Problèmes aux limites en dimension 1

Simon Hergott

May 13, 2023

1 Introduction au problème

Cet article aura pour but de résumer les résultats et applications concernant les problèmes aux limites en dimension 1. Nous reviendrons sur la méthode des différences finies, fondant l'approximation des solutions des problèmes aux limites, et nous traiterons de plus quelques applications les plus courantes pour cette catégorie de problèmes. Éventuellement, nous verrons brièvement les méthodes de résolution de ces problèmes en dimension 2. Rappellons d'abord l'énoncé du modèle du problème aux limites en dimension 1:

$$-u''(x) = f(x) \qquad \forall x \in]0,1[\tag{1}$$

$$u(0) = u(1) = 0 \tag{2}$$

Ce problème est donc simplement caractérisé par l'équation (1) munie des conditions limites (2). Une partie importante lors de la résolution de ce problème est donc de trouver les valeurs propres et les fonctions propres (fonctions ne subissant qu'une transformation scalaire lors de leur utilisation comme solution) de l'équation différentielle

$$X'' + \alpha X = 0 \qquad \forall X \in]0,1[, \forall \alpha \in \mathbb{R}* \tag{3}$$

Nous pouvons alors écrire les solutions de (1) comme les fonctions de la forme

$$u(x) = c_1 + c_2 x - \int_0^x F(s)ds \quad c_1, c_2 \in \mathbb{R} \tag{4}$$

selon le théorème fondamental de l'Analyse, avec

$$F(s) = \int_0^s f(t)dt \tag{5}$$

Nous verrons que l'on peut écrire u sous la forme

$$u(x) = \int_0^1 G(x,s)f(s)ds \quad x \in [0,1] \tag{6}$$

ce qui nous permettra d'introduire avec G le concept de *Fonction de Greene*.

Les applications du problème aux limites en dimension 1 sont multiples, et relativement nombreuses dans la physique où sa résolution permet la simulation de plusieurs phénomènes tels que la conduction de la chaleur, et les défomrations élastiques.

Historiquement parlant, les problèmes aux limites sont apparus avec le développement de la physique: plusieurs mathématiciens tels que Jean Baptiste Fourier et Carl Gauss ont largement participé au développement de cette classe de problèmes au 19e siècle, dans le cadre de résolution de problèmes pour la physique. Les problèmes aux limites sont encore actuellement largement utililisés en recherche, notamment en physique où la réslution d'équations différentielles est courante.

2 Méthode des différences finies

Développée en 1715 par Brook Taylor dans son ouvrage *Methodus incrementorum directa et inversa*, la méthode des diffrences finies est un procédé courant utilisé pour approcher la solution d' équations différentielles. En résumé, on établit une grille de points généralement uniforme sur l'espace de recherche pour discrétiser le problème (le réduire à un nombre fini de pas), puis on réduit la distance entre les points pour approcher au maximum la solution. Formellement, on utilise la formule de Taylor pour discrétiser les différentielles n-ièmes : on peut alors choisir la formule de Taylor-Young, ou la formule de Taylor avec reste intégral pour évaluer les erreurs (la discrétisation induit une approximation, qui engendre des erreurs). Dans le cas de Taylor-Young simple (on appellera ici de cette manière la formule de Taylor sans reste intégral), on a:

$$f(x_0 + h) = f(x_0) + \sum_{0<i\leq n} \frac{f^{(i)}(x_0)}{i!} h^i$$

La méthode des différences finies sert à approximer la dérivée de certaines fonctions par des valeurs numériques, qui ne peut parfois pas être calculées par des méthodes formelles classiques. En effet, généralement les fonctions ne peuvent pas être formulées analytiquement ce qui rend le calcul de leurs dérivées impossible.

Pour appliquer la méthode des différences finies, on se place sur un segment $[0, \alpha]$ (on verra plus tard que dans le cas du problème aux limites en dimension 1, $\alpha = 1$)sur lequel on ajoute un pas de discrétisation h et une suite de points $x_n | n \in [0, N]$ avec $Nh = \alpha$. Ici, le pas est uniforme mais il peut tout à fait être défini non uniformément sur chaque x_i pour tout $i \in [0, N-1]$ comme h_i, avec $\sum_{0\leq i<N} h_i = \alpha$. On appellera x_n la *grille*.

Posons une fonction régulière u telle que: $u : [0, \alpha] \longrightarrow \mathbb{R}$, et on appellera u_n l'évaluation de u en chaque point x_n de la grille.

On appelle *une différence finie à p points* une combinaison linéaire de p u_n, servant à approximer au point x_n les dérivées de u. Considérons Du une différence finie, on dira qu'elle *approche* $u^{(l)}(x_n)$ *à l'ordre* q si :

$$\exists C > 0, h_0 > 0 \quad | \quad \forall h \in [0, h_0] \quad ||Du - u^{(l)}(x_n)|| \leq Ch^q \tag{7}$$

Cette méthode permet donc de décomposer les équations différentielles ordinaires en un système d'équations linéaires, solvable en utilisant l'algèbre linéaire.

2.1 Un exemple : la méthode d'Euler

Nous nous focaliserons maintenant sur un exemple pour définir la stabilité des méthodes numériques en général : nous utiliserons la méthode d'Euler, aussi appellée *méthode de la tangente*. Pour cela, nous poserons le problème à valeur initiale

$$\frac{dy}{dt} = f(t, y) \tag{8}$$

avec la condition initiale

$$y(t_0) = y_0 \tag{9}$$

Nous pouvons remarque que ce problème ressemble fortement au problème aux limites, sauf qu'ici il n'y en a qu'une. Dans Méthodes Numériques d'Alfio Quarteroni, cet exemple est utilisé en donnant $\frac{dy}{dt} = 0$, la résolution reste la même. Nous supposerons ici que les fonctions f et y sont continues dans le compact contenant le point (t_0, y_0) Nous savons alors qu'il existe une solution ans un certain intervalle autour de t_0. Nous supposerons la la solution est unique, en partant du principe que la fonction f est linéaire sans quoi l'intervalle pourrait être difficile à déterminer.

La méthode d'Euler est relativement simple : si on réécrit l'équation (8) au point t_n (toujours dans le compact de définition), nous pouvons arriver à la forme

$$\frac{d\phi}{dt}(t_n) = f(t_n, \phi(t_n)) \tag{10}$$

qui nous permet de la réécrire par le quotient de la différence aux points t_{n+1} et t_n:

$$\frac{\phi(t_{n+1}) - \phi(t_n)}{t_{n+1} - t_n} \approx f(t_n, \phi(t_n)) \tag{11}$$

Enfin, en remplaçant ϕ par y et en notant $y(t_n) = y_n$, nous pouvons écrire

$$y_{n+1} = y_n + f(t_n, y_n)(t_{n+1} - t_n) \quad \forall n \in \mathbb{N} \tag{12}$$

qui, en supposant le pas uniforme et de valeur h se simplifie en

$$y_{n+1} = y_n + h * f_n \quad \forall n \in \mathbb{N} \tag{13}$$

La méthode d'Euler consiste à calculer l'équation (13) en boucle en utilisant le résultat de l'étape précédente pour la nouvelle évaluation. L'algorithme est très simple, et nous pouvons dès à présent constater qu'il est extrêmement important d'éviter l'amplification d'erreurs d'étape en étape, en évaluant leur propagation. Afin de comprendre les erreurs apparaissant lors de l'utilisation de méthodes numériques d'approximation, nous allons voir quelques méthodes alternatives pour voir la méthode d'Euler sous un autre angle. Plusieurs méthodes sont possibles pour y parvenir, l'une d'entre elles est d'écrire le problème (8) sous la forme d'une intégrale.En supposant $y = \phi(t)$ la solution de notre problème en satisfant la condition (9), nous avons

$$\int_{t_n}^{t_{n+1}} \phi'(t)dt = \int_{t_n}^{t_{n+1}} f(t, \phi(t))dt \tag{14}$$

ce qui revient à dire

$$\phi(t_{n+1}) = \phi(t_n) + \int_{t_n}^{t_{n+1}} f(t, \phi(t))dt \tag{15}$$

L'intégrale de l'équation ci dessus (15) peut être représentée graphiquement comme l'aire sous la courbe de f entre t_n et t_{n+1}. En approwimant f par sa fonction discrète en posant $f(t, \phi(t)) \approx f(t_n, \phi(t_n))$ et en gardant l'hypothèse que nous sommes sur une *grille* (si ce concept n'a pas encore été défini, il le sera tout à l'heure) de pas uniforme h, nous avons alors:

$$\phi(t_{n+1}) \approx \phi(t_n) + hf(t_n, \phi(t_n)) \tag{16}$$

Une autre approche consiste à supposer que la solution $\phi(t)$ est développaple au sens de Taylor (en série de Taylor) en t_n:

$$\phi(t_n + h) = \phi(t_n) + \phi'(t–n)h + \phi''(t_n)\frac{h^2}{2!} + ... \tag{17}$$

Cette expression suppose que le développement de Taylor comprend plus que 2 termes, sinon nous aurions retrouvé une écriture précédente de la formule d'Euler.

L'utilisation de méthodes numériques telles que la formule d'Euler engendre des erreurs, qui doivent être identifiées et dont l'importance doit être vérifiée avant de pouvoir utiliser la solution approximée comme satisfaisante. En effet, un problème se pose pour la *convergence* de la solution : lorsque le pas h (en supposant qu'il soit uniforme, sinon considérer la distance entre les points de la grille) tend vers 0, est ce que al distance entre les solutions approximéees y_n en les points t_n tend vers 0? Et approchent-elles les véritables solutions du problème? Même en répondant à ces questions, nous pouvons encore nous demander avec quelle vitesse les solutions convergent-elles ; reformulé, à quel niveau de précision sur h faut-il aller afin de garantir une certaine marge d'erreur maximum? Certains effets de bord peuvent aussi ne pas être évidents : la diminution de la taille du pas afin de garantir une meilleure précision pourrait causer de plus grandes erreurs, qui annuleraient tout bénéfice.

Il y a un certain nombre de sources fondamentales d'erreurs dans l'approximation numérique d'un problème de valeur initiale (ce qui s'étend aux problèmes aux limites):

1. La formule est une approximation, donc source d'erreur. Dans le cas de la méthode d'Euler, la solution est approximée par une ligne droite en lieu et place de la courbe.

2. Les données utilisées sont souvent elle mêmes des approximations (voir à nouveau la méthode d'Euler)

3. Éventuellement, dans le cas d'un calcul sur ordinateur la précision pour chaque variable est finie (la mémoire ne peut stocker de valeurs analogiques)

En supposant que lors d'un calcul sur ordinateur la précision soit infinie (nous venons de vérifier que ce n'est pas vrai), nous pouvons poser E_n l'erreur d'approximation en chaque point t_n par:

$$E_n = \phi(t_n) - u_n \tag{18}$$

Cette erreur ne repose que sur les deux premiers points de la liste, elle est parfois appellée *erreur globale de troncature*.
En revanche, en relâchant l'hypothèse d'avoir des ordinateurs parfait, nous devons faire face à une erreur supplémentaire, *l'erreur d'arrondi* qui intervient dès lors que l'on calcule avec une arithmétique à *précision finie*. Cette erreur d'arrondi en un point t_n s'exprime comme

$$R_n = y_n - Y_n \tag{19}$$

avecc Y_n la valeur calculée par la machine en utilisant la méthode numérique.
En ajoutant ces erreurs, nous pouvons constater que l'erreur totale est la somme de la valeur absolue des erreurs d'approximation, et des erreurs d'arrondi.

2.2 Consistance

Étant donné ces erreurs, nous pouvons alors évaluer la *consistance* d'un problème: les méthodes génériques pour approcher la solution de (8) (en considérant $\frac{dy}{dt} = 0$ pour simplifier) sont en général toutes constituées d'une suite de problèmes approchés

$$f_n(t_n, y_n) = 0 \tag{20}$$

Nous supposerons que le problème est *bien posé* (que sa solution existe, soit unique et dépende continûment des données), sans quoi il devient difficile de montrer sa consistance. Nous verrrons plus tard que les problèmes aux limites considérés sont eux aussi bien posés, cette hypothèse n'est donc pas gênante ici.
Nous pouvons alors définir la notion de consistance d'une méthode numérique donnée comme:

$$n \to \infty \quad \Rightarrow \quad f_n(t_n, y) - f(t, y) \to 0 \tag{21}$$

Intuitivement, elle correspond à la quantité d'erreur commise par le schéma numérique, au temps t_n et doit tendre vers une valeur infiniment petite. Si cette relation est vraie, la méthode numérique est dite consistante. Sinon, elle ne l'est pas. De plus, une méthode numérique est *fortement consistante* si

$$\forall n \in \mathbb{N} \quad f_n(t, y) - f(t, y) = 0 \tag{22}$$

Nous retrouvons alors la notion de stabilité: une méthode numérique sera considérée *bien posée* ou *stable* si pour tout n il existe une solution t_n *unique* correspondant à la donnée y_n, et que t_n dépend des y_n continûment.

(ici arretés page 38 de quarteroni)

3 Résolution du problème en dimension 1

Revenons à l'introduction de cet article : nous avions vu que pour $u \in C^2[0, 1]$ satisfaisant l'équation (1), on avait u de la forme (4) Pour retrouver la fonction de Greene, nous devons

intégrer par parties $\int_0^x F(s)ds$:

$$\int_0^x F(s)ds = [sF(s)]_0^x - \int_0^x sF'(s)ds = \int_0^x (x-s)f(s)ds \tag{23}$$

D'après (2), les constantes c_1 et c_2 sont respectivement égales à 0 et $\int_0^1 (1-s)f(s)ds$. On peut alors écrire u sous la forme

$$u(x) = x\int_0^1 (1-s)f(s)ds - \int_0^x (x-s)f(s)ds = \int_0^1 G(x,s)f(s)ds \tag{24}$$

avec

$$G(x,s) = \begin{cases} s(1-x) & s \in [0,x] \\ x(1-s) & s \in [x,1] \end{cases} \tag{25}$$

On remarquera qu'on a continuité de la fonction G pour $s = x$, et que G est une fonction *affine* de x à s fixé, et de s à x fixé. Elle se nomme *Fonction de Greene* pour le problème aux limites défini par (1) et (2).

3.1 L'histoire avec Greene

La fonction de Greene est continue et symétrique sur $[0,1]^2$, ainsi que positive non strictement ($G(x,s) = 0 \iff x = 0 \vee s = 0$). Son existence dans le problème générique (1) (nous verrons plus tard des applications précises du problème comme la déformation d'une corde élastique) est assurée par sa définition sur tout l'intervalle entre les limites (dans notre cas, (2) mais on pourrait fixer d'autres limites).
On voit alors que lorsque G existe et qu'elle est connue formellement, on peut écrire explicitement les soutions du problème aux limites dans une forme très simple. Un des principaux avantages de la représentation (24) des solutions du problème est qu'elle élimine la dépendance au terme $f(s)$, qui n'est dépendant que de l'équation différentielle en (1) et des limites qui nous sont imposées : une fois que l'on aura déterminé G, les solutions seront connues selon $f(s)$ pour peu pqu'on puisse écrire les solutions sous la forme (24). Nous verrons dans les applications que cette forme n'est pas forcément admissible sous toutes les conditions.
De plus, la forme intégrale (24) est bien plus propice à l'analyse numérique que l'équation différentielle (1), ce qui rend le traitement par ordinateur bien plus simple et efficace. Les motivations principales sont donc de passer d'une recherche de u à une recherche de G : il faut alors trouver un moyen d'exprimer G sans avoir à résoudre l'équation (1).
Nous pouvons alors lister quelques propriétés de la fonction de Greene, qui nous seront utiles par la suite dans la résolution du problème:

1. G satisfait l'équation homogène de (1):

$$G'' = 0 \tag{26}$$

sur les intervalles $0 \leq s < x$ et $x < s \leq L$ avec dans notre cas L = 1. Généralement, on a bien continuité en $s = x$, nous le prouverons dans le cas du problème de l'élasticité d'une corde. (ou pas? Remettre la démo éventuellement)

2. $G(x,0) = G(x,L)$, elle satisfait donc les conditions limites (2).

3. $G(x,s) = G(s,x)$, G est symétrique sur ses arguments.

A partir de ces propriétés, nous pouvons commencer la résolution du problème aux limites en supposant l'existence d'une fonction de Greene G. Si cette supposition est valide, alors nous pouvons retrouver (24) depuis (1), ainsi que ses propriétés listées précédemment. En effet, d'après (1), après une multiplication des deux côtés par G nous pouvons intégrer comme suit:

$$\int_0^1 u''G(x,s)dx = -\int_0^1 f(x)G(x,s)dx \tag{27}$$

avec comme seule supposition la continuité lorsque $s \to x$. Pour être plus formel, nous pouvons exclure de l'intervalle d'intégration le point $x = s$ et éviter toute intégrale impropre.

$$\int_0^1 u''G(x,s)dx = \lim_{\epsilon \to s^-} \int_0^\epsilon u''G(x,s)dx + \lim_{\eta \to s^+} \int_\eta^1 u''G(x,s)dx \tag{28}$$

En intégrant deux fois par parties les deux intégrales à droite, nous avons:

$$\int_0^\epsilon u''G(x,s)dx = [Gu' - G'u]_0^\epsilon + \int_0^\epsilon uG''(x,s)dx \tag{29}$$

$$\int_\eta^1 u''G(x,s)dx = [Gu' - G'u]_0^\epsilon + \int_\eta^1 uG''(x,s)dx \tag{30}$$

Or, en choisissant $G(x,s)$ de manière à satisfaire $G'' = 0$ en tant que fonction de x dans les intervalles des intégrales ($[0,\epsilon] \cup [\eta,1]$), alors les intégrales à droite sont nulles. En ajoutant les termes restants (qu'on avait enlevés afin d'éviter les intégrales impropres), on revient en supposant G symétrique à:

$$\begin{aligned} -\int_0^1 f(x)G(x,s)dx = (G(s,s^-)u'(s^-) - G'(s,s^-)u(s^-) - G(0,s)u'(0) + G(1,s)u'(1) \\ - G(s,s^+)u'(s^+) + G'(s,s^+)u(s^+)) \end{aligned} \tag{31}$$

En supposant que G vérifie les conditions limites et que u est continu en s, nous pouvons écrire:

$$\int_0^1 f(x)G(x,s)dx = -u(y)(G'(s,s^+) - G'(s,s^-)) + u'(s)(G(s,s^+) - G(s,s^-)) \tag{32}$$

De plus, G est continue en $x = s$, tandis que G' y est discontinue. Nous aurons besoin de cette proposition pour continuer, nous allons la démontrer.
À partir de (1), nous pouvons utiliser la méthode de la variation des variables pour supposer que, si le problème existe, alors les solutions seront de la forme

$$u(x) = A(x)cos(kx) + B(x)sin(kx). \tag{33}$$

En dérivant deux fois selon x et en supposant qu'on ait $A'(x)cos(kx)+B'(x)sin(kx) = 0$, on trouve que (33) constitue effectivement une solution, à la condition

$$-kA'sin(kx) + kB'cos(kx) = -f(x) \tag{34}$$

En résolvant (34) accompagné de sa condition présupposée, nous pouvons alors exprimer A' et B':

$$A'(x) = \frac{f(x)sin(kx)}{k} \tag{35}$$

$$B'(x) = \frac{-f(x)cos(kx)}{k} \tag{36}$$

Nous pouvons alors écrire la solution de (1) comme:

$$u(x) = \frac{cos(kx)}{k}\int_{c_1}^{x} f(s)sin(ks)ds - \frac{sin(kx)}{k}\int_{c_2}^{x} f(s)cos(ks)ds \tag{37}$$

avec c_1, c_2 constantes bien choisies pour satisfaire les conditions (2). En utilisant la condition en 0 de (2), on trouve $c_1 = 0$. De même, la condition en 1 implique

$$u(x) = \frac{1}{k}\int_{0}^{x} f(s)sin(k(s-x))ds - \frac{sin(kx)}{ksin(kl)}\int_{x}^{1} f(s)sin(k(s-1)ds \tag{38}$$

ce qui nous donne enfin

$$\begin{aligned} u(x) &= \int_0^x \frac{sin(ks)sin(k(1-x))}{ksin(k)}ds + \int_x^1 f(s)\frac{sin(kx)sin(k(1-s))}{ksin(kl)}ds \\ &= \int_0^1 f(s)G(x,s)ds \end{aligned} \tag{39}$$

qui revient au même pour introduire la fonction de Greene, de manière bien détaillée cette fois. On a par la même occasion montré la consistance du problème aux limites (revoir la def!!!!).
Finalement, nous aboutissons à une forme de G intéressante:

$$G(x,s) = \frac{sin(ky)sin(k(1-x)}{ksin(k)} \quad s \in [0,x] \tag{40}$$

$$G(x,s) = \frac{sin(kx)sin(k(1-s)}{ksin(k)} \quad s \in [x,1] \tag{41}$$

À partir d'ici, nous pouvons vérifier aisément que la fonction G est continue en $s = x$, et que sa dérivée est discontinue en ce même point.
Depuis (32), nous pouvons alors conclure que

$$u(s) = \int_0^1 f(x)G(x,s)ds \tag{42}$$

et nous obtenons alors une représentation désirable de la solution duproblème aux limites par une fonction G. Cette fonction vérifie tous les axiomes de la fonction de Greene énoncés auparavant, nous avons donc établi une méthode pour retrouver la solution des problèmes aux limites d'une manière relativement directe. Néanmoins, cette méthode suppose que nous connaissions la fonction G.

3.2 Suite de la resolution : approximation par différences finies

Sur la grille des points $(x_j)_{j=0}^n$ de pas uniforme h (on a donc $x_j = jh$) sur $[0,1]$, l'approximation de la solution est une suite finie $(u_j)_{j=0]}^n$ telle que:

$$-\frac{u_{j+1} - 2u_j + u_{j-1}}{h^2} = f(x_j) \quad j \in [[1, n-1]] \tag{43}$$

avec $u_i = 0$ aux points 0 et n. On a alors u_j approche $u(x_j)$, la valeur de chaque point de la grille. Pour s'en convaincre, on se réfèrera à ... dans la partie traitent des différences finies centrées, en remplaçant $u''(x)$ par son approximation du 2nd ordre.

On posera $u =< u_1, ..., u_n - 1 >$ et $f =< f_1, ..., f_n - 1 >$ vecteurs avec $f_i = f(x_i)$, en utilisant une notation empruntée au C++. Nous pouvons alors voir que (43) peut s'écrire comme

$$A_{df} u = f \tag{44}$$

avec A_{df} matrice carrée de différences finies de taille $(n-1)$ définie par

$$A_{fd} = h^{-2} tridiag_{n-1}(-1, 2, -1) \tag{45}$$

Elle est à diagonale dominante par ligne, et définie positive (démontrable).

Alors, l'équation (44) n'admet qu'une unique solution.
Définissons les M-Matrices comme suit : une M-Matice est une matrice carrée inversible avec tous ses coefficients non sur la diagonale négatifs ou nuls, ainsi que tous les coefficients de son inverse sont positifs ou nuls. Nous pouvons noter que A_{df} est une M-Matrice, ce qui permet de satisfaire la condition de monotonie de la solution exacte $u(x)$: nous avons alors $f > 0 \Longrightarrow u > 0$ (propriété du *maximum discret*).

Le but est maintenant de réécrire (43). Pour cela, nous pouvons considérer V_h ensemble de fonctions discrètes définies sur les x_j points de la grille. Pour tout v_h dans V_h, elle est alors définie en tout point de la grille, et $v_j = v_h(x_j)$.
On posera de plus $V_h^0 = \{v_h \in V_h | v_0 = v_n = 0\}$. On pourra définir L_h comme

$$(L_h v_h)(x_j) = -\frac{v_{j+1} - 2v_j + w_{j-1}}{h^2} \qquad v_h \in V_h \quad j \in [[1, n-1]] \tag{46}$$

On peut alors encore réécrire le problème (43) comme:

$$(L_h u_h)(x_j) = f(x_j) \quad j \in [[1, n-1]] \tag{47}$$

en cherchant $u_h \in V_h^0$: on prend en compte les conditions limites.

3.3 Analyse de la stabilité

Nous allons maintenant analyser la *stabilité* du problème. On peut définir la stabilité par la capacité de l'algorithme de résolution à ne pas amplifier les erreurs, et à produire des résultats cohérents. Ici, vous allons essayer de montrer que la solution renvoyée en appliquant la méthode des différences finies au problème est bornée par une des variables d'entrées.

Pour cela, nous aurons besoin de quelques notions supplémentaires. Nous définirons le *produit scalaire discret* comme

$$(w_h, v_h)_h = h \sum_{k=0}^{n} c_k w_k v_k \quad \forall v_h, w_h \in V_h \tag{48}$$

avec $c_i = 1 \forall i \in [[1, n-1]]$ et $c_0 = c_n = 1/2$. On peut retrouver le produit scalaire discret par différences finies, avec la *formule composite du trapèze* (voir partie sur les différences finies). On peut alors définir une norme sur V_h par la racine du produit scalaire d'un opérateur de V_h avec lui même.

Nous pouvons alors émettre plusieurs propriétés sur ce produit scalaire. Tout d'abord, il est symétrique et défini positif pour l'opérateur L_h dans V_h^0:

$$(L_h w_h, v_h)_h = (v_h, L_h w_h)_h \quad (L_h v_h, v_h)_h \geq 0 (0 \Leftrightarrow v_h = 0) \tag{49}$$

(démonstration possible)

On définit la norme $|||.|||_h$ sur V_h^0 par:

$$|||v_h|||= \left\{ h \sum_{j=0}^{n-1} (\frac{v_{j+1} - v_j}{h})^2 \right\}^{1/2} \tag{50}$$

On peut alors constater que

$$(L_h v_h)_h = |||v_h|||_h^2 \quad \forall v_h \in V_h^0 \tag{51}$$

De plus, on a:

$$||v_h||_h^2 \leq \frac{1}{\sqrt{2}} |||v_h|||_h \quad \forall v_h \in V_h^0 \tag{52}$$

(démonstration possible, on utilise l'inégalité de Minkowski en repartant de la définition des v_j)
De même, on peut écrire la version discrète de *l'inégalité de poincaré*: en notant $v_h^{(1)}$ la fonction discrète de v_h dans V_h^0 prenant ses valeurs sur la grille par $(v_{j+1} - v_j)/h$ avec $j \in [[0, n-1]]$, on peut la voir commme la dérivée discrète de v_h.
On peut alors écrire l'inégalité comme

$$||v_h||_h \leq \frac{1}{\sqrt{2}} ||v_h^{(1)}||_h \forall v_h \in V_h^0 \tag{53}$$

À partir de (52), on peut multiplier chaque équation par (47) pour avoir

$$(L_h u_h, u_h)_h = (f, u_h)_h \tag{54}$$

En reprenant la notation introduite dans (46), on peut définir f_d fonction discrète égale à f en chaque poitnde la grille. On a alors $f_d(x_i) = f(x_i)$. À partir de (51), on peut trouver

$$|||u_h|||_h^2 \leq ||f_d||_h ||u_h||_h \tag{55}$$

à partir de l'inégalité de Schwarz. Alors, on a

$$||u_h||_h \leq \frac{1}{2}||f_d||_h \tag{56}$$

Regarder d'où sortent les uh qui trainent par ici!!!!!!!!!!!!!!!!!!!!!!!!!!
ce qui conclut que la seule solution correspondant à $f_d = 0$ est $u_h = 0$, donc le problème aux différences finies ne possède qu'une seule solution. La stabilité, titre de cette section, est quant à elle donnée par la majoration (ou *borne*, comme on est sur une norme) de la solution par f_d qui est une des données du problème.

3.4 Consistance

Nous devons maintenant prouver la convergence du problème aux limites. Nous devrons introduire ici la notion de *consistance* : elle a été définie dans le cadre général par (21). Dans notre cas, si $f \in C^0([0,1])$ et la solution de (1) u est de classe C^2 sur ce même segment (dont les bornes correspondent aux cas limites) alors nous pouvons considérer *l'erreur de troncature locale* τ_h définie par

$$\tau_h(x_j) = (L_h u)(x_j) - f(x_j) \quad j \in [[1, n-1]] \tag{57}$$

En développant par Taylor, on a:

$$\tau_h(x_j) = (L_h u)(x_j) + u''(x) = -\frac{h^2}{24}(u^{(iv))}(\chi_j) + u^{(iv))}(\nu_j)) \tag{58}$$

avec $\nu_j \in]x_j, x_{j+1}[\wedge \chi_j \in]x_{j-1}, x_j[$.
On peut définir la *norme discrète du maximum* par

$$||v_h||_{h,\infty} = \max_{0 \leq j \leq n} |v_h(x_j)| \tag{59}$$

ce qui nous permet de déduire de (3.4) que

$$||\tau_h||_{h,\infty} \leq \frac{||f''||_\infty}{12} * h^2 \tag{60}$$

à partir du moment où on a $f \in C^2([0,1])$ dans (1). De plus, dans notre cas avec les conditions (2) nous avons $\lim_{h \to 0} \tau_{h,\infty} = 0$: la méthode des différences finies est donc consistante avec notre problème.

3.5 Convergence

Cette partie sera laissée de côté dans cette analyse, car elle est relativement complexe et ne présente pas de réel intérêt : le résultat de la convergence reprend les arguments de stabilité et de consistance, et montre que l'erreur de discrétisation est comparable à l'erreur de stabilité.
Nous définirons en revanche les fonctions w_h pour toute fonction discrète $g \in V_0^h$, qui nous seront utiles par la suite dans la résolution du problème en une dimension:

$$w_h = \sum_{k=1}^{n-1} g(x_k) G^k \tag{61}$$

où nous posons G fonction de Greene.

Nous laisons le soin au lecteur de voir p 428 d'alfio quarteroni.

3.6 différences finies pour les problèmes en dimension 1

Dans cette partie, on utilisera une généralisation du problème aux limites (1) (**??**) en une dimension:

$$Lu(x) = -(J(u)(x))' + \gamma(x)u(x) = f(x) \quad \forall x \in]0,1[\tag{62}$$

avec les conditions limites d_0 et d_1:

$$u(0) = d_0 \quad u(1) = d_1 \tag{63}$$

en posant $J(u)(x) = \alpha(x)u'(x)$ et α, $\gamma \geq 0$, f des fonctions connues et continues sur l'intervalle $[0,1]$. L'hypothèse de continuité est généralement vérifiée dans les applications physiques. Nous appellerons $J(u)$ le *flux associé à u*, il nous sera utile dans la partie de résolution des applications.

Pour l'application, nous suivrons la méthode décrite dans Alfio Quarteroni : nous reprendrons la grille uniforme précédente des (x_j), et nous créerons une nouvelle grille utilisant les *points milieu* de cette grille: nous définirons les point milieu comme les $x_{j+1/2} = (x_j + x_{j+1})/2 \quad \forall j \in [[1, n-1]]$. Alors, nous pouvons introduire un nouveau schéma aux différences finies pour approcher (62): nous cherchons $u_h \in V_h \quad |$

$$L_h u_h(x_j) = f(x_j) \quad \forall j \in [[1, n-1]] \tag{64}$$

avec les conditions limites

$$u_h(x_0) = d_0 \quad \wedge u_h(x_n) = d\text{–}1 \tag{65}$$

en définissant L_h à partir de la fonction γ précédente et des w_k de (61) par:

$$L_h w_h(x_j) = -\frac{J_{j+1/2}(w_h) - J_{j-1/2}(w_h)}{h} + \gamma_j w_j \quad \forall j \in [[1, n-1]] \tag{66}$$

On a repris ici la notation habituelle $\gamma_j = \gamma(x_j)$. Nous pouvons alors poser $\alpha_{j+1/2} = \alpha(x_{j+1/2})$, afin de définir les *flux approchés* $J_i \quad \forall i \in [[1/2, n-1/2]]$:

$$J_{j+1/2}(w_h) = \alpha_{j+1/2} * \frac{w_{j+1} - w_j}{h} \tag{67}$$

À partir de cette formulation avec des flux approchés, nous pouvons alors remettre en forme le problème (64) ainsi que ses conditions limites avec les flux approchés J_i. Pour cela, nous aurons besoin de définir la matrice de différences finies A_{df} comme:

$$A_{fd} = h^{-2} tridiag_{n-1}(a, d, a) + diag_{n-1}(c) \tag{68}$$

avec:

- $a = -[\alpha_{3/2}, ..., \alpha_{n-3/2}]^T$ (indices de pas 1)
- $d = [\alpha_{1/2} + \alpha_{3/2}, ..., \alpha_{n-3/2} + \alpha_{n-1/2}]^T$
- $c = [\gamma_1, ..., \gamma_{n-1}]^T$

Avec les facteurs de a, d, c dans $\mathbb{R}$. La démonstration est assez longue et calculatoire, et globalement peu intéressante ici, nous renvoyons à Alfio Quarteroni?.

Nous pouvons dès lors remarquer que la matrice A_{fd} de (68) est définie positive, symétrique, et à diagonale strictement dominante (la valeur avsolue de chaque terme sur la diagonale est supérieur non strictement à la somme des valeurs absolues des autres termes de sa ligne) pourvu que $\gamma > 0$. On peut analyser la convergence de ce schéma aux différences finies en reprenant la technique de la Méthode de l'Energie vue précédemment. Il est possible enfin de définir différentes conditions limites, plus générales que celles définies au début de cette section : une pléthore de ces conditions ont vu le jour dans des applications physiques, nous nous focaliserons principalement sur les conditions dites de *Neumann* et de *Dirichlet*. Ces conditions s'expriment respectivement comme

$$J(u)(1) = g_1 \quad ; \quad u(0) = d_0 \tag{69}$$

Nous allons discrétiser la condition de Neumann, en utilisant la *technique du miroir*, qui consiste à ne garder qu'une partie simple d'une fonction ψ et à la développant linéairement:
Posons ψ fonction régulière, nous pouvons la développer en sérien de Taylor en x_n comme

$$\psi_n = \frac{\psi_{n-1/2} + \psi_{n+1/2}}{2} - \frac{h^2}{16}(\psi''(\eta_n) + \psi''\zeta_n)) \tag{70}$$

en choisissant η_n dans $[x_{n-1/2}, x_n[$ et ζ_n dans $]x_n, x_{n+1/2}]$. On peut alors utiliser la condition limite (69),

$$\psi = J(u) \quad \Rightarrow \quad J_{n+1/2}(u_h) = 2g_1 - J_{n-1/2}(u_h) \tag{71}$$

Le point $x_{n+1/2} = x_n + h/2$ n'existe pas comme il sort de la grille (on suppose le pas uniforme, et différent de 0) : il est appellé *point fantôme* dans Alfio Quarteroni. Nous obtenons le flux approché correspondant par prolongation linéaire des deux flux précédents $J_{n-1/2}$ et J_n. En x_n, l'équation (66) se reformule

$$\frac{J_{n-1/2}(u_h) - J_{n+1/2}(u_h)}{h} + \gamma_n u_n = f_n \tag{72}$$

Nous pouvons alors reprendre le point fantôme vu dans (71) pour que $J_{n+1/2}$ existe (ou pas, finalement...), et on obtient une approximation à l'ordre 2 en développant l'équation précédente:

$$-\alpha_{n-1/2}\frac{u_{n-1}}{h^2} + (\frac{\alpha_{n-1/2}}{h^2} + \frac{\gamma_n}{2})u_n = \frac{g_1}{h} + \frac{f_n}{2} \tag{73}$$

À partir de cette formule, nous pouvons alors modifier simplement les coefficients de la matrice de (44), ce qui termine l'utilisation des différences finies pour la résolution du problème en dimension 1.

3.7 Introduction à la formulation intégrale

Nous utiliseraons dans cette partie les notations d'alfio Quarteroni. Le problème aux limites (1) peut être généralisé comme

$$-(\alpha u')'(x) + (\beta u')(x) + (\gamma u)(x) = f(x) \quad \forall u \in]0,1[\tag{74}$$

en vérifiant les conditions limites (2) en posant $u(0) = u(1) = 0$, avec α, β, γ des fonctions continues sur $]0,1[$, et en supposant l'existence d'une constante α_0 telle que $\alpha(x) \geq \alpha_0 > 0 \quad \forall x \in [0,1]$. Nous pouvons alors utiliser une méthode dérivée de la fonction test : posons une fonction v de classe C^1 sur $[0,1]$ que l'on choisira en tant que fonction test. Nous pouvons alors multiplier (74) par v et l'intégrer sur l'intervalle $[0,1]$:

$$\int_0^1 \alpha u'v'dx + \int_0^1 \beta u'vdx + \int_0^1 \gamma uvdx = \int_0^1 fvdx + [\alpha u'v]_0^1 \tag{75}$$

après une intégration par parties sur le premier terme. En appliquant les conditions limites sur v (on oblige v à être nulle en 0 et en 1), on a alors $[\alpha u'v]_0^1 = 0$ ce qui nous donne

$$\int_0^1 \alpha u'v'dx + \int_0^1 \beta u'vdx + \int_0^1 \gamma uvdx = \int_0^1 fvdx \tag{76}$$

Voir si on garde le truc avant, ca sert peu!!!

On posera V l'espace des fonctions test telles v, en prenant compte les conditions limites : cet espace contient des fonctions continues, s'annulant en 0 et en 1, et de dérivée continue par morceaux. C'est alors un espace vectoriel, que l'on peut aussi noter

$$H_0^1(]0,1[) = \{v \in L^2(]0,1[) \quad | \quad v' \in L^2(]0,1[) \wedge v(0) = v(1) = 0\} \tag{77}$$

en notant v' la *dérivée au sens des distributions*, dont la définition sera donnée dans la section suivante (bases des distributions).

Nous venons de montrer que si une fonction u de $C^2([0,1])$ satisfait la généralisation (74) du problème, alors u est aussi la solution du problème

$$a(u,v) = (f,v) \quad u \in V \quad \forall v \in V \tag{78}$$

où on définit $(f,v) = \int_0^1 fvdx$ comme le produit scalaire de $L^2(]0,1[)$, avec de plus

$$a(u,v) = \int_0^1 \alpha u'v'dx + \int_0^1 \beta u'vdx + \int_0^1 \gamma uvdx \tag{79}$$

forme bilinéaire par rapport à (u,v).

On appellera *formulation faible* le problème (78) : cette formulation est plus généaliste, car ellle permet d'étudier des cas où u n'est pas de classe C^2, car les dérivées sont toutes simples. Cette formulation permet de décrire par exemple le problème de la déformation d'une corde élastique, qui sera étudié par la suite.

En cas de non homogénité des conditions limites dans (74) (ce qui arrive dans le cas général), il est toujours possible de se ramener à la forme (78) en utlisant des *conditions limites de Neumann* et une interpolation des extrémités par une fonction affine (détails dans quarteroni, page 433).

3.8 Bases des distributions

Cette section est dédiée à la présentation des bases de la théorie des distributions, qui nous sera utile ors de l'étude des propriétés de la méthode de Galerkin. En résumé, la théorie des distributions est un outil servant à généraliser la notion de dérivée d'une fonction, afin de pouvoir résoudre certaines équations différentielles (par exemple, un problème aux limites en dimension 1). Les distributions utilisent une méthode différente pour évaluer une fonction : au lieu d'évaluer la fonction en un point, on évalue la fonction sur un voisinage du point tout en effectuant une moyenne (posiblement pondérée) sur ces points afin de réduire le poids des discontinuité dans le résultat.

Supposons X un espace vectoriel normé complet (c'est à dire un *espace de Banach*). Définissons une *forme* $T : X \to \mathbb{R}$ continue linéaire sur X. Nous utiliserons la notation de *dualité*, utilisée dans la théorie des distributions:

$$\langle T, x\rangle = T(x) \tag{80}$$

Nous pouvons également définir la notation de l'espace des fonctions indéfiniment dérivables de support compact sur $[a,b]$: $C_0^\infty(]a,b[)$. Nous considérons que tous les compacts seront dans $]0,1[$ à des fins pratiques, comme cet intervalle correspond aux limites de notre problème.
Nous pouvons maintenant introduire la notion d' *espace dual* : c'est l'espace défini par les formes linéaires $C_0^\infty(]0,1[)$, que l'on notera $D'(]0,1[)$. Les éléments le composant sont des *distributions* : toute fonction localement intégrable f est associée à une distribution notée ϕ et définie par

$$\langle f, \phi\rangle = \int_0^1 f\phi \tag{81}$$

Nous avons maintenant les bases pour introduire les *dérivées au sens des distributions* : soit $T \in D'(]0,1[)$. Alors, $\forall k \in \mathbb{N}*$, $T^{(k)}$ est une distribution telle que

$$\langle T^{(k)}, \phi\rangle = (-1)^k \langle T, \phi^{(k)}\rangle \quad \forall \phi \in C_0^\infty(]0,1[) \tag{82}$$

Un court exemple est disponible dans Alfio Quarteroni.

3.9 Propriétés de la méthode de Galerkin

La méthode de Galerkin permet la résolution du problème aux limites par une autre approche que les différences finies: on utilise ici la résolution du problème (78), appellée *formulation faible*.
En supposant V_h un sous espace vectoriel de V tel que $dim(V_h) < \infty$, la méthode de Galerkin consiste à approximer le problème (78) par:

$$u_h \in V_h \quad | \quad a(u_h, v_h) = (f(v_h) \quad \forall v_h \in V_h \tag{83}$$

Qui devient un autre problème lorsqu'on cherche u_h. Il n'est pas forcément évident que ce problème soit en dimension finie (EXPLIQUER PQ?), nous allons le montrer. Soit $\{\phi_1, ..., \phi_N\}$ base de V_h, avec $dim(V_h) = N$. On peut alors dire que

$$u_h(x) = \sum_{j=1}^{N} u_j \phi_j(x) \tag{84}$$

Avec $v_h = \phi_i$ dans (**??**), le problème devient la recherche de N réels $\{u_1, ..., u_N\}$ |

$$\sum_{j=1}^{N} u_j a(\phi_j, \phi_i) \quad \forall i \in [[1, N]] \tag{85}$$

On rappellera que $a(.,.)$ est linéaire par rapport à la première place. Nous introduirons la matrice $A_G = (a_{ij}) \quad | \quad a_{ij} = a(\phi_j, \phi_i)$, un vecteur colonne inconnu solution de notre ptoblème $u = [u_1, ..., u_N]$ et le *second membre* colonne $f_G = [f_1, ..., f_N] \quad | \quad f_i = (f, \phi_i)$: nous pouvons alors voir que comme pour (44), nous pouvons compacter l'équation précédente en

$$A_G u = f_G \tag{86}$$

La précision de u_h dépend de la forme des fonctions de la base ϕ, qui est elle même déterminée par V_h.

Nous pourrions démontrer par la suite que la méthode de Galerkin est bornée par rapport à la dimension de V_h, ce qui montre sa stabilité. La méthode de Galerkin permet alors de trouver des solutions au problème aux différences finies, pour peu que l'on connaisse V_h. Une méthode possible pour trouver V_h est la méthode des *éléments finis*: en résumé, elle consiste ici à discrétiser l'espace de recherche, puis d'approximer la solution de l'équation différentielle interpolant la solution en chaque point discrétisé de manière à avoir un résultat satisfaisant (continu notamment).
Nous ne traiterons pas cette méthode ici par manque de temps.

4 Introduction à la résolution du problème en dimension 2

5 Applications

5.1 Conduction de la chaleur

5.2 Déformation d'une corde élastique

Dans le quarteroni page 433

5.3 Problème stationnaire elliptique

5.4 Problème hyperbolique