

Implémentation des méthodes

4.1	Algorithme Brennan-Schwartz	49
4.2	La méthode PSOR	65

Ce chapitre est consacré à la résolution numérique du problème de complémentarité linéaire issue de la discrétisation par différences finies ou éléments finis. Vu que, la méthode des éléments finis et de différences finies génèrent des matrices tridiagonales et du faite que la contrainte est imposée, alors ils existent des méthodes efficaces et appropriés pour résoudre ce type du problème. Parmi ces méthodes, la méthode de Brennan-Schwartz et la méthode de sur-relaxation avec projections successives (PSOR "Projected

Successive Over-Relaxation"). En sus, nous justifions théoriquement l'adaption de l'algorithme de Brennan-Schwartz au cas de discrétisation par éléments finis dans la première section. Dans la deuxième section, nous présentons l'algorithme de PSOR, selon [36], a fin de justifier numériquement la convergence de l'algorithme de Brennan-Schwartz dans le prochain chapitre observations numériques.

4.1 Algorithme Brennan-Schwartz

M.J. Brennan and E.S. Schwartz ont introduit un algorithme de résolution des options américaines put avec conditions aux bords mixtes Dirichlet et Neumann (voir [9]) par différences finies. P. Jaillet, D. Lembreton et B. Lapeyre ont reformulé le problème avec des conditions aux bords Dirichlet, puis ils ont donné la justification de cet algorithme introduit par [9] pour le cas de différences finies.

Nous avons remarqué que lors de la discrétisation soit par la méthode des éléments finis ou différences finies que le type de matrice qui apparaît est tridiagonale, de la forme :

$$\mathbf{T} = \begin{pmatrix} b_1 & c_1 & 0 & \cdots & 0 \\ a_2 & b_2 & c_2 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & a_{m-1} & b_{m-1} & c_{m-1} \\ 0 & \cdots & 0 & a_m & b_m \end{pmatrix}.$$

La proposition 4.1.1 donne un algorithme directe de résolution du PCL de type

$$\begin{aligned} \mathbf{T}u &\geq \varrho, \\ u &\geq \xi, \\ (\mathbf{T}u - \varrho, u - \xi) &= 0, \end{aligned} \tag{4.1}$$

dans le cas où les éléments c_j de $\mathbf{T}$ sont nuls.

Proposition 4.1.1 (voir [10]) Une matrice du type $\mathbf{T}$ avec $c_j = 0$ et $b_j > 0$ pour $j = 1, \dots, m$, alors, pour tous vecteurs ξ et ϱ le PCL suivant :

$$\begin{aligned} \mathbf{T}u &\geq \varrho, \\ u &\geq \xi, \\ (\mathbf{T}u - \varrho, u - \xi) &= 0, \end{aligned} \tag{4.2}$$

a une solution unique donnée par les équations de récurrence :

$$\begin{aligned} u_1 &= \max \left\{ \frac{\varrho_1}{b_1}, \xi_1 \right\}, \\ u_j &= \max \left\{ \frac{1}{b_j} (\varrho_j - a_j u_{j-1}), \xi_j \right\}, \quad 2 \leq j \leq m. \end{aligned}$$

La preuve de ce résultat découle des égalités : $(\mathbf{T}u)_1 = b_1 u_1$ et $(\mathbf{T}u)_j = b_j u_j + a_j u_{j-1}$ (voir [10]).

Afin de vérifier les hypothèses de la proposition 4.1.1, nous allons procéder à la décompositions LU . Dans le papier de S. Ikonnen et J. Toivannan ont combiné, sans justifier, l'algorithme de Brennan-Schwartz avec la décompositions LU (voir [11]) pour un problème de complémentarité linéaire issu de la discrétisation par différence finies pour l'option call. Dans ce qui suit, nous justifions cet l'algorithme, dans le cas des éléments finis pour l'option put.

Factorisation LU standard et UL

La factorisation standard LU de la matrice A est donné par $A = LU$, alors pour résoudre le système $Ax = b$ consiste simplement à résoudre successivement les deux systèmes triangulaires

$$Ly = b,$$

$$Ux = y.$$

Soit la matrice de permutation P avec des 1 sur l'anti-diagonal et des zéros partout ailleurs

$$P = \begin{pmatrix} & & 1 \\ & 1 & \\ 1 & & \end{pmatrix},$$

Alors $P^{-1} = P^T = P$, et soit $A_2 = PAP$

Exemple 4.1.1 Pour

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \text{ alors } A_2 = \begin{pmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{pmatrix},$$

Pour le système d'équations, cela implique

$$Ax = b \Leftrightarrow PAPPx = Pb \Leftrightarrow A_2Px = Pb$$

Maintenant, examinons la factorisation LU pour la matrice A_2

$$\begin{aligned} A_2 &= PAP = L_2 U_2, \\ A &= PA_2 P = (PL_2)(U_2 P) = U_3 L_3, \end{aligned}$$

D'où, en remplaçant $A = U_3 L_3$ dans le système d'équations $Ax = b$, nous mène à résoudre les deux systèmes

$$\begin{aligned} U_3 y &= b, \\ L_3 x &= y. \end{aligned}$$

Cette dernière est la factorisation UL . Nous allons adopter ce principe pour la matrice B . Rappelons que,

$$B = \begin{pmatrix} b & c & 0 & \cdots & 0 \\ a & b & c & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & a & b & c \\ 0 & \cdots & 0 & a & b \end{pmatrix},$$

avec

$$b = \frac{2}{3} r h k + \frac{4h}{6}, \quad (4.3)$$

$$c = -\frac{1}{2} \beta k + \frac{1}{6} r h k - \frac{1}{2} \sigma^2 k, \quad (4.4)$$

$$a = \frac{1}{2} \beta k + \frac{1}{6} r h k - \frac{1}{2} \sigma^2 k. \quad (4.5)$$

La décompositions UL de la matrice $\mathbf{B}$ pour l'option dite put.

$$\mathbf{B} = \mathbf{U}\mathbf{L} = \begin{pmatrix} 1 & c'_1 & 0 & \cdots & 0 \\ 0 & 1 & c'_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & 1 & c'_{m-1} \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b'_1 & 0 & 0 & \cdots & 0 \\ a'_1 & b'_2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 0 & 0 \\ \vdots & \ddots & a'_{m-1} & b'_{m-1} & 0 \\ 0 & \cdots & 0 & a'_m & b'_m \end{pmatrix}. \quad (4.6)$$

Pour déterminer les coefficients a'_j , b'_j et c'_m , nous multiplions les deux matrices $\mathbf{U}$ et $\mathbf{L}$. En effectuant un calcul rétrograde simple :

— Dans la dernière ligne donne des équations ci-dessus

$$b = 1 \cdot b'_m \text{ alors } b'_m = b,$$

$$a = 1 \cdot a'_m \text{ alors } a'_m = a.$$

— Dans l'avant dernière ligne donne des équations ci-dessus

$$c = c'_{m-1} b'_m \text{ alors } c'_{m-1} = \frac{c}{b'_m},$$

$$b = 1 \cdot b'_{m-1} + c'_{m-1} \cdot a'_m \text{ alors } b'_{m-1} = b - \frac{a \cdot c}{b'_{m-1}},$$

$$a = 1 \cdot a'_m \text{ alors } a'_m = a.$$

— Dans la $(j + 1)$ -ème ligne nous obtenons

$$c = c'_{m-j} b'_{m-j+1} \text{ alors } c'_{m-j} = \frac{c}{b'_{m-j+1}},$$

$$b = 1 \cdot b'_{m-j} + c'_{m-j} \cdot a'_{m-j+1} \text{ alors } b'_{m-j} = b - \frac{a \cdot c}{b'_{m-j+1}},$$

$$a = 1 \cdot a'_{m-j} \text{ alors } a'_{m-j} = a,$$

Concernant, le second membre définit par

$$\mathbf{d}^n = \mathbf{M}u_h^n - g \in \mathbb{R}^m,$$

on lui associe le vecteur suivant, pour l'option put, s'écrit :

$$\tilde{\mathbf{d}}_m^n = \mathbf{d}_m^n, \quad \tilde{\mathbf{d}}_i^n = \mathbf{d}_i^n - \tilde{\mathbf{d}}_{i+1}^n \frac{c}{b'_{i+1}},$$

qui solution de

$$\mathbf{U}\tilde{\mathbf{d}}^n = \mathbf{d}^n.$$

Maintenant, il est possible de décrire l'algorithme en deux boucles seulement.

Algorithm 1 Brennan-Schwartz pour l'option put

- Pour un vecteur donné d^n résoudre $\mathbf{U}y = d^n$
 - pour $i = 1$ à m , calculé
- $$u_i^{n+1} = \max\{(y_i - a'_i u_{i-1}^n)/b'_i, \psi_i\}$$
-

Remarque 4.1.1 Pour le cas de l'option call, nous utilisons la décompositions LU de la matrice $\mathbf{B}$

$$\mathbf{B} = \mathbf{L}\mathbf{U} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ a'_1 & 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 0 & 0 \\ \vdots & \ddots & a'_{m-1} & 1 & 0 \\ 0 & \cdots & 0 & a'_m & 1 \end{pmatrix} \begin{pmatrix} b'_1 & c'_1 & 0 & \cdots & 0 \\ 0 & b'_2 & c'_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & b'_{m-1} & c'_{m-1} \\ 0 & \cdots & 0 & 0 & b'_m \end{pmatrix}$$

De la même façon que précédemment le cas de put mais, le calcul sera progressive. Le second membre dans le l'option call, s'écrit :

$$\tilde{\mathbf{d}}_1^n = \mathbf{d}_1^n, \quad \tilde{\mathbf{d}}_i^n = \mathbf{d}_i^n - \tilde{\mathbf{d}}_{i-1}^n \frac{c}{b'_{i-1}}.$$

Il aussi est possible de décrire l'algorithme en deux boucles seulement.

Algorithm 2 Brennan-Schwartz pour l'option call

- Pour un vecteur donné d^n résoudre $\mathbf{U} y = d^n$
 - pour $i = m$ à 1, calculé

$$u_i^{n+1} = \max\{(y_i - a'_i u_{i+1}^n)/b'_i, \psi_i\}$$
-

Il nous reste à vérifier que $b'_j > 0$, dans les hypothèses de la proposition 4.1.1.

Lemme 4.1.1 Si $\mathbf{B}$ est une matrice coercive, les coefficients diagonaux de la matrice $\mathbf{L}$ sont strictement positifs.

Dans la preuve on se restreint au cas de l'option put.

Preuve. Du faite que,

$$\det \mathbf{B} = \det \mathbf{U} \det \mathbf{L},$$

et que

$$\det \mathbf{U} = 1, \quad \det \mathbf{L} = \prod_{j=1}^m b'_j,$$

alors

$$\det \mathbf{B} = \prod_{j=1}^m b'_j.$$

Puisque la matrice $\mathbf{B}$ est coercive alors les n mineurs principaux dominants soient strictement positifs, autrement dit que $\det \mathbf{B} > 0$, par récurrence sur m on déduit que $b'_j > 0$

pour $j = 1, \dots, m$. ■

La proposition suivante est une extension de la proposition 4.1.1, pour tenir en compte de la frontière libre pour l'option call où put.

Proposition 4.1.2 *Assurer que*

- $\mathbf{B}$ est une matrice tridiagonale et $c \leq 0$.
- La solution u_h^{n+1} de

$$(I) \left\{ \begin{array}{l} \mathbf{B} u_h^{n+1} \geq \mathbf{d}^n, \\ u_h^{n+1} \geq \psi_h, \\ (\mathbf{B} u_h^{n+1} - \mathbf{d}^n, \psi_h - u) = 0, \end{array} \right.$$

satisfait, $(u_h^{n+1})_j = (\psi_h)_j$, pour $j = 1, 2, 3, \dots, k$ et $(u_h^{n+1})_j > (\psi_h)_j$ pour $j = k + 1, \dots, m$ Alors u_h^{n+1} est aussi solution de $(\tilde{I})$

$$(\tilde{I}) \left\{ \begin{array}{l} \mathbf{L} u_h^{n+1} \geq \tilde{\mathbf{d}}^n, \\ u_h^{n+1} \geq \psi_h, \\ (\mathbf{L} u_h^{n+1} - \tilde{\mathbf{d}}^n, \psi_h - u_h^{n+1}) = 0. \end{array} \right.$$

Remarque 4.1.2 De la proposition précédente, la segmentation en sous ensemble détermine la frontière libre pour l'option put, c'est-à-dire,

$$u_h^{n+1} = (\psi_h)_j \text{ pour } j \leq k, \text{ et } \mathbf{B} u_h^{n+1} = \mathbf{d}^n \text{ pour } j > k.$$

Cette condition de segmentation, nous assure que

$$(\mathbf{B} u_h^{n+1} - \mathbf{d}^n, \psi_h - u) = 0.$$

Pour le cas de l'option call, la condition de segmentation

$$u_h^{n+1} = (\psi_h)_j \text{ pour } j > k, \text{ et } \mathbf{B} u_h^{n+1} = \mathbf{d}^n \text{ pour } j \leq k.$$

Preuve. Puisque $c \leq 0$, on aura $c'_j = \frac{c}{b'_{j-1}} \leq 0$. De la définition du second membre nous avons donc des constantes positives $\lambda_{j,j+1}, \dots, \lambda_{j,m}$, pour $j = 1, \dots, m-1$ tels que

$$(\tilde{\mathbf{d}}^n)_j = (\mathbf{d}^n)_j + \lambda_{j,j+1} (\mathbf{d}^n)_{j+1} + \dots + \lambda_{j,m} (\mathbf{d}^n)_m,$$

et

$$(\mathbf{L} u_h^{n+1})_j = (\mathbf{B} u_h^{n+1})_j + \lambda_{j,j+1} (\mathbf{B} u_h^{n+1})_{j+1} + \dots + \lambda_{j,m} (\mathbf{B} u_h^{n+1})_m,$$

par conséquent, $\mathbf{B} u_h^{n+1} \geq \mathbf{d}^n$, cela implique $\mathbf{L} u_h^{n+1} \geq \tilde{\mathbf{d}}^n$. Pour $j > k$, nous savons que $(u_h^{n+1})_j > (\psi_h)_j$, alors $(\mathbf{B} u_h^{n+1})_j = (\mathbf{d}^n)_j$. Cela implique $(\mathbf{L} u_h^{n+1})_j = (\tilde{\mathbf{d}}^n)_j$. ■

Pour assuré la solution de problème de complémentarité linéaire, il est nécessaire que la matrice soit coercive. Le lemme suivant nous assure, alors, la coercivité de $\mathbf{B}$.

Lemme 4.1.2 Si $h < \frac{3\sigma^2 k}{1+r k}$, alors la matrice $\mathbf{B}$ est coercive et pour, $x \in \mathbb{R}^m$

$$\mathbf{B} x \geq 0 \text{ et } x \neq 0 \Rightarrow x > 0.$$

Preuve. Nous remarquons que la matrice $\mathbf{B}$ n'est pas symétrique, mais nous pouvons travailler avec sa partie symétrique puisque

$$\mathbf{P} = \frac{1}{2} (\mathbf{B} + \mathbf{B}^T).$$

En utilisant,

$$2\langle x, \mathbf{P}x \rangle = \langle x, \mathbf{B}x \rangle + \langle x, \mathbf{B}^T x \rangle = \langle x, \mathbf{B}x \rangle + \langle \mathbf{B}x, x \rangle = 2\langle x, \mathbf{B}x \rangle.$$

Nous remarquons aussi que $\mathbf{B}$ est définie positive si $\mathbf{P}$ est définie positive. Maintenant, examinons $\mathbf{P}$

$$\mathbf{P} = \begin{pmatrix} b & -\hat{a} & & & \\ -\hat{a} & b & -\hat{a} & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -\hat{a} \\ & & & -\hat{a} & b \end{pmatrix},$$

avec $b > 0$. Alors

$$\begin{aligned} (\mathbf{P}x, x) &= -\hat{a} \sum_{i=2}^m x_{i-1} x_i + b \sum_{i=1}^m x_i^2 - \hat{a} \sum_{i=1}^{m-1} x_{i+1} x_i \\ &\geq -\frac{|\hat{a}|}{2} \sum_{i=2}^m (x_{i-1}^2 + x_i^2) + b \sum_{i=1}^m x_i^2 - \frac{|\hat{a}|}{2} \sum_{i=1}^{m-1} (x_{i+1}^2 + x_i^2) \\ &\geq (b - 2|\hat{a}|) \sum_{i=1}^m x_i^2 = (b - 2|\hat{a}|) \|x\|^2. \end{aligned}$$

Dans notre cas, nous avons

$$\begin{aligned} \hat{a} &= \frac{-h}{6} + k \left(\frac{\sigma^2}{2h} - \frac{rh}{6} \right), \\ b &= \frac{4h}{6} + k \left(\frac{\sigma^2}{h} + \frac{2}{3} rh \right). \end{aligned}$$

Pour que $\hat{a} > 0$, soit vérifié il faut

$$\begin{aligned} 0 &< \frac{-h}{6} + k \left(\frac{\sigma^2}{2h} - \frac{r h}{6} \right) \\ 0 &< -h^2 + k (3\sigma^2 - r h^2) \\ h^2 + r h^2 k &< 3\sigma^2 k \\ h^2 &< \frac{3\sigma^2 k}{1 + r k}. \end{aligned}$$

Avec $b - 2\hat{a} = h + r \Delta t h$ et $r > 0$, nous obtenons

$$(\mathbf{P}x, x) \geq (b - 2\hat{a}) \|x\|^2 = h(1 + r \Delta t) \|x\|^2.$$

Assumons que $\mathbf{B}x = \mathbf{d}^n \geq 0$. Utilisons $\mathbf{B} = \mathbf{U} \cdot \mathbf{L}$ et déterminer y par $\mathbf{U}\tilde{\mathbf{d}}^n = \tilde{\mathbf{d}}^n$. Puisque $x \neq 0$ on conclue $\mathbf{d}^n \neq 0$ et $\tilde{\mathbf{d}}^n \neq 0$. Nous utilisons la preuve de la proposition 4.1.2 pour conclure $\tilde{\mathbf{d}}^n \geq 0$ et $(\tilde{\mathbf{d}}^n)_1 > 0$. Puisque, $\mathbf{L}x = \tilde{\mathbf{d}}^n \geq 0$, nous avons

$$\mathbf{L}\vec{x} = \begin{pmatrix} b_1 & 0 & 0 & \cdots & 0 \\ a_2 & b_2 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \cdots & a_{m-1} & b_{m-1} & 0 \\ 0 & \cdots & 0 & a_m & b_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{pmatrix} = \tilde{\mathbf{d}}^n.$$

Utilisant $b_i > 0$ et $a_i < 0$ nous obtenons $x_i > 0$. ■

Remarque 4.1.3 Dans la condition $h < \frac{3\sigma^2 k}{1 + r k}$, le pas spatiale h doit être suffisamment petite, Selon le pas du temps k . Dans [10], la condition dans le cas différence finis est

$$\beta = r - q - \frac{\sigma^2}{2} < \frac{\sigma^2}{h}, \text{ ou } h < \frac{\sigma^2}{\beta}.$$

Observez que la condition sur h dans [10] est indépendante de pas de temps k .

La difficulté, en pratique, réside dans la vérification de l'hypothèse de segmentation comme le montre cet exemple suivant introduit par [10].

Exemple 4.1.2 Soit $m = 3$ et $\mathbf{B} = \begin{pmatrix} 1 & -1 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, avec $\mathbf{d} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ et $\psi = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$.

Alors, la solution du système $(\tilde{I})$, selon la proposition 4.1.2, est donné par : $u = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$, nous remarquons la condition de segmentation est vérifié $u_1 = \psi_1, u_2 = \psi_2$ et $u_3 > \psi_3$ mais elle n'est pas solution (I) .

Le lemme suivant nous assure la justification de l'hypothèse de segmentation pour la proposition 4.1.2.

Lemme 4.1.3 Soit $\psi(x) = \max\{E - E \exp(x), 0\}$, l'ensemble

$$\Gamma(\psi) = \left\{ j \in \mathbb{Z}, \psi(jh) > 0 \text{ et } (\mathbf{B}\psi)_j \geq (\mathbf{M}\psi)_j \right\},$$

vérifié cette propriété,

$$j_0 \in \Gamma(\psi) \implies \forall j \leq j_0, j \in \Gamma(\psi).$$

Remarque 4.1.4

— L'ensemble donné dans les travaux de P. Jaillet D. Lamberton et B. Lapeyre (voir [10, 23]) est

$$\Gamma(\psi) = \left\{ j \in \mathbb{Z}, \psi(jh) > 0 \text{ et } \tilde{a}\psi((j-1)h) + \tilde{b}\psi(jh) + \tilde{c}\psi((j+1)h) \geq \psi(jh) \right\},$$

avec $\tilde{a}$, $\tilde{b}$ et $\tilde{c}$ les coefficients de la matrice de discrétisation $\mathcal{A}_h$ par différence finies en 3.31, 3.32 et 3.33.

— Pour l'option call, la propriété que l'ensemble $\Gamma(\psi)$ est

$$j_0 \in \Gamma(\psi) \implies \forall j \geq j_0, j \in \Gamma(\psi).$$

Preuve. Utilisons

$$\begin{aligned} \exp((j+1)h) - \exp((j-1)h) &= \exp(jh) 2 \sinh(h), \\ \exp((j+1)h) + \exp((j-1)h) &= \exp(jh) 2 \cosh(h), \\ -\exp((j+1)h) + 2\exp(jh) - \exp((j-1)h) &= \exp(jh) (2 - 2 \cosh(h)). \end{aligned}$$

Sachant que de (4.3) – (4.5), cela implique

$$\begin{aligned} f_1 &= a \exp((j-1)h) + b \exp(jh) + c \exp((j+1)h) + \\ &\quad - \frac{h}{6} \exp((j-1)h) - \frac{4h}{6} \exp(jh) - \frac{h}{6} \exp((j+1)h) \\ &= \left(a - \frac{h}{6}\right) \exp((j-1)h) + \left(b - \frac{4h}{6}\right) \exp(jh) + \left(c - \frac{h}{6}\right) \exp((j+1)h) \\ &= k \exp(jh) \frac{h}{6} r (2 \cosh(h) + 2) + \frac{\sigma^2 k}{2h^2} \exp(jh) (2 - 2 \cosh(h)) + \\ &\quad - \frac{\beta}{2} k \sinh(h) 2 \exp(jh) \\ &= k \exp(jh) \left[\frac{\sigma^2}{h^2} (1 - \cosh(h)) - \beta \sinh(h) + \frac{h}{3} r (\cosh(h) + 1) \right] \\ f_2 &= a E + b E + c E - \frac{h}{6} E - \frac{4h}{6} E - \frac{h}{6} E \\ &= E r h k \\ f(j) &= f_2 - E f_1 = k E \left(r h + \exp(jh) \left[\frac{\sigma^2}{h^2} (\cosh(h) - 1) + \beta \sinh(h) - \frac{h}{3} r (\cosh(h) + 1) \right] \right). \end{aligned}$$

Vue que $f(j) = f_2 - E f_1$ est une fonction monotone de j de limite $E r h k > 0$ lorsque, $j \rightarrow -\infty$. Avec $\psi(x) = \max(E - E \exp(x), 0)$ et $\bar{\psi}(x) = E - E \exp(x)$. De ce qui précède,

— Si $j_0 \in \Gamma(\psi)$ alors $\psi(j_0 h) > 0$ et puisque $\psi(x)$ décroissant, alors

$$j_0 \in \Gamma(\bar{\psi}) \implies \forall j \leq j_0, j \in \Gamma(\bar{\psi}).$$

— Si $j_0 \in \Gamma(\psi)$ et $\psi((j_0 + 1)h) \geq 0$, la monotonie de $f(j)$, implique $j \in \Gamma(\psi), \forall j \leq j_0$.

— Si $j_0 \in \Gamma(\psi)$ et $\psi((j_0 + 1)h) = (E - E \exp((j_0 + 1)h)) + R = 0$, pour un certain $R > 0$, en utilisant $c \leq 0$ et

$$0 \leq \left(a - \frac{h}{6}\right) \psi_{j_0-1} + \left(b - \frac{4h}{6}\right) \psi_{j_0} + \left(c - \frac{h}{6}\right) \psi_{j_0+1} = f(j_0) + \left(c - \frac{h}{6}\right) R,$$

et alors $f(j_0) > 0$. Maintenant, la monotonie de $f(j)$ implique $j \in \Gamma(\psi), \forall j \leq j_0$.

■

En utilisant lemme 4.1.3 et la proposition 4.1.3 suivante, nous arrivons à vérifier l'hypothèse de la segmentation de la proposition 4.1.2, ce qui nous assure la validité de l'algorithme de Brennan-Schwartz pour l'approche par élément finis.

Proposition 4.1.3 *Les vecteurs $u^0, u^1, u^2, \dots, u^N$, suite de solutions du problème complémentarité linéaire (3.28) vérifient les propriétés suivantes :*

- $\forall n \in \{1, \dots, N\} \ u^n \geq u^{n-1}$
- Pour chaque entier n , il existe un indice k_i tel que

$$u_j^n = \psi_j \quad \text{si } 1 \leq j \leq k_i \quad \text{et } u_j^n > \psi_j \quad \text{si } j > k_i$$

Preuve.

- La première propriété se démontre par récurrence sur n . Il est clair que l'on a, $u^1 \geq u^0 = \psi$ car

$$\mathbf{B}u^1 \geq \mathbf{d}^0, \quad (4.7)$$

$$u^1 \geq \psi = u^0. \quad (4.8)$$

Supposons que, l'inégalité $u^n \geq u^{n-1}$ est acquise. Alors u^{n+1} vérifié :

$$\mathbf{B}u^{n+1} \geq \mathbf{d}^{n-1}.$$

Vue que, u^n solution de

$$\mathbf{B}u^n \geq \mathbf{d}^{n-1}.$$

Et utilisant la définition 1.3.4 et la proposition 1.3.2 : $u^{n+1} \geq u^n$.

- Pour la deuxième assertion, on traite d'abord le cas : $n = 1$. Puisque ψ est non nul, on a, par le lemme 4.1.2, $u_j^1 \geq 0$ pour tout $j \in \{1, \dots, m\}$. Le raisonnement par absurde, en supposant qu'il existe, un intervalle d'entiers $[l_1, l_2]$, avec $1 \leq l_1 \leq l_2 \leq m$, tel que :

$$u_{l_2}^1 = \psi_{l_2} \text{ et } u_j^1 > \psi_j \text{ pour } j \in [l_1, l_2[.$$

Quitte à diminuer l_1 , on peut supposer que, $l_1 = 1$, ou

$$u_{l_1-1}^1 = \psi_{l_1-1}. \quad (4.9)$$

De l'inégalité, $(\mathbf{B}u^1)_{l_2} \geq (\mathbf{d}^0)_{l_2} = (\mathbf{M}u^0)_{l_2} = (\mathbf{M}\psi)_{l_2}$ et de l'égalité $u_{l_2}^1 = \psi_{l_2}$, on déduit :

$$au_{l_2-1}^1 + b\psi_{l_2} + cu_{l_2+1}^1 \geq \frac{h}{6}\psi_{l_2-1} + \frac{4h}{6}\psi_{l_2} + \frac{h}{6}\psi_{l_2+1},$$

et puisque, les constantes a et c sont négatives :

$$a\psi_{l_2-1} + b\psi_{l_2} + c\psi_{l_2+1} \geq \frac{h}{6}\psi_{l_2-1} + \frac{4h}{6}\psi_{l_2} + \frac{h}{6}\psi_{l_2+1}. \quad (4.10)$$

Il en résulte que, $\lambda_2 \in \Gamma(\psi)$, ce qui entraîne par le lemme 4.1.3 $[1, l_2] \subset \Gamma(\psi)$ et, par conséquent :

$$(\mathbf{B}\psi)_j \geq (\mathbf{M}\psi)_j \text{ pour } j \in [1, l_2].$$

Comme pour $j \in [l_1, l_2[$ on a : $(\mathbf{B}u^1)_j = (\mathbf{M}\psi)_j$, on voit que le vecteur vérifié $(\mathbf{B}\psi)_j \geq (\mathbf{B}u^1)_j$ pour $j \in [l_1, l_2[$. Le vecteur $v = \psi - u^1$ vérifié donc $(\mathbf{B}v)_j \geq 0$ pour $j \in [l_1, l_2[$, avec, de plus $v_{\lambda_2} = 0$ et, si $l_1 > 1$, $v_{l_1-1} = 0$. On déduit, en appliquant le lemme 4.1.2 à une matrice extraite de $\mathbf{B}$, que l'on a : $v_j \geq 0$, pour $j \in [l_1, l_2[$. D'où : $\psi_j \geq u_j^1$ pour $j \in [l_1, l_2[$, ce qui est contradictoire. Si maintenant on suppose que la propriété est vraie à l'ordre n , et si $[l_1, l_2]$ est intervalle tel que :

$$u_{l_2}^{n+1} = \psi_{l_2} \text{ et } u_j^{n+1} > \psi_j \text{ pour } j \in [l_1, l_2[,$$

alors, on a,

$$u_{l_2}^n = \psi_{l_2},$$

et par conséquent : $u_j^n = \psi_j$ pour $j \leq l_2$ et

$$au_{l_2-1}^{n+1} + b\psi_{l_2} + cu_{l_2+1}^{n+1} \geq \frac{h}{6}\psi_{l_2-1} + \frac{4h}{6}\psi_{l_2} + \frac{h}{6}\psi_{l_2+1},$$

et puisque a et c sont négatives :

$$a\psi_{l_2-1} + b\psi_{l_2} + c\psi_{l_2+1} \geq \frac{h}{6}\psi_{l_2-1} + \frac{4h}{6}\psi_{l_2} + \frac{h}{6}\psi_{l_2+1}$$

Il en résulte que $l_2 \in \Gamma(\psi)$, ce qui entraîne par le lemme 4.1.3 $[1, l_2] \subset \Gamma(\psi)$ et, par conséquent de :

$$(\mathbf{B}\psi)_j \geq (\mathbf{M}\psi)_j \text{ pour } j \in [1, l_2].$$

Comme pour $j \in [l_1, l_2[$ on a : $(\mathbf{B}u^n)_j = (\mathbf{M}\psi)_j$, on voit que le vecteur vérifie $(\mathbf{B}\psi)_j \geq (\mathbf{B}u^{n+1})_j$ pour $j \in [l_1, l_2[$. Le vecteur $v = \psi - u^{n+1}$ vérifie donc $(\mathbf{B}v)_j \geq 0$ pour $j \in [l_1, l_2[$, avec, de plus $v_{l_2} = 0$ et, si $l_1 > 1$, $v_{l_1-1} = 0$. On déduit, en appliquant le lemme 10 à une matrice extraite de $\mathbf{B}$, que l'on a : $v_j \geq 0$, pour $j \in [l_1, l_2[$. D'où : $\psi_j \geq u_j^{n+1}$ pour $j \in [l_1, l_2[$, ce qui est contradictoire.

■

4.2 La méthode PSOR

L'algorithme de relaxation, PSOR "Projected Successive Over-Relaxation" est un procédé itératif qui, introduit par [37] et l'extension de méthode de relaxation pour la résolution des systèmes linéaires sous la forme :

$$\mathbf{A}v = \mathbf{f},$$

où $\mathbf{A} \in \mathbb{R}^{m \times m}$, $v \in \mathbb{R}^m$ et $\mathbf{f} \in \mathbb{R}^m$.

La méthode de relaxation est basée sur la décomposition

$$\mathbf{A} = \mathbf{D} - \mathbf{L} - \mathbf{U},$$

où $\mathbf{D}$, la matrice diagonale, $\mathbf{L}$, matrice triangulaire inférieure stricte et $\mathbf{U}$, matrice triangulaire supérieure, alors v est solution de $\mathbf{A}v = \mathbf{f}$, si et seulement, si v vérifie

$$(\mathbf{D}/\omega - \mathbf{L})v = ((1/\omega - 1)\mathbf{D} + \mathbf{U})v + \mathbf{f}.$$

Pour des valeurs de ω dans l'intervalle $]0, 2[$, on introduit la méthode de sur-relaxation successive (ou méthode SOR pour successive over relaxation) (voir [38, 39])

$$\frac{1}{\omega}a_{ii}x_i^{n+1} + \sum_{j < i} a_{ij}v_j^{n+1} = f_i + \left(\frac{1}{\omega} - 1\right)a_{ii}v_i^n - \sum_{j > i} a_{ij}v_j^{n+1}, \quad (4.11)$$

pour tout, $i \in \{1, \dots, m\}$. On peut écrire (4.11), sous la forme vectorielle :

$$(\mathbf{I} - \omega\mathbf{D}^{-1}\mathbf{L})v^{n+1} = [(1 - \omega)\mathbf{I} + \omega\mathbf{D}^{-1}\mathbf{U}]v^n + \omega\mathbf{D}^{-1}\mathbf{f},$$

d'où on déduit la matrice d'itération

$$\mathbf{G}_\omega = (\mathbf{I} - \omega\mathbf{D}^{-1}\mathbf{L})^{-1} [(1 - \omega)\mathbf{I} + \omega\mathbf{D}^{-1}\mathbf{U}].$$

La méthode SOR converge si ([38])

$$\rho(\mathbf{G}_\omega) < 1, \quad \text{pour } \omega \in]0, 2[,$$

avec $\rho(\mathbf{G}_\omega)$ est le rayon spectral. Du faite que, le paramètre ω influence sur la vitesse de convergence (voir [38]), le pas optimal est donné par cet relation :

$$\omega_{opt} = \frac{2}{1 + \sqrt{1 - \rho(\mathbf{G}_{Jacobi})^2}},$$

avec $\mathbf{G}_{Jacobi} = \mathbf{D}^{-1}(\mathbf{A} - \mathbf{D})$ la matrice d'itération de Jacobi (voir [38]).

En utilisant le théorème de Gerschgorin (voir [40]), On peut a priori borner les valeurs

propres par

$$\rho(\mathbf{G}_{Jacobi}) \approx \max_i \frac{1}{A_{ii}} \sum_{j \neq i} |A_{ij}|.$$

Nous intéressons maintenant, à appliquer la méthode PSOR, au problème de complémentarité linéaire suivant

$$\begin{aligned} \mathbf{B} u_h^{n+1} &\geq d^n, \\ u_h^{n+1} &\geq \psi_h, \\ (\mathbf{B} u_h^{n+1} - d^n)(u_h^{n+1} - \psi_h) &= 0, \end{aligned} \tag{4.12}$$

pour $n = 0, \dots, N$.

Le problème 4.12, est équivalent à (voir [41])

$$\min \left\{ \mathbf{B} u_h^{n+1} - d^n, u_h^{n+1} - \psi_h \right\} = 0,$$

pour, $n = 0, \dots, N$.

Posons, $\mathbf{A} = \mathbf{B}$, $v = u_h^{n+1}$ et $f = d^n$, on aura (voir [36])

$$\min_v \left\{ v - \mathbf{A}^{-1} f, v - \psi_h \right\} = 0.$$

Où encore

$$v = \max \left\{ \mathbf{A}^{-1} f, \psi_h \right\}.$$

Ce problème peut être résolu par PSOR, ce dernier a été suggéré par Cryer (voir[37]) pour une discussion détaillée sur la méthode de Cryer voir [14]. Le principe est d'appliquer à la contrainte de maximisation (4.2) l'algorithme SOR. Alors

$$\forall i \in \{1, \dots, m\}, v_i = \max \{y_i, (\psi_h)_i\},$$

et les y_i sont définis par

$$\frac{1}{\omega} a_{ii} y_i^{n+1} + \sum_{j < i} a_{ij} y_j^{n+1} = f_i + \left(\frac{1}{\omega} - 1 \right) a_{ii} y_i^n - \sum_{j > i} a_{ij} y_j^{n+1}.$$

L'algorithme de PSOR appliqué à 4.12, pour une itération du temps, est décrit par

Algorithm 3 PSOR

for $j = 1$ to m do

$$r_j = d_j - \sum_{i=1}^m B_{ji}u_i$$

$$u_j = u_j + \omega \frac{r_j}{B_{jj}}$$

$$u_j = \max[u_j, \psi_j],$$

end for
