de 10^{-7}. Pour S n'entourant que du vide, on obtient -3.4×10^{-7} (au lieu de zéro). On vérifie que l'erreur semble tendre vers zéro avec le nombre d'itérations SOR. L'erreur relative ne peut toutefois pas être inférieure à la précision machine : les erreurs d'arrondi empêchent d'aller à des précisions encore meilleures.

Un test similaire, mais avec un réseau plus grossier, conduit au même résultat qualitatif (mais avec un nombre d'itérations GSSOR moins élevé).

Ces tests montrent que, pour le schéma numérique considéré, la loi de Gauss pour le champ électrique est satisfaite "exactement" (à la précision machine), indépendamment des erreurs de troncature, pour autant que l'on résolve le système algébrique linéaire "exactement" (à la précision machine) et que l'on choisisse des surfaces fermées S passant par les milieux des cellules du réseau. Lorsqu'on choisit d'autres surfaces S, les erreurs de troncature réapparaissent à cause des interpolations que l'on doit faire.

Satisfaire la loi de Gauss pour certaines surfaces S bien choisies avec une précision machine ne veut pas dire que la précision sur la solution du problème est atteinte à la précision machine quelle que soit la taille du maillage. Il reste les erreurs de troncature. On les examinera en exercice, en considérant la solution ϕ en des endroits particuliers, et en observant comment la solution converge en ces endroits, en prenant des maillages de plus en plus fins.

Suggestion d'exercice. Calculer le potentiel créé par une paire de conducteurs minces de taille finie placés dans une boîte conductrice rectangulaire. Vérifier le théorème de Gauss en prenant différentes surfaces fermées entourant l'un ou l'autre conducteur, ou les deux. On montre à la FIG. 3.7 un exemple de configuration asymétrique. Le flux du champ électrique pour une surface S entourant les deux conducteurs intérieurs est-il nul ? Le flux du champ électrique est-il le même, au signe près, pour S_1 autour du conducteur de gauche et pour S_2 autour du conducteur de droite ?

Suggestion d'exercice. Calculer le potentiel créé par une charge ponctuelle, en coordonnées sphériques. Indication : la singularité du problème en $r = 0$ peut être traitée en considérant une charge non pas parfaitement ponctuelle, mais avec une taille finie. Il faut veiller à ce que cette taille ne soit pas plus petite que la taille du réseau Δr, afin d'avoir plusieurs points de discrétisation sur la "particule". Comparer avec le résultat analytique exact.

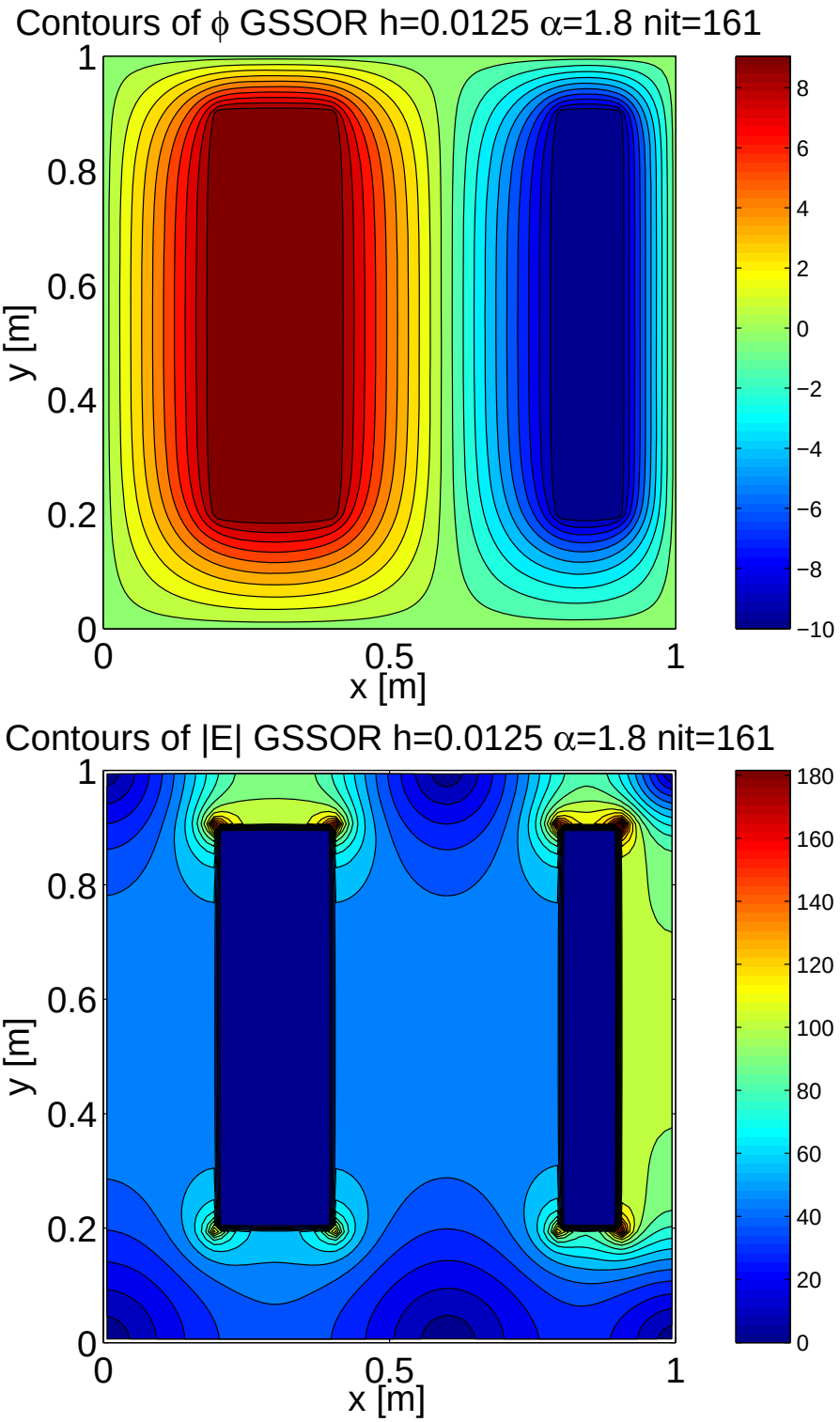


FIGURE 3.7 – *Condensateur rectangulaire asymétrique. Conditions aux bords* $\phi = 10\text{V}$ *sur le conducteur intérieur de gauche,* $\phi = -10\text{V}$ *sur le conducteur intérieur de droite,* $\phi = 0$ *sur le conducteur extérieur. Méthode de différences finies, maillage* $N_x = 81$, $N_y = 81$ *. Problème matriciel résolu avec Gauss-Seidel et SOR, paramètre de sur-relaxation* $\alpha = 1.8$*, précision requise : résidu* $r < \epsilon = 10^{-4}$*. En haut : lignes de niveau du potentiel (de -10V à +10V). En bas : lignes de niveau de* $|\vec{E}|$ *(de 0 à 191 V/m).*

3.2.5 Optimisation et complexité de l'algorithme

Dans le problème de l'équation de Laplace 2-D, avec N points de maillage dans chaque direction, on a au minimum de l'ordre de N^2 opérations à effectuer, en supposant un algorithme "idéal", qui donnerait la solution ϕ_{ij} en une seule itération. Un tel algorithme n'existe généralement pas, sauf bien sûr dans les cas où une solution analytique exacte du problème peut être trouvée.

Avec le schéma aux différences finies et l'algorithme de Gauss-Seidel, on trouve que le nombre d'itérations requis pour atteindre une précision donnée est proportionnel à N^2 environ. Donc, puisque chaque pas de Gauss-Seidel implique de l'ordre de N^2 opérations, on a un algorithme qui coûte de l'ordre de N^4 opérations. Pour des tailles de maillage importantes, ceci peut vite devenir prohibitif.

Avec la sur-relaxation (SOR), on peut améliorer considérablement les choses. La FIG. 3.8 (haut) montre le nombre d'itérations requis pour une précision sur le résidu $r < \epsilon = 10^{-3}$ en fonction du paramètre de sur-relaxation α, pour différentes tailles N du maillage $N \times N$ utilisé. Le cas physique est celui de la FIG. 3.7. Le choix optimal de α dépend de N, et est empiriquement donné par $\alpha_{\rm opt} \approx 2 - {\rm const}/N$ (image du milieu). Le nombre d'itérations requis pour atteindre une précision donnée est *proportionnel à* N (image du bas). Le coût de l'algorithme SOR à l'optimum est donc d'ordre N^3, et non N^4 comme Gauss-Seidel sans sur-relaxation.

On pourrait résoudre le système algébrique linéaire $\mathbf{A}\Phi = \mathbf{b}$, Eq.(3.35), "d'un seul coup", c'est-à-dire en une itération. Ce sont les méthodes dites *directes* qui permettent de le faire, par exemple l'élimination de Gauss, ou via la décomposition de Cholesky $\mathbf{A} = \mathbf{L}\mathbf{L}^T$. Si on résout ainsi, en considérant toute la matrice $\mathbf{A}$ commme une matrice carrée $N^2 \times N^2$, l'algorithme requiert de l'ordre de $(N^2)^3$ opérations arithmétiques. On obtient un coût qui est de l'ordre de N^6, qui devient vite exorbitant. On peut faire mieux en considérant le fait que la matrice $\mathbf{A}$ a, pour un maillage recangulaire structuré, une structure de matrice de *bande* : voir Eq.(3.36). La largeur de la bande est proportionnelle au nombre de points de maillage dans une direction, N. Les algorithmes directs d'élimination de Gauss ou de factorisation de Cholesky nécessitent un nombre d'opérations proportionnel au cube de la largeur de bande. On aboutit donc à un coût de l'algorithme proportionnel à N^4. Il y a un prix à payer supplémentaire, en terme de mémoire, puisqu'il faut stocker la matrice $\mathbf{A}$ et sa décomposition.

La conclusion de cette discussion est que l'algorithme SOR semble le plus performant pour ce genre de problème (Laplacien). Des difficultés apparaissent lorsque le domaine de résolution est de géométrie plus complexe et nécessite un maillage non-équidistant, auquel cas il peut être dificile de trouver un α optimal. Lorsque d'autres opérateurs sont considérés, il se peut que les méthodes itératives simples, comme SOR, ne convergent tout

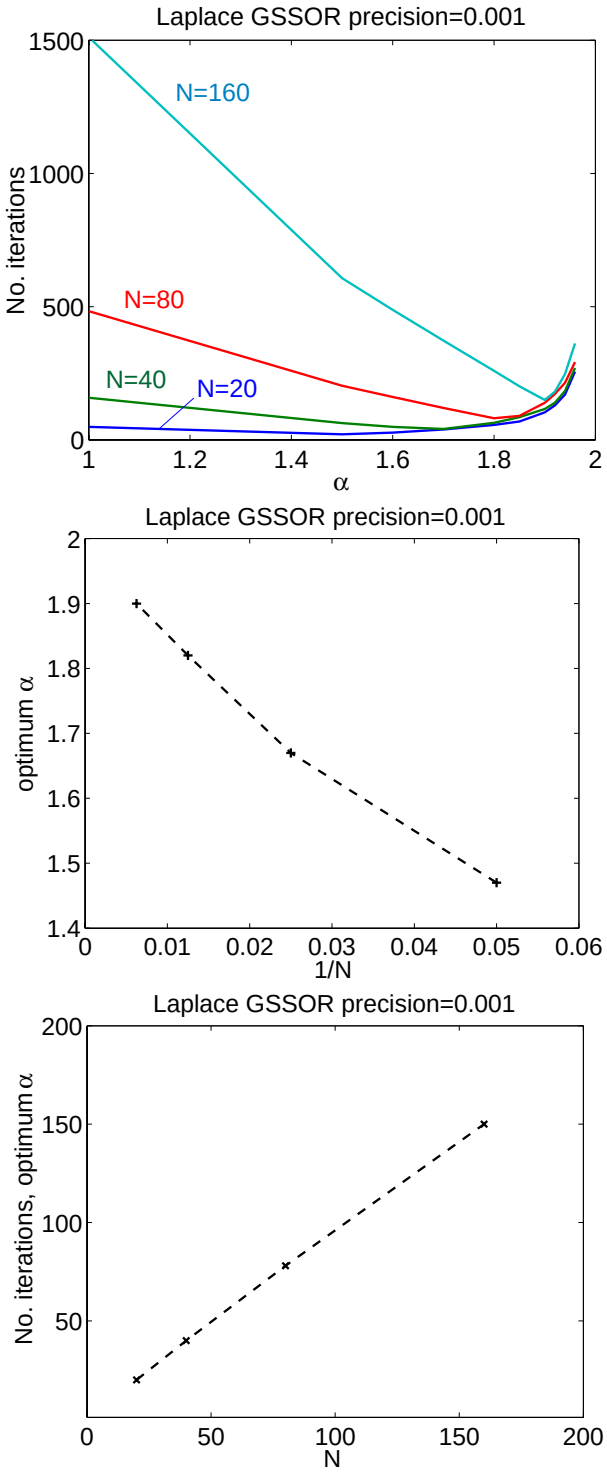


FIGURE 3.8 – *Nombre d'itérations SOR en fonction de α, pour l'équation de Laplace 2D résolue par différences finies sur un maillage $N \times N$ (haut). Paramètre SOR α optimal en fonction de $1/N$ (milieu). Nombre d'itérations SOR requis pour une précision $\epsilon = 10^{-3}$ en fonction de N (bas). Le cas physique correspond au condensateur rectangulaire asymétrique de la FIG. 3.7.*

simplement pas ![2] Dans ce cas, les méthodes directes peuvent être les plus appropriées.

3.2.6 Géométrie plus complexe

Considérons le problème d'une paire électrodes conductrices aux potentiels V_a et V_b, placées dans le vide à l'intérieur d'une boîte rectangulaire conductrice au potentiel 0. C'est la même situation qu'à la FIG. 3.7, mais cette fois on considère des électrodes de formes non rectangulaires.

On résout, comme précédemment, avec la méthode des différences finies sur un maillage cartésien (x_i, y_j) et l'algorithme GS-SOR. Par exemple, choisissons des électrodes elliptiques placées avec une orientation quelconque par rapport aux axes (x, y). La FIG. 3.9 montre la solution numérique obtenue pour $V_a = +10\text{V}$, $V_b = -5\text{V}$, des ellipses de demi-axes $a \times b = 0.35 \times 0.1\text{m}$ et $0.25 \times 0.15\text{m}$, centrées en $(0.25, 0.5)\text{m}$ et $(0.7, 0.3)\text{m}$, avec leurs grands-axes inclinés de 80 et 150 degrés par rapport à l'axe x, respectivement. La boîte extérieure est carrée de côté $L = 1\text{m}$. Le maillage est 160×160, et la paramètre de surrelaxation est $\alpha = 1.9$. Pour atteindre un résidu inférieur à $\epsilon = 10^{-3}$, 147 itérations GS-SOR sont nécessaires.

Si l'algorithme converge bien, dans le sens que la solution numérique $\phi_{i,j}$ converge avec les itérations SOR, il y a un problème avec le champ électrique au voisinage des surfaces des électrodes. Des irrégularités apparaissent, qui ne sont pas physiques, mais qui sont dues au fait que ces conducteurs sont représentés sur un maillage cartésien rectangulaire, et que la surface des conducteurs n'est pas alignée avec les lignes de coordonnées. Cela implique que la représentation numérique de la surface, supposée lisse en réalité, est en "marche d'escalier", et on voit en fait un effet de pointe purement numérique aux coins de ces "marches d'escalier". *La solution pour $\vec{E}$ n'est donc pas bonne au voisinage des surfaces des électrodes.*

La solution à ce problème dépasse le cadre de ce cours. Mentionnons quand même la méthode des *éléments finis*, qui peut être utilisée avec des maillages dont les noeuds peuvent être plaçés le long des surfaces. On peut également utiliser les éléments finis sur des systèmes de coordonnées curvilignes. Des méthodes de raffinement du maillage dans les régions de fort gradient de la solution ont également été développées.

2. Ces méthodes ne convergent que si la matrice $\mathbf{A}$ est symétrique positive définie.

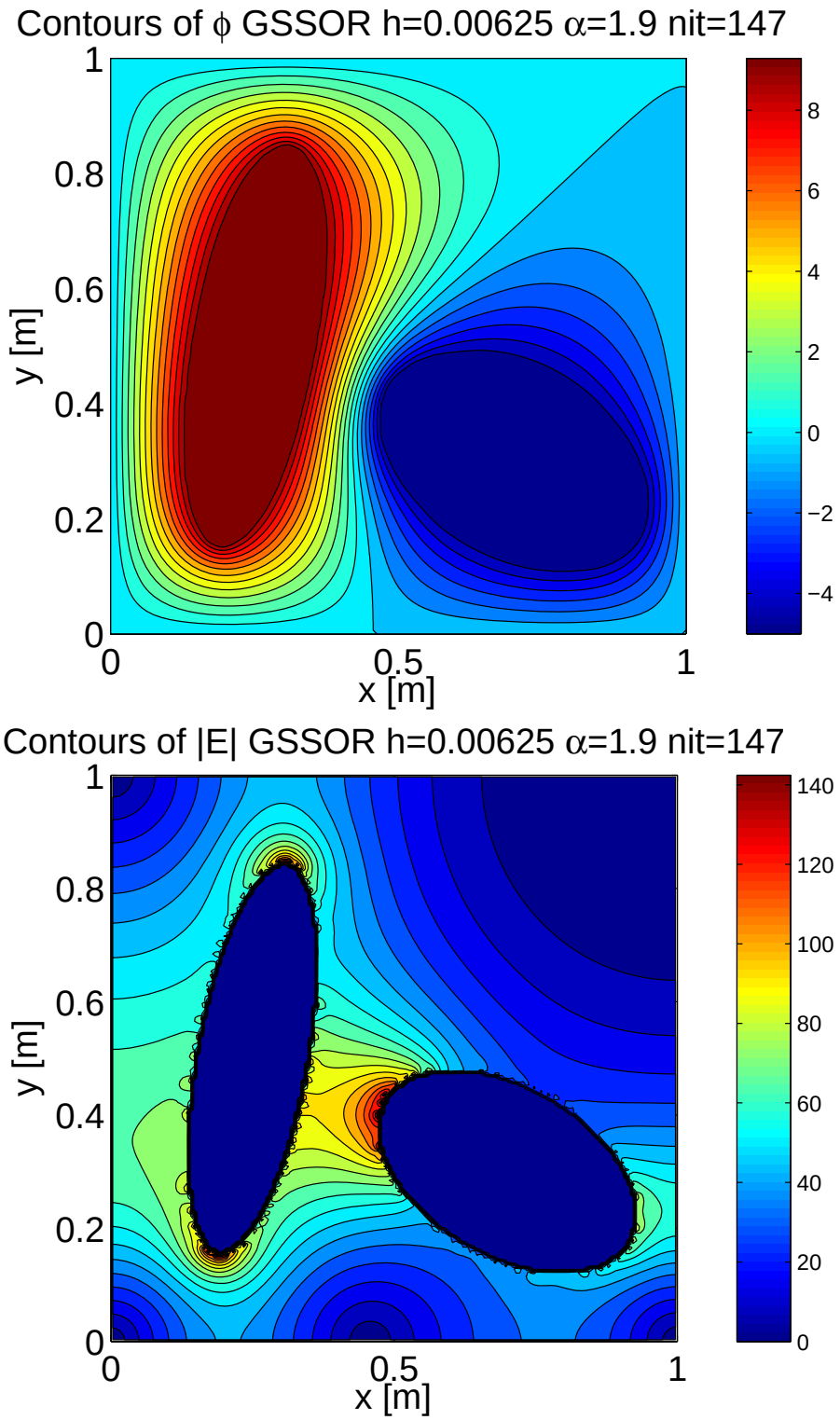


FIGURE 3.9 – *Condensateur avec électrodes elliptiques. Conditions aux bords* $\phi = 10\text{V}$ *sur le conducteur intérieur de gauche,* $\phi = -5\text{V}$ *sur le conducteur intérieur de droite,* $\phi = 0$ *sur le conducteur extérieur. Méthode de différences finies, maillage* $N_x = 161$, $N_y = 161$ *. Problème matriciel résolu avec Gauss-Seidel et SOR, paramètre de sur-relaxation* $\alpha = 1.9$*, précision requise : résidu* $r < \epsilon = 10^{-3}$*. En haut : lignes de niveau du potentiel (de -5V à +10V). En bas : lignes de niveau de* $|\vec{E}|$ *(de 0 à 150 V/m).*

3.3 Forme variationnelle. Eléments finis

Soit un système d'équations aux dérivées partielles (EDP) avec conditions aux limites de type Dirichlet [3] :

$$\boxed{\mathcal{L}(\phi(\vec{x})) = b(\vec{x})}, \quad \boxed{\forall \vec{x} \in \Omega}; \quad \boxed{\phi(\vec{x}) = V(\vec{x}) \ , \ \forall \vec{x} \in \partial\Omega}, \tag{3.45}$$

avec $\mathcal{L}$ un opérateur différentiel linéaire.

3.3.1 Description de la méthode

La méthode des éléments finis pour obtenir une approximation de la solution à ces EDP est construite sur les bases suivants, dont la **forme variationnelle** des équations est l'un des piliers, l'autre étant l'approximation des fonctions en développant sur une **base de fonctions élémentaires de support fini**. La démarche est consituée des points suivants.

1. La définition d'un produit scalaire

$$(\eta, \phi) = \int_\Omega \eta(\vec{x})\phi(\vec{x}) d^3x \tag{3.46}$$

   et de la norme $||\phi|| = \sqrt{(\phi, \phi)}$.

2. La construction d'une forme variationnelle (dite "faible"), en choisissant une *fonction test* $\eta(\vec{x})$, multipliant l'Eq.(3.45) par $\eta(\vec{x})$, puis intégrant sur le domaine Ω. L'EDP avec conditions aux limites, Eq.(3.45), est donc équivalente au problème variationnel suivant : trouver $\phi \in \mathcal{C}^n(\Omega)$ telle que

$$\boxed{(\eta, \mathcal{L}(\phi)) = (\eta, b)}, \ \boxed{\forall \eta(\vec{x}) \in \mathcal{C}^n(\Omega) | \eta(\vec{x}) = 0, \forall \vec{x} \in \partial\Omega}; \boxed{\phi(\vec{x}) = V(\vec{x}), \forall \vec{x} \in \partial\Omega}. \tag{3.47}$$

3. Une intégration par parties. Pour illustration, nous prendrons le cas de l'opérateur de Laplace, $\mathcal{L} = \nabla^2$:

$$(\eta, \nabla^2\phi) = \int_\Omega \eta \nabla^2\phi \, d^3x = \int_\Omega \left(-\nabla\eta \cdot \nabla\phi + \nabla \cdot (\eta\nabla\phi)\right) d^3x$$

   On applique ensuite le théorème de Gauss (appelé aussi *théorème de la divergence*) au dernier terme, pour obtenir :

$$(\eta, \nabla^2\phi) == -\int_\Omega \nabla\eta \cdot \nabla\phi \, d^3x + \int_{\partial\Omega} \eta\nabla\phi \cdot d\vec{\sigma} \ . \tag{3.48}$$

   Le dernier terme est parfois nul, selon les conditions aux bords : on parle dans ce cas de *conditions aux bords naturelles*. Pour les opérateurs $\mathcal{L}$ symétriques que

3. Ici de type Dirichlet, mais on peut avoir d'autres conditions : Neuman, mixtes ou périodiques.

nous considèrerons dans ce cours, l'intégration par parties conduit à symétriser explicitement la forme variationnelle :

$$-(\hat{\mathcal{L}}(\eta), \hat{\mathcal{L}}(\phi)) = (\eta, b) \ , \quad \forall \eta(\vec{x}) \in \mathcal{C}^n(\Omega) \ . \tag{3.49}$$

Pour notre illustration, $\mathcal{L} = \nabla^2$ et $\hat{\mathcal{L}} = \nabla$.

4. Une approximation numérique du problème variationnel. L'idée de base est de considérer un sous-espace de l'espace des fonctions, noté $\mathcal{C}^p(\Omega_h)$, qui est celui des fonctions continues différentiables d'ordre p par morceaux, représentables sur une *base de fonctions* $\Lambda_i(\vec{x})$ ayant un support de taille finie, définie sur une discrétisation (un maillage) de l'espace :

$$\phi(\vec{x}) = \sum_j \phi_j \Lambda_j(\vec{x}) \ . \tag{3.50}$$

On fait de même pour la fonction test η. Si la fonction test est choisie avec les mêmes fonctions de base,

$$\eta(\vec{x}) = \sum_i \eta_i \Lambda_i(\vec{x}) \ , \tag{3.51}$$

on obtient ce qui s'appelle la *méthode de Galerkin*. Les fonctions de base Λ_i sont généralement des polynômes par morceaux. Pour fixer les idées, on a représenté à la FIG. 3.10 le cas 1-D des éléments finis linéaires ($p = 1$), sur un maillage $\{x_j\}$. Noter que les points du maillage ne doivent *pas forcément être équidistants :* c'est une des souplesses importantes que permet la méthode des éléments finis (par rapport aux différences finies).

5. La substitution de cette approximation numérique, Eqs.(3.50-3.51), dans le problème variationnel (3.49) :

$$-\sum_i \sum_j \eta_i \left(\int_\Omega \hat{\mathcal{L}}(\Lambda_i) \hat{\mathcal{L}}(\Lambda_j) \, d^3x \right) \phi_j = \sum_i \eta_i \left(\int_\Omega \Lambda_i b \, d^3x \right) \ , \quad \forall \eta_i \ . \tag{3.52}$$

Définissons la matrice $\mathbf{A}$ avec $A_{i,j}$ = expression entre les grandes parenthèses du membre de gauche, le vecteur Φ des inconnues ϕ_j, et le vecteur $\mathbf{b}$ avec b_i = expression entre parenthèses du membre de droite. Le problème variationnel discrétisé ci-dessus, Eq.(3.52), doit être vérifié pour tout η_i, ce qui veut dire que l'égalité doit être satisfaite pour chaque terme de la somme sur i. On obtient ainsi un système d'équations linéaires *algébriques*

$$\boxed{\mathbf{A}\Phi = \mathbf{b}} \ . \tag{3.53}$$

$\mathbf{A}$ est une matrice $N \times N$, N étant le nombre de points du maillage. L'étape de construction des éléments de matrice $A_{i,j}$ nécessite le calcul d' intégrales faisant intervenir des produits des fonctions de base, $\Lambda_i \Lambda_j$, et de leurs dérivées, par exemple $(\partial\Lambda_i/\partial x)(\partial\Lambda_j/\partial x)$. Comme ces fonctions de base ont un support fini, ces intégrales sont nulles sauf pour i voisin de j. En conséquence, la matrice $\mathbf{A}$ a une structure de *matrice de bande*, dont la largeur dépend de l'ordre des éléments et de la dimensionalité du problème. Dans le cas d'éléments finis linéaires 1-D, FIG. 3.10, chaque élément Λ_i ne recouvre que les plus proches voisins, Λ_{i-1} et Λ_{i+1}, en plus de lui-même ; la matrice $\mathbf{A}$ est alors *tridiagonale*.

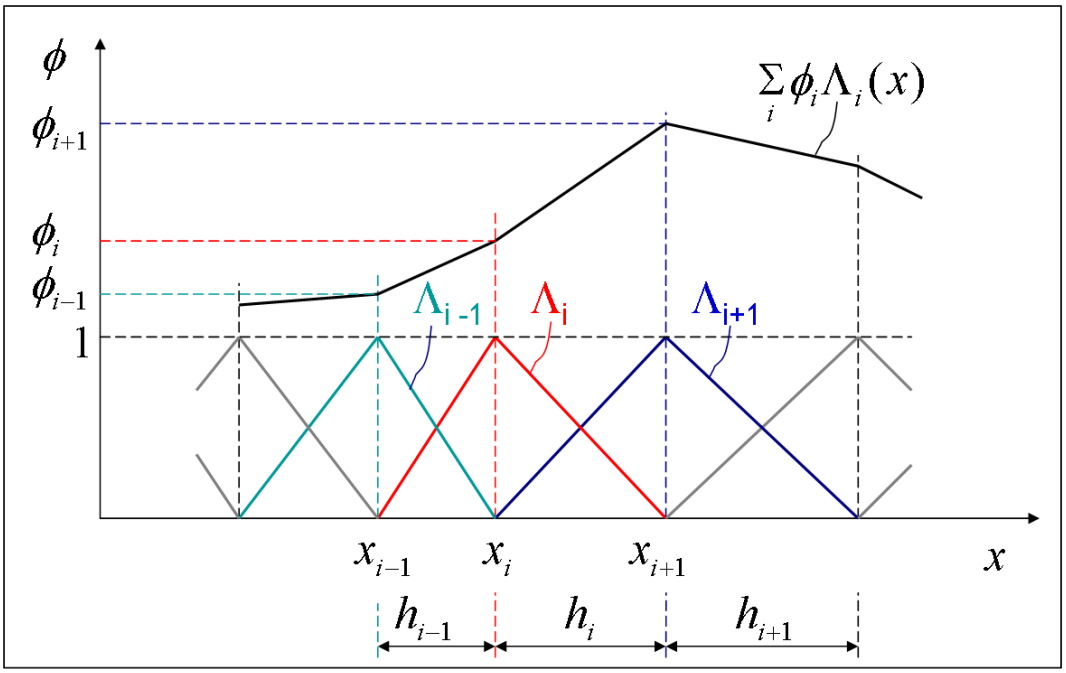


FIGURE 3.10 – *Eléments finis linéaires 1-D. Fonctions de base $\Lambda_i(x)$ et représentation (approximation) d'une fonction $\phi(x)$ par ces éléments finis.*

6. L'imposition des conditions aux bords *explicites*. On le fait généralement au niveau du système algébrique linéaire $\mathbf{A}\Phi = \mathbf{b}$. Par exemple, dans le cas de conditions de Dirichlet et d'éléments finis linéaires, on a

$$A_{i,i} = 1.0; \quad A_{i,j\neq i} = 0; \quad A_{j\neq i,i} = 0; \quad b_i = V(\vec{x}(x_i)) \tag{3.54}$$

pour tout i tel que $\vec{x}(x_i)$ soit un point sur la surface $\partial\Omega$. Dans le cas de conditions aux bords plus compliquées, et/ou d'éléments finis d'ordre plus élevé, leur application implique généralement de faire des combinaisons linéaires des lignes et colonnes de $\mathbf{A}$ et du membre de droite $\mathbf{b}$.

7. La résolution du système algébrique linéaire (3.53). Voir Section 3.2.3.

On notera au passage que l'intégration par parties permet de faire décroître l'ordre de l'opérateur différentiel : par exemple, pour une équation du 2e ordre, la forme variationnelle intégrée par parties ne fera intervenir que les dérivées du 1er ordre : cela permet l'utilisation de fonctions de base *linéaires* pour une équation différentielle qui au départ fait intervenir les *2e dérivées*.

Intégration numérique des éléments de matrice et du terme de droite

On se restreindra au cas 1-D. L'algorithme de construction de ces éléments consiste à calculer, pour chaque intervalle $[x_k, x_{k+1}]$, de taille $h_k = x_{k+1} - x_k$ (les h_k peuvent être tous différents), la contribution aux éléments de matrice et du terme de droite. Ce sont

des intégrales du type

$$\int_{x_k}^{x_{k+1}} f(x)dx \tag{3.55}$$

On peut utiliser la règle du point milieu ou celle des trapèzes (Annexe B) ou, encore mieux, un mélange des deux, avec un paramètre p compris entre 0 et 1 :

$$\int_{x_k}^{x_{k+1}} f(x)dx \approx h_k \left[p\,\frac{f(x_k)+f(x_{k+1})}{2} + (1-p)\, f\left(\frac{x_k+x_{k+1}}{2}\right)\right] . \tag{3.56}$$

3.3.2 Elements finis - Equation de Poisson 1-D

Soit Ω un intervalle $[x_a, x_b]$. On place des électrodes en $x = x_a$ et $x = x_b$, aux potentiels V_a et V_b. Entre les deux électrodes se trouve une distribution de charge, de densité $\rho(x)$ donnée. On aimerait calculer le potentiel $\phi(x)$ et le champ électrique $E_x(x)$ entre les deux électrodes.

Il faut donc trouver une solution au problème :

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}(x) = -\frac{\rho(x)}{\varepsilon_0} \;, \forall x \in]x_a, x_b[, \qquad \phi(x_a) = V_a,\ \phi(x_b) = V_b\ . \tag{3.57}$$

En suivant la méthode présentée à la section précédente, on construit la forme variationnelle en multipliant l'équation de Poisson (3.57) par une fonction test $\eta(x)$ et en intégrant entre x_a et x_b :

$$\int_{x_a}^{x_b} \eta \frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}\, dx = \int_{x_a}^{x_b} -\eta\rho/\varepsilon_0\, dx\ . \tag{3.58}$$

Intégrant par parties,

$$\int_{x_a}^{x_b} \frac{\mathrm{d}\eta}{\mathrm{d}x}\frac{\mathrm{d}\phi}{\mathrm{d}x}\, dx - \left[\eta\frac{\mathrm{d}\phi}{\mathrm{d}x}\right]_{x_a}^{x_b} = \int_{x_a}^{x_b} \eta\rho/\varepsilon_0\, dx\ . \tag{3.59}$$

Le terme intégré, aux bornes de l'intervalle Ω, peut être considéré comme nul ; en effet, comme ϕ est connu aux bords, par les conditions aux limites, il n'est pas nécessaire de faire la variation en ces points. Autrement dit, on peut choisir η nul en $x = x_a$ et en $x = x_b$. Le problème variationnel s'énonce alors comme suit : trouver une fonction $\phi(x)$ telle que l'équation

$$\int_{x_a}^{x_b} \frac{\mathrm{d}\eta}{\mathrm{d}x}\frac{\mathrm{d}\phi}{\mathrm{d}x}\, dx = \int_{x_a}^{x_b} \eta\rho/\varepsilon_0\, dx\ . \tag{3.60}$$

soit satisfaite *pour toute fonction $\eta(x)$ à valeur nulle en x_a et en x_b*, et telle que $\phi(x_a) = V_a,\ \phi(x_b) = V_b$.

Attention aux conditions sur la fonction-test η. Un exemple d'erreur de raisonnement est le suivant. Prenons pour simplifier la cas du vide, $\rho(x) = 0, \forall x \in [x_a, x_b]$. L'équation de Poisson devient une équation de Laplace :

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} = 0$$

Donc $\mathrm{d}\phi/\mathrm{d}x = C$, avec $C = \text{const}$, et $\phi(x) = Cx + D$, avec $D = \text{const}$. Avec les conditions aux limites, on trouve facilement la solution $\phi(x) = V_a + (V_b - V_a)(x - x_a)/(x_b - x_a)$. En considérant la forme variationnelle, Eq.(3.60) avec $\rho = 0$, et en posant $g = \mathrm{d}\eta/\mathrm{d}x$, on a :

$$\int_{x_a}^{x_b} g \frac{\mathrm{d}\phi}{\mathrm{d}x} \mathrm{d}x = 0, \forall g. \tag{3.61}$$

On en conclut

$$\frac{\mathrm{d}\phi}{\mathrm{d}x} = 0, \forall x \ .$$

Mais évidemment cela contredit la solution du problème ! Où est donc l'erreur ?

L'erreur est que nous avons oublié de transcrire la condition sur η : *(pour toute fonction $\eta(x)$ à valeur nulle en x_a et en x_b)* en une condition correspondante sur la fonction g. En effet, la condition sur η implique la condition suivante pour g :

$$\forall g | \int_{x_a}^{x_b} g(x) \mathrm{d}x = 0 \ . \tag{3.62}$$

Pour de telles fonctions, la forme variationnelle, Eq.(3.61) admet comme solution pour ϕ

$$\frac{\mathrm{d}\phi}{\mathrm{d}x} = \text{const}. \tag{3.63}$$

La valeur de cette constante n'est pas nécessairement zéro !

Choisissons un maillage de N points x_i, $i = 1..N$, **pas forcément équidistants**, avec $h_i = x_{i+1} - x_i > 0$, $i = 1..n$, où $n = N - 1$ est le nombre d'intervalles. On utilise ensuite l'approximation numérique des éléments finis, Eqs.(3.50- 3.51), avec des fonctions de base linéaires, FIG. 3.10. On obtient, sur le modèle de l'Eq.(3.52) :

$$\sum_i \sum_j \eta_i \left(\int_{x_a}^{x_b} \frac{\mathrm{d}\Lambda_i}{\mathrm{d}x} \frac{\mathrm{d}\Lambda_j}{\mathrm{d}x} \, dx \right) \phi_j = \sum_i \eta_i \left(\int_{x_a}^{x_b} \frac{\rho}{\varepsilon_0} \Lambda_i \, dx \right) \ , \quad \forall \eta_i \ . \tag{3.64}$$

On obtient donc le système algébrique linéaire $\mathbf{A}\Phi = \mathbf{b}$, avec

$$A_{ij} = \int_{x_a}^{x_b} \frac{\mathrm{d}\Lambda_i}{\mathrm{d}x} \frac{\mathrm{d}\Lambda_j}{\mathrm{d}x} \, dx \tag{3.65}$$

$$b_i = \int_{x_a}^{x_b} \frac{\rho}{\varepsilon_0} \Lambda_i \, dx \ . \tag{3.66}$$

On a deux méthodes algorithmiques de construire la matrice et le membre de droite. La première méthode consiste à effectuer une **boucle sur les intervalles** ($k = 1..n$) et à **ajouter** à la matrice $\mathbf{A}$ et au membre de droite $\mathbf{b}$ la contribution de l'intervalle numéro k aux intégrales (3.65,3.66). Plus spécifiquement,

$$A_{ij} = \int_{x_a}^{x_b} \frac{\mathrm{d}\Lambda_i}{\mathrm{d}x} \frac{\mathrm{d}\Lambda_j}{\mathrm{d}x} \, dx = \sum_{k=1}^{n} \int_{x_k}^{x_{k+1}} \frac{\mathrm{d}\Lambda_i}{\mathrm{d}x} \frac{\mathrm{d}\Lambda_j}{\mathrm{d}x} \, dx \tag{3.67}$$

$$b_i = \int_{x_a}^{x_b} \frac{\rho}{\varepsilon_0} \Lambda_i \, dx = \sum_{k=1}^{n} \int_{x_k}^{x_{k+1}} \frac{\rho}{\varepsilon_0} \Lambda_i \, dx \ . \tag{3.68}$$

Dans les intégrales ci-dessus, seuls les termes $(i,j) = (k,k), (k,k+1), (k+1,k), (k+1,k+1)$ pour la matrice et les termes $i = k$ et $i = k+1$ pour le membre de droite sont non nuls. L'algorithme de cette première méthode de construction de $\mathbf{A}$ et $\mathbf{b}$ consiste en une **boucle sur les intervalles** $k = 1..n$

$$\mathbf{A} = \mathbf{A} + \begin{pmatrix} & & {\scriptstyle (k)} & {\scriptstyle (k+1)} & \\ & . & . & & \\ {\scriptstyle (k)} & . & 1/h_k & -1/h_k & \\ {\scriptstyle (k+1)} & & -1/h_k & 1/h_k & . \\ & & & . & . \end{pmatrix} \tag{3.69}$$

$$b_k = b_k + h_k \left(p\,\frac{\rho(x_k)}{2\varepsilon_0} + (1-p)\frac{\rho(x_{k+1/2})}{2\varepsilon_0} \right) \tag{3.70}$$

$$b_{k+1} = b_{k+1} + h_k \left(p\,\frac{\rho(x_{k+1})}{2\varepsilon_0} + (1-p)\frac{\rho(x_{k+1/2})}{2\varepsilon_0} \right) . \tag{3.71}$$

Les contributions au membre de droite, b_k et b_{k+1} ci-dessus, ont été écrites en utilisant la formule d'intégration (3.56).

Une deuxième méthode de construction de la matrice et du membre de droite consiste à calculer dans une **boucle sur les équations du système algébrique linéaire, donc sur les lignes de la matrice**, l'indice i dénotant le numéro de la ligne, (qui représente aussi le numéro de l'élément fini de la fonction test). Il est facile de calculer les éléments de matrice exactement, notant que

$$\mathrm{d}\Lambda_i/\mathrm{d}x = -1/h_i,\ \forall x \in]x_i, x_{i+1}[\ ;\ \ 1/h_{i-1},\ \forall x \in]x_{i-1}, x_i[\ ;\ \ 0\ \text{ ailleurs.} \tag{3.72}$$

On obtient

$$\begin{aligned} A_{i,i} &= \frac{1}{h_{i-1}} + \frac{1}{h_i}\ ,\ i = 2..n;\ A_{11} = \frac{1}{h_1}\ , A_{NN} = \frac{1}{h_n}\ ; \\ A_{i,i+1} &= -\frac{1}{h_i}\ ,\ i = 1..n; \\ A_{i,i-1} &= -\frac{1}{h_{i-1}}\ ,\ i = 2..N; \\ A_{i,j} &= 0\ , \forall j \notin \{i-1, i, i+1\}\ . \end{aligned} \tag{3.73}$$

$$b_i = h_{i-1}\left[p\,\frac{\rho(x_i)}{2\varepsilon_0} + (1-p)\frac{\rho(x_{i-1/2})}{2\varepsilon_0} \right] + h_i \left[p\,\frac{\rho(x_i)}{2\varepsilon_0} + (1-p)\frac{\rho(x_{i+1/2})}{2\varepsilon_0} \right]\ , i = 2..n. \tag{3.74}$$

Les lignes $i = 1$ et $i = N$ doivent être traités séparément : l'intégrale de la forme variationnelle va de x_a à x_b, et pour le premier et le dernier point de maillage, il n'y a que la moitié de la fonction de base correspondante qui contribue. On obtient

$$b_1 = h_1 \left[p\,\frac{\rho(x_1)}{2\varepsilon_0} + (1-p)\frac{\rho(x_{1+1/2})}{2\varepsilon_0} \right] \tag{3.75}$$

$$b_N = h_n \left[p\,\frac{\rho(x_N)}{2\varepsilon_0} + (1-p)\frac{\rho(x_{N-1/2})}{2\varepsilon_0} \right] . \tag{3.76}$$

L'avantage de la première méthode (raisonner par intervalles) par rapport à la deuxième (raisonner par ligne du système linéaire) est que toutes les contributions à $\mathbf{A}$ et $\mathbf{b}$ de l'intervalle k, Eqs.(3.69-3.71), ne font intervenir que h_k, et non un mélange de h_i et h_{i-1} de la deuxième méthode, Eqs.(3.73-3.74). De plus, la première méthode ne nécessite pas de traitement spécial des points du bord du domaine, alors que c'est le cas de la deuxième, Eqs.(3.75-3.76). [Mais, dans tous les cas, les conditions aux bords doivent être appliquées, voir ci-dessous !]

La matrice $\mathbf{A}$ est tridiagonale, et il n'est pas nécessaire de la stocker sous la forme d'une matrice carrée pleine. On ne stocke que la diagonale principale (d), la diagonale inférieure(a) et la diagonale supérieure (c) :

$$\mathbf{A} = \begin{pmatrix} d_1 & c_1 & & & & & & \\ a_1 & d_2 & c_2 & & & & & \\ & \cdot & \cdot & \cdot & & & & \\ & & \cdot & \cdot & \cdot & & & \\ & & & a_{k-1} & d_k & c_k & & \\ & & & & \cdot & \cdot & \cdot & \\ & & & & & a_{n-1} & d_n & c_n \\ & & & & & & a_n & d_N \end{pmatrix} . \tag{3.77}$$

La correspondance entre les éléments de matrice A_{ij} et les composantes de d, a et c est

$$A_{k,k} \to d_k \ , \quad A_{k,k+1} \to c_k \ , \quad A_{k+1,k} \to a_k \ , \quad A_{k+1,k+1} \to d_{k+1} \ . \tag{3.78}$$

Il faut imposer les *conditions aux bords*, Eq.(3.57), *explicitement* sur l'équation matricielle. La première équation doit être remplacée par $\phi_1 = V_a$ et la dernière équation par $\phi_N = V_b$. On le fait en posant :

$$d_1 = 1;\ c_1 = 0;\ b_1 = V_a; \quad \text{et} \quad d_N = 1;\ a_n = 0;\ b_N = V_b \ . \tag{3.79}$$

La résolution du système algébrique linéaire $\mathbf{A}\Phi = \mathbf{b}$ se fait par méthode directe (élimination de Gauss, ici en Matlab®) :

```
for k=2:N
   piv=a(k-1)/d(k-1);
   d(k)=d(k)-piv*c(k-1);
   b(k)=b(k)-piv*b(k-1);
end
phi=b./d;
for k=n:-1:1
   phi(k)=(b(k)-c(k)*phi(k+1))/d(k);
end
```

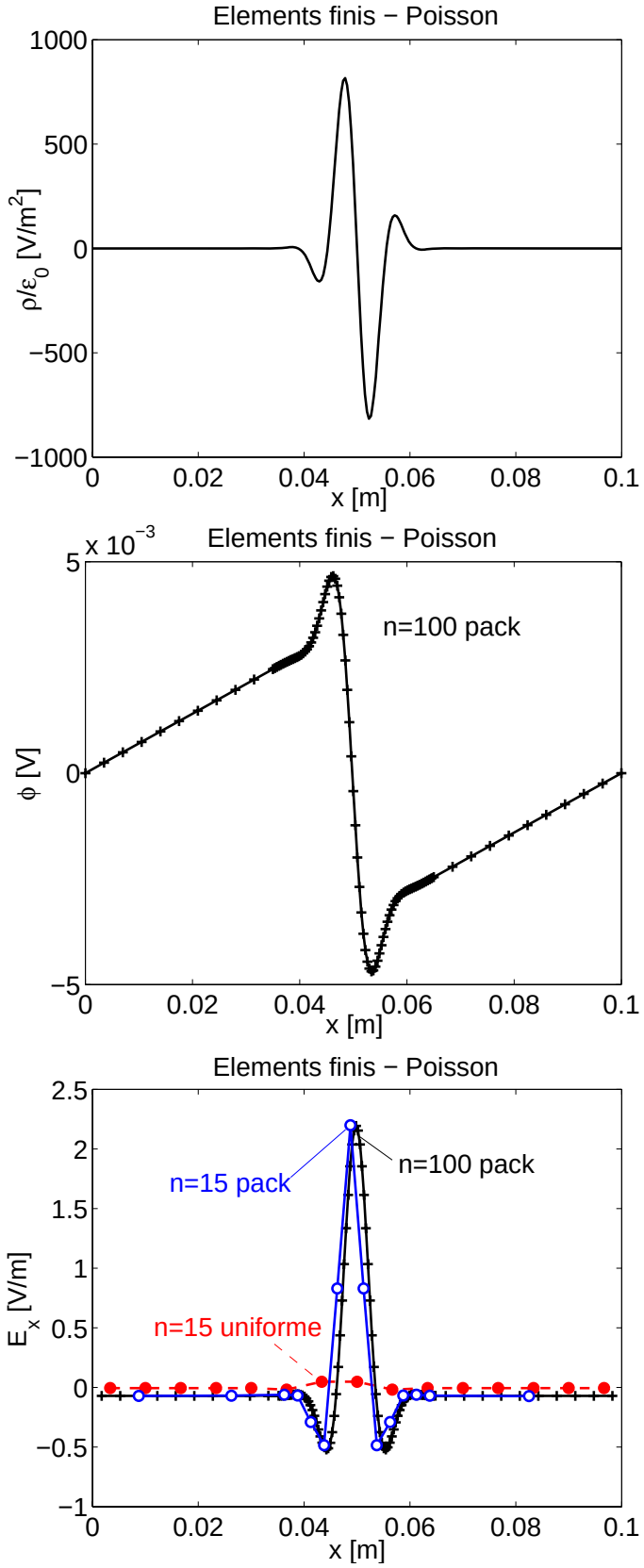


FIGURE 3.11 – *Résolution de l'équation de Poisson avec la méthode des éléments finis linéaires 1-D sur un maillage non-uniforme. Densité (haut), potentiel (milieu) et champ électrique (bas). Champ électrique obtenu avec un maillage uniforme $n = 15$ (traitillés rouge), avec un maillage non uniforme $n = 15$ (bleu avec 'o') et $n = 100$ (noir '+').*

L'avantage de pouvoir définir un maillage non équidistant peut s'avérer crucial lorsque la physique que l'on veut étudier présente des structures très localisées dans l'espace. Prenons le cas d'une distribution de charge

$$\rho(x) = \varepsilon_0 a_0 \sin(k_x x) \exp^{-\frac{(x-x_0)^2}{2\sigma^2}} \tag{3.80}$$

représentée à la FIG. 3.11 pour $a_0 = 1000\text{V/m}^2$, $k_x = 18\pi/x_b\text{m}^{-1}$, $x_0 = 0.05\text{m}$, $\sigma = 0.004\text{m}$. On résout Poisson entre $x_a = 0\text{m}$ et $x_b = 0.1\text{m}$, où les deux électrodes sont mises à la terre, $V_a = V_b = 0\text{V}$. On utilise l'intégration Eq.(3.56) pour le membre de droite, avec le paramètre $p = 1/3$. Le maillage est choisi uniforme par morceaux dans les intervalles $[0, 0.035]$, $[0.035, 0.065]$ et $[0.065, 1]$. On répartit 80% des points dans l'intervalle central et 10% dans chacun des autres intervalles. La solution, FIG. 3.11, montre clairement comment la haute densité de points du maillage dans les régions où la solution exhibe une structure localisée est appropriée : pour comparaison, on a représenté le champ électrique obtenu avec un maillage uniforme partout.

3.4 Magnétostatique - Biot-Savart

Les équations pour le potentiel électrostatique ϕ et pour le potentiel vecteur $\vec{A}$, Eqs. (3.25) et (3.27), sont toutes deux de la forme d'une équation de Poisson. On peut donc utiliser les méthodes présentées aux sections précédentes, différences finies et éléments finis, pour résoudre aussi les problèmes de magnétostatique.

On mentionnera quand même deux autres approches. La première est utilisée dans le vide (cas $\vec{j} = 0$), où $\nabla \times \vec{B} = 0$ implique l'existence d'un potentiel scalaire [4] $\Psi(\vec{x})$ tel que $\vec{B} = \nabla\Psi$. Avec $\nabla \cdot \vec{B} = 0$, on a

$$\nabla^2 \Psi = 0 \tag{3.81}$$

et les méthodes pour résoudre l'équation de Laplace peuvent être utilisées.

La deuxième approche utilise la formule de Biot-Savart. Si on a une distribution de courant $\vec{j}(\vec{x}')$ donnée, on peut trouver la solution explicite de l'Eq.(3.27). En se rappelant que la solution de l'équation de Poisson (3.25) pour un élément de charge $\rho(\vec{x}')\, d^3x'$ est $\phi(\vec{x}) = \rho(\vec{x}')\, d^3x'/4\pi\varepsilon_0 r$, avec $r = |\vec{x} - \vec{x}'|$, la solution de (3.27) s'obtient en substituant formellement ρ/ε_0 par $\mu_0\vec{j}$ et en intégrant sur tout l'espace :

$$\vec{A}(\vec{x}) = \frac{\mu_0}{4\pi} \int\int\int \frac{\vec{j}(\vec{x}')d^3x'}{|\vec{x} - \vec{x}'|} \,. \tag{3.82}$$

En effectuant $\vec{B} = \nabla \times \vec{A}$ (attention, l'opérateur ∇ opère sur $\vec{x}$ mais pas sur $\vec{x}'$), et en

4. Il ne s'agit évidemment PAS du potentiel électrostatique.

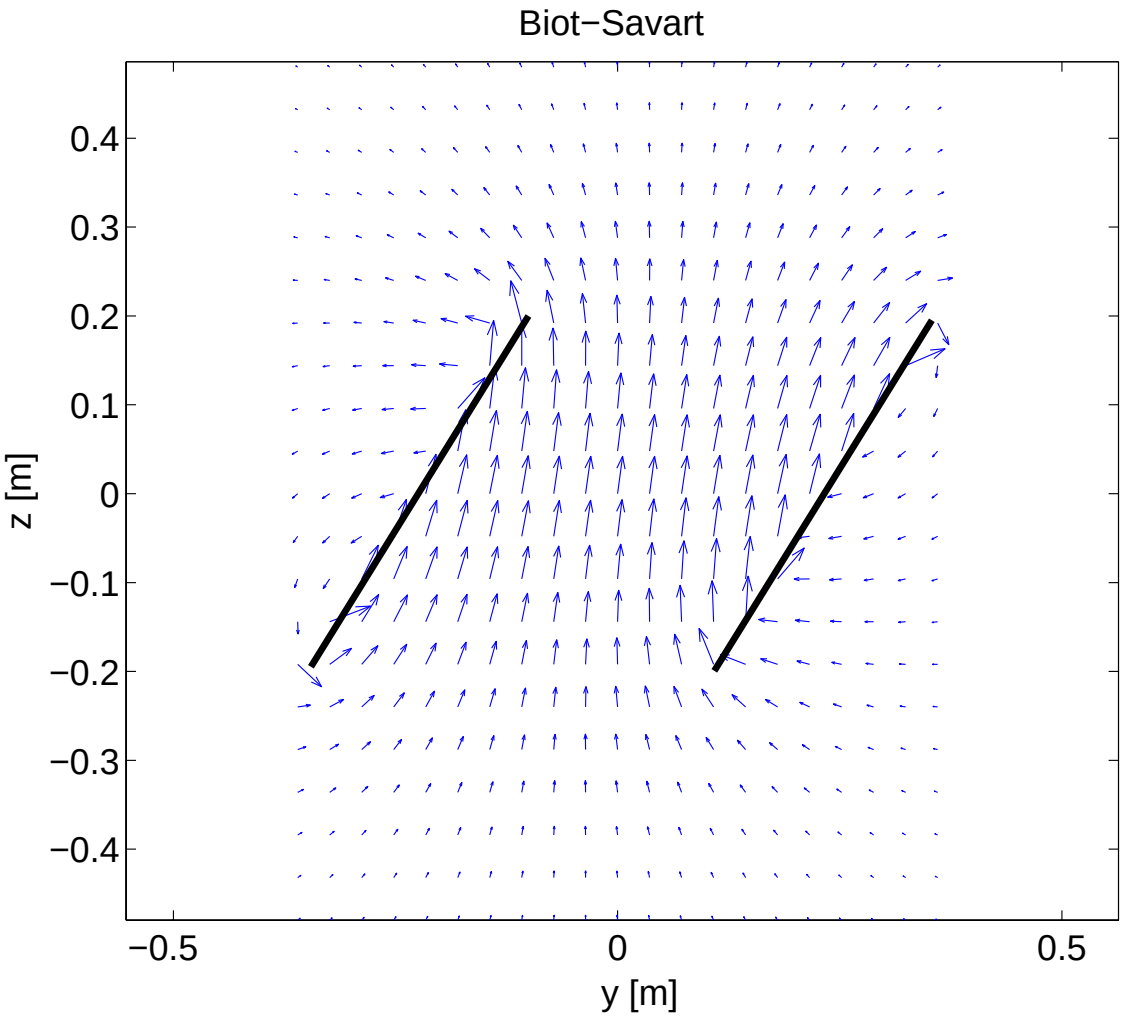


FIGURE 3.12 – *Champ magnétique créé par une bobine inclinée, dont la position est symbolisée par les deux traits noirs obliques. Les flèches sont proportionnelles au champ* $\vec{B}$.

posant $\vec{e}_r = (\vec{x} - \vec{x}')/r$, on obtient la **formule de Biot-Savart** :

$$\boxed{\vec{B}(\vec{x}) = \frac{\mu_0}{4\pi} \int\int\int \frac{\vec{j}(\vec{x}') \times \vec{e}_r}{r^2} d^3x'} \,. \tag{3.83}$$

Dans le cas d'une boucle de courant (fil mince), courbe Γ, courant I, $\vec{j}d^3x' = I\vec{dl}' = I\vec{e}_t dl'$ et on a, de (3.82)(3.83) :

$$\boxed{\vec{A}(\vec{x}) = \frac{\mu_0 I}{4\pi} \oint_\Gamma \frac{\vec{dl}'}{r}} \,, \quad \boxed{\vec{B}(\vec{x}) = \frac{\mu_0 I}{4\pi} \oint_\Gamma \frac{\vec{e}_t \times \vec{e}_r}{r^2} dl'} \,, \tag{3.84}$$

où $\vec{e}_t$ est le vecteur unité tangent au fil en tout point.

L'utilisation de la formule de Biot-Savart implique donc d'effectuer des **intégrales**. **Quelques méthodes d'intégration numérique sont présentées à l'Annexe B.**

Lorqu'il y a plusieurs circuits ("boucles de courant"), on applique le principe de superposition.

Un exemple est illustré à la FIG. 3.12. On a calculé le champ $\vec{B}$ créé par une bobine constituée de 40 boucles de fil de forme circulaire de rayon $R = 0.24$m parcourues par un courant $I = 1$A, et empilées les unes sur les autres avec un décalage. A la FIG. 3.12, on a représenté le champ magnétique dans un plan (y, z) par des flèches proportionnelles au champ $\vec{B}$. La position de la bobine est symbolisée par les deux traits obliques. Chaque boucle de la bobine est discrétisée avec $N = 128$ points pour l'intégration de Biot-Savart. Pour cette figure, la règle des trapèzes a été utilisée. Des tests numériques standards de convergence de la solution avec le nombre de points de discrétisation montrent que la méthode a une erreur en h^2 (donc en $1/N^2$), comme prévu par la théorie (Annexe B).

On peut encore effectuer des vérifications sur la qualité de la physique du résultat. On vérifiera la précision avec laquelle la loi d'Ampère, $\oint_\Gamma \vec{B} \cdot \vec{dl} = \mu_0 I$, et la loi de Gauss pour le champ magnétique, $\oiint_\Sigma \vec{B} \cdot \vec{d\sigma} = 0$, sont vérifiées pour tout parcours fermé Γ et pour toute surface fermée Σ, respectivement.

Chapitre 4

Intégration Spatio-Temporelle

4.1 Advection-diffusion

4.1.1 Advection

L'advection désigne le transport d'une quantité physique, décrite par un champ scalaire $f(\vec{x}, t)$, dans un écoulement décrit par un champ de vitesses $\vec{v}(\vec{x}, t)$. Il peut s'agir, par exemple, de la concentration d'un polluant, ou de l'humidité dans l'air, etc. Le flux de cette quantité physique est $\vec{j} = f\vec{v}$. De l'équation de continuité

$$\frac{\partial f}{\partial t} + \nabla \cdot \vec{j} = 0 \tag{4.1}$$

et en supposant, de plus, un écoulement incompressible, $\nabla \cdot \vec{v} = 0$, on obtient l'équation décrivant l'évolution spatio-temporelle de f, appelée *équation d'advection :*

$$\frac{\partial f(\vec{x}, t)}{\partial t} + \vec{v} \cdot \nabla f(\vec{x}, t) = 0 \tag{4.2}$$

On se restreindra dans ce chapitre au cas à une dimension d'espace. La solution de cette équation, pour v constant, est triviale :

$$f(x, t) = f_0(x - vt) \tag{4.3}$$

où f_0 est la condition initiale, $f_0(x) = f(x, 0)$. La solution est donc une simple translation dans l'espace, à la vitesse v, de la condition initiale. Etant donné une solution exacte si simple, on peut se demander pourquoi développer des méthodes numériques pour résoudre le prboblème de l'advection. En fait, la solution n'est pas triviale si la vitesse v n'est pas uniforme ou non constante. De plus, le phénomène d'advection est souvent combiné à celui de la *diffusion*, sujet traité dans la section suivante.

La résolution numérique d'une équation apparamment si simple n'est cependant pas si triviale.

Définissons les quantités suivantes :

- le nombre de particules

$$N(t) = \int f(x,t)dx \tag{4.4}$$

- la position moyenne

$$< x > (t) = \frac{1}{N} \int x f(x,t)dx \tag{4.5}$$

- la variance σ^2 et l'écart quadratique moyen, ou écart-type, σ

$$< x^2 > (t) = \frac{1}{N} \int x^2 f(x,t)dx \ , \quad \sigma^2(t) =< x^2 > (t) - (< x > (t))^2 \tag{4.6}$$

Les intégrales dans les expressions ci-dessus sont à effectuer sur tout le domaine spatial de définition de f.

Advection en différences finies

La méthode est de discrétiser l'espace et le temps sur un maillage équidistant (x_i, t_j) et d'utiliser les approximations en différences finies des opérateurs $\partial/\partial t$ et $\partial/\partial x$.

$$\frac{\partial f}{\partial t}(x_i,t_j) = \frac{f(x_i,t_{j+1}) - f(x_i,t_j)}{\Delta t} + \mathcal{O}(\Delta t) \tag{4.7}$$

$$\frac{\partial f}{\partial x}(x_i,t_j) = \frac{f(x_i,t_j) - f(x_{i-1},t_j)}{\Delta x} + \mathcal{O}(\Delta x) \tag{4.8}$$

Le lecteur attentif aura remarqué que l'on fait des différences finies "forward", Eq.(A.22), pour la première dérivée temporelle, alors que l'on fait des différences finies "backward" pour la première dérivée spatiale. En fait, le schéma ci-dessus va être stable si $v \geq 0$, mais il est *instable si* $v < 0$ *!* Pour $v < 0$, on utilise

$$\frac{\partial f}{\partial x}(x_i,t_j) = \frac{f(x_{i+i},t_j) - f(x_i,t_j)}{\Delta x} + \mathcal{O}(\Delta x) \tag{4.9}$$

On remarque que dans les 2 cas, cela revient à prendre la première dérivée spatiale "dans la direction d'où vient l'écoulement v", d'ou le nom **upwind scheme** (up-the-wind) pour ce schéma.

Pour simplifier les notations, on notera dans la suite $f_{i,j} = f(x_i, t_j)$.

On définit le **paramètre CFL** (Courant-Friedrichs-Lewy) [1]

$$\boxed{\beta = v\frac{\Delta t}{\Delta x}}\,. \tag{4.10}$$

On aboutit ainsi au schéma suivant :

$$\begin{aligned} f_{i,j+1} &= f_{i,j} - \beta\,(f_{i,j} - f_{i-1,j}) \quad \text{si} \;\; \beta \geq 0\,, \\ f_{i,j+1} &= f_{i,j} - \beta\,(f_{i+1,j} - f_{i,j}) \quad \text{si} \;\; \beta < 0\,. \end{aligned} \tag{4.11}$$

Ce schéma est dit **explicite** : on obtient la solution au temps $j+1$ en fonction des valeurs de f aux points de maillage spatial au temps précédent j. Il est dit "à 2 niveaux", car il fait intervenir deux temps consécutifs. En résumé, il s'agit du *schéma différences finies explicite upwind à 2 niveaux pour l'équation d'advection.*

On montre un exemple aux FIGS.4.1-4.2, avec une distribution initiale de densité gaussienne (ou "normale")

$$f(x,0) = \frac{N}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right) \tag{4.12}$$

centrée en $x_0 = 0$, écart-type $\sigma = 0.2$m. Les paramètres sont les suivants : $v = 1$m/s, 64 points de maillage pour $x \in [-2, 2]$, $\Delta t = 0.01$s, donnant un paramètre CFL $\beta = 0.16$. On voit clairement que la solution initiale se propage à la bonne vitesse (en moyenne), mais que le profil de densité s'étale, avec une **variance σ^2 proportionnelle au temps t.** Cela n'est pas physique, la solution devrait conserver sa forme. En fait, on assiste à un phénomène de *diffusion numérique*, qui ressemble à de la diffusion physique, voir Section suivante, mais qui est dû ici à *l'amortissement du schéma numérique.*

Il y a pire, si on prend des intervalles spatiaux Δx plus petits (dans l'intention d'obtenir une solution approximée de meilleure qualité) et/ou des intervalles temporels Δt plus grands (dans l'intention de faire de plus longues simulations), une **instabilité numérique** qui peut se développer. Un exemple est montré à la FIG. 4.3, pour les mêmes paramètres que la FIG. 4.1 sauf que l'on a pris 128 intervalles en x et un $\Delta t = 0.0375$s donnant un paramètre CFL $\beta = 1.2$. Après un temps fini, une instabilité de courte longueur d'onde se développe, qui est évidemment non physique, de par le fait, notamment, qu'elle fait apparaître des valeurs négatives de la densité (!).

On peut montrer qu'en effet le schéma explicite utilisé ici est **instable si le nombre CFL β est supérieur à 1.** On fera la démonstration de ce critère de stabilité à la section 4.1.3.

Puisque le schéma upwind est stable, mais amorti, pour un CFL $\beta < 1$, et que le schéma "downwind" est instable pour tout CFL β, on pourrait choisir le meilleur des deux mondes

1. Courant, R. ; Friedrichs, K. ; and Lewy, H. "On the Partial Difference Equations of Mathematical Physics." IBM J. 11, 215-234, 1967.

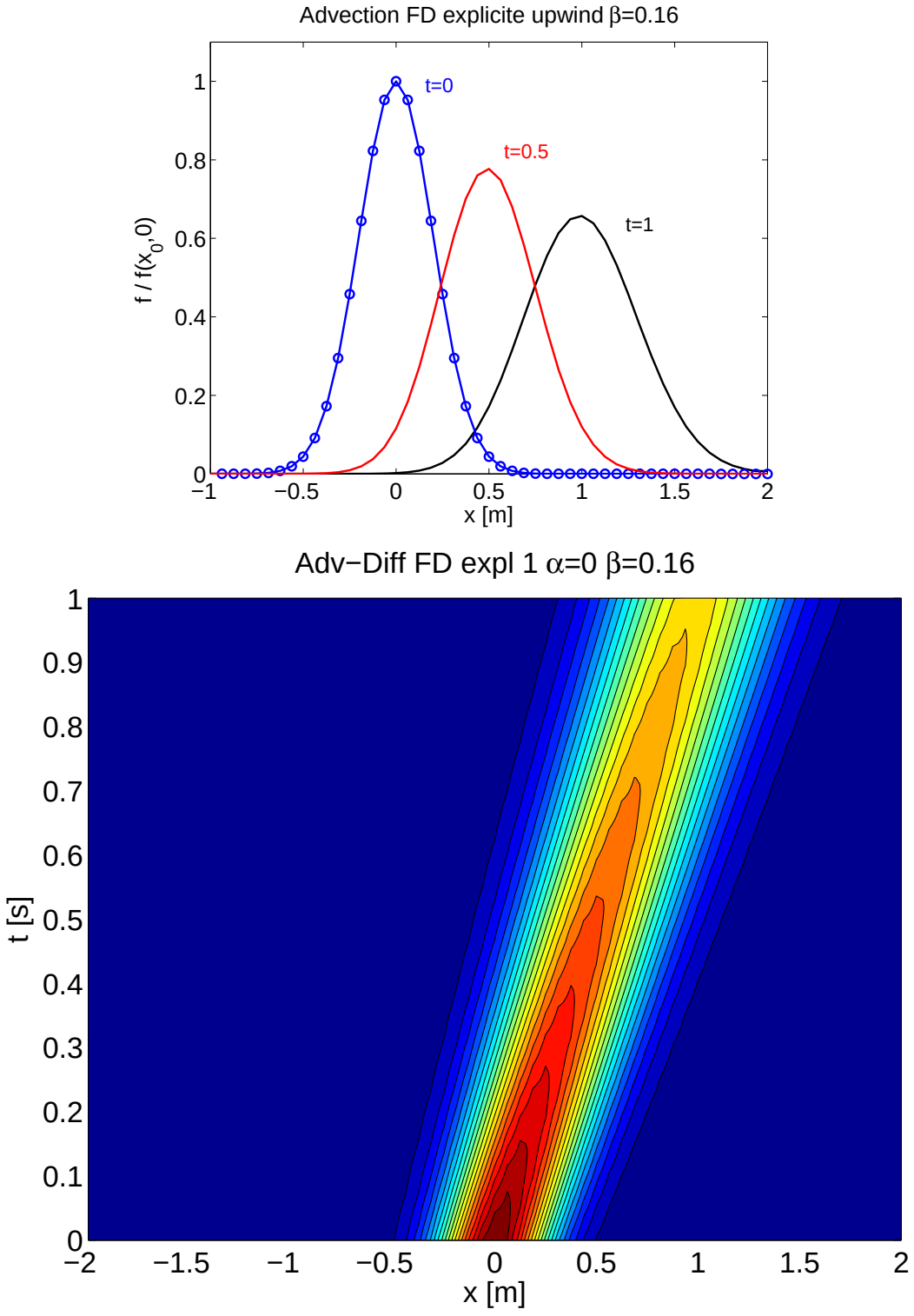


FIGURE 4.1 – *Advection d'une quantité scalaire $f(x,t)$. Différences finies, schéma explicite à 2 niveaux, upwind, Eq.(4.11). Paramètres : $u = 1\mathrm{m/s}$, CFL $\beta = 0.16$, $N_x = 64$. En haut : instantanés de la densité. En bas : contours de f en fonction de x et t. L'amortissement (ici purement numérique !) a pour effet d'étaler la distribution de densité.*

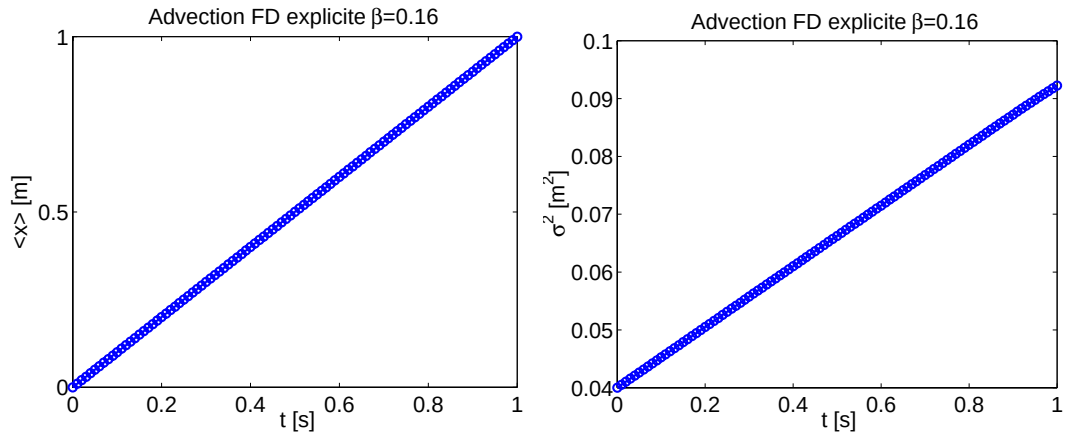


FIGURE 4.2 – *Evolution temporelle de la position moyenne* $< x >$ *et de la variance* σ^2*, pour la simulations de la FIG. 4.1.*

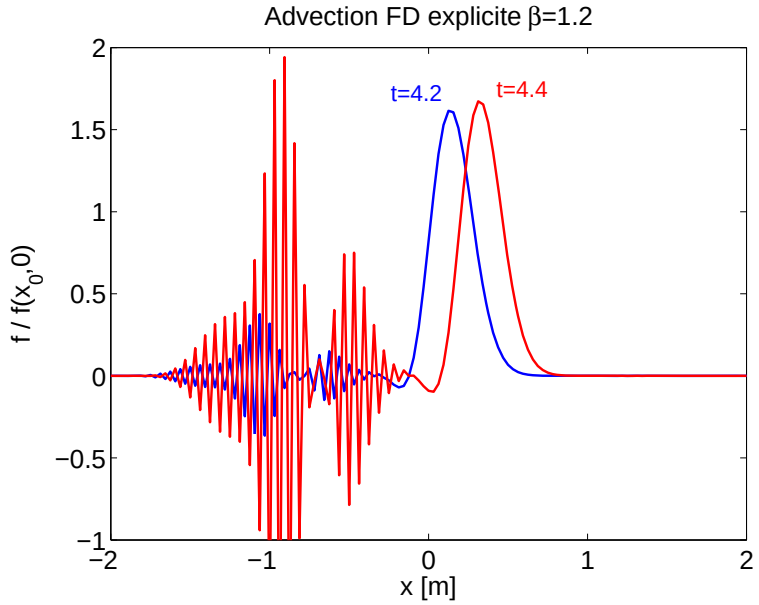


FIGURE 4.3 – *Instabilité du schéma différences finies explicite upwind à 2 niveaux pour l'advection, mêmes paramètres qu'aux FIGS.4.1-4.2, mais avec un paramètre CFL* $\beta = 1.2$*.*

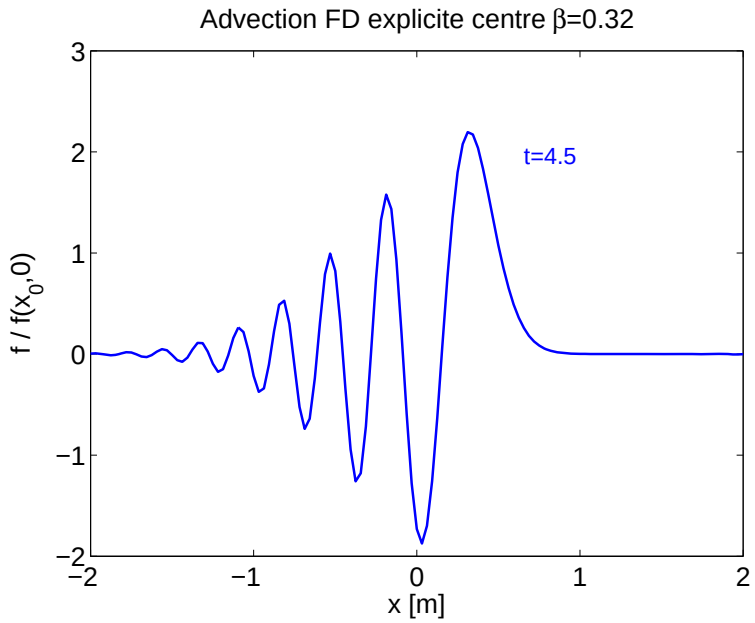


FIGURE 4.4 – *Apparition d'oscillations non physiques dans le schéma différences finies explicite centré, Eq.(4.13), à 2 niveaux pour l'advection, mêmes paramètres qu'aux FIGS.4.1-4.2, sauf le paramètre CFL* $\beta = 0.32$.

en considérant la moyenne de l'upwind et du downwind, autrement dit les différences finies centrées pour $\partial f/\partial x$, soit la moyenne de l'Eq.(4.8) et de l'Eq.(4.9) :

$$\frac{\partial f}{\partial x}(x_i, t_j) = \frac{f(x_{i+i}, t_j) - f(x_{i-1}, t_j)}{2\Delta x} + \mathcal{O}(\Delta x^2) \; , \tag{4.13}$$

ceci quel que soit le signe de v. Malheureusement, si l'amortissement numérique disparaît, on a l'apparition d'oscillations ("overshoots") dans la solution : voir FIG. 4.4. Ces oscillations sont évidemment non physiques : elles exhibent des endroits de densité négative ! De plus, la mesure de $\sigma^2(t)$ fait apparaître une décroissance monotone, comme si le schéma numérique introduisait de "l'anti-diffusion".

La conclusion de cette section est que les schémas numériques pour résoudre l'advection peuvent introduire de la **diffusion (ou anti-diffusion) numérique**, quant ils ne sont pas carrément instables. Le paramètre de stabilité fondamental est le paramètre CFL ; le **critère de stabilité CFL** est

$$\boxed{\beta = \frac{v\Delta t}{\Delta x} \leq 1} \; . \tag{4.14}$$

4.1.2 Diffusion

Du microscopique au macroscopique

Le processus de diffusion, au niveau microscopique (moléculaire) est dû à l'agitation thermique des particules. C'est un biologiste, Brown, qui a le premier documenté ses observations au microscope du mouvement de grains de pollen. Ce mouvement apparaît désordonné, aléatoire, et ne s'arrête jamais. On lui donne le nom de *mouvement Brownien.*

L'interprétation est que ce mouvmement est dû aux collisions avec les particules. Ces collisions ont lieu de façon **aléatoire**. On décrit donc le phénomène par une approche probabiliste. Le modèle de mouvment Brownien décrit une *marche aléatoire*, succession de déplacements dûs aux collisions successives. On fait les hypothèses suivantes.

- La direction des déplacements suit une loi de probabilité uniforme (isotropie).
- La succession des déplacements est complètement décorrélée : il n'y a pas de dépendance entre une collision et la suivante.
- La norme des déplacements est une variable aléatoire de moyenne non nulle.
- Chaque déplacement obéit à la même loi de probabilité.

On verra à la section suivante comment la simulation numérique peut s'inspirer de cette description probabiliste. Ici, on passe à une description macroscopique, en effectuant des moyennes statistiques sur un grand nombre de particules soumises à cette marche aléatoire. On décrit donc, (voir physique des fluides), la densité des particules par un champ scalaire $n(\vec{x}, t)$. Le phénomène de diffusion peut s'observer pour d'autres quantités physiques que la densité, et on notera ce champ scalaire de façon générique par $f(\vec{x}, t)$.

Le flux de cette quantité physique (par exemple nombre de particules par unité de temps et par unité de surface) est

$$\vec{j} = \frac{1}{V} \sum_{i=1}^{N} \vec{v}_i = f < \vec{v} > , \tag{4.15}$$

où $< \vec{v} >$ est la vitesse moyenne des particules dans un élément de volume V, et N est le nombre de particules dans le volume V. Empiriquement, on mesure que ce flux est proportionnel au gradient de densité, dans la direction opposée à celui-ci, avec une constante de proportionalité D appelée **coefficient de diffusion**.

$$\vec{j} = -D \nabla f \ . \tag{4.16}$$

C'est la loi de Fick. On invoque ensuite le principe de conservation du nombre de particules, exprimé par l'équation de continuité, Eq.(4.1), pour obtenir **l'équation de diffusion**

$$\frac{\partial f}{\partial t} - \nabla \cdot (D \nabla f) = 0 \ . \tag{4.17}$$

On peut combiner l'advection avec la diffusion. [Image : un polluant dans l'atmosphère ; l'advection est le transport de ce polluant par le vent ; la diffusion est dûe aux collisions

au niveau microscopique et a lieu même en l'absence de vent. Dans la réalité, les deux phénomènes ont lieu simultanément.] Avec $\vec{j} = f\vec{v} - D\nabla f$ et l'équation de continuité (4.1), on obtient **l'équation d'advection-diffusion**, qui dans le cas d'un écoulement incompressible ($\nabla \cdot \vec{v} = 0$) s'écrit :

$$\frac{\partial f}{\partial t} + \vec{v} \cdot \nabla f - \nabla \cdot (D \nabla f) = 0 \, . \tag{4.18}$$

Dans cette section, on ne considèrera que les cas à une dimension d'espace $f(x,t)$, et où la vitesse d'advection v et le coefficient de diffusion D sont uniformes et constants. On obtient :

$$\boxed{\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} - D\frac{\partial^2 f}{\partial x^2} = 0} \, . \tag{4.19}$$

On peut trouver la solution analytique à cette équation, par la méthode des fonctions de Green, la transformée de Laplace temporelle et/ou la transformée de Fourier spatiale. On obtient, voir Annexe C :

$$f(x,t) = \frac{N}{2\sqrt{\pi D\, t}} \exp\left(-\frac{(x - x_0 - vt)^2}{4Dt}\right) \tag{4.20}$$

pour une condition initiale $f(x,0) = N\delta(x - x_0)$ (toutes les particules sont initialement en $x = x_0$). [On peut, en exercice, montrer en substituant dans (4.19) qu'elle satisfait bien l'équation d'advection diffusion.] La densité est donc une gaussienne centrée en $x_0 + vt$ (mouvment de translation uniforme, effet de l'advection) et dont la largeur σ augmente comme la *racine carrée du temps* (effet de la diffusion) ; plus précisément :

$$\boxed{< x > (t) = x_0 + vt} \, , \quad \boxed{\sigma(t) = \sqrt{2Dt}} \, . \tag{4.21}$$

Remarque : la diffusion est ici décrite par un modèle déterministe et continu. La solution est unique pour une condition initiale donnée. Elle est représentée par un champ scalaire continu.

Advection-Diffusion en différences finies

Le schéma numérique consiste à approximer les opérateurs de l'Eq. (4.19) par leurs expressions en différences finies sur un maillage de l'espace-temps. On procède pour l'advection comme exposé à la section 4.1.1, en choisissant les différences finies "upwind", Eq. (4.11), ou centrées, Eq.(4.13). Le terme de diffusion implique la deuxième dérivée par rapport à x :

$$\frac{\partial^2 f}{\partial x^2}|_{i,j} = \frac{f_{i-1,j} - 2f_{i,j} + f_{i+1,j}}{\Delta x^2} + \mathcal{O}(\Delta x^2) \, . \tag{4.22}$$

On obtient, dans le cas de l'advection centrée,

$$f_{i,j+1} = f_{i,j} - \beta \left(f_{i+1,j} - f_{i-1,j}\right)/2 + \alpha \left(f_{i-1,j} - 2f_{i,j} + f_{i+1,j}\right) \, , \tag{4.23}$$

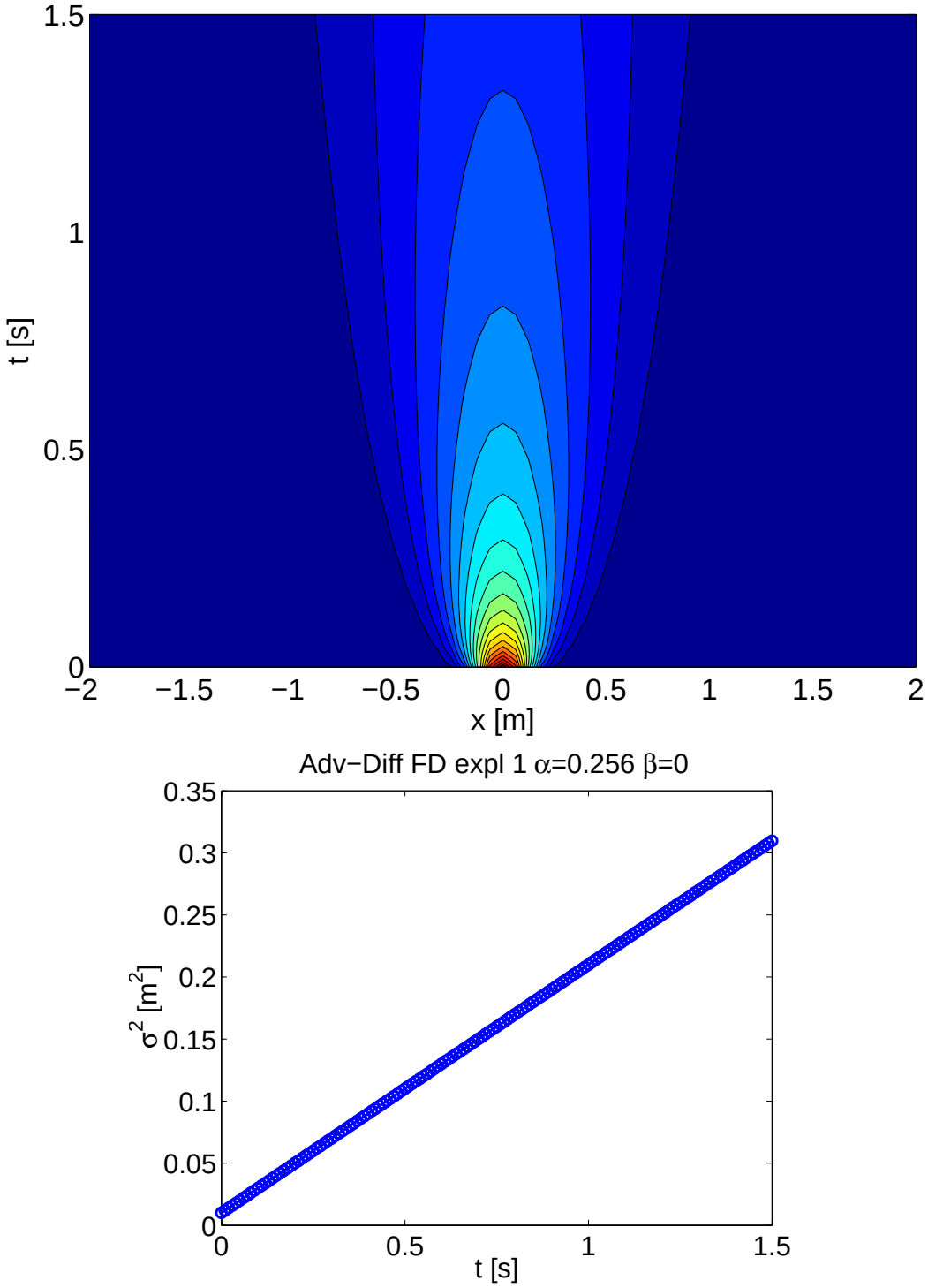


FIGURE 4.5 – *Diffusion d'une quantité scalaire $f(x,t)$. Différences finies, schéma explicite à 2 niveaux, Eq.(4.23). Paramètres : $u = 0$, $D = 0.1\mathrm{m}^2/\mathrm{s}$, $N_x = 64$, $\Delta t = 0.01\mathrm{s}$, $\alpha = 0.256$, distribution initiale Gaussienne avec $\sigma(0) = 0.1$. En haut : contours de f en fonction de x et t. En bas : écart quadratique σ^2 en fonction du temps.*

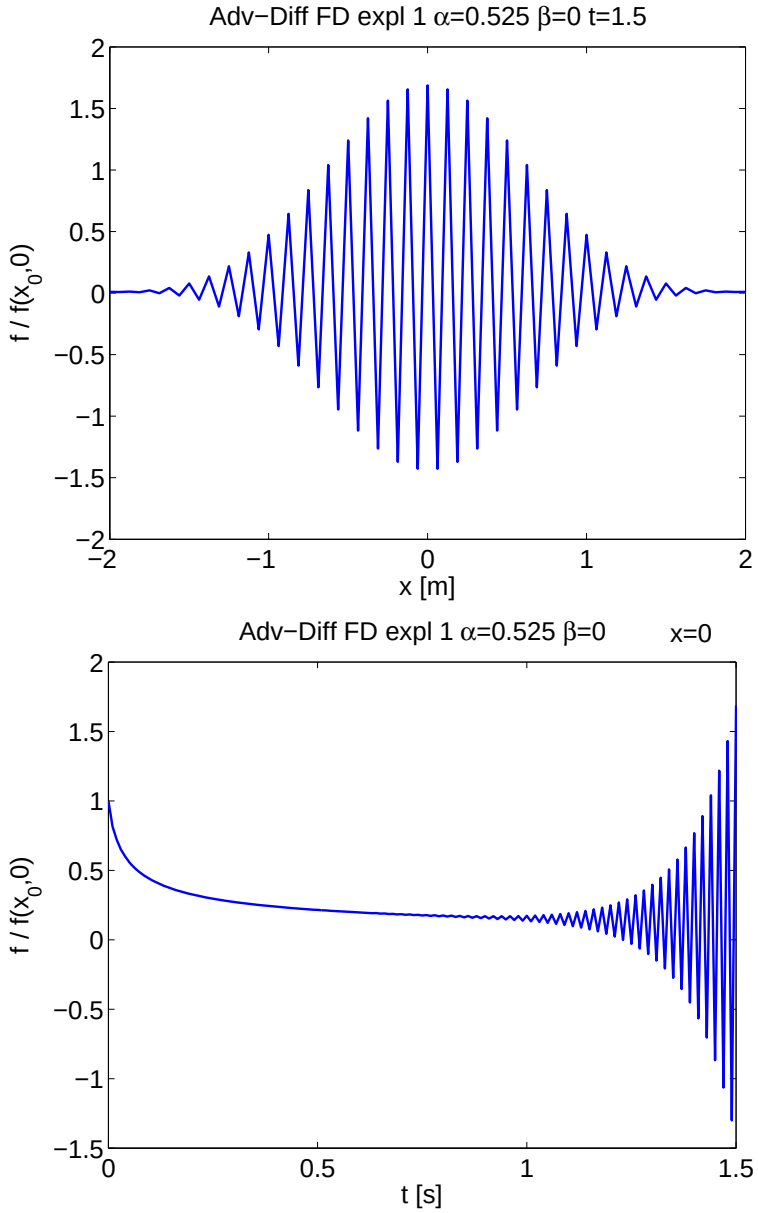


FIGURE 4.6 – *Instabilité du schéma différences finies explicite à 2 niveaux, Eq.(4.23). Paramètres :* $u = 0$, $D = 0.205\mathrm{m}^2/\mathrm{s}$, $N_x = 64$, $\Delta t = 0.01\mathrm{s}$, $\alpha = 0.5248$, *distribution initiale gaussienne avec* $\sigma(0) = 0.1$. *Une oscillation de courte longueur d'onde (2 points de maillage spatial par longueur d'onde) et de haute fréquence (2 points de maillage temporel par période) apparaît avec une amplitude qui croît exponentiellement dans le temps.*

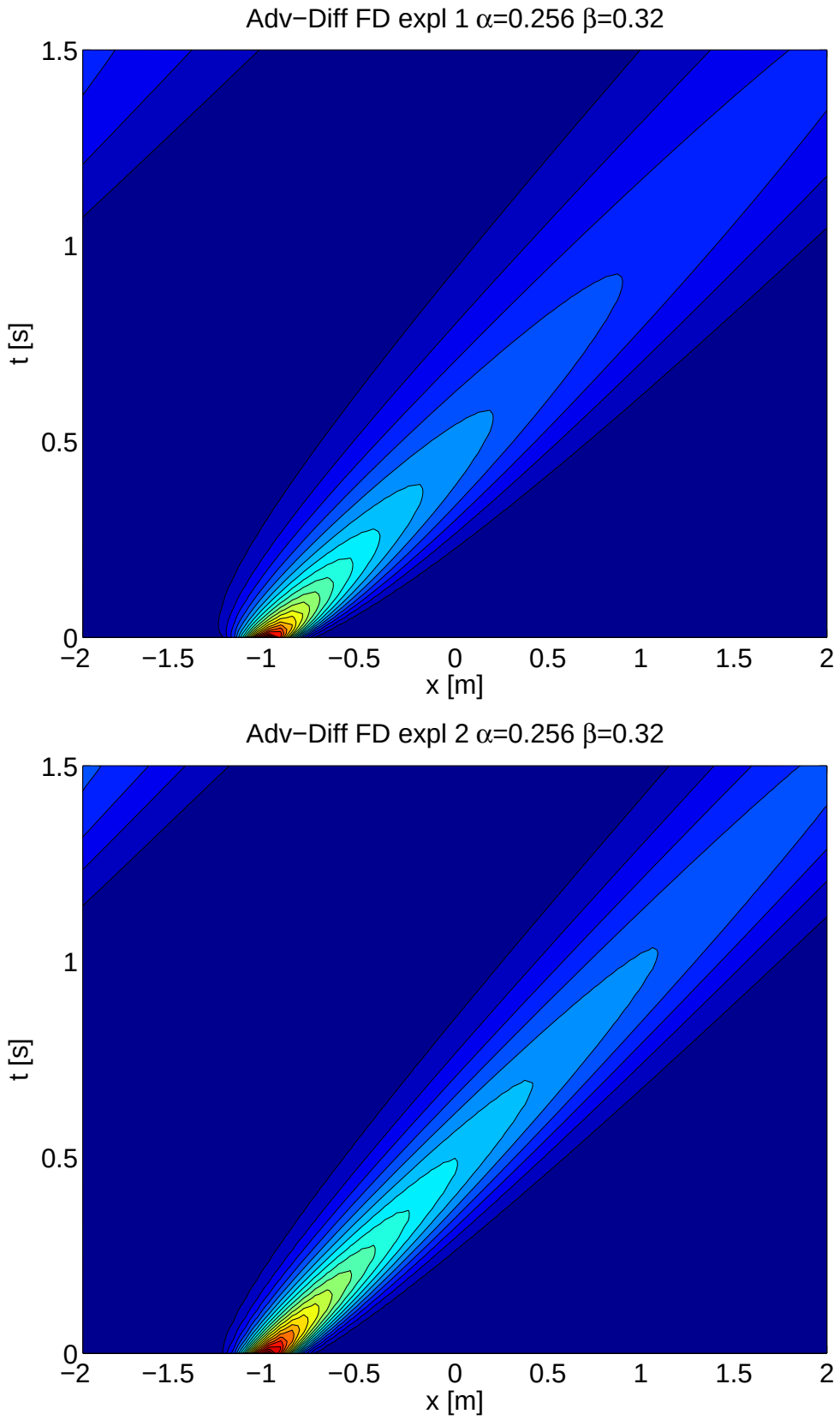


FIGURE 4.7 – *Diffusion et advection d'une quantité scalaire $f(x,t)$. Différences finies, schéma explicite à 2 niveaux. Paramètres : $u = 2\mathrm{m/s}$, $D = 0.1\mathrm{m}^2/\mathrm{s}$, $N_x = 64$, $\Delta t = 0.01\mathrm{s}$, $\alpha = 0.256$, $\beta = 0.32$, distribution initiale Gaussienne avec $\sigma(0) = 0.1$. Contours de f en fonction de x et t, schéma "upwind" (en haut) et schéma centré (en bas).*

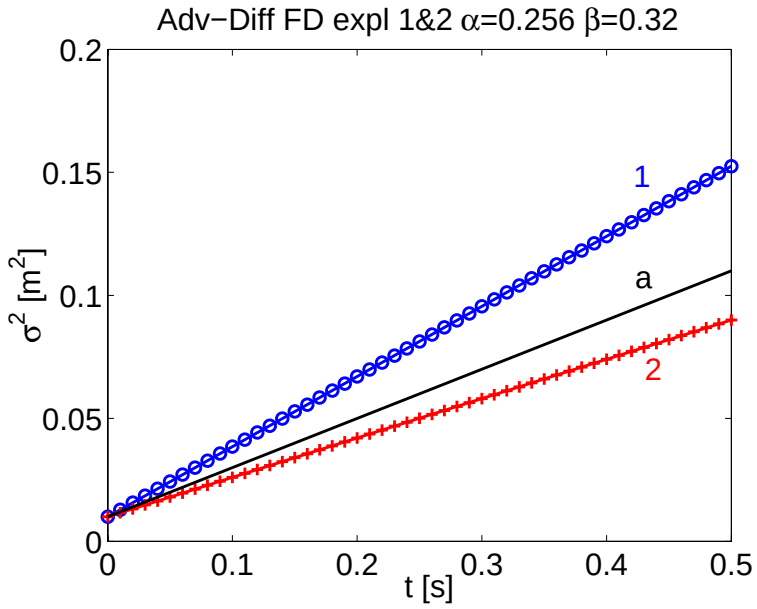


FIGURE 4.8 – *Diffusion et advection d'une quantité scalaire $f(x,t)$. Variance σ^2 en fonction du temps, pour les simulations de la FIG. 4.7, "1" pour "upwind", "2" pour centré, "a" pour la solution analytique,.*

et, dans le cas de l'upwind,

$$\begin{aligned} f_{i,j+1} &= f_{i,j} - \beta\,(f_{i,j} - f_{i-1,j}) + \alpha\,(f_{i-1,j} - 2f_{i,j} + f_{i+1,j}) \ , \quad \text{si} \ \ \beta \geq 0 \ , \\ f_{i,j+1} &= f_{i,j} - \beta\,(f_{i+1,j} - f_{i,j}) + \alpha\,(f_{i-1,j} - 2f_{i,j} + f_{i+1,j}) \ , \quad \text{si} \ \ \beta < 0 \ , \end{aligned} \tag{4.24}$$

où on a défini le paramètre α :

$$\alpha = \frac{D\Delta t}{\Delta x^2} \ . \tag{4.25}$$

Un exemple est montré à la FIG. 4.5, pour un cas sans advection ($\beta = 0$), et pour $\alpha = 0.256$. La densité initiale est une gaussienne centrée en $x = 0$, de largeur $\sigma = 0.1$m. Les paramètres sont : $N_x = 64$ points de maillage spatial pour $x \in [-2, +2]$, $\Delta t = 0.01$s, $D = 0.1\text{m}^2/\text{s}$, $v = 0$. L'étalement du profil de densité est clairement visible. Une mesure de la moyenne du x^2 donne $\sigma^2 = 0.01 + 0.2\,t$, en accord avec la théorie : l'étalement est proportionnel à la racine carrée du temps, avec une variance σ^2 égale à $\sigma_0^2 + 2Dt$.

D'autres tests sont également concluants : la densité est partout strictement positive, et le nombre de particules, $N = \int f(x)dx$, est conservé à la précision machine près.

Le schéma n'est pas toujours stable. Si on augmente le coefficient de diffusion à $D = 0.205\text{m}^2/\text{s}$, gardant tous les autres paramètres, il se développe une oscillation de courte longueur d'onde (2 points de maillage spatial par longueur d'onde) dont l'amplitude croît exponentiellement dans le temps : un exemple est illustré à la FIG. 4.6. Pour cette simulation, le paramètre $\alpha = 0.5248$. On montrera à la section 4.1.3 que le schéma numérique est **instable pour** $\boxed{\alpha > 0.5}$.

Lorsque l'avection et la diffusion sont simultanément présentes, $\alpha \neq 0$ et $\beta \neq 0$, il faut être prudent avec les résultats numériques. En effet, le schéma "upwind" de l'advection introduit de la diffusion numérique, comme montré à la section précédente, et la diffusion mesurée sur les simulations est la somme de la diffusion physique et de cette diffusion numérique. Le schéma centré pour le terme d'advection conduit, quant à lui, à de "l'anti-diffusion". On montre à la FIG. 4.7 un exemple avec $v = 2\text{m/s}$, $D = 0.1\text{m}^2/\text{s}$, maillage $n_x = 64$, $x \in [-2, 2]$, $\Delta t = 0.01$, donnant les paramètres $\alpha = 0.256$ et $\beta = 0.32$. On compare les résultats avec l'advection "upwind" (notée 1) et l'advection centrée (notée 2). La variance obtenue s'écarte de la valeur analytique (notée a). On obtient un coefficient de diffusion résultant $D_{\text{sim}} = 0.1469\text{m}^2/\text{s}$ pour le schéma upwind et $D_{\text{sim}} = 0.08\text{m}^2/\text{s}$ pour le schéma centré, au lieu de la valeur exacte $D = 0.1\text{m}^2/\text{s}$.

4.1.3 Stabilité du schéma numérique : analyse de Von Neumann

L'analyse de la stabilité numérique se fait en examinant comment l'amplitude d'une perturbation sinusoïdale évolue dans le temps par le schéma numérique. On pose donc la solution au temps t comme

$$f(x,t) = e^{i(kx-\omega t)} \; . \tag{4.26}$$

La solution au temps $t + \Delta t$ sera donc

$$f(x, t+\Delta t) = e^{i(kx-\omega(t+\Delta t))} = f(x,t)e^{-i\omega\Delta t} \; . \tag{4.27}$$

L'amplitude de la perturbation au temps $t+\Delta t$ sera donc celle au temps t multipliée par le *gain*

$$G = e^{-i\omega\Delta t} \; . \tag{4.28}$$

La condition de stabilité est

$$|G| \leq 1 \; . \tag{4.29}$$

En effet, si $|G| > 1$, alors l'amplitude est multipliée par un facteur > 1 à chaque pas temporel, ce qui conduit à une croissance exponentielle de la perturbation. Il faut donc trouver et résoudre une équation pour G. On l'obtient en substituant la forme sinusoïdale, Eq.(4.26), dans le schéma numérique , Eq.(4.24). On obtient

$$G = 1 - \beta\left(1 - e^{-ik\Delta x}\right) + \alpha\left(e^{ik\Delta x} - 2 + e^{-ik\Delta x}\right) \tag{4.30}$$

$$G = 1 - \beta\left(1 - e^{-ik\Delta x}\right) - 4\alpha \sin^2\left(\frac{k\Delta x}{2}\right) \; . \tag{4.31}$$

Dans le cas de l'advection pure, $\alpha = 0$, on a

$$|G|^2 \leq 1 \quad \Leftrightarrow \quad (1 - \beta + \beta\cos(k\Delta x))^2 + \beta^2 \sin^2(k\Delta x) \leq 1 \; , \;\; \forall k$$

$$\Leftrightarrow 1 + \beta^2 + \beta^2\cos^2(k\Delta x) - 2\beta + 2\beta\cos(k\Delta x) - 2\beta^2\cos(k\Delta x) + \beta^2\sin^2(k\Delta x) \leq 1 \; , \;\; \forall k$$

$$2\beta^2 - 2\beta + \cos(k\Delta x)(2\beta - 2\beta^2) \leq 0 \; , \;\; \forall k$$

$$\beta(\beta-1)\left(1-\cos(k\Delta x)\right) \leq 0\ ,\quad \forall k$$

$$\Leftrightarrow \beta(\beta-1)2\sin^2\left(\frac{k\Delta x}{2}\right) \leq 0\ ,\quad \forall k \quad \Leftrightarrow \boxed{0 \leq \beta \leq 1}\,. \tag{4.32}$$

Cette condition de stabilité s'appelle le **critère CFL**. Dans le cas de la diffusion pure, $\beta = 0$, on a

$$|G|^2 \leq 1 \quad \Leftrightarrow \quad 1 - 8\alpha\sin^2\left(\frac{k\Delta x}{2}\right) + 16\alpha^2\sin^4\left(\frac{k\Delta x}{2}\right) \leq 1\ ,\quad \forall k$$

$$\Leftrightarrow 8\sin^2\left(\frac{k\Delta x}{2}\right)\alpha(1-2\alpha) \geq 0\ ,\quad \forall k$$

$$\Leftrightarrow \boxed{0 \leq \alpha \leq \frac{1}{2}}\,. \tag{4.33}$$

Les résultats numériques présentés aux sections précédentes vérifient bien ces propriétés : voir notamment les FIGS.4.3 et 4.6.

4.1.4 Diffusion et marche aléatoire

On a vu que le processus de diffusion est dû, au niveau microscopique, aux multiples collisions entre particules du système. Dans cette section, on présente un schéma numérique de simulation de la diffusion qui s'inspire directement de ce caractère aléatoire. Soit une particule du système. Soit un intervalle de temps Δt. Pendant cet intervalle, la particule va subir un certain nombre de collisions, qui auront pour effet de déplacer la particule. On décrit ce déplacement par une variable aléatoire, de distribution de probabilité uniforme en direction, et avec une variance finie pour sa norme. Au cours du temps, on suppose que les collisions successives sont indépendantes, du point de vue probabiliste.

On considère ensuite un ensemble de particules identiques, et on fait l'hypothèse que les collisions de chaque particule sont décrites par la même loi de probabilité, et qu'il n'y a aucune dépendance, au sens des probabilités, entre les collisions subies par ces particules.

Les systèmes réels sont constitués d'un nombre immense de particules, et il est irréaliste de vouloir les décrire toutes. On considère donc un **échantillonnage** de taille N. Chaque "particule numérique", en quelque sorte "représente" un grand nombre de particules réelles.

Pour obtenir une mesure de la densité des particules, on subdivise l'espace en N_{bin} "casiers", et on compte le nombre de particules numériques dans chaque casier, $n_{\text{bin},i}$, $i = 1..N_{\text{bin}}$. On obtient ainsi un histogramme, dont les valeurs sont proportionnelles à la densité.

L'alogorithme, qui fait partie de ce qu'on appelle la méthode de **Monte Carlo** à cause de son caractère aléatoire (tirage au sort, roulette, etc), est le suivant.

1. Initialisation : définition des "casiers", du nombre de particules numériques N, du pas de temps Δt et tirage au sort des positions initiales de chaque particule selon une distribution de probabilité proportionnelle à la densité initiale.
2. Boucle sur le temps (t_j)
3. Boucle sur les particules, $i = 1..N$
4. Tirage d'un nombre aléatoire R selon une loi de probabilité de moyenne nulle et de variance unité
5. Déplacer la particule $x_i \to x_i + \Delta x_i$, avec un déplacement Δx_i proportionnel au nombre aléatoire obtenu R
6. Fin de la boucle sur les particules
7. Compter le nombre de particules dans chaque casier (histogramme)
8. Fin de la boucle sur le temps

Il faut encore déterminer la relation entre la diffusion D et le déplacement Δx des particules. Plus exactement, on établit une relation entre le coefficient de diffusion D et la *variance* du déplacement. Une démonstration est faite à l'Annexe D. On trouve :

$$\sigma^2 =< \Delta x^2 >= 2D\Delta t \; . \tag{4.34}$$

On peut comprende ce résultat en invoquant le théorème central limite : le déplacement est le résultat d'une somme de déplacements indépendants. Donc, sa distribution tend vers une loi de probabilité gaussienne (normale) de variance proportionnelle au nombre de termes de la somme. Supposons qu'il y ait $n_{\rm coll}$ collisions durant l'intervalle de temps Δt. Le déplacement résultant obéira alors à une loi gaussienne de moyenne nulle et de variance proportionnelle au nombre de collisions. Ce nombre de collisions est proportionnel au coefficient de diffusion D et à la durée de l'intervalle Δt. Donc la variance du déplacement est proportionnelle à D et à Δt.

La position de la particule i au temps $j + 1$ (étape no.5 de l'algorithme) est donc

$$x_{i,j+1} = x_{i,j} + R\sqrt{2D\Delta t} \tag{4.35}$$

avec R la réalisation d'une variable aléatoire de moyenne nulle et de variance unité. Si on choisit pour R une distribution normale, on a alors un schéma valable pour des Δt arbitrairement grands.

Il est facile de combiner la diffusion avec une advection de vitesse v :

$$\boxed{x_{i,j+1} = x_{i,j} + v\Delta t + R\sqrt{2D\Delta t}} \; . \tag{4.36}$$

Un exemple est illustré à la FIG. 4.9, avec une distribution initiale gaussienne $x_0 = 0$, $\sigma = 0.2$m, $D = 0.1$m^2/s, $v = 0.1$m/s, $N = 10000$ particules numériques, $\Delta t = 0.05$. L'étalement de la densité est clairement visible, d'abord très rapide, puis ralentissant. Il

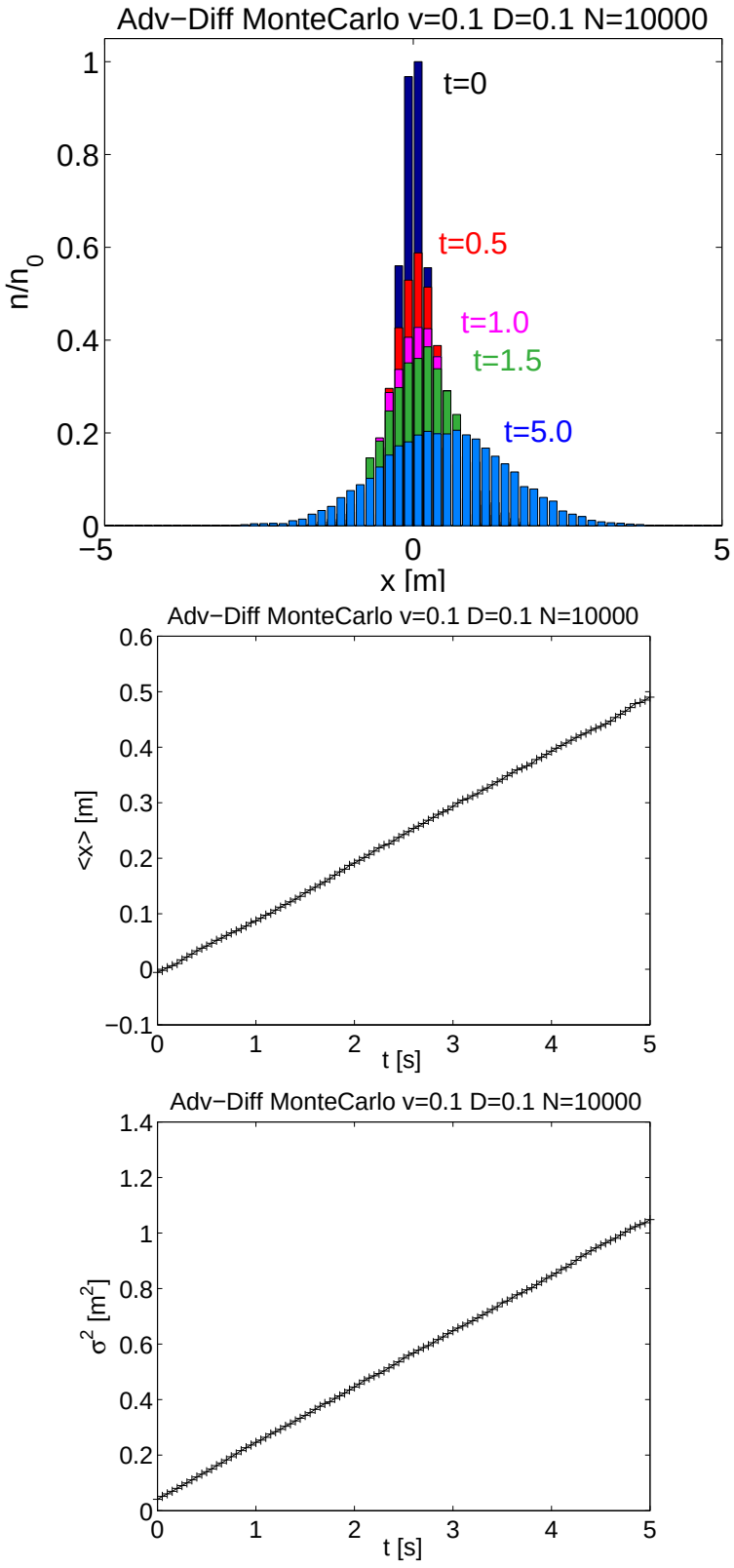


FIGURE 4.9 – *Diffusion et advection d'une densité. Schéma Monte Carlo. Paramètres :* $v = 0.1\mathrm{m/s}$, $D = 0.1\mathrm{m}^2/\mathrm{s}$, $N_{\mathrm{bin}} = 64$, $\Delta t = 0.05\mathrm{s}$. *Instantanés de la densité en fonction de* x *(en haut). Position moyenne* $< x > (t)$ *(au milieu) et variance* $\sigma^2(t)$ *(en bas).*

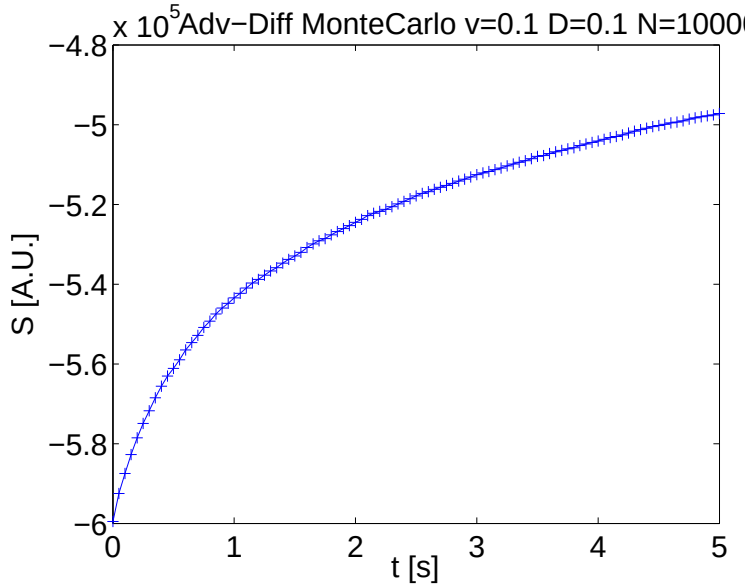


FIGURE 4.10 – *Mesure de l'entropie, pour la simulation de la FIG. 4.10.*

s'accompagne d'une translation de la position moyenne. L'analyse des résultats montre effectivement une variance $\sigma^2(t)$ augmentant linéairement avec le temps, et un mouvement uniforme de la moyenne $< x > (t)$.

Une quantité physique importante est l'entropie S. En physique statistique, une mesure de S est donnée par

$$S = -\sum_{i=1}^{N_{\text{bin}}} n_{\text{bin},i} \log n_{\text{bin},i} \tag{4.37}$$

La FIG. 4.10 montre que, conformément à la théorie, l'entropie est une fonction croissante du temps : le système, initialement très loin de l'équilibre thermodynamique, car présentant une densité avec des gradients très forts, s'approche de l'équilibre thermodynamique (qui est caractérisé par une densité uniforme) en augmentant son entropie. On a en effet un système fermé, et les résultats sont donc en accord avec le deuxième principe de la thermodynamique.

Le grand avantage de ce type de méthodes est qu'**il n'y a pas de limite de stabilité pour le paramètre CFL**. On peut donc, en principe, prendre des pas temporels Δt très grands. Il s'agit en effet d'une **méthode Lagrangienne**, c'est-à-dire que l'on suit l'écoulement avec les "particules numériques", contrairement aux schémas des sections précédentes, qui sont des **méthodes Euleriennes**, où le problème est discrétisé sur un maillage *fixe*.

Le grand désavantage est intrinsèque à la méthode Monte Carlo : chaque simulation doit être considérée comme la réalisation d'un ensemble de variables aléatoires. En termes plus précis, le nombre de particules numériques dans chaque "casier", qui est une mesure

de la densité recherchée, a un écart-type proportionnel à $\sqrt{N}$. Chaque simulation fournit un résultat différent, l'ensemble des résultats produisant une moyenne, mais aussi une dispersion statistique non nulle. L'écart-type sur l'estimation de la densité physique est donc proportionnel à $\boxed{1/\sqrt{N}}$. Voir aussi les remarques sur l'intégration Monte Carlo en Annexe B.3.

4.2 Ondes

Rappel. Une onde est une perturbation qui se propage dans l'espace et le temps. Dans ce cours, nous étudierons les phénomènes ondulatoires dits *linéaires*, c'est-à-dire où l'onde se propage dans un milieu en le perturbant suffisamment peu, pour que cela ne modifie pas les propriétés de propagation de l'onde. Signalons cependant que de nombreux phénomènes peuvent apparaître lorsque l'amplitude de la perturbation devient importante (ondes dites non-linéaires) : ondes de choc, auto-focalisation, désintégration paramétrique, etc.

4.2.1 Ondes en milieu homogène

Dans le cas où le milieu dans lequel l'onde se propage est homogène et isotrope, la perturbation, que nous noterons $f(\vec{x}, t)$, satisfait **l'équation d'Alembert** :

$$\frac{\partial^2 f}{\partial t^2} = u^2 \nabla^2 f \ , \tag{4.38}$$

avec $u = \text{const}$. Dans le cas unidimensionnel dans l'espace, $f(x, t)$, on a

$$\frac{\partial^2 f}{\partial t^2} = u^2 \frac{\partial^2 f}{\partial x^2} \ . \tag{4.39}$$

Au cours de physique, on aborde divers exemples. L'équation (4.39) peut modéliser les vibrations d'une corde, auquel cas f représente le déplacement transversal d'un élément de la corde. Elle peut modéliser les oscillations longitudinales d'un ressort, auquel cas f représente la déformation longitudinale d'un élément du ressort. Elle peut modéliser une onde sonore dans un tuyau, auquel cas f représente la *perturbation* de pression (ou de densité ou de vitesse longitudinale).

La solution générale de l'Eq.(4.39) est la superposition d'une onde dite "progressive" (perturbation propageant vers la droite) et d'une onde dite "rétrograde" (propageant vers la gauche) :

$$f(x, t) = F(x - |u|t) + G(x + |u|t), \tag{4.40}$$

où F et G sont des fonctions arbitraires (suffisamment régulières pour que l'Eq.(4.39) ait un sens).

Pour trouver une solution unique à l'équation d'Alembert, il faut préciser des **conditions aux bords** et des **conditions initiales**. Pour les **conditions aux bords** du domaine spatial $\Omega = [x_l,\ x_r]$, on distinguera 5 cas :

1. Condition de bord fixe : $f(x_e, t) = C,\ \forall t$, avec C une constante et x_e au bord du domaine, $x_e = x_l$ et/ou $x_e = x_r$ (p.ex. extrémité fixe d'une corde de guitare). C'est une condition dite de Dirichlet.

2. Condition de bord libre : $\partial f/\partial x(x_e, t) = 0$, $\forall t$ (p.ex. extrémité ouverte d'un tuyau d'orgue). C'est une condition dite de Neumann.
3. Conditions aux bords périodiques : $f(x_l, t) = f(x_r, t)$. On suppose que le système se répète indéfiniment et périodiquement dans l'espace. On ne discrétisera qu'une seule période spatiale.
4. Condition au bord harmonique : $f(x_e, t) = A\sin(\omega t)$, avec A une amplitude et ω une fréquence données. Cela simule l'excitation du système par une "antenne" de fréquence donnée. Condition dite de Dirichlet dépendante du temps.
5. Condition au bord de sortie de l'onde. Les conditions aux bords 1, 2 et4 ci-dessus conduisent au phénomène de la réflexion. On aimerait trouver une condition permettant la "sortie" de l'onde par les bords, sans provoquer de réflexion ni de retour de l'onde par l'autre bord comme c'est le cas pour des conditions périodiques. L'onde, au bord droite, $(x = x_r)$, sortira du domaine si elle est purement progressive au voisinage de $x = x_r$. Elle sortira du domaine au bord gauche si elle est purement rétrograde au voisinage de $x = x_l$.

Il faut encore déterminer les **conditions initiales**. Comme l'équation d'Alembert est du 2e ordre en temps, il faut préciser 2 conditions initiales. La plus simple à imposer est $f(x, t_0) = f_{\text{init}}(x)$, avec $f_{\text{init}}(x)$ une fonction donnée. Dans l'exemple de la corde vibrante, elle représente la forme de la corde au moment où le musicien "pince" la corde juste avant qu'il ne la lâche.

Mais il faut une deuxième condition initiale. Dans un premier temps, nous considèrerons un système initialement au repos, autrement dit $f(x, t) = f_{\text{init}}(x)$, $\forall x, \forall t \leq t_0$. Nous verrons plus loin comment initialiser le système pour générer une onde propageante soit vers la droite, soit vers la gauche.

Dans cette section, nous allons résoudre numériquement cette équation en utilisant une discrétisation de l'espace, $\{x_j\}_{j=1}^{Nx}$, et du temps $\{t_n\}_{n=1}^{Nt}$. On supposera les maillages en x et en t équidistants, avec $\Delta x = x_{j+1} - x_j$ et $\Delta t = t_{n+1} - t_n$. On approximera les opérateurs différentiels par des différences finies d'espace et de temps.

A l'ordre le plus bas, les différences finies pour les opérateurs deuxièmes dérivées, Eq.(A.7), introduites dans l'Eq. d'Alembert (4.39), donnent :

$$\frac{f(x_i, t_{n+1}) - 2f(x_i, t_n) + f(x_i, t_{n-1})}{(\Delta t)^2} \approx u^2 \left(\frac{f(x_{i+1}, t_n) - 2f(x_i, t_n) + f(x_{i-1}, t_n)}{(\Delta x)^2} \right) \tag{4.41}$$

On définit le **paramètre CFL** (Courant-Friedrichs-Lewy) [2]

$$\boxed{\beta = u\frac{\Delta t}{\Delta x}} \tag{4.42}$$

2. Courant, R. ; Friedrichs, K. ; and Lewy, H. "On the Partial Difference Equations of Mathematical Physics." IBM J. 11, 215-234, 1967.

et on peut réécrire l'expression ci-dessus comme

$$\boxed{f(x_i, t_{n+1}) \approx 2\left(1-\beta^2\right) f(x_i, t_n) - f(x_i, t_{n-1}) + \beta^2 \left[f(x_{i+1}, t_n) + f(x_{i-1}, t_n)\right]} \quad (4.43)$$

Cette expression nous donne une approximation pour f, en chaque point du réseau spatial ($x = x_i$), au temps ultérieur ($t = t_{n+1}$), en fonction de la solution aux instants présent ($t = t_n$) et antérieur ($t = t_{n-1}$), au même point spatial ($x = x_i$) et ses plus proches voisins ($x = x_{i\pm1}$).

Ceci est donc la base de l'algorithme. Il est à différences finies "à 3 niveaux", il y a en effet besoin de stocker 3 niveaux temporels (précédent, actuel et prochain).

Il faut encore exprimer la version discrétisée des conditions aux bords du domaine spatial (voir page précédente).

1. Bord gauche : $f(x_1, t_{n+1}) = f(x_1, t_n)$. Bord droite : $f(x_{Nx}, t_{n+1}) = f(x_{Nx}, t_n)$.
2. Bord gauche : $f(x_1, t_{n+1}) = f(x_2, t_{n+1})$. Bord droite : $f(x_{Nx}, t_{n+1}) = f(x_{Nx-1}, t_{n+1})$.
3. Bord gauche : substituer $i = 1$ et remplacer $i - 1$ par $N_x - 1$ dans l'Eq.(4.43). Bord droite : substituer $i = N_x$ et remplacer $i + 1$ par 2 dans l'Eq.(4.43).
4. En exercice.
5. Bord droite : dérivant f au voisinage de x_r par rapport à t et par rapport à x, on obtient

$$\frac{\partial f}{\partial t}(x_r, t) = \frac{\partial}{\partial t} F(x_r - |u|t) = F'(x_r - |u|t)(-|u|) = -|u| \frac{\partial f}{\partial x}(x_r, t) \quad (4.44)$$

   La version discrétisée de cette condition au bord s'obtient en utilisant les différences finies "backward" d'ordre le plus bas pour la première dérivée par rapport à x, $f_i' \approx (f_i - f_{i-1})/h$, exprimée au point de maillage $i = N_x$, et les différences finies "forward" pour la première dérivée par rapport à t, Eq.(A.22) :

$$\frac{f(x_{Nx}, t_{n+1}) - f(x_{Nx}, t_n)}{\Delta t} = -|u| \frac{f(x_{Nx}, t_n) - f(x_{Nx-1}, t_n)}{\Delta x} \quad (4.45)$$

   et ainsi

$$f(x_{Nx}, t_{n+1}) = f(x_{Nx}, t_n) - |\beta| \left[f(x_{Nx}, t_n) - f(x_{Nx-1}, t_n)\right] \quad (4.46)$$

   Bord gauche : en exercice.

Pour que l'algorithme explicite à 3 niveaux, Eq.(4.43), puisse démarrer, il faut initialiser f non seulement au temps t_0,

$$f(x_i, t_0) = f_{\text{init}}(x_i), \ \forall i, \quad (4.47)$$

mais aussi au temps $t_0 - \Delta t$. On a plusieurs possibilités, selon le problème que l'on vent résoudre.

1. Si on suppose le système immobile pout $t < t_0$, alors on prend

$$f(x_i, t_{-1}) = f(x_i, t_0) = f_{\text{init}}(x_i), \ \forall i. \quad (4.48)$$

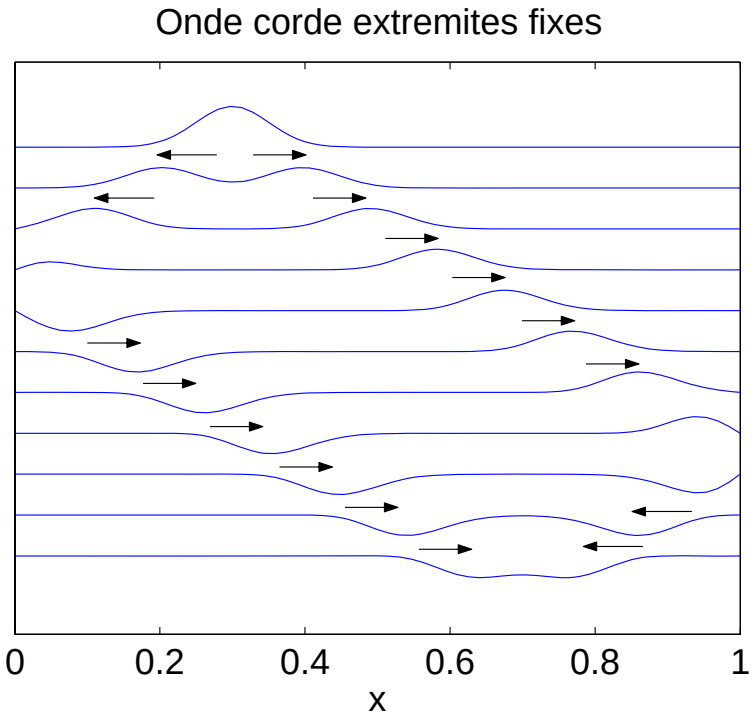


FIGURE 4.11 – *Propagation d'ondes sur une corde vibrante fixée à ses deux extrémités. Schéma explicite à 3 niveaux, Eq.(4.43). Paramètres : $u = 1$, CFL $\beta = 0.5$, $N_x = 65$. La perturbation initiale se décompose en onde progressive et onde rétrograde. Chacune subit une réflexion aux extrémités qui change le signe de la perturbation.*

2. Si on veut initialiser une onde propageant vers la droite, on utilise le fait que la solution doit s'écrire comme $F(x - |u|t)$, et donc

$$f(x_i, t_{-1}) = F(x_i - |u|(-\Delta t)) = f_{\text{init}}(x_i + |u|\Delta t), \ \forall i. \tag{4.49}$$

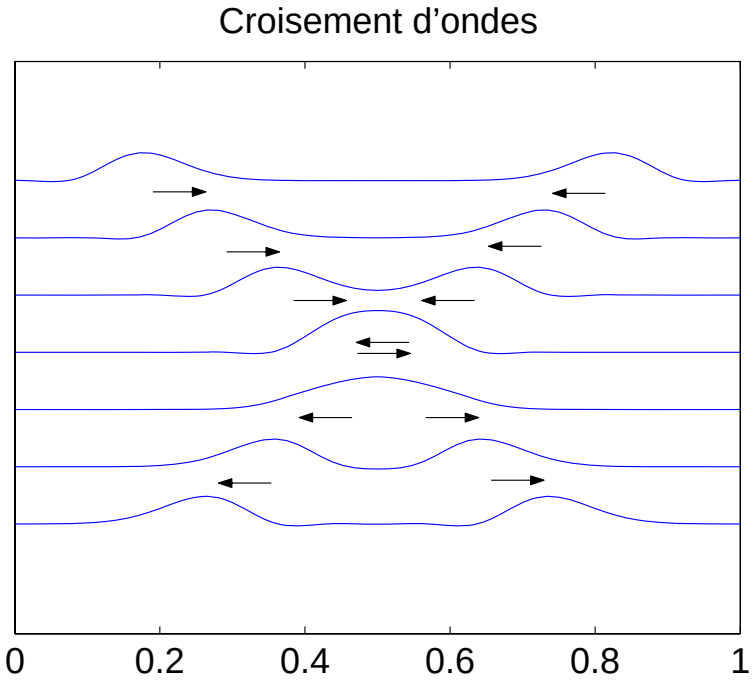
3. Si on veut initialiser une onde propageant vers la gauche, on utilise le fait que la solution doit s'écrire comme $G(x + |u|t)$, et donc

$$f(x_i, t_{-1}) = G(x_i + |u|(-\Delta t)) = f_{\text{init}}(x_i - |u|\Delta t), \ \forall i. \tag{4.50}$$

Un exemple est montré à la FIG. 4.11, pour le cas de conditions aux bords fixes (no.1), une perturbation initiale de forme gaussienne, et une condition initiale de type "système immobile" pour $t \leq 0$. La déformation initiale se sépare en deux "paquets" se propageant l'un à droite et l'autre à gauche. On remarque le changement de signe des perturbations lors de chaque réfléxion. On a utilisé $u = 1$, le paramètre CFL $\beta = 0.5$ et 64 intervalles (donc 65 points) en x.

On peut vérifier que, si on utilise la condition au bord libre (no.2), les perturbations sont réfléchies, mais avec le même signe que la perturbation incidente.

Avec les mêmes paramètres numériques, on peut vérifier le **principe de superposition linéaire** lors d'un croisement d'ondes, FIG. 4.12. Les perturbations propageantes



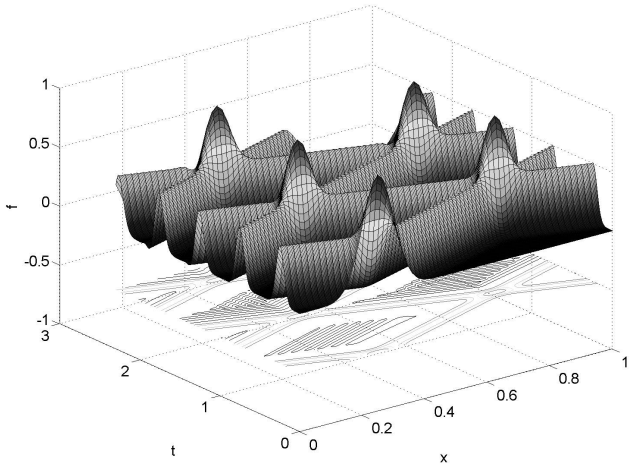


FIGURE 4.12 – *Croisement d'ondes. Schéma explicite à 3 niveaux, Eq.(4.43). Paramètres : $u = 1$, CFL $\beta = 0.5$, $N_x = 65$. Conditions aux bords périodoques. Les ondes progressive et rétrograde se traversent en s'ignorant mutuellement, sans se déformer.*

et rétrogrades se traversent mutuellement sans se déformer. C'est parce que l'équation fondamentale (d'Alembert) est **linéaire** que ce principe est vérifié.

Avec la condition au bord no.3 à gauche (excitation sinusoïdale) et la condition au bord droite no.1 (bord fixe), on peut observer le phénomène de **résonance**. Pour des valeurs bien déterminées de la fréquence d'excitation, on voit l'onde progressive et l'onde rétrograde (créée par la réflexion au bord) se superposer constructivement à chaque passage de l'onde, et on observe, pour des temps très longs, une **onde stationnaire** dont l'amplitude croît au cours du temps. Alors que si on choisit une fréquence entre ces fréquences déterminées, la superposition des ondes progressives et rétrogrades n'arrive pas à construire une onde stationnaire, et la perturbation reste de petite amplitude. Ces fréquences bien déterminées sont les **fréquences propres** du système, et les ondes stationnaires correspondantes sont les **modes propres** du système. On montre un exemple à la FIG. 4.13. On fait le calcul analytique de ces fréquences et modes propres en substituant l' Ansatz

$$f(x,t) = \hat{f}(x)e^{-i\omega t} \tag{4.51}$$

dans l'Eq. d'Alembert et en y appliquant les conditions aux bords. On trouve comme modes propres des fonctions sinusoïdales $\hat{f}_n(x) = sin(n\pi x/L)$, $n = 1, 2, 3, ...$ et des fréquences propres $\omega_n = nu\pi/L$, où $L = x_r - x_l$ est la longueur du système. Voir cours de Physique.

On peut se rendre compte que les fréquences et modes propres dépendent des conditions aux bords. On simulera **(suggestion d'exercice)** les fréquences et modes propres obtenus avec une condition au bord droite libre. On comparera avec les résultats analytiques.

4.2.2 Stabilité du schéma numérique : analyse de Von Neumann

Le paramètre crucial pour la stabilité numérique est le paramètre CFL $\beta = u\Delta t/\Delta x$, Eq.(4.42). On montre un exemple à la FIG. 4.14 d'une simulation avec $\beta = 1.01$, initialisée avec une perturbation gaussienne. La simulation se déroule normalement pour des temps courts, mais soudain une perturbation de courte longueur d'onde (2 points de maillage par longueur d'onde) apparaît, croissant exponentiellement dans le temps et finissant par "noyer" complètement la simulation.

L'analyse de la stabilité numérique se fait en examinant comment l'amplitude d'une perturbation sinusoïdale dans l'espace-temps évolue dans le temps par le schéma numérique, Eq.(4.43). On pose donc la solution au temps t comme

$$f(x,t) = e^{i(kx-\omega t)} \ . \tag{4.52}$$

La solution au temps $t + \Delta t$ sera donc

$$f(x,t+\Delta t) = e^{i(kx-\omega(t+\Delta t))} = f(x,t)e^{-i\omega\Delta t} \ . \tag{4.53}$$

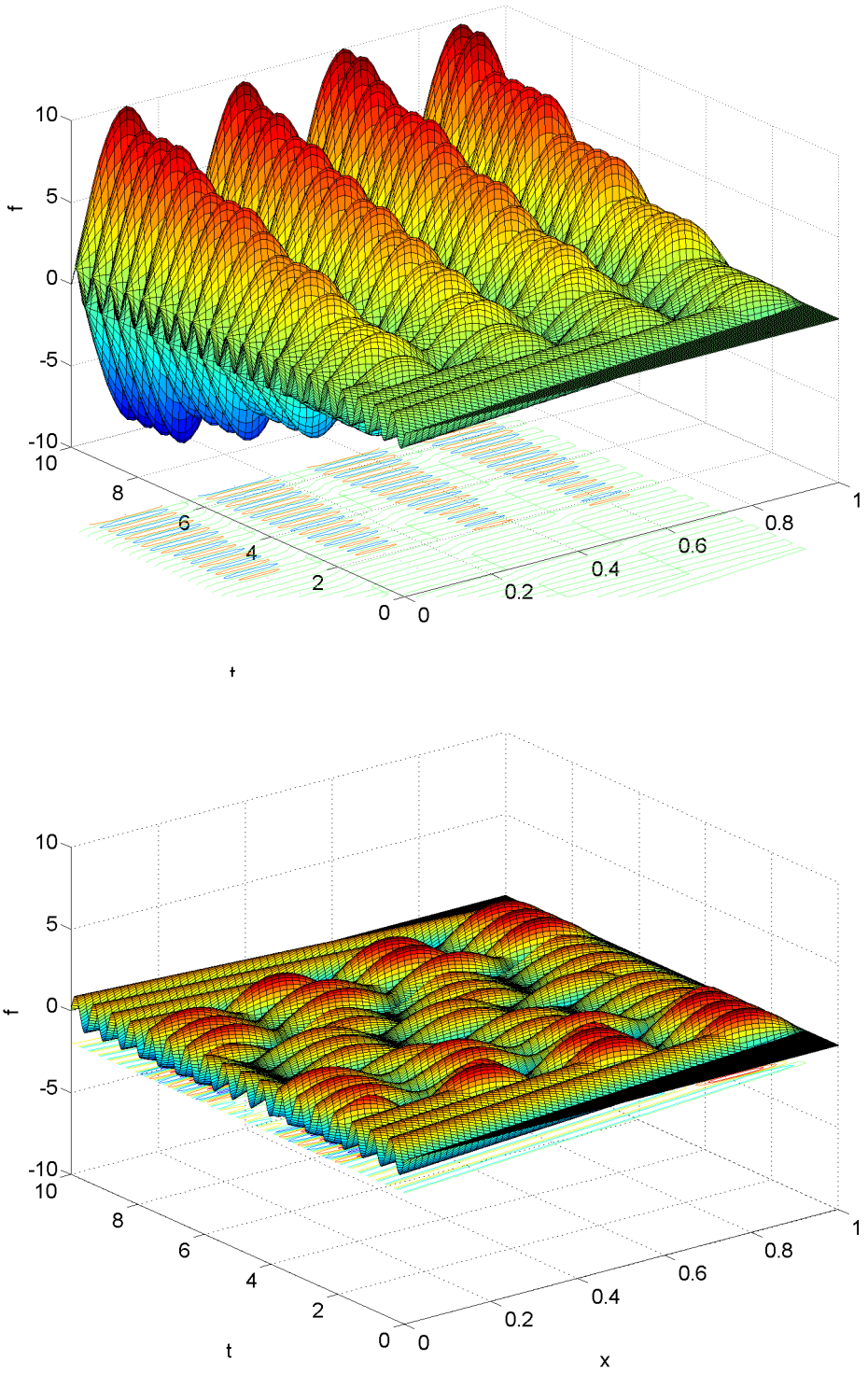


FIGURE 4.13 – *Résonance par excitation d'une fréquence propre su système. Schéma explicite à 3 niveaux, Eq.(4.43). Paramètres : $u = 1$, CFL $\beta = 1.0$, $N_x = 65$. Conditions aux bords fixe à droite, $\sin(\omega t)$ à gauche. Dans le cas où ω est une fréquence propre du système (en haut), il s'établit un mode propre, onde stationnaire, qui est d'amplitude croissante. Si ω n'est pas une fréquence propre su système (en bas), il ne s'établit pas d'onde stationnaire, et les perturbations restent de faible amplitude (en bas).*

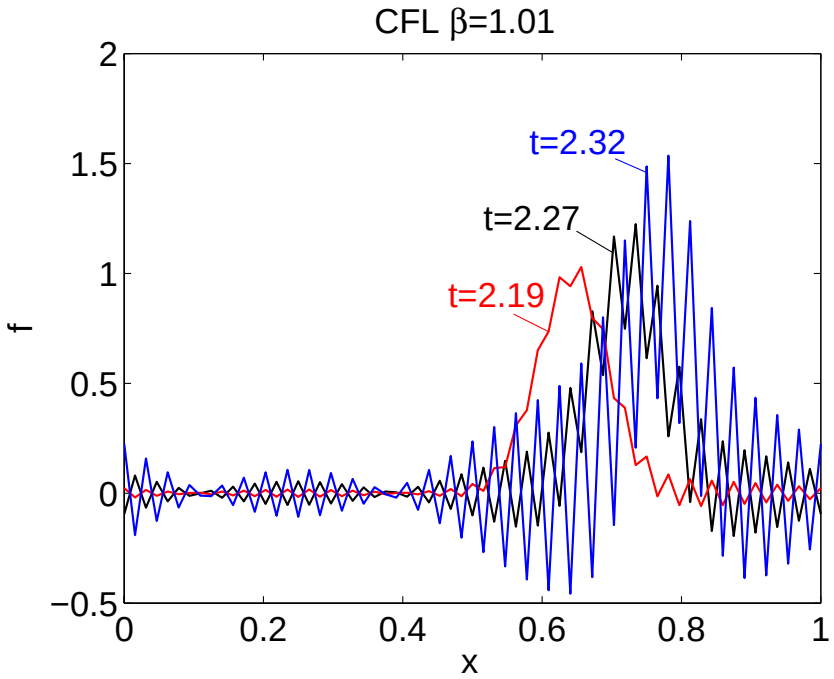


FIGURE 4.14 – *Instabilité du schéma explicite à 3 niveaux, Eq.(4.43). Paramètres : $u = 1$, CFL $\beta = 1.01$, $N_x = 65$. Conditions aux bords périodoques. L'instabilité se manifeste par la croissance exponentielle non physique d'une perturbation de courte longueur d'onde (2 points de maillage par longueur d'onde).*

L'amplitude au temps $t + \Delta t$ sera donc multipliée par le gain

$$G = e^{-i\omega\Delta t} \, . \tag{4.54}$$

La condition de stabilité est

$$|G| \leq 1 \, . \tag{4.55}$$

En effet, si $|G| > 1$, alors l'amplitude est multipliée par un facteur > 1 à chaque pas temporel, ce qui conduit à une croissance exponentielle. Il faut donc trouver et résoudre une équation pour G. On l'obtient en substituant la forme sinusoïdale, Eq.(4.52), dans le schéma numérique , Eq.(4.43).

$$\begin{aligned} e^{i(kx_j - \omega(t_n+\Delta t))} &= 2\left(1-\beta^2\right) e^{i(kx_j-\omega t_n)} - e^{i(kx_j-\omega(t_n-\Delta t))} \\ &+ \beta^2 \left[e^{i(k(x_j+\Delta x)-\omega t_n)} + e^{i(k(x_j-\Delta x)-\omega t_n)} \right] . \end{aligned} \tag{4.56}$$

Simplifiant par $e^{i(kx_j-\omega t_n)}$ et multipliant par G, on obtient

$$G^2 - 2\left[1 - 2\beta^2 \sin^2\left(\frac{k\Delta x}{2}\right)\right] G + 1 = 0. \tag{4.57}$$

Posant

$$\alpha = \frac{k\Delta x}{2} \, , \tag{4.58}$$

on a les solutions

$$G = 1 - 2\beta^2 \sin^2\alpha \pm \sqrt{(1 - 2\beta^2 \sin^2\alpha)^2 - 1} \, . \tag{4.59}$$

Si $|\beta| \leq 1$, alors le discriminant est négatif ou nul, et on a

$$G = 1 - 2\beta^2 \sin^2\alpha \pm i\sqrt{1-(1-2\beta^2\sin^2\alpha)^2} \tag{4.60}$$

et donc

$$|G|^2 = (1-2\beta^2\sin^2\alpha)^2 + 1 - (1-2\beta^2\sin^2\alpha)^2 = 1\ . \tag{4.61}$$

Pour CFL $|\beta| \leq 1$ le schéma numérique explicite à 3 niveaux est marginalement stable pour toute longueur d'onde.

Si $|\beta| > 1$, alors on a $|G| > 1$, au moins pour $\alpha = \pi/2$, ce qui correspond à 2 points de maillage par longueur d'onde. On comprend ainsi pourquoi ce sont ces perturbations-là qui deviennent instables en premier lieu, comme le montre l'exemple de la FIG. 4.14.

Suggestion d'exercice. Vérifier analytiquement et numériquement que le schéma explicite à 3 niveaux est toujours instable dans le cas $\beta^2 < 0$. Ce cas correspond à une onde *évanescente*.

4.2.3 Ondes en milieu inhomogène. Vitesse de phase variable

On peut, avec une modification bénigne de l'algorithme, considérer des cas où le milieu est inhomogène. Cela se traduit par une vitesse de phase qui est fonction de x, $u(x)$. L'équation d'Alembert, Eq.(4.39), doit être modifiée pour tenir compte de la variation de u. L'expression explicite pour $u(x)$ à laquelle on aboutit dépend du système physique considéré. On obtient généralement :

$$\frac{\partial^2 f}{\partial t^2} = \frac{\partial}{\partial x}\left(u^2(x)\frac{\partial f}{\partial x}\right)\ . \tag{4.62}$$

"Tsunami". Les ondes de gravitation dans les fluides incompressibles donnent, dans la limite d'une profondeur $h_0 << \lambda$, où λ est la longueur d'onde, une vitesse de phase

$$u(x) = \sqrt{g h_0(x)}\ . \tag{4.63}$$

Cette limite est appelée "ondes en eaux peu profondes". Ici, $h_0(x)$ est la profondeur de l'océan au repos, c'est-à-dire en l'absence de vagues. L'annexe E montre comment on aboutit à l' Eq.(4.62), avec $u(x)$ donné par l'Eq.(4.63). Il se trouve qu'elle s'applique, au moins partiellement, au cas d'une vague de type de celle qui apparaît lors d'un tsunami. On choisit une profondeur qui varie linéairement de $h_{0,\text{far}} = 7000$m à 1000km des côtes jusquà une profondeur de $h_{0,\text{reef}} = 200$m à 100km des côtes, puis linéairement jusqu'à la profondeur $h_{0,\text{beach}} = 20$m au bord. (On ne peut pas prendre une profondeur nulle : les équations deviennent singulières et, de plus, des phénomènes non-linéaires apparaîssent, dont nous ne tiendrons pas compte dans le cadre de ce cours). On note que pour 7000m de profondeur, la vitesse de propagation de la vague est de plus de 900 km/h !

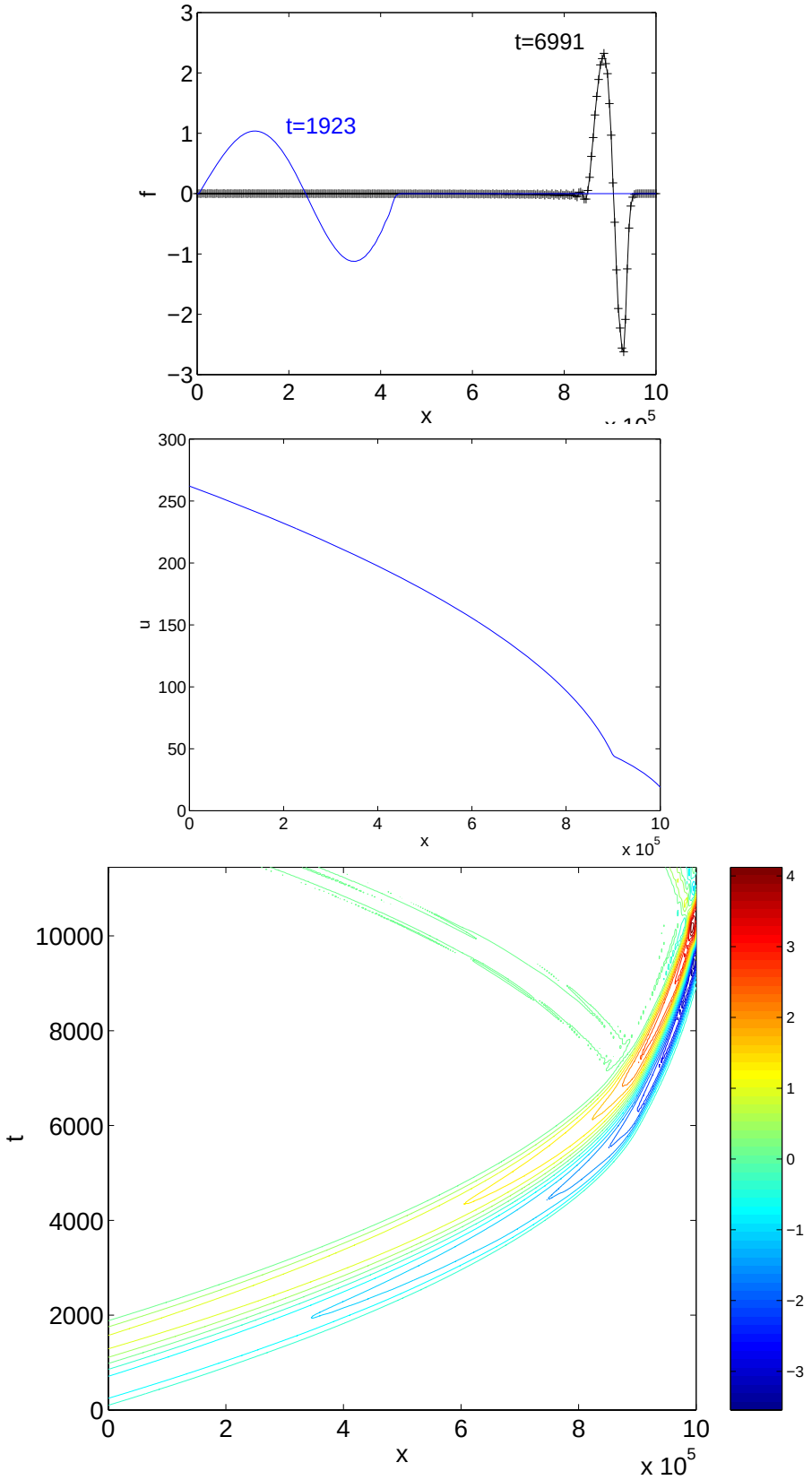


FIGURE 4.15 – *Simulation d'une onde dans l'océan se rapprochant des côtes. En haut : instantanés de la perturbation. Au milieu : vitesse de phase. En bas : lignes de niveau de la perturbation dans l'espace-temps* (x, t).

On intègre numériquement l'équation (4.62) avec le schéma de différences finies présenté à la section précédente, Eq.(4.43), modifié pour y ajouter le terme $(\partial u^2/\partial x)(\partial f/\partial x)$ (Suggestion d'exercice). On considère une excitation au bord gauche de type sinusoïdal (no.3) mais avec une seule période, et une condition au bord droite du type "sortie de l'onde" (no.5). A la FIG. 4.15, on montre deux instantanés de la perturbation. On remarque que la longueur d'onde raccourcit et que l'amplitude augmente lorsque la vague se rapproche de la côte. La vitesse de propagation, par contre, diminue à mesure que la vague se rapproche de la côte. Cela est très clair sur l'image des lignes de niveau de la perturbation en fonction de x et de t.

4.2.4 Approximation analytique : la méthode WKB

La méthode WKB (Wentzel-Kramers-Brillouin) a été développée en 1926 pour décrire le comportement d'une particule dans un potentiel par la mécanique quantique. Jeffreys avait déjà en 1923 développé une méthode générale pour approximer les solutions des équations différentielles linéaires du deuxième ordre, ainsi la méthode est parfois appelée "WKBJ" ou "JWKB". On en esquisse ici les grandes lignes, pour notre problème ondulatoire classique.

1) On considère des solutions sinusoïdales du temps, $f(x,t) = \hat{f}(x)e^{-i\omega t}$. En substituant dans l'Eq.(4.62), on a

$$-\omega^2 \hat{f} = \frac{\mathrm{d}}{\mathrm{d}x}\left(u^2(x)\frac{\mathrm{d}\hat{f}}{\mathrm{d}x}\right) . \tag{4.64}$$

2) On fait l'Ansatz

$$\hat{f}(x) = A(x)e^{iS(x)} . \tag{4.65}$$

La substitution de l'Ansatz dans l'Eq.(4.65) donne, en notant $\mathrm{d}/\mathrm{d}x$ avec le symbole $'$,

$$-\omega^2 A = -(S')^2u^2A + i\left(2S'A'u^2 + S''Au^2 + S'A(u^2)'\right) + A''u^2 + A'(u^2)' . \tag{4.66}$$

3) On suppose que l'amplitude $A(x)$ est une fonction "lentement" variable, alors que la phase $S(x)$ est "rapidement" variable. (Les termes "lentement" et "rapidement" qualifient ici une variation selon x, et non temporelle). On fait l'hypothèse que la "lente" variation de l'amplitude $A(x)$ est liée au fait que le terme $u^2(x)$ varie, lui aussi, "lentement". On résout l'Eq.(4.66) par approximations successives, en supposant l'existence d'un paramètre d'ordre, suffisamment "petit", que nous noterons ϵ. Nous classons ensuite les différents termes apparaissant dans l'équation selon leur ordre en ϵ. Ainsi, $S(x)$ varie "rapidement", ce qui signifie que sa variation selon x est grande. On aura donc

$$S' \sim \epsilon^0 \tag{4.67}$$

De même, u^2 et A ne sont pas "petits", donc

$$u^2 \sim \epsilon^0, \quad A \sim \epsilon^0, \tag{4.68}$$

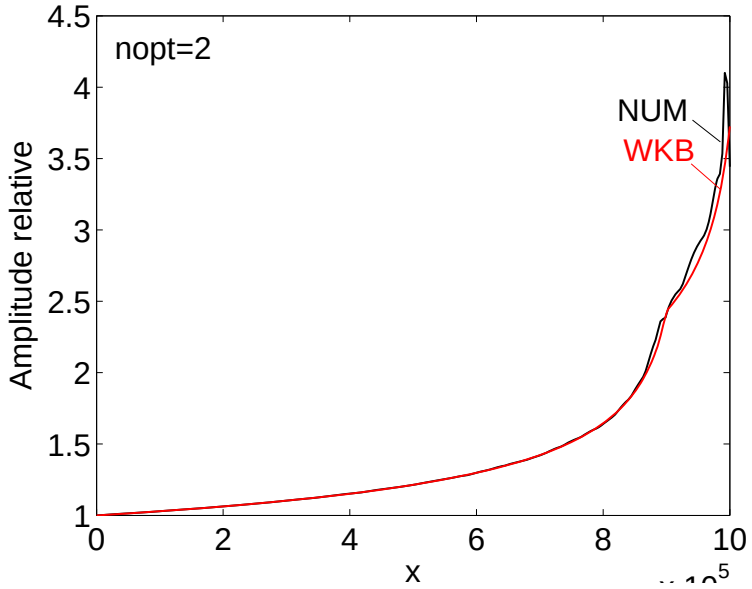


FIGURE 4.16 – *Simulation d'une onde dans l'océan se rapprochant des côtes. Amplitude de la perturbation en fonction de x, obtenue avec le schéma explicite à 3 niveaux (courbe noire "NUM") et comparée avec la solution analytique approchée par la méthode WKB (courbe rouge "WKB"). Les deux approches mettent en évidence l'augmentation de l'amplitude avec la diminution de la vitesse de propagation.*

mais leur variation, i.e. leur dérivée selon x, est "petite", et nous allons la supposer *du même ordre, d'ordre 1*, donc

$$(u^2)' \sim \epsilon^1, \quad A' \sim \epsilon^1. \tag{4.69}$$

La variation d'un terme, i.e. la dérivé selon x d'un terme, s'accompagne de l'augmentation de l'ordre d'une unité. Ainsi,

$$S'' \sim \epsilon^1, \quad , (u^2)'' \sim \epsilon^2, \quad A'' \sim \epsilon^2. \tag{4.70}$$

La multiplication de deux termes additionne leur ordre, ainsi par exemple,

$$(u^2)'A' \sim \epsilon^2, \quad u^2A'S' \sim \epsilon^1, ...etc. \tag{4.71}$$

Revenant à l' Eq.(4.66), on a donc que le membre de gauche et le premier terme du membre de droite sont d'ordre le plus bas ($\sim \epsilon^0$), le deuxième terme est du premier ordre ($\sim \epsilon^1$), alors que les deux derniers termes sont du deuxième ordre ($\sim \epsilon^2$), que nous allons négliger.

La méthode consiste ensuite à résoudre l'Eq.(4.66) *ordre par ordre*. On a donc, à l'ordre 0 :

$$S' = \frac{\omega}{u} \, . \tag{4.72}$$

On définit le "nombre d'onde local"

$$k(x) = \frac{\mathrm{d}S}{\mathrm{d}x} \, , \tag{4.73}$$

et ainsi on a la "relation de dispersion locale"

$$k(x) = \frac{\omega}{u(x)} \ , \tag{4.74}$$

avec $\lambda(x) = 2\pi/k(x)$ définissant une "longueur d'onde locale". Dans l'approximation WKB, on suppose que la longueur d'onde locale varie "lentement", autrement dit varie peu à l'échelle d'une longueur d'onde : $\lambda'/\lambda \sim k'/k = S''/k \sim \epsilon^1$.

A l'ordre 1, annullant le deuxième terme du membre de droite de l'Eq.(4.66), et en y substituant la solution à l'ordre 0, Eq.(4.72), on obtient

$$2uA' + u'A = 0 \ . \tag{4.75}$$

En supposant que u et A ne s'annullent jamais, on a $A'/A = -(1/2)(u'/u)$, donc $(\log A)' = -(1/2)(\log u)'$, et on obtient

$$\boxed{A(x) = \frac{A_0}{\sqrt{u(x)}}}. \tag{4.76}$$

L'amplitude augmente donc lorsque la vitesse de phase dimunue. Pour le cas des vagues en eaux peu profondes, on a $u(x) = \sqrt{gh_0(x)}$ et $A(x) = A_0/(h_0(x))^{1/4}$: lorsque la vague se rapproche des côtes, h_0 diminue et donc la vitesse de propagation u diminue, mais hélas l'amplitude de la vague augmente. On montre à la FIG. 4.16 la comparaison entre la méthode numérique (courbe "NUM") et la solution Eq.(4.76) obtenue par la méthode WKB (courbe "WKB"). L'accord est excellent. Il est intéressant de réaliser ce que représente cette figure : il s'agit de la **comparaison entre une solution numérique approximative et une solution analytique approximative**. Les approximations faites numériquement et analytiquement étant de natures complètement différentes, ce type de comparaison est très utile pour vérifier à la fois le schéma numérique et l'approximation analytique.

Il est absolument crucial que l'équation soit du type de l'Eq.(4.62) pour que le comportement ci-dessus soit correctement décrit. Si l'équation était

$$\frac{\partial^2 f}{\partial t^2} = u^2(x)\frac{\partial^2 f}{\partial x^2} \ , \tag{4.77}$$

l'analyse WKB (en exercice) montre qu'alors on obtiendrait non pas une amplitude $A(x) \propto 1/\sqrt{u(x)} \propto 1/(h_0(x))^{1/4}$, mais $A(x) \propto \sqrt{u(x)} \propto (h_0(x))^{1/4}$: la vague diminuerait d'amplitude en se rapprochant des côtes au lieu d'augmenter ! Si l'équation était

$$\frac{\partial^2 f}{\partial t^2} = \frac{\partial^2}{\partial x^2}\left(u^2(x) f\right) \ , \tag{4.78}$$

l'analyse WKB (en exercice) montre qu'alors on obtiendrait une amplitude $A(x) \propto 1/(u(x))^{3/2} \propto 1/(h_0(x))^{3/4}$, ce qui, dans l'exemple du tsunami, impliquerait qu'une vaguelette de 10cm de haut au large (profondeur 7000m) aurait une amplitude de 8m (!) près des côtes (profondeur 20m).

4.3 Schrödinger

La mécanique quantique ne décrit pas les particules comme des "points matériels", comme en mécanique classique. Les particules ont en fait un comportement

- probabiliste : on ne peut prédire qu'une probabilité de détecter une particule à un endroit donné ;
- corpusculaire : au moment où on la détecte, une particule est indivisible ;
- ondulatoire : la probabilité de présence d'une particule est généralement le résultat d'une interférence.

Voir le cours de Physique IV, puis le cours de Physique Quantique I, pour plus de détails.

On décrit une particule par une *fonction d'onde* $\psi(\vec{x},t)$, à valeurs *complexes*. Une particule de masse m soumise au potentiel $V(x)$ obéit à **l'équation de Schrödinger** :

$$\boxed{i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi\ . + V(\vec{x})\psi} \tag{4.79}$$

Définissant *l'hamiltonien* du système

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V\ , \tag{4.80}$$

l'équation de Schrödinger s'écrit

$$i\hbar\frac{\partial\psi}{\partial t} = H\psi\ . \tag{4.81}$$

On interprète $|\psi(\vec{x},t)|^2$ comme la densité de probabilité de trouver la particule au voisinage de $\vec{x}$ au temps t. Définissant le produit scalaire

$$(\eta,\psi) = \int \eta^*\psi d^3x\ , \tag{4.82}$$

où l'intégrale est sur tout l'espace, on doit avoir

$$(\psi,\psi) = 1\ , \quad \forall t. \tag{4.83}$$

La probabilité que la particule existe "quelque part" est toujours 1 (pas de "disparition" de la particule).

Nous nous limiterons dans la suite au cas unidimensionnel dans l'espace : $\psi(x,t)$.

4.3.1 Schéma semi-implicite de Crank-Nicolson

Etant donné une fonction d'onde, supposée connue à $t = 0$, $\psi(x,0)$, on peut formellement intégrer l'équation de Schrödinger :

$$\psi(x,t) = \exp\left(-\frac{i}{\hbar}tH\right)\psi(x,0)\ , \tag{4.84}$$

où on a défini *l'Hamiltonien*

$$H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x) \,, \tag{4.85}$$

ainsi que l'opérateur exponentiel d'un opérateur A :

$$(\exp A)\,(\psi) = \psi + A(\psi) + \frac{1}{2}A\,(A(\psi)) + ... = \sum_{n=0}^{\infty}\frac{A^n}{n!}(\psi) \,. \tag{4.86}$$

L'opérateur exponentiel a certaines propriétés qui rappellent celles de la fonction exponentielle, par exemple $\exp\left((\lambda_1+\lambda_2)A\right) = \left(\exp(\lambda_1 A)\right)\left(\exp(\lambda_2 A)\right)$ pour tous $\lambda_1, \lambda_2 \in \mathbb{C}$. Attention toutefois, en général les opérateurs ne commutent pas, $[A,B] \equiv AB - BA \neq 0$, et en général $\exp(A+B) \neq \exp(A)\exp(B)$. Une autre propriété est que $\exp(iA)$ est unitaire si et seulement si A est hermitien. Ainsi, l'opérateur d'évolution temporelle apparaissant dans l'Eq.(4.84) ci-dessus, $T = \exp(-itH/\hbar)$, est *unitaire*, car H est hermitien (on dit aussi "auto-adjoint"). Il conserve la probabilité totale :

$$\begin{aligned}(\psi(x,t),\psi(x,t)) &= (T\psi(x,0),T\psi(x,0)) = (\psi(x,0),T^*T\psi(x,0)) \\ &= (\psi(x,0),\exp(+itH/\hbar)\exp(-itH/\hbar),\psi(x,0)) \\ &= (\psi(x,0),\psi(x,0)) \,.\end{aligned} \tag{4.87}$$

La fonction d'onde est de norme 1, et cette norme reste constante au cours du temps. L'approximation numérique de cet opérateur doit aussi avoir cette propriété. Ceci suggère le schéma suivant. On définit un maillage du temps, avec des intervalles équidistants Δt. On a :

$$\begin{aligned}\psi(x,t+\Delta t) &= \exp\left(-(i/\hbar)(t+\Delta t)H\right)\psi(x,0) \\ &= \exp\left(-(i/\hbar)\Delta t\, H\right)\exp\left(-(i/\hbar)t\, H\right)\psi(x,0) \\ &= \exp\left(-(i/\hbar)\Delta t\, H\right)\psi(x,t) \,.\end{aligned} \tag{4.88}$$

Appliquant l'opérateur $\exp\left((i/\hbar)(\Delta t/2)\, H\right)$ à gauche et à droite, on obtient

$$\exp\left(\frac{i}{\hbar}\frac{\Delta t}{2}\,H\right)\psi(x,t+\Delta t) = \exp\left(-\frac{i}{\hbar}\frac{\Delta t}{2}\,H\right)\psi(x,t) \,. \tag{4.89}$$

Jusqu'ici, tout est exact. C'est à ce stade que nous faisons une approximation : nous ne retenons que les termes jusqu'au premier ordre dans le développement définissant l'opérateur exponentiel. On obtient ainsi :

$$\boxed{\left(1+\frac{i}{\hbar}\frac{\Delta t}{2}\,H\right)\psi(x,t+\Delta t) = \left(1-\frac{i}{\hbar}\frac{\Delta t}{2}\,H\right)\psi(x,t)} + \mathcal{O}(\Delta t^2) \tag{4.90}$$

Le schéma (4.90) a été développé par **Crank et Nicolson** en 1947, originellement pour résoudre l'équation de la chaleur dépendante du temps. Il est dit **semi-implicite** : la solution en $t+\Delta t$ dépend en partie *explicitement* de la solution en t (membre de droite de (4.90)), et en partie *implicitement* (membre de gauche). La partie implicite est un opérateur qu'il faut *inverser* pour trouver la solution en $t+\Delta t$:

$$\psi(x,t+\Delta t) = \left(1+\frac{i}{\hbar}\frac{\Delta t}{2}\,H\right)^{-1}\left(1-\frac{i}{\hbar}\frac{\Delta t}{2}\,H\right)\psi(x,t) + \mathcal{O}(\Delta t^2) \tag{4.91}$$

L'opérateur d'évolution temporelle discrétisé ci-dessus conserve la probabilité totale. Posons $\alpha = (\Delta t/2\hbar)H$. Soit l'opérateur d'évolution temporelle discrétisé

$$T_{\Delta t} = (1 + i\alpha)^{-1}(1 - i\alpha) \; . \tag{4.92}$$

Examinons la **réversibilité** de l'algorithme. Changer $t \to -t$ implique $\Delta t \to -\Delta t$ et donc $\alpha \to -\alpha$. De l'Eq.(4.90), l'opérateur d'évolution temporelle "en marche arrière" est

$$T_{-\Delta t} = (1 - i\alpha)^{-1}(1 + i\alpha) \; . \tag{4.93}$$

Première propriété :

$$T_{-\Delta t} = T^*_{\Delta t} \; , \tag{4.94}$$

où on a noté par * l'adjoint de l'opérateur. [Rappel : $A*$ est opérateur adjoint de $A \Leftrightarrow (\eta, A^*\varphi) = (A\eta, \varphi), \forall\eta, \forall\varphi$.] La preuve de cette propriété est la suivante : l'opérateur α est hermitien puisque H l'est : $(H\eta, \varphi) = (\eta, H\varphi)$, donc $(1+i\alpha)^* = (1-i\alpha)$. Soit $A = 1+i\alpha$. On a $T^*_{\Delta t} = (A^{-1}A^*)^* = A(A^*)^{-1}$. Or, $A(A^*)^{-1} = (A^*)^{-1}A$: en effet, multipliant cette dernière relation à gauche et à droite par A^*, on a $A^*A = AA^*$, qui est bien toujours vérifié, puisque égal à $1 + \alpha^2$. Donc $T^*_{\Delta t} = (A^*)^{-1}A = T_{-\Delta t}$.

La deuxième propriété est la **réversibilité** :

$$T_{-\Delta t} = T^{-1}_{\Delta t} \; . \tag{4.95}$$

En d'autres termes, faire un pas temporel en avant, puis un pas temporel en arrière, conduit *exactement* à la condition initiale. La preuve de cette propriété s'exprime come suit :

$$T_{-\Delta t}T_{\Delta t} = (1 - i\alpha)^{-1}(1 + i\alpha)(1 + i\alpha)^{-1}(1 - i\alpha) = (1 - i\alpha)^{-1}1(1 - i\alpha) = 1 \; . \tag{4.96}$$

Les deux propriétés ci-dessus conduisent au fait que l'opérateur $T_{\Delta t}$ est **unitaire** :

$$\boxed{T^{-1}_{\Delta t} = T^*_{\Delta t}} \; . \tag{4.97}$$

Ainsi, l'opérateur d'évolution temporelle discrétisé conserve la probabilité :

$$\begin{aligned} (\psi(x, t + \Delta t), \psi(x, t + \Delta t)) &= (T_{\Delta t}\psi(x, t), T_{\Delta t}\psi(x, t)) = (\psi(x, t), T^*_{\Delta t}T_{\Delta t}\psi(x, t)) \\ &= (\psi(x, t), \psi(x, t)) \; . \end{aligned} \tag{4.98}$$

Pour la discrétisation spatiale de (4.90), nous avons le choix de plusieurs méthodes : par exemple les éléments finis, voir Section 3.3, ou les différences finies. Cette dernière méthode, avec l'Eq.(A.7) pour l'opérateur $\partial^2/\partial x^2$, donne le système algébrique linéaire suivant, écrit sous forme matricielle : **(suggestion d'exercice)**

$$\begin{pmatrix} \cdot \\ -a \\ 1 + 2a + b \\ -a \\ \cdot \end{pmatrix}^T \begin{pmatrix} \cdot \\ \psi_{j-1} \\ \psi_j \\ \psi_{j+1} \\ \cdot \end{pmatrix} (t + \Delta t) = \begin{pmatrix} \cdot \\ a \\ 1 - 2a - b \\ a \\ \cdot \end{pmatrix}^T \begin{pmatrix} \cdot \\ \psi_{j-1} \\ \psi_j \\ \psi_{j+1} \\ \cdot \end{pmatrix} (t) \tag{4.99}$$

avec

$$a = \frac{i\hbar}{4m}\frac{\Delta t}{\Delta x^2} \ , \quad b = \frac{i}{\hbar}\frac{\Delta t}{2}V(x_i) \ . \tag{4.100}$$

C'est un système matriciel tridiagonal du type $\boxed{\mathbf{A}\Psi_{t+\Delta t} = \mathbf{B}\Psi_t}$, où Ψ_t est le vecteur des $\psi(x_i, t)$, valeurs de ψ aux points du maillage spatial x_i, au temps t.

Selon les conditions aux bords, il faudra les imposer explicitement sur le système matriciel. On utilise ensuite une des méthodes standard pour la résolution du système matriciel. Par exemple l'élimination de Gauss, comme à la section 3.3.

4.3.2 Particule libre

Une particule dite "libre" n'est soumise à aucune force. Elle se déplace dans un potentiel $V(x)$ constant, que l'on peut prendre nul. Les relations entre les quantités corpusculaires (quantité de mouvement et énergie) et ondulatoires (nombre d'onde et fréquence) décrivant la particule ont été données par de Broglie :

$$\boxed{\begin{array}{rcl} \vec{p} & = & \hbar\vec{k} \\ E & = & \hbar\omega \end{array}} \tag{4.101}$$

Ces relations sont écrites pour une particule ayant *une* quantité de mouvement $\vec{p}$ et *une* énergie E bien définies. La fonction d'onde correspondante est du type onde plane (sinusoïdale), qui en 1-D s'écrit

$$\psi(x,t) \sim \exp(i(kx - \omega t)) \ , \tag{4.102}$$

où k et ω sont liés par la *relation de dispersion* suivante, obtenue en substituant l'Ansatz onde plane ci-dessus, Eq. (4.102), dans l'équation de Schrödinger, Eq.(4.79), avec $V = 0$:

$$\omega = \omega(k) = \frac{\hbar k^2}{2m} \ . \tag{4.103}$$

Par les relations de de Broglie (4.101), cette relation de dispersion entre quantités ondulatoires ω et k, n'est autre que la relation entre les quantités corpusculaires E et p pour la particule libre :

$$E = \frac{p^2}{2m} \ . \tag{4.104}$$

L'équation de Schrödinger étant linéaire, toute superposition de solutions est aussi solution. Ainsi, on construit la solution de Schrödinger comme une somme d'ondes planes :

$$\psi(x,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \hat{\psi}(k)\exp(i(kx - \omega(k)t))dk \ . \tag{4.105}$$

Du point de vue quantique, la particule libre se comporte donc comme une superposition d'ondes. La vitesse de groupe,

$$v_g = \frac{\partial\omega}{\partial k} = \frac{\hbar k}{m} \ , \tag{4.106}$$

correspond à la vitesse de la particule dans la représentation classique de la particule, via la relation de quantification de de Broglie, Eq.(4.101). La vitesse de phase,

$$v_p = \frac{\omega}{k} = \frac{\hbar k}{2m}\,, \tag{4.107}$$

n'a pas d'équivalent dans la représentation classique. On notera que la vitesse de phase dépend de la longueur d'onde, comme pour un milieu *dispersif*. Les ondes qui se propagent dans de tels milieux sont *déformables* : la forme spatiale de l'onde change au cours du temps. C'est une des propriétés que nous allons examiner plus en détail par la suite.

L'expression de la fonction d'onde à $t = 0$

$$\psi(x,0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{\psi}(k) \exp(ikx) dk \tag{4.108}$$

indique que $\hat{\psi}(k)$ est la **transformée de Fourier** de l'état initial. On a :

$$\hat{\psi}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \psi(x,0) \exp(-ikx) dx \tag{4.109}$$

On peut former un "paquet d'onde" initial en superposant des ondes planes avec des poids $\hat{\psi}(k)$. L'extension spatiale Δx de ce paquet d'onde est liée à l'extension dans l'espace de Fourier Δk par le théorème de Fourier. En définissant précisément Δx comme l'écart-type de la distribution de probabilité pour la position,

$$\Delta x = \sqrt{\langle x^2 \rangle - \left(\langle x \rangle\right)^2} \tag{4.110}$$

avec

$$\langle x^m \rangle = \int_{-\infty}^{+\infty} x^m |\psi(x,t)|^2 dx, \tag{4.111}$$

et Δk comme l'écart-type de la distribution de probabilité pour le nombre d'onde k,

$$\Delta k = \sqrt{\langle k^2 \rangle - \left(\langle k \rangle\right)^2} \tag{4.112}$$

avec

$$\langle k^m \rangle = \int_{-\infty}^{+\infty} k^m |\hat{\psi}(k)|^2 dk, \tag{4.113}$$

le théorème de Fourier s'écrit :

$$(\Delta x)(\Delta k) \geq 1/2 \tag{4.114}$$

ce qui, via la relation de Broglie, $p = \hbar k$, correspond au **principe d'incertitude** de Heisenberg :

$$\boxed{(\Delta x)(\Delta p) \geq \hbar/2}\,. \tag{4.115}$$

Dans la suite, on choisira un système d'unités tel que $\hbar = 1$, et une particule de masse $m = 1/2$.

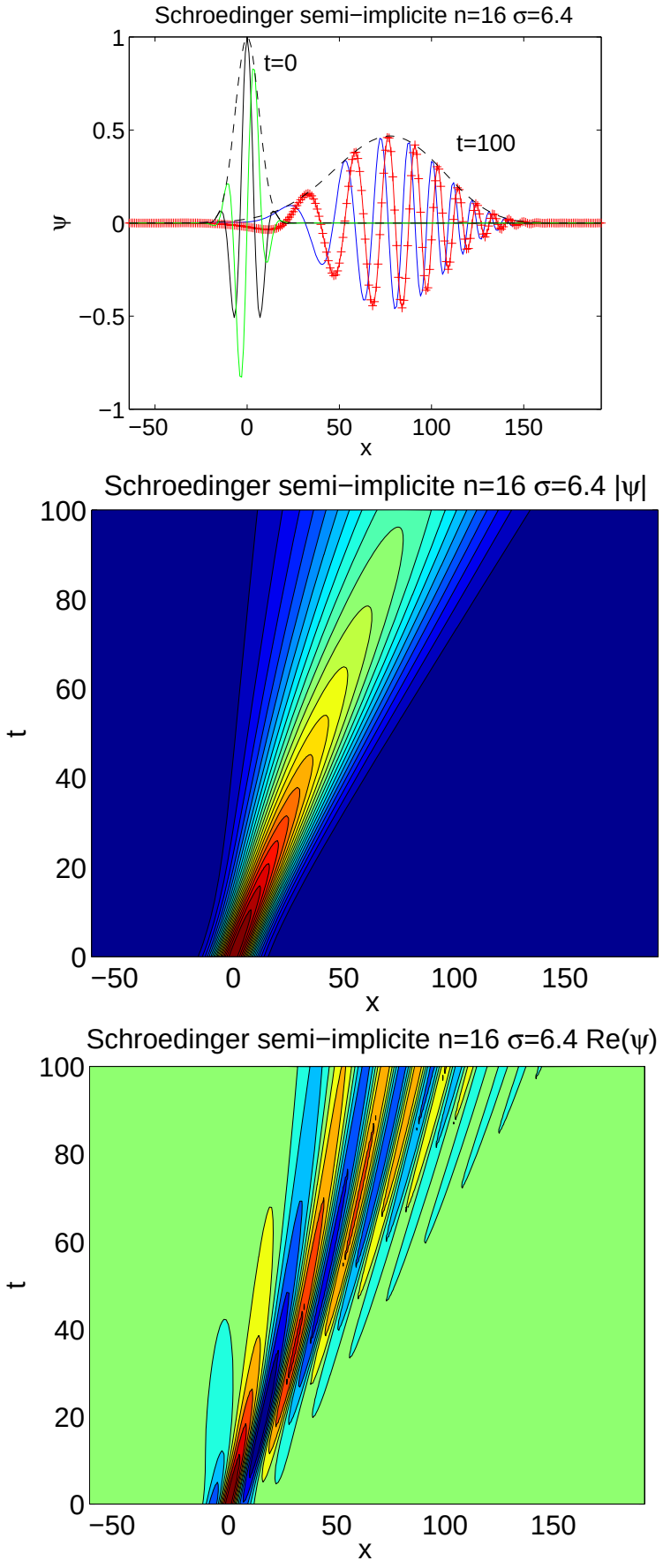


FIGURE 4.17 – *Particule libre avec une incertitude initiale $\Delta x = 6.4$ et un nombre d'onde moyen $n = 16$. Haut : $Re(\psi)$ et $Im(\psi)$ en traits continus, $|\psi|$ en traitillés. Milieu : contours de $|\psi(x,t)|$. Bas : contours de $Re(\psi(x,t))$. Le centre du paquet d'onde (max de $|\psi|$) se déplace à la vitesse $\hbar k_0/m$. L'incertitude sur la position augmente (étalement du paquet d'onde). La vitesse de phase est différente de la vitesse de groupe.*

Les propriétés de la particule libre sont illustrées par la simulation de la FIG. 4.17. On donne le paquet d'onde initial de forme gaussienne

$$\psi(x,0) = C\exp(ik_0x)\exp[-(x-x_0)^2/(2\sigma^2)]\ , \quad k_0 = n2\pi/L\ , \tag{4.116}$$

où $n = 16$ est le nombre d'onde moyen, $L = 256$ est la longueur du domaine de simulation, $x_0 = 0$ est la position initiale du maximum de $|\psi|$, $\sigma = 6.4$ la largeur de la gaussienne et C est une constante de normalisation telle que

$$\int_{-\infty}^{+\infty} |\psi(x,0)|^2 dx = 1\ , \tag{4.117}$$

ce qui donne

$$|C|^2 \int_{-\infty}^{+\infty} e^{-y^2}\sigma\,\mathrm{d}y = 1 \ \Rightarrow\ |C| = \frac{1}{\sqrt{\sigma\sqrt{\pi}}}\ . \tag{4.118}$$

On a appliqué le schéma semi-implicite, Eq.(4.99), avec $\Delta x = 1$, $\Delta t = 0.5$. Le maximum de $|\psi(x,t)|$, autrement dit la position la plus probable de la particule, se déplace à la vitesse $v_g^{\rm num} = 0.775$, en bon accord avec la solution analytique $v_g = \hbar k/m = (2\pi n/L)/(1/2) = 0.785$. L'effet **dispersif** se manifeste par un **étalement** du paquet d'onde au cours du temps : l'incertitude sur la position augmente au cours du temps. On remarque aussi que les composantes de courte longueur d'onde du paquet d'onde se propagent plus rapidement que les composantes de longue longueur d'onde, ce qui est également conforme à l'analyse. L'image de la partie réelle de $\psi(x,t)$ (bas de la FIG. 4.17) montre bien que la vitesse de phase est inférieure à la vitesse de groupe, en accord avec la théorie.

Finalement, on vérifie que la probabilité est conservée : la mesure de $\int |\psi(x,t_j)|^2 dx$ aux temps t_j donne un résultat constant à la précision machine près (10^{-14}).

Etalement du paquet d'onde : de la différence entre diffusion et dispersion

On peut montrer (voir cours de Physique 4 et de Mécanique Quantique I) [3] que la solution exacte de l'équation de Schrödinger pour une particule libre dans l'état initial donné par le paquet d'onde Gaussien (4.116) est telle que

$$|\psi(x,t)|^2 = \sqrt{\frac{1}{\pi}}\frac{1}{\sigma}\frac{1}{\sqrt{1+\frac{\hbar^2t^2}{m^2\sigma^2}}}\exp\left(-\frac{\left(x-\frac{\hbar k_0}{m}t\right)^2}{\sigma^2\left(1+\frac{\hbar^2t^2}{m^2\sigma^4}\right)}\right)\ . \tag{4.119}$$

Cette quantité, rappelons-le, est la densité de probabilité de trouver la particule en x au temps t. La fonction d'onde a donc à tous les temps une forme gaussienne, mais son écart-type varie au cours du temps, ce qui donne une incertitude sur la position, $<\Delta x>$ donnée par

$$<\Delta x>(t) = <\Delta x>(0)\ \sqrt{1+\frac{\hbar^2t^2}{m^2\sigma^4}}\ . \tag{4.120}$$

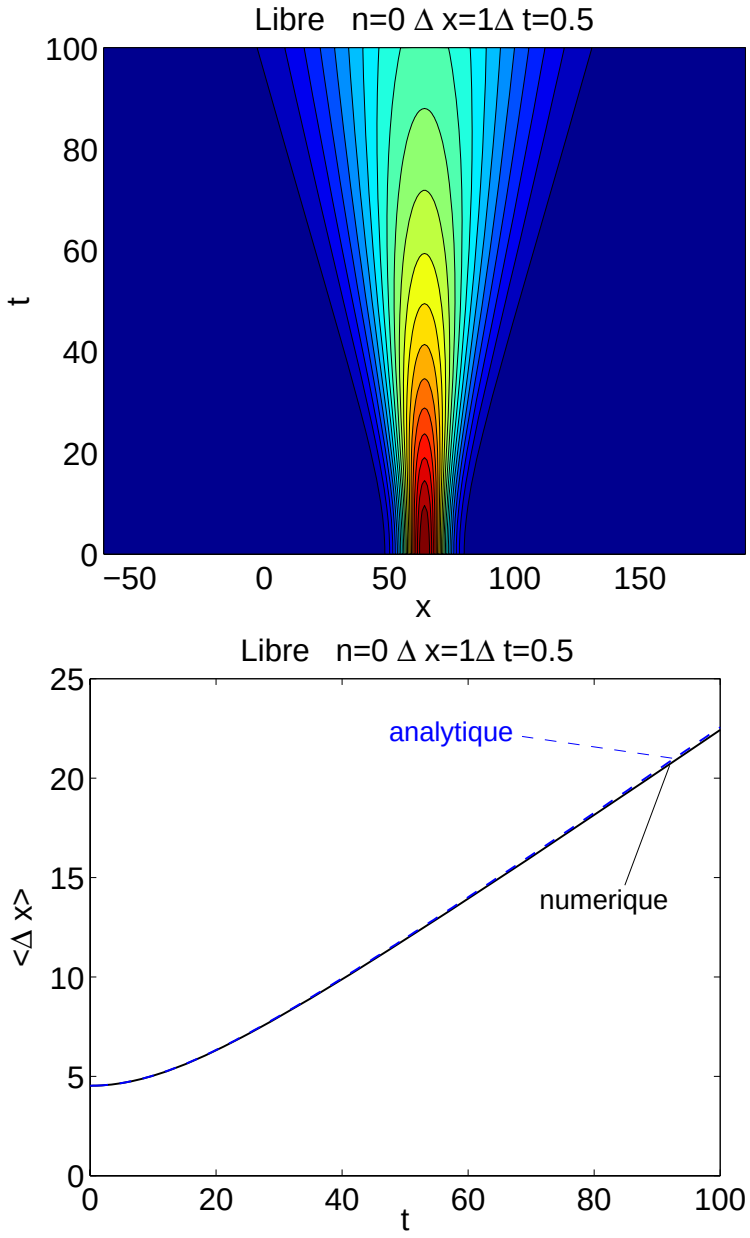


FIGURE 4.18 – *Particule libre avec une fonction d'onde initiale gaussienne de largeur* $\sigma = 6.4$ *et un nombre d'onde moyen* $n = 0$. *En haut : contours de* $|\psi(x,t)|$. *En bas : incertitude sur la position* $< \Delta x > (t)$ *; en traitillés, la solution analytique, Eq.(4.120). L'incertitude sur la position augmente (étalement du paquet d'onde).*

L'incertitude sur la position augmente donc au cours du temps, d'autant plus rapidement que le paquet d'onde initial a une incertitude petite. (N.B. : $< \Delta x > (0) = \sigma/\sqrt{2}$). Cette augmentation, pour des temps longs, est *linéaire avec le temps*. On illustre ceci avec une simulation numérique de mêmes paramètres que pour la FIG. 4.17, sauf que l'on considère une particule initialement "immobile", avec $k_0 = 0$. On montre à la FIG. 4.18 le module de la fonction d'onde dans l'espace et le temps, ainsi que l'incertitude $< \Delta x > (t)$. L'étalement est bien visible. On a reporté aussi la solution analytique en traitillés. Le petit écart entre la solution analytique et la solution numérique est dû à l'effet de la discrétisation spatiale (Δx) et temporelle (Δt).

L'étalement de la fonction d'onde de la particule libre pourrait, à première vue, ressembler à de la diffusion. On pourrait imaginer qu'une particule est constituée d'un grand nombre de points matériels, distribués selon une certaine densité, et qui, à cause de multiples collisions aléatoires entre eux, donnerait lieu à un processus diffusif, résultant en un étalement. Or, cette image est totalement fausse. Dans un processus de diffusion classique, nous avons montré à la section 4.1.2 que la largeur de la fonction augmentait comme la racine carrée du temps, alors qu'en mécanique quantique l'étalement est proportionnel au temps (pour des temps suffisaments longs). On peut comparer la FIG. 4.5 de la diffusion au résultat quantique de la FIG. 4.18.

L'étalement de la fonction d'onde d'une particule libre en mécanique quantique n'est pas dû à de la diffusion, mais à la **dispersion**. L'origine en est la relation de dispersion (4.103), qui indique que la vitesse de phase dépend du nombre d'onde k. Or, un paquet d'onde de largeur finie consiste en une somme d'ondes planes ayant des k différents. Dans notre cas du paquet d'onde gaussien initial, on a un ensemble de valeurs de k centrées autour de k_0. Les composantes ayant un k élevé vont se propager plus vite que les composantes ayant un k plus petit. Traduisons : les longueurs d'onde les plus courtes vont se propager plus vite que les longues longueur d'onde, ce qui est visible sur l'image du haut de la FIG. 4.17. C'est ce phénomène qui, au cours du temps, contribue à "étaler" le paquet d'onde.

4.3.3 Barrière de potentiel : résonances et effet tunnel

On considère une particule incidente sur un potentiel de forme carrée, de hauteur V_0 et d'épaisseur δ. L'état initial est un paquet d'onde de forme gaussienne, Eq.(4.116), de nombre d'onde moyen $n = 32$ et de largeur $\sigma = 0.075$. Le domaine de simulation a une longueur $L = 256$. Les paramètres numériques sont $n_x = 512$, $\Delta x = 0.5$, $\Delta t = 0.5$. Le schéma semi-implicite, Eq.(4.99), est utilisé. Les unités sont choisies avec $\hbar = 1$, et la masse de la particule est $m = 1/2$.

La relation (4.103) entre ω et k, d'une part, et la relation de de Broglie (4.101) entre E

3. Ref. C. Cohen-Tannoudji, Mécanique Quantique I, complément G_I, p.64-67

et ω, d'autre part, donnent une énergie moyenne $E_0 = \hbar^2 k^2/2m = 0.6169$. Remarque : comme on a un paquet d'onde dont la largeur dans l'espace de Fourier Δk est non nulle, la particule n'a pas une énergie bien définie : il y a incertitude non nulle ΔE.

Nous allons étudier le comportement de la particule pour différentes hauteurs V_0 et épaisseurs δ de la barrière de potentiel.

Cas $V_0 < E_0$

On rappelle que la solution analytique de Schrödinger, pour le cas d'ondes planes "monochromatiques", c'est-à-dire ayant une énergie bien déterminée (donc un k de la particule incidente unique), prédit une probabilité généralement non nulle que la particule soit réfléchie par la barrière. Ceci est contraire à la prédiction de la physique classique, pour laquelle la particule passerait *avec certitude* par dessus la barrière si $V_0 < E$.

D'autre part, la mécanique quantique prédit aussi que la probabilité de réflexion de la particule n'augmente pas de façon monnotone avec l'épaisseur δ de la barrière. Notamment, pour des épaisseurs de barrière δ multiples de π/k_t, la probabilité de transmission est 1, donc celle de réflexion est nulle. k_t est le nombre d'onde de la solution ψ dans la barrière,

$$k_t = \sqrt{2m(E_0 - V_0)}/\hbar \, . \tag{4.121}$$

Remarque : comme $E_0 > V_0$, la solution est propageante à l'intérieur de la barrière.

Nous allons illustrer ces propriétés avec des simulations numériques. Soit $V_0 = 0.8E_0$. Pour ces paramètres, $\pi/k_t = 8.94$. Pour une épaisseur $\delta = 4.5$, la figure 4.19 montre que la particule a une probabilité non nulle d'être réfléchie.

Pour une épaisseur plus élevée, $\delta = 18$, la probabilité de réflexion est bien plus petite. Ceci est en accord avec la théorie, on remarque en effet que δ est proche de $2\pi/k_t$, qui est une condition de résonance prédite par la théorie dans le cas d'une particule "monochromatique" d'énergie bien définie. Dans notre cas, la particule n'a pas *une* énergie E_0, son état ayant une incertitude non nulle.

Cas $V_0 > E_0$

Avec les mêmes paramètres pour la particule incidente, mais cette fois $V_0 = 1.2E_0$, les résultats de la FIG. 4.20, en haut, pour une épaisseur de barrière $\delta = 2.5$, montrent que la particule a une probabilité non nulle de traverser la barrière. Ce comportement, appelé **effet tunnel**, est complètement différent de la prédiction de la physique classique,

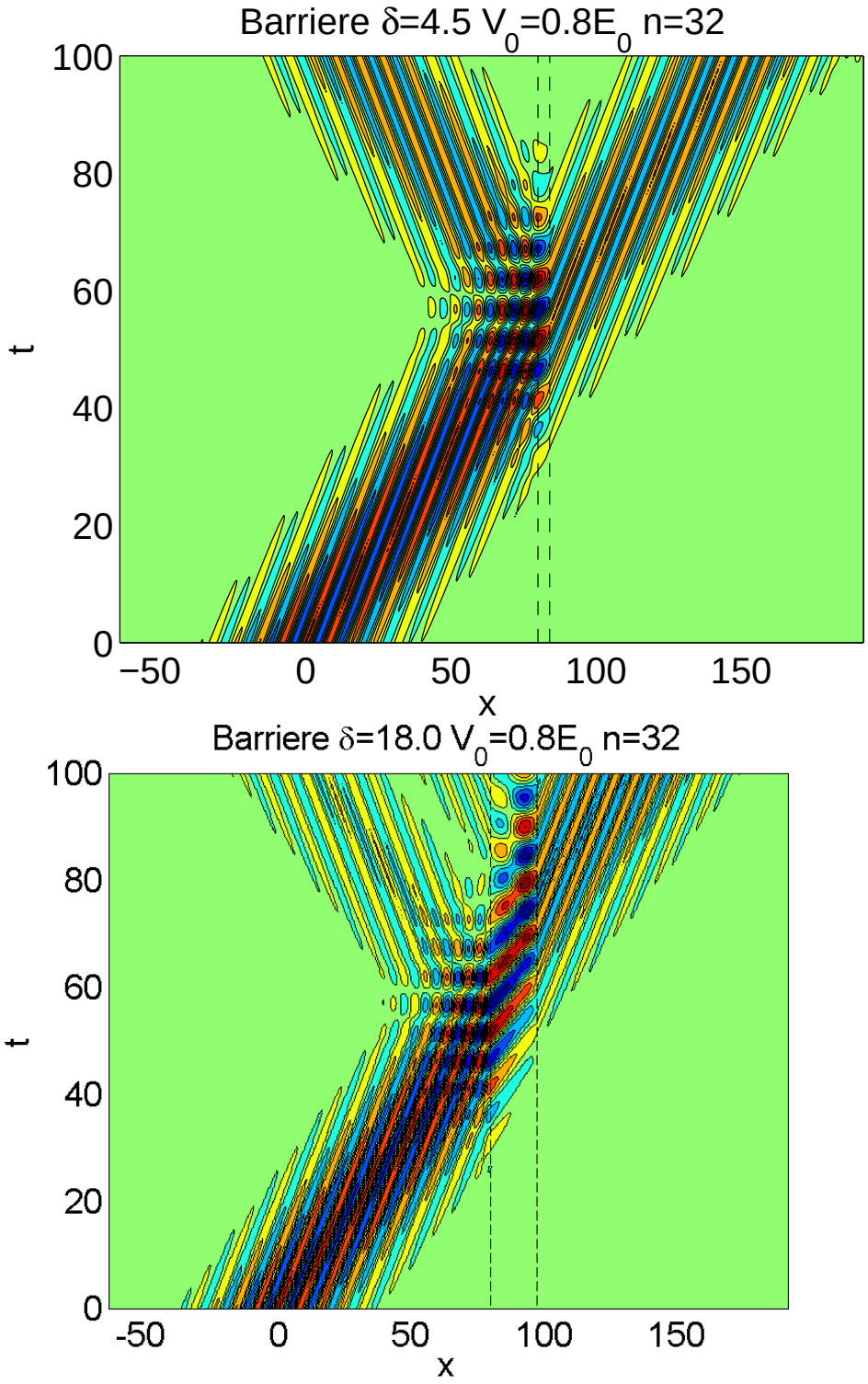


FIGURE 4.19 – *En haut : particule incidente sur une barrière de potentiel de hauteur* $V_0 = 0.8E_0$ *et de largeur* $\delta = 4.5$ *(lignes traitillées). La particule a une probabilité non nulle d'être réfléchie. En bas, pour* $\delta = 18$*, correspondant à peu près à une condition de résonance prédite par la théorie, la probabilité de réflexion est bien plus petite. La quantité représentée est la partie réelle de* $\psi(x,t)$*.*

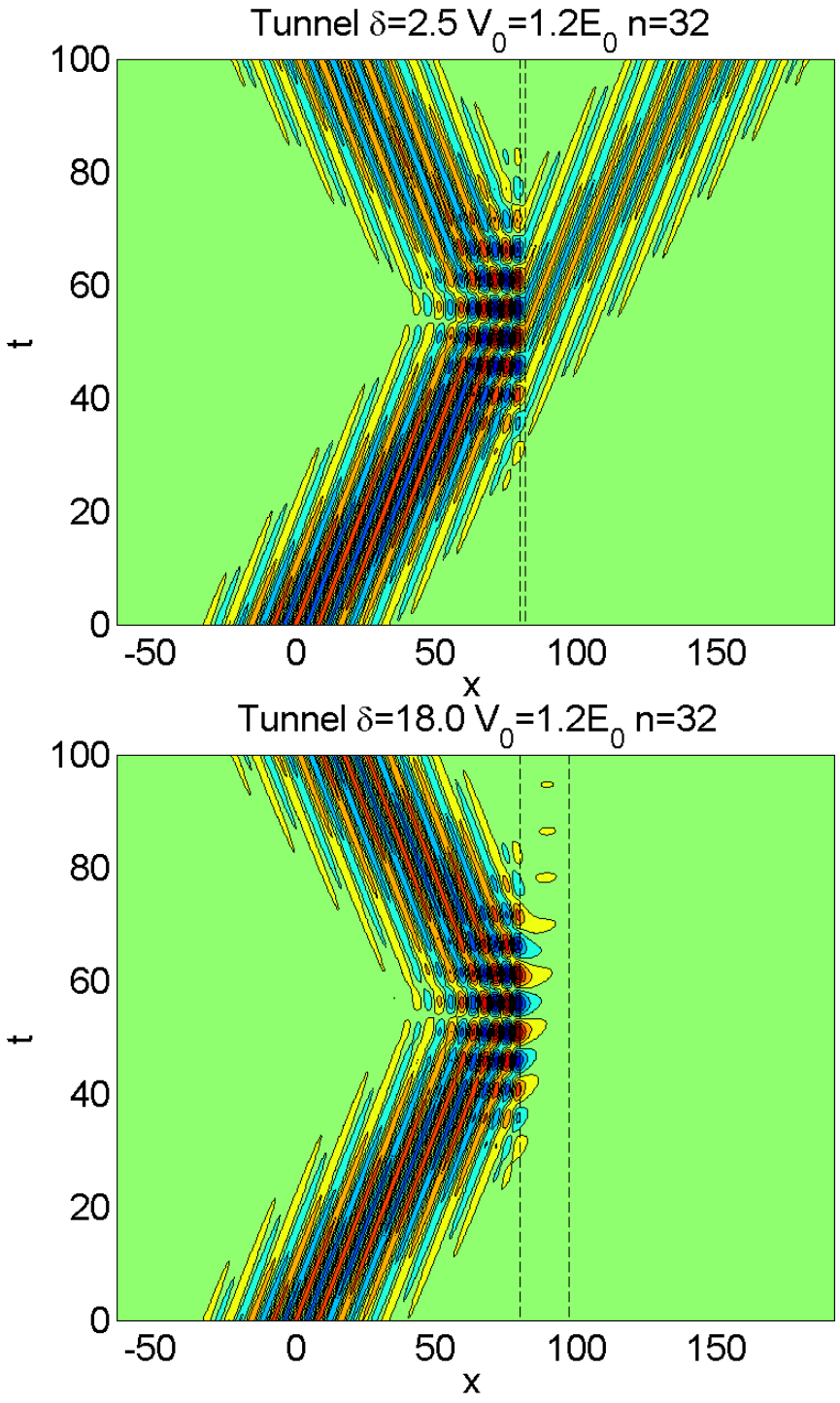


FIGURE 4.20 – *En haut : particule incidente sur une barrière de potentiel de hauteur* $V_0 = 1.2E_0$*, de largeur* $\delta = 2.5$ *(lignes traitillées). La particule a une probabilité non nulle de traverser la barrière (effet tunnel). En bas, pour une largeur plus importante,* $\delta = 18$*, cette probabilité devient exponentiellement petite, et la réflexion est pratiquement totale. La fonction d'onde est évanescente dans la barrière. La quantité représentée est la partie réelle de* $\psi(x,t)$*.*

où la particule serait réfléchie à coup sûr. En bas, pour une largeur plus importante, $\delta = 18$, cette probabilité devient exponentiellement petite, et la réflexion est pratiquement totale. La fonction d'onde est évanescente dans la barrière : elle a une amplitude exponentiellement décroissante en fonction de x

4.3.4 Oscillateur harmonique

Soit une particule de masse m dans un potentiel quadratique

$$V(x) = \frac{1}{2} m\omega_0^2 x^2 \ , \tag{4.122}$$

avec ω_0 une constante donnée. (N.B. : dans le cas classique d'une masse attachée à un ressort de constante K, on a $\omega_0 = \sqrt{K/m}$.) Nous allons calculer le comportement de cette particule tel que la mécanique quantique le prédit, et nous allons essayer de trouver quelles analogies il est possible de faire avec le mouvement prédit par la mécanique classique.

Classiquement, on sait que le mouvement est sinusoïdal, de fréquence angulaire $\omega = \omega_0 = \sqrt{K/m}$. Si l'énergie mécanique de la particule est E, alors son mouvement est confiné entre $x_{\min}$ et $x_{\max}$ donnés par les solutions de $V(x) = E$. On a donc la trajectoire classique

$$x_{\text{class}}(t) = \sqrt{\frac{2E}{m}} \frac{1}{\omega_0} \sin(\omega_0 t + \varphi) \ . \tag{4.123}$$

Quantiquement, on verra dans la section suivante qu'une particule ayant une énergie E bien déterminée ne peut généralement pas exister sauf pour des valeurs bien spécifiques de l'énergie E. Dans cette section, nous considérerons une particule dans un état initial décrit par un "paquet d'onde", comme aux sections précédentes, Eq.(4.116) avec une extension spatiale σ, une position moyenne x_0 et un nombre d'onde moyen k_0 donnés.

Ainsi, la particule n'a pas *une* position, *une* quantité de mouvement, *une* vitesse, et *une* énergie. Mais on peut montrer (théorème d'Ehrenfest) que la valeur *moyenne* de la position, définie par

$$< x > (t) = (\psi, x\psi) = \int \psi^*(x,t) x \psi(x,t) dx \ , \tag{4.124}$$

et la valeur *moyenne* de la quantité de mouvement, définie par

$$< p > (t) = (\psi, p\psi) = \int \psi^*(x,t) (-i\hbar)(\partial/\partial x) \psi(x,t) dx \ , \tag{4.125}$$

satisfont les équations du mouvement classique

$$\frac{\mathrm{d} < p >}{\mathrm{d}t} = < -\frac{\mathrm{d}V}{\mathrm{d}x} > \tag{4.126}$$

$$\frac{\mathrm{d} < x >}{\mathrm{d}t} = < \frac{p}{m} > \ . \tag{4.127}$$

Nous n'allons pas démontrer ce théorème (ce sera fait dans le cours de Quantique), mais nous allons vérifier cette propriété sur des solutions numériques de l'équation de Schrödinger, c'est à dire vérifier que

$$< x > (t) = x_{\rm class}(t) \ . \tag{4.128}$$

On choisit un système d'unités avec $\hbar = 1$ et la masse de la particule $m = 1/2$. On considère un domaine de simulation $x \in [-L/2 + L/2]$, avec $L = 256$, et on choisit un potentiel quadratique

$$V(x) = V_0 \left(\frac{x}{L/2} \right)^2 \tag{4.129}$$

avec un V_0 donné. V_0 n'est autre que la valeur du potentiel aux bords du domaine de simulation. On a

$$\omega_0^2 = \frac{8V_0}{mL^2} \ . \tag{4.130}$$

On place une particule dans un état initial de la forme (4.116), avec $n = 32$, ($k_0 = 0.7854$), $x_0 = 0$, $\sigma = 6.4$, ($< \Delta x > (0) = 4.5255$). On choisit le coefficient V_0 du potentiel de telle sorte qu'il soit égal à 4 fois l'énergie $E_0 = \hbar^2 k_0^2 / 2m$. Ce choix signifie, dans la limite classique, que l'on place une particule au minimum du potentiel, avec une vitesse initiale telle que son énergie cinétique initiale est 1/4 du potentiel aux bords du domaine de simulation. On discrétise avec $n_x = 512$ intervalles ($\Delta x = 0.5$) et $\Delta t = 0.5$. On utilise le schéma semi-implicite, Eq.(4.99).

La FIG. 4.21 montre l'évolution spatio-temporelle du module et de la partie réelle de la fonction d'onde. Les évolutions temporelles de la position moyenne, $< x > (t)$, et de l'incertitude sur la position, $< \Delta x > (t)$, sont affichées à la FIG. 4.22. On a représenté, en traitillés, la solution pour le mouvement classique $x_{\rm class}(t)$. Le mouvement de la posiiton moyenne $< x > (t)$ est bien une oscillation sinusoïdale. La différence avec la solution classique est une fréquence un peu plus basse. Cette différence est due aux erreurs de discrétisation (Δx et Δt finis). On peut montrer que la solution numérique pour $< x > (t)$ tend bien vers la solution classique $x_{\rm class}(t)$ dans la limite $\Delta x \to 0$ et $\Delta t \to 0$. Ainsi, les résultats numériques sont en bon accord avec la théorie.

L'évolution de l'incertitude $< \Delta x > (t)$ montre qu'elle ne croît pas indéfiniment au cours du temps, contrairement au cas de la particule libre. L'incertitude oscille autour d'une valeur moyenne. La théorie, qui sera faite au cours de mécanique quantique [4], montre qu'il existe des états de la particule dans un potentiel harmonique tels que leur incertitude est constante au cours du temps. On appelle ces états *quasi-classiques*, ou états *"cohérents"*. Ils sont constitués de paquets d'ondes de forme gaussienne, avec une incertitude sur la position donnée par

$$< \Delta x >_{\rm quasi-class} = \sqrt{\frac{\hbar}{2m\omega_0}} \tag{4.131}$$

4. voir p.ex. Cohen-Tannoudji, Mécanique Quantique I, complément G_V, p.560-575.

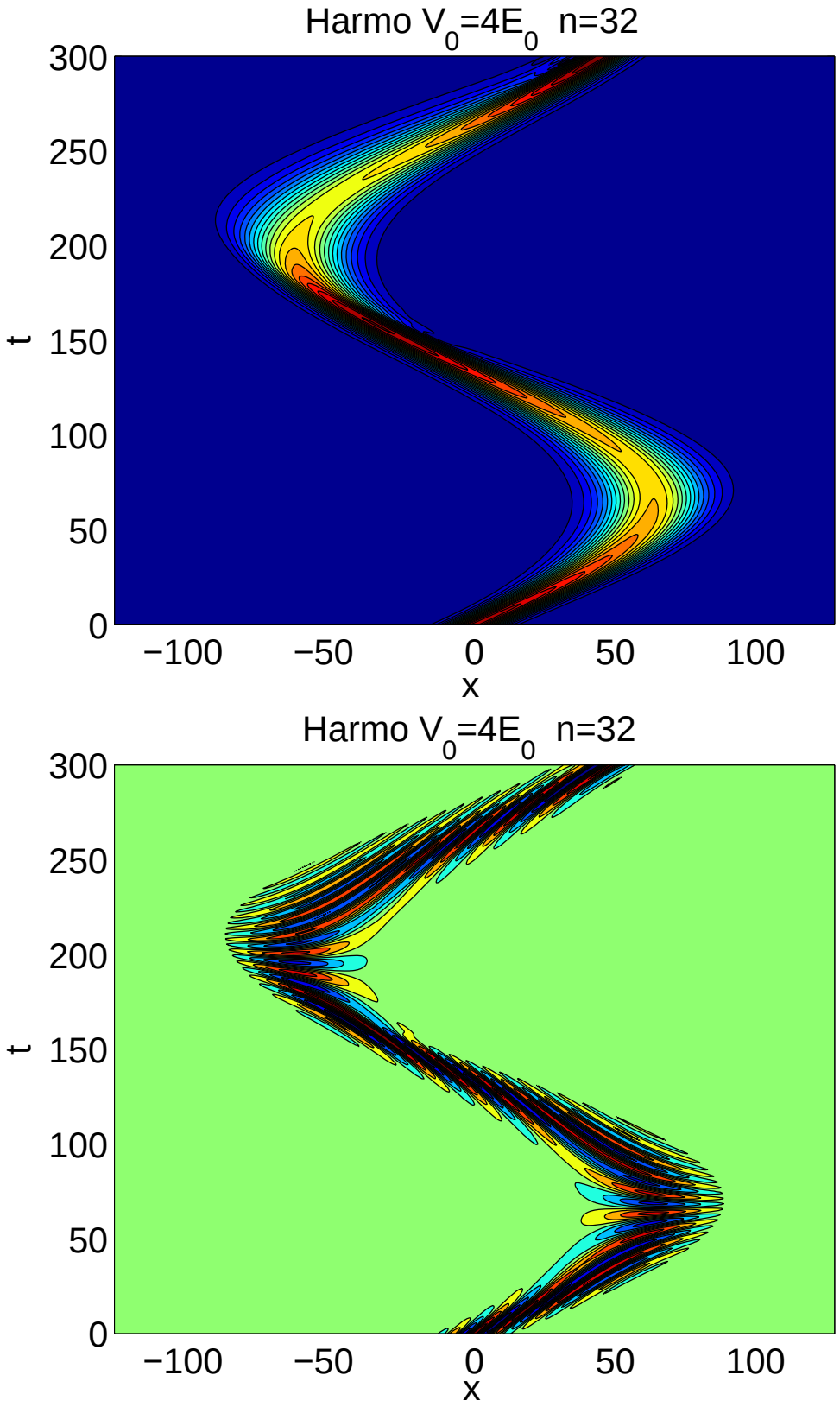


FIGURE 4.21 – *Particule dans un potentiel harmonique. En haut,* $|\psi(x,t)|$*. En bas,* $Re(\psi(x,t))$*.*

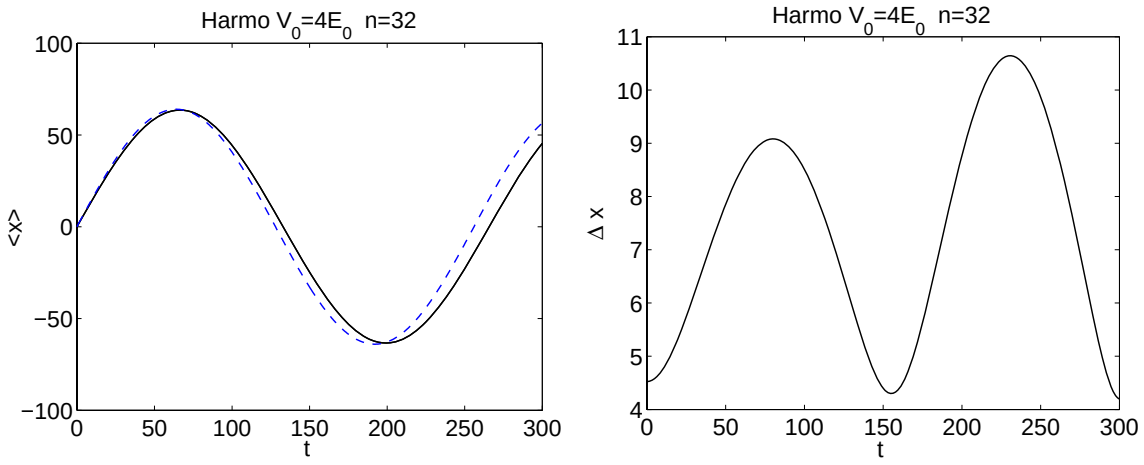


FIGURE 4.22 – *Particule dans un potentiel harmonique (même simulation que la FIG. 4.21). A gauche : position moyenne $< x > (t)$, avec en traitillés la solution de la physique classique. A droite, incertitude sur la position $< \Delta x > (t)$.*

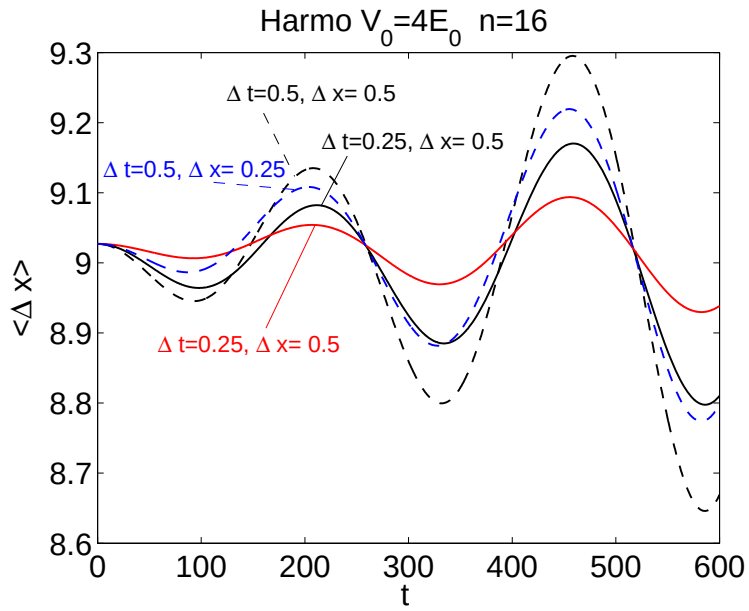


FIGURE 4.23 – *Particule dans un potentiel harmonique dans un état dit quasi-classique, simulée avec diverses résolutions spatiales et temporelles. L'incertitude sur la position $< \Delta x > (t)$, qui devrait selon la théorie être constante, présente des oscillations. L'amplitude de ces oscillations diminue avec des Δx et/ou des Δt plus petits.*

Les simulations numériques utilisant le schéma semi-implicite (4.99) montrent que si on choisit une largeur de la gaussienne initiale conformément à l'expression ci-dessus, l'incertitude sur la position $< \Delta x > (t)$ reste approximativement constante. Il subsiste néanmoins une oscillation de $< \Delta x > (t)$, qui n'est pas physique, mais d'origine purement numérique. On montre un exemple à la FIG. 4.23, avec les paramètres : $n = 16$, $L = 256$, $V_0 = 4E_0$. Quatre simulations sont effectuées, avec $\Delta x = 0.5,\ 0.25$ et $\Delta t = 0.5,\ 0.25$. L'amplitude des oscillations de $< \Delta x > (t)$ décroît lorsque l'on diminue Δt et/ou Δx.

En conclusion, les résultats numériques basés sur le schéma semi-implicite, Eq.(4.99), permettent de mettre en évidence le comportement parfois inattendu, parfois contraire à la physique classique, des particules. Nous avons aussi illustré, pour l'oscillateur harmonique, à quel point les prédictions de la mécanique quantique sont, dans un certain sens seulement, analogues à celles de la mécanique classique : les valeurs moyennes se comportent comme des particules classiques. Ces résultats numériques sont en bon accord avec les calculs analytiques, qui seront faits au cours de Physique et de Mécanique Quantique, pour lesquels ils peuvent servir d'illsutrations.

4.3.5 Etats stationnaires ou états propres de la particule

Soit une particule dans un potentiel $V(\vec{x})$. On aimerait trouver une solution $\psi(x,t)$ de l'équation de Schrödinger, Eq.(4.79), qui donne une énergie bien déterminée de la particule.

Par la relation de de Broglie, $E = \hbar\omega$, Eq.(4.101), dire que l'énergie E est donnée implique que la fréquence ω est donnée. On cherchera donc des solutions de l'Eq. de Schrödinger de la forme :

$$\boxed{\psi(\vec{x},t) = \Psi(\vec{x}) \exp(-i\omega t)} \tag{4.132}$$

Ces solutions sont appelées **états stationnaires** : en effet, la densité de probabilité, $|\psi|^2$, est une fonction de l'endroit $(\vec{x})$ mais pas du temps. La probabilité est stationnaire, dans le même sens que l'intensité (moyennée sur une période) d'une onde stationnaire ne dépend pas du temps.

Introduisant cet Ansatz dans l'Eq. de Schrödinger (4.79), on a

$$\boxed{-\frac{\hbar^2}{2m}\nabla^2\Psi + V(\vec{x})\Psi = E\Psi}\,. \tag{4.133}$$

C'est **l'équation de Schrödinger stationnaire, ou "équation de Schrödinger indépendante du temps"**. Dans la limite classique, elle exprime simplement le principe de conservation de l'énergie mécanique ($p^2/2m + V = E_{mec}$).

Avec la définition de l'opérateur Hamiltonien

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(\vec{x}) \ , \tag{4.134}$$

l'équation de Schrödinger stationnaire s'écrit

$$\boxed{H(\Psi) = E\Psi} \ . \tag{4.135}$$

Cette équation indique que **les énergies possibles d'une particule dans un potentiel $V(\vec{x})$ sont les valeurs propres de l'Hamiltonien**. Les états d'énergie donnée correspondants à ces valeurs propres sont les fonctions propres de cet Hamiltonien. On les appelle donc états propres.

Trouver les états propres et les énergies possibles d'une particule revient donc à "diagonaliser" l'Hamiltonien du système.

Méthodes numériques

Il existe plusieurs méthodes pour trouver des états et énergies propres. On peut par exemple utiliser les outils déjà développés dans les sections précédentes, à savoir la méthode des différences finies ou celle des éléments finis, appliquée à la discrétisation spatiale de l'opérateur Hamiltonien. La nuance est que les fonctions recherchées sont à valeurs complexes, et non plus réelles.

Cette opération de discrétisation numérique conduit à approximer l'opérateur Hamiltonien, qui est différentiel, par un opérateur algébrique. les inconnues étant par exemple les valeurs de Ψ aux points du réseau $x_j, j = 1..N$. La problème se réduit donc à un problème matriciel :

$$\sum_j A_{ij}\Psi_j = E\Psi_i \tag{4.136}$$

où Ψ_j est le vecteur des inconnues $\Psi(x_j)$.

Trouver des approximations numériques des états et des énergies propres d'une particule revient donc à **diagonaliser la matrice A, autrement dit à trouver ses vecteurs propres et valeurs propres**.

Par exemple, le schéma de différences finies (A.7) appliqué à Schrödinger stationnaire 1-D conduit à la matrice

$$A = \text{tridiag}\,(C \qquad -2C + V(x_i) \qquad C) \quad , \tag{4.137}$$

avec

$$C = -\frac{\hbar^2}{2m(\Delta x)^2} \ . \tag{4.138}$$

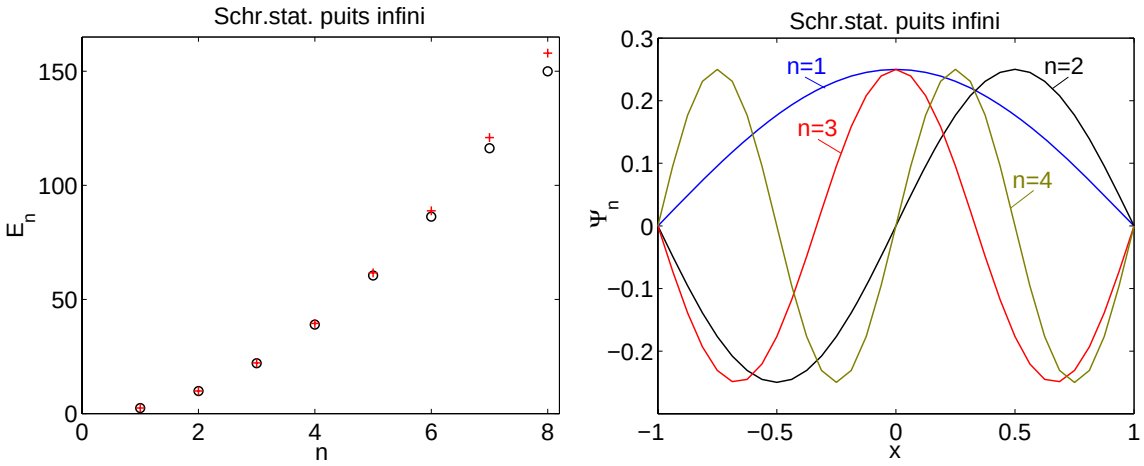


FIGURE 4.24 – *Spectre des énergies propres (à gauche) et les 4 premiers états propres (à droite) pour une particule confinée dans un puits de potentiel de profondeur infinie.*

Particule dans une boîte

La modélisation la plus simple d'une particule confinée dans une boîte est de dire que la particule n'a aucune chance de se trouver en dehors de la boîte. A l'intérieur de la boîte, elle ne subit aucune force, c'est-à-dire que le potentiel est constant. Le problème à résoudre est donc simplement Schrödinger stationnaire avec $V(x) = 0$ et des conditions aux bords de type Dirichlet :

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}\Psi}{\mathrm{d}x^2} = E\Psi \ , \quad \Psi(0) = 0 \ , \quad \Psi(L) = 0 \ , \tag{4.139}$$

où L est la taille de la boîte. Il est facile de trouver les solutions analytiques : on trouve les fonctions propres et valeurs propres

$$\Psi_n(x) = \sin\left(n\frac{\pi x}{L}\right) \tag{4.140}$$

$$E_n = \frac{\hbar^2}{2m}\frac{n^2\pi^2}{L^2} \tag{4.141}$$

La FIG. 4.24 montre les résultats numériques avec les différences finies et $n_x = 32$ intervalles, pour une particule de masse $m = 1/2$, confinée dans une boîte de taille $L = 2$. Comme précédemment, les unités avec $\hbar = 1$ ont été utilisées. La différence entre la solution analytique (croix) et la solution numérique (cercles) est due à la discrétisation. On peut montrer (exercice) que cette erreur diminue avec le nombre de points de maillage.

Une modélisation un peu plus réaliste considère une particule confinée par un "puits" de potentiel de profondeur finie. On montre à la FIG. 4.25 le spectre et les états propres des états d'énergies les plus basses. On a pris un potentiel fe forme carrée, $V(x) = -100$ entre $x_a = -0.5$ et $x_b = +0.5$, zéro ailleurs. Le domaine de calcul a été pris entre $x = -2$ et $x = +2$. Il est en effet nécessaire de prendre un domaine plus large que la boîte :

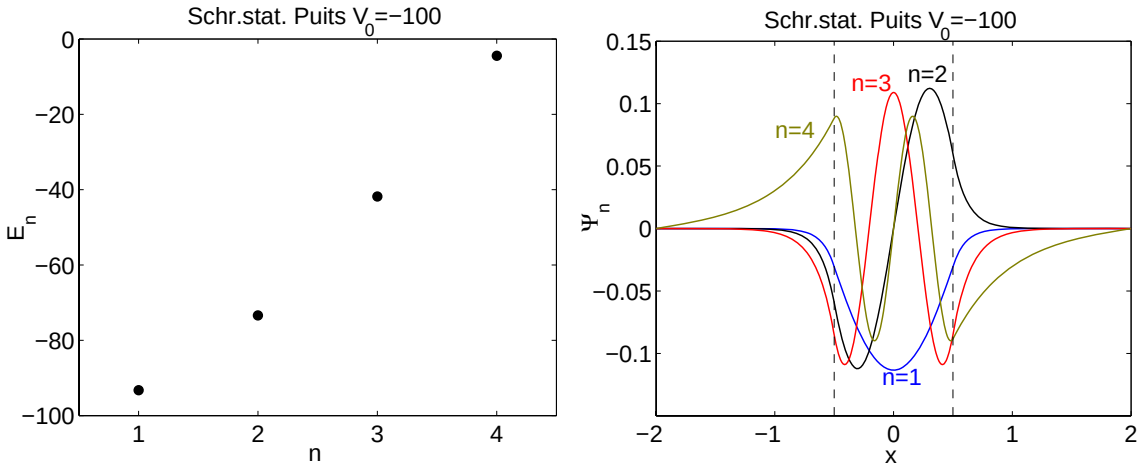


FIGURE 4.25 – *Spectre des énergies propres (à gauche) et les 4 premiers états propres (à droite) pour une particule confinée dans un puits de potentiel de profondeur finie,* $V_0 = -100$*, entre* $x = -0.5$ *et* $x = +0.5$ *(lignes traitillées).*

les résultats montrent que la particule a une *probabilité non nulle de se trouver quelque peu en dehors de la boîte*, même si son énergie est plus petite que zéro ! L'autre résultat important, que l'on aurait pu déjà constater sur le cas précédent, est que l'état d'énergie le plus bas, appelé état fondamental, est *d'énergie plus élevée que la valeur minimum du potentiel.*

Particule dans un potentiel périodique : physique du solide

L'état solide est caractérisé, au niveau microscopique, par un arrangement régulier, périodique, d'atomes. La cohésion du solide est assurée par certains électrons du système, alors que d'autres électrons participent éventuellement à la condution électrique et de chaleur.

La structure des énergies possibles des électrons dans un solide est étonnante : elle présente des bandes séparées par des "bandes interdites" (en anglais : **gap**), où *aucun* électron ne se trouve. Nous allons essayer de comprendre pourquoi grâce à l'approche numérique.

On modélise un électron dans un solide par une particule dans un potentiel périodique. Ce potentiel représente l'effet des noyaux atomiques et des autres électrons. On néglige l'interaction entre électrons. On prend un potentiel

$$V(x) = V_0 \sin\left(n_{\text{pot}} \frac{2\pi x}{L}\right) \tag{4.142}$$

où L est la taille du solide. Dans la réalité L est beaucoup plus grand que la taille inter-atomique. Il serait irréaliste (et irréalisable) de simuler tout un solide de taille

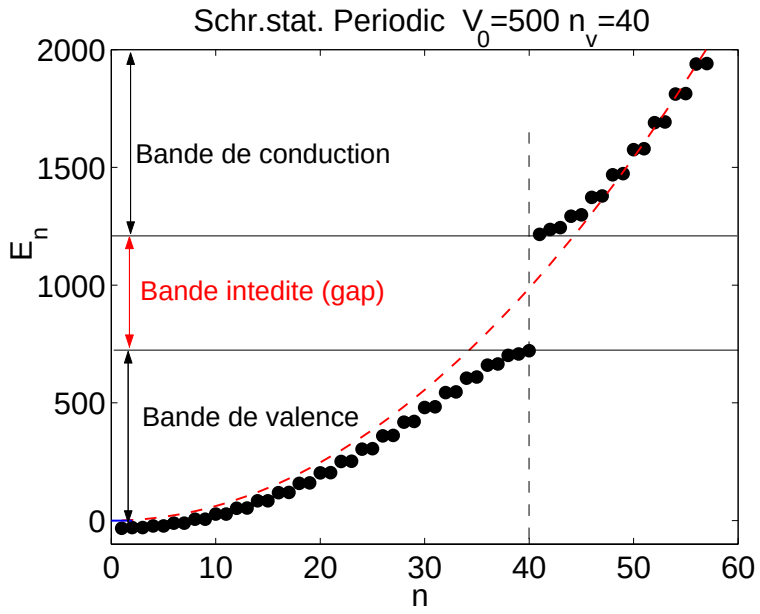


FIGURE 4.26 – *Spectre des énergies propres pour une particule dans un potentiel périodique sinusoïdal d'amplitude*$V_0 = 500$. *Le domaine de simulation est périodique. Sa taille est de* 40 *périodes du potentiel (donc une taille de 40 couches interatomiques). La ligne rouge en traitillés indique le spectre en l'absence de potentiel périodique (particule libre).*

macroscopique. Prenons les paramètres suivants : $L = 4$, $n_{\text{pot}} = 40$ (ce qui veut dire que l'on simule une tranche de 40 atomes). On prendra, de plus, un système périodique de période L. Cela nécessite une petite modification de l'algorithme **(exercice)**.

Avec $V_0 = 500$ et un maillage de $n_x = 512$ intervalles, on obtient les résultats de la FIG. 4.26. Par comparaison, on montre en traitillés le spectre d'une particule libre, autrement dit le cas $V_0 = 0$. Il est clair que l'effet de la perturbation périodique du potentiel est de créer une bande d'énergie interdite. La taille de cette bande interdite est d'environ 490, soit du même ordre que l'amplitude V_0 de la perturbation sinusoïdale du potentiel. De plus, c'est pour le mode numéro $n = 40$ que le saut en énergie a lieu. L'analyse de la fonction d'onde correspondante montre qu'elle a une **longueur d'onde double de la distance interatomique**. On constate que cela correspond à la condition de Bragg : soit une onde incidente de longueur d'onde λ sur un réseau périodique de période spatiale d ; les ondes réfléchies par les couches successives seront en phase (interférence constructive) si $2d \sin\theta = N\lambda$, où N est un nombre entier strictement positif qui est l'ordre de l'interférence. Ici, $\theta = 0$, car nous sommes en 1-D. Pour l'ordre d'interférence le plus bas ($N = 1$), on a bien $\lambda = 2d$. L'onde stationnaire résultante a un module dont les maxima coïncident soit avec les maxima, soit avec les minima du potentiel, résultant en une énergie soit plus élevée, soit plus basse ($\pm V_0/2$) que l'énergie de la particule libre correspondante.

Paquet d'onde dans un potentiel périodique

Il est intéressant de revenir au problème dépendant du temps. Pour le même potentiel périodique qu'au paragraphe précédent, on résout cette fois l'équation de Schrödinger, Eq.(4.79), avec la méthodes de différences finies et le schéma semi-implicite, Eq.(4.99). On utilise 512 points de discrétisation spatiale et un pas temporel $\Delta t = (\hbar/E_0)/8$, où E_0 est l'énergie moyenne du paquet d'onde $E_0 \approx \hbar^2 k_0^2/2m$, où k_0 est la valeur centrale du nombre d'onde. Ceci correspond à $8 \times 2\pi$ pas temporels par période d'oscillation.

Comme condition initiale, nous prenons un paquet d'onde Gaussien, Eq.(4.116), avec une position moyenne $x_0 = -0.6$, et $\sigma = 0.4$. Le nombre d'onde moyen est choisi pour trois cas différents, $n = 14$, $n = 20$ et $n = 26$. Le premier cas correspond à une particule dont l'énergie moyenne est dans la bande de valence. Le deuxième cas à une particule dont lénergie serait dans la bande interdite. Le troisième cas correspond à une particule d'énergie moyenne dans la bande de conduction.

Les résultats de ces trois simulations sont représentés aux FIGS. 4.27-4.28. Les particules dans la bande de valence ($n = 14$) et dans la bande de conduction ($n = 26$) se propagent bien à travers le système. Il y a une certaine modulation due au potentiel périodique, mais la position moyenne est en mouvement (presque) uniforme, (presque) comme si la particule était libre. Pour la particule dans la bande interdite ($n = 20$), les choses se passent tout différemment. La fonction d'onde ne propage plus ! La position moyenne de la particule est pratiquement immobile. On peut montrer que ce comportement est conforme à la théorie : au voisinage du gap, la vitesse de groupe tend vers zéro.

Pour en savoir plus : Bibliographie, Refs. [17]-[19].

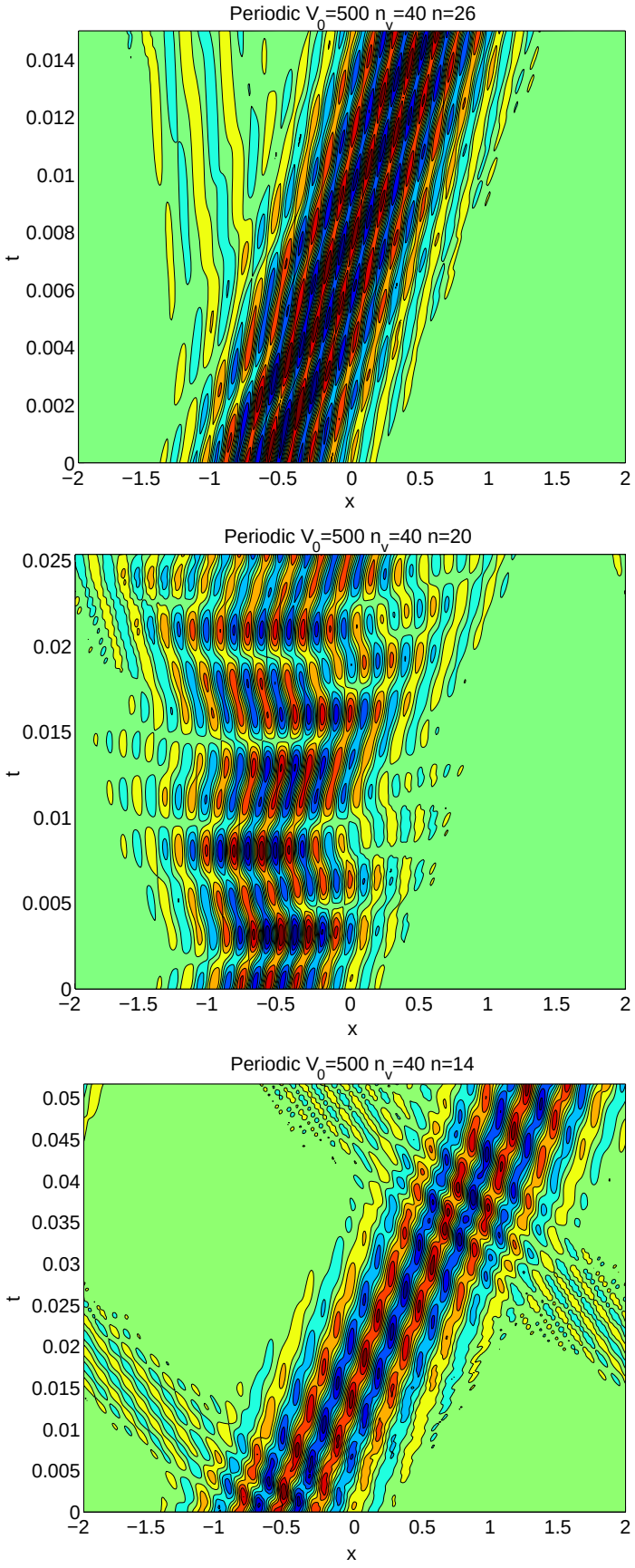


FIGURE 4.27 – *Particule dans un potentiel périodique, dans la bande de conduction ($n = 26$, haut), dans la bande interdite ($n = 20$, milieu) et dans la bande de valence ($n = 14$, bas). Contours de $Re(\psi(x,t))$.*

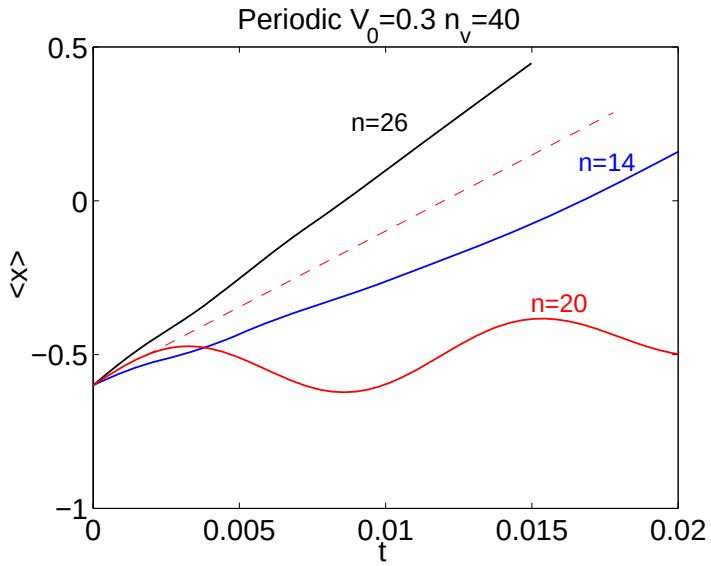


FIGURE 4.28 – *Positions moyennes de trois particules dans un potentiel périodique, respectivement dans la bande de conduction ($n = 26$), dans la bande interdite ($n = 20$) et dans la bande de valence ($n = 14$), correspondant aux trois simulations de la FIG. 4.27. La ligne en traitillés représente le mouvement qu'aurait une particule libre avec $n = 20$.*

Chapitre 5

Méthodes statistiques

Dans ce chapitre, nous étendons notre analyse à des systèmes contenant un grand nombre de particules en interaction. Dans la partie du cours consacrée au phénomène de diffusion, nous avons mis en évidence le caractère aléatoire, au niveau microscopique, du processus. La grande simplification que nous avons faite alors est de supposer l'indépendance totale des collisions individuelles. En d'autres termes, nous avons négligé les interactions.

Nous verrons que ces interactions jouent un rôle essentiel dans le comportement de ces systèmes, notemment dans les phénomènes de *transition de phase*, tels la solidification, la liquéfaction, la condensation et l'évaporation. Dan les matériaux solides il existe aussi de nombreuses transitions de phases, par exemple l'apparition du ferromagnétisme.

Dans tous ces phénomènes, le concept de *température* joue un rôle central. C'est pourquoi les concepts de thermodynamique et de physique statistique sont invoqués pour les décrire.

L'approche numérique adoptée dans ce chapitre s'inspire du caractère aléatoire de l'état microscopique : on parle de *méthodes de Monte Carlo*, que nous avons déja rencontrées pour des problèmes simples, comme la désintégration ou la diffusion.

5.1 Modèle d'Ising

Le ferromagnétisme apparaît dans certains matériaux à cause de l'interaction entre les moments magnétiques des atomes. Dans ces matériaux, l'énergie d'interaction est minimisée si ceux-ci sont alignés dans la même direction.

Les atomes ont la propriété fondamentale d'avoir un moment cinétique intrinsèque, appelé le *spin*, auquel est associé un moment magnétique. L'existence du spin est un effet purement quantique et ne sera pas discuté ici. Nous ne retiendrons que sa propriété d'être quantifié, c'est-à-dire que sa valeur projetée selon un axe donné ne peut donner que des valeurs discrètes. Nous prendrons ces valeurs, pour simplifier, comme

$$s = \pm 1 \tag{5.1}$$

Dans une simplification supplémentaire, nous ne considèrerons l'énergie d'interaction qu'entre les plus proches voisins. De plus, on supposera les atomes disposés régulièrement sur un réseau. L'énergie du système s'écrit donc

$$E = -J \sum_{<ij>} s_i s_j \tag{5.2}$$

où $< ij >$ désigne une paire d'atomes voisins, et J est appelée constante de couplage. Cette description s'appelle le *modèle d'Ising*, dans son expression la plus simple.

Si $J > 0$, cela signifie que l'énergie est minimisée quand tous les spins sont alignés. C'est la situation pour une substance ferromagnétique à très basse température. Thermodynamiquement parlant, c'est un état d'entropie S minimale.

A plus haute température, l'agitation thermique va contribuer a rompre l'alignement parfait des spins, et l'amantation (somme des moments magnétiques) va diminuer. Il existe une température critique T_c au delà de laquelle l'aimantation est nulle (en moyenne statistique sur un grand nombre $N \to \infty$ d'atomes). A $T >> T_c$, l'état du système alors maximise son entropie S alors que son énergie interne U est à peu près nulle (en tous cas, elle n'est pas minimale) [1]. Une conséquence du deuxième principe est que, à toute température, le système à l'équilibre *minimise son énergie libre* $F = U - TS$.

Si on rajoute un champ magnétisant H au système, l'énergie devient :

$$E = -J \sum_{<ij>} s_i s_j - \mu H \sum_i s_i \ , \tag{5.3}$$

où μ est le moment magnétique associé à chaque spin. Le champ H tend à aligner les spins parallélement à $\vec{H}$, puisque cela contribue à diminuer l'énergie.

5.1.1 Statistique de Boltzmann

L'état d'un système de N spins est donc caractérisé par une séquence de $+$ et de $-$. Un système à 4 spins, par exemple, peut être dans l'état $(++--)$, ou dans l'état $(+--+)$,

1. Dans la limite thermodynamique, l'énergie interne U est reliée à la moyenne statistique de l'énergie, $< E >$, sur tous les états microscopiques possibles.

ou dans l'état $(--+-)$, etc. A chaque état, que nous numéroterons avec la lettre α, correpond une énergie E_α.

Un résultat fondamental de la physique statistique est que la **pour une température T du système, la probabilité P_α de trouver le système dans l'état numéro α est donnée par :**

$$\boxed{P_\alpha = C\, e^{-E_\alpha/k_B T}}\,, \tag{5.4}$$

où $k_B = 1.38066 \times 10^{-23}$J/K est la constante de Boltzmann et C est une constante de normalisation telle que

$$\sum_\alpha P_\alpha = 1\,, \tag{5.5}$$

la somme portant sur tous les états microscopiques possibles du système.

Dans notre cas, on peut ainsi calculer l'aimantation du système (macroscopique) en fonction des aimantations M_α de chaque état microscopique et de leurs probabilités respectives P_α par leur moyenne statistique

$$M = \sum_\alpha M_\alpha P_\alpha\,. \tag{5.6}$$

Si on ne prend pas de précaution pour effectuer cette moyenne, on trouvera toujours la valeur *nulle* pour un système de taille finie, même pour un état ferromagnétique où tous les spins sont alignés : en effet, la probabilité de trouver tous les spins à $+1$ est égale à celle de trouver tous les spins à -1. On doit effectuer la limite en rajoutant un champ H extérieur au système de taille finie L, cacluler une moyenne $M(L,H)$, prendre la limite d'un système de taille infinie, et ensuite faire tendre H vers zéro, soit par valeurs positives, soit par valeurs négatives :

$$M_+ = \lim_{H\to 0+}\Big(\lim_{L\to\infty} M(L,H)\Big) \tag{5.7}$$

$$M_- = \lim_{H\to 0-}\Big(\lim_{L\to\infty} M(L,H)\Big) \tag{5.8}$$

Pour un système de N spins, le nombre d'états microscopiques possibles est 2^N, et il devient vite prohibitif de calculer tous ces états possibles. L'approche passe par une simplification majeure, expliquée dans la section suivante.

5.1.2 Théorie du champ moyen

Considérons un système constitué d'un seul spin s_i, dont les valeurs possibles sont ± 1, plongé dant un champ magnétisant extérieur H. La statistique de Boltzmann, Eq.(5.4), et l'expression de l'énergie du système, Eq.(5.3), impliquent que les probabilités de trouver le système dans chacun des deux états possibles $(\pm)$ sont données par

$$P_+ = Ce^{+\mu H/k_B T} \tag{5.9}$$

$$P_- = Ce^{-\mu H/k_B T} \tag{5.10}$$

avec la constante de normalisation

$$C = \left(e^{+\mu H/k_B T} + e^{-\mu H/k_B T}\right)^{-1} \ . \tag{5.11}$$

La moyenne statistique du spin est donc

$$< s_i >= \sum_{s_i=\pm 1} s_i P_\pm = P_+ - P_- = \tanh\left(\frac{\mu H}{k_B T}\ .\right) \tag{5.12}$$

Considérons maintenant ce spin s_i étant l'un parmi un système de N spins. L'approximation de la théorie du champ moyen consiste à faire l'hypothèse que l'interaction de ce spin avec ses voisins est équivalente à la présence d'un champ magnétisant effectif H_{eff}. H_{eff} représente donc le champ moyen crée par les *autres* spins à l'endroit du spin s_i. On a donc

$$< s_i >= \tanh\left(\frac{\mu H_{\text{eff}}}{k_B T}\right) \ . \tag{5.13}$$

Nous pouvons écrire l'expression de l'énergie du système, Eq. (5.3), comme

$$E = -\left(J \sum_{<ij>} s_j\right) s_i - \mu H s_i \ . \tag{5.14}$$

Dans cette dernière expression H est un champ magnétisant *extérieur* au système de N spins, alors que le terme entre parenthèses représente l'effet des autres spins $s_{j,\, j\neq i}$ du système. L'approximation du champ moyen consiste à remplacer ce terme entre parenthèses par μH_{eff},

$$\left(J \sum_{<ij>} s_j\right) \approx \mu H_{\text{eff}} \ , \tag{5.15}$$

et à supposer que les spins individuels s_j peuvent être remplacés par leur valeur moyenne $< s_j >$. Comme tous les spins sont des particules identiques, leur valeur moyenne est identique, et on peut donc omettre l'indice j : $< s_j >=< s >$, $\forall j$. On a donc :

$$H_{\text{eff}} \approx \frac{J}{\mu} \sum_{<ij>} < s > \ . \tag{5.16}$$

Dans la somme, $< ij >$ signifie une somme sur tous les plus proches voisins. Si n est le nombre de plus proches voisins

$$H_{\text{eff}} \approx \frac{nJ}{\mu} < s > \ . \tag{5.17}$$

On a donc, de l'Eq.(5.13),

$$< s >= \tanh\left(\frac{nJ < s >}{k_B T}\right) \ . \tag{5.18}$$

Cette dernière expression est une équation non triviale, non algébrique pour $< s >$. On peut la résoudre numériquement (exercice) par exemple avec la méthode de Newton -

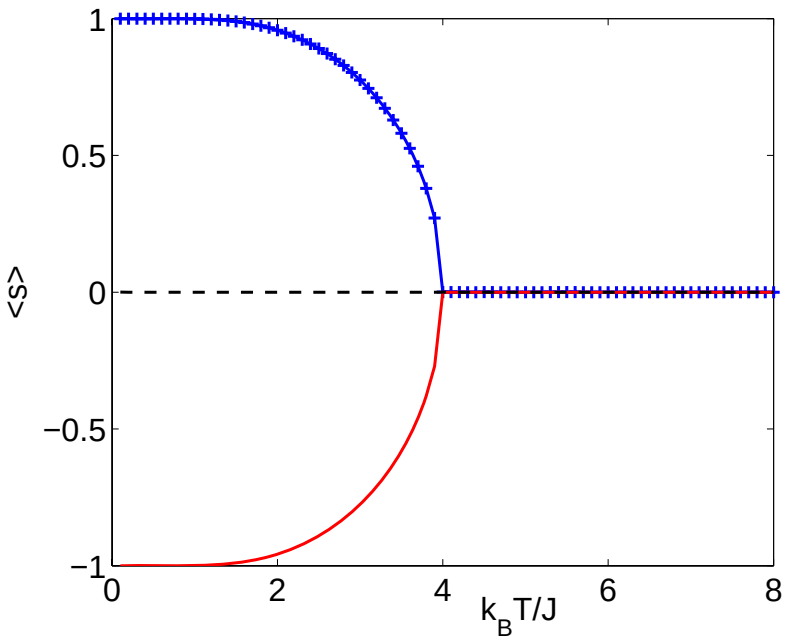


FIGURE 5.1 – *Valeur moyenne du spin en fonction de la température prédite par la théorie du champ moyen appliquée au modèle d'Ising 2-D à $n = 4$ plus proches voisins.*

Raphson. Le résultat est illustré à la FIG.5.1, en fonction de la température. On s'aperçoit de l'existence d'une *température critique* T_c, au voisinage de laquelle la valeur moyenne du spin, et donc l'aimantation, varie brutalement. Pour $T > T_c$, il n'y a que la solution $< s >= 0$, et l'aimantation est nulle : le système est paramagnétique. Pour $T < T_c$, il y a 3 solutions ($< s >= 0$, et deux solutions non nulles de même valeur absolue, l'une positive et l'autre négative). La solution $< s >= 0$ est instable car elle correspond à un maximum local de l'énergie libre du système. Les deux solutions symétriques, $< s > > 0$ et $< s > < 0$, représentent l'aimantation permanente du système, qui est donc dans l'état ferromagnétique.

Ce qui se passe au voisinage de $T = T_c$ est un exemple de *transition de phase*. L'aimantation, proportionnelle à $< s >$, joue le rôle de paramètre d'ordre. A basse température, l'aimantation moyenne est non nulle, ce qui signifie une tendance à aligner les spins dans la même direction, et on a un système dans un état ordonné. A haute température, l'aimantation moyenne est nulle, ce qui signifie que les spins perdent leur alignement mutuel, et le système est dans un état désordonné.

Pour le modèle d'Ising 2-D à n plus proches voisins, on peut montrer que la température de transition est $T_c = nJ/k_B$. Au voisinage de $T = T_c$, et pour $T < T_c$, la valeur du spin moyen est

$$< s >= \sqrt{\frac{3}{T}\left(\frac{T}{T_c}\right)^3}(T_c - T)^{1/2} \, . \tag{5.19}$$

L'exposant 1/2 est appelé exposant critique de la transition. En fait, l'approximation du

champ moyen est incorrecte de ce point de vue : la solution exacte du modèle d'Ising donne un exposant critique 1/8.

5.1.3 Monte Carlo, algorithme de Metropolis

La méthode de Monte Carlo utilise une approche dite *stochastique*, c'est-à-dire résultant de processus aléatoires. Dans notre modèle d'Ising de N spins en interaction, à la température T, l'algorithme dit de Metropolis consiste en :

1. Initialiser un état microscopique quelconque, donc une séquence $(++-+--+...)$.
2. Pour chaque spin s_i, calculer l'énergie nécessaire à faire basculer le spin, ΔE. Dans notre modèle, cette énergie ne dépend que des plus proches voisins, voir Eq.(5.3).
3. Si $\Delta E < 0$, basculer le spin
4. Si $\Delta E > 0$, générer un nombre aléatoire r selon une distribution de probablité uniforme entre 0 et 1.
5. Si $r \leq \exp(-\Delta E/k_B T)$, basculer le spin. Autrement, le laisser inchangé.
6. Une fois tous les spins du systèmes traités de cete manière, calculer la nouvelle valeur de l'énergie et la nouvelle valeur du spin moyen.
7. Répéter les étapes $2-6$ un nombre suffisant de fois.

On peut comprendre qualitativement comment l'algorithme est capable de représenter la physique. Si on basculait les spins chaque fois que $\Delta E < 0$ et jamais si $\Delta E > 0$, le système évoluerait vers un état d'énergie minimale, où tous les spins sont alignés. C'est ce qui se passe dans la limite $T \to 0$. La température finie introduit la possibilité pour le système d'évoluer vers un état d'énergie plus élevée. A basse T, le facteur $\exp(-\Delta E/k_B T)$ est proche de 0, et la probabilité qu'un spin bascule est petite : le système aura tendance à rester dans une phase ferromagnétique. A mesure que T augmente, le facteur $\exp(-\Delta E/k_B T)$ augmente, et avec lui la probabilité de basculement : le système a alors de plus en plus tendance à rompre l'alignement des spins, et donc tend vers un état paramagnétique.

La statistique de Boltzmann (5.4) implique que le rapport de la probabilité P_1 d'avoir un spin basculé par rapport à la probabilité P_2 d'avoir un spin non basculé est

$$\frac{P_1}{P_2} = e^{-\frac{\Delta E}{k_B T}} \tag{5.20}$$

L'alorithme de Metropolis conduit ainsi à une situation dans laquelle les probabilités relatives de trouver des états microscopiques différents sont données correctement selon la statistique de Boltzmann.

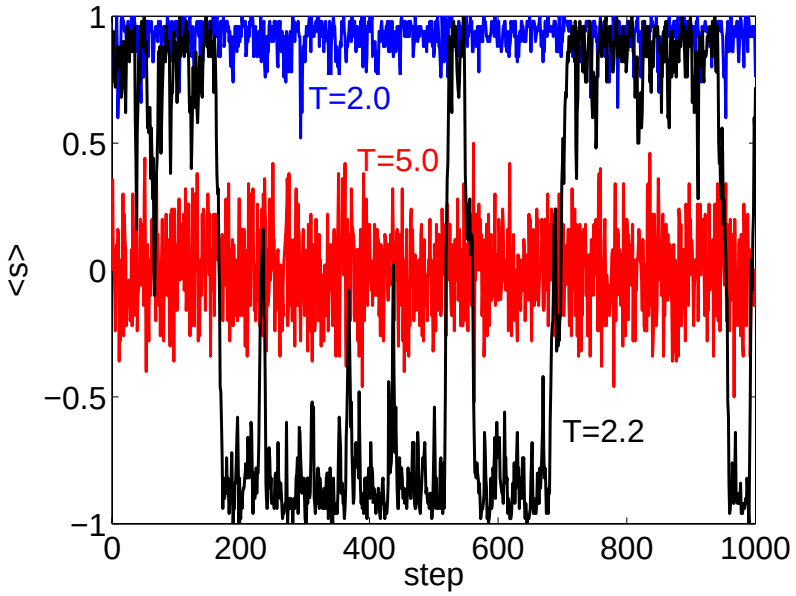


FIGURE 5.2 – *Spin moyen* $\bar{s} = \sum s_i/N$ *au cours des étapes de l'algorithme de Metropolis, pour différentes valeurs de la température. Modèle d'Ising 2-D à* $n = 4$ *plus proches voisins sur un réseau périodique de* 10×10 *spins.*

On montre à la FIG. 5.2 les résultats de 3 simulations pour un réseau périodique de 10×10 spins, pour 3 valeurs différentes de la température. Après chaque étape (numéro k) de l'algorithme de Metropolis (balayage complet de tous les spins du système), on calcule le spin moyen,

$$\bar{s}_{(k)} = \frac{1}{N} \sum_{i=1}^{N} s_{i(k)} \tag{5.21}$$

Pour $T = 2$, on observe que les spins sont toujours presque tous alignés. Pour $T = 2.2$, le spin moyen fluctue énormément, avec de brusques basculements d'une valeur positive à une valeur négative. Pour $T = 5$, le spin moyen fluctue autour d'une valeur nulle.

Le spin moyen $\bar{s}$ est une variable aléatoire dont on obtient un échantillon statistique $\{\bar{s}_{(k)}\}, k = 1..N_{\text{sweep}}$ avec la simulation numérique, N_{sweep} désignant le nombre d'étapes de l'algorithme de Metropolis. On obtient une estimation statistique de sa valeur moyenne et de sa variance par :

$$< \bar{s} > \;=\; \frac{1}{N_{\text{sweep}}} \sum_{k=1}^{N_{\text{sweep}}} \bar{s}_{(k)} \tag{5.22}$$

$$\sigma^2 \;=\; \frac{1}{N_{\text{sweep}}} \sum_{k=1}^{N_{\text{sweep}}} \bar{s}^2_{(k)} - \left(< \bar{s} >\right)^2 \tag{5.23}$$

Il faut faire attention que la séquence d'étapes de l'algorithme est généralement *corrélée* : l'état de l'étape $k+1$ dépend de celui de l'étape k. C'est notamment visible au voisinage

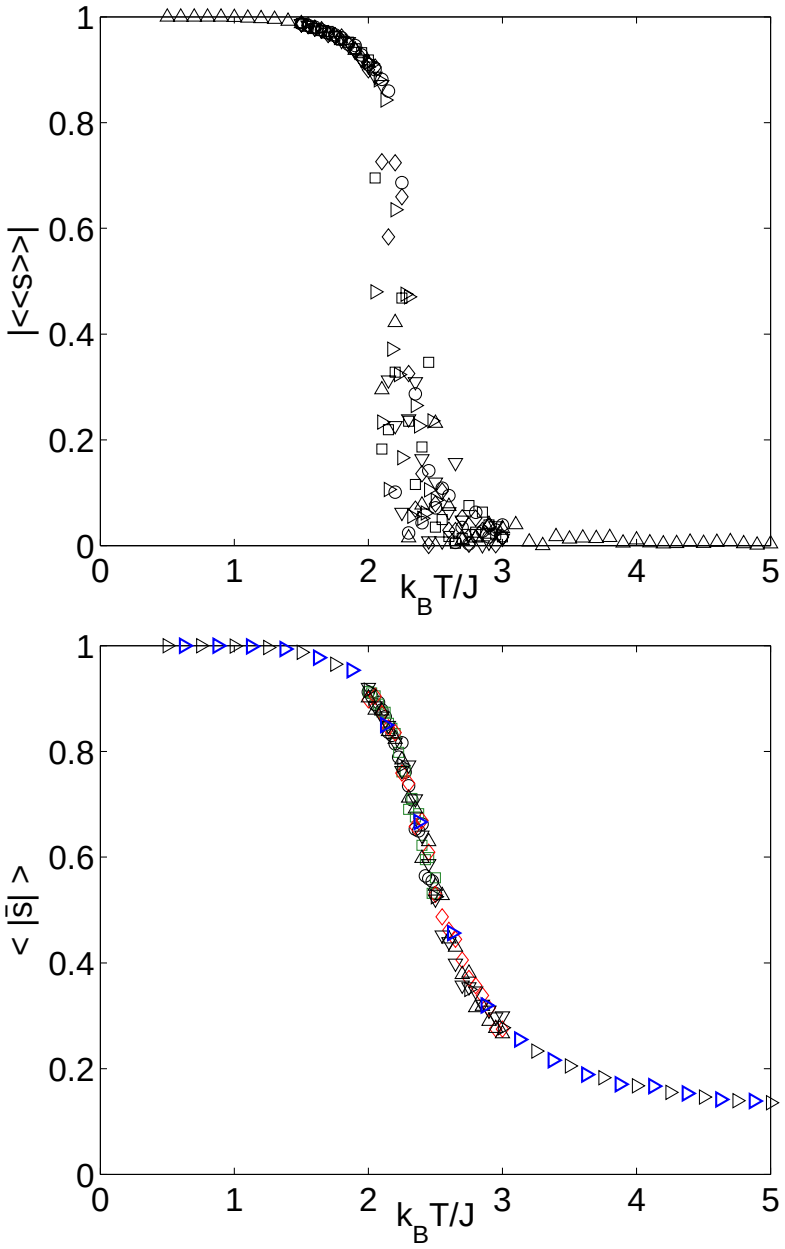


FIGURE 5.3 – *En haut : valeur absolue de la moyenne du spin moyen en fonction de la température, pour une série de simulations Metropolis. Modèle d'Ising 2-D à $n = 4$ plus proches voisins sur un réseau périodique de 10×10 spins. Pour chaque simulation, on a pris la moyenne du spin moyen sur 901 états microscopiques produits aux différantes étapes (balayages) de l'algorithme. En bas : idem, sauf que l'on a pris la moyenne de la valeur absolue du spin moyen.*

du point critique, par exemple $T = 2.2$ sur la FIG. 5.2, où le système met un nombre élevé d'étapes pour faire basculer le spin moyen. Pour s'assurer que les moyennes et variances ont un sens statistiquement correct, on doit faire *plusieurs* simulations indépendantes du point de vue des probabilités, c'est-à-dire un nombre N_{bin} de simulations complètes ayant chacune N_{sweep} étapes, à partir de conditions initiales différentes et non corrélées. On calcule ensuite la moyenne et la variance sur cet ensemble de N_{bin} simulations, l'écart-type nous donnant une estimation de la barre d'erreur du résultat. De plus, pour chaque simulation, on laisse un certain nombre détapes pour que le système "oublie" sa condition initiale. On ne prend les mesures des grandeurs physiques qu'après cette phase de la simulation. Dans ce qui suit, on a pris 100 étapes dans cette phase.

En effectuant une moyenne de la valeur absolue [2] du spin moyen $\bar{s}$ sur les étapes, $< |\bar{s}| >$, on obtient une quantité proportionnelle à l'aimantation du système. En effectuant plusieurs simulations à plusieurs températures, on obtient les résultats de la FIG. 5.3. Autour de $T \approx 2.3$, on remarque la chute abrupte de l'aimantation, indiquant une transition de phase.

Le modèle d'Ising peut être résolu analytiquement, donnant une température de transition $T_c = 2.27$, et un comportement au voisinage de cette température $< |\bar{s}| > \sim (Tc - T)^{\beta}$ avec un exposant critique $\beta = 1/8$. La simulation de Metropolis donne donc des résultats en bien meilleur accord avec la solution exacte que la solution obtenue avec l'approximation du champ moyen, qui donne, elle, $T_c = 4$ et $\beta = 1/2$, comparer les FIGS. 5.1 et 5.3, et voir l'Eq.(5.19).

Une mesure de la fluctuation de $\bar{s}$ est fournie par la variance σ^2 de cette quantité. Le théorème de fluctuation - dissipation de la mécanique statistique donne le résultat que la susceptibilité magnétique est donnée par $\chi_m = \sigma^2 \mu^2 / k_B T$. Les résultats d'une série de simulations avec les mêmes paramètres qu'à la FIG. 5.3 sont montrés à la FIG. 5.4. On remarque la brutale augmentation de cette quantité au voisinage de la température critique. Notre système de spins du modèle d'Ising présente une susceptibilité magnétique très importante au voisinage de la température de transition de phase. En fait, pour un système de taille infinie (nombre infini de spins), il s'avère même que χ tend vers l'infini lorsque $T \to T_c$.

On peut faire une analyse intéressante de l'énergie du système E en fonction de la température T. Les résultats numériques montrent que la "pente" dE/dT est maximale en $T = T_c$. En fait, cette pente est infinie dans la limite d'un système de taille infinie. Ce qui veut dire que la chaleur spécifique du système tend vers l'infini lorsqu'on s'approche de la température critique T_c. Le théorème de fluctuation - dissipation donne une chaleur spécifique $C = \sigma^2(E)/k_B T^2$, où $\sigma^2(E)$ désigne la variance de l'énergie moyenne par spin.

2. On prend la valeur absolue du spin moyen car, comme noté plus haut, on s'intéresse à l'aimantation en valeur absolue, et non au fait de savoir si cette moyenne est positive ou négative, états qui sont symtriquement probables.

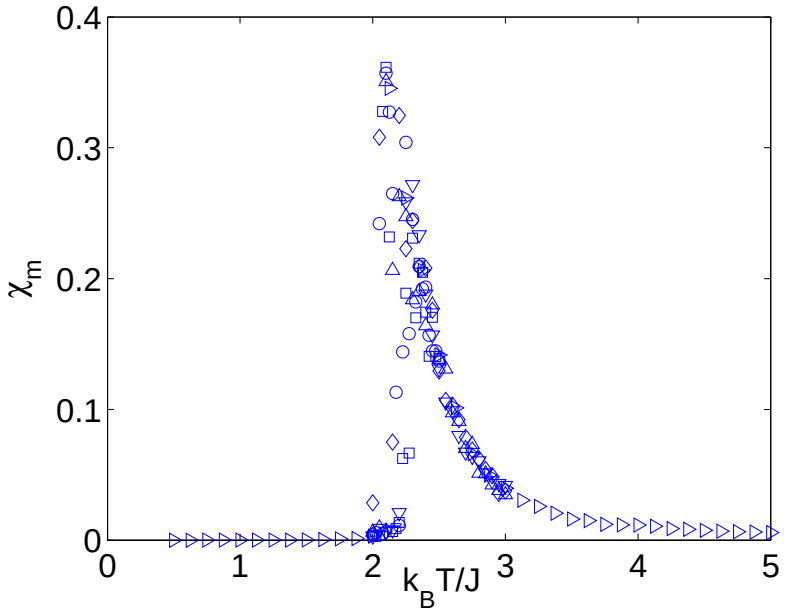


FIGURE 5.4 – *Susceptibilité magnétique χ_m obtenue à partir de la variance σ^2 du spin pour les simulations de la FIG. 5.3 ($\chi_m = \sigma^2\mu^2/k_BT$).*

Un tel comportement au voisinage de la température critique est caractéristique des transitions de phase appelées *du 2e ordre*, parfois aussi qualifiées de *continues* : l'aimantation, qui est la 1e dérivée de l'énergie libre par rapport à H, croît de façon continue à partir d'une valeur nulle à $T = T_c$ lorsque la température décroît. La susceptibilité magnétique et la chaleur spécifique, par contre, 2e dérivées de l'énergie libre, changent de façon discontinue et présentent une singularité au point critique. Les transitions de phase dites du premier ordre présentent des discontinuités de la première dérivée de l'énergie libre. Elles impliquent l'existence d'une chaleur latente. Par exemple, la solidification / liquéfaction et l' évaporation / condensation sont des transitions de phase du premier ordre.

Il est intéressant d'étudier l'effet d'un champ extérieur H sur le modèle d'Ising avec l'algorithme de Metropolis. Pour une température $T = 0.25$, bien inférieure à la température critique $T_c = 2.27$, on montre à la FIG. 5.5 le spin moyen en fonction du champ appliqué H. Il y a transition abrupte, mais le fait remarquable est que la valeur du champ H à laquelle cette transition se produit dépend de l'histoire du système : si on augmente le champ H à partir d'une valeur négative, il faut plus que juste inverser la direction du champ magnétique pour faire basculer les spins dans l'autre sens : on remarque que le spin moyen reste négatif même pour des valeurs de H positives entre 0 et 2.5. Réciproquement, si on fait décroître le champ H à partit d'une valeur positive, on trouve des cas où le spin moyen reste positif alors que le champ H est négatif, entre 0 et -2.5. C'est le phénomène *d'hystérèse*. Le champ nécessaire à faire basculer les spins dans l'autre sens est le *champ de démagnétisation*.

En augmentant la température, on verra (suggestion d'exercice) que la valeur du champ

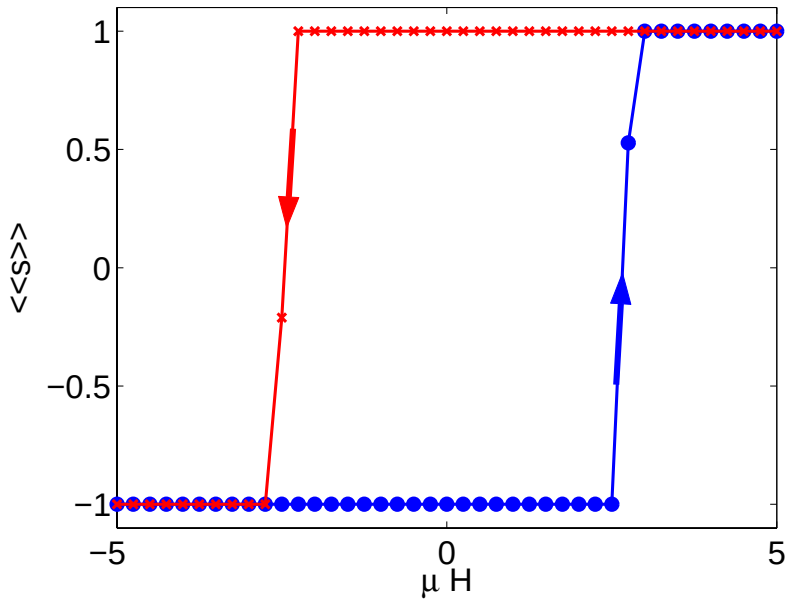


FIGURE 5.5 – *Moyenne du spin en fonction du champ magnétisant appliqué H (normalisé à J/μ) pour une température normalisée $k_BT/J = 0.25$. Modèle d'Ising 2-D à $n = 4$ plus proches voisins sur un réseau périodique de 10×10 spins, 1000 étapes de l'algorithme de Metropolis pour chaque simulation. Les simulations avec cercles bleus ont été effectuées pour des valeurs de H croissantes, celles avec les croix rouges pour des valeurs de H décroissantes. On remarque le phénomène d'hystérèse.*

de démagnétisation diminue, et que donc le cycle d'hystérèse diminue de taille à mesure que l'on s'approche de la température de transition T_c. Pour $T \geq T_c$ la discontinuité de l'aimantation disparaît complètement.

Un comportement similaire se produit pour la transition de phase entre liqude et gaz : il existe une température critique T_c au-delà de laquelle les phases liquide et gazeuse ne sont plus séparées par une transition de phase : on passe continûment de l'une à l'autre.

Remarque : pour une taille finie du système du modèle d'Ising 2-D, on peut montrer que le cycle d'hystérèse en fait disparaît. Si on effectuait des simulations Metropolis très longues, on verrait un basculement des spins au bout d'un certain nombre d'étapes, qui tend exponentiellement vers l'infini pour $T \to 0$. La probabilité de basculement des spins devient en effet exponentiellement petite, et il faudrait faire des simulations infiniment longues pour l'observer. Le fait que l'on observe des valeurs de spin moyen ayant la direction opposée à celle du champ magnétisant signifie qu'en fait la simulation n'a pas permis encore d'atteindre l'état d'équilibre. Le système est dans un état dit *métastable* sur la période de la simulation.

Pour en savoir plus : Bibliographie, Refs. [20]-[24]

Annexe A

From Taylor to Abramowitz to Pascal

A.1 Even order derivatives

We are going to obtain finite difference expressions for derivatives of order 2 and 4. We start from Taylor series expansions around grid points x_j, with $f_j = f(x_j)$:

$$f_{j-2} = f_j - 2hf'_j + 2h^2 f''_j - \frac{8}{6}h^3 f_j^{(3)} + \frac{16}{24}h^4 f_j^{(4)} - \frac{32}{120}h^5 f_j^{(5)} + \mathcal{O}(h^6) \quad \text{(A.1)}$$

$$f_{j-1} = f_j - hf'_j + \frac{1}{2}h^2 f''_j - \frac{1}{6}h^3 f_j^{(3)} + \frac{1}{24}h^4 f_j^{(4)} - \frac{1}{120}h^5 f_j^{(5)} + \mathcal{O}(h^6) \quad \text{(A.2)}$$

$$f_{j+1} = f_j + hf'_j + \frac{1}{2}h^2 f''_j + \frac{1}{6}h^3 f_j^{(3)} + \frac{1}{24}h^4 f_j^{(4)} + \frac{1}{120}h^5 f_j^{(5)} + \mathcal{O}(h^6) \quad \text{(A.3)}$$

$$f_{j+2} = f_j + 2hf'_j + 2h^2 f''_j + \frac{8}{6}h^3 f_j^{(3)} + \frac{16}{24}h^4 f_j^{(4)} + \frac{32}{120}h^5 f_j^{(5)} + \mathcal{O}(h^6) \quad \text{(A.4)}$$

Odd order derivatives will be eliminated by taking sums of pairs of these expressions. Eq.(A.2) + Eq.(A.3) and Eq.(A.1) + Eq.(A.4) give respectively :

$$f_{j-1} + f_{j+1} = 2f_j + h^2 f''_j + \frac{1}{12}h^4 f_j^{(4)} + \mathcal{O}(h^6) \quad \text{(A.5)}$$

$$f_{j-2} + f_{j+2} = 2f_j + 4h^2 f''_j + \frac{4}{3}h^4 f_j^{(4)} + \mathcal{O}(h^6) \quad \text{(A.6)}$$

To obtain first order accurate second order derivative f''_j, we use Eq.(A.5) neglecting $\mathcal{O}(h^4)$:

$$h^2 f''_j = f_{j-1} - 2f_j + f_{j+1} + \mathcal{O}(h^4) \Rightarrow f''_j = \frac{1}{h^2}\left(f_{j-1} - 2f_j + f_{j+1}\right) + \mathcal{O}(h^2) \quad \text{(A.7)}$$

To obtain first order accurate fourth order derivative $f_j^{(4)}$, we eliminate f''_j from 4*Eq.(A.5) - Eq.(A.6) :

$$f_j^{(4)} = \frac{1}{h^4}\left(f_{j-2} - 4f_{j-1} + 6f_j - 4f_{j+1} + f_{j+2}\right) + \mathcal{O}(h^2) \quad \text{(A.8)}$$

To obtain *second order accurate* second order derivative f''_j, we eliminate $f^{(4)}_j$, from 16*Eq.(A.5) - Eq.(A.6) :

$$-f_{j-2} + 16f_{j-1} + 16f_{j+1} - f_{j+2} = 30f_j + 12h^2 f''_j + \mathcal{O}(h^6)$$

$$f''_j = \frac{1}{12h^2}\left(-f_{j-2} + 16f_{j-1} - 30f_j + 16f_{j+1} - f_{j+2}\right) + \mathcal{O}(h^4) \tag{A.9}$$

Remark 1 : For a given order of accuracy in h, increasing the order of the derivative requires an increasingly large number of grid points : compare Eq.(A.7) and Eq.(A.8).
Remark 2 : For a given order of derivative n, increasing the order of accuracy requires an increasingly large number of grid points : compare Eq.(A.7) and Eq.(A.9).

A.2 Odd order derivatives

Remark that the finite difference expressions for odd order derivatives are **centered**, i.e. they are expressed at half-integer grid point numbers, i.e. half way between grid points. The method is therefore the same as for even order derivatives, except that we make Taylor series expansions around half-integer grid point $x_{j+1/2}$, with $f_{j+1/2} = f(x_{j+1/2})$.

$$f_{j-1} = f_{j+1/2} - \frac{3}{2}hf'_{j+1/2} + \frac{9}{8}h^2 f''_{j+1/2} - \frac{27}{48}h^3 f^{(3)}_{j+1/2} + \frac{81}{384}h^4 f^{(4)}_{j+1/2} + \mathcal{O}(h^5) \tag{A.10}$$

$$f_j = f_{j+1/2} - \frac{1}{2}hf'_{j+1/2} + \frac{1}{8}h^2 f''_{j+1/2} - \frac{1}{48}h^3 f^{(3)}_{j+1/2} + \frac{1}{384}h^4 f^{(4)}_{j+1/2} + \mathcal{O}(h^5) \tag{A.11}$$

$$f_{j+1} = f_{j+1/2} + \frac{1}{2}hf'_{j+1/2} + \frac{1}{8}h^2 f''_{j+1/2} + \frac{1}{48}h^3 f^{(3)}_{j+1/2} + \frac{1}{384}h^4 f^{(4)}_{j+1/2} + \mathcal{O}(h^5) \tag{A.12}$$

$$f_{j+2} = f_{j+1/2} + \frac{3}{2}hf'_{j+1/2} + \frac{9}{8}h^2 f''_{j+1/2} + \frac{27}{48}h^3 f^{(3)}_{j+1/2} + \frac{81}{384}h^4 f^{(4)}_{j+1/2} + \mathcal{O}(h^5) \tag{A.13}$$

We eliminate $f_{j+1/2}$ (which is in principle not known to us : f is sampled on integer grid points only) and even order derivatives by taking differences of these expressions, Eq.(A.12)-Eq.(A.11) and Eq.(A.13)-Eq.(A.10) :

$$f_{j+1} - f_j = hf'_{j+1/2} + \frac{1}{24}h^3 f^{(3)}_{j+1/2} + \mathcal{O}(h^5) \tag{A.14}$$

$$f_{j+2} - f_{j-1} = 3hf'_{j+1/2} + \frac{27}{24}h^3 f^{(3)}_{j+1/2} + \mathcal{O}(h^5) \tag{A.15}$$

To obtain first order accurate first order derivative f'_j, we use Eq.(A.14) neglecting $\mathcal{O}(h^3)$:

$$f'_{j+1/2} = \frac{1}{h}\left(f_{j+1} - f_j\right) + \mathcal{O}(h^2) \tag{A.16}$$

To obtain first order accurate third order derivative $f^{(3)}_j$, we eliminate f'_j from Eq.(A.15) - 3*Eq.(A.14) :

$$f^{(3)}_{j+1/2} = \frac{1}{h^3}\left(-f_{j-1} + 3f_j - 3f_{j+1} + f_{j+2}\right) + \mathcal{O}(h^2) \tag{A.17}$$

A.3 Pascal triangle

An alternative way to derive first order accurate finite difference formulae for derivatives is the following. Consider first two neighbouring grid points j and j+1. We have

$$f'_{j+1/2} = \frac{1}{h}\left(f_{j+1} - f_j\right) + \mathcal{O}(h^2) \tag{A.18}$$

We then consider two adjacent grid points j-1 and j. We have

$$f'_{j-1/2} = \frac{1}{h}\left(f_j - f_{j-1}\right) + \mathcal{O}(h^2) \tag{A.19}$$

We apply once more the first order derivative finite difference expression, but this time to the function f', considering the half integer grid points j-1/2 and j+1/2 :

$$f''_j = (f')'_j = \frac{1}{h}\left(f'_{j+1/2} - f'_{j-1/2}\right) + \mathcal{O}(h^2) \tag{A.20}$$

Substituting Eqs.(A.18,A.19) into Eq.(A.20) we get

$$f''_j = \frac{1}{h^2}\left(f_{j-1} - 2f_j + f_{j+1}\right) + \mathcal{O}(h^2) \tag{A.21}$$

To obtain the third order derivative, we write down Eq.(A.21) for f'' at grid point j+1, and use the finite difference expression Eq.(A.18) for the first derivative of f''.

And so on...

We can obtain the coefficients of the finite difference expressions for derivatives by constructing Pascal triangle

$$\begin{array}{ll}
(0) & 1 \\
(1) & 1\ \ 1 \\
(2) & 1\ \ 2\ \ 1 \\
(3) & 1\ \ 3\ \ 3\ \ 1 \\
(4) & 1\ \ 4\ \ 6\ \ 4\ \ 1 \\
\ldots & \ldots\ldots\ldots\ldots
\end{array}$$

and then putting alternate + and - signs

$$\begin{array}{ll}
(0) & +1 \\
(1) & -1\ \ +1 \\
(2) & +1\ \ -2\ \ +1 \\
(3) & -1\ \ +3\ \ -3\ \ +1 \\
(4) & +1\ \ -4\ \ +6\ \ -4\ \ +1 \\
\ldots & \ldots\ldots\ldots\ldots
\end{array}$$

A.4 Forward finite differences

The idea is to start from Taylor expansions around point no j using expressions at forward grid points, i.e. j+1, j+2, ..., see Eqs(A.3, A.4). From Eq.(A.3), neglecting $\mathcal{O}(h^2)$, we obtain the lowest order accurate, first order derivative :

$$f'_j \approx \frac{1}{h}\left(f_{j+1} - f_j\right) \tag{A.22}$$

Retaining terms up to h^2 an neglecting $\mathcal{O}(h^3)$, we obtain from 4* Eq.(A.3)-Eq.(A.4)

$$4f_{j+1} - f_{j+2} = 3f_j + 2hf'_j + \mathcal{O}(h^3)$$

$$\Rightarrow \; f'_j = \frac{1}{2h}\left(-3f_j + 4f_{j+1} - f_{j+2}\right) + \mathcal{O}(h^2) \tag{A.23}$$

Higher order in accuracy and higher order derivative expressions can be obtained by considering further forward points.

Annexe B

Intégration numérique

Soit le segment $[a, b]$ et une discrétisation de N intervalles, avec les points de maillage x_i, $i = 1..N + 1$ équidistants de h. Soit une fonction $f \in C^n([a, b])$, avec n un entier positif "suffisamment grand". Pour obtenir une approximation à

$$\int_a^b f(x)dx \tag{B.1}$$

on a les formules suivantes.

B.1 Point milieu, trapèzes, Simpson

— (a) Règle du point milieu : soit $x_{i+1/2} = (x_i + x_{i+1})/2$;

$$\int_a^b f(x)dx = h\sum_{i=1}^{N} f(x_{i+1/2}) + \mathcal{O}(h^2) \ . \tag{B.2}$$

— (b) Règle des trapèzes :

$$\int_a^b f(x)dx = h\sum_{i=1}^{N} \left(f(x_i) + f(x_{i+1})\right)/2 + \mathcal{O}(h^2) \ . \tag{B.3}$$

— (c) Règle de Simpson :

$$\int_a^b f(x)dx = h\sum_{i=1}^{N} \frac{1}{6}\left(f(x_i) + 4f(x_{i+1/2}) + f(x_{i+1})\right) + \mathcal{O}(h^4) \ . \tag{B.4}$$

Ces formules, ainsi que l'ordre de l'erreur, $\mathcal{O}(h^n)$, s'obtiennent à partir de développements limités de la fonction f. Cela présuppose que f est de régularité suffisante. Considérons

l'intervalle numéro i, $[x_i, x_{i+1}]$. Soit $x_{i+1/2}$ le point milieu de cet intervalle. Le développement de Taylor de f au voisinage de ce point donne

$$f(x_{i+1/2}+\epsilon) = f(x_{i+1/2}) + \epsilon f'_{i+1/2} + \frac{1}{2}\epsilon^2 f''_{i+1/2} + \frac{1}{6}\epsilon^3 f'''_{i+1/2} + \mathcal{O}(\epsilon^4) \ . \tag{B.5}$$

Intégrant sur l'intervalle, on a

$$\int_{x_i}^{x_{i+1}} f(x)dx = \int_{-h/2}^{+h/2} f(x_{i+1/2}+\epsilon)d\epsilon = hf_{i+1/2} + \frac{h^3}{24} f''_{i+1/2} + \mathcal{O}(h^5) \ , \tag{B.6}$$

la contribution des termes de puissance paire en h étant nulle.

La règle du point milieu (a) s'obtient en ne considérant que le premier terme de (B.6). Sur chaque intervalle, l'erreur est ainsi d'ordre h^3. En sommant sur les N intervalles, puisque $N \propto 1/h$, l'erreur sur l'intégrale entre a et b est d'ordre h^2.

La règle des trapèzes (b) s'obtient en ne considérant que le premier terme de (B.6) et en substituant $f_{i+1/2}$ par les développements limités de f autour de $x_{i+1/2}$ en x_i et x_{i+1}, c'est-à-dire l'Eq.(B.5) avec $\epsilon = -h/2$ et $+h/2$, respectivement :

$$f_i = f_{i+1/2} - \frac{h}{2} f'_{i+1/2} + \mathcal{O}(h^2) \ , \tag{B.7}$$

$$f_{i+1} = f_{i+1/2} + \frac{h}{2} f'_{i+1/2} + \mathcal{O}(h^2) \ . \tag{B.8}$$

En faisant la moyenne des deux expressions ci-dessus, on a

$$f_{i+1/2} = \frac{1}{2}\left(f_i + f_{i+1}\right) + \mathcal{O}(h^2) \ , \tag{B.9}$$

et la formule des trapèzes donne une erreur d'ordre h^3 pour chaque intervalle, donc d'ordre h^2 pour l'intégrale entre a et b.

La règle de Simpson (c) s'obtient de (B.6) et de l'expression aux différences finies (A.7) pour $f''_{i+1/2}$ [N.B. substituant $h \to h/2$, $j \to i+1/2$, $j-1 \to i$, $j+1 \to i+1$] :

$$f''_{i+1/2} = \frac{1}{(h/2)^2}\left(f_i - 2f_{i+1/2} + f_{i+1}\right) + \mathcal{O}(h^2) \ . \tag{B.10}$$

On obtient ainsi

$$\int_{x_i}^{x_{i+1}} f(x)dx = hf_{i+1/2} + \frac{h^3}{24}\left[\frac{4}{h^2}\left(f_i - 2f_{i+1/2} + f_{i+1}\right) + \mathcal{O}(h^2)\right] \ , \tag{B.11}$$

$$\int_{x_i}^{x_{i+1}} f(x)dx = \frac{h}{6}\left(f_{i-1} + 4f_{i+1/2} + f_{i-1}\right) + \mathcal{O}(h^5) \ . \tag{B.12}$$

L'erreur de la règle de Simpson est ainsi d'ordre h^5 pour chaque intervalle, et donc d'ordre h^4 pour l'intégrale entre a et b.

B.2 Méthode de quadrature de Gauss

A la section précédente, on obtenait une estimation de l'intégrale en sommant, avec des poids différents, la fonction f évaluée aux points milieux $x_{i+1/2}$ et/ou aux points de bords x_i des intervalles de discrétisation.

L'idée de la méthode de Gauss est de choisir non seulement les poids, mais aussi les abcisses, des points où la fonction f est évaluée. Soit n un entier positif. On écrit la contribution de l'intervalle numéro i, $[x_i, x_{i+1}]$, à l'intégrale :

$$\int_{x_i}^{x_{i+1}} f(x)dx = \frac{h}{2}\sum_{j=1}^{n} w_j f(x_j) + R_n \ , \tag{B.13}$$

avec

$$x_j = x_{i+1/2} + \frac{h}{2}\xi_j \ . \tag{B.14}$$

Les abcisses ξ_j et les poids w_j sont donnés dans la table ci-dessous. Le résidu (erreur) R_n est d'ordre p. La méthode de Gauss consiste à choisir judicieusement les poids w_j et les abcisses ξ_j de telle sorte que la formule d'intégration soit *exacte* pour un polynôme de degré $2n-1$.

n	ξ_j	w_j	p
1	0	2	2
2	$\pm\sqrt{1/3}$	1	4
3	0	$8/9$	6
	$\pm\sqrt{3/5}$	$5/9$	
4	$\pm\sqrt{3/7-\sqrt{120}/35}$	$1/2+5/(3\sqrt{120})$	8
	$\pm\sqrt{3/7+\sqrt{120}/35}$	$1/2-5/(3\sqrt{120})$	
5	0	$128/225$	10
	$\pm\sqrt{245-14\sqrt{70}}/21$	$(322+13\sqrt{70})/900$	
	$\pm\sqrt{245+14\sqrt{70}}/21$	$(322-13\sqrt{70})/900$	

B.3 Intégration de Monte Carlo

L'idée est de choisir les abcisses x_i non pas selon un maillage régulier, mais "au hasard", c'est-à-dire selon une fonction de distribution de probabilité uniforme dans l'intervalle $[a, b]$. Obtenir une séquence de N points aléatoires sur un ordinateur n'est pas si trivial : l'ordinateur ne peut pas "jouer aux dés", pour paraphraser un célebre physicien. Il faut un algorithme, qui par définition est une séquence d'instructions déterministes. On ne peut que simuler le caratère aléatoire : on parle de générateur *pseudo*-aléatoire.

On obtient ensuite une approximation à l'intégrale :

$$\int_a^b f(x)dx \approx \frac{b-a}{N} \sum_{i=1}^{N} f(x_i) \ . \tag{B.15}$$

La question des erreurs de cette méthode est cruciale. Elle se base sur le théorème central limite. Si les $f(x_i)$ sont des variables aléatoires avec une variance non nulle, si elles sont distribuées selon la même densité de probabilité, et si elles sont indépendantes, alors leur somme tend, pour $N \to \infty$, vers une variable aléatoire ayant une fonction de distribution de probabilité **normale**, dont la variance est proportionnelle à $\sqrt{N}$. La variance de l'intégrale de Monte Carlo a donc une **variance σ proportionnelle à $1/\sqrt{N}$.**

La méthode de Monte Carlo pour estimer une intégrale est comparativement avantageuse pour des intégrales à plusieurs dimensions d. Les méthodes des sections précédentes, basées sur des maillages réguliers, nécessitent de l'ordre de $N_{tot} = 1/h^d$ évaluations de la fonction f (à tous les points de maillage). Pour la méthode des trapèzes, par exemple, l'erreur est en h^2. Comme $h = N_{tot}^{-1/d}$, l'erreur va comme $N_{tot}^{-2/d}$. Pour la méthode de Simpson, l'erreur est en h^4 et donc en $N_{tot}^{-4/d}$. L'erreur dans la méthode de Monte Carlo est en $N_{tot}^{-1/2}$ *quel que soit le nombre de dimensions d.* Ainsi, la méthode Monte Carlo devient avantageuse par rapport à la règle des trapèzes pour $d > 4$, et par rapport à la règle de Simpson pour $d > 8$.

Annexe C

Solution analytique de l'équation d'advection-diffusion

Nous allons obtenir la solution analytique de l'équation d'advection-diffusion en 1D, Eq.(4.19) :

$$\boxed{\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} - D\frac{\partial^2 f}{\partial x^2} = 0}\,. \tag{C.1}$$

avec la condition initiale

$$f(x,0) = N\delta(x - x_0)\ , \tag{C.2}$$

avec x_0 donnée, ce qui correspond à placer, en $t = 0$, toutes les N particules à la même position x_0. Soit $\hat{f}(k,t)$ la transformée de Fourier spatiale de $f(x,t)$,

$$\hat{f}(k,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(x,t)e^{-ikx}\,dx\ . \tag{C.3}$$

On a donc la transformée de Fourier inverse :

$$f(x,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \hat{f}(k,t)e^{+ikx}\,dk\ . \tag{C.4}$$

Prendre la dérivée partielle par rapport à x revient, dans l'espace de Fourier, à appliquer une multiplication par ik. L'Eq.(C.1) s'écrit donc, dans l'espace de Fourier :

$$\frac{\partial \hat{f}}{\partial t} + ikv\hat{f} + k^2 D\hat{f} = 0\ . \tag{C.5}$$

La solution pour $\hat{f}$ s'obtient facilement :

$$\hat{f}(k,t) = \hat{f}(k,0)\exp\left[-(ikv + k^2 D)t\right] \tag{C.6}$$

Pour obtenir $n(x,t)$, il faut revenir dans l'espace réel, autrement dit appliquer une transformée de Fourier inverse. Ici, la fonction $\hat{f}$ est sous la forme d'un produit de fonctions : $\hat{f} = \hat{f}(k,0)\hat{G}(k,t)$, avec

$$\hat{G}(k,t) = \exp\left[-(ikv + k^2 D)t\right]\ . \tag{C.7}$$

Or, la transformée de Fourier inverse d'un produit de fonctions est une convolution dans l'espace réel :

$$f(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x',0) G(x-x',t)\, dx' \;, \tag{C.8}$$

où $G(x,t)$ est la transformée de Fourier inverse de $\hat{G}(k,t)$:

$$G(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp -(ikv + k^2 D)t \; e^{+ikx} \; dk \;; \tag{C.9}$$

$$G(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{ik(x-vt)} e^{-k^2 Dt} \; dk \;. \tag{C.10}$$

On utilise la formule

$$\int_{-\infty}^{+\infty} e^{-p^2 k^2} e^{qk} dk = \frac{\sqrt{\pi}}{p} e^{q^2/4p^2} \tag{C.11}$$

avec $p = \sqrt{Dt}$ et $q = i(x - vt)$, pour obtenir :

$$G(x,t) = \frac{1}{\sqrt{2Dt}} \, e^{-(x-vt)^2/4Dt} \;. \tag{C.12}$$

Insérant cette expression, et la condition initiale Eq.(C.2) dans l'expression de convolution Eq.(C.8), on obtient :

$$f(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} N\delta(x' - x_0) \frac{1}{\sqrt{2Dt}} \exp\left[-(x - x' - vt)^2/4Dt\right] dx' \tag{C.13}$$

En utilisant la propriété de la fonction δ,

$$\int_{-\infty}^{+\infty} \delta(x' - x_0) f(x') dx' = f(x_0) \;, \tag{C.14}$$

on obtient

$$\boxed{f(x,t) = \frac{N}{\sqrt{4\pi Dt}} \exp\left[-(x - x_0 - vt)^2/4Dt\right]} \;. \tag{C.15}$$

On obtient bien l'Eq.(4.20).

Annexe D

Coefficient de diffusion et marche aléatoire

Nous allons établir la relation entre la variance d'une marche aléatoire et le coefficient de diffusion.

Soit une marche aléatoire résultant d'une succession de M "pas" ξ_i, variables aléatoires statistiquement indépendantes, de moyenne nulle $< \xi_i >= 0$ et de variance non nulle $< \xi_i^2 > \neq 0$.

La position finale $x = \sum_{i=1}^{M} \xi_i$ est une variable aléatoire de moyenne nulle, $< x >= \sum_{i=1}^{M} < \xi_i >= 0$, mais de variance non nulle :

$$< x^2 >= \left\langle \left(\sum_{i=1}^{M} \xi_i \right) \left(\sum_{j=1}^{M} \xi_j \right) \right\rangle = \sum_{i=1}^{M} < \xi_i^2 > + \sum_{i \neq j} < \xi_i \xi_j >= M < \xi_i^2 > \ . \quad \text{(D.1)}$$

On peut définir le temps caractéristique τ par le temps entre deux "pas", c'est-à-dire entre deux collisions successives, et le libre parcours moyen λ_{mfp} par

$$\lambda_{\text{mfp}} = \sqrt{< \xi_i^2 >} \, . \quad \text{(D.2)}$$

Le nombre de "pas" (de collisions) M pendant un intervalle de durée Δt est donc $\Delta t / \tau$, et on a

$$< x^2 >= \frac{\Delta t}{\tau} \lambda_{\text{mfp}}^2 \, . \quad \text{(D.3)}$$

Considérons maintenant la description continue, soit l'équation de diffusion, Eq.(4.17), que l'on prend ici en 1-D avec un coefficient de diffusion D constant et uniforme , Eq.(4.19) avec une vitesse d'advection nulle ($v = 0$). Prenant le 2e moment de cette équation (multipliant par x^2 et intégrant sur x), le premier terme donne

$$\int_{-\infty}^{+\infty} x^2 \frac{\partial n}{\partial t} dx = \frac{\partial}{\partial t} \int_{-\infty}^{+\infty} x^2 n dx = N \frac{\partial}{\partial t} \bar{x^2} \, , \quad \text{(D.4)}$$

où $N = \int n(x,t)dx$ est le nombre total de particules et

$$\bar{x^2}(t) = \frac{1}{N}\int_{-\infty}^{+\infty} x^2 n(x,t)dx \ . \tag{D.5}$$

Le 3e terme donne, en intégrant par parties,

$$\begin{aligned}\int_{-\infty}^{+\infty} -x^2 D\frac{\partial^2 n}{\partial x^2}dx &= -\left[x^2 D\frac{\partial n}{\partial x}\right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty}\frac{\partial}{\partial x}(x^2 D)\frac{\partial n}{\partial x}dx \\ &= \left[\frac{\partial}{\partial x}(x^2 D)n\right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty}\frac{\partial^2}{\partial x^2}(x^2 D)\, n\, dx \\ &= -2D\int_{-\infty}^{+\infty} n\, dx = -2DN \ . \end{aligned} \tag{D.6}$$

On a donc, de (D.4) et (D.6),

$$N\frac{\partial \bar{x^2}}{\partial t} - 2DN = 0 \quad \Rightarrow \quad \bar{x^2} = \bar{x^2}(0) + 2Dt \ . \tag{D.7}$$

On identifie $\bar{x^2}$ de la description continue (macroscopique) avec la variance $< x^2 >$ de la description de la marche aléatoire (microscopique). Pour une marche aléatoire, la variance de la position initiale est nulle, et on a, pour l'intervalle de temps Δt,

$$\boxed{< x^2 >= 2D\Delta t} \ . \tag{D.8}$$

Ainsi, on peut exprimer le coefficient de diffusion de la description continue (macroscopique) en termes de grandeurs liées à la marche aléatoire (microscopique), à partir de (D.3) et (D.8) :

$$\boxed{D = \frac{\lambda_{\rm mfp}^2}{2\tau}} \ . \tag{D.9}$$

Annexe E

Equations d'ondes en eaux peu profondes

Nous allons établir les équations régissant les ondes à la surface de l'eau, sous certaines hypothèses simplificatrices, appelés "ondes en eaux peu profondes" (shallow water wave equations). A une dimension d'espace, nous montrerons que l'on obtient une équation de la forme de l'Eq.(4.62), avec une vitesse de propagation donnée par l'Eq.(4.63).

Considérons un fluide incompressible de densité (constante) ρ_0. Au repos, la profondeur est donnée par une fonction $h_0(x)$ donnée et la vitesse du fluide est nulle $v_0 = 0$. En présence de perturbation (Fig. E.1), la profondeur et la vitesse sont

$$h(x,t) = h_0(x) + \delta h(x,t) \tag{E.1}$$
$$\vec{v}(x,t) = 0 + \delta\vec{v}(x,t) \tag{E.2}$$

Les équations de base sont obtenues de l'équation du mouvement, ou 2e loi de Newton, pour une particule fluide de la surface de l'eau, et de l'équation de continuité exprimant la conservation de la masse au cours du mouvement :

$$\rho_0 \frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = -\nabla P + \rho_0 \vec{g}\,, \tag{E.3}$$

$$\frac{\partial h}{\partial t} + \nabla \cdot (h\vec{v}) = 0\,. \tag{E.4}$$

Projetant l'Eq.(E.3) sur l'axe vertical y,

$$\rho_0 \frac{\mathrm{d}v_y}{\mathrm{d}t} = -\frac{\partial P}{\partial y} - \rho_0 g \tag{E.5}$$

On fait l'hypothèse que

$$\left|\frac{\mathrm{d}v_y}{\mathrm{d}t}\right| \ll g \tag{E.6}$$

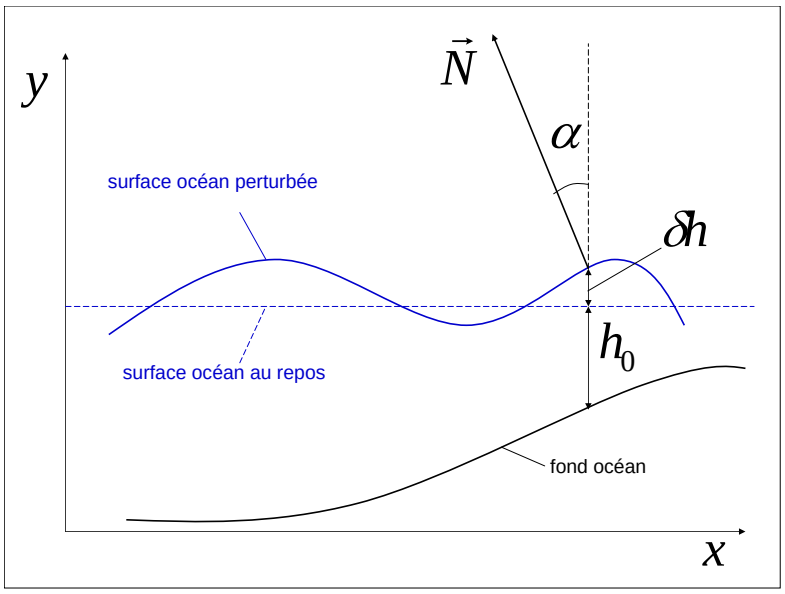


FIGURE E.1 – *Vague sur l'océan.*

qui revient à supposer que le mouvement vertical est suffisamment lent et varie lentement, de sorte que l'accélération verticale est négligeable par rapport à la pesanteur. Ainsi,

$$\frac{\partial P}{\partial y} = -\rho_0 g \ . \tag{E.7}$$

Projetant l'Eq.(E.3) sur l'axe horizontal x,

$$\rho_0 \frac{\mathrm{d}v_x}{\mathrm{d}t} = -\frac{\partial P}{\partial x} \tag{E.8}$$

Ecrivant $\vec{N} = -\nabla P$, et sachant que le gradient de pression est normal aux isobares et que la surface de l'eau est une isobare, et définissant l'angle α par

$$\tan\alpha = \frac{\partial \delta h}{\partial x} \tag{E.9}$$

(voir Fig. E.1), α est l'angle que fait $\vec{N}$ avec la verticale, tel que $N_x = -\partial P/\partial x = -|N|\sin\alpha$, $N_y = -\partial P/\partial y = |N|\cos\alpha$. Avec l'Eq.(E.7), $|N|\cos\alpha = \rho_0 g$, et ainsi, il vient :

$$\rho_0 \frac{\mathrm{d}v_x}{\mathrm{d}t} = -|N|\sin\alpha = -\rho_0 g \tan\alpha = -\rho_0 g \frac{\partial \delta h}{\partial x} \ . \tag{E.10}$$

Ainsi,

$$\frac{\partial v_x}{\partial t} + v_x \frac{\partial v_x}{\partial x} = -g \frac{\partial \delta h}{\partial x} \ . \tag{E.11}$$

En supposant un problème unidimensionnel en x (donc $\partial/\partial y = \partial/\partial z = 0$), l'Eq.(E.4) s'écrit

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(h v_x) = 0 \ . \tag{E.12}$$

Séparant l'équilibre de la perturbation, Eqs.(E.1-E.2), on obtient, après linéarisation :

$$\boxed{\frac{\partial \delta v_x}{\partial t} + g\frac{\partial \delta h}{\partial x} = 0}\,, \tag{E.13}$$

$$\boxed{\frac{\partial \delta h}{\partial t} + \frac{\partial}{\partial x}\left(h_0 \delta v_x\right) = 0}\,. \tag{E.14}$$

Prenant $\partial/\partial t$ de l'Eq.(E.14) et substituant $\partial \delta v_x/\partial t$ de l'Eq.(E.13), on obtient :

$$\boxed{\frac{\partial^2 \delta h}{\partial t^2} - \frac{\partial}{\partial x}\left(g h_0 \frac{\partial \delta h}{\partial x}\right) = 0}\,. \tag{E.15}$$

Il s'agit bien d'une équation de la même forme que l'Eq.(4.62). On identifie ainsi

$$\boxed{u(x) = \sqrt{g h_0(x)}} \tag{E.16}$$

Equation de balance d'énergie

On peut obtenir une équation de type conservatif pour une quantité que l'on identifiera avec l'énergie de l'onde. Multipliant l'Eq.(E.13) par $h_0 \delta v_x$,

$$h_0 \delta v_x \frac{\partial \delta v_x}{\partial t} + g h_0 \delta v_x \frac{\partial \delta h}{\partial x} = 0 \tag{E.17}$$

$$\Rightarrow \frac{\partial}{\partial t}\left(h_0 \frac{\delta v_x^2}{2}\right) + \frac{\partial}{\partial x}\left(g h_0 \delta v_x \delta h\right) - g\frac{\partial}{\partial x}\left(h_0 \delta v_x\right)\delta h = 0 \tag{E.18}$$

De l'Eq.(E.14), on a $\partial(h_0 \delta v_x)/\partial x = -\partial \delta h/\partial t$, et il vient

$$\boxed{\frac{\partial}{\partial t}\left(\frac{1}{2} h_0 (\delta v_x)^2 + \frac{1}{2} g (\delta h)^2\right) + \frac{\partial}{\partial x}\left(g h_0 \delta v_x \delta h\right) = 0}\,. \tag{E.19}$$

C'est une équation de continuité pour la **densité d'énergie de l'onde**

$$\boxed{\mathcal{E} = \frac{1}{2} h_0 (\delta v_x)^2 + \frac{1}{2} g (\delta h)^2} \tag{E.20}$$

et on identifie le **flux d'énergie de l'onde**

$$\boxed{\mathcal{S} = g h_0 \delta v_x \delta h}\,. \tag{E.21}$$

N.B. : Le problème 2D que nous avons résolu ici est tel que la coordonnée z est ignorable. En d'autres termes, on a obtenu une description valable pour une "tranche" d'épaisseur L_z arbitraire. Pour obtenir des quantités en unités physiques habituelles, on notera que $\mathcal{E}\rho_0/L_z$ est une énergie par unité de volume, et $\mathcal{S}\rho_0/L_z$ est une énergie par unité de surface et par unité de temps. Multiplier ces quantités par une constante ρ_0/L_z ne change pas leur propriétés de conservation.

Bibliographie

[1] N.J. Giordano and H. Nakanishi, *Computational Physics (2nd Edition)* Prentice Hall (2006) ISBN 0-13-146990-8

[2] F.J. Vesely, *Computational Physics, An Introduction (2nd Edition)*, Kluwer Academic / Plenum Publishers, New York (2001) ISBN 0-306-46631-7

[3] R. Fitzpatrick, *Computational Physics : An Introductory Course* http://farside.ph.utexas.edu/teaching/329/lectures/lectures.html

[4] Tao Pang, *An Introduction to Computational Physics*, Cambridge University Press (1997) ISBN 0-521-48592-4

[5] D. Yevick, *A First Course in Computational Physics and Object-Oriented Programming with C++*, Cambridge University Press (2005) ISBN 0-521-82778-7

[6] Ce site recense plusieurs liens vers des références librement accessibles online : http://www.freebookcentre.net/Physics/Computational-Physics-Books.html

[7] F.-J. Elmer, *The Pendulum Lab*, Pendule avec effets nonlinéaires. Inclut un "laboratoire virtuel" de simulation http://www.elmer.unibas.ch/pendulum/index.html

[8] P. Falstad, *Math and Physics Applets*, Nombreux applets de simulations de physique. Très utile pour illuster un cours de physique générale. http://www.falstad.com/mathphysics.html

[9] J. Boris, in *Proceedings of the Fourth Conference on Numerical Simulation of Plasmas* (Naval Research Laboratory, Washington D.C., 1970), p. 3

[10] O. Buneman, Journal of Computers Physics **1**, 517 (1967)

[11] C. Birdsall and A. Langdon, *Plasma Physics Via Computer Simulation* (McGraw-Hill, Inc., New York, 1985), p. 356

[12] G.L. Baker and J.P. Gollub, *Chaotic Dynamics : An Introduction*, Cambridge University Press (1996) ISBN 0521476852

[13] J.N. Reddy, *Introduction to the Finite Element Method*, McGraw-Hill (1993) ISBN 0070513554

[14] O.C.Zienkiewicz, R.L. Taylor and J.Z. Zhu, *The Finite Element Method : Its Basis and Fundamentals (6th Edition)*, Elsevier (2005) ISBN 0-7506-6320-0

[15] W.G. Strang and G.J. Fix, *An Analysis of the Finite Element Method (2nd Edition)*, Wellesley Cambridge (2008) ISBN 0980232708

[16] J.M. Thijssen, *Computational Physics*, Cambridge University Press (1999) ISBN 0521575885

[17] Claude Cohen-Tannoudji, Bernard Diu, Franck Laloë, *Mécanique quantique, Vol.1 et 2*, Hermann (1997) ISBN 2705660747, ISBN 270566 1212

[18] Charles Kittel, *Solid State Physics*, Wiley (1995) ISBN 0471111813

[19] Ashcroft et Mermin, *Physique des solides*, Brooks Cole (2003, version française du livre paru en 1976) ISBN 2868835775

[20] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, *Equation of State Calculations by Fast Computing Machines*, J. Chem. Phys. **21**, 1087 (1953)

[21] D.P. Landau and R. Alben, *Monte Carlo Calculations as an Aid in Teaching Statistical Mechanics*, Am. J. Phys. **41**, 394 (1973)

[22] K. Binder and D.W. Heermann, *Monte Carlo Simulation in Statistical Physics*, Springer-Verlag, New York (1992)

[23] W. Greiner, L. Neise and H. Stöcker, *Thermodynamique et mécanique statistique*, Springer-Verlag, Berlin (1999)

[24] A. Pasquarello, *Simulation numérique de systèmes physiques I - II / Computer simulation of physical systems I - II*, cours à option de Master EPFL

[25] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions : with Formulas, Graphs and Mathemetical Tables*, Dover Publications (1965) ISBN 0486612724