

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

Brigitte Bidégaray-Fesquet

ENSIMAG 3A MIF — 2011–2012

Plan du cours

- Partie 1.** Introduction des équations aux dérivées partielles (déterministes) de Black et Scholes et de Heston ;
- Partie 2.** Approximation numérique aux différences finies dans le cas d'un seul actif ;
- Partie 3.** Problèmes liés à l'augmentation du nombre d'actifs ;
- Partie 4.** Généralités sur le splitting d'opérateur ;
- Partie 5.** Définition et analyse de la méthode du splitting pour des systèmes d'évolution linéaires ;
- Partie 6.** Discrétisation en temps de ces systèmes d'évolution linéaires ;
- Partie 7.** L'exemple historique : l'équation de la chaleur ;
- Partie 8.** Application aux modèles de Black et Scholes et de Heston.

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

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

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 N actifs de valeurs $s = (s_1, \dots, s_N) \in (\mathbb{R}^+)^N$, on alors l'équation

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

Les coefficients de la matrice ξ sont

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

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

- $u_0(s) = (K - \frac{1}{N} \sum_{i=1}^N s_i)_+$;
- $u_0(s) = (K - \max_{i=1}^N s_i)_+$;
- $u_0(s) = (K - \min_{i=1}^N s_i)_+$.

1.2 Premières propriétés du modèle de Black et Scholes

On fait des hypothèses de bornitude de σ et r .

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

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

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

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

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

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

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 préférera 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

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

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

1.5 Le modèle de Heston

Dans le modèle de Heston [13], 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 v : $u(t, s, v)$. Il vérifie l'équation parabolique

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

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 v). Ceci donne :

$$\partial_t w - \frac{1}{2} v \partial_{xx}^2 w - \rho v \partial_{xv}^2 w - \frac{1}{2} v \sigma^2 \partial_{vv}^2 w - (r_d(t) - r_f(t) - \frac{1}{2} v) \partial_x w - \kappa(\eta - v) \partial_v w + r_d(t) w = 0.$$

2 Approximation du modèle de Black et Scholes à un actif

Commençons par regarder un problème relativement simple, mais qui est en fait l'essence du modèle plus général. On se place dans le cadre de l'équation de Black et Scholes à un actif avec des coefficients de volatilité et de taux d'intérêt sans risque constants. On regarde donc l'équation

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

Nous voulons faire une simulation déterministe de ces équations. Pour cela, il faut discrétiser l'équation, c'est-à-dire déterminer les valeurs de w sur une grille en "espace"-temps. La discrétisation des équations la plus simple est alors les différences finies. Dans cette méthode, on remplace les opérateurs de dérivation par des opérateurs aux différences, que l'on calcule en fonction de points sur la grille de discrétisation.

2.1 Différences finies pour l'équation de la chaleur en dimension 1 — θ -schéma.

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

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

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 $\delta x = X/(M+1)$ et $x_i = i\delta x$, $i = 0, \dots, M+1$. 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 équation. 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.

En pratique u_0^n et u_{M+1}^n sont "donnés" par les conditions aux bords (cf. infra). On écrit alors le schéma sous forme matricielle en notant $U^n = {}^t(u_1^n, \dots, u_M^n)$ 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

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

ce qui permet de faire des estimations d'erreur. Le θ -schéma se réécrit

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

La matrice $(I - \delta t \theta L)$ a l'avantage d'être tri-diagonale (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. On montre que la condition de stabilité pour ce schéma est

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

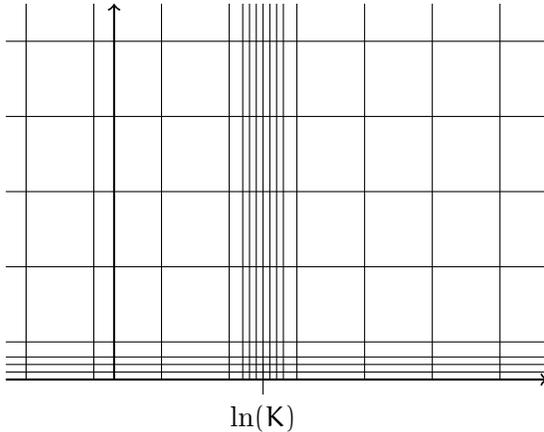
Le système est inconditionnellement stable si $\theta \geq 1/2$ (pas de condition sur le pas de temps). Pour $\theta = 0$, qui a l'avantage d'être explicite, c'est-à-dire de ne pas nécessiter d'inversion de matrice, mais uniquement des produits matrice-vecteur, la condition de stabilité devient $\delta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

2.2 Discrétisation du laplacien sur des grilles irrégulières

Dans le paragraphe précédent, nous avons utilisé une grille régulière en espace. L'utilisation de grilles irrégulières peut se justifier de deux façons dans notre application aux équations de Black et Scholes.

1. Tout d'abord, nous pouvons vouloir discrétiser de manière régulière l'espace des prix dans la variable s . Lorsque l'on passe en variable log, les intervalles réguliers en s deviennent irréguliers en x .

2.



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.

Dans les deux cas, la non-uniformité du maillage oblige à revoir 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 (alors qu'avant absolument aucun stockage n'était nécessaire, à part la valeur de $\delta t / \delta x^2$).

On utilise alors l'approximation de la dérivée seconde en x_i :

$$\partial_{xx}^2 u(x_i) \simeq \frac{2}{x_{i+1} - x_{i-1}} \left(\frac{1}{x_{i+1} - x_i} u(x_{i+1}) - \left(\frac{1}{x_{i+1} - x_i} + \frac{1}{x_i - x_{i-1}} \right) u(x_i) + \frac{1}{x_i - x_{i-1}} u(x_{i-1}) \right),$$

qui généralise l'approximation déjà utilisée dans le cas de pas constants ($\delta x = x_{i+1} - x_i = x_i - x_{i-1}$) :

$$\partial_{xx}^2 u(x_i) \simeq \frac{1}{\delta x^2} (u(x_{i+1}) - 2u(x_i) + u(x_{i-1})).$$

2.3 Les conditions aux bords

Les conditions aux bords doivent être imposées à chaque itération en temps.

Pour simplifier, nous expliquons le traitement des conditions aux bord dans le cas implicite ($\theta = 1$), le cas général se traite de la même manière. Toujours pour simplifier nous considérons le cas du maillage régulier. Si les points du maillage sont les $x_i = i\delta x$, $i = 0, \dots, M+1$, on ne calcule les u_i^{n+1} que pour $i = 1, \dots, M$.

2.3.1 Condition de Dirichlet

Si on a une condition de Dirichlet, $u(t, x = 0) = g(t)$, on remplace la formule générale

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

par

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

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

Le θ -schéma présenté au paragraphe 2.1 considèrerait implicitement une condition de Dirichlet homogène aux bords : $g \equiv 0$.

2.3.2 Condition de Neumann

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

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

ce qui fait que la première ligne de L devient

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

2.3.3 Condition de Robin

Enfin, pour une condition de Robin, $\partial_x u + \beta u = 0$, on écrit (par exemple, cela dépend en fait du signe de β , voir le traitement du centrage au paragraphe 2.4) $(u_1^{n+1} - u_0^{n+1})/\delta x + \beta u_0^{n+1} = 0$, et donc $u_0^{n+1} = u_1^{n+1}/(1 - \beta \delta x)$, la première ligne de L devenant

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

2.3.4 Application à l'équation de Black et Scholes sur grille irrégulière

À gauche, en \underline{X} , nous avons déjà vu que nous avons une condition de type Dirichlet, originellement en $-\infty$, qui est ramenée à distance finie. On a $u_0^n = K \exp(-t^n r)$ (dans le cas d'un taux d'intérêt sans risque constant) et

$$\frac{u_1^{n+1} - u_1^n}{\delta t} = \frac{2}{x_2 - x_0} \left(\frac{1}{x_2 - x_1} u_2^n - \left(\frac{1}{x_2 - x_1} + \frac{1}{x_1 - x_0} \right) u_1^n + \frac{1}{x_1 - x_0} K \exp(-t^n r) \right),$$

et la première ligne de la matrice L est inchangée, la condition de Dirichlet jouant un rôle au second membre.

Si on choisit d'utiliser aussi une condition de Dirichlet à droite, en \bar{X} , celle-ci est alors homogène et la dernière ligne correspond à

$$\frac{u_M^{n+1} - u_M^n}{\delta t} = \frac{2}{x_{M+1} - x_{M-1}} \left(- \left(\frac{1}{x_{M+1} - x_M} + \frac{1}{x_M - x_{M-1}} \right) u_M^n + \frac{1}{x_M - x_{M-1}} u_{M-1}^n \right),$$

Si, en revanche, on choisit une équation de Neumann en \bar{X} , on prend $u_{M+1}^n = u_M^n$ et

$$\frac{u_M^{n+1} - u_M^n}{\delta t} = \frac{2}{(x_{M+1} - x_{M-1})(x_M - x_{M-1})} (-u_M^n + u_{M-1}^n).$$

2.4 Traitement des dérivées d'ordre un

Pour le traitement des dérivées d'ordre un, on peut penser à plusieurs schémas, que l'on va analyser sur un cas simplifié à l'instar d'Achdou et Pironneau [1], à savoir l'équation en dimension 1, indépendante du temps

$$-a u''(x) + b u'(x) + c(x) u(x) = f(x), \text{ pour } x \in]0, 1[,$$

avec des conditions aux bords de Dirichlet homogène ($u(0) = u(1) = 0$) et où a et b sont strictement positifs et c est positive et continue. Trois choix sont naturels :

– le schéma centré

$$-a \frac{u_{i+1} - 2u_i + u_{i-1}}{\delta x^2} + b \frac{u_{i+1} - u_{i-1}}{2\delta x} + c(x_i)u_i = f(x_i),$$

– le schéma décentré amont

$$-a \frac{u_{i+1} - 2u_i + u_{i-1}}{\delta x^2} + b \frac{u_i - u_{i-1}}{\delta x} + c(x_i)u_i = f(x_i),$$

– le schéma décentré aval

$$-a \frac{u_{i+1} - 2u_i + u_{i-1}}{\delta x^2} + b \frac{u_{i+1} - u_i}{\delta x} + c(x_i)u_i = f(x_i),$$

Ces trois schémas sont consistants. Le schéma centré est d'ordre 2, les deux schémas décentrés sont d'ordre 1. Pour que l'on puisse résoudre ce système et obtenir une solution raisonnable, il faut que le système soit inversible et aussi qu'il soit stable. Pour avoir la stabilité, il faut que l'inverse de la matrice à inverser A soit de norme bornée. On s'intéresse à deux normes, les normes (subordonnées) 2 et ∞ .

On obtient les résultats suivants (voir [1] pour les détails). En norme 2, l'analyse est assez simple et on trouve que

- pour le schéma amont, $\|A^{-1}\|_2 \lesssim \frac{1}{\pi^2(a + \frac{\delta x b}{2})}$, la stabilité est donc inconditionnelle ;
- pour le schéma centré, $\|A^{-1}\|_2 \lesssim \frac{1}{\pi^2 a}$, la stabilité est donc inconditionnelle ;
- pour le schéma aval, $\|A^{-1}\|_2 \lesssim \frac{1}{\pi^2(a - \frac{\delta x b}{2})}$, il n'y a stabilité que si $\delta x < 2a/b$.

L'analyse en norme ∞ est beaucoup plus compliquée :

- le schéma amont est inconditionnellement stable ;
- le schéma centré est stable sous la condition $\delta x < 2a/b$;
- le schéma aval est stable sous la condition $\delta x < a/b$.

Selon que la valeur $2a/b$ constitue ou non une borne raisonnable, on sera donc amenés à choisir le schéma centré pour des raisons d'ordre ou le schéma amont pour des raisons de stabilité.

2.5 Application au modèle de Black et Scholes à un actif

Nous allons maintenant regrouper tous les éléments précédents dans le cas du modèle de Black et Scholes à un actif, avec paramètres dépendant du temps, mais sur une grille régulière en la variable x .

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

2.5.1 Écriture explicite du système linéaire

On discrétise l'intervalle de temps $[0, T]$ par des temps $t^n = n\delta t$ où $\delta t = T/N$. On discrétise également l'intervalle $[\underline{X}, \bar{X}]$ par des valeurs discrètes x_i que l'on suppose réparties régulièrement : $x_i = \underline{X} + i\delta x$. On cherche à approcher $w(t^n, x_i)$ par une valeur w_i^n .

On ne connaît pas exactement de le signe de $\sigma^2(t, e^x)/2 - r(t)$, qui d'ailleurs peut changer en fonction de t et x dans notre domaine de calcul. On prend donc un schéma centré en espace pour les termes du premier ordre, ce qui imposerait en gros $\delta x < 1$ si l'étude du modèle simplifié est encore valide ici (ce qui resterait à prouver).

$$\begin{aligned}
& \frac{w_i^{n+1} - w_i^n}{\delta t} - \theta \frac{1}{2} \sigma^2(t^{n+1}, e^{x_i}) \frac{w_{i+1}^{n+1} - 2w_i^{n+1} + w_{i-1}^{n+1}}{\delta x^2} \\
& - (1 - \theta) \frac{1}{2} \sigma^2(t^n, e^{x_i}) \frac{w_{i+1}^n - 2w_i^n + w_{i-1}^n}{\delta x^2} \\
& + \theta \left(\frac{1}{2} \sigma^2(t^{n+1}, e^{x_i}) - r(t^{n+1}) \right) \frac{w_{i+1}^{n+1} - w_{i-1}^{n+1}}{2\delta x} \\
& + (1 - \theta) \left(\frac{1}{2} \sigma^2(t^n, e^{x_i}) - r(t^n) \right) \frac{w_{i+1}^n - w_{i-1}^n}{2\delta x} \\
& + \theta r(t^{n+1}) w_i^{n+1} + (1 - \theta) r(t^n) w_i^n = 0.
\end{aligned}$$

Ceci s'écrit de manière synthétique pour la variable $W^n = {}^t(w_1^n, \dots, w_M^n)$

$$(I - \delta t \theta L^{n+1}) W^{n+1} = (I + \delta t(1 - \theta)L^n) W^n + (\theta S^{n+1} + (1 - \theta)S^n),$$

où L est tridiagonale avec

$$\begin{aligned}
L_{i,i-1}^n &= \frac{\sigma^2(t^n, e^{x_i})}{2\delta x^2} + \frac{\frac{1}{2}\sigma^2(t^n, e^{x_i}) - r(t^n)}{2\delta x}, \\
L_{ii}^n &= -\frac{\sigma^2(t^n, e^{x_i})}{\delta x^2} - r(t^n), \\
L_{i,i+1}^n &= \frac{\sigma^2(t^n, e^{x_i})}{2\delta x^2} - \frac{\frac{1}{2}\sigma^2(t^n, e^{x_i}) - r(t^n)}{2\delta x}.
\end{aligned}$$

On a vu que les conditions aux bords modifiaient la première et la dernière lignes des matrices L^n et définissaient les seconds membres S^n . On a toujours $S_i^n = 0$ pour $i = 2, \dots, M - 1$. À gauche, on fait le choix d'une condition de Dirichlet.

$$S_1^n = \delta t \left(\sigma^2(t^n, e^{x_1}) \left(\frac{1}{2\delta x^2} + \frac{1}{4\delta x} \right) - \left(\frac{1}{2\delta x} + 1 \right) r(t^n) \right) K \exp \left(- \int_0^{t^n} r(\tau) d\tau \right).$$

À droite, si on choisit des conditions de Dirichlet homogène, la dernière ligne des matrices L^n est inchangée et $S_M^n = 0$. En revanche, si on choisit des conditions aux limite de Neumann, $L_{M,M}^n$ est modifié en (somme des termes $L_{M,M}^n$ et " $L_{M,M+1}^n$ " précédents)

$$L_{M,M}^n = -\frac{\sigma^2(t^n, e^{x_M})}{2\delta x^2} - \frac{\frac{1}{2}\sigma^2(t^n, e^{x_M}) - r(t^n)}{2\delta x} - r(t^n),$$

et on a à nouveau $S_M^n = 0$. Dans tous les cas, la matrice à inverser est $(I - \delta t \theta L^{n+1})$ et est tridiagonale.

2.5.2 Résultats théoriques

Cette matrice est à diagonale strictement dominante ($r(t^n) > 0$), elle est donc inversible. On peut montrer (cf. [1]) que la solution de ce système est positive (c'est important pour notre application !) et que son maximum $\max_i(w_i^n)$ décroît avec n . On peut montrer également des estimations d'erreur sur E^n de coefficients $w(t^n, e^{x_i}) - w_i^n$: pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que si $\delta x < \eta$ et $\delta t < \eta$

$$\delta x^{1/2} \max_{0 \leq n \leq N} \|E^n\|_2 \leq \varepsilon.$$

Pour avoir une estimation meilleure, il faut supprimer la singularité de la fonction payoff en $s = K$, et on peut alors espérer (sous réserve que σ et r soient aussi régulières)

$$\delta x^{1/2} \max_{0 \leq n \leq N} \|E^n\|_2 \leq C(\delta t + \delta x).$$

Si on applique par exemple un schéma centré, on a une majoration sur δt pour assurer l'inversibilité de la matrice. Cette condition n'est pas très contraignante dans le cas de grandes volatilités. Si cette condition est vérifiée, et si la fonction payoff est régulière, on a

$$\delta x^{1/2} \max_{0 \leq n \leq N} \|E^n\|_2 \leq C(\delta t + \delta x^2).$$

La bonne solution consiste à centrer toutes les dérivées (Crank–Nicolson et un schéma centré pour le terme d'ordre 1). On obtient alors un schéma toujours inversible d'ordre 2 en temps et en espace.

2.5.3 Inversion pratique du système linéaire — algorithme LS

Étant donné le problème $Ax = b$ pour des vecteurs de taille M et une matrice tri-diagonale A , l'algorithme LS consiste à calculer des coefficients δ_i :

$$\delta_1 = 0 \text{ et } \delta_{i+1} = A_{i,i+1}/(A_{i,i} - A_{i,i-1}\delta_i) \text{ pour } i = 1, \dots, M-1,$$

puis à résoudre un système bi-diagonal inférieur

$$x'_1 = b_1/A_{1,1} \text{ et } x'_i = (b_i - A_{i,i-1}x'_{i-1})/(A_{i,i} - A_{i,i-1}\delta_i) \text{ pour } i = 2, \dots, M,$$

et enfin un système bi-diagonal supérieur

$$x_M = x'_M \text{ et } x_i = x'_i - \delta_i x_{i+1} \text{ pour } i = M-1, \dots, 1.$$

2.5.4 Stockage

Pour effectuer le calcul précédent, il est clair qu'il est inutile de stocker la totalité de la matrice L^{n+1} dont seules trois diagonales sont non nulles. Il faut donc au pire stocker uniquement ces trois diagonales.

Dans le cas de maillages non uniformes, il faut peut-être les stocker effectivement, mais en faisant attention de n'évaluer qu'une seule fois calculant les $\sigma^2(t^n, e^{x_i})$ ou les $t(t^n)$. Le cas de stockage minimal est obtenu dans le cas où le maillage est uniforme et où σ et r ne dépendent pas de t^n et x_i . Dans cas cas là, on ne stocke qu'une poignée de scalaires.

Le calcul de S_1^n doit faire l'objet d'une accumulation pour ce qui est du calcul de l'exponentielle. En effet, on peut noter $e^n = K \exp(-\int_0^{t^n} r(\tau) d\tau)$, qui vérifie la relation de récurrence exacte

$$e^{n+1} = \exp\left(-\int_{t^n}^{t^{n+1}} r(\tau) d\tau\right) e^n,$$

qu'il suffit d'approcher par une condition à l'ordre 1 ou 2 (en fonction de l'ordre des autres méthodes utilisées). Par exemple, la méthode du point milieu :

$$\exp\left(-\int_{t^n}^{t^{n+1}} r(\tau) d\tau\right) \simeq \exp\left(-(t^{n+1} - t^n)r\left(\frac{t^n + t^{n+1}}{2}\right)\right).$$

3 Modèle de Black et Scholes à plusieurs actifs

3.1 Grandes dimensions – stockage

Soit l'équation de la chaleur en dimension 2

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

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 $\partial_{xx}^2 u + \partial_{yy}^2 u$ 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 N , on peut écrire une discrétisation du laplacien en utilisant $1 + 2N$ points.

Faisons une rapide estimation de la dimension des données. Imaginons que l'on considère une équation de Black et Scholes à $N = 4$ 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 $100^N = 10^8$ 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 $1+2N$ valeurs non nulles sur chaque ligne, mais cela fait tout de même la bagatelle $9 \cdot 10^8$ 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.

3.2 L'idée du splitting

Une idée consiste à dire que résoudre $\partial_t u = \partial_{xx}^2 u + \partial_{yy}^2 u$ sur un intervalle de temps $[0, T]$, c'est "presque" comme résoudre $\partial_t u = \partial_{xx}^2 u$ sur cet intervalle de temps, puis prendre le résultat comme nouvelle donnée initiale pour le problème $\partial_t u = \partial_{yy}^2 u$ 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 N , on se ramène à résoudre en N 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^N variables de l'exemple précédent, se trouvent groupées à chaque étape en 100^{N-1} paquets indépendants de 100 variables, permettant une parallélisation naturelle des algorithmes.

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.

3.3 Dérivées mixtes

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 ∂_{xy}^2 . 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.

4 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.

4.1 Une équation scalaire

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

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

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

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

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

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

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.

4.2 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

- les systèmes différentiels linéaires : $\dot{x} = (A + B)x$ (dès que les matrices A et B ne commutent pas),
- les systèmes différentiels linéaires avec deux échelles différentes : $\dot{x} = (\frac{1}{\varepsilon}A + B)x$,
- les équations aux dérivées partielles non linéaires : par exemple $\partial_t u = \Delta u + f(u)$, voir par exemple [33],
- 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, voir par exemple [22].
- 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.

4.3 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

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

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

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

Avec cette notation, on définit sans effort (et sans introduction de variables supplémentaires) deux nouveaux types de splitting : les **splittings de Strang** [32]

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

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 :

$$(P1) \ \mathcal{A}(0) = I, \quad (\text{cf. } \exp(0) = 1),$$

$$(P2) \ \mathcal{A}(t + s) = \mathcal{A}(t)\mathcal{A}(s), \text{ pour tout } t, s \geq 0 \quad (\text{cf. } \exp(a(t + s)) = \exp(at)\exp(as)).$$

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 [27].

5 Le cas des systèmes linéaires

On considère maintenant le système linéaire

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

où $x \in \mathbb{R}^N$ et A et B sont des matrices de \mathcal{M}_N . 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

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

5.1 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.

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

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

5.2 Splitting de Strang

Étudions de même la formule de Strang

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

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

5.3 Splittings d'ordre plus élevé

5.3.1 Calcul de l'ordre et convergence

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

Théorème 1

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

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

soit vrai dans un voisinage de 0. Alors

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

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

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

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

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

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 ?

5.3.2 Une réponse négative

La première réponse est négative dans le cas où on cherche des coefficients positifs. Ce résultat est dû à Michelle Schatzman [29]. 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 $p \geq 3$.

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 $p \geq 3$.

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.

5.3.3 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

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

pour laquelle

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

5.3.4 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 méthode 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

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

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

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

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

$$g(t) = \frac{4}{3} \exp(\frac{1}{4}tA) \exp(\frac{1}{2}tB) \exp(\frac{1}{2}tA) \exp(\frac{1}{2}tB) \exp(\frac{1}{4}tA) - \frac{1}{3} \exp(\frac{1}{2}tA) \exp(tB) \exp(\frac{1}{2}tA).$$

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.

6 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** :

$$D = \text{diag}(d_k). \text{ On a alors } \exp(tD) = \text{diag}(\exp(td_k)).$$

– 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 :

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

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

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

(N.B. Ici et dans l'équation suivante t^k désigne t à la puissance k et pas un point de discrétisation du temps.)

– les **matrices idempotentes** comme les **projecteurs** : $P^n = P$ pour tout $n \geq 1$. 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} .

$$\dot{x} = (A + B)x, \quad x(t^n) = x^n,$$

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 $t = t^n$:

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

On a donc

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

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

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

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 .

6.1 Schéma d'Euler

Pour l'équation (S) de départ, le schéma d'Euler consiste à approcher l'intégrale par la méthode des rectangles à gauche et s'écrit

$$x^{n+1} = \mathcal{E}(A + B, h)x^n = x^n + h(A + B)x^n = (I + h(A + B))x^n.$$

Si on applique le schéma d'Euler à (L1), on a

$$\begin{aligned} \mathcal{E}(B, h)x^n &= x^n + hBx^n = (I + hB)x^n, \\ x^{n+1} &= \mathcal{E}(A, h)\mathcal{E}(B, h)x^n = \mathcal{E}(B, h)x^n + hA\mathcal{E}(B, h)x^n = (I + hA)\mathcal{E}(B, h)x^n \\ &= (I + hA)(I + hB)x^n = (I + h(A + B) + h^2AB)x^n. \end{aligned}$$

On a

$$\mathcal{L}_1\mathcal{E}(h) = I + h(A + B) + h^2AB.$$

Du point de vue de l'ordre, on a

$$\mathcal{L}_1\mathcal{E}(h) - \mathcal{S}(h) = \frac{1}{2}h^2([A, B] - (A^2 + B^2)) + O(h^3).$$

Le schéma de splitting de Lie avec approximation d'Euler est d'ordre 1. On peut dissocier l'erreur d'approximation de l'exponentielle de celle du splitting :

$$\mathcal{L}_1\mathcal{E}(h) - \mathcal{L}_1(h) = -\frac{1}{2}h^2(A^2 + B^2) + O(h^3).$$

Cette erreur d'approximation est symétrique en A et B . Elle ne dépend donc pas du choix de méthode de Lie. On est bien sûr encore d'ordre 1.

6.2 Schéma d'Euler rétrograde

Pour l'équation (S) de départ, le schéma d'Euler rétrograde consiste à approcher l'intégrale par la méthode des rectangles à droite et s'écrit

$$\begin{aligned} x^{n+1} &= \mathcal{R}(A + B, h)x^n = x^n + h(A + B)\mathcal{R}(A + B, h)x^n, \\ \Rightarrow x^{n+1} &= (I - h(A + B))^{-1}x^n. \end{aligned}$$

Il suffit de prendre h suffisamment petit pour que ceci ait un sens : $h \leq 1/\rho(A + B)$.

On peut effectuer un développement limité et

$$\mathcal{R}(A + B, h) = I + h(A + B) + h^2(A + B)^2 + O(h^3).$$

Si on applique le schéma d'Euler rétrograde à (L1), on a

$$\begin{aligned} \mathcal{R}(B, h)x^n &= (I - hB)^{-1}x^n, \\ x^{n+1} &= \mathcal{R}(A, h)\mathcal{R}(B, h)x^n = (I - hA)^{-1}(I - hB)^{-1}x^n. \end{aligned}$$

On obtient l'opérateur

$$\mathcal{L}_1\mathcal{R}(h) = (I - hA)^{-1}(I - hB)^{-1} = I + h(A + B) + h^2(A^2 + AB + B^2) + O(h^3).$$

Du point de vue de l'ordre, on a

$$\mathcal{L}_1\mathcal{R}(h) - \mathcal{S}(h) = \frac{1}{2}h^2([A, B] + (A^2 + B^2)) + O(h^3).$$

Le schémas de splitting de Lie avec approximation d'Euler rétrograde sont d'ordre 1. À nouveau, on peut extraire uniquement l'erreur d'approximation

$$\mathcal{L}_1\mathcal{R}(h) - \mathcal{L}_1(h) = \frac{1}{2}h^2(A^2 + B^2) + O(h^3).$$

qui est à nouveau symétrique en A et B et l'ordre est toujours 1.

6.3 Couplage Euler–Euler rétrograde

Une fois le splitting effectué, rien n'oblige à utiliser le même schéma pour les deux parties. Par exemple avec (L1), on peut écrire :

$$\begin{aligned}\mathcal{L}_1\mathcal{R}\mathcal{E}(h) &= (I - hA)^{-1}(I + hB) = I + h(A + B) + h^2(A^2 + AB) + O(h^3), \\ \mathcal{L}_1\mathcal{E}\mathcal{R}(h) &= (I + hA)(I + hB)^{-1} = I + h(A + B) + h^2(AB + B^2) + O(h^3).\end{aligned}$$

6.3.1 Méthode de Crank–Nicolson

La **méthode de Crank–Nicolson** peut être décrite ainsi. Soit à résoudre $\dot{x} = Ax$, que l'on écrit

$$\dot{x} = \frac{1}{2}Ax + \frac{1}{2}Ax.$$

On utilise une méthode d'Euler pour une partie et d'Euler rétrograde pour l'autre. On est clairement dans le cadre commutatif et

$$\mathcal{C}(h) = (I + \frac{1}{2}hA)(I - \frac{1}{2}hA)^{-1} = (I - \frac{1}{2}hA)^{-1}(I + \frac{1}{2}hA).$$

Ceci est équivalent à l'écriture habituelle de cette méthode qui correspond à utiliser la méthode du trapèze pour approcher la formulation intégrale :

$$\frac{x^{n+1} - x^n}{h} = A \frac{x^{n+1} + x^n}{2}.$$

Cette méthode est bien connue pour être d'ordre 2. Pour le démontrer, il faut pousser un peu plus loin le développement

$$\mathcal{C}(h) = (I + \frac{1}{2}hA)(I + \frac{1}{2}hA + \frac{1}{4}h^2A^2 + \frac{1}{8}h^3A^3 + O(h^4)) = I + hA + \frac{1}{2}h^2A^2 + \frac{1}{4}h^3A^3 + O(h^4).$$

On a

$$\mathcal{C}(h) - \mathcal{A}(h) = \frac{1}{12}h^3A^3 + O(h^4).$$

La méthode est exactement d'ordre 2.

6.3.2 Le θ -schéma

On peut de manière plus générale utiliser la décomposition

$$\dot{x} = \theta Ax + (1 - \theta)Ax$$

en traitant $\dot{x} = \theta Ax$ par un schéma d'Euler rétrograde et $\dot{x} = (1 - \theta)Ax$ par un schéma d'Euler. On obtient le θ -schéma :

$$x^{n+1} = (I - h\theta A)^{-1}(I + h(1 - \theta)A)x^n.$$

On a l'habitude de définir ce schéma sous la forme

$$x^{n+1} = x^n + h((1 - \theta)Ax^n + \theta Ax^{n+1}),$$

c'est-à-dire que l'on approche la formulation intégrale par un barycentre de x^n et x^{n+1} avec les poids $1 - \theta$ et θ . Cette interprétation n'a de sens que si les poids sont positifs et donc $\theta \in [0, 1]$. Sinon, on peut montrer que la méthode n'est pas consistante.

On reconnaît bien sûr, le schéma d'Euler pour $\theta = 0$, le schéma d'Euler rétrograde pour $\theta = 1$ et le schéma de Crank–Nicolson pour $\theta = 1/2$. Seul ce dernier cas lie à un ordre 2, le θ -schéma étant génériquement d'ordre 1.

Ce schéma est très largement utilisé en mathématiques financières.

6.4 Vers l'ordre 2 pour un splitting approché

Pour passer à l'ordre 2, on peut envisager différentes solutions :

- monter à l'ordre 2 sur le schéma de splitting,
- monter à l'ordre 2 sur le schéma d'approximation de l'exponentielle,
- monter à l'ordre 2 sur le schéma de splitting et le schéma d'approximation de l'exponentielle.

On a vu que l'on pouvait dissocier erreur d'approximation de l'exponentielle et de splitting. Il ne faut donc pas espérer de miracle : appliquer un schéma d'Euler (ordre 1) sur une formule de Strang (ordre 2) ou un schéma de Crank–Nicolson (ordre 2) au splitting de Lie (ordre 1), donne un ordre total de 1. Les calculs sont laissés au lecteur.

6.4.1 Combinaison de Strang et Crank–Nicolson

Une façon d'obtenir l'ordre 2 est bien sûr de combiner une méthode de Strang avec une méthode de Crank–Nicolson.

$$\begin{aligned} \mathcal{S}_1 \mathcal{C}(h) &= (I + \frac{h}{4}A)(I - \frac{h}{4}A)^{-1}(I + \frac{h}{2}B)(I - \frac{h}{2}B)^{-1}(I + \frac{h}{4}A)(I - \frac{h}{4}A)^{-1} \\ &= (I + \frac{1}{2}hA + \frac{1}{8}h^2A^2 + \frac{1}{32}h^3A^3 + O(h^4))(I + hB + \frac{1}{2}h^2B^2 + \frac{1}{4}h^3B^3 + O(h^4)) \times \\ &\quad \times (I + \frac{1}{2}hA + \frac{1}{8}h^2A^2 + \frac{1}{32}h^3A^3 + O(h^4)) \\ &= I + h(A + B) + \frac{1}{2}h^2(A^2 + AB + BA + B^2) \\ &\quad + h^3(\frac{3}{16}A^3 + \frac{1}{8}(A^2B + BA^2) + \frac{1}{4}(ABA + AB^2 + B^2A + B^3)) + O(h^4). \end{aligned}$$

L'erreur locale est donnée par

$$\begin{aligned}
S_1\mathcal{C}(h) - S(h) &= h^3\left(\frac{3}{16}A^3 + \frac{1}{8}(A^2B + BA^2) + \frac{1}{4}(ABA + AB^2 + B^2A + B^3)\right) \\
&\quad - \frac{1}{6}h^3(A^3 + A^2B + ABA + BA^2 + BAB + AB^2 + B^2A + B^3) + O(h^4) \\
&= \frac{1}{48}h^3(A^3 - 2(A^2B + BA^2) + 4(ABA + AB^2 + B^2A + B^3) - 8BAB) \\
&\quad + O(h^4).
\end{aligned}$$

La méthode n'est pas plus précise que l'ordre 2 et elle met en œuvre la construction de quatre opérateurs dont trois à inverser, il faut donc lui préférer la méthode de Peaceman–Rachford.

6.4.2 Méthode de Peaceman–Rachford

On peut combiner les opérateurs qui constituent les méthodes d'ordre 1 en

$$\mathcal{P}(h) = (I - \frac{h}{2}A)^{-1}(I + \frac{h}{2}B)(I - \frac{h}{2}B)^{-1}(I + \frac{h}{2}A).$$

Ceci s'appelle la **méthode de Peaceman–Rachford** [28]. On peut interpréter ceci par un demi-pas de temps d'Euler pour A , suivi d'un pas de temps de Crank–Nicolson pour B et enfin un demi-pas de temps d'Euler rétrograde pour A . De nombreuses autres interprétations existent, voir [2]. On calcule

$$\begin{aligned}
\mathcal{P}(h) &= (I + \frac{1}{2}hA + \frac{1}{4}h^2A^2 + \frac{1}{8}h^3A^3 + O(h^4))(I + hB + \frac{1}{2}h^2B^2 + \frac{1}{4}h^3B^3 + O(h^4))(I + \frac{1}{2}hA) \\
&= I + h(A + B) + \frac{1}{2}h^2(A^2 + AB + BA + B^2) \\
&\quad + \frac{1}{4}h^3(A^3 + A^2B + ABA + AB^2 + B^2A + B^3) + O(h^4).
\end{aligned}$$

L'erreur locale est donnée par

$$\begin{aligned}
\mathcal{P}(h) - S(h) &= \frac{1}{4}h^3(A^3 + A^2B + ABA + AB^2 + B^2A + B^3) \\
&\quad - \frac{1}{6}h^3(A^3 + A^2B + ABA + BA^2 + BAB + AB^2 + B^2A + B^3) + O(h^4) \\
&= \frac{1}{12}h^3(A^3 + A^2B + ABA + AB^2 + B^2A + B^3) - \frac{1}{6}h^3(BA^2 + BAB) + O(h^4).
\end{aligned}$$

On obtient une méthode exactement d'ordre 2. Elle met en œuvre la construction de quatre opérateurs dont deux seulement sont à inverser.

On reconnaîtra cette méthode dans la suite sous la forme

$$(I - \frac{h}{2}B)(I - \frac{h}{2}A)x^{n+1} = (I + \frac{h}{2}B)(I + \frac{h}{2}A)x^n.$$

6.5 Et si A et B commutent ?

Nous avons vu que toutes les approximations à base d'exponentielles sont exactes si A et B commutent. Est-ce encore vrai lorsque l'on approche les exponentielles ? Sinon, peut-on espérer au moins un gain en ordre ? Dans tous les cas, l'erreur locale sera la même que l'on compare avec $S(h)$ ou avec la méthode de Lie, de Strang, ... puisqu'elles sont équivalentes. Reprenons toutes les méthodes déjà vues une-à-une.

6.5.1 Schéma d'Euler-Lie

On calcule l'erreur locale

$$\mathcal{L}_1 \mathcal{E}(h) - \mathcal{S}(h) = \mathcal{L}_2 \mathcal{E}(h) - \mathcal{S}(h) = -\frac{1}{2}h^2(A^2 + B^2) + O(h^3).$$

Les deux méthodes $\mathcal{L}_1 \mathcal{E}(h)$ et $\mathcal{L}_2 \mathcal{E}(h)$ sont donc clairement identiques. Aucun gain en ordre n'est observé.

6.5.2 Schéma d'Euler rétrograde-Lie

On calcule l'erreur locale

$$\mathcal{L}_1 \mathcal{R}(h) - \mathcal{S}(h) = \mathcal{L}_2 \mathcal{R}(h) - \mathcal{S}(h) = \frac{1}{2}h^2(A^2 + B^2) + O(h^3).$$

La conclusion est la même que pour Euler.

6.5.3 Schéma de Peaceman-Rachford

La formule d'erreur locale pour la méthode de Peaceman-Rachford est nettement simplifiée si A et B commutent. En effet $A^2B + ABA - 2BA^2 = 0$ et $AB^2 + B^2A - BAB = 0$ donc

$$\mathcal{P}(h) - \mathcal{S}(h) = \frac{1}{12}h^3(A^3 + B^3) + O(h^4).$$

L'ordre n'est cependant pas amélioré.

7 Généralités sur l'équation de la chaleur

La méthode des directions alternées (ADI, Alternate Directions Implicit, en anglais) prend son origine dans l'application initiale par Peaceman et Rachford [28] et par Douglas [8] de la méthode : la résolution de l'équation de la chaleur en décomposant le laplacien en sommes.

Nous allons d'abord donner un cadre favorable dans lequel la formule de Trotter-Kato est valide. Dans ce contexte, tout marche comme dans le cas matriciel.

Dans le cadre continu, où on traite de laplaciens, un problème crucial est celui de la stabilité pour laquelle nous donnerons des définitions et que nous illustrerons par un exemple. Dans les cas pratiques nous allons regarder des approximations du laplacien. Nous serons donc dans le cadre d'espace fonctionnels finis et de matrices. Ces problèmes de stabilité ne nous affecteront pas directement, mais il y en aura d'autres. . .

7.1 Formule de Trotter-Kato

Lemme 1 (Trotter-Kato)

Soient $C \geq 1$ et $\omega > 0$. On suppose que l'on a une suite de générateurs infinitésimaux A_n de semi-groupes uniformément continus $\mathcal{A}_n(t)$ qui satisfont $\|\mathcal{A}_n(t)\| \leq C \exp(\omega t)$. On suppose qu'il existe $\lambda_0 \in \mathbb{C}$ tel que $\text{Re}(\lambda_0) > \omega$ et pour tout $x \in X$, $(\lambda_0 I - A_n)^{-1}x \rightarrow R(\lambda_0)x$, où l'image de $R(\lambda_0)$ est dense dans X .

Alors il existe un générateur de groupe infinitésimal A d'un semi-groupe uniformément continu $\mathcal{A}(t)$ qui satisfait $\|\mathcal{A}(t)\| \leq C \exp(\omega t)$ tel que $R(\lambda_0) = (\lambda_0 I - A)^{-1}$ et $\mathcal{A}_n(t)x \rightarrow \mathcal{A}(t)x$ pour tout $x \in X$. Cette limite est uniforme sur tout intervalle borné en temps.

L'application à notre cadre donne typiquement le résultat :

$$\lim_{n \rightarrow \infty} \left(\mathcal{A}\left(\frac{t}{n}\right) \mathcal{B}\left(\frac{t}{n}\right) \right)^n \equiv \lim_{n \rightarrow \infty} \left(\exp\left(\frac{t}{n}A\right) \exp\left(\frac{t}{n}B\right) \right)^n = \exp(t(A+B)) \equiv \mathcal{S}(t).$$

7.2 Notions de stabilité

On considère l'EDP suivante

$$\frac{du}{dt} = Au + Bu, \quad u(t=0) = u^0 \in H,$$

Pour la résoudre on utilise un schéma donné par son semi-groupe $f(h)$. Dans la démonstration de convergence que nous avons déjà donnée dans le cadre matriciel, on avait vu qu'il fallait de la stabilité en plus de la consistance (ordre local au moins égal à 2).

Classiquement, on dira que le schéma est **fortement stable** si il existe une constante C tel que pour h petit,

$$\|f(h)\|_{\mathcal{B}(H)} \leq 1 + Ch,$$

où $\|\cdot\|_{\mathcal{B}(H)}$ est la norme des opérateurs linéaires bornés sur l'espace de Hilbert H :

$$\|A\|_{\mathcal{B}(H)} = \sup_{u \neq 0} \frac{\|Au\|_H}{\|u\|_H}.$$

Pour l'approximation de Peaceman–Rachford :

$$\mathcal{P}(h) = \left(I - \frac{1}{2}hA\right)^{-1} \left(I + \frac{1}{2}hB\right) \left(I - \frac{1}{2}hB\right)^{-1} \left(I + \frac{1}{2}hA\right),$$

on peut montrer que

- cette approximation est **faiblement stable** au sens où pour tout opérateurs A et B négatifs auto-adjoints et $u \in D(A)$, on a

$$\|\mathcal{P}\left(\frac{t}{n}\right)^n u\|_H \leq \|u\|_H + \tau \|Au\|_H,$$

pour tout n tel que $t/n \leq \tau$ (voir Appendice A.3). Il y a en quelque sens une perte de régularité.

- on peut donner un contre-exemple à la stabilité à base de laplaciens, à savoir des opérateurs A et B pour lesquels

$$\|\mathcal{P}(h)\|_{\mathcal{B}(H)} \geq 2.$$

Le résultat général dont la démonstration dépasse de loin le cadre de ce cours est que, sous certaines hypothèses supplémentaires (voir l'article de Schatzman [30]), on a uniquement

$$\|\mathcal{P}(h)\|_{\mathcal{B}(H)} \leq 1 + C\sqrt{h}.$$

7.3 La méthode des directions alternées

7.3.1 La méthode ADI historique

La méthode ADI historique a été introduite par Douglas [8], Douglas et Rachford [12] et Peaceman et Rachford [28] pour l'équation de la chaleur en dimension 2

$$\partial_t u(t, x, y) = \partial_{xx}^2 u(t, x, y) + \partial_{yy}^2 u(t, x, y)$$

sous la forme :

$$\begin{aligned} \frac{u^{n+1/2} - u^n}{\frac{1}{2}\delta t} &= L_x u^{n+1/2} + L_y u^n, \\ \frac{u^{n+1} - u^{n+1/2}}{\frac{1}{2}\delta t} &= L_x u^{n+1/2} + L_y u^{n+1}, \end{aligned}$$

où maintenant L_x et L_y désignent des matrices de taille $M^2 \times M^2$. Cette méthode consiste à faire deux étapes de Euler–Euler rétrograde sur l’opérateur $(\partial_{xx}^2 + \partial_{yy}^2)/2$. On reconnaît la méthode de Peaceman–Rachford pour $A = L_y$ et $B = L_x$:

$$\begin{aligned} (I - \frac{1}{2}\delta t L_x)u^{n+1/2} &= (I + \frac{1}{2}\delta t L_y)u^n, \\ (I - \frac{1}{2}\delta t L_y)u^{n+1} &= (I + \frac{1}{2}\delta t L_x)u^{n+1/2}, \\ (I - \frac{1}{2}\delta t L_x)(I - \frac{1}{2}\delta t L_y)u^{n+1} &= (I + \frac{1}{2}\delta t L_x)(I + \frac{1}{2}\delta t L_y)u^n. \end{aligned}$$

C’est aussi une généralisation de la méthode de Crank–Nicolson, car prendre $L_y \equiv 0$ redonne la méthode de Crank–Nicolson.

On peut également voir cette méthode comme une méthode de prédiction–correction, pour cela on l’écrit

$$\begin{aligned} \frac{u^{n+1,*} - u^n}{\delta t} &= \frac{1}{2}L_x(u^{n+1,*} + u^n) + L_y u^n, \\ \frac{u^{n+1} - u^n}{\delta t} &= \frac{1}{2}L_x(u^{n+1,*} + u^n) + \frac{1}{2}L_y(u^{n+1} + u^n), \end{aligned}$$

ou encore

$$\begin{aligned} (I - \frac{1}{2}\delta t L_x)u^{n+1,*} &= (I + \frac{1}{2}\delta t L_x + \delta t L_y)u^n, \\ (I - \frac{1}{2}\delta t L_y)u^{n+1} &= u^{n+1,*} - \frac{1}{2}\delta t L_y u^n. \end{aligned}$$

Sous cette forme, on voit que la nouvelle valeur intermédiaire $u^{n+1,*}$ est une approximation par splitting de l’équation complète, où on traite avec Euler rétrograde $\partial_{xx}^2/2$ et avec Euler $\partial_{xx}^2/2 + \partial_{yy}^2$. Le deuxième pas est une correction en implicitant aussi la dérivée $\partial_{yy}^2/2$. À noter que l’on n’avance pas en temps dans cette dernière étape.

C’est sous la forme de méthode de prédiction–correction que la méthode ADI a été généralisée au cas de trois directions d’espace par Douglas [10] en ajoutant une autre variable intermédiaire

$$\begin{aligned} \frac{u^{n+1,*} - u^n}{\delta t} &= \frac{1}{2}L_x(u^{n+1,*} + u^n) + L_y u^n + L_z u^n, \\ \frac{u^{n+1,**} - u^n}{\delta t} &= \frac{1}{2}L_x(u^{n+1,*} + u^n) + \frac{1}{2}L_y(u^{n+1,**} + u^n) + L_z u^n, \\ \frac{u^{n+1} - u^n}{\delta t} &= \frac{1}{2}L_x(u^{n+1,*} + u^n) + \frac{1}{2}L_y(u^{n+1,**} + u^n) + \frac{1}{2}L_z(u^{n+1} + u^n), \end{aligned}$$

qui s’écrit aussi :

$$\begin{aligned} (I - \frac{1}{2}\delta t L_x)u^{n+1,*} &= (I + \frac{1}{2}\delta t L_x + \delta t L_y + \delta t L_z)u^n, \\ (I - \frac{1}{2}\delta t L_y)u^{n+1,**} &= u^{n+1,*} - \frac{1}{2}\delta t L_y u^n, \\ (I - \frac{1}{2}\delta t L_z)u^{n+1} &= u^{n+1,**} - \frac{1}{2}\delta t L_z u^n. \end{aligned}$$

À nouveau toutes les variables $u^{n+1,*}$, $u^{n+1,**}$ et u^{n+1} sont moralement au temps t^{n+1} . Cette forme permet clairement de mettre en évidence les calculs à effectuer en pratique : il faut inverser les matrices $(I - \frac{1}{2}\delta t L_x)$, $(I - \frac{1}{2}\delta t L_y)$ et $(I - \frac{1}{2}\delta t L_z)$ qui sont toutes tridiagonales à une permutation près.

On montre qu’en fait on calcule :

$$(I - \frac{1}{2}\delta t L_x)(I - \frac{1}{2}\delta t L_y)(I - \frac{1}{2}\delta t L_z)u^{n+1} = (I + \frac{1}{2}\delta t L_x)(I + \frac{1}{2}\delta t L_y)(I + \frac{1}{2}\delta t L_z)u^n.$$

Nous restons en terrain connu.

7.3.2 Cas du θ -schéma

Au lieu de couper les opérateurs en deux moitiés égales, on utilise en pratique des θ -schémas pour des questions de stabilité en grande dimensions (cf. infra). Cela revient à calculer, par exemple en dimension 3,

$$\begin{aligned}(I - \theta\delta t L_x)u^{n+1,*} &= (I + (1 - \theta)\delta t L_x + \delta t L_y + \delta t L_z)u^n, \\ (I - \theta\delta t L_y)u^{n+1,**} &= u^{n+1,*} - \theta\delta t L_y u^n, \\ (I - \theta\delta t L_z)u^{n+1} &= u^{n+1,**} - \theta\delta t L_z u^n.\end{aligned}$$

Cette généralisation ne s'écrit pas comme splitting de θ -schémas. En effet, si on développe on obtient

$$\begin{aligned}(I - \theta\delta t L_x)(I - \theta\delta t L_y)(I - \theta\delta t L_z)u^{n+1} &= u^n + (1 - \theta)\delta t(L_x + L_y + L_z)u^n \\ &\quad + \theta^2\delta t^2(L_x L_y + L_x L_z + L_y L_z)u^n - \theta^3\delta t^3 L_x L_y L_z u^n.\end{aligned}$$

Ceci étant, on n'espère pas plus que l'ordre 1 avec $\theta \neq 1/2$ et le second membre est aussi une approximation à l'ordre 1 de $(I + (1 - \theta)\delta t L_x)(I + (1 - \theta)\delta t L_y)(I + (1 - \theta)\delta t L_z)u^n$.

On peut aussi écrire ce schéma sous la forme

$$P_\theta u^{n+1} = (P_\theta + \delta t L)u^n, \quad \text{avec } P_\theta = (I - \theta\delta t L_x)(I - \theta\delta t L_y)(I - \theta\delta t L_z) \text{ et } L = L_x + L_y + L_z.$$

Cette forme sera utilisée plus tard.

7.3.3 Extensions

Une modification de cette méthode est proposée dans [5]. Pour pouvoir appliquer ceci en mathématiques financières, il va falloir être capables d'étendre ce modèle pour pouvoir traiter

- d'autres opérateurs différentiels (du premier ordre notamment);
- des coefficients variables en temps en en « espace »;
- des dérivées secondes croisées (voir [25, 26]);
- des conditions aux bord.

8 Application aux modèles en mathématiques financières

8.1 Traitement des termes mixtes

Pour pouvoir traiter aussi bien le modèle de Black et Scholes à plusieurs actifs que le modèle de Heston, il faut savoir traiter des dérivées secondes croisées (ou mixtes). Comment alterner des directions quand celles-ci sont combinées? Nous allons voir que l'on contourne en fait le problème.

Soit l'équation aux dérivées partielles

$$\partial_t u = \sum_{i,j=1}^N q_{ij} \partial_{x_i x_j}^2 u,$$

où $q = (q_{ij})_{ij}$ est une matrice définie positive, pour assurer le caractère parabolique et bien posé. En prenant un pas d'espace uniforme, on définit les opérateurs discrets (en dimension 2, pour simplifier les écritures)

$$\begin{aligned}(\delta_{xx}^2 u)_{i,j} &= \frac{1}{\delta x^2} (u_{i+1,j} - 2u_{i,j} + u_{i-1,j}), \\ (\delta_{xy}^2 u)_{i,j} &= \frac{1}{4\delta x^2} (u_{i+1,j+1} - u_{i+1,j-1} - u_{i-1,j+1} + u_{i-1,j-1}).\end{aligned}$$

Tous les schémas sont évidemment construits pour avoir au moins l'ordre 1 et donc la consistance. En revanche, pour chaque schéma, il faut se poser la question cruciale de la stabilité. Comme

l'équation est linéaire et à coefficients constants (pour l'instant) il suffit de faire une analyse en fréquence. Pour cela, on suppose que u^n est de la forme

$$u^n(x_1, \dots, x_N) = \xi^n \exp(i(\omega_1 x_1 + \dots + \omega_N x_N)).$$

En pratique, toute solution s'écrit comme une superposition de telles fonctions. On appelle ξ le facteur d'amplification. La stabilité consiste à montrer que pour toute fréquence $\omega = (\omega_1, \dots, \omega_N)$, l'amplitude reste bornée en temps, ce qui suppose que $|\xi| \leq 1$.

8.1.1 La méthode de Douglas

En généralisant en un θ -schéma l'algorithme de Douglas et Gunn [11] (écrit pour $\theta = 1$), on peut écrire

$$\begin{aligned} (I - \theta q_{11} \delta t \delta_{x_1 x_1}^2) u^{n+1(1)} &= \left(I + \delta t (1 - \theta) q_{11} \delta_{x_1 x_1}^2 + \delta t \sum_{i=2}^N q_{ii} \delta_{x_i x_i}^2 + 2\delta t \sum_{i=2}^N \sum_{j=1}^{i-1} q_{ij} \delta_{x_i x_j}^2 \right) u^n, \\ (I - \theta q_{jj} \delta t \delta_{x_j x_j}^2) u^{n+1(j)} &= u^{n+1(j-1)} - \theta q_{jj} \delta t \delta_{x_j x_j}^2 u^n, \quad 2 \leq j \leq N, \\ u^{n+1} &= u^{n+1(N)}. \end{aligned}$$

Le premier pas peut se réécrire

$$(I - \theta q_{11} \delta t \delta_{x_1 x_1}^2) u^{n+1(1)} = (I - \theta q_{11} \delta t \delta_{x_1 x_1}^2) u^n + \delta t L u^n \quad \text{avec } L = \sum_{i,j=1}^N q_{ij} \delta_{x_i x_j}^2,$$

c'est-à-dire que moralement on se donne $u^{n+1(0)} = u^n$ pour initier la récurrence. Toute la partie mixte de l'équation est traitée à ce pas et uniquement à ce pas, de manière explicite.

On peut écrire ce schéma sous la forme compacte

$$P_\theta u^{n+1} = (P_\theta + \delta t L) u^n \quad \text{avec } P_\theta = \prod_{i=1}^N (I - \theta q_{ii} \delta t \delta_{x_i x_i}^2),$$

ou encore

$$(DO) \quad u^{n+1} = P_\theta^{-1} (P_\theta + \delta t L) u^n = (I + \delta t P_\theta^{-1} L) u^n.$$

Étant donné de la complexité du problème, on ne peut espérer une condition nécessaire et suffisante de stabilité pour des q_{ij} donnés. On est capable (voir [6]) de trouver une condition suffisante de stabilité :

$$\theta \geq \bar{\theta} = \begin{cases} \frac{1}{2} & \text{si } N = 2, \\ \frac{N^2(\sqrt{1+(N-2)/N-1})}{2(N-2)} & \text{si } N \geq 3, \end{cases}$$

et une condition nécessaire et suffisante dans le pire cas :

$$\theta \geq \underline{\theta} = \frac{1}{2} \left(\frac{N-1}{N} \right)^{N-1}.$$

On peut prendre le problème dans un autre sens : pour une valeur donnée de θ , on définit la valeur maximale $R_\theta = \max_i (q_{ii}) \delta t / \delta x^2$ qui assure la stabilité. Pour les premières dimensions, on obtient le tableau suivant.

N	$\bar{\theta}$	$\underline{\theta}$	R_0	R_1
2	0,500	0,500	0,250	∞
3	0,696	0,666	0,148	∞
4	0,899	0,844	0,093	∞
5	1,104	1,024	0,064	2,1016
6	1,309	1,206	0,046	0,2163

TABLE 1 – Stabilité du schéma de Douglas

À partir de la dimension 5, on ne peut pas espérer avoir de schéma précis car pour avoir la stabilité, on sort du domaine de validité du θ -schéma sur lequel est basée cette méthode.

8.1.2 La méthode de Craig et Sneyd

La méthode de Craig et Sneyd [6] est une variante de la méthode de Douglas. Pour cela, on appelle L_0 la partie mixte du schéma :

$$L_0 = 2 \sum_{i=2}^N \sum_{j=1}^{i-1} q_{ij} \delta_{x_i x_j}^2.$$

On utilise la méthode de Douglas comme une étape de prédiction

$$P_\theta y^{n+1} = (P_\theta + \delta t L) u^n,$$

à laquelle on associe une étape de correction

$$P_\theta u^{n+1} = (P_\theta + \delta t L) u^n + \lambda \delta t L_0 (y^{n+1} - u^n).$$

Ceci peut s'écrire de manière compacte

$$\begin{aligned} \text{(CS)} \quad u^{n+1} &= P_\theta^{-1} [(P_\theta + \delta t L) u^n + \lambda \delta t L_0 \delta t P_\theta^{-1} L u^n] \\ &= (I + \delta t P_\theta^{-1} L + \lambda \delta t^2 P_\theta^{-1} L_0 P_\theta^{-1} L) u^n \\ &= [I + (I + \lambda \delta t P_\theta^{-1} L_0) \delta t P_\theta^{-1} L] u^n. \end{aligned}$$

Pour l'implémentation, ceci se réécrit

$$\begin{aligned} y^{n+1(0)} &= (I + \delta t L) u^n, \\ (I - \theta q_{jj} \delta t \delta_{x_j x_j}^2) y^{n+1(j)} &= y^{n+1(j-1)} - \theta q_{jj} \delta t \delta_{x_j x_j}^2 u^n, \quad 1 \leq j \leq N, \\ y^{n+1} &= y^{n+1(N)}, \\ u^{n+1(0)} &= y^{n+1(0)} + \lambda \delta t L_0 (y^{n+1} - u^n), \\ (I - \theta q_{jj} \delta t \delta_{x_j x_j}^2) u^{n+1(j)} &= u^{n+1(j-1)} - \theta q_{jj} \delta t \delta_{x_j x_j}^2 u^n, \quad 1 \leq j \leq N, \\ u^{n+1} &= u^{n+1(N)}. \end{aligned}$$

Dans le cas où $\lambda = \theta = 1/2$, ce schéma est d'ordre 2 en temps (et en espace).

Par ailleurs, pour cette méthode, on sait trouver (voir [6, 7]) une condition suffisante de stabilité inconditionnelle reliant λ à la dimension N et à θ . Cette condition s'écrit

$$\lambda \leq \frac{N^{N-1}}{(N-1)^N} \theta.$$

Il faut aussi nécessairement $\theta \geq 1/2$. Pour essayer d'obtenir l'ordre 2 et la stabilité, on va s'intéresser à la valeur minimale θ_{\min} de θ qui assure la stabilité inconditionnelle pour $\lambda = 1/2$.

N	2	3	4	5	6
$\frac{N^{N-1}}{(N-1)^N}$	2,000	1,125	0,790	0,610	0,498
θ_{\min}	0,500	0,500	0,633	0,819	1,005

TABLE 2 – Stabilité du schéma de Craig et Sneyd

Dans le cadre des dimensions 2 et 3, on peut avoir un schéma d'ordre 2 en temps et en espace et inconditionnellement stable. On a gagné une dimension d'espace permettant la stabilité inconditionnelle. Le cas $N = 6$ pose encore problème.

Si on essaie d'appliquer cette méthode au cas sans dérivée mixte ($L_0 = 0$), on trouve que seul le résultat de l'étape de correction sert et $P_\theta u^{n+1} = (P_\theta + \delta t L)u^n$. On retrouve le schéma de Douglas.

8.1.3 La méthode de Hundsdorfer et Verwer

La méthode de Hundsdorfer et Verwer [14, 34] est une autre variante de la méthode de Douglas. À nouveau, on utilise la méthode de Douglas comme une étape de prédiction

$$P_\theta y^{n+1} = (P_\theta + \delta t L)u^n,$$

puis on effectue une étape de correction

$$P_\theta u^{n+1} = P_\theta y^{n+1} + \delta t L u^n + \mu \delta t L (y^{n+1} - u^n).$$

Ceci peut s'écrire de manière compacte

$$\begin{aligned} \text{(HV)} \quad u^{n+1} &= P_\theta^{-1} [(P_\theta + \delta t L)u^n + \delta t L u^n + \mu \delta t L \delta t P_\theta^{-1} L u^n] \\ &= (I + 2\delta t P_\theta^{-1} L + \mu \delta t^2 P_\theta^{-1} L P_\theta^{-1} L) u^n \\ &= [I + (2I + \mu \delta t P_\theta^{-1} L) \delta t P_\theta^{-1} L] u^n. \end{aligned}$$

Pour l'implémentation, ceci se réécrit

$$\begin{aligned} y^{n+1(0)} &= (I + \delta t L)u^n, \\ (I - \theta q_{jj} \delta t \delta_{x_j x_j}^2) y^{n+1(j)} &= y^{n+1(j-1)} - \theta q_{jj} \delta t \delta_{x_j x_j}^2 u^n, \quad 1 \leq j \leq N, \\ y^{n+1} &= y^{n+1(N)}, \\ u^{n+1(0)} &= y^{n+1(0)} + \mu \delta t L (y^{n+1} - u^n), \\ (I - \theta q_{jj} \delta t \delta_{x_j x_j}^2) u^{n+1(j)} &= u^{n+1(j-1)} - \theta q_{jj} \delta t \delta_{x_j x_j}^2 y^{n+1}, \quad 1 \leq j \leq N, \\ u^{n+1} &= u^{n+1(N)}. \end{aligned}$$

Contrairement au schéma de Craig et Sneyd, l'étape de correction a ici un rôle même en absence de dérivée mixte ($L_0 = 0$).

Pour $\mu \geq 1$, le schéma est inconditionnellement stable si $\theta \geq \mu a_N N$, où a_N est définie de manière implicite comme l'unique solution dans $]0, 1/2[$ de l'équation

$$2a \left(1 + \frac{1-a}{N-1} \right)^{N-1} - 1 = 0.$$

L'équivalent du tableau 2 étendu au schéma de Hundsdorfer et Verwer est donné par le tableau 3.

N	2	3	4	5	6	7	8	9
θ_{\min} pour (CS)	0,500	0,500	0,633	0,819	1,005	1,190	1,374	1,559
θ_{\min} pour (HV)	0,293	0,402	0,515	0,630	0,745	0,860	0,975	1,091

TABLE 3 – Comparaison de la stabilité des schémas de Craig et Sneyd (CS) et de Hundsdorfer et Verwer (HV)

Le schéma de Hundsdorfer et Verwer permet de gagner trois dimensions d'espace pour lesquelles on peut effectuer des calculs stables tout en gardant la consistance du θ -schéma.

8.2 Application au modèle de Black et Scholes à plusieurs actifs

Pour le modèle de Black et Scholes à plusieurs actifs, on combine les résultats précédents. Nous allons donc utiliser des méthodes de splitting adaptées aux dérivées mixtes. L'étude du problème mixte seul nous montre qu'il va falloir utiliser un θ -schéma pour traiter la dérivée seconde, avec $\theta \geq 1/2$. On peut aussi le faire pour les dérivées d'ordre un, étant donné que l'on a de toutes façons peu de chance d'obtenir autre chose qu'un schéma d'ordre 1. On peut par exemple répartir le terme linéaire $r(t)u$.

Les termes de bord sont en général traités au premier pas, même si Craig et Sneyd [6] suggèrent d'autres approches pour conserver l'ordre 2.

On se retrouve donc avec un système discrétisé en espace de type

$$U'(t) = AU(t) + S(t),$$

où la matrice A se décompose en $A = A_0 + A_1 + \dots + A_N$. Dans cette décomposition, A_0 contient les termes issus des dérivées mixtes et les A_i , $1 \leq i \leq N$ les termes issus des dérivées selon s_i (et du terme $r(t)u$). Le schéma de Douglas s'écrit alors

$$\begin{aligned} u^{n+1(0)} &= (I + \delta t A)u^n + \delta t S(t^n), \\ (I - \theta \delta t A_j)u^{n+1(j)} &= u^{n+1(j-1)} - \theta \delta t A_j u^n, \quad 1 \leq j \leq N, \\ u^{n+1} &= u^{n+1(N)} \end{aligned}$$

ou encore

$$Q_\theta u^{n+1} = (Q_\theta + \delta t A)u^n + \delta t S(t^n) \quad \text{avec } Q_\theta = \prod_{i=1}^N (I - \theta \delta t A_i).$$

De même, on peut généraliser la méthode de Craig et Sneyd

$$\begin{aligned} Q_\theta y^{n+1} &= (Q_\theta + \delta t A)u^n + \delta t S(t^n), \\ Q_\theta u^{n+1} &= (Q_\theta + \delta t A)u^n + \delta t S(t^n) + \lambda \delta t A_0 (y^{n+1} - u^n) \end{aligned}$$

ou celle de Hundsdorfer et Verwer

$$\begin{aligned} Q_\theta y^{n+1} &= (Q_\theta + \delta t A)u^n + \delta t S(t^n), \\ Q_\theta u^{n+1} &= Q_\theta y^{n+1} + \delta t A u^n + \delta t S(t^n) + \mu \delta t A (y^{n+1} - u^n). \end{aligned}$$

Comment paramétrer ces schémas? Il faut s'inspirer des analyses effectuées dans les cas plus simples. Si on se place dans le cadre du système en variable log, comme $q_{ii}(t, s) = \frac{1}{2} \sigma_{ii}^2(t, s)$, on a un encadrement des coefficients q_{ii} qui permettent d'utiliser les valeurs du tableau 1 pour estimer les paramètres pour le schéma de Douglas.

8.3 Application au modèle de Heston

Pour le modèle de Heston, on fait des choses semblables au cas du modèle de Black et Scholes à plusieurs actifs.

Ici, la dimension "spatiale" est $N = 2$, paradis promis pour obtenir l'ordre 2. Nous voulons aussi bien sûr la stabilité.

- Pour le schéma de Douglas, l'ordre est toujours 1, et il suffit d'avoir $\theta \geq 1/2$ pour assurer la stabilité.
- Pour le schéma de Craig et Sneyd, l'ordre 2 nécessite $\theta = \lambda = 1/2$. Pour la stabilité, il suffit pour $\theta \geq 1/2$ d'avoir de plus $\lambda \leq 2\theta$ (voir tableau 2), ce qui est clairement assuré pour $\lambda = 1/2$.

- Pour le schéma de Hundsdorfer et Verwer, l'ordre 2 nécessite $\mu = 1/2$, indépendamment de θ . Il n'y a pas de résultat complet concernant la stabilité. En absence de convection ($r_d(t) - r_f(t) = 0$ et $\eta - \nu = 0$ — impossible à obtenir même en passant en variable \log en ν), on obtient un schéma inconditionnellement stable pour $\mu = 1/2$ et $\theta \geq 1/(2 + \sqrt{2}) \simeq 0,293$ (ensemble de valeurs pour lesquelles le schéma est d'ordre 2). En revanche en présence de termes de convection, on a uniquement une conjecture de stabilité inconditionnelle pour $\mu = 1/2$ et $\theta \geq 1/2 + 1/2\sqrt{3} \simeq 0,789$ [21].

8.4 Sélection d'articles sur le splitting appliqué en mathématiques financières

8.4.1 Modèle de Black et Scholes à plusieurs actifs

Lo et Hui [23, 24] traitent du modèle de Black et Scholes à plusieurs actifs avec un raisonnement fortement basé sur une théorie algébrique et la commutation des opérateurs. Le splitting est de type Lie et sépare également les dérivées premières et secondes associées aux mêmes directions.

8.4.2 Modèle de Heston

J'ai basé la présentation de mon cours sur un article court de in 't Hout [19] qui traite des schémas de Douglas, Craig et Sneyd et Hundsdorfer et Verwer pour le modèle de Heston pour un call européen avec r_d . Il rappelle les résultats d'ordre et de stabilité de la littérature. Des calculs numériques sont effectués pour un jeu de valeur des paramètres :

$$\rho = 0,8 \text{ (forte corrélation)}, r_d = 0,03, r_f = 0, \eta = 0,2, \kappa = 2, \sigma = 0,3, K = 100, T = 1.$$

Les choix numériques sont la plage de calcul : $s \in [0, 200]$, $\nu \in [0, 1]$, avec $M = 100$ ou 200 points dans la direction s et $M/2$ points dans la direction ν . Les paramètres θ , λ et μ sont choisis pour optimiser ordre et stabilité.

Dans un article plus long, in 't Hout et Foulon [20] traitent du même problème mais en introduisant quelques choix algorithmiques supplémentaires comme par exemple le raffinement du maillage au tour de $s = K$ et de $\nu = 0$. L'analyse des cas test est plus précise et cette fois-ci quatre jeux de paramètres sont proposés incluant des rapports entre paramètres assez différents.

8.4.3 Splitting pour Black et Scholes non linéaire

Ševčovič [31] présente un modèle de Black et Scholes non linéaire :

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

où q est le taux de dividende. On remarque la dépendance de σ en $s^2 \partial_{ss}^2 u$, ce qui rend ce modèle non linéaire. Un splitting est appliqué à cette équation en variable \log , mais cette fois-ci il n'y a pas de directions à alterner. Dans cet article, on fait deux pas, le premier traitant la partie convective (dérivées premières) et le second la partie diffusive non linéaire.

8.4.4 Splitting pour les options américaines

Une série d'articles de Ikonen et Toivanen [15, 16, 17, 18] traitent d'un problème non abordé dans ce cours mais en forte relation : les méthodes de splitting pour les options américaines. Toutes ces méthodes s'écrivent sous la forme d'un problème de complémentarité linéaire

$$\begin{cases} (u_t - Au) \geq 0, u \geq g, \\ (u_t - Au)(u - g) = 0, \end{cases}$$

où g est la fonction putoff. Ici le temps est décroissant. La méthode de splitting est alors dérivée de la formulation avec un multiplicateur de Lagrange (voir n'importe quel livre d'optimisation)

$$\begin{cases} (u_t - Au) = \lambda, \lambda \geq 0, u \geq g, \\ \lambda(u - g) = 0. \end{cases}$$

Les étapes se font en faisant décroître l'indice de temps n , pour des pas de temps variables Plus précisément, on a une première étape de type θ -schéma :

$$(I + \theta^n \delta t^n A) \tilde{u}^n = (I - (1 - \theta^n) \delta t^n A) u^{n+1} + \delta t^n b(t^n) + \delta t^n \lambda^{n+1},$$

suivie d'une étape où tous les points spatiaux sont décorrélés :

$$\begin{cases} u^n - \tilde{u}^n = \delta t^n (\lambda^n - \lambda^{n+1}), \lambda \geq 0, u \geq g, \\ \lambda^n \cdot (v^n - g) = 0. \end{cases}$$

Ces articles traitent basiquement trois types de modèles :

- le modèle de Black et Scholes
- des modèles de diffusion à saut du type

$$Au(t, s) = -\frac{1}{2} \sigma^2(t, s) s^2 \partial_{ss}^2 u(t, s) - (r - \mu \zeta) s \partial_s u(t, s) + (r + \mu) u(t, s) - \mu \int_{\mathbb{R}^+} u(t, ys) f(y) dy,$$

où μ est le taux des sauts, la fonction f définit la distribution des sauts et ζ est l'amplitude moyenne des sauts.

- le modèle de Heston.

Références

- [1] Y. Achdou et O. Pironneau, Computational methods for option pricing, collection Frontiers in applied mathematics, volume 30. Society for Industrial and Applied Mathematics, 2005.
- [2] B. Bidégaray-Fesquet, *Méthodes de splitting pour les problèmes multi-échelles*, cours de M2R, 2007. <http://ljk.imag.fr/membres/Brigitte.Bidegaray/cours2007.pdf>
- [3] B. Bidégaray-Fesquet, *Méthodes de splitting pour les problèmes multi-échelles, complément 2008*, cours de M2R, 2008. <http://ljk.imag.fr/membres/Brigitte.Bidegaray/complement2008.pdf>
- [4] F. Black et M. Scholes, *The pricing of options and corporate liabilities*, Journal of Political Economy, 81(3), 637-654, 1973.
- [5] P. L. T. Brian, *A finite-difference method of high-order accuracy for the solution of three-dimensional transient heat conduction problems*, A.I.Ch.E. Journal, 7(3), 367-370, 1961.
- [6] I.J.D. Craig et A.D. Sneyd, *An alternating-direction implicit scheme for parabolic equations with mixed derivatives*, Computers & Mathematics with Applications, 14(4), 341-350, 1988.
- [7] I.J.D. Craig et A.D. Sneyd. *An alternating direction implicit scheme for parabolic systems of partial differential equations*, Computers & Mathematics with Applications, 20(3), 53-62, 1990.
- [8] J. Douglas, Jr., *On the numerical integration of $\partial^2 u / \partial x^2 + \partial^2 u / \partial y^2 = \partial u / \partial t$ by implicit methods*, Journal of the Society of Industrial and Applied Mathematics, 3, 42-65, 1955.
- [9] J. Douglas, Jr., *On the numerical integration of quasilinear parabolic differential equations*, Pacific Journal of Mathematics, 6(1), 35-42, 1956.
- [10] J. Douglas, Jr., *Alternating direction methods for three space variables*, Numerische Mathematik, 4(1), 41-63, 1962.
- [11] J. Douglas, Jr et J.E. Gunn, *A general formulation of alternating direction methods. Part I. parabolic and hyperbolic problems*, Numerische Mathematik, 6(1), 428-453, 1964.
- [12] J. Douglas, Jr. et H. H. Rachford, Jr., *On the numerical solution of heat conduction problems in two and three space variables*, Transactions of the American Mathematical Society, 82, 421-439, 1956.

- [13] S. L. Heston, *A closed-form solution for options with stochastic volatility with applications to bonds and currency options*, *The Review of Financial Studies*, 6(2), 327–343, 1993.
- [14] W. H. Hundsdorfer, *Accuracy and stability of splitting with stabilizing corrections*, *Applied Numerical Mathematics*, 42(1–3), 213–233, 2002.
- [15] S. Ikonen et J. Toivanen, *Operator splitting methods for American option pricing*, *Applied Mathematics Letters*, 17(7), 809–814, 2004.
- [16] S. Ikonen et J. Toivanen, *An operator splitting method for pricing American option*, in R. Glowinski and P. Neittaanmäki ed., *Partial Differential Equations*, collection *Computational Methods in Applied Sciences*, volume 16, pages 279–292. Springer, 2008.
- [17] S. Ikonen et J. Toivanen, *Efficient numerical methods for pricing American options under stochastic volatility*, *Numerical Methods for Partial Differential Equations*, 24(1), 104–126, 2008.
- [18] S. Ikonen et J. Toivanen, *Operator splitting methods for pricing American options under stochastic volatility*, *Numerische Mathematik*, 113(2), 299–324, 2009.
- [19] K.J. in 't Hout, *ADI schemes in the numerical solution of the Heston PDE*, in T.E. Simos, G. Psihoyios et C. Tsitouras ed., collection *International Conference on Numerical Analysis and Applied Mathematics*, AIP Conference Proceedings, volume 936, pages 10–14. AIP, 2007.
- [20] K.J. in 't Hout et S. Foulon, *ADI finite difference schemes for option pricing in the Heston model with correlation*, *International Journal of Numerical Analysis and Modeling*, 7(2), 303–320, 2010.
- [21] K.J. in 't Hout et B.D. Welfert, *Stability of ADI schemes applied to convection–diffusion equations with mixed derivative terms*, *Applied Numerical Mathematics*, 57(1), 19–35, 2007.
- [22] D. Lanser, J.G. Blom et J.G. Verwer. *Time integration of the shallow water equations in spherical geometry*, *Journal of Computational Physics*, 171(1), 373–393, 2001.
- [23] C.-F. Lo et C.-H. Hui, *Valuation of financial derivatives with time-dependent parameters : Lie-algebraic approach*, *Quantitative Finance*, 1(1), 73–78, 2001.
- [24] C.-F. Lo et C.-H. Hui, *Pricing multi-asset financial derivatives with time-dependent parameters – Lie algebraic approach*, *International Journal of Mathematics and Mathematical Sciences*, 32(7), 401–410, 2002.
- [25] S. McKee et A.R. Mitchell, *Alternating direction methods for parabolic equations in two space dimensions with a mixed derivative*, *The Computer Journal*, 13, 81–86, 1970.
- [26] S. McKee, D.P. Wall et S.K. Wilson, *An alternating direction implicit scheme for parabolic equations with mixed derivative and convective terms*, *Journal of Computational Physics*, 126(1), 64–76, 1996.
- [27] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, *Applied Mathematical Sciences* 44, Springer, 1983.
- [28] D.W. Peaceman et H.H. Rachford Jr., *The numerical solution of parabolic and elliptic differential equations*, *Journal of the Society of Industrial and Applied Mathematics*, 3, 28–41, 1955.
- [29] M. Schatzman, *Higher order alternate direction methods*, *Computational Methods in Applied Mechanical Engineering*, 116, 219–225, 1994.
- [30] M. Schatzman, *Stability of the Peaceman–Rachford approximation*, *Journal of Functional Analysis*, 162, 219–255, 1999.
- [31] D. Ševčovič. *An iterative algorithm for evaluating approximations to the optimal exercise boundary for a nonlinear Black–Scholes equation*, *Canadian Applied Mathematics Quarterly*, 15(1), 77–97, 2007.
- [32] G. Strang, *On the construction and comparison of difference schemes*, *SIAM Journal on Numerical Analysis*, 5(3), 506–517, 1968.
- [33] P. J. van der Houwen et J. G. Verwer. *One-step splitting methods for semi-discrete parabolic equations*, *Computing*, 22(4), 291–309, 1979.
- [34] J. G. Verwer, E. Spee, J. G. Blom et W. H. Hundsdorfer, *A second-order Rosenbrock method applied to photochemical dispersion problems*, *SIAM Journal on Scientific Computing*, 20(4), 1456–1480, 1999.

A Compléments

A.1 Semi-groupes uniformément continus d'opérateurs linéaires bornés

Soit X un espace de Banach. Une famille à un paramètre $\mathcal{A}(t) : X \rightarrow X$, d'opérateurs linéaires bornés, est un **semi-groupe d'opérateurs linéaires bornés** sur X si les propriétés (P1) et (P2) sont vérifiées. Cette famille est **uniformément continue** si de plus

$$(P3) \lim_{t \downarrow 0} \|\mathcal{A}(t) - I\| = 0.$$

On définit l'opérateur linéaire A par son domaine

$$D(A) = \left\{ x \in X; \lim_{t \downarrow 0} \frac{\mathcal{A}(t)x - x}{t} \text{ existe} \right\}$$

et

$$Ax = \lim_{t \downarrow 0} \frac{\mathcal{A}(t)x - x}{t} = \left. \frac{d_+ \mathcal{A}(t)x}{dt} \right|_{t=0} \text{ pour } x \in D(A).$$

Cet opérateur linéaire est le **générateur infinitésimal** du semi-groupe $\mathcal{A}(t)$. Dans cette terminologie se trouve sous-jacent le fait qu'un seul semi-groupe permet de construire A comme ci-dessus. Le lien entre A et \mathcal{A} est bijectif. Si A est un opérateur borné sur X , on peut écrire

$$\mathcal{A}(t) = \exp(tA) = \sum_{n=0}^{\infty} \frac{(tA)^n}{n!}$$

et la série converge bien en norme pour tout $t \geq 0$.

Théorème 4

Il y a équivalence entre

- (i) A est un opérateur linéaire borné,
 - (ii) \mathcal{A} est uniformément continu.
-

Le semi-groupe $\mathcal{A}(t)$ commute avec son générateur A .

Théorème 5

L'application $t \mapsto \mathcal{A}(t)$ est différentiable en norme et

$$\frac{d\mathcal{A}(t)}{dt} = A\mathcal{A}(t) = \mathcal{A}(t)A.$$

Dans notre exemple, la propriété (P3) s'écrit

$$\lim_{t \downarrow 0} |\exp(at) - 1| = 0.$$

Le domaine de A est

$$D(A) = \left\{ x \in \mathbb{R}; \lim_{t \downarrow 0} \frac{\exp(at)x - x}{t} \text{ existe} \right\} = \mathbb{R}$$

et

$$Ax = \lim_{t \downarrow 0} \frac{\exp(at)x - x}{t} = ax.$$

A.2 Preuve du théorème 1

On pose $h = t/n$. On a $x^m = f(h)^m x^0$ et $x(mh) = \exp(mhC)x(0)$ et on prend bien sûr $x^0 = x(0)$. L'erreur locale est l'erreur de méthode et est définie par

$$\eta^{m+1} = (\exp(hC) - f(h))x(mh).$$

L'erreur globale est définie par

$$e^m = x(mh) - x^m.$$

On peut écrire la relation de récurrence

$$\begin{aligned} e^{m+1} &= x((m+1)h) - x^{m+1} \\ &= \exp(hC)x(mh) - f(h)x^m \\ &= f(h)(x(mh) - x^m) + (\exp(hC) - f(h))x(mh) \\ &= f(h)e^m + \eta^{m+1}. \end{aligned}$$

Comme $e^0 = 0$, on a de façon classique

$$\begin{aligned} e^n &= \sum_{k=0}^{n-1} f(h)^k \eta^{n-k}, \\ |e^n| &\leq \sum_{k=0}^{n-1} \|f(h)^k\| \|Rh^{p+1} + O(h^{p+2})\| \max_{s \in [0, T]} |x(s)| \end{aligned}$$

On se place dans le cadre de méthodes stables. Pour des temps en $O(1)$, on a donc $\|f(h)^k\| = O(1)$ et $\max_{s \in [0, T]} |x(s)| = O(1)$. Ainsi

$$|e^n| \leq Cn(h^{p+1} + O(h^{p+2})) = tC(h^p + O(h^{p+1})).$$

Enfin $e^n = (\exp(tC) - f(t/n)^n)x^0$, ce qui prouve le résultat.

A.3 Stabilité faible, avec perte de régularité

On suppose que A et B sont des opérateurs négatifs auto-adjoints. Pour $h > 0$, on définit sur $D(A)$ la norme

$$\|u\|_h = \|(I - \frac{1}{2}hA)u\|_H.$$

Lemme 2

Pour tout $h > 0$ et tout $u \in D(A)$, $\mathcal{P}(h)A \in D(A)$ et

$$\|\mathcal{P}(h)u\|_h \leq \|u\|_h.$$

Preuve:

Si $u \in D(A)$ alors $(I + hA/2)u \in H$. Par ailleurs, comme B est négatif auto-adjoint $(I + hB/2)(I - hB/2)^{-1}$ est une contraction dans H . Enfin $(I - hA/2)^{-1}$ envoie continuellement H dans $D(A)$. Ainsi $\mathcal{P}(h)$ est bien dans $D(A)$ et en outre

$$\begin{aligned} \|\mathcal{P}(h)u\|_h &\leq \|(I + \frac{1}{2}hB)(I - \frac{1}{2}hB)^{-1}(I + \frac{1}{2}hA)(I - \frac{1}{2}hA)^{-1}(I - \frac{1}{2}hA)u\|_H \\ &\leq \|u\|_h. \end{aligned}$$

■

En particulier, on peut appliquer ce lemme à $h = t/n$:

$$\|\mathcal{P}\left(\frac{t}{n}\right)u\|_{\frac{t}{n}} \leq \|u\|_{\frac{t}{n}}.$$

Comme A est un opérateur négatif, on a clairement

$$\|u\|_{t/n} = \|(I - \frac{t}{2n}A)u\|_H \geq \|u\|_H.$$

Par ailleurs pour $t/n \leq \tau$, on a également clairement

$$\|u\|_{t/n} = \|(I - \frac{t}{2n}A)u\|_H \leq \|u\|_H + \tau\|Au\|_H.$$

Ainsi

$$\|\mathcal{P}\left(\frac{t}{n}\right)u\|_H \leq \|\mathcal{P}\left(\frac{t}{n}\right)u\|_{\frac{t}{n}} \leq \|u\|_{\frac{t}{n}} \leq \|u\|_H + \tau\|Au\|_H.$$

Ceci est la formule de stabilité faible.

A.4 Contre-exemple de stabilité forte

On se place dans l'espace fonctionnel $H = L^2(\mathbb{R}^d)^2$ et on considère les opérateurs A et B sont le forme $A = M\Delta$ et $B = N\Delta$, où M et N sont des matrices symétriques définies positives. Les opérateurs A et B sont bien négatifs auto-adjoints. C'est *a priori* un cadre sympathique de type équation de la chaleur, réputé pour être bien posé. En particulier on a $D(A) = D(B) = H^2(\mathbb{R}^d)^2$.

Néanmoins, on a le résultat suivant.

Théorème 6

Pour tout $K > 0$, il existe un choix des matrices M et N tel que

$$\|\mathcal{P}(t)\|_{\mathcal{L}(H)} \geq K.$$

Comme ce théorème est vrai pour des $K > 1$, cela implique que le schéma de Peaceman–Rachford n'est pas fortement stable. Dans ce théorème, la norme utilisée $\|\cdot\|_{\mathcal{L}(H)}$ est celle des opérateurs linéaires sur l'espace de Hilbert H , vus comme des multiplicateurs de Fourier. On utilise la norme subordonnée L^2 dans cet espace :

$$\|A\|_{\mathcal{L}(H)} = \sup_{\xi \in \mathbb{R}^d} \frac{\|\widehat{Au}(\xi)\|_H}{\|\widehat{u}(\xi)\|_H}$$

où la transformée de Fourier est définie par

$$\widehat{u}(\xi) = \int_{\mathbb{R}^d} u(x) \exp(-2i\pi x \cdot \xi) dx.$$

Pour la formule de Peaceman–Rachford, on a

$$\begin{aligned} \|\mathcal{P}(t)\|_{\mathcal{L}(H)} &= \sup_{\xi \in \mathbb{R}^d} \|(I + \frac{t}{2}|\xi^2|M)^{-1}(I - \frac{t}{2}|\xi^2|N)(I + \frac{t}{2}|\xi^2|N)^{-1}(I - \frac{t}{2}|\xi^2|M)\|_{\mathcal{B}(\mathbb{R}^2)} \\ &= \sup_{r \in \mathbb{R}} \|(I + rM)^{-1}(I - rN)(I + rN)^{-1}(I - rM)\|_{\mathcal{B}(\mathbb{R}^2)}. \end{aligned}$$

En particulier, on remarque que $\|\mathcal{P}(t)\|_{\mathcal{L}(H)}$ ne dépend pas du temps t . On choisit les formes de matrice suivantes (symétriques définies positives)

$$M = \begin{pmatrix} a & b \\ b & a \end{pmatrix} \text{ avec } a > b > 0, \quad N = \begin{pmatrix} r^{-2} & 0 \\ 0 & 1 \end{pmatrix}.$$

On aura alors

$$\|\mathcal{P}(t)\|_{\mathcal{L}(\mathcal{H})} \geq \|(I + rM)^{-1}(I - rN)(I + rN)^{-1}(I - rM)\|_{\mathcal{B}(\mathbb{R}^2)}.$$

Nous allons faire tendre r vers $+\infty$. Nous supposons donc qu'il est grand et écrivons des développements limités en r^{-1} . Calculons tout d'abord

$$\begin{aligned} (I - rN)(I + rN)^{-1} &= 2(I + rN)^{-1} - I = 2 \begin{pmatrix} 1 + r^{-1} & 0 \\ 0 & 1 + r \end{pmatrix}^{-1} - I \\ &= \begin{pmatrix} \frac{1-r^{-1}}{1+r^{-1}} & 0 \\ 0 & \frac{1-r}{1+r} \end{pmatrix} = \begin{pmatrix} \frac{1-r^{-1}}{1+r^{-1}} & 0 \\ 0 & -\frac{1-r^{-1}}{1+r^{-1}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + O(r^{-1}). \end{aligned}$$

Par ailleurs

$$\begin{aligned} (I + rM)^{-1} &= \begin{pmatrix} 1 + ra & rb \\ rb & 1 + ra \end{pmatrix}^{-1} = \frac{1}{r} \begin{pmatrix} a + r^{-1} & b \\ b & a + r^{-1} \end{pmatrix}^{-1} \\ &= \frac{1}{r} \frac{1}{a^2 - b^2 + 2r^{-1} + r^{-2}} \begin{pmatrix} a + r^{-1} & -b \\ -b & a + r^{-1} \end{pmatrix} \\ &= \frac{1}{r} \left\{ \frac{1}{a^2 - b^2} \begin{pmatrix} a & -b \\ -b & a \end{pmatrix} + O(r^{-1}) \right\}, \\ (I - rM) &= -r \begin{pmatrix} a - r^{-1} & b \\ b & a - r^{-1} \end{pmatrix} = -r \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} + O(r^{-1}) \right\} \end{aligned}$$

Ainsi

$$\begin{aligned} (I + rM)^{-1}(I - rN)(I + rN)^{-1}(I - rM) &= -\frac{1}{a^2 - b^2} \begin{pmatrix} a & -b \\ -b & a \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} a & b \\ b & a \end{pmatrix} + O(r^{-1}) \\ &= -\frac{1}{a^2 - b^2} \begin{pmatrix} a^2 + b^2 & 2ab \\ -2ab & -(a^2 + b^2) \end{pmatrix} + O(r^{-1}) \end{aligned}$$

Calculer la norme $\mathcal{B}(\mathbb{R}^2)$, c'est calculer la norme subordonnée L^∞ :

$$\|A\|_\infty = \max_j \sum_k |a_{jk}|.$$

On obtient ainsi

$$\|(I + rM)^{-1}(I - rN)(I + rN)^{-1}(I - rM)\|_{\mathcal{B}(\mathbb{R}^2)} = \frac{a + b}{a - b} + O(r^{-1}).$$

On peut choisir a et b de telle façon que $(a + b)/(a - b) \geq K + 1$. Pour r suffisamment grand, on aura bien $\|\mathcal{P}(t)\|_{\mathcal{L}(\mathcal{H})} \geq K$ ce qui démontre le théorème 6.

Le problème tient vraiment au fait que l'on a une EDP non scalaire. Le commutateur $[A, B]$ est un opérateur différentiel d'ordre 4. Le problème ne se pose plus si on travaille dans $L^2(\mathbb{R}^d)^2$ ou pour une EDO comme le montre l'appendice A.5.

A.5 Le problème du laplacien sans laplacien

En effet, si on considère le problème

$$\dot{x} = Mx + Nx, \quad x(0) = x^0 \in \mathbb{R}^2,$$

Il faut cette fois-ci calculer

$$\|\mathcal{P}(t)\|_{\mathcal{L}(\mathcal{H})} = \|(I - \frac{t}{2}M)^{-1}(I + \frac{t}{2}N)(I - \frac{t}{2}N)^{-1}(I + \frac{t}{2}M)\|_{\mathcal{B}(\mathbb{R}^2)}.$$

Maintenant, il suffit de faire des calculs valables pour $\tau = t/2$ petit. On n'utilise plus les équivalents. On a

$$(I + \tau N)(I - \tau N)^{-1} = \begin{pmatrix} \frac{1+\tau^{-1}}{1-\tau^{-1}} & 0 \\ 0 & -\frac{1+\tau^{-1}}{1-\tau^{-1}} \end{pmatrix} = \frac{1+\tau}{1-\tau} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Par ailleurs

$$\begin{aligned} (I - \tau M)^{-1} &= -\frac{1}{\tau} \frac{1}{a^2 - b^2 - 2\tau^{-1} + \tau^{-2}} \begin{pmatrix} a - \tau^{-1} & -b \\ -b & a - \tau^{-1} \end{pmatrix} \\ &= \frac{1}{1 - 2\tau + a^2\tau^2 - b^2\tau^2} \begin{pmatrix} 1 - a\tau & b\tau \\ b\tau & 1 - a\tau \end{pmatrix}, \end{aligned}$$

et

$$(I + \tau M) = \begin{pmatrix} 1 + a\tau & b\tau \\ b\tau & 1 + a\tau \end{pmatrix}.$$

Ainsi

$$\begin{aligned} &(I - \tau M)^{-1}(I + \tau N)(I - \tau N)^{-1}(I + \tau M) \\ &= -\frac{1+\tau}{1-\tau} \frac{1}{1 - 2\tau + a^2\tau^2 - b^2\tau^2} \begin{pmatrix} 1 - (a^2 + b^2)\tau^2 & -2ab\tau^2 \\ 2ab\tau^2 & -(1 - (a^2 + b^2)\tau^2) \end{pmatrix}. \end{aligned}$$

On calcule la norme subordonnée L^∞ pour obtenir

$$\|(I - \tau M)^{-1}(I + \tau N)(I - \tau N)^{-1}(I + \tau M)\|_{\mathcal{B}(\mathbb{R}^2)} = \frac{1+\tau}{1-\tau} \frac{1 - (a+b)^2\tau^2}{1 - 2\tau + a^2\tau^2 - b^2\tau^2}.$$

Il n'y a plus de problème de stabilité.

Table des matières

1	Les modèles d'équations aux dérivées partielles en mathématiques financières	1
1.1	Le modèle de Black et Scholes	1
1.2	Premières propriétés du modèle de Black et Scholes	2
1.3	Conditions au bord	2
1.4	Modèle de Black et Scholes en variable log	2
1.5	Le modèle de Heston	3
2	Approximation du modèle de Black et Scholes à un actif	3
2.1	Différences finies pour l'équation de la chaleur en dimension 1 — θ -schéma.	4
2.2	Discrétisation du laplacien sur des grilles irrégulières	5
2.3	Les conditions aux bords	5
2.3.1	Condition de Dirichlet	5
2.3.2	Condition de Neumann	6
2.3.3	Condition de Robin	6
2.3.4	Application à l'équation de Black et Scholes sur grille irrégulière	6
2.4	Traitement des dérivées d'ordre un	6
2.5	Application au modèle de Black et Scholes à un actif	7
2.5.1	Écriture explicite du système linéaire	7
2.5.2	Résultats théoriques	8
2.5.3	Inversion pratique du système linéaire — algorithme LS	9
2.5.4	Stockage	9
3	Modèle de Black et Scholes à plusieurs actifs	9
3.1	Grandes dimensions – stockage	9
3.2	L'idée du splitting	10
3.3	Dérivées mixtes	10
4	Introduction aux méthodes de splitting	11
4.1	Une équation scalaire	11
4.2	Quand le splitting présente un intérêt	11
4.3	Réécriture des schémas de splitting grâce aux semi-groupes d'évolution	12
5	Le cas des systèmes linéaires	13
5.1	Splitting de Lie	13
5.2	Splitting de Strang	13
5.3	Splittings d'ordre plus élevé	14
5.3.1	Calcul de l'ordre et convergence	14
5.3.2	Une réponse négative	14
5.3.3	Combinaison linéaires d'approximations d'ordre inférieur	15
5.3.4	Extrapolations de Richardson	15
6	Approximation de l'exponentielle	16
6.1	Schéma d'Euler	18
6.2	Schéma d'Euler rétrograde	18
6.3	Couplage Euler–Euler rétrograde	19
6.3.1	Méthode de Crank–Nicolson	19
6.3.2	Le θ -schéma	20
6.4	Vers l'ordre 2 pour un splitting approché	20
6.4.1	Combinaison de Strang et Crank–Nicolson	20
6.4.2	Méthode de Peaceman–Rachford	21
6.5	Et si A et B commutent ?	21

6.5.1	Schéma d'Euler–Lie	22
6.5.2	Schéma d'Euler rétrograde–Lie	22
6.5.3	Schéma de Peaceman–Rachford	22
7	Généralités sur l'équation de la chaleur	22
7.1	Formule de Trotter–Kato	22
7.2	Notions de stabilité	23
7.3	La méthode des directions alternées	23
7.3.1	La méthode ADI historique	23
7.3.2	Cas du θ -schéma	25
7.3.3	Extensions	25
8	Application aux modèles en mathématiques financières	25
8.1	Traitement des termes mixtes	25
8.1.1	La méthode de Douglas	26
8.1.2	La méthode de Craig et Sneyd	27
8.1.3	La méthode de Hundsdorfer et Verwer	28
8.2	Application au modèle de Black et Scholes à plusieurs actifs	29
8.3	Application au modèle de Heston	29
8.4	Sélection d'articles sur le splitting appliqué en mathématiques financières	30
8.4.1	Modèle de Black et Scholes à plusieurs actifs	30
8.4.2	Modèle de Heston	30
8.4.3	Splitting pour Black et Scholes non linéaire	30
8.4.4	Splitting pour les options américaines	30
A	Compléments	33
A.1	Semi-groupes uniformément continus d'opérateurs linéaires bornés	33
A.2	Preuve du théorème 1	34
A.3	Stabilité faible, avec perte de régularité	34
A.4	Contre-exemple de stabilité forte	35
A.5	Le problème du laplacien sans laplacien	36