Méthodes numériques avancées pour la finance.
— Méthodes de splitting —

Brigitte Bidégaray-Fesquet

Cours de M2R — 2010–2011

1 Les modèles d'équations aux dérivées partielles en mathématiques financières

1.1 Le modèle de Black et Scholes

L'équation de Black et Scholes [4] est l'équation satisfaite par le prix d'une option (de vente, put) européenne. On note $u(t,s)$ ce prix qui dépend du temps $t \in]0,T]$ (dont on a retourné le sens) et de la valeur du cours $s > 0$. L'équation de Black et Scholes s'écrit

$$\partial_t u(t,s) - \frac{1}{2}\sigma^2(t,s)s^2\partial^2_{ss}u(t,s) - r(t)s\partial_s u(t,s) + r(t)u(t,s) = 0.$$

Ce modèle est paramétré par la volatilité $\sigma(t,s)$ du prix de l'action et le taux d'intérêt sans risque $r(t)$. Cette équation est assortie d'une donnée initiale, qui est la fonction payoff

$$u(0,s) = u_0(s) = (K-s)_+ \equiv \max(K-s,0),$$

où K est le prix d'exercice de l'option. Pour ce type de put, il existe une solution explicite à l'EDP (qui a valu le prix Nobel d'économie à Robert Merton et Myron Scholes en 1997, Fischer Black étant mort en 1995).

En revanche, ce n'est pas le cas si on complexifie cette équation pour mieux correspondre à la réalité des marchés financiers.

Tout d'abord, il paraît naturel de généraliser au cas d'options portant sur plusieurs actifs de valeurs $s = (s_1,\dots,s_k) \in (\mathbb{R}^+)^k$, on alors l'équation

$$\partial_t u(t,s) - \frac{1}{2}\sum_{i,j=1}^{k} \xi_{ij}(t,s)s_i s_j \partial^2_{s_i s_j}u(t,s) - \sum_{i=1}^{k} r(t)s_i\partial_{s_i}u(t,s) + r(t)u(t,s) = 0.$$

Les coefficients de la matrice ξ sont

$$\xi_{ij}(t,s) = \begin{cases} \sigma^2_{ii}(t,s), & \text{si } i = j, \\ p\sigma_{ii}(t,s)\sigma_{jj}(t,s), & \text{sinon, avec } -\frac{1}{k-1} < p < 1, \end{cases}$$

ce qui est fait que cette matrice est symétrique définie positive. Cette propriété est fondamentale pour que le problème de Cauchy (i.e. aux données initiales) soit bien posé, c'est-à-dire ait une solution unique pour tout temps. Il y a alors différentes façons de choisir la fonction de payoff, classiquement

- $u_0(s) = (K - \frac{1}{k}\sum_{i=1}^{k} s_i)_+$;
- $u_0(s) = (K - \max_i s_i)_+$;
- $u_0(s) = (K - \min_i s_i)_+$.

1.2 Premières propriétés du modèle de Black et Scholes

On fait des hypothèses de bornitude de σ et r.
- Il existe deux constantes $0 < \underline{\sigma} \leqslant \bar{\sigma}$ telles que $\underline{\sigma} \leqslant \sigma(t,s) \leqslant \bar{\sigma}$ pour tout $t \in [0,T]$ et $s > 0$.
- Il existe une constante C_1 telle que $|s\partial_s\sigma(t,s)| \leqslant C_1$ pour tout $t \in [0,T]$ et $s > 0$.
- Il existe une constante C_2 telle que $0 \leqslant r(t) \leqslant C_2$ pour tout $t \in [0,T]$.

Sous ces hypothèses, et avec la condition au bord

$$u(t,0) = K\exp\left(-\int_0^t r(\tau)d\tau\right) \text{ pour tout } t \in]0,T],$$

on peut montrer que l'équation de Black et Scholes (à un actif) admet une unique solution dans un espace fonctionnel bien choisi.

Par ailleurs, le principe du maximum s'applique à cette équation, et on peut montrer que pour tout $t \in [0,T]$ et $s > 0$, on a

$$0 \leqslant u(t,s) \leqslant K\exp\left(-\int_0^T r(\tau)d\tau\right).$$

1.3 Conditions au bord

Pour calculer une solution numérique, il faut tronquer l'espace de calcul en la variable s. On définit S assez grand pour que $u(t,S) \simeq 0$ pour tout $t \in [0,T]$. La condition au bord sera alors
- de type Dirichlet : $u(t,S) = 0$ pour tout $t \in [0,T]$;
- de type Neumann : $\partial_s u(t,S) = 0$ pour tout $t \in [0,T]$;
- de type Robin : $\partial_s u(t,S) + \beta u(t,S) = 0$ pour tout $t \in [0,T]$, cette dernière condition étant plus à même de traduire le fait que, pour un β bien choisi, l'onde u sort du domaine sans se réfléchir sur la paroi artificielle en $s = S$.

1.4 Modèle de Black et Scholes en variable log

Le modèle de Black et Scholes pose des problèmes du fait de la présence de coefficients dépendant de s devant les dérivées. Il s'ensuit une grande variation de ces coefficients pour s grand et une perte de parabolicité en $s = 0$.

Pour parer à ceci, on passe classiquement en variable log. Pour cela, on pose $s = \exp(x)$ et $w(t,x) = u(t,\exp(x))$. On voit immédiatement que

$$\begin{aligned}
\partial_x w(t,x) &= e^x\partial_s u(t,e^x) = s\partial_s u(t,s),\\
\partial^2_{xx} w(t,x) &= e^x\partial_s u(t,e^x) + (e^x)^2\partial_s u(t,e^x) = s\partial_s u(t,s) + s^2\partial^2_{ss} u(t,s).
\end{aligned}$$

En remplaçant dans le modèle de Black et Scholes, on obtient

$$\partial_t w(t,x) - \frac{1}{2}\sigma^2(t,e^x)\partial^2_x w(t,x) + \left(\frac{1}{2}\sigma^2(t,e^x) - r(t)\right)\partial_x w(t,x) + r(t)w(t,x) = 0.$$

On a toujours une dépendance en x des coefficients mais uniquement à travers σ qui est bornée inférieurement et supérieurement. Les coefficients sont donc bornés et la parabolicité est toujours assurée.

On préfèrera donc cette forme pour les simulations numériques. Il faudra à nouveau assortir ce modèle d'une condition initiale

$$w(0,x) = (K - e^x)_+,$$

et surtout de conditions aux bords en tronquant le domaine de calcul à gauche, la valeur $s = 0$ étant envoyée en $x = -\infty$. On se place donc sur un domaine $x \in [\underline{X}, \bar{X}]$, avec par exemple des conditions de Dirichlet $w(t,\bar{X}) = 0$ et $w(t,\underline{X}) = K\exp\left(-\int_0^t r(\tau)d\tau\right)$.

Pour le modèle à plusieurs actifs, on trouve bien sûr

$$\partial_t w(t,\mathbf{x}) - \frac{1}{2}\sum_{i,j=1}^{k} \xi_{ij}(t,\exp(\mathbf{x}))\partial^2_{x_i x_j} w(t,\mathbf{x}) + \sum_{i=1}^{k}\left(\frac{1}{2}\sigma^2_{ii}(t,\exp(\mathbf{x})) - r(t)\right)\partial_{x_i} w(t,\mathbf{x}) + r(t)w(t,\mathbf{x}) = 0.$$

avec $\mathbf{x} = (\ln(s_1),\dots,\ln(s_k))$.

1.5 Le modèle de Heston

Dans le modèle de Heston [11], on suppose que la volatilité est aussi stochastique, alors qu'elle est déterministe dans le modèle de Black et Scholes. Pour une option européenne à un actif, le prix dépend maintenant du temps t, de la valeur du cours s et aussi de la variance ν : $u(t,s,\nu)$. Il vérifie l'équation parabolique

$$\partial_t u - \frac{1}{2}\nu s^2\partial^2_{ss}u - \rho\sigma\nu s\partial^2_{s\nu}u - \frac{1}{2}\nu\sigma^2\partial^2_{\nu\nu}u - (r_d(t) - r_f(t))s\partial_s u - \kappa(\eta-\nu)\partial_\nu u + r_d(t)u = 0.$$

Ce modèle est paramétré par le taux de réversion moyen κ, la moyenne à long terme η, la volatilité de variance $\sigma > 0$, la corrélation des deux mouvements browniens sous-jacents $\rho \in [-1,1]$ et les taux d'intérêt domestique r_d et à l'étranger r_f.

On peut également écrire ce modèle en variable log (uniquement sur la variable s, cela n'a pas de sens sur la variable ν). Ceci donne :

$$\partial_t w - \frac{1}{2}\nu\partial^2_{xx}w - \rho\sigma\nu\partial^2_{x\nu}w - \frac{1}{2}\nu\sigma^2\partial^2_{\nu\nu}w - (r_d(t) - r_f(t) - \frac{1}{2}\nu)\partial_x w - \kappa(\eta-\nu)\partial_\nu w + r_d(t)w = 0.$$

2 Difficultés de la modélisation numérique de ces modèles

Nous allons nous concentrer sur le modèle de Black et Scholes et reviendrons à la fin du cours sur une application au modèle de Heston qui présente les même difficultés.

Nous voulons faire une simulation déterministe de ces équations. Pour cela, il faut discrétiser l'équation, c'est-à-dire déterminer les valeurs de u ou w sur une grille en "espace"–temps. Chaque variable d'espace, c'est-à-dire chaque actif, est indépendante des autres, il est donc naturel de construire une grille régulière dans toutes les directions. La discrétisation des équations la plus simple est alors les différences finies.

2.1 Différences finies pour l'équation de la chaleur en dimension 1

Considérons tout d'abord l'équation de la chaleur en dimension 1

$$\partial_t u(t,x) = \partial^2_{xx}u(t,x), \text{ pour tout } t \in]0,T],\ x \in [0,X].$$

On discrétise le temps et l'espace en définissant un pas de temps $\delta t = T/N$ et $t_n = n\delta t$, $n = 0,\dots,N$, et un pas d'espace $\delta x = X/M$ et $x_i = i\delta x$, $i = 0,\dots,M$. On calcule u_i^n, qui se veut une approximation de $u(t^n,x_i)$, grâce au θ-schéma, pour $\theta \in [0,1]$,

$$\frac{u_i^{n+1} - u_i^n}{\delta t} = \theta\frac{u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}}{\delta x^2} + (1-\theta)\frac{u_{i+1}^{n} - 2u_i^{n} + u_{i-1}^{n}}{\delta x^2}.$$

La condition $\theta \in [0,1]$ est nécessaire et suffisante pour que le schéma soit consistant, c'est-à-dire que l'on approche bien la bonne éqaution. Des cas particuliers de ce schéma sont

- le cas $\theta = 0$, appelé schéma d'Euler,
- le cas $\theta = 1$, appelé schéma d'Euler rétrograde,
- le cas $\theta = 1/2$, appelé schéma de Crank–Nicolson.

On veut écrire ce schéma sous forme matricielle, pour cela on note $U^n = {}^t(u_0^n, \dots, u_M^n)$ et

$$L = \frac{1}{\delta x^2} \begin{pmatrix} -2 & 1 & \dots & \dots & 0 \\ 1 & -2 & 1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & -2 & 1 \\ 0 & \dots & \dots & 1 & -2 \end{pmatrix}.$$

Le θ-schéma se réécrit alors

$$\frac{U^{n+1} - U^n}{\delta t} = \theta L U^{n+1} + (1-\theta) L U^n.$$

La matrice L est symétrique définie négative et on en connaît les valeurs propres

$$\lambda_i = -\frac{4}{\delta x^2} \sin^2\left(\frac{\pi i}{2(M+1)}\right),$$

ce qui permet de faire des estimations d'erreur. La matrice $(I - \delta t \theta L)$ est de plus tridiagonale (seule la diagonale, principale et les deux diagonales adjacentes sont non nulles) et il existe des algorithmes explicites et efficaces pour inverser de telles matrices, voir l'appendice A.6. On montre que la condition de stabilité pour ce schéma est

$$(1-2\theta)\frac{\delta t}{\delta x^2} \leqslant \frac{1}{2}.$$

Le système est inconditionnellement stable si $\theta \geqslant 1/2$ (pas de condition sur le pas de temps). Pour $\theta = 0$, qui a l'avantage d'être explicite, c'est-à-dire de ne pas nécessiter d'inversion de matrice, mais uniquement des produits matrice–vecteur, la condition de stabilité devient $\delta t \leqslant \delta x^2/2$ qui est très contraignante en pratique, obligeant à prendre de nombreux pas de temps.

2.2 Grandes dimensions – gestion du calcul

Par ailleurs, nous voulons traiter l'équation de Black et Scholes à plusieurs actifs. Pour fixer les idées, pour l'équation de la chaleur en dimension 2

$$\partial_t u(t,x,y) = \partial^2_{xx} u(t,x,y) + \partial^2_{yy} u(t,x,y), \text{ pour tout } t \in]0,T],\ x \in [0,X],\ y \in [0,Y],$$

si on utilise cette fois-ci deux pas d'espace δx, δy et une approximation $u_{i,j}$ de $u(x_i, y_j)$ (on omet ici la variable de temps), la discrétisation de l'opérateur $\partial^2_{xx} u + \partial^2_{yy} u$ s'écrit

$$\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{\delta x^2} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{\delta y^2},$$

faisant intervenir 5 points. Dans le cas général d'une équation en dimension k, on peut écrire une discrétisation du laplacien en utilisant $1 + 2k$ points.

Faisons une rapide estimation de la dimension des données. Imaginons que l'on considère une équation de Black et Scholes à $k = 4$ actifs, et que l'on discrétise chaque variable d'espace avec 100 points (valeur tout à fait raisonnable). Sans parler du nombre d'itérations en temps à réaliser, la grille en espace compte déjà $100^k = 10^8$ points de discrétisation, et donc autant de valeurs de u à calculer à chaque pas de temps. Si on stockait la matrice L comme une matrice pleine, on aurait 10^{16} valeurs à stocker. En pratique, on a vu qu'il y avait que 1+2k valeurs non nulles sur chaque ligne, mais cela fait tout de même la bagatelle $9\ 10^8$ valeurs ! En pratique, dans le cas d'un maillage régulier, on ne stocke pas la matrice, mais uniquement les valeurs des $1/\delta x_i^2$. Il faut cependant stocker et manipuler une ou plusieurs copies de la variable u, ce qui n'est pas ingérable, mais demande tout de même de travailler un peu finement.

2.3 L'idée du splitting

Une idée consiste à dire que résoudre $\partial_t u = \partial^2_{xx} u + \partial^2_{yy} u$ sur un intervalle de temps $[0, T]$, c'est "presque" comme résoudre $\partial_t u = \partial^2_{xx} u$ sur cet intervalle de temps, puis prendre le résultat comme nouvelle donnée initiale pour le problème $\partial_t u = \partial^2_{yy} u$ que l'on résout à nouveau sur le même intervalle de temps $[0, T]$. Cette méthode est dite des "directions alternées", qui est un cas particulier des méthodes de splitting, que nous allons dûment justifier dans la suite.

Si on admet pour l'instant cette idée et qu'on l'étend à la dimension k, on se ramène à résoudre en k étapes l'équation de la chaleur, chaque étape faisant intervenir une matrice avec 3 coefficients non nuls sur chaque ligne (donnant accès à des algorithmes efficaces dédiés). De plus, les 100^k variables de l'exemple précédent, se trouvent groupées à chaque étape en 100^{k-1} paquets indépendants de 100 variables, permettant une parallélisation naturelle des algorithmes.

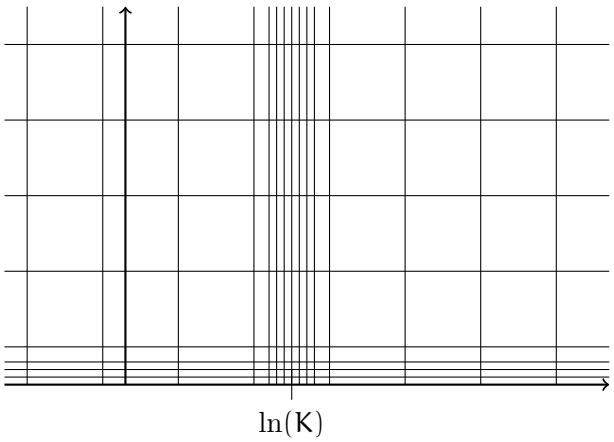
2.4 Problème des autres dérivées

Nous remarquons que, dans l'équation de Black et Scholes et aussi dans l'équation de Heston, nous avons aussi des dérivées secondes mixtes du type ∂^2_{xy}. Pour celles-là, il sera impossible d'alterner les directions et il faut trouver une autre idée, qui constituera, comme nous le verrons, une première étape nécessairement explicite des méthodes proposées, induisant une perte de stabilité en grande dimension.

Les dérivées premières sont quant à elles naturellement décomposables selon les différentes directions. Il faudra seulement être vigilants au signe du coefficient qui les précède pour choisir le schéma au différences finies le plus adapté, décentré en amont ou en aval.

2.5 Problème de la donnée initiale non régulière

La donnée initiale est non dérivable. L'équation de Black et Scholes est une équation parabolique qui régularise immédiatement cette donnée initiale, c'est-à-dire qu'elle devient infiniment dérivable pour tout temps $t > 0$. Suivant la fonction de pay-off, on peut être tenté de raffiner le maillage en espace autour des $s_i = K$ et aussi en temps, pour les premiers pas de temps. Ce raffinement oblige à définir légèrement différemment la discrétisation du Laplacien dont les coefficients vont alors dépendre du point de la grille, ce qui nécessitera un stockage effectif des matrices de type L sous forme de matrice bande.



2.6 Les conditions aux bords

Les conditions aux bords doivent être imposées à chaque itération en temps. Nous sommes ici dans un cas assez simple puisque l'on se ramène à des problèmes en dimension 1 à chaque étape.

Si, par exemple, les points du maillage sont les $x_i = i\delta x$, $i = 0, \dots, M$, on ne calcule les u_i^n que pour $i = 1, \dots, M-1$. Pour simplifier nous regardons le cas explicite ($\theta = 0$), le cas général se traite de la même manière.

Si on a une condition de Dirichlet, $u = g$ en 0, on remplace la formule générale

$$\frac{u_i^{n+1} - u_i^n}{\delta t} = \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\delta x^2}$$

par

$$\frac{u_1^{n+1} - u_1^n}{\delta t} = \frac{u_2^n - 2u_1^n + g}{\delta x^2}$$

pour $i = 1$. La première ligne de la matrice L reste alors identique et on a un second membre dans l'équation matricielle.

Si en revanche, on a une condition de Neumann homogène, $\partial_x u = 0$ en 0, on impose alors $u_0 = u_1$, et donc

$$\frac{u_1^{n+1} - u_1^n}{\delta t} = \frac{u_2^n - u_1^n}{\delta x^2},$$

ce qui fait que la première ligne de L devient

$$\frac{1}{\delta x^2}(-1\ 1\ 0\ \dots\ 0).$$

Enfin, pour une condition de Robin, $\partial_x u + \beta u = 0$, on écrit (par exemple, cela dépend en fait du signe de β) $(u_1^n - u_0^n)/\delta x + \beta u_0^n = 0$, et donc $u_0^n = u_1^n/(1 - \beta\delta x)$, la première ligne de L devenant

$$\frac{1}{\delta x^2}(-\frac{1 - 2\beta\delta x}{1 - \beta\delta x}\ 1\ 0\ \dots\ 0).$$

2.7 Complexité

Le splitting permet de se ramener à des étapes de calcul relativement simples avec uniquement des produits matrice–vecteurs et des méthodes d'inversion très simples d'inversion des matrices tri-diagonales, algorithmes qui sont la plupart du temps implémentés sans stocker la matrice. Étant donné la taille du problème, même pour un nombre modéré d'actifs, une implémentation efficace passe clairement par une réflexion sur le bon équilibre entre calcul et stockage. Par ailleurs, il faut traiter proprement la réduction à chaque étape d'un gros problème en multiple petits sous-problèmes parallélisables.

3 Introduction aux méthodes de splitting

On traduit souvent en français le terme de « méthode de splitting » par « méthode des pas fractionnaires », ce qui de mon point de vue constitue un contresens, car les pas de temps sont souvent entiers, c'est l'opérateur qui est fractionné.

3.1 Une équation scalaire

Considérons l'équation différentielle ordinaire (EDO) scalaire suivante :

$$\text{(EDO) } \dot{x} = (a + b)x, \qquad x(0) = x^0,$$

où a et b sont des scalaires. On connaît la solution exacte de cette équation :

$$\begin{aligned} x(t) = \exp((a+b)t)x^0 &= \exp(at)\exp(bt)x^0 \qquad \text{(méthode 1)} \\ &= \exp(bt)\exp(at)x^0 \qquad \text{(méthode 2).} \end{aligned}$$

Nous pouvons ainsi séparer l'évolution selon l'équation (EDO) en deux temps :

$$\text{(L1)} \left| \begin{array}{ll} \dot{y} = by, & y(0) = x^0, \\ \dot{x} = ax, & x(0) = y(t), \end{array} \right. \qquad \text{(L2)} \left| \begin{array}{ll} \dot{y} = ay, & y(0) = x^0, \\ \dot{x} = bx, & x(0) = y(t). \end{array} \right.$$

Pour le système (L1), on a clairement

$$x(t) = \exp(at)x(0) = \exp(at)y(t) = \exp(at)\exp(bt)y(0) = \exp(at)\exp(bt)x^0.$$

Le calcul pour (L2) se fait de la même manière et donne le même résultat. On appelle **splitting de Lie** les deux méthodes (L1) et (L2). Pour une équation scalaire, ces deux méthodes sont identiques et reviennent au même que de traiter l'équation en une seule fois.

3.2 Quand le splitting présente un intérêt

Pour cette équation scalaire, le splitting n'a pas l'air de présenter un intérêt. En fait, l'exemple donné est à peu près le seul pour lequel le splitting ne change pas la solution. Le splitting a un intérêt et un impact dans de nombreux contextes théoriques comme par exemple

(a) les systèmes différentiels linéaires : $\dot{x} = (A + B)x$ (dès que les matrices A et B ne commutent pas),
(b) les systèmes différentiels linéaires avec deux échelles différentes : $\dot{x} = (\frac{1}{\varepsilon}A + B)x$,
(c) les équations aux dérivées partielles non linéaires : par exemple $\partial_t u = \Delta u + f(u)$,
(d) les systèmes en grande dimension d'espace,

et leur approximation numérique. Pour les cas (b) et (c), on peut se reporter aux cours [2, 3]. Nous nous concentrons ici sur les cas (a) et (d).

Le splitting est utilisé dans de nombreux contextes applicatifs. Nous pouvons citer notamment :
- la **chimie complexe** (traitement séparé des phénomènes de réaction chimique (équations non linéaires) et de diffusion des espèces. Ce sont des systèmes avec un très grand nombre de variables et des échelles de temps très différentes.
- la **météorologie, l'océanographie**. Là encore, la multiplicité des phénomènes mis en jeu, donne lieu à une multiplicité de termes de natures très différentes.
- la **décomposition de domaine**. Celle-ci est utilisée lorsqu'il y a couplage de phénomènes sur des domaines différents (adjacents) ou avec des géométries très différentes. Dans la zone d'interaction, il y a redondances des variables et on peut souvent séparer l'opérateur d'évolution en un opérateur facile à intégrer et un autre petit en un certain sens.
- les **méthodes d'ondelettes**. À nouveau, les termes diagonaux des opérateurs sont prépondérants et faciles à intégrer et les termes extra-diagonaux sont moralement petits, c'est d'ailleurs là tout l'intérêt de la méthode.
- les **mathématiques financières**. Il faut traiter des systèmes d'équations aux dérivées partielles en grande dimension. On sépare donc les dimensions.
- et bien d'autres ...

Deux intérêts principaux du splitting peuvent d'ores et déjà être identifiés :
- le fait de pouvoir résoudre exactement ou numériquement chacune des sous-équations alors que cela est impossible ou difficile avec l'équation entière,
- le fait de pouvoir traiter séparément des variables ou des opérateurs correspondant à des échelles très différentes.

Un inconvénient apparaît aussi immédiatement comme l'impossibilité de conserver les propriétés fines liées à la structure de certaines équations et mettant en œuvre toute l'équation (quantités conservées).

Dans l'analyse générale des méthodes de splitting, nous nous intéressons exclusivement aux semi-discrétisations en temps. La discrétisation en espace peut faire *a priori* appel à n'importe quelle technique adaptée (différences finies, éléments finis, volumes finis, méthodes spectrales,...), et éventuellement différentes pour chacune des parties de l'équation. C'est là tout l'intérêt de la chose.

3.3 Plan du cours

Après avoir défini la notation des semi-groupes pour simplifier les écritures, le plan du cours sera le suivant.

partie 4. Définition et analyse de méthode de splitting pour des systèmes d'évolution linéaires ;

partie 5. Discrétisation en temps de ces systèmes d'évolution linéaires ;

partie 6. Généralités sur l'équation de la chaleur, y compris avec des non-linéarités ;

partie 7. Application aux modèles de Black et Scholes et de Heston.

3.4 Réécriture des schémas de splitting grâce aux semi-groupes d'évolution

Pour exprimer le splitting dans notre exemple simpliste, nous avons été obligés d'introduire une variable intermédiaire y. Ceci s'avérerait à l'usage peu pratique pour des contextes plus compliqués. Nous introduisons donc une notation de type semi-groupe d'évolution. L'application qui à x^0 associe $x(t)$ par le flot de l'EDO est le **semi-groupe d'évolution** que nous noterons $\mathcal{S}(t)$ pour l'équation toute entière :

$$x(t) = \mathcal{S}(t)x^0 = \exp((a+b)t)x^0.$$

Quand l'opérateur est linéaire, on garde aussi souvent la notation avec l'exponentielle. Si on note $\mathcal{A}(t)$ et $\mathcal{B}(t)$, les semi-groupes d'évolution associés aux deux parties de l'équation, à savoir

$$\mathcal{A}(t)x^0 = \exp(at)x^0, \qquad \mathcal{B}(t)x^0 = \exp(bt)x^0,$$

les deux splittings (L1) et (L2) consistent à écrire

$$\text{(L1) } x(t) = \mathcal{A}(t)\mathcal{B}(t)x^0, \qquad \text{(L2) } x(t) = \mathcal{B}(t)\mathcal{A}(t)x^0.$$

Avec cette notation, on définit sans effort (et sans introduction de variables supplémentaires) deux nouveaux types de splitting : les **splittings de Strang** [27]

$$\text{(S1) } x(t) = \mathcal{A}(\frac{t}{2})\mathcal{B}(t)\mathcal{A}(\frac{t}{2})x^0, \qquad \text{(S2) } x(t) = \mathcal{B}(\frac{t}{2})\mathcal{A}(t)\mathcal{B}(\frac{t}{2})x^0.$$

Évidemment, pour notre premier exemple ces deux splittings sont toujours équivalents à l'équation initiale.

Qu'entend-t-on par **semi-groupe d'évolution** ? L'opérateur $\mathcal{A}(t)$ qui est défini de $\mathbb{R}$ dans $\mathbb{R}$ a les deux propriétés suivantes :

$$\begin{aligned} &\text{(P1) } \mathcal{A}(0) = I, && (\text{cf. } \exp(0) = 1),\\ &\text{(P2) } \mathcal{A}(t+s) = \mathcal{A}(t)\mathcal{A}(s), \text{ pour tout } t, s \geqslant 0 && (\text{cf. } \exp(a(t+s)) = \exp(at)\exp(as)). \end{aligned}$$

La propriété (P1) dit que l'évolution pendant un temps nul donne la donnée initiale. La propriété (P2) dit que cela revient au même d'évoluer pendant un temps $t+s$, ou d'évoluer pendant un temps s puis un temps t. Ceci est vrai parce que le système est autonome, sinon cela est faux. Dans notre exemple, on a même un groupe, car $\mathcal{A}(t)$ est défini pour des temps t négatifs, mais cela n'est pas toujours le cas. Certaines équations sont mal posées «en rétrograde», comme typiquement les équations paraboliques que nous allons étudier pour l'application en mathématiques financières.

Les propriétés de semi-groupe peuvent être énoncées dans un cadre beaucoup plus général que celui. Quelques pistes sont données à l'appendice A.1. Pour en savoir plus consulter par exemple le livre de Pazy [22].

4 Le cas des systèmes linéaires

On considère maintenant le système linéaire

$$(S)\ \dot{x} = (A+B)x, \qquad x(0) = x^0,$$

où $x \in \mathbb{R}^d$ et A et B sont des matrices de $\mathcal{M}_d$. C'est une généralisation du cas scalaire. La matrice A est la représentation d'un opérateur linéaire dans la base canonique. Le semi-groupe d'évolution $\mathcal{A}(t)$ s'exprime dans cette même base par la matrice $\exp(tA)$.

Si les matrices A et B commutent, on a, par exemple, $\exp(tA)\exp(tB) = \exp(t(A+B))$, ce qui limite l'intérêt du splitting (du moins de son analyse mathématique). On s'intéressera donc au cas où A et B ne commutent pas et on définira le **commutateur** ou encore **crochet de Lie** par

$$[A,B] = AB - BA.$$

4.1 Splitting de Lie

Dans notre cadre, les matrices A et B jouent des rôles symétriques, on ne regarde donc qu'un seul cas de splitting de Lie.

$$\begin{aligned}
\mathcal{A}(t)\mathcal{B}(t) - \mathcal{S}(t) &= \left(I + tA + \frac{t^2}{2}A^2\right)\left(I + tB + \frac{t^2}{2}B^2\right) - \left(I + t(A+B) + \frac{t^2}{2}(A+B)^2\right) + O(t^3) \\
&= t^2\left(\frac{1}{2}A^2 + AB + \frac{1}{2}B^2\right) - \frac{t^2}{2}\left(A^2 + AB + BA + B^2\right) + O(t^3) \\
&= \frac{t^2}{2}[A,B] + O(t^3).
\end{aligned}$$

Ce développement permet de démontrer l'ordre de la méthode. Nous ferons la preuve plus loin dans un cadre plus général.

4.2 Splitting de Strang

Étudions de même la formule de Strang

$$\begin{aligned}
\mathcal{A}(\frac{t}{2})\mathcal{B}(t)\mathcal{A}(\frac{t}{2}) - \mathcal{S}(t) &= \left(I + \frac{t}{2}A + \frac{t^2}{8}A^2 + \frac{t^3}{48}A^3\right)\left(I + tB + \frac{t^2}{2}B^2 + \frac{t^3}{6}B^3\right) \\
&\quad \times \left(I + \frac{t}{2}A + \frac{t^2}{8}A^2 + \frac{t^3}{48}A^3\right) \\
&\quad - \left(I + t(A+B) + \frac{t^2}{2}(A+B)^2 + \frac{t^3}{6}(A+B)^3\right) + O(t^4) \\
&= t^3\left(\frac{1}{6}A^3 + \frac{1}{8}A^2B + \frac{1}{4}ABA + \frac{1}{8}BA^2 + \frac{1}{4}B^2A + \frac{1}{4}AB^2 + \frac{1}{6}B^3\right) \\
&\quad - \frac{t^3}{6}\left(A^3 + A^2B + ABA + BA^2 + B^2A + BAB + AB^2 + B^3\right) + O(t^4) \\
&= t^3\left(-\frac{1}{24}A^2B + \frac{1}{12}ABA - \frac{1}{24}BA^2 + \frac{1}{12}B^2A - \frac{1}{6}BAB + \frac{1}{12}AB^2\right) + O(t^4) \\
&= t^3\left(-\frac{1}{24}[A,[A,B]] + \frac{1}{12}[B,[B,A]]\right) + O(t^4).
\end{aligned}$$

4.3 Splittings d'ordre plus élevé

4.3.1 Calcul de l'ordre et convergence

Les deux résultats précédents sont deux cas particuliers du résultat suivant.

Théorème 1

Soit $C \in \mathcal{M}_d$ *et* f *une fonction continue définie sur un voisinage de* 0 *dans* $\mathbb{R}$ *à valeurs dans* $\mathcal{M}_d$*, tels qu'il existe une matrice* $R \in \mathcal{M}_d$ *et un entier* p *tels que le développement limité*

$$\text{(DL)} \qquad f(t) - \exp(tC) = Rt^{p+1} + O(t^{p+2})$$

soit vrai dans un voisinage de 0*. Alors*

$$f(\frac{t}{n})^n - \exp(tC) = O((\frac{t}{n})^p).$$

De plus, cette estimation est optimale, sauf si R *est identiquement nulle.*

La preuve de ce théorème est donnée à l'appendice A.2.

Ce théorème assure que n applications de $f(t/n)$ à x^0 fournit ainsi une méthode d'ordre p pour approcher $x(t)$. Les splittings de Lie et de Strang correspondent aux cas $p = 1$ et $p = 2$ respectivement. En effet

$$\left\| \left(\mathcal{A}(\frac{t}{n}) \mathcal{B}(\frac{t}{n}) \right)^n x^0 - x(t) \right\| = O(\frac{t}{n})$$

(cette formule est une version dans le cadre matriciel de la formule de Trotter–Kato que nous verrons plus loin) et

$$\left\| \left(\mathcal{A}(\frac{t}{2n}) \mathcal{B}(\frac{t}{n}) \mathcal{A}(\frac{t}{2n}) \right)^n x^0 - x(t) \right\| = O((\frac{t}{n})^2).$$

Les méthodes de Lie et de Strang sont respectivement d'ordre 1 et 2.

Ceci justifie l'utilisation du splitting comme méthode numérique, en admettant que l'on sache calculer de manière exacte ou suffisamment précise les exponentielles $\exp(tA/n)$ et $\exp(tB/n)$.

L'ordre 1 et *a fortiori* les ordres supérieurs suffisent à montrer la consistance des méthodes. La stabilité sera assurée systématiquement par la stabilité de chacun des schémas pour résoudre les différentes parties du splitting. Il est bien connu (théorème de Lax–Richtmyer) que consistance et stabilité assurent la convergence de la méthode.

Une question se pose : comment construire des méthodes d'ordre plus élevé ?

4.3.2 Une réponse négative

La première réponse est négative dans le cas où on cherche des coefficients positifs. Ce résultat est dû à Michelle Schatzman [24]. On cherche des coefficients α_j et β_j tels que la fonction

$$f_k(t, \alpha, \beta) = \exp(\alpha_1 tA) \exp(\beta_1 tB) \dots \exp(\alpha_k tA) \exp(\beta_k tB)$$

vérifie le développement limité du théorème précédent.

Théorème 2

Si $[A, [A, B]]$ *et* $[B, [B, A]]$ *sont linéairement indépendants, il n'existe aucun choix de* k *et des coefficients* α_j *et* β_j *réels positifs pour obtenir (DL) pour* $p \geqslant 3$*.*

On remarque que les commutateurs qui interviennent sont ceux du reste dans la méthode d'ordre 2 de Strang.

Le résultat est même plus fort, on peut généraliser la forme de la fonction recherchée en

$$f(t) = r_1(tA)s_1(tB)\dots r_k(tA)s_k(tB),$$

où les fonctions r_j et s_j sont analytiques sur $\mathbb{R}$ avec $r_j(0) = s_j(0) = 1$, $r_j'(0) = \alpha_j$ et $s_j'(0) = \beta_j$. Cette généralisation inclut la plupart des approximations numériques de l'exponentielle.

Théorème 3

Si

(i) α_j *et* β_j *sont positifs,*

(ii) $[A, B]$, A^2 *et* B^2 *sont linéairement indépendants,*

(iii) $[A, [A, B]]$, $[B, [B, A]]$, $[A^2, B]$, $[A, B^2]$, A^3 *et* B^3 *sont linéairement indépendants,*

alors il n'existe aucun choix de k *et des fonctions* r_j *et* s_j *pour obtenir (DL) pour* $p \geqslant 3$.

On peut contourner le problème en choisissant certains des α_j ou β_j négatifs, mais cette approche n'est pas intéressante pour l'application qui nous intéressent dans laquelle les équations ne sont pas bien posées en rétrograde.

4.3.3 Combinaison linéaires d'approximations d'ordre inférieur

On se rappelle que

$$\mathcal{A}(t)\mathcal{B}(t) - \mathcal{S}(t) = \frac{t^2}{2}[A, B] + O(t^3).$$

On a donc

$$\frac{1}{2}(\mathcal{A}(t)\mathcal{B}(t) + \mathcal{B}(t)\mathcal{A}(t)) - \mathcal{S}(t) = O(t^3).$$

En poussant plus loin les développements, on voit apparaître des termes faisant intervenir les commutateurs $[A, [A, B]]$ et $[B, [B, A]]$. On peut essayer de les annuler avec ceux de $\mathcal{A}(t/2)\mathcal{B}(t)\mathcal{A}(t/2) - \mathcal{S}(t)$ et $\mathcal{B}(t/2)\mathcal{A}(t)\mathcal{B}(t/2) - \mathcal{S}(t)$. Ceci donne la formule

$$g(t) \quad = \quad \frac{2}{3}\left(\mathcal{A}(\frac{t}{2})\mathcal{B}(t)\mathcal{A}(\frac{t}{2}) + \mathcal{B}(\frac{t}{2})\mathcal{A}(t)\mathcal{B}(\frac{t}{2})\right) - \frac{1}{6}\left(\mathcal{A}(t)\mathcal{B}(t) + \mathcal{B}(t)\mathcal{A}(t)\right),$$

pour laquelle

$$g(t) - \exp(t(A + B)) = -\frac{t^4}{24}[A, B]^2 + O(t^5).$$

4.3.4 Extrapolations de Richardson

L'extrapolation de Richardson est une méthode générale pour accélérer la convergence d'une méthode numérique d'intégration des EDO. Présentons la d'abord dans ce cadre.

On suppose que l'on approche la solution exacte $y(t)$ d'une EDO par une méthode dépendant d'un pas h et calculant une approximation $y(t; h)$ telle que

$$y(t; h) = y(t) + h^p g(t) + O(h^{p+1})$$

qui est donc d'ordre local p. Si au lieu d'utiliser le pas de temps h, on utilise le pas de temps qh, on a

$$y(t; qh) = y(t) + (qh)^p g(t) + O(h^{p+1}).$$

On peut calculer la combinaison linéaire

$$\frac{q^p y(t;h) - y(t;qh)}{q^p - 1} = y(t) + O(h^{p+1})$$

qui donne une nouvelle méthode qui est d'ordre local au moins $p+1$.

Adaptons ceci aux méthodes de splitting. La solution exacte est $S(t)x_0$. Si on calcule avec une seule itération de splitting, on calcule $f(t)x_0$ qui est l'équivalent de $y(t;h)$. On choisit $q = 1/2$, ce qui revient à faire deux itérations de la méthodes avec un pas de temps moitié : $f(t/2)f(t/2)x_0$. La nouvelle méthode s'écrit :

$$\frac{\frac{1}{2^p} f(t) - f(\frac{t}{2})f(\frac{t}{2})}{\frac{1}{2^p} - 1} = \frac{2^p f(\frac{t}{2})f(\frac{t}{2}) - f(t)}{2^p - 1}.$$

Appliquons ceci à un splitting de Lie $\mathcal{L}_1(t) = \mathcal{A}(t)\mathcal{B}(t)$ pour lequel $p = 1$. On obtient le schéma

$$\mathcal{RL}_1(t) = 2\mathcal{A}(\frac{t}{2})\mathcal{B}(\frac{t}{2})\mathcal{A}(\frac{t}{2})\mathcal{B}(\frac{t}{2}) - \mathcal{A}(t)\mathcal{B}(t)$$

qui n'est que d'ordre 2, donc pas meilleur qu'un splitting de Strang tout en demandant plus de calculs. En outre, ce schéma a l'inconvénient de comporter des signes moins. On oublie donc.

Appliquons ceci à un splitting de Strang $\mathcal{S}_1(t) = \mathcal{A}(t/2)\mathcal{B}(t)\mathcal{A}(t/2)$ pour lequel $p = 2$. On obtient le schéma

$$\mathcal{RS}_1(t) = \frac{1}{3}\left(4\mathcal{A}(\frac{t}{4})\mathcal{B}(\frac{t}{2})\mathcal{A}(\frac{t}{4})\mathcal{A}(\frac{t}{4})\mathcal{B}(\frac{t}{2})\mathcal{A}(\frac{t}{4}) - \mathcal{A}(\frac{t}{2})\mathcal{B}(t)\mathcal{A}(\frac{t}{2})\right).$$

On peut combiner les deux $\mathcal{A}(t/4)$ centraux et obtient le nouveau schéma

$$\begin{aligned} g(t) \quad = \quad & \frac{4}{3}\exp(\frac{1}{4}tA)\exp(\frac{1}{2}tB)\exp(\frac{1}{2}tA)\exp(\frac{1}{2}tB)\exp(\frac{1}{4}tA) \\ & -\frac{1}{3}\exp(\frac{1}{2}tA)\exp(tB)\exp(\frac{1}{2}tA). \end{aligned}$$

Cette formule est construite à partir de deux formules d'ordre 2 pour être d'ordre 3. En fait, elle est d'ordre 4 et est utilisée en pratique contrairement à la précédente.

5 Approximation de l'exponentielle

Tous les calculs précédents ne sont applicables en pratique que pour des matrices dont on sait calculer facilement l'exponentielle. Elles ne sont pas si nombreuses. Entrent dans ce cadre

- les **matrices diagonales** :
  $D = \mathrm{diag}(d_k)$. On a alors $\exp(tD) = \mathrm{diag}(\exp(td_k))$.
- les **matrices sous forme de Jordan** :

$$J = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & 0 & & \lambda \end{pmatrix}.$$

Il est tout d'abord facile de calculer leurs puissances :

$$J^k = \begin{pmatrix} \lambda^k & k\lambda^{k-1} & \dots & C_k^{k-(d-1)}\lambda^{k-(d-1)} \\ & \ddots & \ddots & \\ & & \ddots & k\lambda^{k-1} \\ & 0 & & \lambda^k \end{pmatrix}.$$

De manière générique, on a pour $j \geqslant i$, $J_{ij}^k = C_k^{k-(j-i)}\lambda^{k-(j-i)}$. On a alors

$$\begin{aligned}\exp(tJ)_{ij} &= \sum_{k=0}^{\infty}\frac{1}{k!}t^k J_{ij}^k = \sum_{k=j-i}^{\infty}\frac{1}{k!}t^k C_k^{k-(j-i)}\lambda^{k-(j-i)}\\ &= \sum_{k=0}^{\infty}\frac{1}{(k+(j-i))!}t^k t^{j-i}C_{k+(j-i)}^k\lambda^k = \sum_{k=0}^{\infty}\frac{1}{k!(j-i)!}t^k t^{j-i}\lambda^k\\ &= \frac{t^{j-i}}{(j-i)!}\sum_{k=0}^{\infty}\frac{1}{k!}(t\lambda)^k = \frac{t^{j-i}}{(j-i)!}\exp(t\lambda).\end{aligned}$$

– les **matrices idempotentes** comme les **projecteurs** : $P^n = P$ pour tout $n \geqslant 1$. On a alors

$$\exp(tP) = \sum_{n=0}^{\infty}\frac{(tP)^n}{n!} = I + \sum_{n=1}^{\infty}\frac{t^n P}{n!} = I - P + \sum_{n=0}^{\infty}\frac{t^n}{n!}P = I + (\exp(t)-1)P.$$

Dans les autres cas, il faut avoir recours à des approximations à différents ordres de l'exponentielle.

On suppose que l'on discrétise le temps avec un pas constant h : $t_n = nh$. On cherche à calculer une approximation x^n de $x(t_n)$. Il s'agit d'écrire une relation de récurrence qui relie x^{n+1} à x^n. Celle-ci résulte de la discrétisation du système différentiel écrit entre les temps t_n et t_{n+1}.

$$\dot{x} = (A+B)x, \qquad x(t_n) = x^n,$$

ou de son équivalent intégral

$$x(t) = x(t_n) + \int_{t_n}^{t}(A+B)x(\tau)d\tau.$$

Pour pouvoir analyser ce qui suit, donnons le développement limité de x autour de $t = t_n$:

$$x(t_n+h) = (I + (A+B)h + \frac{1}{2}(A+B)^2h^2 + \frac{1}{6}(A+B)^3h^3 + O(h^4))x(t_n).$$

On a donc

$$\begin{aligned}\mathcal{S}(h) &= (I + (A+B)h + \frac{1}{2}(A^2+AB+BA+B^2)h^2\\ &\quad + \frac{1}{6}(A^3+A^2B+ABA+AB^2+BA^2+BAB+B^2A+B^3)h^3 + O(h^4).\end{aligned}$$

Il suffira de calculer la différence de cette quantité avec celles associées aux différents schémas, qui sont toutes des fonctions continues de h au voisinage de 0 et à valeur dans pour obtenir une expression de type

$$\text{(DL) } f(h) - \mathcal{S}(h) = Rh^{p+1} + O(h^{p+2}),$$

ce qui est l'opérateur d'erreur locale et le théorème assure que, pour $t = nh$,

$$f(h)^n - \mathcal{S}(nh) = O(h^p)$$

et la méthode est d'ordre p.

Nous ne donnons les résultats que pour les premiers schémas, (L1) et (S1), les résultats pour les autres schémas s'en déduisant clairement en échangeant le rôle de A et B.