Chapitre 2

Implicitation de la méthode de Relaxation

L'objectif de ce chapitre est de proposer une implicitation en temps de la méthode de Relaxation décrite dans le chapitre précédent pour approcher les solutions stationnaires du système des équations d'Euler à plusieurs pressions indépendantes (1.5).

Motivés par la mise en place d'une technique d'implicitation robuste et peu coûteuse, nous proposons de tirer profit du découplage naturel, obtenu par construction, des entropies du système (1.87) lorsque $\lambda = 0$. Par découplage, nous entendons qu'une fois le choix de la vitesse a effectué conformément aux conditions de Whitham à chaque pas de temps, l'évolution des variables ρ, ρu, $\rho\Sigma$ et ρI est totalement indépendante de l'évolution de chacune des entropies spécifiques. Ce découplage résulte de la procédure de Relaxation sur la pression totale. Il ne vaut évidemment pas pour le système à l'équilibre (1.5) puisque les pressions partielles p_i dépendent intrinsèquement des entropies S_i. Il apparait donc naturel de vouloir impliciter en temps de manière découplée le bloc transport des entropies :

$$\{\partial_t \rho S_i + \partial_x \rho S_i u = 0, \quad i = 1, ..., N-1, \tag{2.1}$$

et le bloc hydrodynamique :

$$\left\{ \begin{array}{l} \partial_t \rho + \partial_x \rho u = 0, \\ \partial_t \rho u + \partial_x (\rho u^2 + \Pi) = 0, \\ \partial_t \rho\Sigma + \partial_x (\rho\Sigma + \Pi) u = 0, \\ \partial_t \rho I + \partial_x \rho I u = 0, \quad I = \Pi + a^2/\rho. \end{array} \right. \tag{2.2}$$

Nous renvoyons le lecteur à la fin de la section 2.1.2 pour une discussion précise sur la détermination de la valeur du paramètre a.

La technique que nous développons ci-dessous nous permettra d'assurer la validité d'un principe du maximum pour chacune des entropies spécifiques S_i, et ce sans restriction CFL. Elle reposera sur la résolution de $(N-1)$ systèmes linéaires à structure tridiagonale scalaire en une dimension d'espace (pentadiagonale en deux dimensions d'espace). La matrice à inverser sera toujours la même. La principale difficulté concernera l'implicitation en temps du bloc hydrodynamique. Une implicitation "naturelle" de ces équations, basée sur la réinterprétation du flux numérique de la méthode de

Godunov dans le formalisme des flux numériques associés à une méthode de Roe, sera dans un premier temps proposée. La robustesse de l'algorithme sera illustrée numériquement. Néammoins, une analyse plus fine établira que cette approche "naturelle" n'autorise pas l'obtention de solutions rigoureusement convergées en temps. La mise en évidence des raisons de ce défaut de convergence permettra alors de déduire une technique de correction simple de cette implicitation "naturelle". L'expérience numérique menée pour des problèmes en deux dimensions d'espace montrera que l'algorithme obtenu est stable, et qu'il permet le calcul de solutions parfaitement stationnaires du système Euler multi-pression. Soulignons que l'algorithme proposé présente un coût d'évaluation très avantageux par rapport à une approche classique reposant sur une implicitation du système Euler multi-pression complet.

De manière à débuter la description de la procédure, nous supposons connue à l'instant t^n une solution discrète à l'équilibre

$$\mathbf{u}_j^n = (\rho_j^n, (\rho u)_j^n, (\rho E)_j^n, \{(\rho S_i)_j^n\}_{i=1,\dots,N-1}), \quad j \in Z,$$

de vecteurs de Θ et nous définissons la solution relaxée associée :

$$\overline{\mathbf{u}}_j^n = (\rho_j^n, (\rho u)_j^n, (\rho E)_j^n, \{(\rho S_i)_j^n\}_{i=1,\dots,N-1}, (\rho I)_j^n), \quad j \in Z,$$

où

$$(\rho I)_j^n = \rho_j^n \sum_{i=1}^{N} p_i(1/\rho_j^n, (\rho S_i)_j^n/\rho_j^n) + a^2, \quad j \in Z.$$

Il s'agit de proposer une mise à jour implicite en temps de cette solution.

2.1 Implicitation "naturelle" de la méthode de Relaxation

Rappelons brièvement la stratégie de remise à jour de la solution discrète dans le cadre d'une formulation explicite en temps. Celle-ci est composée de deux pas, respectivement un pas d'évolution en temps où le paramètre de Relaxation λ est choisi nul (*i.e.* en l'absence du terme de Relaxation) suivi immédiatement d'une projection sur la variété d'équilibre (*i.e.* avec formellement $\lambda = +\infty$). Cette stratégie donne entière satisfaction pour une formulation explicite en temps. Il semble dès lors naturel d'adopter une démarche identique dans l'établissement d'une formulation implicite en temps : une implicitation en temps de l'étape d'évolution avec $\lambda = 0$ conduisant à des valeurs réactualisées de ρ, ρu, ρE, $\{\rho S_i\}_{i=1,\dots,N}$, ρI à une date t donnée, puis de manière à redémarrer l'algorithme, un écrasement à cette même date de la valeur de ρI en posant :

$$\rho I = \rho \sum_{i=1}^{N} p_i(1/\rho, \rho S_i/\rho) + a^2.$$

Ici, il faut évidemment définir à cette date t une nouvelle valeur de a suffisamment grande conformément aux conditions de Whitham, fonctions de la solution à l'équilibre réactualisée à cet instant. Nous renvoyons le lecteur à la fin de la sous-section 2.1.2 pour une brève discussion sur le choix de a. L'objectif de cette section est de proposer une telle implicitation avec le souci de découpler les blocs hydrodynamique et transport des entropies. L'expérience numérique montrera que l'algorithme implicite du bloc hydrodynamique partage des propriétés de robustesse comparable à celle

d'une méthode de Van Leer dans la résolution des équations d'Euler 3×3. Toutefois, nous montrerons que l'implicitation proposée ne peut pas conduire à une solution parfaitement convergée en temps. L'analyse de ce défaut permettra alors de corriger simplement cette première technique d'implicitation pour obtenir un algorithme partageant la même stabilité avec le bénéfice de converger vers une solution rigoureusement stationnaire.
Nous commençons par détailler l'implicitation de l'étape d'évolution en temps puis concluons avec la mise en évidence du défaut de convergence.

2.1.1 Implicitation du bloc transport des entropies

Motivations et choix algorithmiques

Notre objectif est de proposer un algorithme inconditionnellement stable, *i.e.* sans restriction CFL, assurant un principe du maximum pour chacune des entropies spécifiques. En conséquence de cela, nous serons assurés de la positivité de chacune de ces entropies, ainsi que de la préservation des constantes. Rappelons qu'au niveau continu, la validité d'un principe du maximum repose sur l'utilisation conjointe de la densité ρ et de l'entropie ρS_i (ici i est fixé) :

$$\partial_t \rho + \partial_x \rho u = 0, \tag{2.3}$$

$$\partial_t \rho S_i + \partial_x \rho u S_i = 0. \tag{2.4}$$

pour obtenir classiquement l'équation de transport suivante (et donc le principe du maximum recherché) :

$$\partial_t S_i + u \partial_x S_i = 0.$$

Remarquons que cette dernière identité vaut également au sens des distributions pour les solutions Lipschitz-continues par morceaux du système considéré, puisque la vitesse u et l'entropie spécifique S_i ne sont jamais discontinues simultanément. Il n'y a donc pas d'ambiguïté dans le produit non conservatif $u\partial_x S_i$.
Envisageons maintenant la validité des manipulations précédentes dans un contexte discret en temps et en espace. A cette fin, adoptons formellement dans une première étape une écriture conservative totalement implicite en temps pour considérer les analogues discrets de (2.3) et (2.4) suivants :

$$\rho_j^{n+1} = \rho_j^n - \frac{\Delta t}{\Delta x}((\rho u)_{j+1/2}^{n+1} - (\rho u)_{j-1/2}^{n+1}), \tag{2.5}$$

$$(\rho S_i)_j^{n+1} = (\rho S_i)_j^n - \frac{\Delta t}{\Delta x}((\rho u S_i)_{j+1/2}^{n+1} - (\rho u S_i)_{j-1/2}^{n+1}), \tag{2.6}$$

où conformément à la méthode de Relaxation proposée, la formule de Larrouturou [81] a lieu :

$$(\rho u S_i)_{j+1/2}^{n+1} = (\rho u)_{j+1/2,+}^{n+1} (S_i)_j^{n+1} + (\rho u)_{j+1/2,-}^{n+1} (S_i)_{j+1}^{n+1}, \tag{2.7}$$

avec la notation générique :

$$X_- = \min(X, 0), \quad X_+ = \max(X, 0), \quad \forall\, X \in \mathbb{R}. \tag{2.8}$$

Des manipulations algébriques simples conduisent à déduire de (2.5) et (2.6), avec des notations claires :

$$\begin{aligned} \rho_j^n (S_i)_j^{n+1} &= \rho_j^n (S_i)_j^n - \tfrac{\Delta t}{\Delta x} (\rho u)_{j+1/2}^{n+1} ((S_i)_{j+1/2}^{n+1} - (S_i)_j^{n+1}) \\ &- \tfrac{\Delta t}{\Delta x} (\rho u)_{j-1/2}^{n+1} ((S_i)_j^{n+1} - (S_i)_{j-1/2}^{n+1}), \end{aligned} \tag{2.9}$$

soit en vertu de (2.7), la formule suivante :

$$\begin{aligned} (S_i)_j^{n+1} &= (S_i)_j^n - \tfrac{\Delta t}{\Delta x} \tfrac{(\rho u)_{j+1/2,-}^{n+1}}{\rho_j^n} ((S_i)_{j+1}^{n+1} - (S_i)_j^{n+1}) \\ &- \tfrac{\Delta t}{\Delta x} \tfrac{(\rho u)_{j-1/2,+}^{n+1}}{\rho_j^n} ((S_i)_j^{n+1} - (S_i)_{j-1}^{n+1}). \end{aligned} \tag{2.10}$$

Nous montrerons plus loin (voir théorème 8) que sous une forme algébriquement identique, un principe du maximum a nécessairement lieu sans restriction CFL. La validité de (2.10) repose clairement sur l'identité $(\rho S_i)_j^{n+1} = \rho_j^{n+1} \times (S_i)_j^{n+1}$ où les formules de mises à jour de ρ_j^{n+1} et $(\rho S_i)_j^{n+1}$ sont liées par la formule de Larrouturou.
Dans la pratique, il est d'une part illusoire pour des raisons de coût calcul d'envisager une mise à jour de la densité ρ_j^{n+1} répondant à la formule complètement implicite (2.5). Une technique de linéarisation en temps classique lui sera préférée. D'autre part, nos efforts portent sur un découplage du bloc hydrodynamique et du bloc transport des entropies. Il semble difficile dans ces conditions d'opérer une linéarisation en temps des formules (2.5) et (2.6) permettant de garantir une formule algébriquement équivalente à la relation (2.10). Dans les faits, il s'avère difficile de proposer un algorithme assurant la constance éventuelle de l'entropie S_i, ce qui traduit une violation du principe du maximum.
Dans ces conditions, nous proposons d'opérer un découplage complet entre les blocs considérés en résolvant les équations de transport (*i.e.* en l'absence de ρ) :

$$\{\partial_t S_i + u \partial_x S_i = 0, \quad i = 1, ..., N-1. \tag{2.11}$$

La nature non conservative (et non ambiguë) de ces équations n'est pas rédhibitoire pour les applications que nous envisageons ici. Nous renvoyons en particulier le lecteur aux travaux de De Vuyst [44] qui illustrent l'intérêt de cette formulation dans d'autres applications. Nous proposons ci-après un algorithme conduisant naturellement à la satisfaction d'un principe du maximum sans restriction CFL, et possédant la propriété de n'avoir à inverser qu'un seul système linéaire tridiagonal scalaire pour toutes les entropies S_i, $i = 1, ..., N-1$.

Définition de l'algorithme

De manière à motiver l'introduction de l'algorithme implicite en temps, rappelons tout d'abord la discrétisation explicite des équations de transport (2.11) résultant d'une approche décentrée. Un champ de vitesse noté $\{u^*\}_{j+1/2}^n$ étant supposé donné à chaque interface $x_{j+1/2}$, la méthode de Godunov appliquée au système (2.11) s'écrit ici :

$$(S_i)_j^{n+1} = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} \{S_i\}(\overline{\mathbf{u}}_{\Delta x}(x, t^{n+1-})) dx, \quad i = 1, ..., N-1, \quad j \in Z.$$

En intégrant (2.11) sous une condition CFL usuelle (≤ 1) et sur chaque pavé

$$(x_{j-1/2}, x_{j+1/2}) \times (t^n, t^{n+1}),$$

nous obtenons de manière classique la mise à jour décentrée et explicite suivante :

$$\begin{aligned} (S_i)_j^{n+1} &= (S_i)_j^n - \tfrac{\Delta t}{\Delta x}\times \\ (\{u^*\}_{j-1/2,+}^n\{(S_i)_j^n - (S_i)_{j-1}^n\} &+ \{u^*\}_{j+1/2,-}^n\{(S_i)_{j+1}^n - (S_i)_j^n\}). \end{aligned} \tag{2.12}$$

Nous proposons une implicitation linéarisée de (2.12) en figeant à l'instant t^n les vitesses moyennes $\{u^*\}_{j+1/2}^n$. Nous obtenons pour tout $i = 1, ..., N-1$ et pour tout $j \in Z$ la mise à jour implicite suivante :

$$\begin{aligned} (S_i)_j^{n+1} &= (S_i)_j^n - \tfrac{\Delta t}{\Delta x}\times \\ (\{u^*\}_{j-1/2,+}^n\{(S_i)_j^{n+1} - (S_i)_{j-1}^{n+1}\} &+ \{u^*\}_{j+1/2,-}^n\{(S_i)_{j+1}^{n+1} - (S_i)_j^{n+1}\}), \end{aligned} \tag{2.13}$$

permettant d'évaluer les quantités $\{(S_i)_j^{n+1}\}_{j\in Z}$. La mise en œuvre numérique correspondante consiste en la résolution de $(N-1)$ systèmes linéaires (un par entropie) dont la matrice globale notée A possède une structure tridiagonale à coefficients scalaires. De manière plus précise et d'après (2.13), une ligne courante j de la matrice A est constituée des éléments a_j, b_j et c_j vérifiant

$$\begin{aligned} a_j &= -\tfrac{\Delta t}{\Delta x}\{u^*\}_{j-1/2,+}^n, \\ b_j &= 1 + \tfrac{\Delta t}{\Delta x}(\{u^*\}_{j-1/2,+}^n - \{u^*\}_{j+1/2,-}^n), \\ c_j &= \tfrac{\Delta t}{\Delta x}\{u^*\}_{j+1/2,-}^n. \end{aligned} \tag{2.14}$$

La matrice A, identique pour chaque entropie S_i, est clairement une M-matrice à diagonale strictement dominante (par M-matrice nous entendons classiquement une matrice à diagonale positive et à termes extra-diagonaux négatifs ou nuls). Sous des conditions aux limites *adhoc* (non précisées ici) le système linéaire de taille finie résultant est donc inversible, et ce indépendamment du choix du pas de temps. Ceci assure l'existence et l'unicité des $(S_i)_j^{n+1}$. Une analyse plus fine des propriétés de cette matrice permet d'établir la stabilité de la méthode en exhibant les principes du maximum recherchés :

Théorème 8
La mise à jour implicite (2.13) des inconnues $\{(S_i)_j^{n+1}\}_{j\in Z}$ pour tout $i = 1, ..., N-1$ admet une unique solution vérifiant les principes du maximum suivants :

$$\begin{aligned} \min\nolimits_{k\in Z}(S_i)_k^n \leq (S_i)_j^{n+1} &\leq \max\nolimits_{k\in Z}(S_i)_k^n, \\ i = 1, ..., N-1, &\quad j \in Z. \end{aligned} \tag{2.15}$$

Démonstration
Définissons pour tout $n \in N$ et pour tout $i = 1, ..., N-1$ le vecteur $S_i^n = ((S_i)_j^n)_{j\in Z}$.

La mise à jour implicite (2.13) se réexprime sous la forme $AS_i^{n+1} = S_i^n$. Les coefficients de la matrice tridiagonale A étant donnés par (2.14), il est clair que le vecteur $v = (v_j)_{j\in Z}$ défini par $v_j = 1$ pour tout $j \in Z$ est un vecteur propre de la matrice A associé à la valeur propre 1.
Considérons maintenant les vecteurs $m_i = (m_{i,j})_{j\in Z}$ définis pour tout $i = 1, ..., N-1$ par $m_{i,j} = \min_{k\in Z}(S_i)_k^n$ pour tout $j \in Z$. Nous avons en particulier $m_i = \min_{k\in Z}(S_i)_k^n \times v$, de telle sorte que les vecteurs m_i sont également des vecteurs propres de la matrice A associés à la valeur propre 1.
Ainsi, il est clair que :

$$A(S_i^{n+1} - m_i) = AS_i^{n+1} - Am_i = AS_i^{n+1} - m_i = S_i^n - m_i, \quad i = 1, ..., N-1.$$

Par ailleurs, la matrice A étant une M-matrice (à diagonale dominante) son inverse est une matrice positive : tous ses coefficients sont positifs ou nuls (voir [95] pour plus de détails). Les vecteurs $S_i^n - m_i$ étant par construction trivialement positifs, il en est donc de même pour les vecteurs $S_i^{n+1} - m_i$, ce qui signifie que

$$\min_{k\in Z}(S_i)_k^n \leq (S_i)_j^{n+1}, \quad i = 1, ..., N-1, \quad j \in Z.$$

Un raisonnement similaire mené avec le vecteur $M_i = (M_{i,j})_{j\in Z}$ défini pour tout $i = 1, ..., N-1$ par $M_{i,j} = \max_{k\in Z}(S_i)_k^n$, $j \in Z$ permet d'obtenir la majoration attendue, à savoir

$$(S_i)_j^{n+1} \leq \max_{k\in Z}(S_i)_k^n, \quad i = 1, ..., N-1, \quad j \in Z.$$

Ceci conclut la démonstration.

2.1.2 Implicitation du bloc hydrodynamique

Motivations et choix algorithmiques

Par construction, la formulation explicite en temps de la méthode de volumes finis associée au bloc hydrodynamique repose sur la fonction flux numérique de Godunov. Or, il est connu que cette fonction flux n'est que Lipschitz-continue et non différentiable. Rappelons en effet que Larrouturou a établi qu'une méthode de volumes finis capturant exactement des discontinuités stationnaires (ici en l'occurence les discontinuités de contact) ne peut pas être différentiable. Nous avons volontairement privilégié cette propriété de précision (pour des raisons déjà évoquées) au détriment de la propriété de différentiabilité pourtant fort agréable dans le contexte d'une implicitation linéarisée en temps. Puisque pour des raisons de coût, il est indispensable de procéder à une telle linéarisation, il importe d'isoler les défauts de différentiabilité dans la fonction flux considérée, ce qui semble difficile en l'état. De manière à contourner cette difficulté, nous montrons dans cette section que la méthode de Godunov ici développée peut être réinterprétée de manière équivalente comme une méthode de Roe (pour le système de Relaxation (1.87) avec $\lambda = 0$ et non pas bien sûr pour le système d'équilibre (1.1)). L'existence d'une réécriture du flux basée sur une linéarisation de Roe permet alors classiquement d'isoler les défauts de régularité en nous ramenant à des algorithmes implicites linéarisés largement éprouvés dans d'autres applications (voir par exemple les références classiques [108], [147], ...). La réinterprétation en termes de méthode de Roe est suggérée par une observation due à Mehlman [104], consistant à reconnaître que lorsque localement la solution d'un problème de Riemann pour le système des équations d'Euler n'est composée que de discontinuités (choc-contact-choc), alors la méthode de Godunov associée est localement équivalente à une méthode de Roe. Dans notre contexte, la validité de cette réinterprétation est uniforme puisque par la propriété de dégénérescence linéaire partagée par tous les champs du système (1.87) avec $\lambda = 0$, toutes les ondes composant la solution du problème de Riemann sont systématiquement des discontinuités. Nous établissons ce résultat de manière rigoureuse ci-dessous.

Définition de l'algorithme

Nous définissons au préalable quelques notations nécessaires pour la suite. Soient $\overline{\mathbf{u}}_L$ et $\overline{\mathbf{u}}_R$ deux états de $\overline{\Theta}$. Nous introduisons les vecteurs $\overline{\mathbf{w}}_L = \overline{\mathbf{w}}(\overline{\mathbf{u}}_L)$ et $\overline{\mathbf{w}}_R = \overline{\mathbf{w}}(\overline{\mathbf{u}}_R)$ de $I\!R^4$ définis par

$$\begin{aligned} \overline{\mathbf{w}}_L = \overline{\mathbf{w}}(\overline{\mathbf{u}}_L) = (\rho_L, (\rho u)_L, (\rho E)_L, (\rho I)_L), \\ \overline{\mathbf{w}}_R = \overline{\mathbf{w}}(\overline{\mathbf{u}}_R) = (\rho_R, (\rho u)_R, (\rho E)_R, (\rho I)_R), \end{aligned} \tag{2.16}$$

et définissons la solution auto-semblable $\overline{\mathbf{w}}(x/t; \overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)$ du système (2.2) pour la donnée initiale

$$\overline{\mathbf{w}}_0(x) = \begin{cases} \overline{\mathbf{w}}_L & \text{si} \quad x < 0, \\ \overline{\mathbf{w}}_R & \text{si} \quad x > 0. \end{cases} \tag{2.17}$$

En reprenant les notations du théorème 7, nous avons donc

$$\overline{\mathbf{w}}(x/t; \overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R) = \begin{cases} \overline{\mathbf{w}}_L & \text{si} \quad \frac{x}{t} < \lambda_1(\overline{\mathbf{w}}_L), \\ \overline{\mathbf{w}}_1 & \text{si} \quad \lambda_1(\overline{\mathbf{w}}_L) < \frac{x}{t} < \lambda_2(\overline{\mathbf{w}}_1), \\ \overline{\mathbf{w}}_2 & \text{si} \quad \lambda_2(\overline{\mathbf{w}}_2) < \frac{x}{t} < \lambda_3(\overline{\mathbf{w}}_R), \\ \overline{\mathbf{w}}_R & \text{si} \quad \lambda_3(\overline{\mathbf{w}}_R) < \frac{x}{t}, \end{cases} \tag{2.18}$$

avec

$$\overline{\mathbf{w}}_1 = (\rho_1, (\rho u)_1, (\rho\Sigma)_1, (\rho I)_1), \quad \overline{\mathbf{w}}_2 = (\rho_2, (\rho u)_2, (\rho\Sigma)_2, (\rho I)_2).$$

Soit $(\overline{\mathbf{w}}_j^n)_{j\in Z}$ la suite des vecteurs de $I\!R^4$ à l'équilibre déduits de $(\overline{\mathbf{u}}_j^n)_{j\in Z}$ et définis par

$$\overline{\mathbf{w}}_j^n = \overline{\mathbf{w}}(\overline{\mathbf{u}}_j^n) = (\rho_j^n, (\rho u)_j^n, (\rho E)_j^n, (\rho I)_j^n), \quad j \in Z,$$

$$(\rho I)_j^n = \rho_j^n \sum_{i=1}^N p_i(1/\rho_j^n, (\rho S_i)_j^n/\rho_j^n) + a^2.$$

La méthode de Godunov appliquée au système (2.2) se réécrit alors sous la condition CFL ≤ 1 sous la forme conservative suivante :

$$\overline{\mathbf{w}}_j^{n+1} = \overline{\mathbf{w}}_j^n - \frac{\Delta t}{\Delta x}(\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n) - \overline{\mathbf{f}}^g(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n)), \quad j \in Z, \tag{2.19}$$

où le flux numérique $\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n)$ est défini pour tout $j \in Z$ par

$$\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n) = \overline{\mathbf{f}}^e(\overline{\mathbf{w}}(0^+; \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n)). \tag{2.20}$$

Le flux exact $\overline{\mathbf{f}}^e$ du système (2.2) trouve quant à lui la définition

$$\overline{\mathbf{f}}^e(\rho, \rho u, \rho E, \rho I) = (\rho u, \rho u^2 + \Pi, (\rho E + \Pi)u, \rho I u), \quad I = \Pi + a^2/\rho.$$

De manière à réinterpréter le flux numérique de la méthode de Godunov (2.20) dans le formalisme des flux numériques associés à une méthode de Roe, la première étape de notre raisonnement consiste à réexprimer de manière équivalente ce flux numérique selon le résultat suivant :

Proposition 3
Soient $\overline{\mathbf{u}}_L$ et $\overline{\mathbf{u}}_R$ deux états de $\overline{\Theta}$ et $\overline{\mathbf{w}}_L$ et $\overline{\mathbf{w}}_R$ les vecteurs correspondants de $I\!R^4$ définis par (2.16). Soit $\overline{\mathbf{w}}(x/t; \overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)$ la solution auto-semblable définie par (2.18) du problème de Riemann (2.2) pour la donnée initiale (2.17).
Alors, le flux numérique $\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)$ associé à la méthode de Godunov et défini par $\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R) = \overline{\mathbf{f}}^e(\overline{\mathbf{w}}(0^+; \overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R))$ trouve la formulation équivalente suivante :

$$\begin{aligned}\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R) = \tfrac{1}{2}(\overline{\mathbf{f}}^e(\overline{\mathbf{w}}_L) + \overline{\mathbf{f}}^e(\overline{\mathbf{w}}_R)) - \\ \tfrac{1}{2}(|\lambda_1(\overline{\mathbf{w}}_L)|(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_L) - |\lambda_2(\overline{\mathbf{w}}_1)|(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_1) - |\lambda_3(\overline{\mathbf{w}}_R)|(\overline{\mathbf{w}}_R - \overline{\mathbf{w}}_2)).\end{aligned} \tag{2.21}$$

Démonstration
Soit $\overline{\mathbf{w}}(x/t; \overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)$ la solution auto-semblable définie par (2.18) du problème de Riemann (2.2) pour la donnée initiale (2.17). Nous définissons les quantités

$$\overline{\mathbf{w}}_l = \frac{2}{\Delta x}\int_{-\frac{\Delta x}{2}}^{0} \overline{\mathbf{w}}(x/\Delta t; \overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)dx,$$

$$\overline{\mathbf{w}}_r = \frac{2}{\Delta x}\int_{0}^{\frac{\Delta x}{2}} \overline{\mathbf{w}}(x/\Delta t; \overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)dx.$$

La solution du problème de Riemann n'étant composée que d'états constants (voir (2.18)), nous obtenons facilement

$$\begin{aligned}\overline{\mathbf{w}}_l = \overline{\mathbf{w}}_L - \tfrac{2\Delta t}{\Delta x}\times \\ (\lambda_1^-(\overline{\mathbf{w}}_L)(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_L) + \lambda_2^-(\overline{\mathbf{w}}_1)(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_1) + \lambda_3^-(\overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_R - \overline{\mathbf{w}}_2)),\end{aligned}$$

et

$$\begin{aligned}\overline{\mathbf{w}}_r = \overline{\mathbf{w}}_R - \tfrac{2\Delta t}{\Delta x}\times \\ (\lambda_1^+(\overline{\mathbf{w}}_L)(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_L) + \lambda_2^+(\overline{\mathbf{w}}_1)(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_1) + \lambda_3^+(\overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_R - \overline{\mathbf{w}}_2)),\end{aligned}$$

où nous avons de nouveau utilisé les notations génériques (2.8).
En intégrant le système (2.2) sur les rectangles $(-\Delta x/2, 0) \times (0, \Delta t)$ (respectivement $(0, \Delta x/2) \times (0, \Delta t)$), des arguments classiques conduisent par ailleurs à :

$$\overline{\mathbf{w}}_l = \overline{\mathbf{w}}_L - \frac{2\Delta t}{\Delta x}(\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R) - \overline{\mathbf{f}}^e(\overline{\mathbf{w}}_L)),$$

et

$$\overline{\mathbf{w}}_r = \overline{\mathbf{w}}_R - \frac{2\Delta t}{\Delta x}(\overline{\mathbf{f}}^e(\overline{\mathbf{w}}_R) - \overline{\mathbf{f}}^g(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)).$$

Il est ainsi possible d'obtenir deux nouvelles expressions du flux numérique de la méthode de Godunov, à savoir

$$\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R) = \overline{\mathbf{f}}^e(\overline{\mathbf{w}}_L)+$$

$$\lambda_1^-(\overline{\mathbf{w}}_L)(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_L) - \lambda_2^-(\overline{\mathbf{w}}_1)(\overline{\mathbf{w}}_2-\overline{\mathbf{w}}_1) - \lambda_3^-(\overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_R-\overline{\mathbf{w}}_2),$$

et

$$\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R) = \overline{\mathbf{f}}^e(\overline{\mathbf{w}}_R)-$$

$$\lambda_1^+(\overline{\mathbf{w}}_L)(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_L) - \lambda_2^+(\overline{\mathbf{w}}_1)(\overline{\mathbf{w}}_2-\overline{\mathbf{w}}_1) - \lambda_3^+(\overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_R-\overline{\mathbf{w}}_2).$$

Enfin, puisque

$$|\lambda_i| = \lambda_i^+ - \lambda_i^-,$$

nous réexprimons ce flux en prenant la demi-somme des deux expressions précédentes pour obtenir :

$$\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R) = \tfrac{1}{2}(\overline{\mathbf{f}}^e(\overline{\mathbf{w}}_L)+\overline{\mathbf{f}}^e(\overline{\mathbf{w}}_R))-$$

$$\tfrac{1}{2}(|\lambda_1(\overline{\mathbf{w}}_L)|(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_L) - |\lambda_2(\overline{\mathbf{w}}_1)|(\overline{\mathbf{w}}_2-\overline{\mathbf{w}}_1) - |\lambda_3(\overline{\mathbf{w}}_R)|(\overline{\mathbf{w}}_R-\overline{\mathbf{w}}_2)).$$

Ceci conclut la démonstration.

Pour comprendre ce flux de Godunov dans le formalisme des méthodes de Roe, il reste donc à montrer l'existence d'une matrice de Roe notée $\overline{\mathbf{A}}(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R)$ satisfaisant les trois propriétés de consistance suivantes (voir [113] pour plus de détails) :

$$\begin{array}{l} 1.\ \overline{\mathbf{A}}(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R) \quad \text{est diagonalisable à valeurs propres réelles,} \\ 2.\ \overline{\mathbf{A}}(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_R-\overline{\mathbf{w}}_L) = \overline{\mathbf{f}}^e(\overline{\mathbf{w}}_R)-\overline{\mathbf{f}}^e(\overline{\mathbf{w}}_L), \\ 3.\ \overline{\mathbf{A}}(\overline{\mathbf{w}},\overline{\mathbf{w}}) = \nabla\overline{\mathbf{f}}^e(\overline{\mathbf{w}}), \end{array} \tag{2.22}$$

ainsi que

$$\begin{array}{l} 4.\ |\overline{\mathbf{A}}|(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_R-\overline{\mathbf{w}}_L) = \\ (|\lambda_1(\overline{\mathbf{w}}_L)|(\overline{\mathbf{w}}_1-\overline{\mathbf{w}}_L) + |\lambda_2(\overline{\mathbf{w}}_1)|(\overline{\mathbf{w}}_2-\overline{\mathbf{w}}_1) + |\lambda_3(\overline{\mathbf{w}}_R)|(\overline{\mathbf{w}}_R-\overline{\mathbf{w}}_2)). \end{array} \tag{2.23}$$

En effet, dès que cette dernière relation est vérifiée, la proposition précédente montre que la fonction flux de Godunov s'écrit comme une fonction flux de Roe :

$$\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R) = \tfrac{1}{2}(\overline{\mathbf{f}}^e(\overline{\mathbf{w}}_L)+\overline{\mathbf{f}}^e(\overline{\mathbf{w}}_R)) - \tfrac{1}{2}|\overline{\mathbf{A}}|(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_R-\overline{\mathbf{w}}_L). \tag{2.24}$$

L'existence d'une telle linéarisée de Roe résulte du résultat suivant.

Théorème 9

Soient $\overline{\mathbf{u}}_L$ et $\overline{\mathbf{u}}_R$ deux états de $\overline{\Theta}$ et $\overline{\mathbf{w}}_L$ et $\overline{\mathbf{w}}_R$ les vecteurs correspondants de $I\!R^4$ définis par (2.16). En reprenant les notations de la proposition précédente et du théorème 7, nous définissons les vecteurs colonnes R_1, R_2^1, R_2^2, et R_3 en posant

$$R_1 = \begin{pmatrix} 1 \\ u_L - a/\rho_L \\ I_L \\ E_L + \Pi_L/\rho_L - au^\star/\rho_L \end{pmatrix}, \quad R_2^1 = \begin{pmatrix} 1 \\ u^\star \\ \Pi^\star \\ 0 \end{pmatrix},$$

$$R_2^2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad R_3 = \begin{pmatrix} 1 \\ u_R + a/\rho_R \\ I_R \\ E_R + \Pi_R/\rho_R + au^\star/\rho_R \end{pmatrix},$$

de telle sorte que

$$\begin{aligned} \overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_L &= (\rho_1 - \rho_L)R_1, \\ \overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_1 &= (\rho_2 - \rho_1)R_2^1 + (\rho_2\Sigma_2 - \rho_1\Sigma_1)R_2^2, \\ \overline{\mathbf{w}}_R - \overline{\mathbf{w}}_2 &= (\rho_R - \rho_2)R_3. \end{aligned} \tag{2.25}$$

Alors, ces vecteurs forment une base de $I\!R^4$ dès que a vérifie la condition de séparation des valeurs propres (1.90).
De plus, la matrice $\overline{\mathbf{A}}(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)$ définie à partir de sa forme diagonalisée par

$$\overline{\mathbf{A}}(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R) = B(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R) \begin{pmatrix} \lambda_1(\overline{\mathbf{w}}_L) & 0 & 0 & 0 \\ 0 & \lambda_2(\overline{\mathbf{w}}_1) & 0 & 0 \\ 0 & 0 & \lambda_2(\overline{\mathbf{w}}_2) & 0 \\ 0 & 0 & 0 & \lambda_3(\overline{\mathbf{w}}_R) \end{pmatrix} B^{-1}(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R),$$
$$B(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R) = [R_1, R_2^1, R_2^2, R_3](\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R), \tag{2.26}$$

vérifie les quatre propriétés de consistance (2.22)-(2.23).

Démonstration

En utilisant les valeurs explicites (voir théorème 7) des vecteurs $\overline{\mathbf{w}}_1$ et $\overline{\mathbf{w}}_2$, il est facile de vérifier que les trois relations définitions (2.25) conduisent aux vecteurs colonnes proposés. Rappelons que la condition (1.90) sur a impose aux densités intermédiaires d'être strictement positives. Des calculs simples exploitant l'ensemble des invariants vérifiés par la solution du problème de Riemann (voir théorème 7) montrent que ces quatre vecteurs sont nécessairement linéairement indépendants. En conséquence, la matrice B^{-1} existe et la linéarisée proposée est bien définie.
Ensuite, la troisième propriété de consistance est satisfaite, puisqu'un calcul direct montre que les vecteurs propres de la jacobienne du flux exact $\nabla\overline{\mathbf{f}}^e(\overline{\mathbf{w}})$ évalués en $\overline{\mathbf{w}} = (\rho, \rho u, \rho\Sigma, \rho I)$ sont précisément $(1, u - a/\rho, I, \Sigma + \Pi/\rho - au/\rho), (1, u, \Pi, 0), (0, 0, 0, 1)$ et $(1, u + a/\rho, I, \Sigma + \Pi/\rho + au/\rho)$ respectivement associés aux valeurs propres $\lambda_1(\overline{\mathbf{w}})$, $\lambda_2(\overline{\mathbf{w}})$ (valeur propre d'ordre 2), $\lambda_3(\overline{\mathbf{w}})$. Nous rappelons en effet que $\Pi = I - a^2/\rho$.
La première propriété étant vraie par construction, il reste à démontrer la validité des propriétés 2 et 4.
Par construction, les vecteurs R_1, R_2^1, R_2^2, et R_3 sont des vecteurs propres de la matrice $\overline{\mathbf{A}}(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)$.
Ainsi, d'après (2.25) :

$$\begin{aligned} \overline{\mathbf{A}}(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_R - \overline{\mathbf{w}}_L) &= \overline{\mathbf{A}}(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_L) + \overline{\mathbf{A}}(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_1) \\ &+ \overline{\mathbf{A}}(\overline{\mathbf{w}}_L, \overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_R - \overline{\mathbf{w}}_2) \\ &= \lambda_1(\overline{\mathbf{w}}_L)(\overline{\mathbf{w}}_1 - \overline{\mathbf{w}}_L) + \lambda_2(\overline{\mathbf{w}}_1)(\overline{\mathbf{w}}_2 - \overline{\mathbf{w}}_1) \\ &+ \lambda_3(\overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_R - \overline{\mathbf{w}}_2). \end{aligned}$$

Or, les valeurs propres $\lambda_1(\overline{\mathbf{w}}_L)$, $\lambda_2(\overline{\mathbf{w}}_1)$, $\lambda_3(\overline{\mathbf{w}}_R)$ ne sont autres que les vitesses de propagation des discontinuités composant la solution du problème de Riemann. En invoquant les conditions de Rankine-Hugoniot associées, nous déduisons de l'identité précédente :

$$\begin{aligned}\overline{\mathbf{A}}(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R)(\overline{\mathbf{w}}_R-\overline{\mathbf{w}}_L) &= \overline{\mathbf{f}}^e(\overline{\mathbf{w}}_1)-\overline{\mathbf{f}}^e(\overline{\mathbf{w}}_L)+\overline{\mathbf{f}}^e(\overline{\mathbf{w}}_2)-\overline{\mathbf{f}}^e(\overline{\mathbf{w}}_1)\\ &+ \overline{\mathbf{f}}^e(\overline{\mathbf{w}}_R)-\overline{\mathbf{f}}^e(\overline{\mathbf{w}}_2)\\ &= \overline{\mathbf{f}}^e(\overline{\mathbf{w}}_R)-\overline{\mathbf{f}}^e(\overline{\mathbf{w}}_L).\end{aligned}$$

La deuxième propriété de consistance est donc démontrée.
De la même manière et puisque $|\overline{\mathbf{A}}|(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R)$ vaut par définition :

$$|\overline{\mathbf{A}}|(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R)=B\begin{pmatrix}|\lambda_1(\overline{\mathbf{w}}_L)| & 0 & 0 & 0\\ 0 & |\lambda_2(\overline{\mathbf{w}}_1)| & 0 & 0\\ 0 & 0 & |\lambda_2(\overline{\mathbf{w}}_2)| & 0\\ 0 & 0 & 0 & |\lambda_3(\overline{\mathbf{w}}_R)|\end{pmatrix}B^{-1}, \tag{2.27}$$

les vecteurs R_1, R_2^1, R_2^2, et R_3 sont des vecteurs propres de la matrice $|\overline{\mathbf{A}}|(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R)$. Nous obtenons donc la quatrième propriété d'après les relations définitions (2.25), ce qui conclut la démonstration.

Quelques rappels sur l'implicitation linéarisée d'une méthode de Roe et choix de a

Les brefs rappels donnés ci-dessous sont utiles pour préciser le choix du paramètre a entrant dans la définition du système de Relaxation. Ils ne sont en aucun cas une présentation exhaustive des techniques d'implicitation linéarisée associées à la méthode de Roe. Il existe en effet dans la littérature au moins deux implicitations linéarisées de la méthode de Roe (voir par exemple [147]), l'une conservative, l'autre non, sauf à convergence de l'algorithme en temps vers une solution stationnaire. Puisque notre objectif ne concerne que les solutions stationnaires, nous décrivons brièvement l'une d'entre elles, à savoir celle non conservative (dans la phase instationnaire) car elle permet de mieux appréhender les raisons heuristiques du succès dans les applications pratiques de la linéarisation proposée. Nous suivons ici une présentation très classiquement adoptée dans l'immense littérature consacrée au sujet.
Introduisons la notation $\overline{\mathbf{A}}^n_{j+1/2}=\overline{\mathbf{A}}(\overline{\mathbf{w}}^n_j,\overline{\mathbf{w}}^n_{j+1})$. La deuxième propriété de (2.22) implique

$$\overline{\mathbf{f}}^e(\overline{\mathbf{w}}^n_{j+1})-\overline{\mathbf{f}}^e(\overline{\mathbf{w}}^n_j)=\overline{\mathbf{A}}^n_{j+1/2}(\overline{\mathbf{w}}^n_{j+1}-\overline{\mathbf{w}}^n_j),$$

de telle sorte que (2.19) peut être réécrit, grâce à (2.24), sous la forme :

$$\begin{gathered}\overline{\mathbf{w}}^{n+1}_j=\overline{\mathbf{w}}^n_j-\frac{\Delta t}{2\Delta x}\times\\ ((\overline{\mathbf{A}}^n_{j-1/2}+|\overline{\mathbf{A}}|^n_{j-1/2})(\overline{\mathbf{w}}^n_j-\overline{\mathbf{w}}^n_{j-1})+(\overline{\mathbf{A}}^n_{j+1/2}-|\overline{\mathbf{A}}|^n_{j+1/2})(\overline{\mathbf{w}}^n_{j+1}-\overline{\mathbf{w}}^n_j)).\end{gathered} \tag{2.28}$$

En définissant les matrices $\overline{\mathbf{A}}^+(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R)$ et $\overline{\mathbf{A}}^-(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R)$ par

$$\overline{\mathbf{A}}^\pm(\overline{\mathbf{w}}_L,\overline{\mathbf{w}}_R)=B\begin{pmatrix}\lambda_1^\pm(\overline{\mathbf{w}}_L) & 0 & 0 & 0\\ 0 & \lambda_2^\pm(\overline{\mathbf{w}}_1) & 0 & 0\\ 0 & 0 & \lambda_2^\pm(\overline{\mathbf{w}}_2) & 0\\ 0 & 0 & 0 & \lambda_3^\pm(\overline{\mathbf{w}}_R)\end{pmatrix}B^{-1}, \tag{2.29}$$

nous obtenons facilement

$$|\overline{\mathbf{A}}| = \overline{\mathbf{A}}^{+} - \overline{\mathbf{A}}^{-}, \quad \overline{\mathbf{A}} = \overline{\mathbf{A}}^{+} + \overline{\mathbf{A}}^{-}.$$

Ainsi, (2.28) devient

$$\begin{gathered}\overline{\mathbf{w}}_j^{n+1} = \overline{\mathbf{w}}_j^n - \tfrac{\Delta t}{\Delta x}\times \\ (\overline{\mathbf{A}}^{+n}_{j-1/2}(\overline{\mathbf{w}}_j^n - \overline{\mathbf{w}}_{j-1}^n) + \overline{\mathbf{A}}^{-n}_{j+1/2}(\overline{\mathbf{w}}_{j+1}^n - \overline{\mathbf{w}}_j^n)),\end{gathered} \tag{2.30}$$

où l'on a posé

$$\overline{\mathbf{A}}^{+n}_{j+1/2} = \overline{\mathbf{A}}^{+}(\overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n), \quad \overline{\mathbf{A}}^{-n}_{j+1/2} = \overline{\mathbf{A}}^{-}(\overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n), \quad j \in Z.$$

L'implicitation linéarisée de (2.30) est classiquement obtenue en figeant à l'instant t^n les matrices $\overline{\mathbf{A}}^{+n}_{j-1/2}$ et $\overline{\mathbf{A}}^{-n}_{j+1/2}$. Ainsi, la suite $(\overline{\mathbf{w}}_j^{n+1})_{j\in Z}$ est solution de :

$$\begin{gathered}\overline{\mathbf{w}}_j^{n+1} = \overline{\mathbf{w}}_j^n - \tfrac{\Delta t}{\Delta x}\times \\ (\overline{\mathbf{A}}^{+n}_{j-1/2}(\overline{\mathbf{w}}_j^{n+1} - \overline{\mathbf{w}}_{j-1}^{n+1}) + \overline{\mathbf{A}}^{-n}_{j+1/2}(\overline{\mathbf{w}}_{j+1}^{n+1} - \overline{\mathbf{w}}_j^{n+1})),\end{gathered} \tag{2.31}$$

où rappelons-le, $\overline{\mathbf{w}}_j^n = (\rho_j^n, (\rho u)_j^n, (\rho E)_j^n, (\rho I)_j^n)$ pour tout $j \in Z$ et tout $n \in N$. La mise en œuvre numérique consiste en la résolution d'un système linéaire

$$\mathcal{A}\overline{\mathbf{w}}_j^{n+1} = \overline{\mathbf{w}}_j^n \tag{2.32}$$

où la matrice $\mathcal{A}$ possède une structure tridiagonale par blocs, chaque bloc représentant une matrice de taille 4×4 dans notre cas.
Plus précisément, une ligne courante j de la matrice est constituée des blocs élémentaires $\mathbf{a}_j, \mathbf{b}_j$ et $\mathbf{c}_j$ donnés par :

$$\begin{aligned}\mathbf{a}_j &= -\tfrac{\Delta t}{\Delta x}\overline{\mathbf{A}}^{+n}_{j-1/2}, \\ \mathbf{b}_j &= Id + \tfrac{\Delta t}{\Delta x}(\overline{\mathbf{A}}^{+n}_{j-1/2} - \overline{\mathbf{A}}^{-n}_{j+1/2}), \\ \mathbf{c}_j &= \tfrac{\Delta t}{\Delta x}\overline{\mathbf{A}}^{-n}_{j+1/2}.\end{aligned} \tag{2.33}$$

Pour achever la descritpion de l'implicitation du bloc hydrodynamique, il convient de préciser à chaque pas de temps le choix du paramètre a entrant dans la définition de la méthode de Relaxation. A cette fin, reformulons l'algorithme d'implicitation proposé en adoptant une écriture incrémentale en temps classique (voir [147]). Il est facile de constater que la résolution du problème linéaire (2.32) est équivalente à celle du problème incrémental suivant :

$$\begin{gathered}\mathcal{A}\Delta\overline{\mathbf{w}}_j^{n+1/2} = \overline{\mathbf{r}}(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n), \quad \Delta\overline{\mathbf{w}}_j^{n+1/2} = \frac{\overline{\mathbf{w}}_j^{n+1} - \overline{\mathbf{w}}_j^n}{\Delta t}, \\ \overline{\mathbf{r}}(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n) := -\tfrac{1}{\Delta x}\{\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n) - \overline{\mathbf{f}}^g(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n)\},\end{gathered} \tag{2.34}$$

puisque

$$\overline{\mathbf{f}}^g(\overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n) - \overline{\mathbf{f}}^g(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n) = (\overline{\mathbf{A}}^{+n}_{j-1/2}(\overline{\mathbf{w}}_j^n - \overline{\mathbf{w}}_{j-1}^n) + \overline{\mathbf{A}}^{-n}_{j+1/2}(\overline{\mathbf{w}}_{j+1}^n - \overline{\mathbf{w}}_j^n)). \tag{2.35}$$

Pour obtenir cette identité, nous avons évidemment supposé que la valeur de a entrant dans la définition de la matrice $\mathcal{A}$ et du membre de droite (appelé communément résidu explicite) est identique. Toutefois, l'expérience de la mécanique des fluides numérique indique qu'il est souvent préférable d'utiliser une matrice $\mathcal{A}$ provenant de la linéarisation d'une fonction flux numérique dissipative dans un souci de robustesse, tout en conservant une définition du résidu explicite associée à une fonction flux choisie pour ses qualités de résolution. De sorte qu'à convergence en temps, la solution stationnaire est décrite par cette dernière fonction flux. Conformément à cette heuristique, nous évaluons la partie résidu explicite à chaque pas de temps en faisant choix d'une valeur de a locale satisfaisant au plus près les conditions de Whitham exhibées au théorème 5. Concernant la définition de la matrice $\mathcal{A}$, nous avons adopté un choix de a global résultant du maximum des valeurs précédentes.

Ceci achève la description de l'implicitation du bloc hydrodynamique (2.2). Pour conclure cette section, rapportons brièvement les considérations classiques visant à justifier formellement la linéarisation proposée. Les identités (2.33) suggèrent que la matrice tridiagonale par blocs ainsi obtenue pourrait être associée, via un changement de base (conduisant aux variables caractéristiques), à une M-matrice : matrice à diagonale positive (dominante) et à termes extra-diagonaux négatifs. Ceci ne vaut bien entendu que si les parties positive $\overline{\mathbf{A}}^{+}{}_{j+1/2}$ et négative $\overline{\mathbf{A}}^{-}{}_{j+1/2}$ se diagonalisent pour tout j dans la même base. Ceci est faux en toute rigueur mais l'algorithme fonctionne remarquablement bien dans la pratique.

2.1.3 Projection sur la variété d'équilibre et défaut de convergence

Les algorithmes d'implicitation des blocs hydrodynamique et transport des entropies permettent de remettre à jour la solution discrète relaxée $(\rho, \rho u, \rho E, \{\rho S_i\}_{i=1,\dots,N-1}, \rho I)$ et donc de redéfinir à l'issue de cette réactualisation une nouvelle valeur de ρI à l'équilibre à l'instant considéré :

$$\rho I = \rho \sum_{i=1}^{N} p_i(1/\rho, \rho S_i/\rho) + a^2, \tag{2.36}$$

pour un nouveau choix de a tel qu'il vient d'être discuté. Les expériences de validation du bloc hydrodynamique menées dans un même code de calculs ont montré qu'il était possible d'adopter une stratégie de montée en nombre CFL comparable à celle d'une méthode de Van Leer linéairement implicitée (dans le cadre des équations d'Euler usuelles, voir la présentation des résultats numériques). Comme il est communément admis, cette observation indique une bonne stabilité de l'algorithme d'implicitation proposé.

Toutefois, l'historique de décroissance de la norme L^2 des résidus explicites fait apparaître après quelques décades de décroissance une valeur limite, synonyme de non convergence vers une solution parfaitement stationnaire en temps ($||\overline{\mathbf{r}}(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n)|| = 0$, condition de convergence en temps que nous préciserons ci-dessous). L'objectif de cette section est d'expliquer pourquoi l'implicitation en temps du bloc hydrodynamique ne peut pas en règle générale conduire à des solutions rigoureusement stationnaires. La mise en évidence de l'origine de cette impossibilité permettra d'introduire une technique de correction simple et stable (comme le montrent les expériences numériques). Il suffit, ainsi que nous le verrons, de restreindre la discussion au cadre des équations d'Euler 3×3 puisque seule est incriminée ici l'implicitation du bloc hydrodynamique.

Précisons dès maintenant notre critère de convergence en temps vers une solution stationnaire. Dans le cadre classique, une solution est dite stationnaire si la norme L^2 de tous les résidus explicites est nulle. L'inversibilité de la matrice $\mathcal{A}$ (sous des conditions aux limites *adhoc*) implique en conséquence que tous les incréments en temps sont nuls. Toutefois dans notre contexte, seule la convergence en temps vers un stationnaire des variables densité, quantité de mouvement et énergie totale nous intéresse dans la pratique. Selon notre stratégie d'implicitation linéarisée, l'évolution en temps de ces variables est cependant couplée à celle de la variable de Relaxation ρI, variable sur laquelle il sera difficile, sinon impossible, d'exprimer une notion de convergence en temps (voir ci-dessous). Dès lors et de manière à définir un critère de convergence sur ρ, ρu et ρE, nous exigerons non seulement des résidus associés de converger vers zéro, mais également des incréments en temps de ces mêmes variables de converger vers zéro. Ce critère peut paraître exigeant mais il semble malaisé d'en proposer un plus faible lorsque l'une des variables d'évolution, ici ρI, linéairement couplée aux autres via le problème linéaire (2.34) sera vue ne pas pouvoir converger en temps. Avec des notations claires, nous exigeons donc :

$$\overline{\mathbf{r}}_\rho(\overline{\mathbf{w}}^n_{j-1}, \overline{\mathbf{w}}^n_j, \overline{\mathbf{w}}^n_{j+1}) = \overline{\mathbf{r}}_{\rho u}(\overline{\mathbf{w}}^n_{j-1}, \overline{\mathbf{w}}^n_j, \overline{\mathbf{w}}^n_{j+1}) = \overline{\mathbf{r}}_{\rho\Sigma}(\overline{\mathbf{w}}^n_{j-1}, \overline{\mathbf{w}}^n_j, \overline{\mathbf{w}}^n_{j+1}) = 0,$$

et

$$\Delta\rho_j^{n+1/2} = \Delta\rho u_j^{n+1/2} = \Delta\rho\Sigma_j^{n+1/2} = 0.$$

Par opposition aux variables ρ, ρu et ρE, la variable de Relaxation ρI connaît deux mises à jour distinctes au cours de chaque itération : une prédiction via le système de Relaxation puis une correction par projection sur la variété d'équilibre. Supposons dans une première étape que les inconnues ρ, ρu, ρE deviennent solutions stationnaires du système de Relaxation. Montrons alors formellement que la variable ρI ne peut pas être solution stationnaire de ce système. Nous en déduirons par contradiction dans une seconde étape que les inconnues ρ, ρu, ρE ne peuvent pas non plus être en général solutions stationnaires du système les gouvernant.

Première étape Il suffit de supposer que les quatre variables composant le système de Relaxation convergent vers une solution stationnaire, et d'aboutir à une contradiction. La discussion proposée est formelle puisqu'elle repose sur une analyse de consistance au sens des volumes finis. Si la convergence en temps avait lieu, la solution discrète relaxée représenterait une approximation constante par morceaux d'une solution du système

$$\begin{cases} \partial_x \rho u = 0, \\ \partial_x(\rho u^2 + \Pi) = 0, \\ \partial_x(\rho\Sigma u + \Pi u) = 0, \\ \partial_x \rho I u = 0, \quad I = \Pi + a^2/\rho, \end{cases} \tag{2.37}$$

puisque l'hypothèse d'annulation de tous les résidus explicites revient à exiger l'annulation de tous les incréments en temps. D'après la formule (2.35), ceux-ci prennent chacun la forme d'une différence de flux consistante au sens des volumes finis avec l'opérateur concerné dans (2.37). Notons ensuite que la variable de Relaxation ρI ne peut être qu'à l'équilibre en conséquence de la projection sur la variété d'équilibre à chaque itération et de l'hypothèse d'indépendance relativement (aux itérations) au temps. Ceci implique l'identité :

$$I(x) = p(x) + a^2\tau(x). \tag{2.38}$$

De ce fait, la solution stationnaire considérée est nécessairement une approximation d'une solution de :

$$\begin{cases} \partial_x \rho u = 0, \\ \partial_x(\rho u^2 + p) = 0, \\ \partial_x(\rho E u + p u) = 0, \\ \partial_x \rho p u + a^2 \partial_x u = 0. \end{cases} \tag{2.39}$$

Or, la validité des trois premières équations conduit sous l'hypothèse d'une solution régulière, la pression p à satisfaire l'identité suivante :

$$u\partial_x p - \tau\frac{\partial p}{\partial \tau}\partial_x u = 0, \tag{2.40}$$

version stationnaire de la loi d'évolution bien connue de la pression :

$$\partial_t p + u\partial_x p - \tau\frac{\partial p}{\partial \tau}\partial_x u = 0.$$

Rapprochons l'identité (2.40) de la quatrième équation du système (2.39), clairement équivalente à :

$$u\partial_x p + a^2\tau\partial_x u = 0.$$

La soustraction de ces deux EDO pour la pression p conduit à :

$$(a^2 + \frac{\partial p}{\partial \tau})\tau\partial_x u = 0.$$

La constance en x de $(\rho u)(x)$ fait qu'en règle générale $\tau\partial_x u$ est non nul, d'où nécessairement une contradiction avec la condition de Whitham qui exige $a^2 + \partial_\tau p > 0$.

Deuxième étape Supposons la convergence en temps vers un état stationnaire des variables ρ, ρu et ρE de sorte que ρI ne peut pas devenir indépendant du temps au sens de la discussion précédente. Sur la base de notre critère de convergence, nous obtenons par hypothèse :

$$\overline{\mathbf{r}}_\rho(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n) = \overline{\mathbf{r}}_{\rho u}(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n) = \overline{\mathbf{r}}_{\rho\Sigma}(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n) = 0,$$

et

$$\Delta\rho_j^{n+1/2} = \Delta\rho u_j^{n+1/2} = \Delta\rho\Sigma_j^{n+1/2} = 0,$$

tandis que

$$\overline{\mathbf{r}}_{\rho I}(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n) \neq 0 \quad \text{et} \quad \Delta\rho I_j^{n+1/2} \neq 0.$$

En reprenant les notations (2.33), la structure tridiagonale par blocs 4×4 de la matrice $\mathcal{A}$ implique alors localement en espace (*i.e.* sur une maille courante j) l'identité vectorielle suivante :

$$\mathbf{a}_j\Delta\overline{\mathbf{w}}_{j-1}^{n+1/2} + \mathbf{b}_j\Delta\overline{\mathbf{w}}_j^{n+1/2} + \mathbf{c}_j\Delta\overline{\mathbf{w}}_{j+1}^{n+1/2} = \overline{\mathbf{r}}(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n), \tag{2.41}$$

avec pour $l = j-1, j, j+1$:

$$\Delta\overline{\mathbf{w}}_l^{n+1/2} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \Delta\rho I_l^{n+1/2} \end{pmatrix}, \quad \text{et} \quad \overline{\mathbf{r}}(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \overline{\mathbf{r}}_{\rho I}(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n) \end{pmatrix}. \tag{2.42}$$

Les trois incréments $\{\Delta\rho I_l^{n+1/2}\}_{l=j-1,j,j+1}$ sont donc astreints à être localement solutions d'un système linéaire de quatre équations *a priori* indépendantes, *i.e.* avec des notations claires :

$$\left\{ \begin{array}{l} a_j^{k,4}\Delta\rho I_{j-1}^{n+1/2} + b_j^{k,4}\Delta\rho I_j^{n+1/2} + c_j^{k,4}\Delta\rho I_{j+1}^{n+1/2} = 0, \quad k = 1, ..., 3, \\ a_j^{4,4}\Delta\rho I_{j-1}^{n+1/2} + b_j^{4,4}\Delta\rho I_j^{n+1/2} + c_j^{4,4}\Delta\rho I_{j+1}^{n+1/2} = \overline{\mathbf{r}}_{\rho I}(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n), \end{array} \right. \tag{2.43}$$

de sorte que l'incrément vectoriel $\Delta\rho I^{n+1/2}$ doit simultanément être dans le noyau de trois opérateurs linéaires *a priori* distincts tout en étant solution non triviale (*i.e.* $\Delta\rho I^{n+1/2} \neq 0$) d'un système linéaire avec second membre non nul puisque le résidu $\overline{\mathbf{r}}_{\rho I}$ ne peut être nul. Des tests numériques (non rapportés ici) montrent que les quatre équations (2.43) sont généralement incompatibles de sorte que les inconnues ρ, ρu et $\rho\Sigma$ ne peuvent être solutions stationnaires du système relaxé gouvernant leurs évolutions.

2.2 Correction de l'implicitation "naturelle"

L'objectif de cette section est de remédier au défaut de convergence vers un stationnaire de l'algorithme implicite associé au bloc hydrodynamique, décrit dans la section précédente. Puisque notre intérêt porte sur le système Euler multi-pression, nous exposerons le principe de la stratégie de correction directement dans ce cas. Sa restriction au cas des équations Euler 3×3 s'en deduira immédiatement (en posant $N = 1$).

L'impossibilité de converger vers des valeurs stationnaires en temps de ρ, ρu et ρE résulte du fait que la variable de Relaxation ρI devrait satisfaire à convergence, un ensemble sur-contraint de conditions. Il nous faut donc éliminer ces contraintes généralement incompatibles. Nous montrons formellement ci-dessous qu'un recouplage adéquat de l'étape de projection sur la variété d'équilibre avec la résolution en temps du bloc hydrodynamique permet d'éliminer ces contraintes. Un tel recouplage conduira à exprimer localement sur chaque cellule de calcul l'incrément ρI comme une fonction linéaire des incréments en densité, quantité de mouvement, énergie totale et entropies spécifiques. Les incréments en temps sur les entropies spécifiques ayant été évalués au préalable, la structure tridiagonale par blocs 4×4 associée au bloc hydrodynamique deviendra une structure toujours tridiagonale mais cette fois composée de blocs 3×3, le membre de droite constitué des résidus explicites recevant une modification précisée ci-dessous. Le bénéfice est donc double puisque la convergence vers une solution rigoureusement stationnaire sera assurée (ainsi que le montre l'expérience), et ce au prix de l'inversion d'un problème matriciel de taille moindre.

Nous donnons ici la conséquence formelle de la technique de recouplage envisagée puisque celle-ci doit être menée de manière linéarisée en temps pour des raisons de coût d'évaluation, et ce dans le régime asymptotique d'un paramètre de Relaxation $\lambda \to +\infty$. Une justification de cette procédure réclame des calculs fastidieux ainsi que des hypothèses difficilement vérifiables dans la pratique. Nous ne les rapporterons pas ici dans le détail faute de temps. La remarque concluant cette section donne les grandes lignes directrices de la démarche sous-tendant ces calculs.

Au terme de ces calculs, il s'avère que tout se passe comme s'il fallait prendre en compte l'étape de remise à l'équilibre lors de la résolution du système (2.34). Il s'agit dans les faits d'abandonner le choix $\lambda = 0$ lors de l'évolution en temps de la variable de Relaxation ρI pour lui substituer la résolution implicite linéarisée en temps de l'équation :

$$\partial_t \rho I + \partial_x \rho I u = \lambda \rho (p - \Pi),$$

dans le régime asymptotique $\lambda = +\infty$. Après division par λ et passage à la limite formel sur ce paramètre, une telle démarche conduit à devoir imposer à ρI de vérifier une version linéarisée en temps de la condition d'équilibre :

$$(\rho I)_j^{n+1} = \rho_j^{n+1} \sum_{i=1}^{N} p_i(1/\rho_j^{n+1}, (S_i)_j^{n+1}) + a^2, \quad j \in Z. \tag{2.44}$$

Ici, l'entropie $(S_N)_j^{n+1} := S_N(\rho_j^{n+1}, (\rho u)_j^{n+1}, (\rho\Sigma)_j^{n+1}, \{(S_i)_j^{n+1}\}_{i=1,...,N-1})$ est déterminée par la loi de fermeture sur l'énergie totale (1.3). Notons qu'en définissant la fonction $\mathbf{u} \to f(\mathbf{u})$ par

$$f(\mathbf{u}) = \rho \sum_{i=1}^{N} p_i(1/\rho, S_i) + a^2, \quad \mathbf{u} = (\rho, \rho u, \rho\Sigma, \{S_i\}_{i=1,...,N-1}),$$

avec $S_N = S_N(\rho, \rho u, \rho\Sigma, \{S_i\}_{i=1,...,N-1})$, (2.44) se réécrit $(\rho I)_j^{n+1} = f(\mathbf{u}_j^{n+1})$. Conformément à l'approche adoptée pour l'implicitation du bloc hydrodynamique, la linéarisation envisagée repose sur un développement de Taylor à l'ordre 1 de la fonction f autour de l'instant t^n :

$$\begin{aligned}(\rho I)_j^{n+1} &= (\rho I)_j^n + (\partial_\rho f)_j^n \Delta\rho_j^{n+1/2} + (\partial_{\rho u} f)_j^n \Delta\rho u_j^{n+1/2} \\ &+ (\partial_{\rho\Sigma} f)_j^n \Delta\rho\Sigma_j^{n+1/2} + \textstyle\sum_{i=1}^{N-1} (\partial_{S_i} f)_j^n \Delta(S_i)_j^{n+1/2}.\end{aligned} \tag{2.45}$$

Dès lors, les incréments $\Delta\rho I_j^{n+1/2}$ ne sont plus des inconnues indépendantes mais s'expriment comme des fonctions affines des incréments inconnus $\Delta\rho_j^{n+1/2}$, $\Delta\rho u_j^{n+1/2}$ et $\Delta\rho\Sigma_j^{n+1/2}$. Précisons en effet que les incréments $\Delta(S_i)_j^{n+1/2}$, $i = 1, ..., (N-1)$ intervenant dans (2.45) sont déterminés par l'implicitation découplée du bloc transport des entropies et donc connus à ce niveau de description de l'algorithme. De même, la fonction f et l'ensemble de ses dérivées partielles sont explicitement connues puisqu'évaluées à l'instant t^n.
Sur une maille courante j et de manière à rendre compte de la relation algébrique (2.45) liant les incréments, le système linéaire à résoudre (2.34), déduit de (2.41)-(2.45), s'écrit maintenant sous la forme :

$$\tilde{\mathbf{a}}_j \Delta\mathbf{w}_{j-1}^{n+1/2} + \tilde{\mathbf{b}}_j \Delta\mathbf{w}_j^{n+1/2} + \tilde{\mathbf{c}}_j \Delta\mathbf{w}_{j+1}^{n+1/2} = \tilde{\mathbf{r}}(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n),$$

où nous avons posé $\mathbf{w} = (\rho, \rho u, \rho\Sigma)$. Les matrices impliquées $\tilde{\mathbf{a}}_j$, $\tilde{\mathbf{b}}_j$ et $\tilde{\mathbf{c}}_j$ sont déduites de (2.41)-(2.45) et sont définies avec des notations claires, colonnes par colonnes, par

$$\tilde{a}_j^{k,1} = a_j^{k,1} + a_j^{k,4}(\partial_\rho f)_j^n, \quad \tilde{a}_j^{k,2} = a_j^{k,2} + a_j^{k,4}(\partial_{\rho u} f)_j^n, \quad \tilde{a}_j^{k,3} = a_j^{k,3} + a_j^{k,4}(\partial_{\rho\Sigma} f)_j^n, \tag{2.46}$$

pour tout $k = 1, ..., 3$, tandis que les résidus explicites s'écrivent :

$$\tilde{\mathbf{r}}_\rho = \overline{\mathbf{r}}_\rho - a_j^{1,4} \sum_{i=1}^{N-1} (\partial_{S_i} f)_j^n \Delta(S_i)_j^{n+1/2}, \quad \tilde{\mathbf{r}}_{\rho u} = \overline{\mathbf{r}}_{\rho u} - a_j^{2,4} \sum_{i=1}^{N-1} (\partial_{S_i} f)_j^n \Delta(S_i)_j^{n+1/2}, \tag{2.47}$$

et

$$\tilde{\mathbf{r}}_{\rho\Sigma} = \overline{\mathbf{r}}_{\rho\Sigma} - a_j^{3,4} \sum_{i=1}^{N-1} (\partial_{S_i} f)_j^n \Delta(S_i)_j^{n+1/2}. \tag{2.48}$$

Ainsi, l'implicitation du bloc hydrodynamique permettant d'évaluer la suite

$$(\rho_j^{n+1}, (\rho u)_j^{n+1}, (\rho E)_j^{n+1})_{j\in Z},$$

consiste en la résolution d'un système linéaire dont la matrice possède une structure tridiagonale par blocs, chaque bloc étant de taille 3×3. Les contraintes incompatibles qui portaient sur l'incrément $\Delta\rho I_j^{n+1/2}$ ont donc été éliminées au prix d'une modification des matrices élémentaires gouvernant les incréments en ρ, ρu, ρE. En ce sens, aucun argument ne s'oppose plus à la convergence de l'algorithme proposé vers un stationnaire en ces variables. Notons en effet que lorsque les implicitations des blocs transport des entropies et hydrodynamique convergent simultanément vers une solution stationnaire (cette situation est bien évidemment celle qui nous intéresse), l'annulation des incréments en temps $\Delta(S_i)_j^{n+1/2} = 0$ conduit à l'annulation des résidus explicites $\overline{\mathbf{r}}_\rho$, $\overline{\mathbf{r}}_{\rho u}$ et $\overline{\mathbf{r}}_{\rho\Sigma}$.

La remarque concluant cette section a pour but d'indiquer que la procédure proposée n'est en fait qu'une prise en compte de la projection sur la variété d'équilibre (*i.e.* avec $\lambda = +\infty$) mais linéarisée en temps. Il est dès lors indispensable de corriger cette linéarisation de manière à redémarrer l'algorithme à chaque itération sur une solution rigoureusement à l'équilibre, en réactualisant la valeur de ρI localement sur chaque cellule selon (2.44).

De manière à résumer la stratégie d'implicitation en temps des équations d'Euler multi-pression, il s'agit dans un premier temps de résoudre $(N-1)$ systèmes linéaires à structure tridiagonale scalaire pour l'étape de convection des entropies $(S_i)_j^{n+1}$ dont la connaissance est indispensable à la définition des résidus (2.47)-(2.48). Dans un second temps, il s'agit de résoudre un système linéaire dont la matrice possède une structure tridiagonale par blocs de taille 3×3 (pour le calcul des quantités $\rho_j^{n+1}, (\rho u)_j^{n+1}, (\rho E)_j^{n+1}$) et enfin de redéfinir ρI à l'équilibre. Ainsi et *stricto sensu*, l'évolution du bloc hydrodynamique se retrouve être faiblement couplée avec celle du bloc transport des entropies. Ce couplage faible n'intervenant que dans la définition des résidus "explicites" est bien évidemment moins coûteux qu'une approche implicite linéarisée classique consistant à résoudre un système linéaire à structure tridiagonale par blocs de taille $(N+2) \times (N+2)$. Notons enfin que si l'on adoptait une condition

CFL ≤ 1 construite sur la vitesse de convection des entropies (moins restrictive que la condition CFL construite sur les ondes acoustiques), l'évaluation des résidus (2.47)-(2.48) serait complètement explicite puisque les incréments $\Delta(S_i)^{n+1/2}$ connaîtraient une mise à jour explicite stable. Nous avons privilégié ici une convergence vers l'état stationnaire la plus rapide possible, et donc exempte de toute restriction de principe sur le nombre CFL. L'expérience numérique (voir ci-dessous) montre que la correction proposée conduit à une méthode implicite linéarisée robuste permettant d'obtenir une annulation des résidus explicites satisfaite au zéro machine.

Remarque 1 Ce qui suit n'est qu'une justification formelle de la procédure de correction que nous venons d'introduire. Une telle justification requiert des hypothèses d'existence de solutions de problèmes algébriques non linéaires puis de problèmes linéaires avec perturbation singulière dont la vérification *a priori* semble être hors de portée. Par ailleurs et même au prix de ces hypothèses, les calculs permettant d'aboutir à la technique de correction proposée sont fastidieux et le temps nous a manqué pour en faire le détail ici. Nous nous contenterons de présenter les grandes lignes conduisant à ces calculs. En considérant un paramètre de Relaxation $\lambda > 0$ fixé, recoupler l'étape de projection sur la variété d'équilibre avec l'évolution en temps du bloc hydrodynamique revient formellement à devoir résoudre un système algébrique non linéaire localement défini sur chaque cellule de calcul j par les identités vectorielles suivantes :

$$\mathbf{a}_j \Delta \overline{\mathbf{w}}_{j-1}^{n+1/2} + \mathbf{b}_j \Delta \overline{\mathbf{w}}_j^{n+1/2} + \mathbf{c}_j \Delta \overline{\mathbf{w}}_{j+1}^{n+1/2} = \overline{\mathbf{r}}(\overline{\mathbf{w}}_{j-1}^n, \overline{\mathbf{w}}_j^n, \overline{\mathbf{w}}_{j+1}^n) + \lambda \mathcal{R}_j^{n+1},$$

où le terme de Relaxation s'écrit

$$\mathcal{R}_j^{n+1} = {}^T(0,0,0,\rho_j^{n+1} p(\mathcal{T}_j^{n+1}, \{(S_i)_j^{n+1}\}_{i=1,\dots,N-1}, (\rho\mathcal{S}_{\mathcal{N}})_j^{n+1}) - (\rho\Pi)_j^{n+1}).$$

Ici et conformément à l'analyse du système de Relaxation considéré, nous avons

$$(\rho\Pi)_j^{n+1} = (\rho I)_j^{n+1} - a^2,$$

et il faut comprendre le relèvement $\mathcal{T}$ du covolume τ et le relèvement $\mathcal{S}_{\mathcal{N}}$ de l'entropie S_N comme des fonctions non linéaires de ρ_j^{n+1}, $(\rho u)_j^{n+1}$, $(\rho I)_j^{n+1}$, $(\rho\Sigma)_j^{n+1}$ et $\{(\rho S_i)_j^{n+1}\}_{i=1,\dots,N-1}$ définies dans la sous-section 1.6.2. Il aurait fallu paramétrer l'éventuelle solution de ce problème non linéaire relativement à λ, ce que nous n'avons pas fait pour simplifier les notations. De manière à contourner les non linéarités précédentes, nous proposons d'opérer une linéarisation en temps du terme de Relaxation $\mathcal{R}$ à λ fixé, et ce à l'instant t^n. Nous sommes dès lors conduit à devoir considérer un problème linéaire en l'incrément $\Delta\overline{\mathbf{w}}_\lambda^{n+1/2}$. La matrice associée à ce problème linéaire est tridiagonale par blocs 4×4, la dernière ligne de chaque bloc composant la diagonale dépendant explicitement du paramètre λ. Après division par λ, la dernière ligne de chacun des blocs constituant le système linéaire dépend donc de $1/\lambda$. Suivant la procédure de remise à l'équilibre, nous faisons tendre λ vers l'infini dans ce système linéaire avec perturbation singulière. Des calculs montrent que localement sur chaque cellule de calcul, la ligne où figure le résidu explicite en ρI (lui même divisé par λ) dégénère pour conduire à la relation algébrique (2.45) fixant l'incrément $\Delta\rho I_j^{n+1/2}$ en fonction des incréments des autres variables sur la même cellule. Dès lors, la dernière colonne de chaque bloc liant les incréments en $j-1$, j et $j+1$ conduit à devoir modifier le système limite selon (2.46)-(2.47)-(2.48).

2.3 Résultats numériques

Nous considérons dans cette section l'approximation numérique des solutions stationnaires du système des équations d'Euler multi-pression en deux dimensions d'espace. L'objectif consiste à tester la stabilité de l'implicitation proposée dans les sections précédentes. Le système (1.1) s'écrit sous la forme :

$$\begin{cases} \partial_t\rho + \nabla\cdot\rho\omega = 0, \\ \partial_t\rho\omega + \nabla\cdot\rho\omega\otimes\omega + \nabla\sum_{i=1}^{N} p_i = 0, \\ \partial_t\rho E + \nabla\cdot(\rho E + \sum_{i=1}^{N} p_i)\omega = 0, \\ \partial_t\rho S_i + \nabla\cdot\rho S_i\omega = 0, \quad i = 1, ..., N-1, \end{cases} \tag{2.49}$$

et l'inégalité d'entropie (1.4) sous la forme :

$$\partial_t\rho S_N + \nabla\cdot\rho S_N\omega \leq 0, \tag{2.50}$$

où $\omega = (u, v) \in I\!R^2$ représente le champ de vitesse dans le repère usuel (O, Ox, Oy). Nous considérons une fois de plus des lois de pression de gaz parfaits polytropiques (les coefficients adiabatiques associés sont notés $\gamma_i > 1$, $i = 1, ..., N$) de telle sorte que la loi de fermeture sur l'énergie totale ρE s'écrit :

$$\rho E = \frac{1}{2}\rho|\omega|^2 + \sum_{i=1}^{N} \frac{p_i}{\gamma_i - 1}, \quad |\omega|^2 = u^2 + v^2. \tag{2.51}$$

L'approche que nous adoptons ici exploite très classiquement l'invariance par rotation du système (2.49). Nous nous contenterons juste de rappeler que dans ce contexte, nous avons uniquement besoin de définir une fonction flux numérique "quasi-1D" consistante avec la projection des équations du système dans une direction fixée. La fonction flux numérique dans l'autre direction se déduit simplement de la précédente sous l'action d'une rotation sur le champ de vitesse, suivie de la rotation inverse sur les composantes en impulsion de la fonction flux numérique. Il s'agit là de considérations très classiques qui valent pour la formulation explicite en temps, mais également implicite linéarisée en temps et nous ne rentrons pas plus dans les détails. Nous renvoyons le lecteur à [57] pour une présentation générale dans le contexte de méthodes explicites en temps, et par exemple à l'ouvrage [68] pour une discussion dans le cadre de méthodes implicites linéarisées en temps. Pour fixer les idées, nous donnons la projection dans la direction Ox (sans aucune restriction) du système (2.49)-(2.50) :

$$\begin{cases} \partial_t\rho + \partial_x\rho u = 0, \\ \partial_t\rho u + \partial_x(\rho u^2 + \sum_{i=1}^{N} p_i) = 0, \\ \partial_t\rho v + \partial_x\rho v u = 0, \\ \partial_t\rho E + \partial_x(\rho E + \sum_{i=1}^{N} p_i)u = 0, \\ \partial_t\rho S_i + \partial_x\rho S_i u = 0, \quad i = 1, ..., N-1, \end{cases} \tag{2.52}$$

$$\partial_t \rho S_N + \partial_x \rho S_N u \leq 0. \tag{2.53}$$

Une équation de transport supplémentaire sur la vitesse tangentielle v vient donc s'ajouter aux équations de transport sur les entropies S_i du système (1.1). L'approximation numérique de cette équation est rigoureusement identique à celle des entropies S_i.

L'ordre 2 d'approximation en espace est obtenu par une approche MUSCL (voir [57]) menée ici sur les variables primitives, à savoir la densité, le vecteur vitesse et les pressions partielles $\{p_i\}_{i=1,...,N}$. Le limiteur de pente utilisé est celui proposé par Van Albada et al. ($\phi(r) = (r^2 + r)/(1 + r^2)$ avec des notations classiques).

Afin de tester la stabilité de l'implicitation proposée, nous approchons numériquement plusieurs écoulements stationnaires, supersoniques à l'infini amont, et arrivant sur un cylindre. L'interaction d'un tel écoulement avec le cylindre donne naissance à un fort choc détaché devant celui-ci. Pour cela, nous utilisons un maillage curviligne structuré avec ni points de discrétisation sur la paroi du cylindre et nj points de discrétisation sur l'axe Ox (direction de la ligne d'arrêt). Pour chaque cas test, après avoir précisé les données correspondantes, nous présentons les iso-valeurs des champs de pression et de densité. Les distributions des pressions partielles sur la ligne de point d'arrêt seront également présentées.

Précisons enfin que nous avons utilisé les conditions aux limites classiques suivantes pour effectuer l'ensemble des simulations :

- condition de glissement sur la paroi du cylindre,
- condition de sortie supersonique en aval,
- condition de symétrie sur la ligne d'arrêt.

La prise en compte de ces conditions aux limites au niveau discret a été menée conformément au code de calculs CELHYO développé à l'ONERA. Il s'agit d'une stratégie classique reposant sur l'utilisation d'états miroirs (paroi et condition de symétrie) que nous ne rappellerons pas ici.

Expérience 1

Pour ce premier cas test, nous faisons le choix $N = 1$. Le système (2.49) correspond aux équations d'Euler habituelles. Le maillage utilisé (et représenté sur la figure 2.1) est tel que $ni = 32$ et $nj = 41$. Le coefficient adiabatique γ est quant à lui fixé à 1.2, tandis que le champ amont arrivant sur le cylindre est tel que :

Expérience 1

	ρ	u	v	p
Champ amont	7.032e-4.	4318.	0.	109.

Le nombre de Mach associé à ce calcul est environ égal à 10. Le cas test est donc sévère, comme l'illustre la figure 2.4 mettant en évidence un rapport de pression supérieur à 100 à la traversée du choc détaché composant la solution stationnaire. Dans ce contexte, l'implicitation proposée de la méthode de Relaxation s'est avérée être d'une robustesse supérieure à la méthode de Roe classique, mais inférieure à une méthode de Van Leer linéairement implicitée. La robustesse est ici évaluée relativement à une stratégie de montée en nombre CFL au cours des itérations. Pour une stratégie donnée, nous avons cependant observé pour ce cas test une convergence plus rapide vers la solution stationnaire pour la méthode de Relaxation que pour la méthode de Van Leer. Plus que la robustesse, c'est cette dernière propriété qui nous est apparue importante. Cette propriété est illustrée dans le tableau ci-dessous qui compare à pas de temps donné le nombre d'itérations et le temps CPU nécessaires pour obtenir des résidus explicites L^2 de l'ordre de 10^{-8}. Nous donnons à titre indicatif les performances de la méthode de Relaxation complètement explicite en temps (certes sous une condition CFL incomparable mais en l'absence de toute procédure d'inversion de système linéaire). Nous nous sommes restreints ici à un critère de convergence de 10^{-8} pour rendre ce calcul explicite d'un coût non rédhibitoire.

	Itérations	Temps CPU	Temps CPU / itération
Relaxation explicite	11900	19642.176	1.65
Relaxation implicite	655	3791.925	5.79
Van Leer implicite	725	4950.689	6.828

Nous tenons à préciser que nous ne sommes pas en mesure d'affirmer laquelle des méthodes de Relaxation et de Van Leer est la plus performante en temps calculs. Le nombre de simulations à mener pour de telles conclusions est trop important. Notre propos est simplement d'énoncer les tendances observées pour ce cas test difficile, à savoir : une utilisation de pas de temps plus élevés pour la méthode de Van Leer et une convergence vers l'état stationnaire plus rapide pour la méthode de Relaxation.

A titre d'illustration, nous testons maintenant la méthode en considérant plusieurs pressions indépendantes (*i.e.* $N > 1$). Le code CELHYO de l'ONERA dans lequel la méthode par Relaxation a été implémentée ne traitant pas à l'origine plusieurs lois de pression indépendantes, aucune comparaison de type stabilité ou vitesse de convergence avec une autre méthode n'a pu être effectuée sur ces cas tests.
Nous n'avons pas rencontré de difficulté particulière dans la réalisation des simulations numériques proposées. Soulignons que celles-ci sont relativement sévères puisqu'elles s'accompagnent de taux de compression élevés. Le critère de convergence vers la solution stationnaire a été fixé à 10^{-12} dans tous les cas.

Expérience 2

Pour ce problème, nous considérons $N = 2$ pressions partielles p_1 et p_2 associées respectivement aux coefficients adiabatiques $\gamma_1 = 1.3$ et $\gamma_2 = 1.2$. Le maillage utilisé est le même que pour la simulation précédente, et le champ amont considéré est le suivant :

Expérience 2

	ρ	u	v	p_1	p_2
Champ amont	7.032e-4.	4318.	0.	36.	70.

Le nombre de Mach est toujours environ égal à 10.

Expérience 3

Pour ce cas test, nous considérons trois pressions partielles. Les coefficients adiabatiques vérifient $\gamma_1 = 1.4$, $\gamma_2 = 1.3$ et $\gamma_3 = 1.2$. Le maillage utilisé est tel que $ni = 32$ et $nj = 88$. Le champ arrivant sur le cylindre est le suivant :

Expérience 3

	ρ	u	v	p_1	p_2	p_3
Champ amont	4.018e-4.	964.	0.	10.	26.	40.

Pour ce calcul, le nombre de Mach est approximativement 2.

Présentation des résultats

L'examen des résultats exhibés ci-après indique que les solutions discrètes proposées ne présentent pas de pathologie particulière. La structure de l'écoulement dans le cas de plusieurs pressions indépendantes est largement conforme à l'intuition.

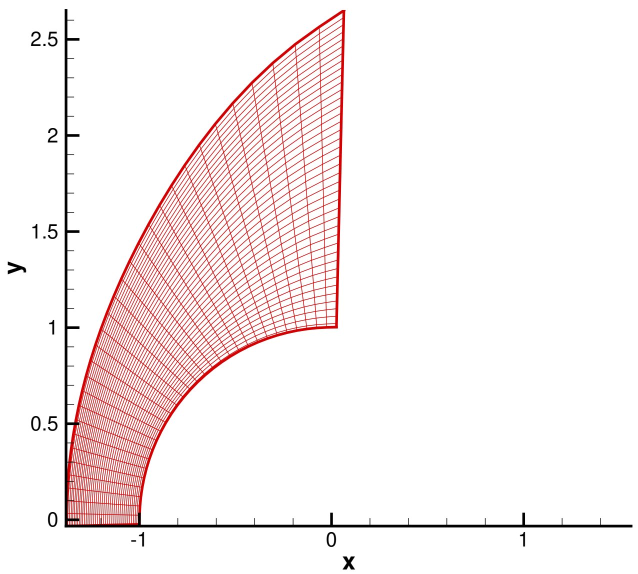

FIG. 2.1 – *Expérience 1. Maillage 32×41*

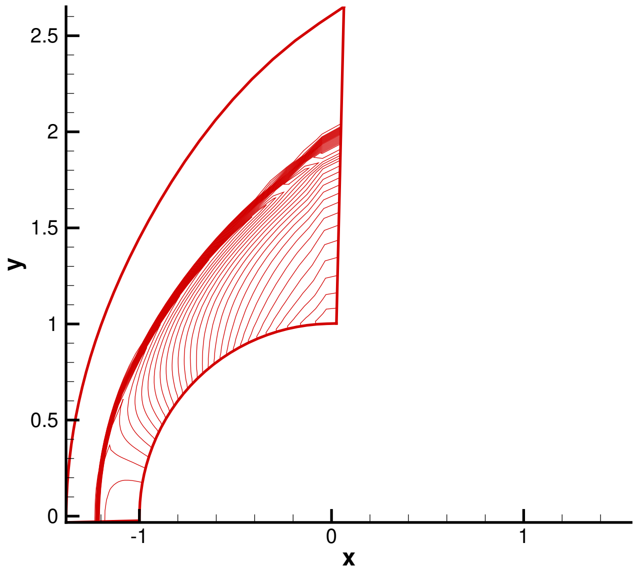

Fig. 2.2 – *Expérience 1. Densité : iso-valeurs*

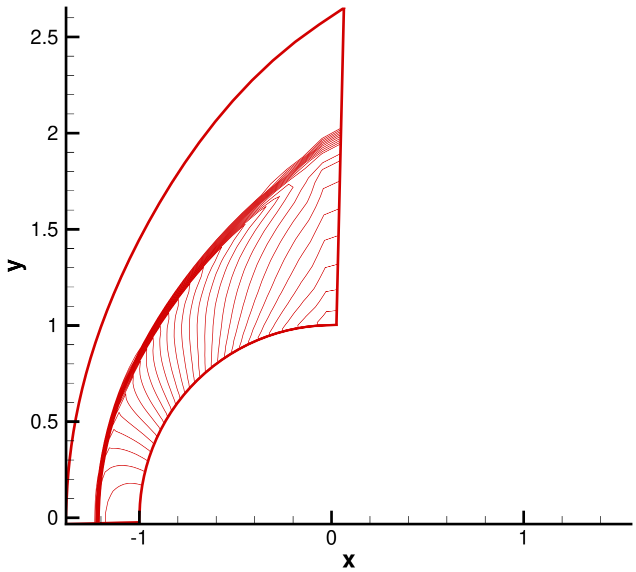

Fig. 2.3 – *Expérience 1. Pression : iso-valeurs*

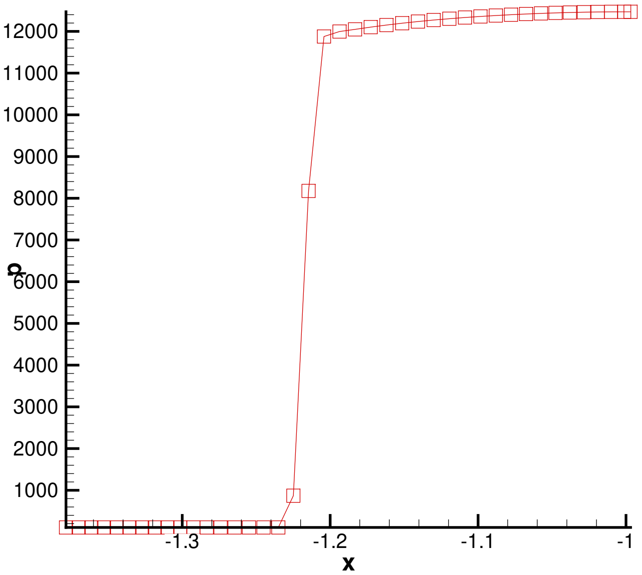

Fig. 2.4 – *Expérience 1. Pression sur la ligne d'arrêt*

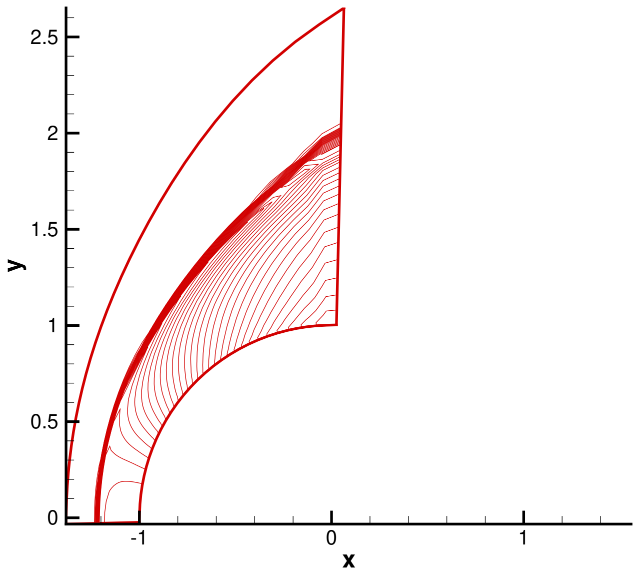

Fig. 2.5 – *Expérience 2. Densité : iso-valeurs*

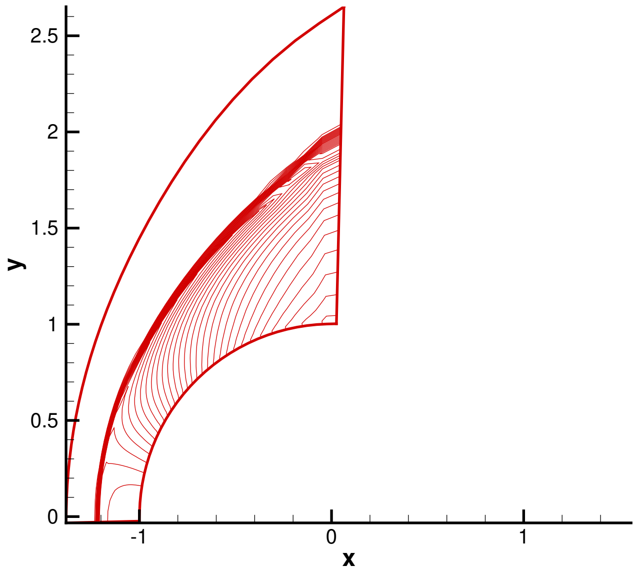

Fig. 2.6 – *Expérience 2. Pression p_1 : iso-valeurs*

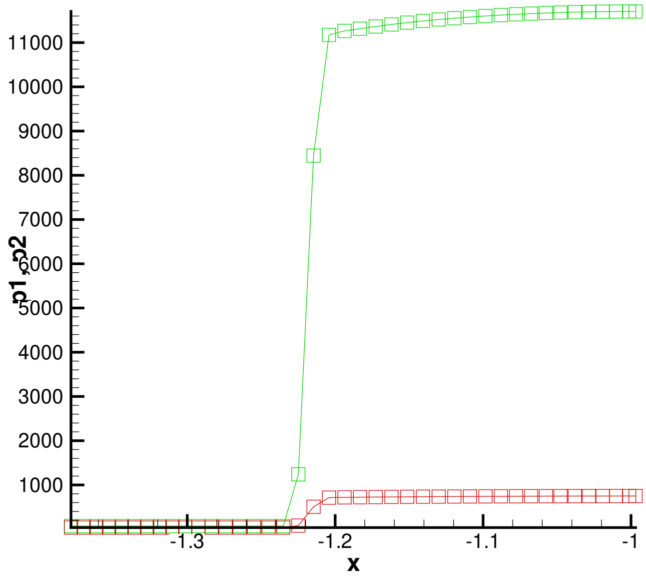

Fig. 2.7 – *Expérience 2. Pressions sur la ligne d'arrêt*

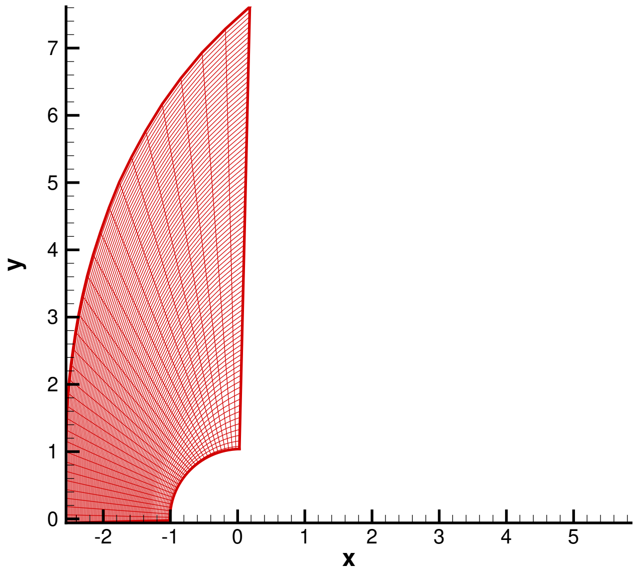

Fig. 2.8 – *Expérience 3. Maillage 32×88*

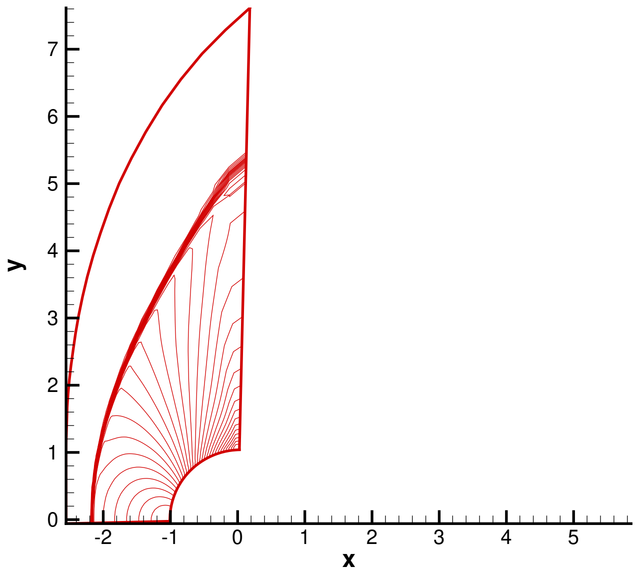

Fig. 2.9 – *Expérience 3. Densité : iso-valeurs*

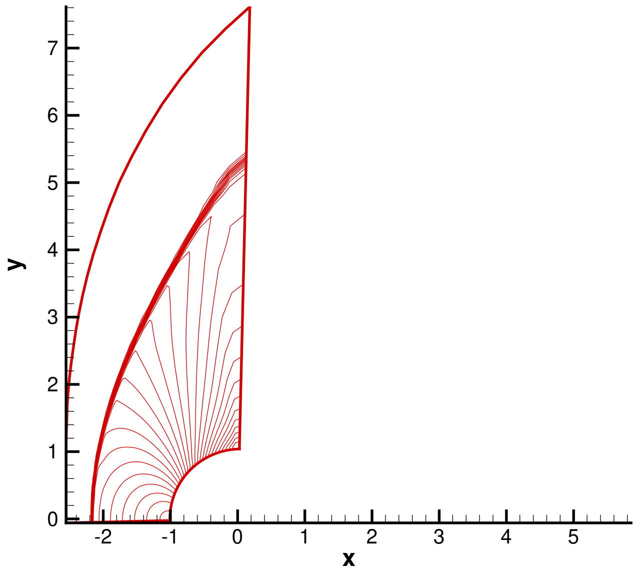

FIG. 2.10 – *Expérience 3. Pression p_2: iso-valeurs*

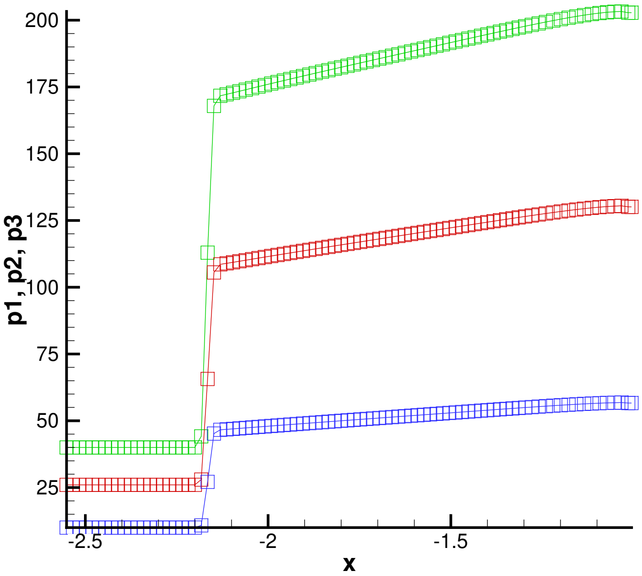

FIG. 2.11 – *Expérience 3. Pressions sur la ligne d'arrêt*