

Notes de cours
Fondements des méthodes numériques (4M006)

M1 - Master mention Mathématiques et Applications

Albert Cohen
Sorbonne Université

15 septembre 2022

Table des matières

1	Introduction	7
1.1	Présentation du cours	7
1.2	Modélisation par les EDP	9
1.3	Quelques solutions explicites	13
1.4	Problèmes aux limites du second ordre en dimension 1	21
1.5	Quelques rappels sur les matrices	29
2	Différences finies pour les problèmes aux limites	35
2.1	Principe de la méthode	35
2.2	Étude de la convergence	41
2.3	Une excursion en dimension 2	47
2.4	La méthode des différences finies en dimension 2	50
3	Différences finies pour les problèmes d'évolution	59
3.1	Approximation de l'équation de la chaleur	59
3.2	Approximation de l'équation de transport	67
3.3	Analyse des schémas	71
4	Formulation variationnelle des problèmes aux limites	81
4.1	La formulation variationnelle : d'où vient-elle ?	81
4.2	Problèmes variationnels abstraits et espaces de Hilbert	84
4.3	Espaces de Sobolev	93
4.4	Application au problème aux limites	104
5	Les méthodes d'approximation variationnelle	109
5.1	Définition et premières propriétés	109
5.2	Forme matricielle de la méthode de Galerkin	112
5.3	Perturbation d'une approximation variationnelle	113
6	La méthode des éléments finis	117
6.1	Définition de la méthode dans le cas dit \mathbb{P}_1	117
6.2	Analyse de la convergence dans le cas \mathbb{P}_1	123
6.3	Éléments finis de degré et régularité plus élevés	127
6.4	Analyse générale de convergence	131
6.5	Compléments	135

7	Méthodes de bases hilbertiennes	141
7.1	Bases hilbertiennes	141
7.2	Bases de Fourier	143
7.3	Bases polynomiales	148
7.4	Bases d'ondelettes	156

Avant-propos

Ces notes reprennent pour l'essentiel celles rédigées par Hervé Le Dret puis par Pascal Frey, Edwige Godlewski, Laurent Boudin, et Sidi-Mahmoud Kaber, qui ont été en charge de ce cours avant moi. Je les remercie de m'avoir ainsi aidé dans la préparation de ce polycopié.

Bibliographie pour les notions prérequis :

- G. Allaire et S. M. Kaber : Algèbre linéaire numérique - Cours et exercices. Ellipses 2002
- G. Allaire, S. M. Kaber : Introduction à Scilab, exercices pratiques corrigés d'algèbre linéaire, Ellipses, 2002
- J.-P. Chancelier, F. Delebecque, C. Gomez, M. Goursat, R. Nikoukhah, S. Steer : Introduction à Scilab, Springer, 2001
- P. G. Ciarlet : Introduction à l'analyse numérique matricielle et à l'optimisation. Dunod, 2006
- M. Crouzeix et A.-L. Mignot : Analyse Numérique des Equations Différentielles, Masson, 1989
- S. Delabrière et M. Postel : Méthodes d'approximation. Equations différentielles, applications Scilab, Ellipses, 2004
- J.-P. Demailly : Analyse numérique et équations différentielles, EDP Sciences, 2006
- P. Lascaux, R. Théodor : Analyse numérique matricielle appliquée à l'art de l'ingénieur, Dunod 2000
- M. Schatzman : Analyse numérique - Une approche mathématique, Dunod 2001.

Pour la suite du cours, on pourra consulter

- G. Allaire : Analyse numérique et optimisation, Editions de l'Ecole Polytechnique, Palaiseau 20
- B. Mohammadi, J.-H. Saïac, Pratique de la simulation numérique, Dunod, 2003
- Laurent Di Menza : Analyse numérique des équations aux dérivées partielles, Cassini 2009
- F. Filbet : Analyse numérique, algorithmes et étude mathématique, Dunod, 2009
- F. Hubert et J. Hubbard, Calcul scientifique 2, de la théorie à la pratique, Vuibert, 2006
- H. Le Dret, B. Lucquin, Partial differential equations : modeling, analysis and numerical approximation, Birkhäuser, 2016
- B. Lucquin : Equations aux dérivées partielles et leurs approximations, Ellipses, 2004
- B. Mohammadi, J.-H. Saïac, Pratique de la simulation numérique, Dunod, 2003
- P.-A. Raviart, J.-M. Thomas : Introduction à l'analyse numérique des équations aux dérivées partielles, Dunod, 2004
- L. Sainsaulieu : Calcul scientifique, Masson enseignement des mathématiques, Dunod, 2000

Chapitre 1

Introduction

“En quelques décennies, les simulations numériques sont devenues un outil privilégié d’investigation dans les sciences et les technologies. Elles ont pour but de reproduire par le calcul le comportement d’un système décrit par un modèle, très souvent constitué d’équations aux dérivées partielles. Ces équations correspondent à la traduction mathématique de lois scientifiques. L’essor des simulations numériques renforce donc la nécessité de l’étude mathématique (analyse) de ces équations et de leur résolution numérique.”

Extrait de la leçon inaugurale de Pierre-Louis Lions au Collège de France (Fayard, 2003)

1.1 Présentation du cours

On va présenter des méthodes numériques de façon très complète sur quelques exemples génériques simples. On n’abordera pas toutes les méthodes (comme celles de résolution de systèmes linéaires, d’optimisation, les méthodes probabilistes, statistiques, ...), mais un certain nombre de celles qui permettent le calcul de solutions approchées d’équations comportant des dérivées partielles, ce qui représente déjà un domaine très important dont nous verrons quelques exemples. Donnons les grandes lignes de la démarche.

On introduit le principe d’une méthode (prenons l’exemple de la méthode des différences finies, nous verrons aussi la méthode des éléments finis, celle des volumes finis); on la décrit précisément sur l’exemple générique à laquelle elle est associée (exemple : le Laplacien pour l’équation de Poisson), cet exemple étant d’abord analysé, pour s’assurer de l’existence et de l’unicité de la solution. Par exemple, pour les différences finies, le *principe* de la méthode est de calculer les valeurs approchées de la solution aux points d’une grille, en approchant les dérivées par des différences finies (nous détaillerons bien sûr plus loin ce que cela signifie); on écrit alors le (ou un) *schéma* qui permet le calcul d’une valeur approchée de la solution (par exemple, on définit un “Laplacien discret”), on en fait l’*analyse numérique*, en étudiant ses propriétés (consistance, stabilité, convergence, estimation de l’erreur); on effectue la *mise en oeuvre* en écrivant un algorithme et un programme informatique qui permet la résolution numérique, c’est à dire le calcul effectif de la solution approchée; le programme est écrit de telle sorte que pour changer le cas d’application, il suffit de changer quelques données en entrée, cela conduit ainsi à un *code*. Le code permet alors de vérifier sur des exemples particuliers (des cas test) dont on connaît la solution exacte les propriétés démontrées théoriquement (l’ordre de

convergence) : de telle manière, on assure la *validation* de la méthode. Lorsqu'il y a plusieurs méthodes ou plusieurs schémas possibles, on les compare sur des cas test. Ces grandes lignes sont les "bases" de l'étude des méthodes numériques.

Les codes utilisés dans la pratique (industrie, recherche appliquée) utilisent des méthodes numériques peut-être plus élaborées mais qui ont été développées à partir d'exemple simplifiés et ont suivi le même processus de validation. Cependant, si la mise en oeuvre sur des cas réels suit cette même démarche, elle est plus compliquée et nécessite d'autres procédures complémentaires (construction du modèle mathématique, choix des variables pour la modélisation, ajustement des paramètres, discrétisation du domaine et choix du modèle numérique, traitement des échelles multiples, comparaison avec les mesures, interprétation des résultats, ...). Par ailleurs, si la taille des cas réels, le nombre de paramètres et de variables, conduit à faire exploser le nombre de calculs à effectuer, il est alors indispensable de développer des méthodes performantes que nous n'aborderons pas, permettant de mener à bien les calculs (décomposition de domaine, calcul parallèle, adaptation de maillages, réduction de modèles, bases réduites,...).

Les systèmes complexes sont modélisés le plus souvent par parties (éléments géométriques ou composants), ou éléments correspondant à un type de fonctionnement, de comportement, de phénomène physique que l'on peut différencier,... qui sont ensuite *couplés* ; chaque "brique élémentaire" suppose une validation, de même que le couplage doit être validé sur des cas test. Les équations ne sont plus des équations génériques aussi simples que celles que nous allons étudier, cependant le processus même de modélisation fait qu'elles rentrent dans de grandes catégories identifiées puisque les équations traduisent des lois "physiques" (celles du domaine d'application), des processus plus ou moins connus (certains doivent encore faire l'objet de nouvelles études approfondies). Un exemple très actuel est celui de la modélisation et la simulation numérique appliquées à différentes problématiques liées au projet de construction du réacteur expérimental ITER.

Le travail du mathématicien appliqué intervient dans le processus de modélisation (en relation avec la discipline d'application ou le métier), pour traduire le phénomène observé par des équations, dans l'analyse des équations, du comportement des solutions (par rapport au temps, leur dépendance par rapport à des paramètres, des données,...), le couplage, dans le choix des méthodes de discrétisation, l'analyse numérique de ces méthodes et aussi dans la mise en oeuvre. Il est donc utile d'illustrer toute la chaîne de cette activité sur des exemples bien compris, avant de suivre cette démarche sur les cas complexes, d'aborder des cas "industriels", ou encore d'envisager la définition de nouvelles méthodes numériques. Mais le domaine de recherche et d'applications se développe continuellement : on étudie de nouveaux contextes, de nouveaux couplages multiphysiques, de nouvelles méthodes dont le comportement est robuste par rapport à un changement d'échelle (passage à limite dans l'équation quand un paramètre devient très petit), on veut traiter les problèmes en grande dimension, on cherche à contrôler la propagation des incertitudes sur les données, introduire l'incertitude sur les modèles....

Nous allons donc présenter les principales méthodes numériques utilisées dans le contexte de l'approximation des EDP et en détailler l'étude dans le cas de la dimension 1. Nous supposerons connues les méthodes classiques d'interpolation par un polynôme, celles permettant d'approcher une dérivée en un point (différences finies), d'approcher une intégrale (formules de quadrature), ou de calculer de façon approchée la solution d'une EDO (équation différentielle ordinaire), comme la méthode d'Euler. L'algorithme et la mise en oeuvre seront détaillées en travaux pratiques.

1.2 Modélisation par les EDP

Ce cours se propose donc d'introduire un certain nombre de méthodes d'approximation numérique des solutions de problèmes d'équations aux dérivées partielles (EDP). Il s'agit d'un exposé relativement élémentaire et les exemples traités seront essentiellement des exemples en dimension 1 d'espace, ce qui relativise singulièrement le terme *dérivée partielle*, bien sûr. Les "vrais" problèmes d'EDP, lesquels sont posés en dimension d'espace 2, 3 (voire plus encore), et peuvent aussi faire intervenir une variation par rapport au temps, nécessitent des concepts et des techniques mathématiques qui se situent largement au delà des limites de ce cours. Mais pourquoi s'intéresse-t-on à l'approximation des solutions d'EDP ?

Il se trouve que la *modélisation* de très nombreux problèmes issus de la physique, de la mécanique et des sciences de l'ingénieur, font qu'ils se présentent sous forme d'EDP. Plus précisément, bon nombre de phénomènes physiques font intervenir des grandeurs qui sont, plus ou moins approximativement, solution (de systèmes) d'EDP. Donnons en quelques exemples.

Si on cherche à modéliser l'évolution d'un polluant dans un milieu donné (atmosphère, océan...), la grandeur qui nous intéresse est la concentration en polluant en fonction de la position $x \in \mathbb{R}^3$ et de l'instant $t \in \mathbb{R}$. Notons

$$(x, t) \mapsto u(x, t),$$

cette fonction. Afin d'étudier son évolution dans le temps, considérons un domaine $D \subset \mathbb{R}^3$ quelconque à frontière régulière (par exemple une boule). La quantité de polluant contenue dans D au temps t est donc donnée par l'intégrale

$$\int_D u(x, t) dx.$$

En l'absence de production de polluant à l'intérieur de D , l'évolution de cette quantité dépend de ce qui entre ou sort de D . Ceci se modélise par un *flux* c'est à dire une fonction vectorielle

$$(x, t) \mapsto q(x, t) = (q_1(x, t), q_2(x, t), q_3(x, t)),$$

telle que

$$\frac{d}{dt} \left(\int_D u(x, t) dx \right) = - \int_{\partial D} q(x, t) \cdot n(x) ds.$$

Dans cette *équation de bilan*, le second membre est une intégrale de surface sur la frontière de ∂D de D et la quantité intégrée est le "flux sortant", c'est à dire le produit scalaire du flux avec le vecteur normal unitaire dans la direction extérieure au domaine, qu'on a noté $n(x)$. On comprend par exemple que si le flux était tangent à la frontière domaine, rien ne rentre ni ne sort et la quantité reste constante.

En supposant que u et q sont suffisamment régulières, on peut appliquer la formule de la divergence pour obtenir

$$\frac{d}{dt} \left(\int_D u(x, t) dx \right) = - \int_D \operatorname{div}(q(x, t)) dx, \quad (1.1)$$

où $\operatorname{div}(q) = \frac{\partial q_1}{\partial x_1} + \frac{\partial q_2}{\partial x_2} + \frac{\partial q_3}{\partial x_3}$. On peut aussi faire rentrer la dérivée en temps dans l'intégrale ce qui donne

$$\int_D \left(\frac{\partial}{\partial t} u(x, t) + \operatorname{div}(q(x, t)) \right) = 0.$$

Finalement, on remarque que pour tout fonction continue v , si on a

$$\int_D v(y) dy = 0.$$

alors v est uniformément nulle (exercice), ce qui nous montre que u et q sont liées par l'équation aux dérivées partielles

$$\frac{\partial u}{\partial t} + \operatorname{div}(q) = 0.$$

On utilise parfois la notation plus légère $\partial_t u$ pour $\frac{\partial u}{\partial t}$.

On va aboutir à une équation posée uniquement sur u en introduisant une *loi de comportement* qui décrit la dépendance de q en fonction de u . Celle-ci dépend du phénomène physique étudié. Lorsque le fluide est en mouvement, en notant $b(x) = (b_1(x), b_2(x), b_3(x))$ sa vitesse au point x , le flux est alors donné par le produit de ce champs de vitesse par la quantité de polluant, c'est à dire

$$q = bu$$

On arrive ainsi à l'équation dite de *transport* ou *convection* :

$$\partial_t u + \operatorname{div}(bu) = 0.$$

On rappelle que $\operatorname{div}(bu) = \operatorname{div}(b)u + b \cdot \nabla u$. Ainsi, si b est tel que $\operatorname{div}(b) = 0$ ce qui correspond à un fluide incompressible (par exemple dans le cas d'un champ de vitesse constant) l'équation prend la forme

$$\partial_t u + b \cdot \nabla u = 0.$$

Une autre source possible de flux est la diffusion du polluant qui peut se produire même en l'absence de mouvement du fluide. La loi de Fick indique que le flux est négativement proportionnel au gradient de u , c'est à dire

$$q = -\sigma \nabla u,$$

où $\sigma > 0$ ce qui intuitivement signifie que le polluant se déplace des zones où il est le plus concentré vers celles où il l'est moins. La constante σ décrit la diffusivité du fluide. Plus généralement σ peut varier en fonction du point x lorsque cette diffusivité n'est pas uniforme. On obtient ainsi l'équation de la *diffusion* qui s'écrit

$$\partial_t u - \operatorname{div}(\sigma \nabla u) = 0,$$

qui dans le cas où σ est une constante s'écrit aussi

$$\partial_t u - \sigma \Delta u = 0, \tag{1.2}$$

où $\Delta u = \operatorname{div}(\nabla u) = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \frac{\partial^2 u}{\partial x_3^2}$ est le Laplacien.

On peut bien entendu considérer les deux phénomènes en même temps, ce qui donne l'équation de transport -diffusion

$$\partial_t u + b \cdot \nabla u - \sigma \Delta u = 0,$$

Il faut bien noter que dans ces équations, la seule inconnue est u : les grandeurs b et σ sont des paramètres physiques que l'on suppose connus.

Dans le cas d'injection de polluant dans certaines régions du fluide, il faut modifier l'équation de bilan (1.1) en ajoutant au second membre un terme de la forme $\int_D f(x, t) dx$ où la fonction f représente la densité de polluant créée à l'instant t et au point x . Par le même raisonnement, cela nous conduit à une équation de la forme

$$\partial_t u + b \cdot \nabla u - \sigma \Delta u = f.$$

On dit que f est un *terme source*. Une autre source d'évolution de la quantité polluant peut être un phénomène de réaction qui sera modélisé en ajoutant au second membre un terme de la forme $-\int_D c(x, t)u(x, t) dx$ où la fonction c représente le taux de réaction à l'instant t et au point x . On arrive ainsi à une équation de transport-diffusion-réaction avec terme source.

$$\partial_t u + b \cdot \nabla u - \sigma \Delta u + cu = f.$$

Bien entendu, on peut considérer des modèles incluant uniquement une partie de ces termes, diffusion avec terme source, transport-réaction, etc.

Le propre des mathématiques est de permettre de modéliser des phénomènes physiques de natures différentes par des équations qui sont parfois du même type. Ainsi l'équation de diffusion (1.2) modélise aussi l'évolution de la température lorsque la chaleur se diffuse dans un milieu de conductivité thermique σ , ce qui explique qu'elle soit souvent appelée *équation de la chaleur*.

Il faut bien noter les analogies et différences avec les équations différentielles ordinaires (EDO) qui décrivent l'évolution d'une grandeur scalaire $u(t) \in \mathbb{R}$ suivant

$$u'(t) = F(t, u(t)), \quad t > 0, \quad u(0) = u_0 \in \mathbb{R}$$

où F est une fonction donnée. Dans les EDP que nous venons de décrire, on décrit l'évolution en temps d'une fonction toute entière $u(\cdot, t)$ qui représente la répartition du polluant (ou de la température pour l'équation de la chaleur) au temps t . Comme pour les EDO, il est naturel d'introduire une condition initiale

$$u(\cdot, 0) = u_0,$$

c'est à dire $u(x, 0) = u_0(x)$ pour tout x , où u_0 est une fonction donnée, et de considérer l'EDP uniquement pour $t > 0$. On parle alors de *problème de Cauchy*.

En pratique le problème est souvent posé sur un domaine borné $\Omega \subset \mathbb{R}^d$ et il est alors nécessaire de spécifier les *conditions aux limites* qu'on impose sur u à la frontière $\partial\Omega$ du domaine. Les plus fréquemment rencontrées sont les suivantes :

1. *Conditions de Dirichlet* : on impose la valeur

$$u|_{\partial\Omega} = g,$$

où g est une fonction donnée définie sur $\partial\Omega$. Dans le cas de l'équation de la chaleur cela signifie qu'on impose la valeur de la température au bord du domaine.

2. *Conditions de Neumann* : en notant $n(x)$ le vecteur normal extérieur à la frontière en un point $x \in \partial\Omega$, on définit sur $\partial\Omega$ la fonction $\frac{\partial u}{\partial n} = \nabla u \cdot n$ appelée *dérivée normale*, et on impose sa valeur

$$\frac{\partial u}{\partial n} = h,$$

où h est une fonction donnée définie sur $\partial\Omega$. Dans le cas de l'équation de la chaleur cela signifie qu'on impose le flux de chaleur au bord du domaine.

Lorsque g ou h sont identiquement nulles (ce qui correspond à des conditions isothermes ou adiabatiques pour la chaleur), on parle de conditions homogènes.

On s'intéresse aussi aux équations qui décrivent un *régime stationnaire*, c'est à dire lorsque la solution u ne dépend plus du temps t . On obtient ainsi pour la fonction $x \mapsto u(x)$ des équations de la forme générale

$$b \cdot \nabla u - \sigma \Delta u + cu = f,$$

assorties de conditions aux limites au bord du domaine considéré. Le cas particulier de l'équation du *Laplacien*

$$-\Delta u = f, \tag{1.3}$$

est fondamental, tant pour l'étude mathématique que pour la multiplicité des phénomènes que cette équation peut représenter en physique :

1. Diffusion : concentration de polluant u à l'équilibre en présence d'une source f .
2. Chaleur : température à l'équilibre en présence d'une source f .
3. Mécanique : déplacement d'une membrane élastique sous l'action d'une force f .
4. Electromagnétisme : potentiel électrique créée par une distribution de charge f .
5. Gravitation : potentiel gravitationnel créée par une distribution de masse f .

Pour être plus précis, il faut bien sûr préciser les conditions aux limites et les unités de mesure physiques utilisées.

Certains problèmes font intervenir la dérivée seconde (accélération) en temps. C'est le cas pour l'équation des ondes

$$\frac{\partial^2 u}{\partial t^2} - c^2 \Delta u = 0,$$

qui décrit par exemple l'évolution du déplacement d'une membrane élastique (un tambour), soumise à des conditions au bord, où la constante c représente la vitesse de propagation. Comme pour les EDO impliquant la dérivée seconde en temps, le problème de Cauchy doit alors spécifier deux conditions initiales : la valeur initiale $u(\cdot, 0) = u_0$ et la vitesse initiale $\partial_t u(\cdot, 0) = u_1$.

L'ordre d'une edp est l'ordre de la plus grande dérivée présente dans l'équation : par exemple l'équation des ondes est d'ordre deux en espace et en temps, alors que l'équation de la chaleur est d'ordre deux en espace et un en temps, l'équation de transport est d'ordre 1 en espace et en temps.

Les équations qu'on a décrit ci-dessus sont linéaires : u dépend *linéairement* de la donnée f . Elles s'écrivent sous la forme $Lu = f$ où l'opérateur L est un opérateur aux dérivées partielles linéaires. Cela se traduit par des propriétés importantes, telles que le *principe de superposition* : la solution u pour une donnée $f = f_1 + f_2$ au second membre est la somme des solutions u_1 et u_2 pour les données f_1 et f_2 (il faut aussi sommer les données des conditions initiales et conditions aux limites). On voit aussi que l'unicité de la solution est équivalente au fait que la seule solution de $Lu = 0$ est la solution nulle. Bien entendu, de nombreuses EDP rencontrées en physique sont non-linéaires ce qui rend leur étude souvent plus difficile. L'étude des EDP linéaires est une première étape fondamentale, aussi bien d'un point de vue théorique que pratique car ces équations peuvent parfois être utilisées comme des modèles approchés d'équations non-linéaires.

On utilise souvent le vocabulaire suivant pour caractériser la nature des EDP d'ordre inférieur ou égal à 2 : elliptique (exemple : équation de Laplace, de Poisson), parabolique (exemple :

équation de la chaleur), hyperbolique (exemple : équation des ondes). Pour deux variables (avec des variables notées x, y), l'opérateur L a la forme générale suivante

$$Lu \equiv a \frac{\partial^2 u}{\partial x^2} + b \frac{\partial^2 u}{\partial x \partial y} + c \frac{\partial^2 u}{\partial y^2} + d \frac{\partial u}{\partial y} + e \frac{\partial u}{\partial x} + a_0$$

En supposant ici les coefficients constants, on regarde le symbole $P(x, y) = ax^2 + bxy + cy^2 + dx + ey + a_0$ et le signe de $b^2 - 4ac$. Le cas dit elliptique correspond à $b^2 - 4ac < 0$, parabolique à $b^2 - 4ac = 0$, hyperbolique à $b^2 - 4ac > 0$.

Un problème est dit *bien posé* si pour toute donnée (c'est à dire la source f et les conditions initiales et/ou aux limites), il possède une unique solution qui dépend continûment des données (ce qui suppose avoir bien précisé sur quel espace de fonctions on définit l'opérateur L). Les trois catégories précédentes correspondent à des problèmes bien posés à condition d'avoir bien défini le type de conditions initiales et/ou aux limites.

1.3 Quelques solutions explicites

Dans toute la suite de ce cours on va se placer en dimension 1 d'espace, c'est à dire $x \in \mathbb{R}$ ce qui simplifie considérablement les problèmes considérés, et permet dans certains cas d'avoir accès à la forme explicite des solutions.

Considérons tout d'abord l'équation de transport qui prend la forme

$$\frac{\partial u}{\partial t} + b \frac{\partial u}{\partial x} = 0, \quad (1.4)$$

ou avec les notations allégées

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

et qu'on suppose ici posée sur tout \mathbb{R} avec b constant indépendant de x et t , et avec donnée initiale

$$u(x, 0) = u_0(x), \quad x \in \mathbb{R}.$$

Il est immédiat de vérifier que pour toute fonction $x \mapsto v(x)$ dérivable, la fonction

$$u(x, t) := v(x - bt),$$

est solution de l'équation, puisqu'on a alors

$$\partial_t v(x, t) = -bv'(x - bt) = -b \partial_x v(x, t).$$

La condition initiale nous indique qu'il faut prendre $v = u_0$. Par conséquent, la solution est donnée par

$$u(x, t) = u_0(x - bt).$$

Visuellement, on voit que le graphe de la fonction $x \mapsto u(x, t)$ se translate à la vitesse $|b|$ dans le sens du signe de b ce qui correspond bien au transport d'un polluant par un flot de vitesse constante, voir la figure 1.1.

Un argument plus complet que nous laissons en exercice permet de montrer que c'est l'unique solution. Indication : le changement, linéaire, de variable et d'inconnues

$$X = x - bt, \quad T = t, \quad U(X, T) = u(x, t)$$

transforme l'équation en $\frac{\partial U}{\partial T} = 0$.

Une propriété importante est que

$$\max_{x \in \mathbb{R}} u(x, t) = \max_{x \in \mathbb{R}} u_0(x)$$

Le maximum en x de la fonction reste conservé. Cette propriété reste vérifiée lorsque la vitesse de propagation b n'est pas constante : c'est le *principe du maximum* pour l'équation de transport. On note aussi que, contrairement au cas de l'équation de la chaleur qu'on verra plus loin, la régularité de la solution initiale est conservée : si u_0 est dérivable k fois, et pas dérivable $k + 1$ fois, il en est de même des fonctions $t \mapsto u(x, t)$ et $x \mapsto u(x, t)$.

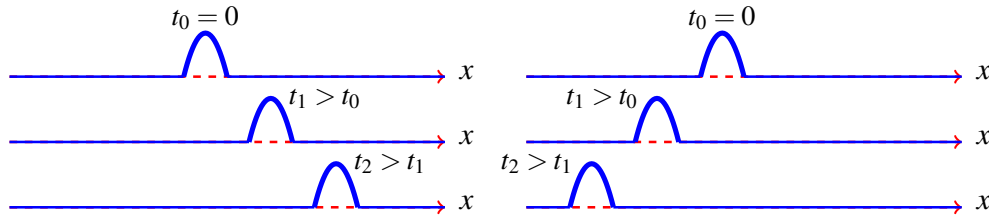


FIGURE 1.1 – Solution du problème de Cauchy pour l'équation de transport (1.4). Exemple d'évolution d'une donnée initiale à support compact. Cas $b > 0$ (gauche) et $b < 0$ (droite).

Comme on peut déterminer explicitement la solution de l'équation (1.4), on peut se demander, à juste titre, pourquoi alors définir des méthodes numériques pour l'approcher. Il se trouve que ces méthodes pourront se généraliser (plus ou moins directement) aux équations de transport non linéaires de la forme

$$\begin{cases} \frac{\partial u}{\partial t} + \frac{\partial}{\partial x} F(u) = 0, & x \in \mathbb{R}, \quad t > 0 \\ u(x, 0) = u_0(x), & x \in \mathbb{R}, \end{cases}$$

où F est une fonction donnée, appelée *flux*. Dans l'équation de Burgers, voir l'exemple 2, $F(u) = u^2/2$. Dans le cas linéaire, le flux est $F(u) = bu$, on retrouve l'équation (1.4) si a est constant (indépendant de x et t). Les équations linéaires servent alors de cas test pour valider les méthodes utilisées dans des cas plus généraux des équations non linéaires, voire des systèmes d'EDP non linéaires. En effet, une première idée pour commencer à étudier ces équations consiste à les linéariser autour d'un état constant, même si cette approche ne suffit pas, bien sûr, à comprendre les phénomènes non linéaires.

Introduisons à présent un outil fondamental pour l'étude des équations de transport, en particulier lorsque la vitesse b est non-constante.

Définition 1.3.1 On appelle caractéristique de l'EDP linéaire

$$\frac{\partial u}{\partial t} + b(x, t) \frac{\partial u}{\partial x} = 0, \quad (E')$$

issue du point x_0 au temps $t = 0$ et on note C_{x_0} , la courbe paramétrée $t \mapsto (X(t), t)$ où X est solution de l'Équation Différentielle Ordinaire (EDO)

$$\begin{cases} X'(t) = b(X(t), t) \\ X(0) = x_0. \end{cases}$$

On supposera toujours que ce problème a une solution unique (par le théorème de Cauchy-Lipschitz).

Proposition 1.3.1 Soit $u \in \mathcal{C}^1(\mathbb{R} \times [0, \infty))$ une solution de (E') . La restriction de u à une caractéristique est constante.

Démonstration. Soit u une fonction de classe \mathcal{C}^1 et $\varphi(t) = u(X(t), t)$ sa restriction à une caractéristique. La variation de $\varphi(t)$ est

$$\frac{d\varphi}{dt}(t) = \left\{ \frac{\partial u}{\partial x} \frac{dX}{dt} + \frac{\partial u}{\partial t} \right\}(X(t), t) = \left\{ \frac{\partial u}{\partial t} + b \frac{\partial u}{\partial x} \right\}(X(t), t)$$

donc si u est solution de (E') , on a $\varphi' = 0$ et $\varphi(t) = \varphi(t_0) = u(x_0, 0) = u_0(x_0)$. \square

Retour à notre équation modèle (1.4) dont les caractéristiques sont des droites ; la caractéristique passant par x_0 à l'instant $t = 0$ a pour équation dans le plan (x, t)

$$x = x_0 + bt.$$

Le point x_0 est appelé “pied” de la caractéristique, voir la figure 1.2.

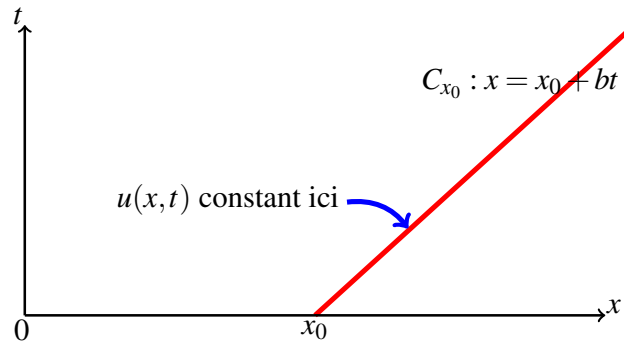


FIGURE 1.2 – Une caractéristique de l'équation (1.4).

La méthode des caractéristiques permet, parfois, de déterminer explicitement la solution de certaines EDP du premier ordre. Plus précisément d'EDP quasilineaires (c'est-à-dire linéaires en les dérivées de la fonction et éventuellement non linéaires en la fonction elle-même) et du premier ordre (ne faisant intervenir que des dérivées premières). Voici un exemple.

Exemple 1 Considérons l'EDP linéaire, mais à coefficients variables

$$\frac{\partial u}{\partial t} + x \frac{\partial u}{\partial x} = 0, \quad x \in \mathbb{R}, \quad t > 0. \quad (1.5)$$

Il est facile de voir que les caractéristiques de cette EDP sont des exponentielles $X(t) = x_0 e^t$. Voir la figure 1.3. On en déduit que la solution du problème de Cauchy associé est $u(x, t) = u_0(xe^{-t})$. Rien qu'en traçant les caractéristiques, et sans calculer la solution explicitement, on voit que, pour tout x , $\lim_{t \rightarrow +\infty} u(x, t) = u_0(0)$. Pourquoi ?

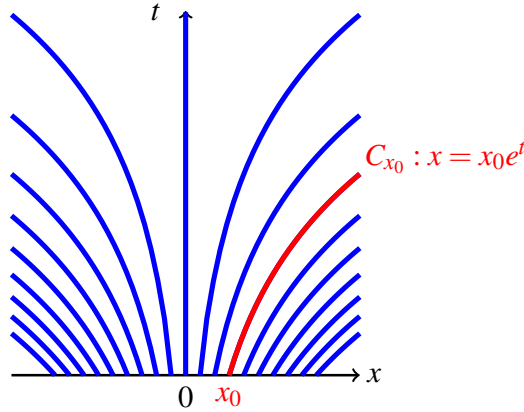


FIGURE 1.3 – Caractéristiques de l'équation (1.5).

La méthode des caractéristiques peut être efficace aussi pour des équations non linéaires de transport, *i.e.*, équations dont la vitesse de propagation dépend de la solution u . Dans ce cas l'équation s'écrit

$$\frac{\partial u}{\partial t} + b(x, t, u) \frac{\partial u}{\partial x} = 0, \quad (E')$$

et l'équation différentielle de la caractéristique C_{x_0} issue du point x_0 est donnée par

$$\begin{cases} X'(t) = b(X(t), t, u(X(t), t)) \\ X(0) = x_0. \end{cases}$$

Le même calcul que dans le cas linéaire nous montre que $u(X(t), t)$ est constant le long de la caractéristique, et donc égal à $u_0(x_0)$. Ceci montre que l'équation de la caractéristique s'écrit de la manière plus simple

$$\begin{cases} X'(t) = b(X(t), t, u_0(x_0)) \\ X(0) = x_0. \end{cases}$$

Nous pouvons appliquer cette méthode à l'équation de Burgers qui est un exemple modèle d'équation de transport non linéaire.

Exemple 2 (Équation de Burgers)

$$\begin{cases} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0, & x \in \mathbb{R}, \quad t > 0, \\ u(x, 0) = u_0(x), & x \in \mathbb{R}. \end{cases} \quad (1.6)$$

Les caractéristiques de cette équation, passant à l'instant initial par x_0 , sont définies par

$$\begin{cases} X'(t) = u(X(t), t) = u(X(0), 0) = u_0(x_0) \\ X(0) = x_0. \end{cases}$$

Comme u est constant le long des caractéristiques, on a $X'(t) = u(X(0), 0) = u_0(x_0)$. Les caractéristiques sont donc des droites d'équations

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

Pour déterminer $u(x, t)$, on note que

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

Il suffit donc de résoudre l'équation (dont l'inconnue est $z = u(x, t)$)

$$z = u_0(x - zt).$$

Cette équation peut avoir une seule solution (c'est le cas favorable), ne pas en avoir ou en avoir plusieurs ! Voir la figure 1.4. Cette situation se produit typiquement au delà d'un temps critique t^* au delà duquel la solution cesse d'être régulière et développe des discontinuités appelées ondes de choc. Dans ce dernier cas, la méthode des caractéristiques doit être combinée à d'autres techniques pour sélectionner parmi toutes solutions possibles, la bonne. Nous n'aborderons pas ce problème difficile ici dont le niveau dépasse le niveau de ce cours.

Fin de l'excursion non linéaire et retour à l'équation linéaire modèle (1.4). On s'intéresse maintenant au problème posé dans un intervalle borné en espace, par exemple l'intervalle $[0, 1]$. On suppose que $b > 0$. Les caractéristiques de l'EDP sont présentées sur la figure 1.5. Pour $x \in [0, 1]$, on peut calculer $u(x, t)$ pour tout $t \leq x/b$ en "remontant" les caractéristiques

$$u(x, t) = u_0(x - bt).$$

Pour $t > x/b$, on ne peut plus calculer $u(x, t)$ par la méthode des caractéristiques car le pied de la caractéristique issue de (x, t) est situé à l'extérieur de $[0, 1]$.

Avec des conditions aux limites périodiques

$$u(x + 1, t) = u(x, t),$$

on peut considérer que le problème est posé sur tout l'espace. La restriction précédente sur le calcul de $u(x, t)$ est levée : on peut déterminer la solution du problème en tout point x et pour tout temps t .

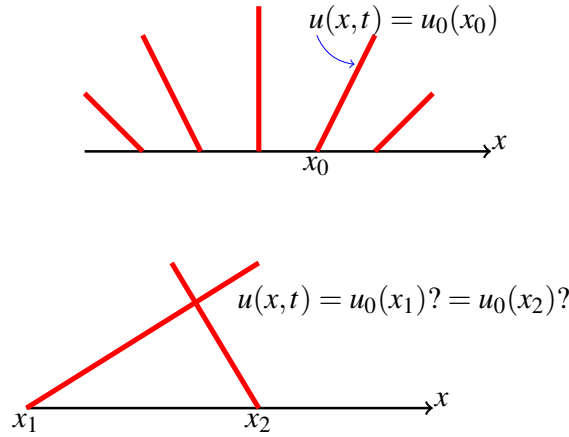


FIGURE 1.4 – Caractéristiques de l'équation de Burgers (1.6). Figure du haut : cas d'une donnée initiale croissante en prenant la valeur nulle. En chaque point du plan (x, t) , passe une et une seule caractéristique. Figure du bas : cas d'une donnée initiale décroissante. Les caractéristiques issues de x_1 et x_2 se coupent, la valeur de $u(x, t)$ n'est donc pas définie de façon univoque.

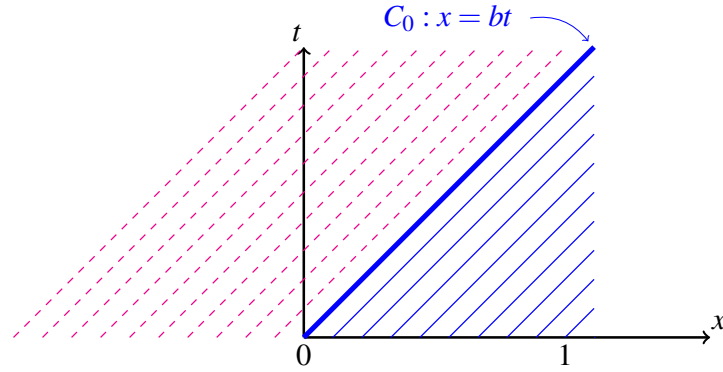


FIGURE 1.5 – Caractéristiques de l'équation de transport (1.4) dans le cas $b > 0$.

Revenons aux conditions aux limites non-périodiques, en supposant encore $b > 0$. Peut-on imposer une condition à la limite, à droite du domaine, en $x = 1$? La réponse est “non” car cela entrerait en conflit avec le fait que, pour $t \leq 1/b$, le pied de la caractéristique issue du point $(1, t)$, noté x_0 , se trouve dans l'intervalle $[0, 1]$ et on doit avoir dans ce cas $u(x, 1) = u_0(x_0)$.

Peut-on imposer une condition à la limite, à gauche du domaine, en $x = 0$? La réponse est “oui” et cela permettra de déterminer $u(x, t)$ comme nous allons le voir. Imposons donc la condition à la limite

$$u(0, t) = u_G(t), \quad t > 0 \quad (1.7)$$

où u_G est une fonction donnée. Pour $t > x/b$, on a $u(x, t) = u_G(\tau)$ où τ est déterminé par l'intersection de la caractéristique issue de (x, t) et l'axe $\{x = 0\}$. On a donc

$$u(x, t) = u_G(t - x/b).$$

Conclusion : $u(x, t)$ est parfaitement déterminé (voir la figure 1.6) :

$$u(x, t) = \begin{cases} u_0(x - bt) & 0 \leq t \leq x/a, \\ u_G(t - x/b) & t > x/b. \end{cases}$$

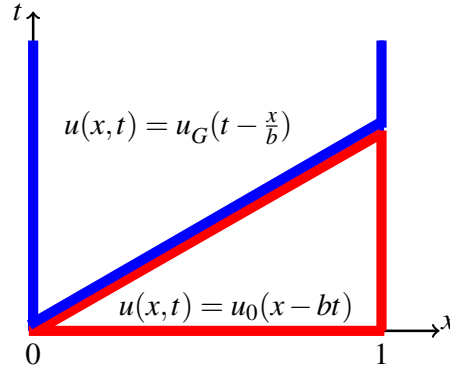


FIGURE 1.6 – Solution de l'équation de transport (1.4) par la méthode des caractéristiques dans le cas $b > 0$. La fonction u_G est la condition à la limite en $x = 0$. La droite $y = bx$ divise le "cylindre" $[0, 1] \times [0, +\infty[$ en deux parties : dans l'une, u est déterminé par la condition initiale, dans l'autre, u est déterminé par la condition à la limite.

Remarque : pour $b < 0$, on impose une condition à la limite, à droite du domaine, en $x = 1$:

$$u(x, t) = \begin{cases} u_0(x - bt) & 0 \leq t \leq (x - 1)/b \\ u_D(t + (1 - x)/b) & t > (x - 1)/b \end{cases}$$

Considérons à présent l'équation des ondes

$$\frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} = 0,$$

On vérifie que pour toute paire de fonctions (v, w) deux fois différentiables, la fonction

$$u(x, t) := v(x - ct) + w(x + ct),$$

est solution sur tout \mathbb{R} (il suffit pour cela de vérifier que chacun des termes est solution, puisque l'équation est linéaire et avec second membre nul). Comme pour l'équation de transport, les conditions initiales permettent d'identifier v et w .

Considérons finalement l'équation de la chaleur

$$\frac{\partial u}{\partial t} - \sigma \frac{\partial^2 u}{\partial x^2} = 0,$$

qu'on suppose ici posée sur tout \mathbb{R} avec $\sigma > 0$ constant indépendant de x et t , et avec donnée initiale

$$u(x, 0) = u_0(x), \quad x \in \mathbb{R}.$$

On suppose que u_0 est continue et uniformément bornée. La solution est alors donnée pour tout $t > 0$ par le produit de convolution

$$u(x, t) = \int_{-\infty}^{+\infty} g_t(x-y)u_0(y)dy = (u_0 * g_t)(x)$$

où g_t est la fonction gaussienne

$$g_t(x) := \frac{1}{\sqrt{4\pi\sigma t}} \exp\left(-\frac{x^2}{4\sigma t}\right).$$

Pour vérifier que l'équation est vérifiée lorsque u est définie par la formule ci-dessous, il suffit de calculer les dérivées $\partial_t u$ et $\frac{\partial^2 u}{\partial x^2}$ en faisant passer la dérivée dans l'intégrale : on trouve ainsi

$$\frac{\partial^2 u}{\partial x^2}(x, t) = \int_{-\infty}^{+\infty} \left(\frac{(x-y)^2}{4\sigma^2 t^2} g_t(x-y) - \frac{1}{2\sigma t} g_t(x-y) \right) u_0(y) dy,$$

et

$$\partial_t u(x, t) = \int_{-\infty}^{+\infty} \left(\frac{(x-y)^2}{4\sigma t^2} g_t(x-y) - \frac{1}{2t} g_t(x-y) \right) u_0(y) dy,$$

ce qui montre que l'équation est vérifiée.

Pour montrer que la condition initiale est vérifiée, étudions la limite de l'intégrale de convolution lorsque $t \rightarrow 0$. On remarque tout d'abord que g_t a la propriété d'*approximation de l'identité* lorsque t tends vers 0 :

(i) g_t est positive et $\int_{\mathbb{R}} g_t = 1$.

(ii) pour tout $\delta > 0$ on a $\lim_{t \rightarrow 0} \int_{|x| > \delta} g_t(x) dx = 0$

Autrement dit, l'intégrale de g_t est de plus en plus concentrée autour de l'origine lorsque $t \rightarrow 0$. En utilisant l'hypothèse que u_0 est continue et uniformément bornée sur \mathbb{R} , ceci permet d'établir (exercice) que

$$\lim_{t \rightarrow 0} u_0 * g_t(x) = u_0(x), \quad x \in \mathbb{R},$$

ce qui montre que la condition initiale est bien vérifiée. Puisque g_t est positive et d'intégrale égale à 1, on voit que pour tout $x \in \mathbb{R}$

$$u(x, t) \leq \max_{y \in \mathbb{R}} u_0(y) \int_{\mathbb{R}} g_t(x-y) dy = \max_{y \in \mathbb{R}} u_0(y),$$

autrement dit

$$\max_{x \in \mathbb{R}} u(x, t) \leq \max_{x \in \mathbb{R}} u_0(x),$$

principe du maximum pour l'équation de la chaleur. De la même manière on obtient que $\min_{x \in \mathbb{R}} u(x, t) \geq \min_{x \in \mathbb{R}} u_0(x)$.

On remarque que pour tout $t > 0$ la fonction $v := u(\cdot, t) = u_0 * g_t$ est dérivable puisqu'on peut dériver la gaussienne à l'intérieur de l'intégrale de convolution, ce qui donne $v' = u_0 * g'_t$. Par itération, on voit que $u(\cdot, t)$ est infiniment dérivable, même si u_0 n'était que continue : c'est l'effet régularisant instantané de l'équation de la chaleur.

A titre d'exercice, le lecteur pourra chercher une formule de convolution permettant de décrire la solution de l'équation de transport-diffusion

$$\frac{\partial u}{\partial t} + b \frac{\partial u}{\partial x} - \sigma \frac{\partial^2 u}{\partial x^2} = 0,$$

en essayant de combiner formule pour l'équation de la chaleur avec la translation due au transport.

Pour de nombreuses EDP en dimension supérieure mais aussi en dimension 1, les solutions explicites ne sont malheureusement pas disponibles. On n'a par conséquent pas d'autre choix que de calculer des *approximations numériques* de ces solutions. On distingue ainsi plusieurs étapes :

1. Analyse théorique : comprendre si le problème est bien posé (existence et unicité de la solution, dépendance continue de la cette solution en fonction des données).
2. Mise au point et analyse numérique : les méthodes d'approximation considérées font intervenir un paramètre de discrétisation, car on ne peut en pratique calculer effectivement qu'un *nombre fini* de valeurs numériques. Ce paramètre prend typiquement la forme d'un pas d'espace δx ou h (et d'un pas de temps δt dans le cas d'un problème d'évolution). Il s'agit de quantifier la convergence de l'approximation vers la solution exacte quand $h \rightarrow 0$. En d'autres termes, on veut démontrer que l'on a bien déterminé des approximations de la solution, et cela en un sens précis.
3. Implémentation effective des méthodes sur un ordinateur. Se posent alors des questions d'algorithmique, de complexité et de programmation, questions que nous n'aborderons pas en détail ici.

Finalement, remarquons qu'il reste pertinent de tester les méthodes numériques sur des problèmes simples dont on connaît la solution exacte : on peut ainsi observer exactement l'erreur commise, et la comparer à l'erreur prévue par l'analyse de convergence.

1.4 Problèmes aux limites du second ordre en dimension 1

Par la suite, nous nous intéresserons essentiellement aux problèmes en dimension 1 d'espace. L'analogue du problème du laplacien (1.3) avec conditions de Dirichlet homogènes en dimension 1 est le suivant : on prend $\Omega =]0, 1[$ (l'intervalle $]0, 1[$ est habituellement choisi comme intervalle de référence, la généralisation à un intervalle quelconque est immédiate), on se donne $f:]0, 1[\rightarrow \mathbb{R}$ et l'on cherche $u:]0, 1[\rightarrow \mathbb{R}$ telle que

$$-u''(x) = f(x) \text{ dans }]0, 1[\quad \text{et} \quad u(0) = u(1) = 0. \quad (1.8)$$

Notons qu'ici la frontière est formée de deux points, $\partial\Omega = \{0, 1\}$. Considérant le segment Ω comme un sous-ensemble de \mathbb{R}^2 qui représente un fil élastique tendu entre deux points et si f modélise une densité de forces verticales que l'on applique sur ce fil divisée par la tension du fil, alors $u(x)$ sera le déplacement vertical que subit le point x du fil à l'équilibre (voir la figure 1.7). Il est très important de bien réaliser que ce *problème aux limites* est très différent du *problème*

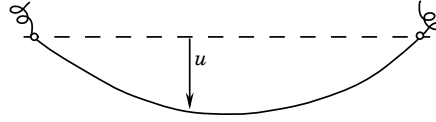


FIGURE 1.7 – Le fil élastique

de Cauchy pour l'équation différentielle ordinaire (EDO) :

$$-u''(x) = f(x) \text{ dans }]0, 1[\quad \text{et} \quad u(0) = \alpha, u'(0) = \beta. \quad (1.9)$$

Dans le problème de conditions initiales, on se donne des *valeurs initiales*, notées ici α, β , pour u et u' au même point 0. Il faut penser plutôt dans ce cas à une évolution d'un système après un instant initial, la variable x s'interprétant alors comme le temps, $u(x)$ étant la position à l'instant x , $u'(x)$ la vitesse et $u''(x)$ l'accélération. Par conséquent, $u(0)$ et $u'(0)$ sont respectivement la position et la vitesse initiales.

Dans le problème aux limites, on se donne au contraire la valeur de u aux *deux extrémités* de l'intervalle et la variable x s'interprète plutôt comme une variable d'espace, comme le montre l'exemple du fil élastique ci-dessus. En raison de cette différence essentielle de nature, les méthodes d'analyse des EDO (équations différentielles ordinaires) ne s'appliquent pas directement et les méthodes d'analyse numérique adaptées aux EDO comme les méthodes d'Euler progressive et rétrograde, les méthodes de Runge-Kutta, les méthodes à pas multiples, etc., ne sont adaptées à l'approximation des problèmes aux limites et ne peuvent pas être utilisées telles quelles. Pour ces derniers, il nous faudra mettre en œuvre des idées et des méthodes parfois entièrement différentes.

Nous commencerons par étudier le problème aux limites pour une équation elliptique en dimension 1 dans une formulation classique, c'est à dire dans des espaces usuels de fonctions C^k , et d'abord avec les outils traditionnels, ceux déjà connus pour les EDO, ce qui est possible en dimension 1 (mais ne l'est plus en dimension supérieure), puis par des méthodes qui, elles, seront généralisables en dimension supérieure. Outre l'existence et l'unicité, il est utile de connaître les propriétés de la solution, avant de chercher à l'approcher, afin de construire des méthodes numériques qui vont essayer de conserver l'analogie discret de ces propriétés.

L'approximation numérique de ces problèmes aux limites fait intervenir la résolution de systèmes linéaires et nous faisons donc auparavant quelques rappels d'algèbre linéaire matricielle.

L'exemple (1.8) de l'équation de Poisson en dimension un d'espace, qui contient une EDO (équation différentielle ordinaire), nous conduit à faire des rappels sur l'important théorème de Cauchy-Lipschitz pour des EDO plus générales. Nous rappelons son énoncé dans le cas où u est un vecteur de \mathbb{R}^n , et F continue Lipschitzienne, même si on peut le poser dans un contexte plus général (espaces de Banach) et avec des hypothèses plus faibles (F localement Lipschitzienne).

Théorème 1.4.1 Soit $I = [t_1, t_2]$ un intervalle compact d'intérieur non vide de \mathbb{R} et $F \in C^0(I \times \mathbb{R}^n, \mathbb{R}^n)$ une fonction continue Lipschitzienne par rapport à la deuxième variable :

$$\forall t \in I, \forall u, v \in \mathbb{R}^n, \|F(t, v) - F(t, u)\| \leq L\|v - u\|$$

Alors $\forall t_0 \in I, \forall u_0 \in \mathbb{R}^n$, il existe une unique fonction $u \in C^1(I, \mathbb{R}^n)$ qui vérifie

$$\begin{cases} u'(t) = F(t, u(t)) & \text{dans }]0, 1[, \\ u(t_0) = u_0. \end{cases} \quad (1.10)$$

La démonstration utilise le théorème du point fixe pour les applications contractantes.

Rappelons aussi qu'on peut ramener une EDO d'ordre p

$$u^{(p)}(t) = F(t, u, u', \dots, u^{(p-1)}), \quad (1.11)$$

avec $u \in C^p(I, \mathbb{R}^n)$, à une EDO du premier ordre en introduisant un nouveau vecteur inconnu $U = (u, u', \dots, u^{(p-1)})^T$ et en travaillant dans \mathbb{R}^{np} . Pour le problème de Cauchy, il faut se donner une condition initiale sur U c'est à dire les p valeurs $U(t_0) = (u, u', \dots, u^{(p-1)})^T(t_0)$ de u et de ses dérivées d'ordre non plus que $p - 1$ en t_0 .

Dans la suite de cette section on s'intéresse à l'existence et unicité de la solution d'un problème aux limites. On va considérer un problème modèle légèrement plus général que le problème du fil élastique (1.8). Les données du problème sont deux fonctions f et c continues sur l'intervalle $\bar{\Omega} = [0, 1]$ et deux nombres réels α et β . On va chercher à résoudre le problème aux limites suivant : trouver $u \in C^0([0, 1]) \cap C^2(]0, 1[)$ tel que

$$(P) \quad \begin{cases} -u''(x) + c(x)u(x) = f(x) & \text{dans } \Omega =]0, 1[, \\ u(0) = \alpha, \quad u(1) = \beta. \end{cases}$$

Remarque 1.4.1 *Un petit commentaire sur les notations : u appartient à $C^0([0, 1]) \cap C^2(]0, 1[)$ signifie que u est une fonction continue sur $[0, 1]$ fermé, dont la restriction à $]0, 1[$ ouvert est deux fois continûment dérivable. La première condition est indispensable pour qu'imposer des conditions aux limites de Dirichlet non homogènes comme celles de la deuxième ligne du problème (P) ait un sens. La deuxième condition nous autorise à parler de dérivée seconde de u dans l'intervalle ouvert, ce qui donne un sens à la première ligne de (P). Cependant, l'équation $u'' = cu - f$ entraîne immédiatement que u'' se prolonge par continuité en 0 et 1 (tout comme u') : la solution de (P) est de ce fait automatiquement de classe $C^2([0, 1])$.*

Dans cet exemple, la fonction correspondante à la notation de (1.11) est $F(x, u, u') = c(x)u - f(x)$, donc F ne dépend pas de la dernière variable, et le problème de Cauchy suppose de se donner $u(0)$ et $u'(0)$. Dans la suite, pour étudier l'existence de solutions de l'équation de Poisson en dimension un d'espace, nous allons effectivement utiliser le théorème 1.4.1 en l'appliquant à un problème de Cauchy (1.9). Cependant, notons tout de suite qu'une autre approche sera développée ultérieurement pour étudier ce type d'équation puisqu'il s'agit d'un *problème aux limites* et non d'un problème de *condition initiale* (ou problème de Cauchy).

ATTENTION! Contrairement au cas des EDO, il se peut que ce problème n'ait aucune solution ou pas de solution unique. Ainsi, par exemple, si $c(x) = -\pi^2$, on voit que l'équation homogène

$$-\pi^2 u + u'' = 0,$$

avec conditions homogènes de Dirichlet $u(0) = u(1) = 0$ admet une solution non-triviale $v(x) = \sin(\pi x)$. Ceci montre que si u est solution de (P) alors $u + \lambda v$ l'est aussi pour tout $\lambda \in \mathbb{R}$ et il n'y a donc pas unicité. Toujours avec ce choix de c et avec $f(x) = 1$ et $\alpha = \beta = 0$, il n'existe pas de solution au problème (P). Pour le montrer, on suppose qu'il en existe une, $u \in C^0([0, 1]) \cap C^2(]0, 1[)$, donc telle que $-u'' - \pi^2 u = 1$ dans $]0, 1[$ et $u(0) = u(1) = 0$. Remarquons tout de suite que, comme $u'' = -1 - \pi^2 u$, la fonction u est en fait de classe $C^2([0, 1])$ (i.e., sur l'intervalle fermé, et pas seulement sur l'intervalle ouvert), nous reviendrons plus en détail sur ce point plus loin. On multiplie d'abord l'équation différentielle par $\sin \pi x$, ce qui donne

$$-u''(x) \sin \pi x - \pi^2 u(x) \sin \pi x = \sin \pi x \quad \text{dans }]0, 1[,$$

puis on intègre l'égalité résultante entre 0 et 1. Il vient

$$-\int_0^1 u''(x) \sin \pi x dx - \pi^2 \int_0^1 u(x) \sin \pi x dx = \int_0^1 \sin \pi x dx = \frac{2}{\pi}.$$

D'un autre côté, en intégrant par parties la première intégrale, ce que l'on a le droit de faire puisque $u \in C^2([0, 1])$, on voit que

$$-\int_0^1 u''(x) \sin \pi x dx = -[u'(x) \sin \pi x]_0^1 + \pi \int_0^1 u'(x) \cos \pi x dx.$$

Comme $\sin 0 = \sin \pi = 0$, i.e., la fonction $\sin \pi x$ satisfait les conditions aux limites, le terme tout intégré s'en va. On intègre par parties une seconde fois et l'on trouve

$$-\int_0^1 u''(x) \sin \pi x dx = [u(x) \cos \pi x]_0^1 + \pi^2 \int_0^1 u(x) \sin \pi x dx = \pi^2 \int_0^1 u(x) \sin \pi x dx$$

puisque la fonction u satisfait les conditions aux limites. On a donc obtenu l'égalité $0 = 2/\pi$, qui est absurde. Il n'existe par conséquent pas de solution u à ce problème.

Conclusion : on n'aura pas existence de solution aux problèmes aux limites pour n'importe quelles données. On remarque la différence frappante avec le problème de Cauchy pour l'EDO, lequel a toujours une solution unique pour n'importe quel choix de c et f continues.

On va néanmoins utiliser les résultats connus sur le problème de Cauchy pour donner un résultat d'existence et d'unicité pour le problème aux limites dans le cas où la fonction c ne prend que des valeurs positives (il est possible de faire mieux).

Théorème 1.4.2 *Pour toutes $f, c \in C^0([0, 1])$ avec $c \geq 0$ et pour tous $\alpha, \beta \in \mathbb{R}$, il existe une solution du problème (P) et une seule. Cette solution appartient à $C^2([0, 1])$.*

Démonstration. Remarquons pour commencer que si u est solution de (P) alors nécessairement, $u \in C^2([0, 1])$, i.e., les dérivées première et seconde de u se prolongent par continuité à $[0, 1]$ fermé. En effet, comme $u''(x) = c(x)u(x) - f(x)$ pour $x \in]0, 1[$ et que c, f et u sont continues sur $[0, 1]$ fermé par hypothèse, u'' se prolonge par continuité en 0 et en 1, donc est bornée sur $[0, 1]$, et par conséquent en intégrant u'' à partir de $x = 1/2$ par exemple, il en va de même pour u' .

Démontrons d'abord l'*unicité*. On suppose donc que l'on a deux solutions du problème (P), u_1 et u_2 . Posant $v = u_1 - u_2$, on voit facilement que v est solution du problème aux limites :

$$\begin{cases} -v''(x) + c(x)v(x) = 0 & \text{dans }]0, 1[, \\ v(0) = v(1) = 0. \end{cases}$$

Multiplions l'équation par v et intégrons le résultat sur $]0, 1[$ (un procédé qui sera utilisé de nombreuses fois dans la suite). Il vient

$$-\int_0^1 v''(x)v(x) dx + \int_0^1 c(x)v(x)^2 dx = 0.$$

Intégrant la première intégrale par parties, on obtient donc

$$-[v'v]_0^1 + \int_0^1 (v'(x)^2 + c(x)v(x)^2) dx = 0.$$

Comme v s'annule en $x = 0$ et en $x = 1$, $[v'v]_0^1 = 0$. La fonction $v'(x)^2 + c(x)v(x)^2$ est continue sur $[0, 1]$, positive puisque c est supposée positive *par hypothèse* et d'intégrale nulle d'après le calcul qui précède. Elle est donc identiquement nulle. C'est une somme de deux termes positifs, donc chacun de ces termes est nul. En particulier $v'(x) = 0$, donc v est constante, ce qui implique que $v(x) = v(0) = 0$ pour tout $x \in [0, 1]$. En d'autres termes, $u_1 = u_2$ et il y a unicité.

Remarque 1.4.2 On déduit aussi du calcul ci-dessus que $c(x)v(x)^2 = 0$ dans $[0, 1]$. Ceci n'implique pas directement que $v = 0$ car les hypothèses n'empêchent pas c de s'annuler sur un ensemble plus ou moins grand. Il faut passer par v' et la condition aux limites pour pouvoir conclure. On voit aussi que si c prend des valeurs strictement négatives, le raisonnement ne marche plus car la fonction intégrée n'a plus de raison d'être positive et peut donc être d'intégrale nulle sans être identiquement nulle.

Pour l'*existence*, on utilise la *méthode de tir*. Le théorème 1.4.1 de Cauchy-Lipschitz nous assure que pour tout $\alpha \in \mathbb{R}$ et tout $\lambda \in \mathbb{R}$, il existe un unique $u_\lambda \in C^2([0, 1])$ solution du problème de Cauchy

$$\begin{cases} -u_\lambda''(x) + c(x)u_\lambda(x) = f(x) & \text{dans }]0, 1[, \\ u_\lambda(0) = \alpha, \quad u_\lambda'(0) = \lambda. \end{cases}$$

En effet, on peut écrire cette EDO sous forme d'un système d'EDO linéaire du premier ordre dans \mathbb{R}^2

$$\begin{cases} u_\lambda'(x) = -w_\lambda(x) \\ w_\lambda'(x) = -c(x)u_\lambda(x) + f(x) \end{cases}$$

soit

$$\begin{pmatrix} u_\lambda \\ w_\lambda \end{pmatrix}'(x) = A(x) \begin{pmatrix} u_\lambda \\ w_\lambda \end{pmatrix}(x) + \begin{pmatrix} 0 \\ f(x) \end{pmatrix}$$

avec la matrice $A(x) = \begin{pmatrix} 0 & -1 \\ -c(x) & 0 \end{pmatrix}$ et la condition initiale $(u_\lambda, w_\lambda)(0) = (\alpha, -\lambda)$. Notons $U = (u, w)^T$, $B(x) = (0, f(x))^T$ et $G(x, U) = A(x)U + B(x)$. La fonction G est bien Lipschitzienne par rapport à la variable U

$$\|G(x, U) - G(x, V)\| \leq \|A(x)\| \|U - V\|$$

pour tout norme matricielle subordonnée, et, par exemple, si on prend la norme ℓ^∞ (norme max), $\|A(x)\| \leq \max(1, |c(x)|)$ est borné uniformément en x car c est supposée continue sur le compact $[0, 1]$.

On aura manifestement construit une solution du problème (P) si l'on réussit à trouver un nombre λ tel que $u_\lambda(1) = \beta$. Comme ceci doit être possible pour tout $\beta \in \mathbb{R}$ donné, on se ramène donc à prouver que l'application

$$\begin{cases} g: \mathbb{R} \rightarrow \mathbb{R}, \\ \lambda \mapsto u_\lambda(1), \end{cases}$$

est *surjective*. On va en fait montrer que l'application g est *affine*, c'est-à-dire qu'il existe deux constantes a et b telles que $g(\lambda) = a\lambda + b$, avec $a \neq 0$ pour la surjectivité. Si cela est vérifié, on aura $g(0) = b$ c'est à dire $u_0(1) = b$, la notation u_0 signifie que u_0 est la solution de $-u_0'' + cu_0 = f$ avec $u_0(0) = \alpha$ et $u_0'(0) = 0$.

Posons donc $b = u_0(1)$, et $v_\lambda(x) = u_\lambda(x) - u_0(x)$. Alors on a

$$\begin{cases} -v_\lambda''(x) + c(x)v_\lambda(x) = 0 & \text{dans }]0, 1[, \\ v_\lambda(0) = 0, \quad v_\lambda'(0) = \lambda, \end{cases}$$

si bien que si l'on introduit la fonction

$$\begin{cases} l: \mathbb{R} \rightarrow \mathbb{R}, \\ \lambda \mapsto v_\lambda(1), \end{cases}$$

on a visiblement

$$g(\lambda) = l(\lambda) + b$$

et il suffit de montrer que l est *linéaire* et *surjective*.

Soient donc deux nombres λ_1 et λ_2 . La fonction $v_{\lambda_1} + v_{\lambda_2}$ satisfait clairement

$$\begin{cases} -(v_{\lambda_1} + v_{\lambda_2})''(x) + c(x)(v_{\lambda_1} + v_{\lambda_2})(x) = 0 & \text{dans }]0, 1[, \\ (v_{\lambda_1} + v_{\lambda_2})(0) = 0, \quad (v_{\lambda_1} + v_{\lambda_2})'(0) = \lambda_1 + \lambda_2, \end{cases}$$

c'est-à-dire le problème de Cauchy correspondant aux données initiales 0 pour la fonction et $\lambda_1 + \lambda_2$ pour sa dérivée. Or le théorème de Cauchy-Lipschitz assure l'*unicité* de la solution de ce problème de Cauchy, que nous avons précédemment notée $v_{\lambda_1 + \lambda_2}$. On a donc $v_{\lambda_1 + \lambda_2}(x) = v_{\lambda_1}(x) + v_{\lambda_2}(x)$ pour tout $x \in [0, 1]$. On montre de la même façon que $v_{\mu\lambda_1} = \mu v_{\lambda_1}$. En particulier, en prenant $x = 1$, on obtient

$$l(\lambda_1 + \lambda_2) = l(\lambda_1) + l(\lambda_2) \quad \text{et} \quad l(\mu\lambda_1) = \mu l(\lambda_1),$$

i.e., l est linéaire de \mathbb{R} dans \mathbb{R} .

Une application linéaire d'un espace vectoriel de dimension finie dans lui-même est surjective si et seulement si elle est injective. Il suffit donc de montrer que l est injective, c'est-à-dire, puisque c'est une application linéaire, montrer que son noyau $\ker l = \{\lambda \in \mathbb{R}, l(\lambda) = 0\} = \{\lambda \in \mathbb{R}, v_\lambda(1) = 0\}$ est réduit au vecteur nul. Or, si $\lambda \in \ker l$, il vient que

$$\begin{cases} -v_\lambda''(x) + c(x)v_\lambda(x) = 0 & \text{dans }]0, 1[, \\ v_\lambda(0) = v_\lambda(1) = 0. \end{cases}$$

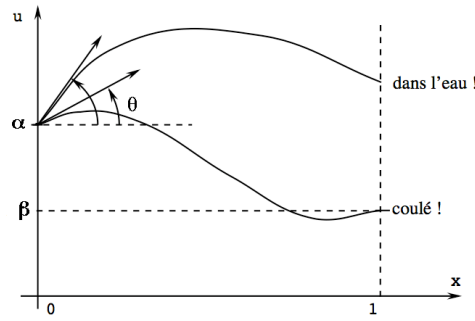


FIGURE 1.8 – Il suffit de bien viser

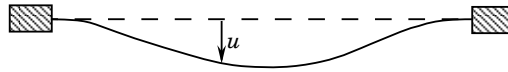


FIGURE 1.9 – Poutre encastrée en flexion

Comme $c \geq 0$, le même calcul que pour l'unicité montre donc que $v_\lambda(x) = 0$, ce qui implique bien sûr que $v'_\lambda(x) = 0$. Or, par définition de v_λ , on a $\lambda = v'_\lambda(0)$, par conséquent, on a montré que $\lambda = 0$, i.e., $\ker l = \{0\}$, l est injective, donc surjective. \square

Remarque 1.4.3 i) L'hypothèse $c \geq 0$ ne joue aucun rôle dans l'existence d'une solution u_λ au problème de Cauchy par le théorème de Cauchy-Lipschitz. Par contre, elle joue un rôle crucial dans l'unicité (qui implique ici l'existence) pour le problème aux limites. L'exemple $c(x) = -\pi^2$ montre d'ailleurs que l'on ne peut se passer de faire des hypothèses restrictives sur c pour pouvoir espérer un théorème d'existence (cependant l'hypothèse $c \geq 0$ n'est pas optimale).

ii) L'origine de la terminologie "méthode de tir" est claire (voir la figure 1.8) : on part de la valeur $u(0) = \alpha$ et on cherche à ajuster "l'angle de tir" (noté θ dans la figure 1.8, dont la tangente est λ) pour toucher la valeur cible β en $x = 1$, la trajectoire de notre projectile fictif entre les deux points étant gouvernée par l'équation différentielle du problème (P).

Malheureusement, cette méthode est strictement monodimensionnelle et n'a aucun analogue en dimension supérieure. Les théories d'existence sont dans ce cas beaucoup plus sophistiquées et sortent largement du cadre de ce cours ; nous en donnerons simplement un exemple (dans un cas simplifié) dans le chapitre suivant pour le problème aux limites en dimension 2 d'espace.

iii) La méthode s'applique également à des problèmes plus compliqués, par exemple

$$\begin{cases} u''''(x) - (a(x)u'(x))' + c(x)u(x) = f(x) & \text{dans }]0, 1[, \\ u(0) = u(1) = u'(0) = u'(1) = 0, \end{cases}$$

qui modélise l'équilibre d'une poutre élastique en flexion encastrée à ses extrémités sous l'action d'une force f (schématisé dans la figure 1.9). Ici, on prendra $a \geq 0$ et $c \geq 0$.

iv) On a montré l'existence et l'unicité d'une solution au problème aux limites, mais, sauf pour des données c et f très particulières, on n'a pas de formule "analytique" permettant d'évaluer numériquement $u(x)$, d'où le besoin de définir des méthodes d'approximation.

On va maintenant montrer que si on suppose que les données f et c sont plus régulières, alors la solution u du problème aux limites est aussi plus régulière (la régularité des données se propage en un certain sens à la solution du problème). On suppose encore $c \geq 0$, condition qui nous permet d'utiliser le théorème 1.4.2.

Théorème 1.4.3 *Si f et c sont de classe $C^m([0, 1])$ avec $m \in \mathbb{N}$, alors u appartient à $C^{m+2}([0, 1])$.*

Démonstration. Par récurrence sur m .

Pour $m = 0$, le résultat est inclus dans le théorème 1.4.2.

Supposons donc le résultat vrai pour $m - 1$, $m \geq 1$. Si f et c sont de classe $C^m([0, 1])$, alors, *a fortiori*, $f, c \in C^{m-1}([0, 1])$. Par hypothèse de récurrence, on a donc $u \in C^{(m-1)+2}([0, 1]) = C^{m+1}([0, 1]) \subset C^m([0, 1])$. Par conséquent, $u'' = cu - f \in C^m([0, 1])$ (car c'est une algèbre), c'est-à-dire $u \in C^{m+2}([0, 1])$. \square

Remarque 1.4.4 *Le résultat est correct et se démontre très simplement en dimension 1. Il est faux tel quel en dimension supérieure dans les espaces $C^m(\bar{\Omega})$ (par exemple, il existe des fonctions u sur un ouvert Ω de \mathbb{R}^2 telles que $\Delta u \in C^0(\bar{\Omega})$ mais $u \notin C^2(\bar{\Omega})$). Par contre, on a des théorèmes de régularité analogues à celui de la dimension 1, mais beaucoup plus difficiles à montrer et valables dans des espaces plus compliqués comme les espaces de fonctions höldériennes $C^{m,\alpha}(\bar{\Omega})$, $0 < \alpha < 1$, ou les espaces de Sobolev $H^m(\Omega)$.*

Donnons maintenant quelques résultats que l'on regroupe sous le terme générique de "principe du maximum".

Théorème 1.4.4 *Supposons que $f \geq 0$ dans $[0, 1]$, $\alpha \geq 0$ et $\beta \geq 0$. Alors la solution u de (P) est telle que $u(x) \geq 0$ pour tout $x \in [0, 1]$.*

Démonstration. On raisonne par l'absurde. Supposons qu'il existe un point $x_0 \in [0, 1]$ tel que $u(x_0) < 0$. Comme $u(0) = \alpha \geq 0$ et $u(1) = \beta \geq 0$, on a nécessairement $x_0 \in]0, 1[$ et comme u est continue, il existe un intervalle $[a, b] \subset [0, 1]$ contenant x_0 tel que $u(x) \leq 0$ sur $[a, b]$ et $u(a) = u(b) = 0$ (ceci par le théorème des valeurs intermédiaires). Sur cet intervalle, on a donc $u''(x) = c(x)u(x) - f(x) \leq 0$, puisque $c \geq 0$ par hypothèse. En d'autres termes, u est concave sur cet intervalle, c'est-à-dire $u(ta + (1-t)b) \geq tu(a) + (1-t)u(b) = 0$ pour tout $t \in [0, 1]$. En particulier, pour $t = (b - x_0)/(b - a)$, qui appartient bien à $[0, 1]$ car $x_0 \in [a, b]$, il vient $u(x_0) \geq 0$, ce qui contredit l'hypothèse $u(x_0) < 0$. \square

Remarque 1.4.5 *Il s'agit d'un résultat de dépendance monotone de la solution u par rapport aux données du problème f, α, β . Il a une interprétation physique très agréable : si on tire sur le fil élastique vers le haut (ou vers le bas), le fil se déplace vers le haut (resp. vers le bas).*

On déduit du théorème précédent une borne *a priori* pour la solution du problème (P). Attention, dans le résultat qui suit, il n'y a aucune hypothèse de signe sur α, β et f .

Théorème 1.4.5 *Supposons qu'il existe $\eta > 0$ tel que $c(x) \geq \eta$ dans $[0, 1]$. Alors la solution u de (P) est telle que*

$$\max_{x \in [0,1]} |u(x)| \leq \max \left\{ |\alpha|, |\beta|, \eta^{-1} \max_{x \in [0,1]} |f(x)| \right\}.$$

Démonstration. Posons $\Gamma = \max \{ |\alpha|, |\beta|, \eta^{-1} \max_{x \in [0,1]} |f(x)| \}$.

i) Soit $v^-(x) = u(x) - \Gamma$. Alors $v^-(0) = u(0) - \Gamma = \alpha - \Gamma \leq \alpha - |\alpha| \leq 0$. De même, $v^-(1) = \beta - \Gamma \leq 0$. Enfin,

$$-v^{--}(x) + c(x)v^-(x) = f(x) - \Gamma c(x) \leq f(x) - \Gamma \eta \leq f(x) - \max_{x \in [0,1]} |f(x)| \leq 0.$$

Le théorème 1.4.4 nous donne donc $v^-(x) \leq 0$ dans $[0, 1]$, c'est-à-dire $u(x) \leq \Gamma$ dans $[0, 1]$.

ii) Soit $v^+(x) = u(x) + \Gamma$. On voit de la même façon que $v^+(0) \geq 0$, $v^+(1) \geq 0$ et

$$-v^{++}(x) + c(x)v^+(x) \geq 0.$$

Par conséquent, $v^+(x) \geq 0$ dans $[0, 1]$, c'est-à-dire $u(x) \geq -\Gamma$ dans $[0, 1]$. □

Dans la suite de ces notes on utilisera parfois la notation

$$\|\varphi\|_{L^\infty} = \sup_{x \in \Omega} |\varphi(x)|,$$

pour une fonction φ définie sur un ensemble Ω . Le résultat précédent peut ainsi s'écrire

$$\|u\|_{L^\infty} \leq \max \{ |\alpha|, |\beta|, \eta^{-1} \|f\|_{L^\infty} \}.$$

1.5 Quelques rappels sur les matrices

Nous donnons ici quelques rappels sur les matrices qui nous seront utiles dans la suite. D'abord, pour des raisons pratiques, on écrira souvent un vecteur (colonne) V de \mathbb{R}^n

$$V = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

sous la forme du transposé d'un vecteur ligne

$$V = (v_1, v_2, \dots, v_n)^T,$$

et alors $V^T = (v_1, v_2, \dots, v_n)$.

Pour tout entier $n \in \mathbb{N}^*$, on note $M_n(\mathbb{R})$ l'espace des matrices carrées $n \times n$ à coefficients réels. Si l'on munit \mathbb{R}^n de sa base canonique, c'est-à-dire, du n -uplet de vecteurs e_i , $1 \leq i \leq n$, définis par $e_1 = (1, 0, \dots, 0)^T$, $e_2 = (0, 1, \dots, 0)^T$, ..., $e_n = (0, 0, \dots, 1)^T$, l'espace $M_n(\mathbb{R})$ s'identifie à l'espace des endomorphismes de \mathbb{R}^n : si $A \in M_n(\mathbb{R})$ est la matrice de coefficients $a_{i,j}$,

$A = (a_{i,j})$, $1 \leq i, j \leq n$, les vecteurs colonnes $Ae_k = (a_{1,k}, a_{2,k}, \dots, a_{n,k})^T$, $1 \leq k \leq n$ déterminent complètement l'endomorphisme correspondant. Nous utiliserons cette identification sans vergogne et noterons $\ker A$, $\operatorname{Im} A$, AV le noyau, l'image et l'action de l'endomorphisme associé à la matrice A sur un vecteur V de \mathbb{R}^n . Dans ce dernier cas, il s'agit aussi du produit de A , matrice carrée $n \times n$, avec la matrice colonne (à n lignes) des coordonnées de V dans la base canonique, ce produit matriciel donnant le vecteur colonne des coordonnées de AV dans la base canonique. Ce type d'abus de langage n'est pas dangereux pour la classe d'applications que l'on a en vue, tant que l'on s'est fixé une base de \mathbb{R}^n et, plus fondamentalement, que l'on sait ce que l'on fait.

L'espace $M_n(\mathbb{R})$ est lui-même un \mathbb{R} -espace vectoriel (de dimension n^2) pour l'addition des matrices et la multiplication par un scalaire. On peut donc parfaitement définir des normes sur cet espace, c'est-à-dire des applications $\|\cdot\| : M_n(\mathbb{R}) \rightarrow \mathbb{R}_+$ qui satisfont les trois propriétés :

$$\left\{ \begin{array}{l} \forall A, B \in M_n(\mathbb{R}), \|A + B\| \leq \|A\| + \|B\| \text{ (inégalité triangulaire),} \\ \forall A \in M_n(\mathbb{R}), \forall \lambda \in \mathbb{R}, \|\lambda A\| = |\lambda| \|A\|, \\ \|A\| = 0 \iff A = 0. \end{array} \right.$$

En plus de sa structure d'espace vectoriel, l'espace $M_n(\mathbb{R})$ est doté d'une structure multiplicative qui en fait un anneau non commutatif. En général, une norme sur $M_n(\mathbb{R})$ n'a aucune raison d'être compatible avec cette structure multiplicative. Les normes qui prennent en compte la structure multiplicative présentent un intérêt particulier.

Définition 1.5.1 Une norme sur $M_n(\mathbb{R})$ est appelée norme matricielle si elle satisfait

$$\forall A, B \in M_n(\mathbb{R}), \quad \|AB\| \leq \|A\| \|B\|.$$

Remarque 1.5.1 Il n'existe aucune norme qui satisfasse la relation plus forte $\|AB\| = \|A\| \|B\|$ pour tout couple de matrices (A, B) . En effet, il existe des couples (A, B) tels que $A \neq 0$, $B \neq 0$ mais $AB = 0$ (l'anneau $M_n(\mathbb{R})$ admet des diviseurs de zéro). Il suffit pour cela que l'on ait $\operatorname{Im} B \subset \ker A$.

Après avoir posé la définition 1.5.1, encore faut-il vérifier qu'il existe des normes qui la satisfont ! En fait, nous allons voir qu'il en existe beaucoup.

Proposition 1.5.1 (et définition) Soit $\|\cdot\|$ une norme sur \mathbb{R}^n . L'application

$$\left\{ \begin{array}{ll} M_n(\mathbb{R}) & \longrightarrow \mathbb{R}_+ \\ A & \longmapsto \|A\| := \sup_{V \in \mathbb{R}^n \setminus \{0\}} \frac{\|AV\|}{\|V\|} \end{array} \right.$$

définit une norme matricielle sur $M_n(\mathbb{R})$ appelée norme subordonnée à la norme vectorielle $\|\cdot\|$.

Démonstration. Soit A une matrice donnée. Posant, pour $V \in \mathbb{R}^n \setminus \{0\}$, $W = V/\|V\|$, on voit que

$$\|A\| = \sup_{V \in \mathbb{R}^n \setminus \{0\}} \frac{\|AV\|}{\|V\|} = \sup_{V \in \mathbb{R}^n \setminus \{0\}} \left\| A \left(\frac{V}{\|V\|} \right) \right\| = \sup_{W \in \mathbb{R}^n, \|W\|=1} \|AW\|.$$

Comme la sphère unité de \mathbb{R}^n est compacte (on est dans un espace de dimension finie) et que la norme choisie est une fonction continue, on en déduit d'abord que $\|A\| < +\infty$, donc l'application

$A \mapsto \|A\|$ envoie bien $M_n(\mathbb{R})$ dans \mathbb{R}_+ . Montrons dans un premier temps qu'il s'agit d'une norme. Considérons un couple de matrices (A, B) et un scalaire λ . Pour tout $V \in \mathbb{R}^n \setminus \{0\}$,

$$\frac{\|(A+B)V\|}{\|V\|} = \frac{\|AV + BV\|}{\|V\|} \leq \frac{\|AV\| + \|BV\|}{\|V\|} = \frac{\|AV\|}{\|V\|} + \frac{\|BV\|}{\|V\|} \leq \|A\| + \|B\|,$$

puisque $\|\cdot\|$ satisfait l'inégalité triangulaire. Prenant le sup du membre de gauche, on obtient

$$\|A+B\| \leq \|A\| + \|B\|.$$

De même,

$$\|\lambda A\| = \sup_{V \in \mathbb{R}^n \setminus \{0\}} \frac{\|\lambda AV\|}{\|V\|} = \sup_{V \in \mathbb{R}^n \setminus \{0\}} |\lambda| \frac{\|AV\|}{\|V\|} = |\lambda| \|A\|.$$

Enfin, si $\|A\| = 0$, on a $AV = 0$ pour tout $V \in \mathbb{R}^n$, d'où $A = 0$ évidemment. On a donc bien affaire à une norme sur $M_n(\mathbb{R})$.

Pour montrer que c'est une norme matricielle, on remarque d'abord que pour tout $V \in \mathbb{R}^n$,

$$\|AV\| \leq \|A\| \|V\|, \quad (1.12)$$

ce qui découle immédiatement de la définition. On prend alors deux matrices A et B et l'on note que pour tout $V \in \mathbb{R}^n \setminus \{0\}$,

$$\|ABV\| = \|A(BV)\| \leq \|A\| \|BV\| \leq \|A\| \|B\| \|V\|,$$

d'où en divisant par $\|V\|$,

$$\frac{\|ABV\|}{\|V\|} \leq \|A\| \|B\|,$$

et l'on conclut en prenant le sup du membre de gauche sur tous les V dans $\mathbb{R}^n \setminus \{0\}$. \square

Proposition 1.5.2 *On a également la caractérisation suivante de la norme subordonnée :*

$$\|A\| = \inf\{\mu \in \mathbb{R}_+ : \|AV\| \leq \mu \|V\|, V \in \mathbb{R}^n\}.$$

Démonstration. Par la remarque (1.12) de la démonstration précédente, il est clair que

$$\|A\| \geq \inf\{\mu \in \mathbb{R}_+ : \|AV\| \leq \mu \|V\|, V \in \mathbb{R}^n\}.$$

Soit donc $\mu \in \mathbb{R}_+$ tel que $\|AV\| \leq \mu \|V\|$ pour tout $V \in \mathbb{R}^n$. On a alors

$$\|A\| = \sup_{V \in \mathbb{R}^n \setminus \{0\}} \frac{\|AV\|}{\|V\|} \leq \sup_{V \in \mathbb{R}^n \setminus \{0\}} \frac{\mu \|V\|}{\|V\|} = \mu,$$

d'où

$$\|A\| \leq \inf\{\mu \in \mathbb{R}_+ : \|AV\| \leq \mu \|V\|, V \in \mathbb{R}^n\},$$

et le résultat. \square

Remarque 1.5.2 i) Toute norme matricielle subordonnée satisfait $\|\text{Id}\| = 1$.
 ii) Il existe des normes matricielles non subordonnées, par exemple

$$\|A\| = (\text{tr}(A^T A))^{1/2} = \left(\sum_{ij} |a_{ij}|^2 \right)^{1/2},$$

(le vérifier).

iii) On introduit parfois la notion de norme subordonnée en prenant le sup sur $\mathbb{C}^n \setminus \{0\}$ au lieu de $\mathbb{R}^n \setminus \{0\}$ (en complexifiant donc l'espace). Il peut arriver, même pour des matrices réelles, que les deux expressions ne coïncident pas, le sup «réel» étant toujours naturellement inférieur au sup «complexe». Cela n'est pas le cas pour les exemples que nous allons considérer plus loin.

Proposition 1.5.3 (et définition) Pour tout $p \in [1, +\infty]$, la fonction $\|V\|_p : \mathbb{R}^n \rightarrow \mathbb{R}_+$ définie par

$$\begin{cases} \|V\|_p = \left(\sum_{i=1}^n |v_i|^p \right)^{1/p} & \text{pour } p < +\infty, \\ \|V\|_\infty = \max_{i=1, \dots, n} |v_i| & \text{pour } p = +\infty, \end{cases}$$

est une norme vectorielle sur \mathbb{R}^n appelée «norme p ».

Nous ne démontrerons pas ce point. Pour $p = 2$, on retrouve la norme euclidienne usuelle. On notera $\|\cdot\|_p$ la norme matricielle subordonnée à la norme p . On sait évaluer plus ou moins explicitement ces normes subordonnées pour trois valeurs de p . On note $\rho(A)$ le *rayon spectral* de la matrice A , c'est-à-dire le module de la valeur propre de A de plus grand module. C'est le rayon du plus petit disque de \mathbb{C} centré en 0 qui contient toutes les valeurs propres de A .

Théorème 1.5.1 On a

$$\begin{aligned} \|A\|_1 &= \max_j \left(\sum_i |a_{ij}| \right), \\ \|A\|_2 &= \sqrt{\rho(A^T A)}, \\ \|A\|_\infty &= \max_i \left(\sum_j |a_{ij}| \right) \end{aligned}$$

Pour une matrice symétrique, on a $\|A\|_2 = \rho(A) = \max_{V \neq 0} \frac{\langle AV, V \rangle}{\|V\|^2}$.

Démonstration. Commençons par le cas $p = +\infty$. Soit V un vecteur non nul. Comme $(AV)_i = \sum_{j=1}^n a_{ij} v_j$, il vient

$$\|AV\|_\infty = \max_i \left| \sum_{j=1}^n a_{ij} v_j \right| \leq \max_i \left(\sum_{j=1}^n |a_{ij}| |v_j| \right)$$

par l'inégalité triangulaire. Or pour tout indice j , $|v_j| \leq \|V\|_\infty$ par définition de la norme ∞ . On en déduit donc que

$$\|AV\|_\infty \leq \left(\max_i \sum_{j=1}^n |a_{ij}| \right) \|V\|_\infty,$$

d'où, grâce à la proposition 1.5.2,

$$\|A\|_\infty \leq \max_i \sum_{j=1}^n |a_{ij}|.$$

Pour conclure, il suffit par conséquent de trouver un vecteur $W \in \mathbb{R}^n$, non nul, qui soit tel que $\|AW\|_\infty = (\max_i \sum_{j=1}^n |a_{ij}|) \|W\|_\infty$. Pour cela, on note i_0 un indice de ligne qui satisfait

$$\max_i \sum_{j=1}^n |a_{ij}| = \sum_{j=1}^n |a_{i_0j}|$$

(il en existe évidemment au moins un). On introduit alors le vecteur W de composantes

$$w_j = \frac{a_{i_0j}}{|a_{i_0j}|} \text{ si } a_{i_0j} \neq 0, \quad w_j = 1 \text{ si } a_{i_0j} = 0.$$

Dans tous les cas $|w_j| = 1$ et donc $\|W\|_\infty = 1$. De plus

$$(AW)_{i_0} = \sum_{j=1}^n a_{i_0j} w_j = \sum_{j=1}^n |a_{i_0j}|,$$

par construction des w_j . On en déduit que

$$\|AW\|_\infty \geq \sum_{j=1}^n |a_{i_0j}| = \max_i \sum_{j=1}^n |a_{ij}|,$$

et W répond donc à la question.

Le cas $p = 1$ est analogue et est laissé en exercice. Traitons le cas $p = 2$. On a

$$\|AV\|_2^2 = (AV)^T AV = V^T A^T AV, \text{ et } \|V\|_2^2 = V^T V.$$

La matrice $A^T A$ est symétrique et positive (puisque $\|AV\|_2^2 \geq 0$). Par conséquent, elle est orthogonalement diagonalisable et ses valeurs propres λ_i sont réelles positives. On convient de les ordonner en ordre croissant, ce qui implique que $\rho(A^T A) = \lambda_n$. On note f_i une base orthonormée de vecteurs propres de $A^T A$. Décomposant V sur cette base $V = \sum_{i=1}^n v_i f_i$, on voit que $A^T AV = \sum_{i=1}^n \lambda_i v_i f_i$, d'où

$$\|V\|_2^2 = \sum_{i=1}^n v_i^2 \quad \text{et} \quad \|AV\|_2^2 = \sum_{i=1}^n \lambda_i v_i^2.$$

D'une part

$$\frac{\|AV\|_2^2}{\|V\|_2^2} = \frac{\sum_{i=1}^n \lambda_i v_i^2}{\sum_{i=1}^n v_i^2} \leq \frac{\lambda_n \sum_{i=1}^n v_i^2}{\sum_{i=1}^n v_i^2} = \lambda_n$$

d'où $\|A\|_2 \leq \sqrt{\lambda_n}$, et d'autre part

$$\|A f_n\|_2^2 = \lambda_n,$$

d'où $\|A\|_2 \geq \sqrt{\lambda_n}$, ce qui conclut la démonstration. \square

Remarque 1.5.3 *Attention aux évidences trompeuses. Il ne faut pas confondre les normes subordonnées $\|A\|_p$ et les "normes p " sur l'espace vectoriel des matrices $M_n(\mathbb{R})$ (de dimension n^2 , donc isomorphe à \mathbb{R}^{n^2}). On voit ici que l'on a $\|A\|_1 \neq \sum_{ij} |a_{ij}|$, $\|A\|_\infty \neq \max_{ij} |a_{ij}|$ et $\|A\|_2 \neq (\sum_{ij} |a_{ij}|^2)^{1/2} \dots$*

