Etude de la solution approchée de problèmes quasilinéaires et analyse d'un problème en théorie du signal

Mohamed Amrani

Centre d'Analyse Non Linéaire
Université de Metz

Thèse présentée pour l'obtention du
Doctorat de l'Université de Metz
en Mathématiques

spécialité : Analyse non linéaire et numérique,

par Mr **Mohamed AMRANI.**

Titre de la thèse :

Etude de la solution approchée de problèmes quasilinéaires et analyse d'un problème en théorie du signal

Soutenue publiquement le 27.11.1995.

Devant le ju y composé de :

A. BECHLER	:	Directeur des développements exploratoires. Société LANDIS & GYR - Energy Management (France). Examinateur.
B. BRIGHI	:	Maître de Conférences à l'Université de Metz. Examinateur.
M. CHIPOT	:	Professeur à l'Université de Metz. Directeur de thèse.
J.M. CROLET	:	Professeur à l'Université de Franche-Comté, Besançon. Rapporteur.
B.P. RAO	:	Professeur à l'Université Louis Pasteur, Strasbourg. Rapporteur.
I. SHAFRIR	:	Professeur à l'Université de Metz. Examinateur.

Centre d'Analyse Non Linéaire
Université de Metz

Thèse présentée pour l'obtention du
Doctorat de l'Université de Metz
en Mathématiques

spécialité : Analyse non linéaire et numérique,
par Mr **Mohamed AMRANI.**

Titre de la thèse :

Etude de la solution approchée de problèmes quasilinéaires et analyse d'un problème en théorie du signal

Soutenue publiquement le 27.11.1995.

Devant le jury composé de :

A. BECHLER : Directeur des développements exploratoires.
Société LANDIS & GYR - Energy Management (France).
Examinateur.

B. BRIGHI : Maître de Conférences à l'Université de Metz.
Examinateur.

M. CHIPOT : Professeur à l'Université de Metz.
Directeur de thèse.

J.M. CROLET : Professeur à l'Université de Franche-Comté, Besançon.
Rapporteur.

B.P. RAO : Professeur à l'Université Louis Pasteur, Strasbourg.
Rapporteur.

I. SHAFRIR : Professeur à l'Université de Metz.
Examinateur.

REMERCIEMENTS

Le travail présenté dans ce mémoire, a été réalisé au département de Mathématiques de l'Université de Metz, sous la direction du Professeur M. Chipot.
Je tiens à exprimer ma gratitude et une grande reconnaissance à M. Chipot qui m'a toujours soutenu depuis que j'ai commencé cette thèse et dont j'ai pu profiter des compétences scientifiques.

Je voudrais remercier A.Bechler pour ses invitations à la société Landis & Gyr, son accueil chaleureux, ainsi que pour les discussions fructueuses que j'ai pu avoir avec lui, tant sur le plan scientifique que sur le plan industriel.

Je remercie J.M. Crolet et B.P. Rao qui m'ont fait l'honneur d'être les rapporteurs de ce travail, et I. Shafrir et B. Brighi qui ont accepté de participer à ce jury.

Enfin je remercie mes parents pour leur soutien moral et financier.

Table des matières

INTRODUCTION GENERALE

Dans la première partie de ce travail, nous nous sommes proposés de compléter les résultats de N. André et de M. Chipot sur l'unicité des solutions du problème approché de l'équation quasilinéaire elliptique suivante :

$$\begin{cases} -\dfrac{\partial}{\partial x_i}(a(x,u)\dfrac{\partial u}{\partial x_i}) = f \quad dans\ \Omega \\ \qquad u \in H_0^1(\Omega) \end{cases}$$

dans laquelle Ω désigne un ouvert borné de R^n, $n \geq 1$, $a(x,u)$ est une fonction de Carathéodory satisfaisant :

$$0 < \alpha \leq a(x,u) \leq \beta, \quad \text{p.p} \quad x \in \Omega, \quad \forall u \in R$$

avec α, β deux constantes positives, et $f \in L^p(\Omega), p \geq 2$.
Plus particulièrement, en dimension 2, on a réussi à prouver unicité de la solution avec une hypothèse optimale sur les angles de la triangulation; par ailleurs on a étudié également, en dimension quelconque, la régularité du problème approché.

Dans la deuxième partie, on s'est intéressé à l'étude d'un problème industriel proposé par la société Landis & Gyr.
Ce problème consiste à trouver une technique numérique pour mesurer l'énergie d'un signal électrique, avec une erreur qui n'excède pas 0,05 %; on a réussi à donner une méthode permettant un calcul exact.

Notations

D'une maniére générale, on a utilisé les notations habituelles pour les espaces L^p, les espaces de Sobolev et leurs duaux, ainsi que pour les normes qui y sont définies.

Première partie : Etude de la solution approchée de problèmes quasilinéaires

CHAPITRE 1 :

Introduction :

Soit Ω un ouvert borné de R^n, $n \geq 1$, de frontière Γ; considérons le problème :

$$\begin{cases} -\dfrac{\partial}{\partial x_i}(a(x,u)\dfrac{\partial u}{\partial x_i}) = f \quad \text{dans } \Omega \\ \qquad u \in H_0^1(\Omega) \end{cases} \tag{1.1}$$

où $f \in H^{-1}(\Omega)$ et $a(x,u)$ est une fonction de Carathéodory satisfaisant :

$$0 < \alpha \leq a(x,u) \leq \beta, \quad \text{p.p } x \in \Omega, \quad \forall u \in R \tag{1.2}$$

avec α , β deux constantes positives,
et

$$|a(x,u) - a(x,v)| \leq C.|u-v| \ \forall u,v \in R, \quad \text{p.p } x \in \Omega \tag{1.3}$$

pour un certain $C > 0$.

On peut approcher la solution de (1.1) en utilisant une méthode simple d'éléments finis : une triangulation τ_h est définie sur $\bar{\Omega}$, i.e, $\bar{\Omega}$ est décomposé en une union de n-simplexes T d'intérieurs deux à deux disjoints, et tel que, pour un n-simplexe donné T, chacune de ses faces est soit face d'un autre simplexe, soit une partie de la frontière Γ.
On construit un analogue discret de (1.1), on suppose que $\bar{\Omega}$ est un domaine polyhédral de R^n i.e. sa frontière Γ est une union finie de $(n-1)$-simplexes fermés.
Etant donné une triangulation τ_h, on lui associe l'espace défini ci-dessous :

$V_0^h = \{u \in C^0(\bar{\Omega}); u_{/T}$ est affine sur chaque $T \in \tau_h$ et $u_{/\Gamma} = 0\}$.

Soient K_i, $1 \leq i \leq N$, les sommets intérieurs de la triangulation τ_h, et soient ϕ_i, $1 \leq i \leq N$, les fonctions de V_0^h satisfaisant

$$\phi_i(K_j) = \delta_{ij}, \quad 1 \leq i \leq N. \tag{1.4}$$

les fonctions ϕ_i ,$1 \leq i \leq N$ forment une base de V_0^h.

Le problème discret consiste à trouver une fonction $u_h \in V_0^h$ telle que:

$$\int_\Omega a(x,u_h)\nabla u_h.\nabla v dx =< f,v > \quad \forall v \in V_0^h \tag{1.5}$$

où $<,>$ est le produit de dualité entre $H^{-1}(\Omega)$ et $H_0^1(\Omega)$.

Soit un n-simplexe T de la triangulation τ_h, soient b_r, $1 \leq r \leq n+1$, ses sommets et soient λ_r, $1 \leq r \leq n+1$, les coordonnées barycentriques d'un point $x \in T$ par rapport aux points b_r.
Au n-simplexe T, on associe les paramètres :

$$h_T = \text{diamètre}(T) \ , \quad \sigma_T = \max_{r \neq s} \cos(\nabla\lambda_r, \nabla\lambda_s) \tag{1.6}$$

$$\nabla\lambda_r = (\frac{\partial\lambda_r}{\partial x_1}, \frac{\partial\lambda_r}{\partial x_2}, ..., \frac{\partial\lambda_r}{\partial x_n}) \ , \ 1 \leq r \leq n+1$$

et

$$\cos(\nabla\lambda_r, \nabla\lambda_s) = \frac{\nabla\lambda_r.\nabla\lambda_s}{|\nabla\lambda_r||\nabla\lambda_s|}$$

où . et $|.|$ sont respectivement le produit scalaire euclidien et la norme euclidienne de R^n.
A la triangulation τ_h , on associe les paramètres :

$$h = \max_{T \in \tau_h} h_T \ , \quad \sigma_h = \max_{T \in \tau_h} \sigma_T \tag{1.7}$$

et ϱ_T = le supremum des diamètres des sphères inscrites dans T (1.8)

On dit qu'une suite (τ_h) de triangulations est une famille régulière si et seulement si, quand $h \longrightarrow 0$, il existe δ indépendant de h tel que :

$$0 < \delta \leq \min_{T \in \tau_h} \frac{\varrho_T}{h_T}.$$

On suppose que :

$$\sigma_h \leq 0 \quad \forall\, h \tag{1.9}$$

ou

$$\sigma_h < 0 \quad \forall\, h. \tag{1.9'}$$

On peut donner une interprétation géométrique pour $n = 2$: cette condition sera satisfaite si et seulement si tous les angles des triangles de τ_h sont $\leq \frac{\pi}{2}$.

Soit u_h la solution approchée, alors il est connu que ([$A.C._2$])

$$\lim_{h\to 0} u_h = u \tag{1.10}$$

dans $H_0^1(\Omega)$-fort, où u est la solution de (1.1) et

$$h = \max_{T\in\tau_h} .diam(T) \tag{1.11}$$

Proposition 1.1:

On pose :

$$a_{ij}^{u_h} = \int_\Omega a(x,u_h)\nabla\phi_i.\nabla\phi_j dx \tag{1.12}$$

où u_h est la solution du problème (1.5).
alors, si (1.9) est vérifiée, on a :

$$\begin{cases} a_{ij}^{u_h} \le 0 & pour\ i\neq j,\ 1\le i,j\le N \\ \sum_{j=1}^N a_{ij}^{u_h} \ge 0\ , & 1\le i\le N. \end{cases} \tag{1.13}$$

Preuve :

Considérons un n-simplexe $T\in\tau_h$ et soit ϕ_i une fonction de base alors :
- ou $T\subset supp.\phi_i$, ce qui implique que $\phi_{i/T}=\lambda_r$,
- ou $T\not\subset supp.\phi_i$, dans ce cas $\phi_{i/T}=0$;
Pour $i\neq j$, le coefficient $a_{ij}^{u_h}$ est réduit à une somme finie d'intégrales de la forme

$$\chi_{rs} = \int_T a(x,u_h)\nabla\lambda_r.\nabla\lambda_s dx \tag{1.14}$$

et les indices r et s sont toujours différents dans la somme .

Puisque les fonctions λ_r sont linéaires et satisfont $0\le\lambda_r\le 1$, $1\le r\le n+1$, on a :

$$\chi_{rs} \le \alpha(\nabla\lambda_r.\nabla\lambda_s)mes(T)\le 0 \quad \forall r\neq s.$$

Si la condition (1.9) est satisfaite, les inégalités (1.13) sont satisfaites .
en effet :

$$\sum_{j=1}^N a_{ij}^{u_h} = \int_\Omega a(x,u_h)\nabla\phi_i.\nabla\psi dx \quad \text{où} \quad \psi=\sum_{j=1}^N\phi_j$$

Soit :

$$\tau_h^\Gamma = \{T \in \tau_h \ / \ T \cap \Gamma \neq \emptyset\} \tag{1.15}$$

alors

$$\sum_{j=1}^{N} a_{ij}^{u_h} = \int_{\tau_h^\Gamma} a(x, u_h) \nabla \phi_i . \nabla \psi dx$$

• si $supp.\phi_i \cap \tau_h^\Gamma \neq \emptyset$ alors

$$\begin{cases} \phi_{i/T} = \lambda_r \\ \text{et} \\ \psi_{/T} = 1 - \sum_{l=1}^{m} \lambda_l \end{cases}$$

avec $b_1, b_2, ..., b_m$ les sommets de T appartenant également à Γ.
par conséquent : $\nabla \phi_i . \nabla \psi \geq 0$ car $r \notin \{1, .., m\}$; il s'en suit que : $\sum_{j=1}^{N} a_{ij}^{u_h} \geq 0$.
• si $supp.\phi_i \cap \tau_h^\Gamma = \emptyset$ alors $\phi_{i/\tau_h^\Gamma} = 0$
d'où : $\sum_{j=1}^{N} a_{ij}^{u_h} = 0$.

■

Remarque et définition :

Dans le cas d'un problème linéaire, si le problème discret associé vérifie la propriété (1.13) alors il est dit de type non négatif; on pourra donc dire que notre problème non linéaire est de type non négatif.

CHAPITRE 2 :

Cas de la dimension 1 :

1. Approximation par éléments finis P_1 :

On se propose de montrer l'unicité du problème approché en utilisant comme fonctions tests les fonctions de base de V_0^h.
Dans cette partie on suppose que $\Omega = (0,1)$ et on considère une subdivision de Ω,

$$0 = x_0 < x_1 < .... < x_N < x_{N+1} = 1. \tag{2.1}$$

Par ailleurs on pose

$$h = max\ (x_i - x_{i-1}), i = 1, 2, ..., N+1. \tag{2.2}$$

Dans ce cas
$V_0^h = \{u : \Omega \to R$, continue Ω, u(0)=u(1)=0, u affine sur chaque $(x_{i-1}, x_i)\}$
Il est clair que V_0^h est un sous espace de $H_0^1(\Omega)$ de dimension N .

L'unicité de la solution approchée en dimension un a été montrée dans $[A.C._2]$.
On se propose de la montrer ici en n'utilisant comme fonctions tests que les fonctions de base de V_0^h et on améliore également l'estimation de l'intervalle sur lequel on a l'unicité.

Théorème 2.1:

Soit u_h la solution du problème (1.5); on suppose que $f \in L^1(\Omega)$, $a(x,u)$ est une fonction de Carathéodory satisfaisant (1.2) et (1.3), alors :

$$|u_h'|_\infty \leq \frac{\|f\|_1}{\alpha}. \tag{2.3}$$

Preuve :

on a :

$$\int_0^1 a(x, u_h) u_h'.\phi' dx = \int_0^1 f\phi\ dx, \quad \forall \phi \in V_0^h$$

soit $\phi = \phi_i$, $1 \leq i \leq N$ alors, en la prenant comme fonction test, on obtient :

$$\int_{x_{i-1}}^{x_{i+1}} a(x, u_h) u_h'.\phi_i' dx = \int_{x_{i-1}}^{x_{i+1}} f\phi_i dx$$

d'où :

$$\lambda_i^u u_{i,h}' - \lambda_{i+1}^u u_{i+1,h}' = f_i, i = 1, 2, ..., N$$

avec

$$\lambda_i^u = \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} a(x, u_h)dx \quad \geq \alpha, \quad f_i = \int_{x_{i-1}}^{x_{i+1}} f\phi_i dx$$

et

$$u'_{i,h} = u'_{h/[x_{i-1},x_i]}$$

où $u'_h/[x_{i-1}, x_i]$ désigne la restriction de u'_h à $[x_{i-1}, x_i]$.
Soit i_0 tel que $u'_{i_0,h} = |u'_h|_\infty > 0$, donc on a en particulier

$$\lambda_{i_0}^u |u'_h|_\infty - \lambda_{i_0+1}^u u'_{i_0+1,h} = f_{i_0}.$$

- Si $u'_{i_0+1,h} < 0$ alors on a $|u'_h|_\infty \leq \frac{\|f\|_1}{\alpha}$.
- Sinon, il existe $j_0 \neq i_0$ tel que $u'_{j_0,h} < 0$

Supposons par exemple que $j_0 > i_0$, soit la fonction test $\psi = \sum_{k=i_0}^{j_0-1} \phi_k$.
On a :

$$\lambda_{i_0}^u |u'_h|_\infty - \lambda_{j_0}^u u'_{j_0,h} = \int_{x_{i_0-1}}^{x_{j_0}} f\psi dx$$

d'où :

$$|u'_h|_\infty \leq \frac{\|f\|_1}{\alpha}$$

■

Corollaire 2.2:

Sous les hypothèses précédentes on a :

$$|u_h(x)| \leq \frac{\|f\|_1}{\alpha}. \tag{2.4}$$

Preuve :

on a : $|u_h(x)| \leq \int_0^x |u'_h(t)|dt$

ce qui fournit (2.4) grâce à (2.3).

■

Théorème 2.3:

On suppose que $f \in L^1(\Omega)$, $a(x,u)$ est une fonction de Carathéodory satisfaisant (1.2) et (1.3).
Si

$$h < \frac{\alpha^2}{C.\|f\|_1} \tag{2.5}$$

alors il existe une unique solution de (1.5).
Preuve :

Soient $u_{1,h}$, $u_{2,h}$ deux solutions du problème (1.5),
on a :

$$\int_0^1 a(x,u_{1,h})u'_{1,h}.\phi' dx = \int_0^1 a(x,u_{2,h})u'_{2,h}.\phi' dx \ , \forall \phi \in V_0^h.$$

Les fonctions $\phi_i, 1 \leq i \leq N$, forment une base de V_0^h.
On a :

$$\phi'_i(x) = \begin{cases} \frac{1}{x_i - x_{i-1}} & \text{sur } [x_{i-1}, x_i] \\ \frac{-1}{x_{i+1} - x_i} & \text{sur } [x_i, x_{i+1}]. \end{cases} \tag{2.6}$$

On obtient donc pour tout $i = 1, 2, .., N$:

$$\frac{1}{x_{i+1} - x_i} \int_{x_i}^{x_{i+1}} a(x,u_{1,h})u'_{1,h} - a(x,u_{2,h})u'_{2,h} dx$$

$$=$$

$$\frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} a(x,u_{1,h})u'_{1,h} - a(x,u_{2,h})u'_{2,h} dx = K \tag{2.7}$$

où K est une constante indépendante de i.
On suppose que K est différent de zéro, par exemple $K > 0$.
Posons

$$w_h = u_{1,h} - u_{2,h}; \quad w'_i = w'_{h/[x_{i-1},x_i]}, \quad (u'_{l,h})_i = u'_{l,h/[x_{i-1},x_i]}, \ 1 \leq i \leq N+1$$

on a :

$$K = w'_{i+1}\gamma_i + L_i, \quad 0 \leq i \leq N \tag{2.8}$$

où

$$\gamma_i = \frac{1}{x_{i+1} - x_i} \int_{x_i}^{x_{i+1}} a(x, u_{1,h}) dx \ , \ 0 \leq i \leq N \tag{2.9}$$

et

$$L_i = \frac{1}{x_{i+1} - x_i} \int_{x_i}^{x_{i+1}} [a(x, u_{1,h}) - a(x, u_{2,h})] u'_{2,h} dx \ , \ 0 \leq i \leq N. \tag{2.10}$$

D'autre part, on a aussi :

$$\begin{cases} \text{sur} & [x_0, x_1] & w_h(x) = w'_1(x - x_0) \\ \text{sur} & [x_N, x_{N+1}] & w_h(x) = w'_{N+1}(x - x_{N+1}). \end{cases}$$

Sur chaque $[x_i, x_{i+1}]$ où : $w_h(x) = w'_{i+1}(x - x_*)$, pour un certain $x_* \in [x_i, x_{i+1}]$, on va montrer que :

$$|L_i| < |w'_{i+1}| \gamma_i \tag{2.11}$$

si h assez petit.
En effet, d'après (1.3), (2.10) :

$$|L_i| \leq C |w'_{i+1}| . |(u'_{2,h})_{i+1}| \frac{1}{x_{i+1} - x_i} \int_{x_i}^{x_{i+1}} |x - x_*| dx$$

d'où

$$|L_i| \leq C |w'_{i+1}| (x_{i+1} - x_i) |(u'_{2,h})_{i+1}|.$$

Ceci implique que

$$|L_i| < C |w'_{i+1}| h \frac{\|f\|_1}{\alpha}$$

et

$$|L_i| \leq \alpha C |w'_{i+1}| h \frac{\|f\|_1}{\alpha^2}$$

et finalement (2.11).
En particulier :

$$\begin{cases} w'_1 \gamma_0 + L_0 = K > 0, & |L_0| \leq |w'_1| \gamma_0 & \Longrightarrow & w'_1 > 0 \\ w'_{N+1} \gamma_N + L_N = K > 0, & |L_N| \leq |w'_{N+1}| \gamma_N & \Longrightarrow & w'_{N+1} > 0. \end{cases}$$

Il existe donc $x_* \in]x_{i_0}, x_{i_0+1}[\ \subset \ ]x_1, x_N[\ / \ w_h(x_*) = 0.$

Alors sur $[x_{i_0}, x_{i_0+1}]$, on a $w_h(x) = w'_{i_0+1}(x - x_*)$ et

d'après (2.8) et (2.11), on déduit que $w'_{i_0+1} > 0$ ce qui implique:

$$w_h(x_{i_0+1}) > w_h(x_{i_0}) > 0$$

on obtient une contradiction.

On peut donc affirmer que $K = 0$.
D'où :

$$\frac{w'_1}{x_1 - x_0} \int_{x_0}^{x_1} a(x, u_{1,h})dx = -L_0$$

et

$$|w_h(x_1)| \leq Ch\frac{\|f\|_1}{\alpha^2}|w_h(x_1)| \tag{2.12}$$

si (2.5) est vérifiée , on obtient une contradiction si $w_h(x_1) \neq 0$.
On conclut que $w_h(x_1) = 0$.
On prouve alors successivement que $w_h(x_i) = 0$, $i = 1, ..., N$, d'où l'unicité.

■

2. Approximation par éléments finis P_2 :

Dans cette partie on suppose également que $\Omega = (0,1)$ et on considère une subdivision de Ω,

$$0 = x_0 < x_1 < .... < x_N < x_{N+1} = 1$$

et on pose

$$h = max\ (x_i - x_{i-1}), i = 1, 2, ..., N+1.$$

On va considérer le cas où :

$V_0^h = \{u : \Omega \to R$, continue sur Ω, u(0)=u(1)=0, u est un polynôme de degré 2 sur chaque $(x_{i-1}, x_i)\}$

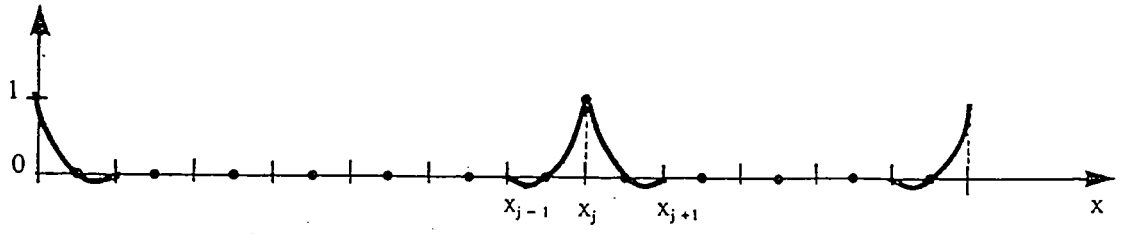
V_0^h est un sous espace vectoriel de $H^1(\Omega)$ de dimension $2N+1$, on peut paramétrer une fonction v par ses valeurs aux points $\{x_j\}$ ainsi que par ses valeurs aux milieux $\{x_{j+\frac{1}{2}}\}$ des segments $[x_j, x_{j+1}]$.

$$v = \sum_{j=1}^{N} v(x_j)\phi_j(x) + \sum_{j=0}^{N} v(x_{j+\frac{1}{2}})\phi_{j+\frac{1}{2}}(x)$$

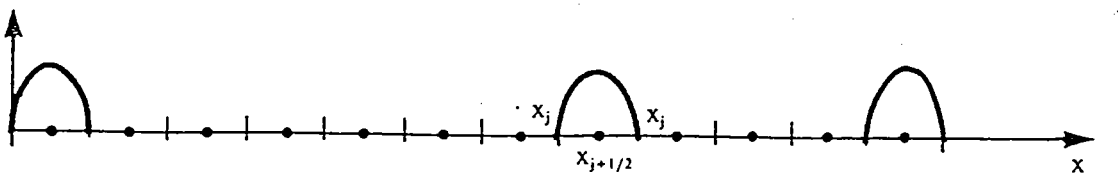
les fonctions $\{\phi_j\}$,$\{\phi_{j+\frac{1}{2}}\}$ sont les fonctions de base associées à cette paramétrisation

on représente ci-dessous le graphe des deux familles de fonctions de base :

- les fonctions de base :$\{\phi_j\}$



- les fonctions de base :$\{\phi_{j+\frac{1}{2}}\}$



Les supports des fonctions de base sont de taille $2h$ ou h la paramétrisation est locale.

Calculs préliminaires :

On considére l'ensemble des polynômes $\hat{P}_2(\hat{x})$ de degré 2 définis sur $[0,1]$.
On peut paramétrer un tel polynôme par les valeurs $\hat{P}_2(0)$, $\hat{P}_2(1)$, $\hat{P}_2(\frac{1}{2})$.

	$\hat{P}_0$	$\hat{P}_{\frac{1}{2}}$	$\hat{P}_1$
$\hat{P}_2(0)$	1	0	0
$\hat{P}_2(\frac{1}{2})$	0	1	0
$\hat{P}_2(1)$	0	0	1
	$(2\hat{x}-1)(\hat{x}-1)$	$4\hat{x}(1-\hat{x})$	$\hat{x}(2\hat{x}-1)$

On a : $\hat{P}_0'(\hat{x}) = 4\hat{x} - 3$, $\hat{P}_{\frac{1}{2}}'(\hat{x}) = 4 - 8\hat{x}$, $\hat{P}_1'(\hat{x}) = 4\hat{x} - 1$.

$|\hat{P}_0'(\hat{x})| \leq 3$, $|\hat{P}_{\frac{1}{2}}'(\hat{x})| \leq 4$, $|\hat{P}_{\frac{1}{2}}'(\hat{x})| \leq 3$.

$$\begin{aligned}\int_0^1 (\hat{P}_0'(\hat{x}))^2 d\hat{x} = \int_0^1 (4\hat{x}-3)^2 d\hat{x} &= \frac{1}{12}[(4\hat{x}-3)^3]_0^1 \\ &= \frac{1}{12}\{1-(-3)^3\} \\ &= \frac{28}{12} = \frac{7}{3},\end{aligned}$$

$$\begin{aligned}\int_0^1 (\hat{P}_1'(\hat{x}))^2 d\hat{x} = \int_0^1 (4\hat{x}-1)^2 d\hat{x} &= \frac{1}{12}[(4\hat{x}-1)^3]_0^1 \\ &= \frac{1}{12}\{3-(-1)^3\} \\ &= \frac{7}{3},\end{aligned}$$

$$\begin{aligned}\int_0^1 (\hat{P}_{\frac{1}{2}}'(\hat{x}))^2 d\hat{x} = \int_0^1 (-8\hat{x}+4)^2 d\hat{x} &= \frac{-1}{24}[(-8\hat{x}+4)^3]_0^1 \\ &= \frac{1}{24}\{4^3-(-4)^3\} \\ &= \frac{16}{3},\end{aligned}$$

$$\begin{aligned}\int_0^1 \hat{P}_0'(\hat{x}).\hat{P}_1'(\hat{x}) d\hat{x} = \int_0^1 (4\hat{x}-3)(4\hat{x}-1) d\hat{x} &= \int_0^1 16\hat{x}^2 - 16\hat{x} + 3 \; d\hat{x} \\ &= [\frac{16}{3}\hat{x}^3 - 8\hat{x}^2 + 3\hat{x}]_0^1 \\ &= \frac{1}{3},\end{aligned}$$

$$\begin{aligned}\int_0^1 \hat{P}'_{\frac{1}{2}}(\hat{x}).\hat{P}'_1(\hat{x})d\hat{x} &= \int_0^1 (4\hat{x}-1)(4-8\hat{x})d\hat{x} \\ &= \int_0^1 (-32\hat{x}^2+24\hat{x}-4)d\hat{x} \\ &= [-\frac{32}{3}\hat{x}^3+12\hat{x}^2-4\hat{x}]_0^1 \\ &= -\frac{8}{3},\end{aligned}$$

$$\begin{aligned}\int_0^1 \hat{P}'_{\frac{1}{2}}(\hat{x}).\hat{P}'_0(\hat{x})d\hat{x} &= \int_0^1 (4\hat{x}-3)(4-8\hat{x})d\hat{x} \\ &= \int_0^1 (-32\hat{x}^2+40\hat{x}-12)d\hat{x} \\ &= [-\frac{32}{3}\hat{x}^3+20\hat{x}^2-12\hat{x}]_0^1 \\ &= -\frac{8}{3}.\end{aligned}$$

On a alors :

$$\begin{cases} \phi_j(x) = \hat{P}_0(\frac{x-x_j}{h}) & si\ x \in [x_j, x_{j+1}] \\ \phi_j(x) = \hat{P}_1(\frac{x-x_{j-1}}{h}) & si\ x \in [x_{j-1}, x_j] \\ \quad \& \\ \phi_{j+\frac{1}{2}}(x) = \hat{P}_{\frac{1}{2}}(\frac{x-x_j}{h}) & si\ x \in [x_j, x_{j+1}]. \end{cases}$$

On se propose d'étudier l'unicité du problème approché (1.5), on impose les mêmes hypothèses que dans le cas de l'approximation par éléments finis P_1, et on suppose de plus que $a(x,u)$ est lipchitzienne en x, i.e :

$$|a(x,u)-a(y,u)| \leq K.|x-y| \ \forall x,y \in \Omega,\ \forall\, u \in R. \tag{2.13}$$

Notation :

$$\begin{cases} \lambda_{ii}^k = \int_0^1 a(x_k+th, u_h(x_k+th))(\hat{P}'_i(t))^2 dt \\ \quad et \\ \lambda_{ij}^k = \int_0^1 a(x_k+th, u_h(x_k+th))\hat{P}'_i(t).\hat{P}'_j(t)dt \end{cases}$$

lemme 2.4 :

On suppose que, $a(x,u)$ est une fonction de Carathéodory satisfaisant (1.2), (1.3) et (2.13). Si

$$h^{\frac{1}{2}} < \frac{1}{144} inf(\frac{1}{\beta}, 1).\frac{\alpha^2}{sup(\alpha K, C\|f\|_*)} \tag{2.14}$$

alors

$$\begin{cases} |\lambda_{ij}^k| > \frac{\alpha}{6} \\ |\lambda_{ii}^k||\lambda_{jj}^k| - |\lambda_{ij}^k|^2 > \frac{24\alpha}{9}. \end{cases} \tag{2.15}$$

Preuve :

Pour $i \neq j$ on a :

$$\lambda_{ij}^k = \int_0^1 a(th + x_k, u_h(th + x_k))\hat{P}_i'(t).\hat{P}_j'(t)dt$$

$$\lambda_{ij}^k = \int_0^1 [a(th + x_k, u_h(th + x_k)) - a(x_k, u_h(x_k))]\hat{P}_i'(t).\hat{P}_j'(t) + a(x_k, u_h(x_k)) \int_0^1 \hat{P}_i'(t).\hat{P}_j'(t)dt.$$

D'où

$$\lambda_{ij}^k = a(x_k, u_h(x_k))c_{ij} + \nu(h)$$

avec

$$\begin{cases} \nu(h) = \int_0^1 [a(th + x_k, u_h(th + x_k)) - a(x_k, u_h(x_k))]\hat{P}_i'(t).\hat{P}_j'(t)dt \\ \text{et} \\ c_{ij} = \int_0^1 \hat{P}_i'(t).\hat{P}_j'(t)dt. \end{cases}$$

On a :

$$|\nu(h)| \leq c \int_0^1 |a(th + x_k, u_h(th + x_k)) - a(x_k, u_h(x_k))|dt$$

où $c = |\hat{P}_i'|_\infty.|\hat{P}_j'|_\infty \leq 12$.
Or

$$|a(th + x_k, u_h(th + x_k)) - a(x_k, u_h(x_k))| \leq$$
$$|a(th + x_k, u_h(th + x_k)) - a(th + x_k, u_h(x_k))| + |a(th + x_k, u_h(x_k)) - a(x_k, u_h(x_k))|$$

avec

$$|a(th + x_k, u_h(x_k)) - a(x_k, u_h(x_k))| \leq Kh$$
$$|a(th + x_k, u_h(th + x_k)) - a(th + x_k, u_h(x_k))| \leq C|u_h(th + x_k) - u_h(x_k)|$$

et

$$|u_h(th+x_k)-u_h(x_k)| \leq \int_{x_k}^{th+x_k} |u_h'(\xi)|d\xi \leq h^{\frac{1}{2}}|u_h'|_2 \leq \frac{\|f\|_*}{\alpha}h^{\frac{1}{2}}.$$

D'où, puisque $h<1$,

$$|\nu(h)| \leq c(C\frac{\|f\|_*}{\alpha}h^{\frac{1}{2}}+Kh) \leq 24\ sup(K, C\frac{\|f\|_*}{\alpha}).h^{\frac{1}{2}} \qquad (2.16).$$

Il vient alors :

$$\begin{aligned}|\lambda_{ij}^k| &\geq \alpha|c_{ij}|-|\nu(h)| \\ &\geq \frac{\alpha}{3}-|\nu(h)|\end{aligned}$$

et si h est tel que :

$$|\nu(h)| \leq \frac{\alpha}{6} \qquad (2.17)$$

alors on aura :

$$|\lambda_{ij}^k| \geq \frac{\alpha}{6}.$$

Pour avoir (2.17), il suffit de prendre :

$$h^{\frac{1}{2}} < \frac{\alpha^2}{144\ sup(\alpha K, C\|f\|_*)}.$$

• On pose :

$$Z = |\lambda_{ii}^k||\lambda_{jj}^k| - |\lambda_{ij}^k|^2.$$

On a :

$$\begin{aligned}|\lambda_{ii}^k| &= |a(x_k,u_h(x_k))c_{ii}+\nu_1(h)|, \\ |\lambda_{jj}^k| &= |a(x_k,u_h(x_k))c_{jj}+\nu_2(h)|, \\ |\lambda_{ij}^k| &= |a(x_k,u_h(x_k))c_{ij}+\nu(h)|,\end{aligned}$$

avec des notations évidentes pour ν_1 et ν_2.
On a :

$$\begin{aligned}|\lambda_{ii}^k||\lambda_{jj}^k| &= |a^2(x_k,u_h(x_k))c_{ii}c_{jj}+a(x_k,u_h(x_k))(c_{ii}\nu_2(h)+c_{jj}\nu_1(h))+v_1(h)\nu_2(h)| \\ &\geq a^2(x_k,u_h(x_k))c_{ii}c_{jj}-a(x_k,u_h(x_k))|c_{ii}\nu_2(h)+c_{jj}\nu_1(h)|-|v_1(h)\nu_2(h)|\end{aligned}$$

$$|\lambda_{ij}^k|^2 = a^2(x_k,u_h(x_k))c_{ij}^2+2a(x_k,u_h(x_k))c_{ij}\nu(h)+\nu^2(h)$$

D'où :

$$Z \geq a^2(x_k, u_h(x_k))(c_{ii}c_{jj} - c_{ij}^2) - \theta(h)$$

où

$$\theta(h) = a(x_k, u_h(x_k))\{|\nu_1(h)|c_{jj} + |\nu_2(h)|c_{ii} + 2|\nu(h)||c_{ij}|\} + \nu^2(h).$$

On remarque que pour $i \neq j$:

$$\begin{aligned} c_{ii}c_{jj} - c_{ij}^2 = c_{00}c_{11} - c_{01}^2 &= \frac{49}{9} - \frac{1}{9} = \frac{48}{9} \text{ pour } (i,j) = (0,1) \\ = c_{00}c_{\frac{1}{2}\frac{1}{2}} - c_{0\frac{1}{2}}^2 &= -\frac{7}{3}\frac{16}{3} - \frac{64}{9} = \frac{48}{9} \text{pour } (i,j) = (0,\frac{1}{2}) \\ = c_{11}c_{\frac{1}{2}\frac{1}{2}} - c_{1\frac{1}{2}}^2 &= \frac{7}{3}\frac{16}{3} - \frac{64}{9} = \frac{48}{9} \text{ pour}(i,j) = (1,\frac{1}{2}). \end{aligned}$$

D'où

$$Z \geq \frac{48\alpha}{9} - \theta(h).$$

Or

$$\begin{aligned} \theta(h) &\leq 16\beta.sup(\nu_1(h), \nu_2(h), \nu(h)) + \nu^2(h) \\ &\leq 384\beta.sup(K, C\frac{\|f\|_*}{\alpha})h^{\frac{1}{2}} \end{aligned}$$

d'après (2.16).
si h est tel que :

$$|\theta(h)| \leq \frac{24\alpha}{9} \tag{2.18}$$

alors on aura :

$$Z \geq \frac{24\alpha}{9}$$

pour avoir (2.18), il suffit de prendre :

$$h^{\frac{1}{2}} < \frac{\alpha^2}{144\beta.\ sup(\alpha K, C\|f\|_*)}$$

■

Remarque :

Avec les mêmes hypothèses que le lemme précédent, on a, pour h vérifiant (2.14):

$$\lambda_{01} > 0, \quad \lambda_{1\frac{1}{2}} < 0, \quad \lambda_{0\frac{1}{2}} < 0.$$

Preuve :

$$\lambda_{01} = \frac{1}{3}a(x_k, u_h(x_k)) + \nu_1(h)$$
$$\lambda_{0\frac{1}{2}} = -\frac{8}{3}a(x_k, u_h(x_k)) + \nu_2(h)$$
$$\lambda_{1\frac{1}{2}} = -\frac{8}{3}a(x_k, u_h(x_k)) + \nu_3(h).$$

D'après ces expressions, on voit facilement que, pour h suffisamment petit, on a bien les signes de l'énoncé.

■

Etudions l'Unicité du problème approché (1.5) :

Théorème 2.5 :

On suppose que $f \in H^{-1}(\Omega)$, $a(x,u)$ est une fonction de Carathéodory satisfaisant (1.2), (1.3) et (2.13), si h est suffisamment petit i.e.satisfait (2.14), alors il existe une unique solution de (1.5).

Preuve :

Soient $u_{1,h}$, $u_{2,h}$ deux solutions du problème (1.5),
on a :

$$\int_0^1 a(x, u_{1,h})u'_{1,h}.\phi' dx = \int_0^1 a(x, u_{2,h})u'_{2,h}.\phi' dx \ , \forall \phi \in V_0^h$$

d'où

$$X = \int_0^1 a(x, u_{1,h})w'_h.\phi' dx = \int_0^1 [a(x, u_{2,h}) - a(x, u_{1,h})]u'_{2,h}.\phi' dx \ , \forall \phi \in V_0^h.$$

On en déduit

$$X \leq C. \int_0^1 |w_h||u'_{2,h}||\phi'| dx.$$

On va construire une fonction test telle que sur chaque intervalle $[x_j, x_{j+1}]$ où $\phi' \neq 0$ on aura :

$$\int_{x_j}^{x_{j+1}} a(x, u_{1,h})w'_h.\phi' dx \geq \nu(|w_h(x_j)| + |w_h(x_{j+\frac{1}{2}})| + |w_h(x_{j+1})|) \qquad (2.19)$$

Soit ϕ la fonction test suivante :

$$\begin{cases} \phi(x_j) & = \begin{cases} 1 \quad si \;\; w_h(x_j) > 0 \\ 0 \qquad sinon \end{cases} \\ \phi(x_{j+\frac{1}{2}}) & = \quad c_j \; \grave{a} \; determiner \end{cases}$$

- si $w_h(x_j) > 0$ et $w_h(x_{j+1}) > 0$ on choisit : $\phi(x_{j+\frac{1}{2}}) = 1$
- si $w_h(x_j) \leq 0$ et $w_h(x_{j+1}) \leq 0$ on choisit : $\phi(x_{j+\frac{1}{2}}) = 0$
- si $w_h(x_j) \leq 0$ et $w_h(x_{j+1}) > 0$ ou si $w_h(x_j) > 0$ et $w_h(x_{j+1}) \leq 0$

alors :

$$\phi(x_{j+\frac{1}{2}}) = c_j$$

où c_j est à déterminer, de telle manière à avoir (2.19).
On aura donc $\phi' = 0$ sur chaque intervalle $[x_j, x_{j+1}]$ où $w_h(x_j)w_h(x_{j+1}) > 0$.
D'où :

$$X = \sum_j \int_{x_j}^{x_{j+1}} a(x, u_{1,h}) w_h' . \phi' dx$$

où la somme est prise sur les j tels que $w_h(x_j).w_h(x_{j+1}) \leq 0$

Sur un tel intervalle : par le changement de variable $t = \frac{x - x_j}{h}$, on peut travailler sur $[0,1]$ avec l'hypothèse :

$$\tilde{w}_h(0).\tilde{w}_h(1) \leq 0$$

où $\tilde{w}_h(t) = w_h(x_j + th)$ pour $t \in [0,1]$. On pose : $\tilde{\phi}(t) = \phi(x_j + th)$.
Supposons par exemple que :

$$\tilde{w}_h(0) > 0 \;\; , \;\; \tilde{w}_h(1) \leq 0$$

Il y trois cas :

(i) premier cas :

$$\tilde{w}_h(\frac{1}{2}) < 0.$$

d'où :

$$\tilde{\phi}(0) = 1 \;\; , \;\; \tilde{\phi}(1) = 0 \;\; , \; et \;\; \tilde{\phi}(\frac{1}{2}) = c$$

où c est à déterminer.

On a alors :

$$\tilde{w}_h(t) = w_h(x_j)\hat{P}_0(t) + w_h(x_{j+\frac{1}{2}})\hat{P}_{\frac{1}{2}}(t) + w_h(x_{j+1})\hat{P}_1(t)$$

et

$$\tilde{\phi}(t) = \hat{P}_0(t) + c.\hat{P}_{\frac{1}{2}}(t)$$

on choisit c tel que :

$$\begin{cases} A = \int_0^1 a(.,.)\hat{P}_0'(t)(\hat{P}_0'(t) + c.\hat{P}_{\frac{1}{2}}'(t))dt > 0 \\ B = \int_0^1 a(.,.)\hat{P}_{\frac{1}{2}}'(t)(\hat{P}_0'(t) + c.\hat{P}_{\frac{1}{2}}'(t))dt < 0 \\ D = \int_0^1 a(.,.)\hat{P}_1'(t)(\hat{P}_0'(t) + c.\hat{P}_{\frac{1}{2}}'(t))dt < 0 \end{cases}$$

On a ainsi :

$$\int_{x_j}^{x_{j+1}} a(x, u_{1,h})w_h'.\phi' dx = |A||w_h(x_j)| + |B||w_h(x_{j+\frac{1}{2}})| + |C||w_h(x_{j+1})|$$

Sachant que :

$$\hat{P}_0(\hat{x}) + \hat{P}_{\frac{1}{2}}(\hat{x}) + \hat{P}_1(\hat{x}) = 1$$

et par conséquent :

$$\hat{P}_0'(\hat{x}) + \hat{P}_{\frac{1}{2}}'(\hat{x}) + \hat{P}_1'(\hat{x}) = 0$$

on déduit alors :

$$(E)\begin{cases} \lambda_{00} + \lambda_{\frac{1}{2}0} + \lambda_{10} = 0 & [E-1] \\ \lambda_{0\frac{1}{2}} + \lambda_{\frac{1}{2}\frac{1}{2}} + \lambda_{1\frac{1}{2}} = 0 & [E-2] \\ \lambda_{01} + \lambda_{\frac{1}{2}1} + \lambda_{11} = 0 & [E-3] \end{cases}$$

et on cherche c telle que :

$$(S)\begin{cases} \lambda_{00} + c\lambda_{0\frac{1}{2}} > 0 & [S-1] \\ \lambda_{0\frac{1}{2}} + c\lambda_{\frac{1}{2}\frac{1}{2}} < 0 & [S-2] \\ \lambda_{01} + c\lambda_{1\frac{1}{2}} < 0 & [S-3]. \end{cases}$$

Or

$[S-1] \Longleftrightarrow c\lambda_{0\frac{1}{2}} > -\lambda_{00}, \Longleftrightarrow c < \frac{|\lambda_{00}|}{|\lambda_{0\frac{1}{2}}|},$

$[S-2] \Longleftrightarrow c|\lambda_{\frac{1}{2}\frac{1}{2}}| < -\lambda_{0\frac{1}{2}}, \Longleftrightarrow c < \frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|},$

$[S-3] \Longleftrightarrow c|\lambda_{1\frac{1}{2}}| > |\lambda_{01}|, \Longleftrightarrow c > \frac{|\lambda_{01}|}{|\lambda_{1\frac{1}{2}}|}.$

Finalement, on choisit c tel que :

$$\frac{|\lambda_{01}|}{|\lambda_{1\frac{1}{2}}|} < c < min(\frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|}, \frac{|\lambda_{00}|}{|\lambda_{0\frac{1}{2}}|})$$

ou encore :

$$\frac{|\lambda_{01}|}{|\lambda_{1\frac{1}{2}}|} < c < \frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|}$$

car $\frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|} < \frac{|\lambda_{00}|}{|\lambda_{0\frac{1}{2}}|}$.

Par ailleurs montrons que :

$$\frac{|\lambda_{01}|}{|\lambda_{1\frac{1}{2}}|} < \frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|}$$

en effet :

$$\begin{cases} [E-3] \implies |\lambda_{01}| + |\lambda_{11}| = |\lambda_{1\frac{1}{2}}| \implies \frac{|\lambda_{01}|}{|\lambda_{1\frac{1}{2}}|} = 1 - \frac{|\lambda_{11}|}{|\lambda_{1\frac{1}{2}}|} \\ [E-2] \implies |\lambda_{1\frac{1}{2}}| + |\lambda_{0\frac{1}{2}}| = |\lambda_{\frac{1}{2}\frac{1}{2}}| \implies \frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|} = 1 - \frac{|\lambda_{1\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|} \end{cases} \tag{2.20}$$

et comme $|\lambda_{1\frac{1}{2}}|^2 < |\lambda_{11}||\lambda_{\frac{1}{2}\frac{1}{2}}|$, alors c existe.

On prend :

$$c = \frac{1}{2}(\frac{|\lambda_{01}|}{|\lambda_{1\frac{1}{2}}|} + \frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|}).$$

On a alors :

$$\begin{aligned} A &= |\lambda_{00}| - c|\lambda_{0\frac{1}{2}}| \\ &= |\lambda_{0\frac{1}{2}}|(\frac{|\lambda_{00}|}{|\lambda_{0\frac{1}{2}}|} - c) \\ A &\geq |\lambda_{0\frac{1}{2}}|(\frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|} - c), \text{d'après l'inégalité de Cauchy-Schwartz} \\ &\geq \frac{1}{2}|\lambda_{0\frac{1}{2}}|(\frac{|\lambda_{11}|}{|\lambda_{1\frac{1}{2}}|} - \frac{|\lambda_{1\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|}) \\ &\geq \frac{1}{2}|\lambda_{0\frac{1}{2}}|(\frac{|\lambda_{11}||\lambda_{\frac{1}{2}\frac{1}{2}}| - |\lambda_{1\frac{1}{2}}|^2}{|\lambda_{1\frac{1}{2}}||\lambda_{\frac{1}{2}\frac{1}{2}}|}) \end{aligned}$$

donc $A > C^{te}$ (voir (2.15)).

$$|B| = |\lambda_{0\frac{1}{2}}| - c|\lambda_{\frac{1}{2}\frac{1}{2}}|$$
$$= |\lambda_{\frac{1}{2}\frac{1}{2}}|(\frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|} - c)$$

$$|B| \geq \frac{1}{2}|\lambda_{\frac{1}{2}\frac{1}{2}}|(\frac{|\lambda_{11}||\lambda_{\frac{1}{2}\frac{1}{2}}| - |\lambda_{1\frac{1}{2}}|^2}{|\lambda_{1\frac{1}{2}}||\lambda_{\frac{1}{2}\frac{1}{2}}|})$$

donc $|B| > C^{te}$.

$$|D| = -|\lambda_{01}| + c|\lambda_{1\frac{1}{2}}|$$
$$= |\lambda_{1\frac{1}{2}}|(-\frac{|\lambda_{01}|}{|\lambda_{1\frac{1}{2}}|} + c)$$
$$\geq \frac{1}{2}|\lambda_{1\frac{1}{2}}|(\frac{|\lambda_{11}||\lambda_{\frac{1}{2}\frac{1}{2}}| - |\lambda_{1\frac{1}{2}}|^2}{|\lambda_{1\frac{1}{2}}||\lambda_{\frac{1}{2}\frac{1}{2}}|})$$

donc $|D| > C^{te}$.

(ii) deuxième cas :

$$\tilde{w}_h(\frac{1}{2}) > 0$$

d'où :

$$\tilde{\phi}(0) = 1, \quad \tilde{\phi}(1) = 0, \quad \tilde{\phi}(\frac{1}{2}) = c$$

où c est à déterminer.

On choisit c tel que :

$$\begin{cases} A' = \int_0^1 a(.,.)\hat{P}'_0(t)(\hat{P}'_0(t) + c.\hat{P}'_{\frac{1}{2}}(t))dt > 0 \\ B' = \int_0^1 a(.,.)\hat{P}'_{\frac{1}{2}}(t)(\hat{P}'_0(t) + c.\hat{P}'_{\frac{1}{2}}(t))dt > 0 \\ D' = \int_0^1 a(.,.)\hat{P}'_1(t)(\hat{P}'_0(t) + c.\hat{P}'_{\frac{1}{2}}(t))dt < 0 \end{cases}$$

soit encore c tel que :

$$(S') \begin{cases} \lambda_{00} + c\lambda_{0\frac{1}{2}} > 0 & [S'-1] \\ \lambda_{0\frac{1}{2}} + c\lambda_{\frac{1}{2}\frac{1}{2}} > 0 & [S'-2] \\ \lambda_{01} + c\lambda_{1\frac{1}{2}} < 0 & [S'-3]. \end{cases}$$

$[S'-1] \Longleftrightarrow c\lambda_{0\frac{1}{2}} > -\lambda_{00}, \quad \Longleftrightarrow c < \frac{|\lambda_{00}|}{|\lambda_{0\frac{1}{2}}|},$

$[S'-2] \Longleftrightarrow c|\lambda_{\frac{1}{2}\frac{1}{2}}| > -\lambda_{0\frac{1}{2}}, \quad \Longleftrightarrow c > \frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|}.$

$[S'-3] \Longleftrightarrow c|\lambda_{1\frac{1}{2}}| > |\lambda_{01}|, \quad \Longleftrightarrow c > \frac{|\lambda_{01}|}{|\lambda_{1\frac{1}{2}}|},$

et on choisit c tel que :

$$(\frac{|\lambda_{01}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|} <) \frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|} < c < \frac{|\lambda_{00}|}{|\lambda_{0\frac{1}{2}}|}.$$

On prend :

$$c = \frac{1}{2}(\frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|} + \frac{|\lambda_{00}|}{|\lambda_{0\frac{1}{2}}|}).$$

D'une manière analogue au cas (i) on a :

$$\begin{aligned} A' &= |\lambda_{00}| - c|\lambda_{0\frac{1}{2}}| \\ &= |\lambda_{0\frac{1}{2}}|(\frac{|\lambda_{00}|}{|\lambda_{0\frac{1}{2}}|} - c) \\ &\geq \frac{1}{2}|\lambda_{0\frac{1}{2}}|(\frac{|\lambda_{00}|}{|\lambda_{0\frac{1}{2}}|} - \frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|}) \\ &\geq \frac{1}{2}|\lambda_{0\frac{1}{2}}|(\frac{|\lambda_{00}||\lambda_{\frac{1}{2}\frac{1}{2}}| - |\lambda_{0\frac{1}{2}}|^2}{|\lambda_{0\frac{1}{2}}||\lambda_{\frac{1}{2}\frac{1}{2}}|}) \end{aligned}$$

donc $A' > C^{te}$.

$$\begin{aligned} |B'| &= -|\lambda_{0\frac{1}{2}}| + c|\lambda_{\frac{1}{2}\frac{1}{2}}| \\ &= |\lambda_{\frac{1}{2}\frac{1}{2}}|(-\frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|} + c) \end{aligned}$$

$$|B'| \geq \frac{1}{2}|\lambda_{\frac{1}{2}\frac{1}{2}}|(\frac{|\lambda_{00}||\lambda_{\frac{1}{2}\frac{1}{2}}| - |\lambda_{0\frac{1}{2}}|^2}{|\lambda_{0\frac{1}{2}}||\lambda_{\frac{1}{2}\frac{1}{2}}|})$$

donc $|B'| > C^{te}$.

$$\begin{aligned} |D'| &= -|\lambda_{01}| + c|\lambda_{1\frac{1}{2}}| \\ &= |\lambda_{1\frac{1}{2}}|(-\frac{|\lambda_{01}|}{|\lambda_{1\frac{1}{2}}|} + c) \\ &\geq |\lambda_{1\frac{1}{2}}|(-\frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|} + c) \\ &\geq |\lambda_{1\frac{1}{2}}|(\frac{|\lambda_{00}||\lambda_{\frac{1}{2}\frac{1}{2}}| - |\lambda_{0\frac{1}{2}}|^2}{|\lambda_{0\frac{1}{2}}||\lambda_{\frac{1}{2}\frac{1}{2}}|}) \end{aligned}$$

donc $|D'| > C^{te}$.

(iii) troisième cas :

$$\tilde{w}_h(\frac{1}{2}) = 0$$

d'où :

$$\tilde{\phi}(0) = 1, \quad \tilde{\phi}(1) = 0, \quad \tilde{\phi}(\frac{1}{2}) = c$$

où c est à déterminer. On choisit c tel que :

$$\begin{cases} A" = \int_0^1 a(.,.)\hat{P}_0'(t)(\hat{P}_0'(t) + c.\hat{P}_{\frac{1}{2}}'(t))dt > 0 \\ B" = \int_0^1 a(.,.)\hat{P}_{\frac{1}{2}}'(t)(\hat{P}_0'(t) + c.\hat{P}_{\frac{1}{2}}'(t))dt = 0 \\ D" = \int_0^1 a(.,.)\hat{P}_1'(t)(\hat{P}_0'(t) + c.\hat{P}_{\frac{1}{2}}'(t))dt < 0 \end{cases}$$

soit encore c tel que :

$$(S") \begin{cases} \lambda_{00} + c\lambda_{0\frac{1}{2}} > 0 & [S" - 1] \\ \lambda_{0\frac{1}{2}} + c\lambda_{\frac{1}{2}\frac{1}{2}} = 0 & [S" - 2] \\ \lambda_{01} + c\lambda_{1\frac{1}{2}} < 0 & [S" - 3]. \end{cases}$$

$[S" - 1] \Longleftrightarrow c\lambda_{0\frac{1}{2}} > -\lambda_{00}, \iff c < \frac{|\lambda_{00}|}{|\lambda_{0\frac{1}{2}}|},$

$[S" - 2] \Longleftrightarrow c|\lambda_{\frac{1}{2}\frac{1}{2}}| = -\lambda_{0\frac{1}{2}}, \iff c = \frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|},$

$[S" - 3] \Longleftrightarrow c|\lambda_{1\frac{1}{2}}| > |\lambda_{01}|, \iff c > \frac{|\lambda_{01}|}{|\lambda_{1\frac{1}{2}}|}.$

D'une manière analogue au cas (i) on a, d'autre part :

$$\begin{aligned} A" &= |\lambda_{00}| - c|\lambda_{0\frac{1}{2}}| \\ &= \frac{|\lambda_{00}||\lambda_{\frac{1}{2}\frac{1}{2}}| - |\lambda_{0\frac{1}{2}}|^2}{|\lambda_{0\frac{1}{2}}||\lambda_{\frac{1}{2}\frac{1}{2}}|} \end{aligned}$$

donc $A" > C^{te}$.

$$\begin{aligned} |D"| &= -|\lambda_{01}| + c|\lambda_{1\frac{1}{2}}| \\ &= |\lambda_{1\frac{1}{2}}|(-\frac{|\lambda_{01}|}{|\lambda_{1\frac{1}{2}}|} + c) \\ &= |\lambda_{1\frac{1}{2}}|(-\frac{|\lambda_{01}|}{|\lambda_{1\frac{1}{2}}|} + \frac{|\lambda_{0\frac{1}{2}}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|}) \\ &= |\lambda_{1\frac{1}{2}}|(\frac{|\lambda_{11}||\lambda_{\frac{1}{2}\frac{1}{2}}| - |\lambda_{1\frac{1}{2}}|^2}{|\lambda_{1\frac{1}{2}}||\lambda_{\frac{1}{2}\frac{1}{2}}|}) \end{aligned}$$

d'après (3.8), donc $|D"| > C^{te}$.

D'autre part dans les trois cas on a :

$$|c| < \sqrt{\frac{|\lambda_{00}|}{|\lambda_{\frac{1}{2}\frac{1}{2}}|}}.$$

D'après les calculs effectués on déduit que, pour des constantes ν, μ strictement positives :

$$\nu \sum_j s_j \leq X \leq \mu \sum_j s_j \int_{x_j}^{x_{j+1}} |u'_{2,h}| dx$$

avec

$$s_j = |w_h(x_j)| + |w_h(x_{j+\frac{1}{2}})| + |w_h(x_{j+1})|.$$

Puisque :

$$\int_{x_j}^{x_{j+1}} |u'_{2,h}| dx \leq |u'_{2,h}|_2 . h^{\frac{1}{2}} \leq \frac{\|f\|_*}{\alpha} h^{\frac{1}{2}}$$

on obtient :

$$\sum_j s_j \leq C^{te} \frac{\|f\|_*}{\alpha^2} h^{\frac{1}{2}} \sum_j s_j.$$

conclusion :
pour h suffisamment petit, on aurait une contradiction, donc $w_h \geq 0$ et changeant les rôles de u_h et v_h, on obtient $w_h = 0$, d'où l'unicité.

■

UNIQUENESS FOR THE APPROXIMATE SOLUTION OF A CLASS OF QUASILINEAR ELLIPTIC EQUATIONS

M. Amrani & M. Chipot
Centre d'Analyse Non Linéaire
Université de Metz, URA CNRS 399
Ile du Saulcy, 57045 METZ Cedex 01
(FRANCE)

Abstract : We present a new technique to prove uniqueness for the approximate solution of some class of elliptic problems when the mesh size approaches 0 and when some angles of the triangulation of the domain are allowed to exceed $\frac{\pi}{2}$.

Key words : Approximation, quasilinear elliptic equations, finite elements.

AMS Subject Classification : 35JXX, 65N30 .

Abbreviated title : Uniqueness for Approximate Solution.

1. Introduction

Let Ω be a polyhedral bounded open subset of $\mathbf{R}^n, n \geq 1$. Assume that $a(x,u)$ is a Carathéodory function satisfying

$$0 < \alpha \leq a(x,u) \leq \beta \quad \text{a.e. } x \in \Omega, \quad \forall u \in \mathbf{R} \tag{1.1}$$

where α, β are two positive constants. For $f \in H^{-1}(\Omega)$, by application of the Schauder fixed point theorem (see for instance [C.M.]), it is easy to show that there exists u solution to the problem

$$\begin{cases} -\dfrac{\partial}{\partial x_i}(a(x,u)\dfrac{\partial u}{\partial x_i}) = f & \text{in } \Omega, \\ u \in H_0^1(\Omega). \end{cases} \tag{1.2}$$

We use here the summation convention and we refer to [G.T.] or [K.S.] for the definition of the Sobolev spaces used throughout the paper.

Moreover, when a is Lipschitz continuous in u then the weak solution to (1.2) is unique. More precisely we have :

THEOREM 1.1 : Assume that for some positive constant C one has

$$|a(x,u) - a(x,v)| \leq C|u-v| \quad \forall u,\ v\ \in \mathbf{R},\ \text{a.e. } x \in \Omega \tag{1.3}$$

then the problem (1.2) has a unique solution.

We refer to [A.C.$_1$], [T.] or [G.T.] for a proof. Note that some extensions of this theorem in terms of the modulus of continuity of a are possible (see [C.M.], [A.C.$_1$], [B.K.S.], [C.C.], [M.]) and in the case where (1.3) fails then uniqueness might fail as well (see [A.C.$_1$]).

In this paper we would like to address the question of uniqueness for the approximate solution of (1.2). Let us denote by V_0^h a finite dimensional subspace of $H_0^1(\Omega)$. Then, under the above assumptions we can introduce u_h solution to

$$\begin{cases} \int_\Omega a(x,u_h)\nabla u_h \cdot \nabla v\ dx =< f,v > & \forall v \in V_0^h, \\ u_h \in V_0^h \end{cases} \tag{1.4}$$

where $< , >$ denotes the duality bracket between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$. We have :

THEOREM 1.2 : If $f \in H^{-1}(\Omega)$ and if a is a Carathéodory function satisfying (1.1), (1.3), then, there exists a solution u_h to (1.4). Moreover, if

$$\forall v \in H_0^1(\Omega), \exists v_h \in V_0^h \quad \text{such that} \quad v_h \to v \text{ in } H_0^1(\Omega) \text{ when } h \to 0 \tag{1.5}$$

then we have :

$$\lim_{h\to 0} u_h = u \tag{1.6}$$

in $H_0^1(\Omega)$-strong.

Proof : The existence part is a straight application of the Brower fixed point theorem. We refer the reader to [A.C.$_2$] for details and a proof of the convergence of u_h toward u.

Even though the limit problem has a unique solution it has been established in [A.C.$_2$] that the corresponding approximation (1.4) might fail to have a unique solution. However, if one allows the mesh size h to be small enough then uniqueness holds again. We will restrict here to the case of P_1-Lagrange finite elements in dimension 2 refering the reader to [Am], [A.] for complements and other finite element methods. Let us first precise the framework of our results.

Let Ω be a polygonal domain of $\mathbf{R}^2$ with boundary Γ. We denote by τ_h a regular triangulation of Ω (see for instance [C.]). Recall that the mesh size h is given by

$$h = \max_{K\in\tau_h} h_K \tag{1.7}$$

where h_K denotes the diameter of the triangle K.

We denote by V_0^h the finite dimensional subspace of $H_0^1(\Omega)$ defined by

$$V_0^h = \{v : \Omega \to \mathbf{R}, \text{ continuous, } v = 0 \text{ on } \Gamma, \ v|_K \in P_1 \ \forall K \in \tau_h\} \tag{1.8}$$

where P_1 denotes the space of polynomials of degree 1, $v|_K$ the restriction of v to K.

If h is small enough and if all the angles θ of the triangles of the triangulation of Ω are such that

$$0 < \delta < \theta \leq \frac{\pi}{2} - \delta \tag{1.9}$$

then it has been shown in [A.C.$_2$] that uniqueness holds for the solution to (1.4) corresponding to V_0^h given by (1.8). The goal of this note is to improve the latter assumption and to allow in particular some of the angles to exceed $\frac{\pi}{2}$. Thus, we will need a slightly stronger assumption than τ_h to be regular in the sense that we will only allow in τ_h triangles with angles θ satisfying

$$0 < \delta_1 < \theta \leq \frac{\pi}{2} - \delta_2 \tag{1.10}$$

or

$$\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2} + \delta_3 \tag{1.11}$$

where δ_1, δ_2, δ_3 are positive constants that will be precised latter on.

Under these conditions we will prove

THEOREM 1.3 : Assume that $f \in H^{-1}(\Omega)$, a is a Carathéodory function satisfying (1.1), (1.3) and that (1.10) or (1.11) holds for any angle of any triangle in τ_h. Moreover assume that

$$\beta\{\frac{4\pi}{\delta_1} - 1\} \tan\delta_3 < \frac{\alpha}{2} \tan\delta_2 \quad , \quad \beta\{\frac{\pi}{\delta_1} - 1\} \tan\delta_3 < \frac{\alpha \sin\delta_1}{\cos\delta_2}, \tag{1.12}$$

and

$$\frac{[(\frac{2\pi}{\delta_1} - 2)\beta \tan\delta_3][\frac{2\pi}{\delta_1}\beta M + \frac{\pi}{\delta_1}\beta \tan\delta_3]}{[\frac{\alpha \sin\delta_1}{\cos\delta_2} - (\frac{\pi}{\delta_1} - 1)\beta \tan\delta_3]} < \frac{\alpha}{4} \tan\delta_2 \tag{1.13}$$

where $M = \max(\frac{1}{2\tan\delta_1}, \cot\delta_1)$. Then, if h is small enough the approximated problem (1.4) has a unique solution.

Our assumptions allow to have in τ_h triangles with angles between δ and $\frac{\pi}{2} - \delta$ as in [A.C.$_2$] but also, provided the triangulation is regular, to have triangles with angles equal to $\frac{\pi}{2}$ which was excluded in [A.C.$_2$]. Indeed in this latter case $\delta_3 = 0$ and if the angles θ of a straight triangle satisfy $\theta \geq \delta$ they also satisfy, except for the straight angle,

$$\theta \leq \frac{\pi}{2} - \delta$$

so that (1.10) holds. Of course to allow straight angles is very important for the applications since many triangulations are generated by the splitting of squares. These kind of triangulation is thus in the framework of our results. Is also suitable, by (1.12), (1.13) any triangulation obtained by the splitting of parallelograms provided that their largest angles are not to far away from $\frac{\pi}{2}$ (note that for δ_1, δ_2 fixed (1.12), (1.13) hold provided δ_3 is taken small enough).

2. Preliminary lemmas

Let T be a triangle with vertices labelled by 1, 2, 3. Let θ_i be the inside angle of T at i and l_i the lengh of the side of T opposite to i. We denote by λ_i the affine function of T such that

$$\lambda_i(j) = \delta_{i,j} \quad \forall i, j = 1, 2, 3. \tag{2.1}$$

Then :

LEMMA 2.1 : Under the above assumptions we have

$$\nabla\lambda_2 \cdot \nabla\lambda_3 = -\frac{cos\theta_1}{sin^2\theta_1 l_2 l_3}, \tag{2.2}$$

$$|\nabla\lambda_2| = \frac{1}{sin\theta_1 |l_3|}. \tag{2.3}$$

Proof : Easy. See also for instance [A.C.$_2$].

■

We will label the nodes of the triangulation by roman letters i, j, k,.... A basis of V_0^h is given by the shape functions φ_i defined as the unique functions of V_0^h such that

$$\varphi_i(j) = \delta_{i,j} \quad \forall i \neq j. \tag{2.4}$$

LEMMA 2.2 : Let T be a triangle having one of his angles θ satisfying (1.11). Let i, j be the vertices of T opposite to the angle θ. Then for any function u

$$0 \leq \int_T a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx \leq \frac{\beta}{2}\tan\delta_3 \tag{2.5}$$

Proof : Denote by l_i, l_j the lengths of the sides of T opposite to i, j respectively. Then due to (2.2) one has

$$\nabla\varphi_i \cdot \nabla\varphi_j = -\frac{\cos\theta}{\sin^2\theta \; l_i l_j} \quad \text{on} \quad T.$$

Hence

$$0 \leq \int_T a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx = \int_T -a(x,u)\frac{\cos\theta}{\sin^2\theta \; l_i l_j} \, dx \leq -\frac{\cos\theta}{\sin^2\theta \; l_i l_j}\beta|T|$$

where $|T|$ denotes the measure of T. Since $|T| = \frac{1}{2}\sin\theta \; l_i l_j$ we obtain

$$0 \leq \int_T a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx \leq -\frac{\beta}{2\tan\theta} \leq -\frac{\beta}{2\tan(\frac{\pi}{2}+\delta_3)} = \frac{\beta}{2}\tan\delta_3.$$

■

Next we have

LEMMA 2.3 : Let T be a triangle having one of his angles θ satisfying (1.10). Let i, j be the vertices of T opposite to the angle θ. Then for any function u

$$\frac{\alpha}{2}\tan\delta_2 \leq -\int_T a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx \leq \frac{\beta}{2\tan\delta_1}. \tag{2.6}$$

Proof : With the same notation than in the preceding lemma and due to (2.2) one has

$$\frac{\cos\theta}{\sin^2\theta \; l_i l_j}\alpha|T| \leq -\int_T a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx = \int_T a(x,u)\frac{\cos\theta}{\sin^2\theta \; l_i l_j} \, dx \leq \frac{\cos\theta}{\sin^2\theta \; l_i l_j}\beta|T|.$$

Hence

$$\frac{\alpha}{2\tan\theta} \leq -\int_T a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx \leq \frac{\beta}{2\tan\theta}.$$

Then (2.6) follows by (1.10).

■

LEMMA 2.4 : Let i, j be two nodes of τ_h on the same triangle. Assume that the angles of τ_h satisfy (1.10) or (1.11) with

$$\tan\delta_3 < \frac{\alpha}{\beta}\tan\delta_2. \tag{2.7}$$

Then if

$$\int_\Omega a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx \geq 0 \tag{2.8}$$

one has

$$\nabla\varphi_i \cdot \nabla\varphi_j \geq 0. \tag{2.9}$$

Moreover, the two angles of the support of $\varphi_i \cdot \varphi_j$ opposite to the segment ij are exeeding $\frac{\pi}{2}$, i.e. satisfy (1.11).

Proof : The support of $\varphi_i \cdot \varphi_j$ or $\nabla\varphi_i \cdot \nabla\varphi_j$ is composed of two triangles (see figure 1) :

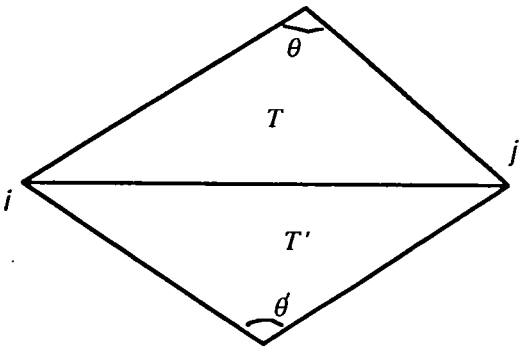


figure 1

So, by (2.8), on one of the triangles -say T- one has necessarilly

$$\nabla\varphi_i \cdot \nabla\varphi_j \geq 0.$$

Let us assume that $\nabla\varphi_i \cdot \nabla\varphi_j < 0$ on the other ones, then by (2.2) one has

$$\cos\theta \leq 0 \quad \text{on} \quad T \quad , \quad \cos\theta' > 0 \quad \text{on} \quad T'$$

i.e. θ satisfies (1.11), θ' satisfies (1.10). Then applying lemmas 2.2, 2.3 one deduces

$$\begin{aligned}\int_\Omega a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx &= \int_T a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx + \int_{T'} a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx \\ &\leq \frac{\beta}{2}\tan\delta_3 - \frac{\alpha}{2}\tan\delta_2 < 0\end{aligned}$$

which contradicts (2.8). This completes the proof of this lemma.

■

We say that j is a neighbour of i if i, j belong to the same triangle. If N denotes the number of neighbours of i and if all the triangles of the triangulation have angles that satisfy (1.10) then one has

$$N\delta_1 \leq 2\pi. \tag{2.10}$$

In the following lemma, for any function u we consider expressions of the form

$$\sum_{j\in S(i)} \int_\Omega a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx \tag{2.11}$$

where j is running on a subset $S(i)$ of neighbours of i.

We have :

LEMMA 2.5 : Assume that the angles of τ_h satisfy (1.10) or (1.11) with

$$\beta\{\frac{4\pi}{\delta_1} - 1\}\tan\delta_3 < \frac{\alpha}{2}\tan\delta_2. \tag{2.12}$$

Then if

$$\sum_{j\in S(i)} \int_\Omega a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx \geq 0 \tag{2.13}$$

each term of the above sum is nonnegative. If (2.13) fails then

$$\sum_{j\in S(i)} \int_\Omega a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx \leq -\frac{\alpha}{4}\tan\delta_2. \tag{2.14}$$

Proof : First remark that if (2.12) holds then by (2.10) one has

$$\beta(2N-1)\tan\delta_3 \leq \beta\{\frac{4\pi}{\delta_1} - 1\}\tan\delta_3 < \frac{\alpha}{2}\tan\delta_2. \tag{2.15}$$

Next, if (2.13) fails then one term of the sum at least is negative. This means that for one j at least one has

$$\nabla\varphi_i \cdot \nabla\varphi_j < 0$$

on one of the triangles of the support of $\nabla\varphi_i \cdot \nabla\varphi_j$. If one bounds from above all the other integrals of the sum as if they were positive and as if $S(i)$ had a maximum number of elements which is N we obtain by lemmas 2.2, 2.3

$$\begin{aligned}\sum_{j\in S(i)} \int_\Omega a(x,u)\nabla\varphi_i \cdot \nabla\varphi_j \, dx &\leq -\frac{\alpha}{2}\tan\delta_2 + \frac{\beta}{2}\tan\delta_3 + (N-1)2\frac{\beta}{2}\tan\delta_3 \\ &= -\frac{\alpha}{2}\tan\delta_2 + (2N-1)\frac{\beta}{2}\tan\delta_3 \\ &\leq -\frac{\alpha}{2}\tan\delta_2 + \frac{\alpha}{4}\tan\delta_2 < 0 \end{aligned} \tag{2.16}$$

(we used (2.15)). This of course contradicts (2.13). In the case where (2.13) fails then (2.14) follows from (2.16).

■

Remark 2.1 : Note that if (2.12) holds, then (2.7) holds as well, see (2.15).

If i is a node of the triangulation inside Ω we denote by D_i the open set defined by

$$D_i = \{x \in \Omega \mid \varphi_i(x) > 0\} \tag{2.17}$$

i.e; D_i is the cell built by the triangles surrounding i.

We have:

LEMMA 2.6 : Let i be a node of the triangulation. Let us denote by H_M the lengh of the largest side of the triangles of D_i and by H_m the lengh of the smallest one. One has

$$H_m \leq H_M \leq (\frac{1}{\sin\delta_1})^{\frac{\pi}{\delta_1}+\frac{1}{2}} H_m \tag{2.18}$$

Proof : The first inequality is clear. For the second one first note that in a triangle of τ_h if one denotes by A, B, C the angles and by a, b, c the lenghs of the sides opposite to these angles, then one has

$$ab \sin C = ac \sin B. \tag{2.19}$$

Moreover, due to (1.10), each of the angles of the triangle satisfies

$$\delta_1 \leq A, B, C$$

and (2.19) implies

$$b \sin \delta_1 \leq b \sin C = c \sin B \leq c. \tag{2.20}$$

If the smallest and the largest side of the triangles of D_i are in the same triangle then (2.18) follows directly from (2.20). Else one has

$$H_M \leq \frac{1}{\sin \delta_1} H_N \tag{2.21}$$

where H_N denotes a side of a neighbouring triangle to the one where the largest side belongs. Applying repeatedly (2.21) turning around i in the most favorable sense we arrive to

$$H_M \leq (\frac{1}{\sin \delta_1})^{\frac{N-1}{2}+1} H_m$$

where N denotes the number of neighbours of i. Then (2.18) follows by (2.10).

■

3. Proof of Theorem 1.3

Let $u_{i,h}$, $i = 1, 2$ be two solutions to (1.4). By subtraction one gets

$$\int_\Omega a(x, u_{1,h}) \nabla (u_{1,h} - u_{2,h}) \cdot \nabla v \, dx$$

$$= \int_\Omega (a(x, u_{2,h}) - a(x, u_{1,h})) \nabla u_{2,h} \nabla v \, dx \quad \forall v \in V_0^h. \tag{3.1}$$

Setting

$$w_h = u_{1,h} - u_{2,h} \tag{3.2}$$

and using (1.3) we obtain

$$\int_\Omega a(x, u_{1,h}) \nabla w_h \cdot \nabla v \, dx \leq C \int_\Omega |w_h| |\nabla u_{2,h}| |\nabla v| \, dx \quad \forall v \in V_0^h. \tag{3.3}$$

Consider next the function v of V_0^h defined by

$$v(i) = \begin{cases} 1 \text{ if } w_h(i) > 0, \\ \\ 0 \text{ else.} \end{cases} \tag{3.4}$$

One has

$$\int_\Omega a(x, u_{1,h}) \nabla w_h \cdot \nabla v \; dx = \sum_i w_h(i) \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla v \; dx \tag{3.5}$$

where the summation is extended to all the inside nodes in Ω. For the right hand side of (3.3) one has

$$\begin{aligned} C \int_\Omega |w_h| |\nabla u_{2,h}| |\nabla v| \; dx &= C \int_\Omega |\sum_i w_h(i) \varphi_i| |\nabla u_{2,h}| |\nabla v| \; dx \\ &\leq \sum_i |w_h(i)| \; C \int_{D_i} |\nabla u_{2,h}| |\nabla v| \; dx \end{aligned} \tag{3.6}$$

Let us first show :

LEMMA 3.1 : There exists a constant $\epsilon(h)$, independent on i, converging toward 0 when $h \to 0$ and such that

$$C \int_{D_i} |\nabla u_{2,h}| |\nabla v| \; dx \leq \epsilon(h). \tag{3.7}$$

Proof : Due to the definition of v one has -see (2.3) and lemma 2.6 for the notation-

$$|\nabla v| \leq \frac{1}{\sin \delta_1 H_m} \quad \text{on} \quad D_i. \tag{3.8}$$

Hence

$$\int_{D_i} |\nabla u_{2,h}| |\nabla v| \; dx \leq \frac{1}{\sin \delta_1 H_m} \int_{D_i} |\nabla u_{2,h}| \; dx. \tag{3.9}$$

We know that $u_{2,h}$ converges toward u the solution to (1.2) when $h \to 0$. So,

$$\begin{aligned} \int_{D_i} |\nabla u_{2,h}| \; dx &\leq \int_{D_i} |\nabla u_{2,h} - u| \; dx + \int_{D_i} |\nabla u| \; dx \\ &\leq |D_i|^{\frac{1}{2}} \{ (\int_{D_i} |\nabla u_{2,h} - u|^2 \; dx)^{\frac{1}{2}} + (\int_{D_i} |\nabla u|^2 \; dx)^{\frac{1}{2}} \} \end{aligned} \tag{3.10}$$

we used Cauchy-Schwarz Inequality. Combining (2.18), (3.9), (3.10) we deduce that for some constant K independent on i

$$C \int_{D_i} |\nabla u_{2,h}| |\nabla v| \; dx \leq K \{ (\int_\Omega |\nabla u_{2,h} - u|^2 \; dx)^{\frac{1}{2}} + (\int_{D_i} |\nabla u|^2 \; dx)^{\frac{1}{2}} \} = \epsilon(h)$$

(see (1.6)). This gives (3.7).

■

Set

$$I^+ = \{i \mid w_h(i) > 0\} \quad , \quad I^- = \{i \mid w_h(i) < 0\}, \tag{3.11}$$

$$S^+(i) = \{j \mid j \text{ is a neighbour of } i \ , \ w_h(j) > 0\}, \tag{3.12}$$

$$S^-(i) = \{j \mid j \text{ is a neighbour of } i \ , \ w_h(j) \leq 0\}. \tag{3.13}$$

If $i \in I^+$ one has

$$v = 1 - \sum_{j \in S^-(i)} \varphi_j \quad \text{on} \quad D_i$$

(indeed the two above functions coincide on the nodes of D_i, see (3.4)). Thus, in this case

$$\int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla v \ dx = - \sum_{j \in S^-(i)} \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \ dx. \tag{3.14}$$

In the case where $i \in I^-$ one has

$$v = \sum_{j \in S^+(i)} \varphi_j \quad \text{on} \quad D_i$$

and thus

$$\int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla v \ dx = \sum_{j \in S^+(i)} \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \ dx. \tag{3.15}$$

Combining the analysis above, we have arrived to the inequality

$$\sum_{i \in I^+} w_h(i) \ \{ \sum_{j \in S^-(i)} - \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \ dx \}$$

$$+ \sum_{i \in I^-} w_h(i) \ \{ \sum_{j \in S^+(i)} \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \ dx \}$$

$$\leq \epsilon(h) \sum_{i \in I^- \cup I^+} |w_h(i)| \tag{3.16}$$

Our strategy will be the following : assume that we can write the above inequality (see (3.7)) as

$$c \sum_i |w_h(i)| \leq \epsilon(h) \sum_i |w_h(i)| \tag{3.17}$$

where the summation is extended to the same set of indices on the left or the right, $c > 0$ being some constant. If the set of the above indices contains a i such that $w_h(i) \neq 0$ then we arrive to a contradiction provided that h is small enough. So, in the sequel, we are going to try to establish (3.17).

For that, consider first a $i \in I^+$ i.e. such that $w_h(i) > 0$. Several cases could occur.

Case 1 : $w_h(j) > 0$ *for any neighbour* j *of* i. In this case, (see (3.4)), $\nabla v = 0$ in D_i and the coefficients of $w_h(i)$ are vanishing on both sides of (3.16).

Case 2 : *There exists a neighbour* j *of* i *such that* $w_h(j) \leq 0$, *and*

$$\sum_{j \in S^-(i)} \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \, dx < 0. \tag{3.18}$$

Then in this case -by lemma 2.5-

$$\sum_{j \in S^-(i)} - \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \, dx \geq \frac{\alpha}{4} \tan \delta_2 = c \tag{3.19}$$

which will leads to (3.17)

The delicate case is when

Case 3 : *There exists a neighbour* j *of* i *such that* $w_h(j) \leq 0$ *and*

$$\sum_{j \in S^-(i)} \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \, dx \geq 0. \tag{3.20}$$

In other words the coefficent in front of $w_h(i)$ in the first term of the left hand side of (3.16) is nonpositive. The idea is then to transfer this term on the other side of (3.16) noticing that the coefficient in front of $w_h(i)$ is small when δ_3 is.

Similarly, if $i \in I^-$ i.e. $w_h(i) < 0$ several cases could occur.

Case 1 : $w_h(j) \leq 0$ *for any neighbour* j *of* i. In this case, (see (3.4)), $\nabla v = 0$ in D_i and the coefficients of $w_h(i)$ are vanishing on both sides of (3.16) (see (3.6)).

Case 2 : *There exists a neighbour* j *of* i *such that* $w_h(j) > 0$, *and*

$$\sum_{j \in S^+(i)} \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \, dx < 0. \tag{3.21}$$

Then in this case -by lemma 2.5-

$$\sum_{j \in S^+(i)} \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \, dx \leq -\frac{\alpha}{4} \tan \delta_2 = -c \tag{3.22}$$

which will leads to (3.17)

The last case is when

Case 3 : *There exists a neighbour j of i such that $w_h(j) > 0$ and*

$$\sum_{j \in S^+(i)} \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \, dx \geq 0. \tag{3.23}$$

In other words the coefficent in front of $-w_h(i) > 0$ in the first term of the left hand side of (3.16) is nonpositive. We will also transfer this term on the other side of (3.16). Let us cary this out.

Let us denote by I_2^+ (respectively I_2^-) the set of $i \in I^+$ (respectively I^-) corresponding to case 2 i.e.

$$I_2^+ = \{i \in I^+ \mid (3.18) \text{ holds } \} \quad , \quad I_2^- = \{i \in I^- \mid (3.21) \text{ holds } \}. \tag{3.24}$$

and by I_3^+ (respectively I_3^-)the set of $i \in I^+$ (respectively I^-) corresponding to case 3 i.e.

$$I_3^+ = \{i \in I^+ \mid (3.20) \text{ holds } \} \quad , \quad I_3^- = \{i \in I^- \mid (3.23) \text{ holds } \}. \tag{3.25}$$

Let us now show :

LEMMA 3.2 : Let $i \in I_3^+$ (resp. I_3^-). Consider j a neighbour of i such that $w_h(j) \leq 0$ (resp. $w_h(j) > 0$). Then the situation is described by the figure 2 below. The two angles opposite to the segment ij are greater or equal to $\frac{\pi}{2}$, $l, m \in I_2^+$ (resp. I_2^-), and $j \in I_2^-$ (resp. I_2^+).

Proof : Consider $i \in I_3^+$ (resp. I_3^-). Due to lemma 2.5, we know that each terms of the sum (3.20) (resp. (3.23)) are nonnegative. Thus, by lemma 2.4, for such a term the situation is described by the figure 2 i.e. the two opposite angles to the segment ij are greater or equal to $\frac{\pi}{2}$.

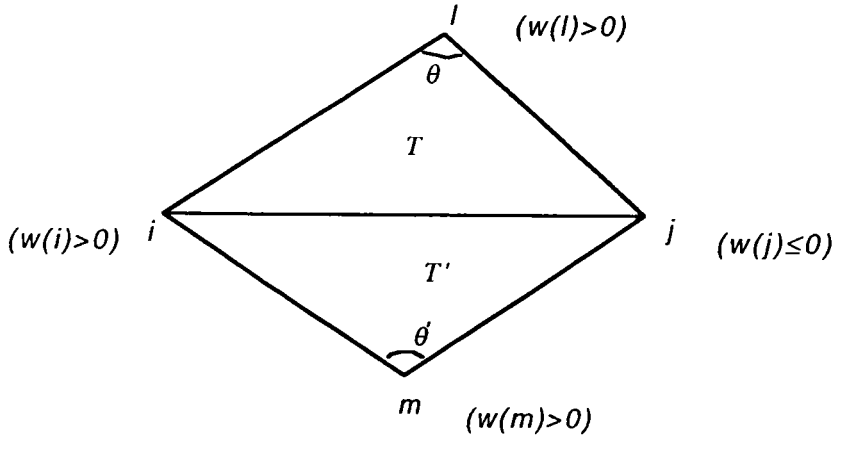


figure 2 (we replaced w_h by w)

Moreover, (see figure 2), we claim that the points l, m are such that $w_h(l),\ w_h(m) > 0$ (resp. $w_h(l),\ w_h(m) \leq 0$) and their coefficients in the left hand side of (3.6) are like in case 2 i.e. $l,\ m \in I_2^+$ (resp. $l,\ m \in I_2^-$). Indeed, due to (2.2), we have

$$\int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_l \, dx < 0. \tag{3.26}$$

If $w_h(l) \leq 0$ (resp. $w_h(l) > 0$) this contradicts (3.20) (resp. (3.23)), due to (2.9), (2.13). Thus $w_h(l) > 0$ (resp. $w_h(l) \leq 0$). If now

$$\sum_{k \in S^-(l)} \int_{D_l} a(x, u_{1,h}) \nabla \varphi_l \cdot \nabla \varphi_k \, dx \geq 0, \tag{3.27}$$

(resp.

$$\sum_{k \in S^+(l)} \int_{D_l} a(x, u_{1,h}) \nabla \varphi_l \cdot \nabla \varphi_k \, dx \geq 0) \tag{3.27'}$$

then again (2.9), (2.13) would imply a contradiction to (3.26) written for j in place of i. Thus, l -and m by the same arguments- are points in I_2^+ (resp. I_2^-). One has $j \in I_2^-$ (resp. I_2^+) since $j \in I_3^-$ (resp. I_3^+) would imply $l \in I_2^-$ (resp. I_2^+). This completes the proof of the lemma.

■

Now, if $i \in I_3^+$ (resp. I_3^-) by (2.5), it is clear that if N denotes the number of neighbours of i

$$0 \leq \sum_{j \in S^-(i)} \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \, dx \leq (N-2)\beta \tan \delta_3.$$

(resp

$$0 \leq \sum_{j \in S^+(i)} \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \, dx \leq (N-2)\beta \tan \delta_3).$$

(Note that $l, m \notin S^-(i)$, (resp. $l, m \notin S^+(i)$)). Hence,

$$0 \leq \sum_{j \in S^-(i)} \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \, dx \leq (\frac{2\pi}{\delta_1} - 2)\beta \tan \delta_3 \tag{3.28}$$

(resp.

$$0 \leq \sum_{j \in S^+(i)} \int_{D_i} a(x, u_{1,h}) \nabla \varphi_i \cdot \nabla \varphi_j \, dx \leq (\frac{2\pi}{\delta_1} - 2)\beta \tan \delta_3). \tag{3.28'}$$

At this stage, if we summarize (3.16), (3.19), (3.22) and (3.28) and if we transfer to the right the terms of the left hand side of (3.16) for which $i \in I_3^+$ and $i \in I_3^-$ we end up with

$$c \cdot \sum_{i \in I_2^+ \cup I_2^-} |w_h(i)| \leq \epsilon(h) \sum_i |w_h(i)| + \{(\frac{2\pi}{\delta_1} - 2)\beta \tan \delta_3\} \sum_{i \in I_3^+ \cup I_3^-} |w_h(i)|. \tag{3.29}$$

We would like now to estimate

$$\sum_{i\in I_3^+} |w_h(i)| \quad \text{and} \quad \sum_{i\in I_3^-} |w_h(i)|.$$

For this we have

LEMMA 3.3 : Set

$$M = \max(\frac{1}{2\tan\delta_1}, \cot\delta_1). \tag{3.30}$$

Then we have

$$\{\frac{\alpha\sin\delta_1}{\cos\delta_2} - (\frac{\pi}{\delta_1} - 1)\beta\tan\delta_3\} \sum_{i\in I_3^+} |w_h(i)| \leq \frac{2\pi}{\delta_1}\beta M \sum_{i\in I_2^+} |w_h(i)| + \frac{\pi}{\delta_1}\beta\tan\delta_3 \sum_{i\in I_2^-} |w_h(i)| + \epsilon(h) \sum_{i\in I_3^+\cup I_2^+\cup I_2^-} |w_h(i)| \tag{3.31}$$

$$\{\frac{\alpha\sin\delta_1}{\cos\delta_2} - (\frac{\pi}{\delta_1} - 1)\beta\tan\delta_3\} \sum_{i\in I_3^-} |w_h(i)| \leq \frac{2\pi}{\delta_1}\beta M \sum_{i\in I_2^-} |w_h(i)| + \frac{\pi}{\delta_1}\beta\tan\delta_3 \sum_{i\in I_2^+} |w_h(i)| + \epsilon(h) \sum_{i\in I_3^-\cup I_2^-\cup I_2^+} |w_h(i)| \tag{3.32}$$

Proof : Let us show formula (3.31). For $i \in I_3^+$ the set D_i is of the type of figure 3 below. It is built with couples of triangles (T, T') like in figure 2, completed with some other triangles.

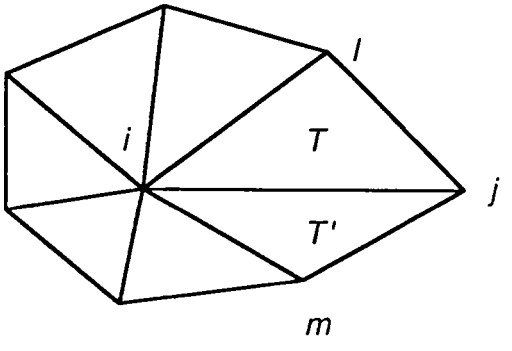


figure 3

The neighbours j of i such that $w_h(j) \leq 0$ are necessarily vertices of such a couple (T, T'), T, T' being located on each sides of the segment ij. Since the points $l, m \in I_2^+$,

(see lemma (3.2)), a couple (T, T') belongs to only one D_i such that $i \in I_3^+$. Moreover, if a triangle is part of D_i but not one of these triangles T, T', then it has a vertex belonging to I_3^+ (i) and the two others cannot satisfy $w_h \leq 0$, so that they are in I_2^+ or I_3^+.

In (3.3) we consider the test function $v = v^+$ defined by :

$$v^+(i) = \begin{cases} 1 \text{ if } i \in I_3^+, \\ 0 \text{ else.} \end{cases} \tag{3.33}$$

Let us evaluate each sides of (3.3). First on D_i we are going to integrate on couples like (T, T'). On such triangles one has (see lemma 3.2) $v^+ = \varphi_i$ so that (we use the notation of figure 3 and set from now on $a = a(x, u_{1,h})$) :

$$\int_T a\nabla w_h \cdot \nabla v^+ \, dx = w_h(i) \int_T a|\nabla \varphi_i|^2 \, dx + w_h(l) \int_T a\nabla\varphi_l \cdot \nabla\varphi_i \, dx + w_h(j) \int_T a\nabla\varphi_j \cdot \nabla\varphi_i \, dx. \tag{3.34}$$

To evaluate $|\nabla\varphi_i|^2$ note that if $\theta_i, \theta_j, \theta_l$ denote the angles of T at i, j, l and l_i, l_j, l_l the lenghs of the sides of T opposite to i, j, l by (2.3) one has

$$|\nabla\varphi_i| = \frac{1}{\sin\theta_l l_j} = \frac{1}{\sin\theta_j l_l}.$$

Hence

$$|\nabla\varphi_i|^2 = \frac{1}{\sin\theta_l l_j \sin\theta_j l_l} = \frac{\sin\theta_i}{2|T|\sin\theta_l \sin\theta_j} = \frac{\sin(\pi - \theta_l - \theta_j)}{2|T|\sin\theta_l \sin\theta_j} = \frac{1}{2|T|}(\cot\theta_l + \cot\theta_j).$$

It follows easilly that

$$\frac{\sin\delta_1}{2|T|\cos\delta_2} \leq |\nabla\varphi_i|^2 \leq \frac{1}{|T|}\cot\delta_1. \tag{3.35}$$

Combining this with lemma 2.2, 2.3 one deduces from (3.34)

$$\int_T a\nabla w_h \cdot \nabla v^+ \, dx \geq \frac{\alpha\sin\delta_1}{2\cos\delta_2} w_h(i) - \frac{\beta}{2\tan\delta_1} w_h(l) + \frac{\beta}{2}\tan\delta_3 w_h(j). \tag{3.36}$$

Similarly, integrating on T' one gets :

$$\int_{T'} a\nabla w_h \cdot \nabla v^+ \, dx \geq \frac{\alpha\sin\delta_1}{2\cos\delta_2} w_h(i) - \frac{\beta}{2\tan\delta_1} w_h(m) + \frac{\beta}{2}\tan\delta_3 w_h(j). \tag{3.37}$$

Hence, adding (3.36) and (3.37) we obtain by (3.30)

$$\int_{T\cup T'} a\nabla w_h \cdot \nabla v^+ \, dx \geq \frac{\alpha\sin\delta_1}{\cos\delta_2} w_h(i) - \beta M\{w_h(l) + w_h(m)\} + \beta\tan\delta_3 w_h(j). \tag{3.38}$$

Considering now the other side of (3.3), and integrating on T first we obtain :

$$C\int_T |w_h||\nabla u_{2,h}||\nabla v^+|\ dx \leq C\{|w_h(i)| + |w_h(l)| + |w_h(j)|\}\int_T |\nabla u_{2,h}||\nabla\varphi_i|\ dx.$$

Arguing as in the proof of (3.7) we deduce

$$C\int_T |w_h||\nabla u_{2,h}||\nabla v^+|\ dx \leq \epsilon(h)\{|w_h(i)| + |w_h(l)| + |w_h(j)|\}$$

where $\epsilon(h)$ is independent of i and converges toward 0 with h. Of course the same type of inequalities holds on T' so that we have :

$$C\int_{T\cup T'} |w_h||\nabla u_{2,h}||\nabla v^+|\ dx \leq \epsilon(h)\{|w_h(i)| + |w_h(l)| + |w_h(m)| + |w_h(j)|\}. \tag{3.39}$$

Besides the couples (T,T') when one integrates on D_i one has to integrate on three types of triangles :

1) *Triangles having its three vertices in* I_3^+. On such a triangle one has $v^+ = 1$, hence $\nabla v^+ = 0$. Thus, on such a triangle T_1

$$\int_{T_1} a\nabla w_h\cdot\nabla v^+\ dx = C\int_{T_1} |w_h||\nabla u_{2,h}||\nabla v^+|\ dx = 0. \tag{3.40}$$

2) *Triangles having two vertices in* I_3^+, *one in* I_2^+. Let i,p,q be these three vertices, with $i,p\in I_3^+$, $q\in I_2^+$. On such a triangle -say T_2- one has clearly $v^+ = 1-\varphi_q$ so that

$$\begin{aligned}\int_{T_2} a\nabla w_h\cdot\nabla v^+\ dx = &-w_h(i)\int_{T_2} a\nabla\varphi_i\cdot\nabla\varphi_q\ dx\\ &- w_h(p)\int_{T_2} a\nabla\varphi_p\cdot\nabla\varphi_q\ dx - w_h(q)\int_{T_2} a|\nabla\varphi_q|^2\ dx.\end{aligned}$$

Using lemmas 2.2, 2.3 and (3.35) and noting that when a term is nonnegative in the above sum one can bound it from below by 0 we obtain :

$$\begin{aligned}\int_{T_2} a\nabla w_h\cdot\nabla v^+\ dx &\geq -\frac{\beta}{2}\tan\delta_3 w_h(i) - \frac{\beta}{2}\tan\delta_3 w_h(p) - \beta\cot\delta_1 w_h(q)\\ &\geq -\frac{\beta}{2}\tan\delta_3 w_h(i) - \frac{\beta}{2}\tan\delta_3 w_h(p) - \beta M w_h(q)\end{aligned} \tag{3.41}$$

Of course, with the same arguments than before one has

$$C\int_{T_2} |w_h||\nabla u_{2,h}||\nabla v^+|\ dx \leq \epsilon(h)\{|w_h(i)| + |w_h(p)| + |w_h(q)|\}. \tag{3.42}$$

3) *Triangles having one vertex in I_3^+, the two others in I_2^+.* i is of course the vertex in I_3^+. Denote by $p, q \in I_2^+$ the two others. We have in this case $v^+ = \varphi_i$ so that if T_3 denotes the triangle on which we are integrating

$$\int_{T_3} a\nabla w_h \cdot \nabla v^+ \, dx = w_h(i) \int_{T_3} a|\nabla \varphi_i|^2 \, dx + w_h(p) \int_{T_3} a\nabla \varphi_p \cdot \nabla \varphi_i \, dx + w_h(q) \int_{T_3} a\nabla \varphi_q \cdot \nabla \varphi_i \, dx.$$

Using lemma 2.2 and bounding from below by 0 terms which are nonnegative we obtain

$$\begin{aligned} \int_{T_3} a\nabla w_h \cdot \nabla v^+ \, dx &\geq -\frac{\beta}{2\tan\delta_1} w_h(p) - \frac{\beta}{2\tan\delta_1} w_h(q) \\ &\geq -\beta M w_h(p) - \beta M w_h(q). \end{aligned} \tag{3.43}$$

As before we have also

$$C \int_{T_3} |w_h||\nabla u_{2,h}||\nabla v^+| \, dx \leq \epsilon(h)\{|w_h(i)| + |w_h(p)| + |w_h(q)|\}. \tag{3.44}$$

If we combine all the above inequalities, noting that if N denotes the maximum number of neigbours of a point we have :

-each D_i has at least a couple (T, T'); it cannot belong to another D_k,

-each D_i has at most $N-2$ other triangles,

-a point belongs at most to N triangles,

we obtain

$$\begin{aligned} \int_\Omega a\nabla w_h \cdot \nabla v^+ \, dx &\geq \{\frac{\alpha \sin\delta_1}{\cos\delta_2} - (\frac{\pi}{\delta_1} - 1)\beta\tan\delta_3\} \sum_{i\in I_3^+} |w_h(i)| \\ &\quad - N\beta M \sum_{i\in I_2^+} |w_h(i)| - N\frac{\beta}{2}\tan\delta_3 \sum_{i\in I_2^-} |w_h(i)| \\ &\geq \{\frac{\alpha \sin\delta_1}{\cos\delta_2} - (\frac{\pi}{\delta_1} - 1)\beta\tan\delta_3\} \sum_{i\in I_3^+} |w_h(i)| \\ &\quad - \frac{2\pi}{\delta_1}\beta M \sum_{i\in I_2^+} |w_h(i)| - \frac{\pi}{\delta_1}\beta\tan\delta_3 \sum_{i\in I_2^-} |w_h(i)| \end{aligned} \tag{3.45}$$

Changing also if necessary $N\epsilon(h)$ in $\epsilon(h)$ we obtain also

$$C \int_\Omega |w_h||\nabla u_{2,h}||\nabla v^+| \, dx \leq \epsilon(h) \sum_{i\in I_3^+ \cup I_2^+ \cup I_2^-} |w_h(i)|. \tag{3.46}$$

Combining (3.45), (3.46) we obtain (3.31). (3.32) is obtained by the same way using a function v^- in (3.3) that is equal to -1 on I_3^- and vanishes elsewhere. One can note also exchanging the roles of $u_{1,h}$ and $u_{2,h}$ that $-w_h$ satisfies an inequality of type (3.3). This completes the proof of the lemma.

■

From (3.31), (3.32) we deduce

$$\{[\frac{\alpha \sin \delta_1}{\cos \delta_2} - (\frac{\pi}{\delta_1} - 1)\beta \tan \delta_3] - \epsilon(h)\} \sum_{i \in I_3^+ \cup I_3^-} |w_h(i)|$$

$$\leq \{[\frac{2\pi}{\delta_1}\beta M + \frac{\pi}{\delta_1}\beta \tan \delta_3] + \epsilon(h)\} \sum_{i \in I_2^+ \cup I_2^-} |w_h(i)|. \quad (3.47)$$

Assuming $\epsilon(h)$ small enough i.e. h small enough in such a way that

$$\{[\frac{\alpha \sin \delta_1}{\cos \delta_2} - (\frac{\pi}{\delta_1} - 1)\beta \tan \delta_3] - \epsilon(h)\} > 0$$

we obtain

$$\sum_{i \in I_3^+ \cup I_3^-} |w_h(i)| \leq \frac{[\frac{2\pi}{\delta_1}\beta M + \frac{\pi}{\delta_1}\beta \tan \delta_3] + \epsilon(h)}{[\frac{\alpha \sin \delta_1}{\cos \delta_2} - (\frac{\pi}{\delta_1} - 1)\beta \tan \delta_3] - \epsilon(h)} \sum_{i \in I_2^+ \cup I_2^-} |w_h(i)|. \quad (3.48)$$

Going back to (3.29) we obtain

$$\{c - c(h)\} \sum_{i \in I_2^+ \cup I_2^-} |w_h(i)| \leq \epsilon(h) \sum_{i \in I_2^+ \cup I_2^-} |w_h(i)| \quad (3.49)$$

with

$$c(h) = \frac{[(\frac{2\pi}{\delta_1} - 2)\beta \tan \delta_3 + \epsilon(h)][\frac{2\pi}{\delta_1}\beta M + \frac{\pi}{\delta_1}\beta \tan \delta_3 + \epsilon(h)]}{[\frac{\alpha \sin \delta_1}{\cos \delta_2} - (\frac{\pi}{\delta_1} - 1)\beta \tan \delta_3 - \epsilon(h)]}.$$

Letting $h \to 0$ we see that

$$c - c(h) \to \frac{\alpha}{4} \tan \delta_2 - \frac{[(\frac{2\pi}{\delta_1} - 2)\beta \tan \delta_3][\frac{2\pi}{\delta_1}\beta M + \frac{\pi}{\delta_1}\beta \tan \delta_3]}{[\frac{\alpha \sin \delta_1}{\cos \delta_2} - (\frac{\pi}{\delta_1} - 1)\beta \tan \delta_3]} > 0$$

(see (1.13)). Thus, if I_2^+ is not empty, provided that h is taken small enough (3.49) gives us a contradiction. It follows then from lemma 3.2 that if I_2^+ is empty so is I_3^+. Thus the only possibility is to have $w_h \leq 0$. Exchanging the roles of $u_{1,h}$ and $u_{2,h}$ we deduce that $w_h = 0$. This completes the proof of the theorem 1.3.

■

Remark 3.1 : We did not assume here the mesh to be uniform. If one assumes the triangulation to be regular, then our assumptions are somehow optimal in the sense that if one triangle has an angle equal to $\frac{\pi}{2} - \delta_2$ with δ_2 close to 0 then, unless the triangulation fails to be regular the other angles cannot exceed $\frac{\pi}{2} + \delta_3$ with δ_3 small.

3.Version approchée du Théorème de Meyers :

Théorème :

Sous les hypothèses du théorème (1.3), et en supposant que $f \in L^2(\Omega)$, la solution u_h du problème discret (1.4) vérifie :

$$\|u_h\|_{1,p} \leq c\|f\|_2 \tag{3.1}$$

pour $0 < h < h_0$ et $2 < p < p_0$.

Preuve :

D'après Meyers [Me.] on a :

$$\|u\|_{1,p} \leq c_p sup_{w \in S_q}(\nabla u, \nabla w) \quad \forall u \in W_0^{1,p}(\Omega) \tag{3.2}$$

où $S_q = \{u \in W_0^{1,q}(\Omega) : \|u\|_{1,q} = 1\}$ et c_p est une fonction log-convexe en $\frac{1}{p}$.

Soit v_h la solution du problème discret suivant :

$$\begin{cases} \int_\Omega \nabla v_h . \nabla \phi \, dx = \int_\Omega F . \nabla \phi \, dx, & \forall \phi \in V_0^h \\ v_h \in V_0^h \end{cases} \tag{3.3}$$

où $F \in (L^p(\Omega))^2$.

Alors on a :

$$\|v_h\|_{1,p} \leq c_p(\delta + 1)\|F\|_p \tag{3.4}$$

pour $0 < h < h_0$.

En effet :

$$\begin{aligned} \|v_h\|_{1,p} &\leq c_p sup_{w \in S_q}(\nabla v_h, \nabla w) \\ &= c_p sup_{w \in S_q}(\nabla v_h, \nabla w_h) \end{aligned}$$

avec w_h la solution de

$$\int_\Omega \nabla w_h . \nabla v_h \, dx = \int_\Omega \nabla w . \nabla v_h \, dx, \quad \forall v_h \in V_0^h,$$

d'où :

$$\begin{aligned} \|v_h\|_{1,p} &\leq c_p sup_{w \in S_q}(F, \nabla w_h) \\ &\leq c_p \|F\|_p \|w_h\|_{1,q} \\ &\leq c_p(\delta + 1)\|F\|_p \end{aligned}$$

pour $0 < h < h_0$.
Soit $G \in W_0^{1,2}(\Omega)$ la solution du problème suivant :

$$-\Delta G = f \tag{3.5}$$

avec $f \in L^2(\Omega)$. On a alors :

$$\|\nabla G\|_r \leq c\|f\|_2, \quad \forall\, r \geq 1. \tag{3.6}$$

En effet on a $|\nabla^2 G| \in L^2(\Omega)$ et en particulier $|\nabla G| \in L^r(\Omega)$, $\forall\ r \geq 1$. Par suite $\|\nabla G\|_r \leq c\|f\|_2$.

Soient $g_h \in V_0^h$ et $v_h \in W_0^{1,2}(\Omega)$ la solution du problème discret :

$$\begin{cases} \int_\Omega a(x, g_h)\nabla v_h.\nabla\phi\ dx = \int_\Omega \nabla G.\nabla\phi\ dx, \quad \forall\phi \in V_0^h. \\ v_h \in V_0^h \end{cases} \tag{3.7}$$

Considérons la transformation linéaire :

$$\tau : W_0^{1,p}(\Omega) \longrightarrow W_0^{1,p}(\Omega) \tag{3.8}$$

définie comme suit : τv_h est l'unique solution w_h de

$$\int_\Omega \nabla w_h.\nabla\phi\ dx = \int_\Omega (1 - \frac{a(x, g_h)}{\beta})\nabla v_h.\nabla\phi\ dx + \frac{1}{\beta}\int_\Omega \nabla G.\nabla\phi\ dx, \quad \forall\phi \in V_0^h.$$

On a d'après ce qui précède :

$$\|w_h\|_{1,p} \leq c_p(\delta + 1)\{\frac{1}{\beta}\|\nabla G\|_p + (1 - \frac{\alpha}{\beta})\|v_h\|_{1,p}\} \tag{3.9}$$

pour $0 < h < h_0$.

Pour $w_{i,h} = \tau v_{i,h}$, $(i = 1, 2)$,

$$\|w_{1,h} - w_{2,h}\|_{1,p} \leq c_p(\delta + 1)(1 - \frac{\alpha}{\beta})\|v_{1,h} - v_{2,h}\|_{1,p} \tag{3.10}$$

pour $0 < h < h_0$; donc pour $2 < p < p_0$, $c_p(\delta + 1)(1 - \frac{\alpha}{\beta}) < 1$, et par conséquent τ est une contraction qui admet un point fixe v_h. (Rappelons que par (3.2) on a $c_2 = 1$)
La solution du problème (3.7) est dans $W_0^{1,p}(\Omega)$ pour $2 < p < p_0$ et vérifie :

$$\|v_h\|_{1,p} \leq c\|f\|_2.$$

Par ailleurs, considérons l'application suivante :

$$T : \ g_h \in V_0^h \longrightarrow v_h \in V_0^h \tag{3.11}$$

où v_h est la solution du problème(3.7). Cette application envoie la boule $B(0, c\|f\|_2)$ de V_0^h sur elle même.
De plus, il est facile de montrer que T est continue, donc d'après le théorème du point fixe de Brouwer, (cf.[G.T]), T admet un point fixe.

■

CHAPITRE 4 :

1. Unicité en dimension quelconque :

Définition 4.1 :

Soit p un nombre réel tel que $1 < p < \infty$. Alors p^* désigne le réel défini par $\frac{1}{p^*} = \frac{1}{p} - \frac{1}{n}$ si $p < n$ et est défini par tout nombre appartenant à $]1, \infty[$ si $p \geq n$.

Soit Ω un ouvert borné de $\mathbb{R}^n$, $n \geq 2$, de frontière Γ de classe C^1. On considère le problème discret suivant :

$$\begin{cases} \int_\Omega a(x, u_h) \nabla u_h . \nabla v \, dx = \int_\Omega f v \, dx, \quad \forall v \in V_0^h \\ u_h \in V_0^h \end{cases} \tag{4.1}$$

où

$$f \in L^p(\Omega), \quad p > n \tag{4.2}$$

(il suffit de prendre $p > \frac{n}{2}$)
et $a(x, u)$ est une fonction de Carathéodory satisfaisant :

$$0 < \alpha \leq a(x, u) \leq \beta, \quad \text{p.p } x \in \Omega, \quad \forall u \in \mathbb{R} \tag{4.3}$$

où α , β sont deux constantes positives,
et

$$|a(x, u) - a(x, v)| \leq C.|u - v| \ \forall u, v \in \mathbb{R}, \quad \text{p.p } x \tag{4.4}$$

$$|a(x, u) - a(y, u)| \leq K.|x - y| \ \forall x, y \in \Omega, \ \forall u \in \mathbb{R} \tag{4.5}$$

où C, K sont des constantes positives.

Théorème 4.2 : Principe du maximum discret

Sous les hypothèses (4.2), (4.3), (4.4), (4.5) et (1.9'); on a :

$$\|u_h\|_\infty \leq C_3 \|f\|_p \tag{4.6}$$

où u_h est la solution du problème discret (4.1).

Preuve : (Voir Annexe).
On a adapté la démonstration faite dans [C.R.].

■

Théorème 4.3 : Estimation L^p du Gradient

Sous les mêmes hypothèses que dans le Théorème 4.2, il existe $h_0 > 0$ tel que la solution du problème discret (4.1) vérifie :

$$\|u_h\|_{1,p} \leq C_4 \|f\|_p \tag{4.7}$$

pour $0 < h < h_0$ et une certaine constante C_4.
Preuve :

Considérons la forme bilinéaire suivante, pour $\phi, \psi \in C_0^\infty(\Omega)$:

$$B[\phi, \psi] = \int_\Omega a(x, u_h(x)) \nabla\phi . \nabla\psi \, dx \tag{4.8}$$

D'après la définition 2 donnée en Annexe 1, B est une forme biliniéaire de Dirichlet uniformément elliptique d'ordre 1 sur Ω.
Du fait que $a(x,u)$ est lipchitzienne par rapport à u, uniformément continue en x et que $u_h(x)$ est continue, $a(x, u_h(x))$ est continue par rapport à x; donc B est régulière au sens de la définition 3 donnée en annexe 1.
D'après le Théorème 2 (Annexe2), $\exists$ $C_1, C_2 > 0$ tels que :

$$\forall u \in W_0^{1,p}(\Omega), \quad C_1 \|u\|_{1,p} \leq sup_{\phi \in S_q} |B[u, \phi]| + C_2 \|u\|_p \tag{4.9}$$

où $S_q = \{v \in W_0^{1,q}(\Omega) \ : \ \|v\|_{1,q} = 1\}$ et $\frac{1}{p} + \frac{1}{q} = 1$.
$u_h \in W_0^{1,\infty}(\Omega) \subset W_0^{1,p}(\Omega)$, donc on peut écrire :

$$C_1 \|u_h\|_{1,p} \leq sup_{\phi \in S_q} |B[u_h, \phi]| + C_2 \|u_h\|_p \tag{4.10}$$

ou encore :

$$C_1 \|u_h\|_{1,p} \leq sup_{\phi \in S_q} |B[u_h, \phi_h]| + C_2 \|u_h\|_p \tag{4.11}$$

si ϕ_h désigne la solution de

$$\int_\Omega a(x, u_h(x)) \nabla\phi_h . \nabla v_h \, dx = \int_\Omega a(x, u_h(x)) \nabla\phi . \nabla v_h \, dx \quad \forall v_h \in V_0^h \tag{4.12}$$

D'où :

$$C_1 \|u_h\|_{1,p} \leq sup_{\{\phi \in S_q\}} \int_\Omega a(x, u_h(x)) \nabla u_h . \nabla\phi_h \, dx + C_2 \|u_h\|_p. \tag{4.13}$$

Par ailleurs, considérons le problème suivant :

$$\begin{cases} -\Delta w = f \\ w \in H_0^1(\Omega) \end{cases} \tag{4.14}$$

$|\nabla^2 w| \in L^p(\Omega) \Longrightarrow |\nabla w| \in W^{1,p}(\Omega) \subset L^{p^*}(\Omega)$ et $\|\nabla w\|_{p^*} \leq c\|f\|_p$.
On a d'après l'inégalité de Hölder :

$$\int_\Omega a(x, u_h(x))\nabla u_h.\nabla \phi_h \, dx \leq \|\nabla w\|_{p^*} \|\phi_h\|_{1,q^*} \tag{4.15}$$

avec : $\frac{1}{p^*} + \frac{1}{q^*} = 1$. D'autre part on a :

$$\begin{aligned} \|\phi_h\|_{1,q^*} &\leq \|\phi - \phi_h\|_{1,q^*} + \|\phi\|_{1,q^*} \\ &\leq C_*\|\phi - \phi_h\|_{1,q} + 1 \\ &\leq \delta + 1 \end{aligned}$$

pour $0 < h < h_0$.
Ce qui nous permet de conclure que pour $0 < h < h_0$, et en utilisant le principe du maximum discret dans (4.13), on a (4.7) pour une certaine constante C_4.

■

lemme 4.4 :

Soit un n-simplexe T de la triangulation τ_h vérifiant (1.9), soient b_r, $1 \leq r \leq n+1$, ses sommets et soient λ_r, $1 \leq r \leq n+1$, les coordonnées barycentriques d'un point $x \in T$ par rapport aux points b_r.
Alors pour $i, j = 1, 2, ..., n+1$ il existe des constantes $c_1, c_2 > 0$ telles que :

$$\begin{cases} \nabla\lambda_i.\nabla\lambda_j \leq -\frac{c_1}{h_T^2} < 0 \\ \frac{1}{h_T} \leq |\nabla\lambda_i| \leq \frac{c_2}{h_T}. \end{cases} \tag{4.16}$$

Preuve :

Soit P l'hyperplan engendré par $b_2, b_3, ..., b_{n+1}$ et soit ν un vecteur unitaire orthogonal à P. On suppose que $b_1 \neq 0$ alors :

$$\lambda_1(x) = \frac{x.\nu}{b_1.\nu}.$$

D'autre part, soit P' le plan parallèle à P passant par b_1;
on a : $|b_1.\nu|$ = distance entre P et P'.

D'où : $h_T \geq |b_1.\nu| \geq \rho_T \implies \frac{1}{h_T} \leq |\nabla\lambda_1| \leq \frac{1}{\rho_T} \leq \frac{c_2}{h_T}$ et finalement $\frac{1}{h_T} \leq |\nabla\lambda_i| \leq \frac{c_2}{h_T}$.
Par ailleurs :

$\nabla\lambda_i.\nabla\lambda_j = |\nabla\lambda_i||\nabla\lambda_j|cos(\nabla\lambda_i, \nabla\lambda_j) \leq -\frac{c_1}{h_T^2}$.

■

Théorème 4.5 :

Sous les mêmes hypothèses que dans le Théorème 4.2, si

$$h < min(h_0, (M\|f\|_p)^{\frac{p}{p-n}}) \tag{4.17}$$

alors pour une certaine constante M, le problème discret (4.1) admet une unique solution.

Preuve :

Supposons que u_h et v_h soient deux solutions de (4.1). On a donc :

$$\int_\Omega a(x,u_h)\nabla(u_h - v_h).\nabla\phi\ dx = \int_\Omega [a(x,v_h) - a(x,u_h)]\nabla v_h.\nabla\phi\ dx, \quad \forall\phi \in V_0^h \tag{4.18}.$$

On pose : $w_h = u_h - v_h$. En utilisant (4.4), on obtient :

$$\int_\Omega a(x,u_h)\nabla w_h.\nabla\phi\ dx \leq C\int_\Omega |w_h||\nabla v_h||\nabla\phi|\ dx, \quad \forall\phi \in V_0^h \tag{4.19}$$

Soient K_i les sommets intérieurs de la triangulation et considérons la fonction test suivante

$$\phi(K_i) = \begin{cases} 1 & si\ w_h(K_i) > 0 \\ 0 & \text{sinon.} \end{cases}$$

Sur chaque n-simplexe T, on a : $\nabla\phi = 0$, sauf si w_h change de signe sur T i.e. :

$$\begin{cases} w_h(b_{i_1}) > 0, w_h(b_{i_2}) > 0, ..., w_h(b_{i_l}) > 0 \\ w_h(b_{i_{l+1}}) \leq 0, w_h(b_{i_{l+2}}) \leq 0, ..., w_h(b_{i_{n+1}}) \leq 0 \end{cases} \tag{4.20}$$

pour $l = 1, 2, ..., n$; i_1, i_2, ..., i_{n+1} désignant une permutation de 1, 2, ..., $n+1$.
On déduit alors que (4.19) s'écrit

$$\sum_T \int_T a(x,u_h)\nabla w_h.\nabla\phi\ dx \leq C\sum_T \int_T |w_h||\nabla v_h||\nabla\phi|\ dx, \quad \forall\phi \in V_0^h \tag{4.21}$$

où la sommation est étendue aux n-simplexes satisfaisant (4.20).

Si T satisfait (4.20), il est clair que w_h s'annule sur T en un certain point y et on a :

$$w_h(x) = w_h(x) - w_h(y) = \nabla w_h.(x-y), \quad \forall x \in T$$

ce qui implique

$$\sum_T \int_T a(x,u_h)\nabla w_h.\nabla\phi \, dx \leq C \sum_T h_T \int_T |\nabla w_h||\nabla v_h||\nabla\phi| \, dx, \quad \forall \phi \in V_0^h. \tag{4.22}$$

On pose :

$$a_j^T = w_h(b_{i_j}).$$

Pour un tel n-simplexe $T \in \tau_h$, on a :

$$w_h = \sum_{j=1}^{n+1} a_j^T \lambda_j, \ \nabla w_h = \sum_{j=1}^{n+1} a_j^T \nabla\lambda_j.$$

Sur T on a :

$$\begin{cases} \phi &= \sum_{k=1}^{l} \lambda_k \\ \nabla\phi &= \sum_{k=1}^{l} \nabla\lambda_k \quad = -\sum_{k=l+1}^{n+1} \nabla\lambda_k \end{cases} \tag{4.23}$$

$$\nabla w_h = \sum_{j=1}^{l} a_j^T \nabla\lambda_j + \sum_{j=l+1}^{n+1} a_j^T \nabla\lambda_j.$$

On a encore :

$$\begin{aligned} (\sum_{j=1}^{l} a_j^T \nabla\lambda_j).(-\sum_{k=l+1}^{n+1} \nabla\lambda_k) &= \sum_{j=1}^{l} a_j^T \sum_{k=l+1}^{n+1} (-\nabla\lambda_j.\nabla\lambda_k) \\ &\geq \frac{(n-l+1)c_1}{h_T^2} \sum_{j=1}^{l} |a_j^T|. \end{aligned}$$

On a également :

$$\begin{aligned} (\sum_{j=l+1}^{n+1} a_j^T \nabla\lambda_j).(\sum_{k=1}^{l} \nabla\lambda_k) &= \sum_{j=l+1}^{n+1} a_j^T \sum_{k=1}^{l} \nabla\lambda_j.\nabla\lambda_k \\ &\geq \frac{lc_1}{h_T^2} \sum_{j=l+1}^{n+1} |a_j^T| \end{aligned}$$

Ceci nous permet de conclure que :

$$\nabla w_h.\nabla\phi \geq min(l, n-l+1)\frac{c_1}{h_T^2} \sum_{j=1}^{n+1} |a_j^T| \ .$$

Par ailleurs, sur chaque n-simplexe T on a :

$$|\nabla w_h| \leq \frac{c_2}{h_T} \sum_{j=1}^{n+1} |a_j^T|$$

$$|\nabla\phi| \leq \frac{min(l, n-l+1).c_2}{h_T}.$$

Finalement, on obtient :

$$\alpha c_1 \sum_T \frac{1}{h_T^2} \int_T \sum_{j=1}^{n+1} |a_j^T| \, dx \leq C c_2^2 \sum_T \frac{1}{h_T} \int_T (\sum_{j=1}^{n+1} |a_j^T|)(|\nabla v_h|) \, dx.$$

On obtient, en utilisant la régularité de la triangulation

$$\sum_T h_T^{n-2} \sum_{j=1}^{n+1} |a_j^T| \leq C' \sum_T h_T^{n-2} \sum_{j=1}^{n+1} |a_j^T| . \frac{1}{h_T^{n-1}} \int_T |\nabla v_h| \, dx$$

où C' est une certaine constante. Or, en utilisant l'inégalité de Hölder,

$$\begin{aligned} \frac{1}{h_T^{n-1}} \int_T |\nabla v_h| \, dx \leq & h_T^{\frac{p-n}{p}} (\int_T |\nabla v_h|^p \, dx)^{\frac{1}{p}} \\ \leq & C_4 \|f\|_p h_T^{\frac{p-n}{p}} \end{aligned}$$

pour $0 < h < h_0$.
Finalement, pour $0 < h < h_0$:

$$\sum_T h_T^{n-2} \sum_{j=1}^{n+1} |a_j^T| \leq h^{\frac{p-n}{p}} M \|f\|_p \sum_T h_T^{n-2} \sum_{j=1}^{n+1} |a_j^T|$$

D'où si

$$h < min(h_0, (M\|f\|_p)^{\frac{p}{p-n}})$$

on obtient une contradiction. On a donc $w_h \geq 0$ et changeant les rôles de u_h et v_h, $w_h = 0$. D'où l'unicité.

■

2. Estimation de la convergence

Théorème 4.6 :

Soient u et u_h respectivement les solutions des problèmes (1.1) et (4.1), alors sous les hypothèses (4.2), (4.5) et si on suppose de plus que :

$$\|f\|_p < \frac{\alpha}{C.C_4.c_*} \tag{4.24}$$

où c_* est une constante provenant d'une injection de Sobolev, alors on a :

$$\|u - u_h\|_{1,2} \leq M_1 \|u - v_h\|_{1,2} \quad \forall v_h \in V_0^h \tag{4.25}$$

pour $0 < h < h_0$ et une certaine constante M_1.

Preuve :

Avec les mêmes notations que précédemment, on a :

$$\begin{aligned} B[u - u_h, v_h; u] &= \int_\Omega a(x,u)\nabla(u - u_h).\nabla v_h \, dx \\ &= \int_\Omega a(x,u_h)\nabla u_h.\nabla v_h \, dx - \int_\Omega a(x,u)\nabla u_h.\nabla v_h \, dx\,. \\ &= \int_\Omega [a(x,u_h) - a(x,u)]\nabla u_h.\nabla v_h \, dx \end{aligned}$$

Par ailleurs :

$$\begin{aligned} B[u - u_h, u - u_h; u] &= B[u - u_h, u - v_h; u] + B[u - u_h, v_h - u_h; u] \\ &= B[u - u_h, u - v_h; u] + \int_\Omega [a(x,u_h) - a(x,u)]\nabla(v_h - u_h).\nabla u_h \, dx \\ &\leq \beta \|u - u_h\|_{1,2}.\|u - v_h\|_{1,2} + C \int_\Omega |u - u_h||\nabla(v_h - u_h).\nabla u_h| \, dx \\ &\leq \beta \|u - u_h\|_{1,2}.\|u - v_h\|_{1,2} + C\| \, |u - u_h|. \, | \, \nabla u_h| \, \|_2.\|v_h - u_h\|_{1,2} \\ &\leq \beta \|u - u_h\|_{1,2}.\|u - v_h\|_{1,2} + C\|u - u_h\|_q \|u_h\|_{1,p}.\|v_h - u_h\|_{1,2} \\ &\leq \|u - u_h\|_{1,2}(\beta\|u - v_h\|_{1,2} + Cc_*\|u_h\|_{1,p}\|v_h - u_h\|_{1,2}) \end{aligned}$$

avec $\frac{1}{q} + \frac{1}{p} = \frac{1}{2}$.

Par conséquent :

$$\|u - u_h\|_{1,2} \leq \frac{\beta}{\alpha}\|u - v_h\|_{1,2} + \frac{C.c_*}{\alpha}\|u_h\|_{1,p}.\|v_h - u_h\|_{1,2} \tag{4.26}$$

$$\|u-u_h\|_{1,2} \leq (\frac{\beta}{\alpha} + \frac{C.c_*}{\alpha}\|u_h\|_{1,p})\|u-v_h\|_{1,2} + (\frac{C.c_*}{\alpha}\|u_h\|_{1,p})\|u-u_h\|_{1,2}$$

d'où :

$$(1 - \frac{C.C_4.c_*}{\alpha}\|f\|_p)\|u-u_h\|_{1,2} \leq (\frac{\beta}{\alpha} + \frac{C.C_4.c_*}{\alpha}\|f\|_p)\|u-v_h\|_{1,2} \qquad (4.27)$$

pour $0 < h < h_0$.

Donc si (4.24) est vérifiée, on a (4.25) pour une certaine constante M_1.

■

ANNEXE 1

Définition 1:

Soit Ω un ouvert de R^n, soient $n \geq 2$ et $m \geq 1$ des entiers. Pour tout $s \in Z_+^n$ avec $|s| \leq 2m$, soient $a_s(.)$ des fonctions complexes définies sur Ω. Alors,

$$L = \sum_{|s| \leq 2m} a_s(.) D^s \qquad (A.1)$$

est un opérateur différentiel d'ordre $2m$ défini sur Ω.

• L est dit uniformément elliptique sur Ω, si

(i) $\exists E > 0$ tel que

$$| \sum_{|s|=2m} a_s(.) l^s | \geq E |l|^{2m} \qquad (A.2)$$

pour tout $x \in \Omega$ et pour tout $l \in R^n$, et si

(ii) la condition suivante est satisfaite (" root condition ") :
$\forall l' = (l_1, l_2, .., l_{n-1}) \in R^{n-1}$ et $\forall x \in \Omega$ le polynôme en $\tau \in C$,

$$P(\tau; l'; x') = \sum_{|s|=2m} a_s(x) l'^{s'} \tau^{s_n}, \quad s = (s', s_n) \qquad (A.3)$$

admet exactement m racines à parties imaginaires positives et m racines à parties imaginaires négatives.

• L est dit uniformément fortement elliptique sur Ω, si L est uniformément elliptique et s'il existe une constante $E' > 0$ telle que :

$$(-1)^m Re \sum_{|s|=2m} a_s(.) l^s \geq E' |l|^{2m} \qquad (A.4)$$

pour tout $x \in \Omega$ et pour tout $l \in R^n$. E et E' sont appelées les constantes d'ellipticité de L.

Définition 2:

Soit Ω un ouvert de R^n, soit $n \geq 2$ et $m \geq 1$ des entiers. Pour $|\alpha| \leq m$, $|\beta| \leq m$, soient $a_{\alpha\beta}(.)$ des fonctions complexes mesurables bornées définies sur Ω.
Pour $\phi, \psi \in C_0^\infty(\Omega)$ soient :

$$B[\phi, \psi] = \sum_{|\alpha|,|\beta| \leq m} (a_{\alpha\beta} D^\alpha \phi, D^\beta \psi) \qquad (A.5)$$

et

$$L_B = (-1)^m \sum_{|\alpha|=|\beta|=m} a_{\alpha\beta} D^\alpha D^\beta \qquad (A.6)$$

Alors B est dite :

• une forme bilinéaire de Dirichlet uniformément elliptique sur Ω, si L_B défini par (1.6) est uniformément elliptique sur Ω au sens de la définition 1.
• une forme bilinéaire de Dirichlet uniformément fortement elliptique sur Ω, si L_B est uniformément fortement elliptique sur Ω.

Définition 3:

Soit $\Omega \subset R^n$ un ouvert borné et soit

$$B[\phi, \psi] = \sum_{|\alpha|,|\beta| \leq m} (a_{\alpha\beta} D^\alpha \phi, D^\beta \psi), \quad \phi, \psi \in C_0^\infty(\Omega) \qquad (A.7)$$

une forme bilinéaire de Dirichlet uniformément elliptique d'ordre m. On dit que $B[\phi, \psi]$ est régulière, si pour $|\alpha| \leq m$, $|\beta| \leq m$, $a_{\alpha\beta}(.) \in L^\infty(\Omega)$. Par ailleurs, on suppose que $a_{\alpha\beta}(.)$ avec $|\alpha| = |\beta| = m$ sont continues sur $\bar{\Omega}$, ou ce qui est équivalent, $a_{\alpha\beta}(.)$ uniformément continues sur Ω pour $|\alpha| = |\beta| = m$.

ANNEXE 2

Théorème 1 : (Généralisation du théorème de Lax-Milgram)

Soit $1 < p < \infty$, $1 < q < \infty$ des nombres réels avec $\frac{1}{p} + \frac{1}{q} = 1$ et soit $m \geq 1$ un entier. Supposons que $\Omega \subset R^n$ soit un ouvert borné de frontière Γ de classe C^m et que $B[.,.]$ soit une forme bilinéaire sur $W_0^{m,p}(\Omega) \times W_0^{m,q}(\Omega)$.
Supposons qu'il existe $C_i > 0 (i = 1, 2)$ telles que :

$$C_1 \|u\|_{m,p} \leq sup_{\phi \in S_q} |B[u, \phi]|, \quad \forall u \in W_0^{m,p}(\Omega) \tag{A.8}$$

où $S_q = \{u \in W_0^{m,q}(\Omega) \ : \ \|u\|_{m,q} = 1\}$,
et

$$C_2 \|v\|_{m,q} \leq sup_{\psi \in S_p} |B[\psi, v]|, \quad \forall v \in W_0^{m,q}(\Omega) \tag{A.9}$$

où $S_p = \{u \in W_0^{m,p}(\Omega) \ : \ \|u\|_{m,p} = 1\}$. Alors pour $\forall F^* \in W_0^{m,q}(\Omega)^*$ $(\forall G^* \in W_0^{m,p}(\Omega)^*)$, $\exists! u \in W_0^{m,p}(\Omega)$, $(\exists! v \in\in W_0^{m,q}(\Omega))$,
tels que :

$$B[u, \phi] =< F^*, \phi >, \quad \forall \phi \in W_0^{m,q}(\Omega) \tag{A.10}$$

$$\bar{B}[\psi, v] =< G^*, \psi >, \quad \forall \psi \in W_0^{m,p}(\Omega) \tag{A.11}$$

et

$$\|u\|_{m,p} \leq \tfrac{1}{KC_1} \|F^*\|^*_{m,q} \tag{A.12}$$

$$\|v\|_{m,q} \leq \tfrac{1}{KC_2} \|G^*\|^*_{m,p} \tag{A.13}$$

■

Théorème 2 : (Généralisation de l'Inégalité de G$\mathring{a}$rding)

Soit $1 < p < \infty$, $1 < q < \infty$ des nombres réels avec $\frac{1}{p} + \frac{1}{q} = 1$ et soit $m \geq 1$ un entier. Supposons que $\Omega \subset R^n$ soit un ouvert borné de frontière Γ de classe C^m et que $B[.,.]$ soit une forme bilinéaire régulière de Dirichlet uniformément elliptique, d'ordre m.
Alors $\exists$ $C_1 = C_1(n, m, p, \Omega, E) > 0$ et $C_2 = C_2(n, m, p, \Omega, E, a_{\alpha\beta}) \geq 0$ tels que :

$$\forall u \in W_0^{m,p}(\Omega), \quad C_1 \|u\|_{m,p} \leq sup_{\phi \in S_q} |B[u, \phi]| + C_2 \|u\|_p \tag{A.14}$$

■

Théorème 3 :

Sous les mêmes hypothèses que le théorème 2, et en supposant de plus que $B[.,.]$ est uniformément fortement elliptique.
Alors, pour $x_0 \in \bar{\Omega}$ il existe un voisinage ouvert Ω_{x_0} de x_0, de frontière $\partial\Omega_{x_0}$ de classe C^m, tel que il existe une constante $K = K(p, n, m, \Omega, \Omega_{x_0}, a_{\alpha,\beta}, E) > 0$ avec :

$$K\|u\|_{m,p} \leq sup_{\{\phi \in W_0^{m,q}(\Omega_{x_0}), \|\phi\|_{m,q}=1\}} |B[u,\phi]|, \quad \forall u \in W_0^{m,p}(\Omega_{x_0}) \tag{A.15}$$

et

$$K\|v\|_{m,q} \leq sup_{\{\psi \in W_0^{m,p}(\Omega_{x_0}), \|\psi\|_{m,p}=1\}} |B[\psi,v]|, \quad \forall u \in W_0^{m,q}(\Omega_{x_0}) \tag{A.16}$$

■

Théorème 4 :

Soit $m \geq 1$ un entier et soit $\Omega \subset R^n$ $(n \geq 2)$ un ouvert borné de frontière Γ de classe C^m. Soit $B[\phi,\psi]$ une forme bilinéaire de Dirichlet régulière d'ordre m uniformément fortement elliptique et soient p, q des nombres réels avec $1 < p, q < \infty$ et $\frac{1}{p} + \frac{1}{q} = 1$.
Alors pour tout $x_0 \in \bar{\Omega}$ il existe un voisinage ouvert Ω_{x_0} de x_0, de frontière $\partial\Omega_{x_0}$ de classe C^m, tel que pour $\forall F \in W_0^{m,q}(\Omega_{x_0})^*$ $(\forall H \in W_0^{m,p}(\Omega_{x_0})^*)$, $\exists! u \in W_0^{m,p}(\Omega_{x_0})$, $(\exists! v \in\in W_0^{m,q}(\Omega_{x_0}))$ avec

$$B[u,\phi] = F(\phi), \quad \forall \phi \in W_0^{m,q}(\Omega_{x_0}) \tag{A.17}$$

$$B[\psi,v] = \bar{H}(\psi), \quad \forall \psi \in W_0^{m,p}(\Omega_{x_0}). \tag{A.18}$$

Par ailleurs, il existe une constante $C = C(x_0) > 0$ telle que:

$$C\|u\|_{m,p} \leq \|F\|_{W_0^{m,q}(\Omega_{x_0})^*} \tag{A.19}$$

$$\|v\|_{m,q} \leq \|H\|_{W_0^{m,p}(\Omega_{x_0})^*}. \tag{A.20}$$

Si B admet des coefficients constants $a_{\alpha\beta} \in C$ satisfaisant $a_{\alpha\beta} = 0$ pour $|\alpha| + |\beta| \leq 2m - 1$, alors on peut choisir $\Omega = \Omega_{x_0}$.

■

Théorème 5 :

Sous les hypothèses du théorème 4. Pour $\lambda \in R$ soit :

$$B^{\lambda}[u,\phi] = B[u,\phi] + \lambda < u,\phi > \quad pour\ (u,\phi) \in W_0^{m,p}(\Omega) \times W_0^{m,q}(\Omega). \tag{A.21}$$

Alors il existe $\lambda_0 \geq 0$ tel que $\forall \lambda \geq \lambda_0$ on a pour une certaine constante $C(\lambda,p) > 0$:

$$C(\lambda,p)\|u\|_{m,p} \leq sup_{\phi \in S_q} |B^{\lambda}[u,\phi]|, \quad \forall u \in W_0^{m,p}(\Omega) \tag{A.22}$$

$$C(\lambda,p)\|v\|_{m,q} \leq sup_{\psi \in S_p} |B^{\lambda}[\psi,v]|, \quad \forall v \in W_0^{m,q}(\Omega) \tag{A.23}$$

■

ANNEXE 3

Principe du maximum discret :

On considére le problème discret suivant :

$$\begin{cases} \int_\Omega a(x,u_h)\nabla u_h.\nabla\phi\ dx = \int_\Omega f_o\phi + \sum_{k=1}^n f_k\frac{\partial\phi}{\partial x_k}\ dx, \quad \forall\phi\in V_0^h, \\ u_h\in V_0^h. \end{cases}$$

On prend les mêmes hypothèses que pour le problème (4.1), sauf que $f_k \in L^p(\Omega)$, $0 \leq k \leq n$, $2 \leq n < p$.

lemme 1 :

Soit K un n-simplexe de R^n, non dégénéré de sommets a_r, $1 \leq r \leq n+1$; soit $v : K \longrightarrow R$ une fonction linéaire positive sur K.
Alors $\exists C_1 > 0$, indépendante de K et v telle que :

$$C_1 mes(K)\sum_{r=1}^{n+1}(v(a_r))^p \leq \|v\|_p^p. \tag{A.24}$$

Preuve :

v est positive ou nulle, on a :

$$\sum_r v(a_r)\lambda_r(x) \geq v(a_r)\lambda_r, \quad \forall r$$

$$(\sum_r v(a_r)\lambda_r(x))^p \geq (v(a_r)\lambda_r)^p, \quad \forall r.$$

En sommant ces inégalités il vient

$$(\sum_r v(a_r)\lambda_r(x))^p \geq \frac{1}{n+1}\sum_r (v(a_r)\lambda_r)^p,$$

d'où :

$$\|v\|_p^p = \int_K (\sum_{r=1}^{n+1} v(a_r)\lambda_r(x))^p dx \geq \frac{1}{n+1}\sum_{r=1}^{n+1}(\int_K (\lambda_r(x))^p dx)(v(a_r))^p.$$

Soit $\hat{K}$ un n-simplexe de référence de R^n, de sommets $\hat{a}_r$, $1 \leq r \leq n+1$.
Alors $\exists B$ une matrice inversible d'ordre n et $b \in R^n$ tels que K est l'image de $\hat{K}$ par une application $F : \hat{x} \longrightarrow F(\hat{x}) = B\hat{x} + b$, et cette application peut être choisie telle que: $F(\hat{a}_r) = a_r$, $1 \leq r \leq n+1$.

Notons par $\hat{\lambda}_r$ les coordonnées barycentriques d'un point $\hat{x} \in \hat{K}$.
on obtient :

$$\int_K (\lambda_r(x))^p dx = \int_{\hat{K}} (\lambda_r(B\hat{x}+b))^p |Jac\ F(\hat{x})|\ d\hat{x}$$
$$= \frac{mes(K)}{mes(\hat{K})} \int_{\hat{K}} (\hat{\lambda}_r(\hat{x}))^p d\hat{x}$$

on déduit finalement $(A.24)$ avec $C_1 = \frac{1}{n+1}(mes\hat{K})^{-1} min\{\int_{\hat{K}} (\hat{\lambda}_r(\hat{x}))^p d\hat{x},\ 1 \leq r \leq n+1\}$.

■

lemme 2 :

Soit τ_h une triangulation de type non négatif, soient $\xi_i,\ 1 \leq i \leq N$ des nombres réels et soit $\mu \in R$;
si on définit $\ell_i,\ 1 \leq i \leq N$ par : $\ell_i = min(\mu, \xi_i),\ 1 \leq i \leq N$ alors on a :

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_{i,j}^{u_h} (\xi_i - \ell_i)\ell_j \geq 0. \qquad (A.25)$$

Preuve :

Soit

$$I = \{1 \leq i \leq N/\ \ \xi_i > \mu\}, \quad J = \{1 \leq i \leq N/\ \ \xi_i \leq \mu\}$$

on a :

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_{i,j}^{u_h} (\xi_i - \ell_i)\ell_j = \sum_{i\in I} \sum_{j\in J} a_{i,j}^{u_h} (\xi_i - \ell_i)\ell_j + \mu \sum_{i\in I} \sum_{j\in I} a_{i,j}^{u_h} (\xi_i - \ell_i)$$

$$= \sum_{i\in I} \sum_{j\in J} a_{i,j}^{u_h} (\xi_i - \mu)\xi_j + \mu \sum_{i\in I} \sum_{j=1}^{N} a_{i,j}^{u_h} (\xi_i - \mu) - \mu \sum_{i\in I} \sum_{j\in J} a_{i,j}^{u_h} (\xi_i - \mu)$$

et ceci car $\sum_{j\in I} = \sum_{j=1}^{N} - \sum_{j\in J}$, et donc on a bien $(A.25)$.

■

lemme 3 :

Soient τ_h une triangulation de type non négatif, $u_h \in V_0^h$ et $\mu \in R$.
On pose $u_{h,\mu}$ la fonction de V_0^h définie par :

$$u_{h,\mu}(a_i) = min\{\mu, u_h(a_i)\},\ \ 1 \leq i \leq N$$

alors la fonction $v_{h,\mu} = u_h - u_{h,\mu} \in V_0^h$ et

$$B[v_{h,\mu}, v_{h,\mu}; u_h] \leq B[u_h, v_{h,\mu}; u_h]$$

avec $B[u, v; u_h] = \int_\Omega a(x, u_h) \nabla u . \nabla v\ \ dx$.

preuve :

$$B[u_h, v_{h,\mu}; u_h] = B[v_{h,\mu}, v_{h,\mu}; u_h] + B[u_{h,\mu}, v_{h,\mu}; u_h]$$

si on pose $\xi_i = u_h(a_i)$ et $\ell_i = u_{h,\mu}(a_i)$, $1 \leq i \leq N$ alors :

$$B[u_{h,\mu}, v_{h,\mu}; u_h] = \sum_{i=1}^{N}\sum_{j=1}^{N} a_{i,j}^{u_h}(\xi_i - \ell_i)\ell_j \geq 0 \qquad (A.26)$$

■

lemme 4 :

Soit $\psi : [0, +\infty) \longrightarrow R_+$ une fonction décroissante satisfaisant :

$$\psi(h) \leq (\frac{c}{h-k})^{\nu_1}.\psi(k)^{\nu_2}$$

pour $h > k$, où c, ν_1, ν_2 sont des constantes positives.
Alors:

- *si* $\nu_2 > 1$, *on a* $\psi(d) = 0$ *pour* $d = c.\psi(0)^{\frac{\nu_2-1}{\nu_1}}.2^{\frac{\nu_2}{\nu_2-1}}$
- *si* $\nu_2 < 1$, *on a* $\psi(h) \leq (\frac{2^{\frac{1}{1-\nu_2}} - c}{h})^{\nu_3}$

avec $\nu_3 = \frac{\nu_1}{1-\nu_2}$,pour $h > 0$.

preuve : voir [Ch.].

■

Théorème :

Sous les hypothèses précèdentes , $\exists$ une constante D, qui est la même pour toutes les triangulations de type non négatif, telle que :

$$\|u_h\|_\infty \leq D(\sum_{k=0}^{n} \|f\|_p) \qquad (4.27)$$

où u_h est la solution du problème discret.

preuve :

Soient μ un nombre réel positif et $v_{h,\mu}$ définie dans le lemme 3, on sait que :

$$B[v_{h,\mu}, v_{h,\mu}; u_h] \leq B[u_h, v_{h,\mu}; u_h] = (f, v_{h,\mu})$$

puisque $v_{h,\mu} \in V_0^h \subset W_0^{1,2}(\Omega)$, les fonctions $v_{h,\mu}$ et $\frac{\partial v_{h,\mu}}{\partial x_k} \in L^{p'}(\Omega)$, $1 \leq k \leq n$ avec $\frac{1}{p} + \frac{1}{p'} = 1$, ($p' < 2$ car $2 \leq n < p$).

D'après l'inégalité de Hölder, on obtient :

$$(f, v_{h,\mu}) \leq |f_0|_{p'} \|v_{h,\mu}\|_{p'} + \sum_{k=1}^{n} |f_k|_{p'} \| \frac{\partial v_{h,\mu}}{\partial x_k} \|_{p'}$$

soit : $E(\mu) = \{x \in \Omega, \quad v_{h,\mu}(x) > 0\}$

il s'en suit que : $v_{h,\mu} = \frac{\partial v_{h,\mu}}{\partial x_k} = 0, \quad 1 \leq k \leq n$ sur $\Omega - \bar{E}(\mu)$,

puisque $\bar{E}(\mu) = \bigcup_{K \in \tau_h} K$.

En utilisant l'inégalité de Hölder, on obtient :

$$\|v_{h,\mu}\|_{p'} = \|v_{h,\mu}\|_{L^{p'}(E(\mu))} \leq \|v_{h,\mu}\|_{L^2(E(\mu))} . mes(E(\mu))^{\frac{1}{2}-\frac{1}{p}}$$

d'où :

$$(f, v_{h,\mu}) \leq (n+1)(\sum_{k=0}^{n} \|f_k\|_p) \|v_{h,\mu}\|_{1,2} . mes(E(\mu))^{\frac{1}{2}-\frac{1}{p}}$$

ainsi :

$$\|v_{h,\mu}\|_{1,2} \leq \frac{n+1}{\alpha} (\sum_{k=0}^{n} \|f_k\|_p) . mes(E(\mu))^{\frac{1}{2}-\frac{1}{p}}$$

D'après l'inclusion de Sobolev, on sait que :

$$W^{1,2}(\Omega) \subset L^{2^*}(\Omega) \; avec \begin{cases} \frac{1}{2^*} = \frac{1}{2} - \frac{1}{n} & si \quad n > 2 \\ 2^* = réel \geq 1 & si \quad n = 2 \end{cases}$$

avec une injection continue, telle qu'on obtient :

$$\|v_{h,\mu}\|_{2^*} \leq C_2 (\sum_{k=0}^{n} \|f_k\|_p) . mes(E(\mu))^{\frac{1}{2}-\frac{1}{p}}$$

pour une constante C_2 indépendante de la triangulation τ_h.

Soit $\gamma > \mu$, de la même manière on définit $E(\gamma)$.

En utilisant le lemme 1, on obtient :

$$\begin{aligned} \|v_{h,\mu}\|_{2^*}^{2^*} &= \sum_{K \in E(\mu)} \int_K (v_{h,\mu}(x))^{2^*} \, dx \\ &\geq C_1 \sum_{a_i \in E(\mu)} (v_{h,\mu}(a_i))^{2^*} . mes(supp\phi_i) \\ &\geq (\gamma - \mu)^{2^*} \sum_{a_i \in E(\gamma)} mes(supp\phi_i) \\ &= C_1 (\gamma - \mu)^{2^*} . mes(E(\gamma)) \end{aligned}$$

Considérons la fonction :

$$\psi : \mu \geq 0 \longrightarrow \psi(\mu) = mes(E(\mu))$$

elle a les propriétés suivantes :

- $\psi \geq 0$ sur $[1, \infty[$,
- ψ est décroissante sur $[1, \infty[$,
- $\forall \gamma > \mu$, on a :

$$\psi(\gamma) \leq \frac{C_3}{(\gamma - \mu)^{2^*}} \psi(\mu)^{\nu}$$

avec $C_3 = (C_2 \sum_{k=0}^{n} \|f_k\|_p)^{2^*} / C_1$ et $\nu = 2^*(\frac{1}{2} - \frac{1}{p})$.
D'après le lemme 4, on déduit que $\psi(d) = 0$ avec $d = (2^{\frac{\nu}{\nu-1}} C_3 \psi(0)^{2^*(\nu-1)})^{\frac{1}{2^*}}$
donc $\forall x \in \bar{\Omega}$:

$$u_h(x) \leq (2^{\frac{\nu}{\nu-1}} C_3 \psi(0)^{\nu-1})^{\frac{1}{2^*}}$$
$$\leq D(\sum_{k=0}^{n} \|f\|_p)$$

avec $D = C_1^{\frac{-1}{2^*}} C_2 . 2^{\frac{\nu}{2^*(\nu-1)}} (mes(\Omega))^{\frac{1}{2^*}(\nu-1)}$;
la conclusion du théorème suit en observant qu'on peut similairement prouver une inégalité opposée.

■

REFERENCES

[A.] N. André : Thesis, University of Metz, (1993).

[A.C.$_1$] N. André, M. Chipot : A remark on uniqueness for quasilinear elliptic equations. Proceedings of the Banach Center, to appear.

[A.C.$_2$] N. André, M. Chipot : Uniqueness and non uniqueness for the approximation of quasilinear elliptic equations. To appear in SIAM J. of Num. Anal.

[Ar.] M. Artola : Sur une classe de problèmes paraboliques quasilinéaires.
Bollettino UMI, (6), 5-B, (1986), p. 51-70.

[B.K.S.] H. Brezis, D. Kinderlehrer and G. Stampacchia : Sur une nouvelle formulation du problème de l'écoulement à travers une digue.
C.R. Acad. Sc. Paris Série A 287, (1978), p. 711-714.

[C.] P. G. Ciarlet : The finite Element Method for Elliptic Problems.
North Holland, Amsterdam, (1987).

[C.R.] P. G. Ciarlet, P.-A.Raviart : Maximum principle and uniform convergence for the finite element method.
Comp.Met.App.Mec.Eng.2, Amsterdam, (1973).

[C.C.] J. Carrillo, M. Chipot : On nonlinear elliptic equations involving derivative of the nonlinearity.
Proc. Roy. Soc. Edinburgh, 100 A, (1985), p. 281-294.

[C.M.] M. Chipot, G. Michaille : Uniqueness results and monotonicity properties for the solution of some variational inequalities.
Annali della Scuola Norm. Sup. Pisa, Serie IV, 16, 1 (1989), p. 137-166.

[Ch.] M.Chipot : Variationnal Inequalities and Flow in Porous Media
Springer Verlag, (1984).

[F.] P. Faure : Analyse, Optimisation et Filtrage Numérique.
Cours de l'Ecole Polytechnique, 1988-89.

[G.T.] D. Gilbarg, N.S. Trudinger : Elliptic Partial Differential Equations of Second Order.
Springer Verlag, Berlin, (1985).

[K.S.] D. Kinderlehrer, G. Stampacchia: An Introduction to Variational Inequalities.
Academic Press, (1980), New York.

[Me.] N. G. Meyers : An L^p-estimate for the gradient of solutions of second order elliptic divergence equations.
Ann.Sc.Norm.Pisa (3), 17, (1963), p.189-206.

[M.] G. Michaille : Thesis, University of Metz, (1988).

[T.] N.S. Trudinger : On the comparison principle for quasilinear divergence structure equations.
Arch. Rat. Mech. Anal. (57), (1974), p. 128-133.

[S.] C. G. Simader : On Dirichlet's Boundary Value Problem
Lecture Notes in Mathematics.
Springer Verlag, (1972).

Deuxième partie :
Traitement Numérique du Signal

0. Introduction :

Dans cette partie, on se propose de résoudre un problème industriel, qui a été proposé par la société Landis & Gyr.
Landis & Gyr est une entreprise spécialisée dans la conception du matériel électrique et a comme principal client Electricité de France.
Cette entreprise souhaitait pouvoir mesurer à l'aide d'une technique numérique les puissances, les énergies actives et réactives consommées sur le réseau de distribution d'électricité par un client d'E.D.F.
A température ambiante et pour des formes d'onde courant-tension parfaites, la précision sur la mesure de l'énergie doit être inférieure à 0,5 %.
D'après les spécifications E.D.F relatives au comptage d'énergie :

- le signal issu du réseau peut comporter des fréquences harmoniques et intermédiaires comprises entre 10 Hz et 600 Hz;

- le taux d'harmonique peut atteindre 40 % et 10 % du niveau du fondamental sur le courant et la tension respectivement, ou 20 % sur chacune.

La principale difficulté qui se posait est que le signal électrique circulant sur le réseau oscille rapidement; si l'on pouvait l'échantillonner à une fréquence élevée (par exemple à la fréquence de Shanon), on aurait une bonne précision sur le calcul des grandeurs à mesurer en utilisant simplement une méthode des rectangles, mais celle-ci aurait l'inconvénient d'imposer une fréquence de travail rapide pour une application de comptage triphasé et exigera donc le choix d'une unité de calcul performante si l'on voulait garder des possibilités de satisfaire d'éventuelles demandes de calculs complémentaires (Qualité de la tension,etc,..).

On propose donc, une technique permettant de mesurer exactement ces grandeurs physiques, sans être obligé d'échantillonner le signal à une fréquence rapide.

1.Méthode de mesure d'énergie :

lemme 1.1 :

Soit g une fonction périodique de classe C^p, $p \geq 1$ de période T, alors pour tout m inférieur ou égal à p, il existe une constante C_m telle que:
Pour tout réel k non nul :

$$| \hat{g}(k) | \leq \frac{C_m}{| k |^m} \tag{1.1}$$

où $\hat{g}(k) = \frac{1}{T}\int_0^T g(t)\exp(-i2\pi\frac{kt}{T})dt$.

Preuve:

Une intégration par parties nous donne :

$$\hat{g}(k) = (\frac{1}{T})[\frac{-T}{2i\pi k}g(t)\exp(\frac{-i2\pi kt}{T})]_0^T + \frac{1}{2i\pi k}\int_0^T g'(t)\exp(\frac{-i2\pi kt}{T})dt$$
$$\hat{g}(k) = \frac{T}{2i\pi k}\hat{g'}(k).$$

Une récurrence immédiate montre que :

$$\hat{g}(k) = (\frac{T}{2i\pi k})^m \hat{g}^{(m)}(k)$$

et la constante de l'énoncé peut être prise égale à $(\frac{T}{2\pi})^m \| g^{(m)} \|_1$
avec $\| g^{(m)} \|_1 = \int_0^T | g^{(m)}(x) | dx$.

Théorème 1.2 :

Soit f une fonction périodique de classe C^m $m \geq 2$ de période T, si on lui applique la formule des rectangles en n points uniformement répartis sur $[0, T]$,on a l'estimation d'erreur:

$$| \int_0^T f(x)dx - \frac{T}{n}\sum_{j=1}^{n} f(\frac{jT}{n}) | \leq \frac{C(m,f)}{n^m}. \tag{1.2}$$

Démonstration :

On écrit le développement en série de Fourier de f :

$$f(x) = \sum_{k=-\infty}^{+\infty} \hat{f}(k)\exp(i2\pi\frac{kx}{T}).$$

On pose:

$$E_n(f) = \frac{T}{n}\sum_{j=1}^{n} f(\frac{jT}{n}) - \int_0^T f(x)dx \tag{1.3}$$

On a :

$$\begin{cases} E_n(\exp(i2\pi\frac{kx}{T})) &= T \quad \text{si k non nul et n divise k} \\ E_n(\exp(i2\pi\frac{kx}{T})) &= 0 \quad \text{dans le cas contraire .} \end{cases} \tag{1.4}$$

En effet si $k = 0$ on a $E_n(1) = 0$.
Si n ne divise pas k ,

$$\sum_{j=1}^{n} \exp(\frac{i2\pi kj}{n}) = 0$$

et par suite $E_n(\exp(i2\pi\frac{kx}{T})) = 0$.
Par contre s'il existe l entier tel que $k = nl$, alors

$$\sum_{j=1}^{n} \exp(i2\pi(\frac{k}{T})(\frac{jT}{n})) = \sum_{j=1}^{n} \exp(i2\pi lj) = n.$$

On en déduit dans ce cas $E_n(f) = T$. D'où, le développement en série de f converge

$$\begin{aligned} E_n(f) &= \frac{T}{n}\sum_{j=1}^{n}\sum_{k=-\infty}^{+\infty} \hat{f}(k)\exp(\frac{i2\pi kj}{n}) - \sum_{k=-\infty}^{+\infty} \hat{f}(k)\int_0^T \exp(i2\pi\frac{kx}{T})dx \\ &= \sum_{k=-\infty,k\neq 0}^{+\infty} \hat{f}(k)E_n(\exp(i2\pi\frac{kx}{T})) \\ &= T\sum_{l=-\infty,l\neq 0}^{+\infty} \hat{f}(ln) \end{aligned}$$

D'après le lemme pour tout réel $k \neq 0$, on a $\mid \hat{f}(k) \mid \leq \frac{C_m}{\mid k\mid^m}$.
On en déduit que :

$$\mid E_n(f) \mid \leq \frac{2TC_m}{n^m}\sum_{l\geq 1}(\frac{1}{l^m}) \tag{1.5}$$

Comme la série $\sum_{l\geq 1}(\frac{1}{l^m})$ converge pour $m \geq 2$, on a démontré le résultat énoncé.

Remarque 1.3 :

Si on écrit

$$f(x) = \frac{a_0}{2} + \sum_{j\geq 1} a_j \cos\frac{2\pi jx}{T} + b_j \sin\frac{2\pi jx}{T} \tag{1.6}$$

comme $a_j = \hat{f}(j) + \hat{f}(-j)$, on déduit que :

$$E_n(f) = T\sum_{l=1}^{+\infty}[\hat{f}(ln) + \hat{f}(-ln)] = T\sum_{l=1}^{+\infty} a_{ln} \tag{1.7}.$$

Le signal électrique parcourant le réseau d'EDF est de la forme :

$$S(t) = \sum_{j=1}^{m} A_j \sin\frac{2\pi jt}{T}. \tag{1.8}$$

On étudie le signal particulier suivant :

$$s(t) = A\sin\frac{2\pi t}{T} + B\sin\frac{2\pi mt}{T} \tag{1.9}$$

et on cherche à calculer l'énergie $\int_0^T s^2(t)dt$,

$$\begin{aligned} f(t) = \; & s^2(t) \\ = \; & \frac{A^2+B^2}{2} - \frac{A^2}{2}\cos\frac{4\pi t}{T} - \frac{B^2}{2}\cos\frac{4\pi mt}{T} \\ & + AB[\cos\frac{2\pi(m-1)t}{T} - \cos\frac{2\pi(m+1)t}{T}]. \end{aligned} \tag{1.10}$$

Ce signal électrique est une fonction rapidement oscillante, et on est lié à une contrainte, à savoir la fréquence d'échantillonnage.
Soient T' la période d'échantillonnage et $K = \frac{T}{T'}$ la constante d'échantillonnage.
On peut penser, dans un premier temps, à appliquer directement une technique classique d'intégration numérique, (méthode des rectangles, trapèzes, Simpson), mais du fait que le signal étudié est rapidement oscillant, ces méthodes donneraient une mauvaise approximation de l'énergie, car on ne peut avoir suffisamment d'échantillons sur une période.
Si on prend un échantillon tous les T' et on le ramène "par périodicité" sur $[0,T]$, on obtient une suite de points uniformément répartis sur cet intervalle, et on est conduit ainsi à appliquer (d'une manière indirecte) la méthode des rectangles avec une estimation d'erreur fournie par (1.2), et ceci en remarquant que les méthodes des rectangles, trapèzes et de Simpson coincident pour une fonction périodique.

Supposons $T = KT'$, avec K décimal. On écrit K en base 10 :

$$K = \sum_{i=0}^{q} n_i 10^{-i}. \tag{1.11}$$

Considérons w l'infimum des entiers positifs z tels que zK soit entier et posons $N = wK$.

Proposition 1.4 :

Soit $x_j = jT'(modT)$, $(x_j)_{1 \leq j \leq N}$ est une suite de points distincts sur $[0, T]$

Preuve:

Supposons qu'il existe deux entiers m_1, m_2; $m_1 > m_2 \in \{1, 2, ..., N\}$ tels que :

$$x_{m_1} = x_{m_2}.$$

On a donc :

$$m_1 T' = m_2 T' \quad (modT),$$

soit encore :

$$(m_1 - m_2)T' = ZT = ZKT'.$$

On a donc

$$N \geq m_1 - m_2 = ZK \in N.$$

D'où nécessairement $Z = w$. Mais alors

$$m_1 = m_2 + wK = m_2 + N > N$$

ce qui est impossible.

Proposition 1.5 :

Soient

$$U = \{x_j / 1 \leq j \leq wK\}, \quad V = \{\frac{jT'}{w} / 0 \leq j \leq wK - 1\}$$

alors on a $U = V$.

Démonstration:

soit u un élément de U, il s'écrit alors $u = jT'(modT)$,

il existe deux entiers ξ, ν avec $\nu < wK$ tels que $wj = \xi wK + \nu \Rightarrow j = \xi K + \nu/w$
d'où $u = \xi T + \frac{\nu T'}{w}$ et finalement : $u = \frac{\nu T'}{w}$ qui appartient clairement à V.
On a montré que $U \subset V$, comme $CardU = CardV$, on conclut que $U = V$.

Conclusion 1.6 :

Il y a deux points importants :

1) $E_n(f) = T \sum_{l=-\infty, l\neq 0}^{+\infty} \hat{f}(ln)$.

2) prendre wK échantillons est équivalent à diviser T en wK parties.

Ce qui nous permet de conclure que :

$$\begin{aligned} E_{wK}(f) &= \frac{T}{wK} \sum_{j=1}^{wK} f(\frac{jT}{wK}) - \int_0^T f(x)dx \\ &= \frac{T}{wK} \sum_{j=1}^{wK} f(x_j) - \int_0^T f(x)dx \\ &= \frac{T}{wK} \sum_{j=1}^{wK} f(jT') - \int_0^T f(x)dx \end{aligned}$$

d'après ce qui a été démontré en (1.7).

$$E_{wK}(f) = T(a_{wK} + a_{2wK} + ...) \tag{1.12}$$

Si K est bien choisi c.a.d tel que :

$$wK > sup\{y \ entiers \ / \ a_y \neq 0 \text{ dans le développement de } s^2(t)\} \tag{1.13}$$

alors $E_{wK}(s^2) = 0$ et par suite il n' y a pas d'erreur mathématique commise si l'on remplace $\int_0^T f(x)dx$ par $\frac{T}{wK} \sum_{j=1}^{wK} f(jT')$.
D'aprés ce qui précède on a le résultat suivant:

Conclusion 1.7 :

Si (1.13) est vérifiée, alors on a

$$\int_0^T s^2(t)dt = \frac{T}{wK} \sum_{j=1}^{wK} s^2(jT') \tag{1.14}$$

Remarque 1.8 :

Lorsqu'on est en présence d'un signal pourvu de plusieurs fréquences harmoniques la méthode de ci-dessus reste valable.
En effet:

$$S(t) = \sum_{j=1}^{m} A_j \sin \frac{2\pi jt}{T}$$

$$S^2(t) = \sum_{j=1}^{m} A_j^2 \sin^2 \frac{2\pi jt}{T} + 2A_pA_q \sum_{p<q} \sin \frac{2\pi pt}{T} \sin \frac{2\pi qt}{T}$$

$$S^2(t) = a_0/2 + \sum_{j=1}^{m}(-A_j^2/2)cos\frac{4\pi jt}{T} + \sum_{p<q} A_pA_q(\cos \frac{2\pi(p-q)t}{T} - \cos \frac{2\pi(p+q)t}{T})$$

$$sup\{y \ entiers \ / \ a_y \neq 0 \text{ dans le développement de } S^2(t)\} = 2m$$

Le problème posé ici ne diffère pas de celui qu'on aura avec un signal pourvu d'une fréquence harmonique de rang m.
Si on choisit un K tel que $wK > 2m$ on a la valeur exacte de $\int_0^T S^2(t)dt$.

Proposition 1.9 :

Soit S(t) un signal électrique pourvu de plusieurs fréquences harmoniques. Supposons que wK soit non supérieur à 2m, alors il existe $l \geq 1$ tel que :

$$\int_0^T S^2(t)dt = \tau \sum_{i=0}^{l-1} \sum_{j=1}^{wK} S^2(jT' + i\tau) \tag{1.15}$$

avec $\tau = \frac{T}{lwK}$.

Preuve :

Considérons $E_{wK} = T(a_{wK} + a_{2wK} + ...) \neq 0$, posons

$$W = \{x \ entiers / a_{xwK} = 0 \ dans \ E_{wK}\}. \tag{1.16}$$

Cet ensemble est non vide, il suffit de considérer un x tel que $xwK > 2m$.

Soit $l = minW$, l est le premier entier x tel que $a_{xwK} = 0$, on a alors $E_{lwK} = 0$, $(l > 1)$

$$\{0, 1, 2, ..., lwK - 1\} = \bigcup_{i=0}^{l-1} \{lk + i/0 \leq k \leq wK - 1\} \tag{1.17}$$

on a :

$$\sum_{j=0}^{lwK-1} S^2(\frac{jT}{lwK}) = \sum_{i=0}^{l-1} \sum_{k=0}^{wK-1} S^2(\frac{kT}{wK} + i\tau).$$

D'après ce qui précède

$\{\frac{kT}{wK} + i\tau/0 \leq k \leq wK - 1\} = \{jT' + i\tau(modT)/0 \leq j \leq wK - 1\}$

Ce qui nous permet d' affirmer que:

$$\sum_{k=0}^{wK-1} S^2(\frac{kT}{wK} + i\tau) = \sum_{j=0}^{wK-1} S^2(jT' + i\tau)$$

on obtient finalement (1.15).

2.Calcul des Coefficients de Fourier :

Le signal parcourant une ligne électrique est de la forme:

$$S(t) = \sum_{j=1}^{m} A_j \sin \frac{2\pi jt}{T}.$$

On étudie le signal particulier suivant:

$$s(t) = A \sin \frac{2\pi t}{T} + B \sin \frac{2\pi mt}{T}$$

$A = \frac{2}{T} \int_0^T s(t). \sin \frac{2\pi t}{T} dt$

On pose : $g_1(t) = s(t). \sin \frac{2\pi t}{T}$,

$$g_1(t) = \frac{A}{2} - \frac{A}{2} \cos \frac{4\pi t}{T} + \frac{B}{2} \cos \frac{2\pi(m-1)t}{T} - \frac{B}{2} \cos \frac{2\pi(m+1)t}{T}$$

$E_N(g_1) = T \sum_{l=-\infty}^{+\infty} \hat{g}_1(lN)$.

Si on choisit un K tel que $N = wK > 2m$, on obtient :

$$A = \frac{2}{N} \sum_{j=1}^{N} s(jT') \sin(\frac{2j\pi}{K})$$

$B = \int_0^T s(t). \sin \frac{2\pi mt}{T} dt.$

On pose : $g_2(t) = s(t). \sin \frac{2\pi.mt}{T}$,

$$g_2(t) = \frac{B}{2} - \frac{B}{2} \cos \frac{4\pi mt}{T} + \frac{A}{2} \cos \frac{2\pi(m-1)t}{T} - \frac{A}{2} \cos \frac{2\pi(m+1)t}{T}$$

et

$$B = \frac{2}{N} \sum_{j=1}^{N} s(jT') \sin(\frac{2j.m\pi}{K})$$

Plus généralement pour un signal S(t) pourvu de plusieurs fréquences harmoniques $S(t) = \sum_{j=1}^{m} A_j \sin \frac{2\pi jt}{T}$

Pour tout $k \in \{1, 2, ..., m\}$, on a :

$$A_k = \frac{2}{T} \int_0^T s(t). \sin \frac{2\pi kt}{T} dt$$

$S(t). \sin \frac{2\pi kt}{T} = \frac{1}{2} \sum_{j=1}^{m} A_j [\cos \frac{2\pi(j-k)t}{T} - \cos \frac{2\pi(j+k)t}{T}]$
et on conclut que :

$$A_k = \frac{2}{N} \sum_{j=1}^{N} S(jT') \sin(\frac{2j.k\pi}{K})$$

Méthode de calcul rapide :

Si on pose : $C_k = \sum_{j=1}^{N} S(jT') \exp(-ikj\frac{2w\pi}{N})$
on a alors :

$$\begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_m \end{pmatrix} = \begin{pmatrix} V & V^2 & \dots & V^{N-1} \\ V^2 & V^4 & \dots & V^{2N-2} \\ \vdots & \ddots & \ddots & \vdots \\ V^m & V^{2m} & \dots & V^{m(N-1)} \end{pmatrix} \cdot \begin{pmatrix} S(T') \\ S(2T') \\ \vdots \\ S((N-1)T') \end{pmatrix}$$

Posons pour k variant de $m+1$ à $N-1$: $C_k = \sum_{j=1}^{N} S(jT') \exp(-ikj\frac{2w\pi}{N})$
En complétant par des élements $C_{m+1}, C_{m+2}, ..., C_{N-1}$, on obtient une matrice carrée et donc on peut essayer d'utiliser les algorithmes "Transformée de fourier rapide" et ne s'intéresser qu'aux valeurs $C_1, ..., C_m$.
On étudie le cas où N est un nombre Premier :
$\forall k \in \{1, 2, ..., N-1\}$ on a :

$$C_k = \sum_{n=1}^{N-1} x_n V^{nk}$$

avec

$$\begin{cases} x_n &= S(nT') \\ V &= \exp(-i\frac{2w\pi}{N}) \end{cases}$$

Théorème 2.1 :

Etant donné un nombre premier N, il existe un entier g (appelé Racine primitive modulo N) tel que pour u prenant une fois et une seule les valeurs entre 0 et $N-2$, $n = g^u \ mod N$ prenne toutes les valeurs possibles entre 1 et $N-1$.

En posant $n = g^u \ \ modN$ et $n = g^v mod \ \ N$, la relation donnant les valeurs C_k devient :

$$C_{g^v} = \sum_{u=0}^{N-2} x_{g^u} V^{g^{u+v}}$$

On reconnait la formule de la corrélation circulaire entre les suites de terme général x_{g^u} et V^g .
exemple : N=5

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{pmatrix} = \begin{pmatrix} V & V^2 & V^3 & V^4 \\ V^2 & V^4 & V & V^3 \\ V^3 & V & V^4 & V^2 \\ V^4 & V^3 & V^2 & V \end{pmatrix} \cdot \begin{pmatrix} S(T') \\ S(2T') \\ S(3T') \\ S(4T') \end{pmatrix}$$

2 est racine primitive mod 5.

$$\begin{cases} 2^0 = 1 mod 5 \\ 2^1 = 2 mod 5 \\ 2^2 = 4 mod 5 \\ 2^3 = 3 mod 5 \end{cases}$$

On en déduit l'isomorphisme :

$$\begin{cases} 0 \longrightarrow 1 \\ 1 \longrightarrow 2 \\ 2 \longrightarrow 4 \\ 3 \longrightarrow 3 \end{cases}$$

Ce qui nous permit d'écrire :

$$\begin{pmatrix} C_1 \\ C_2 \\ C_4 \\ C_3 \end{pmatrix} = \begin{pmatrix} V & V^2 & V^4 & V^3 \\ V^2 & V^4 & V^3 & V \\ V^4 & V^3 & V & V^2 \\ V^3 & V & V^2 & V^4 \end{pmatrix} \cdot \begin{pmatrix} S(T') \\ S(2T') \\ S(4T') \\ S(3T') \end{pmatrix}$$

La matrice, après permutation présente une allure circulante et la phase du calcul du produit d'une telle matrice avec le vecteur de composantes $x(g^u)$ se transpose en un problème particulier de calcul de produits de polynômes.
Le problème de calcul de la convolution cyclique de deux ensembles de N points
$a_0, a_1, ..., a_{N-1}$ et $g_0, g_1, ..., g_{N-1}$ est défini par le produit :

$$\begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{N-1} \end{pmatrix} = \begin{pmatrix} g_0 & g_1 & \cdots & g_{N-1} \\ g_1 & g_2 & \cdots & g_0 \\ \vdots & \ddots & \ddots & \vdots \\ g_{N-1} & g_0 & \cdots & g_{N-2} \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{pmatrix}$$

On peut vérifier que les N composantes de ce produit correspondent aux coefficients du produit polynômial

$$G(z).A(z)\ mod(z^N - 1) \qquad (*)$$

où

$$G(z) = g_0 + g_1 z + ... + g_{N-1} z^{N-1}$$
$$A(z) = a_0 + a_{N-1} z + ... + a_1 z^{N-1}.$$

Un théorème de Winograd démontre que le nombre minimal de multiplications nécessaires pour programmer le produit (*) est $2N - k$ où k est le nombre de diviseurs de N (cf. [Win.1]).

Dans le cas de la transformée de Fourier, Winograd a montré que le nombre de multiplications pouvait être ramené à $2N - d(N-1) - 3$ où $d(N-1)$ désigne le nombre de diviseurs de $N - 1$ (cf [Win.2]).

■

3. Remarque Générale :

1• L'erreur de mesure est dûe à un grand nombre de causes distinctes, et il est naturel d'admettre que ces causes d'erreur sont indépendantes entre elles.
D'autre part, si nous éliminons de nos considérations les erreurs systématiques qui agissent dans un sens déterminé, pour ne retenir que les erreurs dûes au hasard, nous sommes amenés à considérer les erreurs comme des variables aléatoires centrées (c'est à dire grosso modo comme pouvant être aussi bien positives que négatives).
Nous sommes ainsi conduits à penser que l'erreur de mesure est la somme d'un grand nombre de variables aléatoires centrées dont aucune n'est sans doute prépondérente.
Ceci nous conduit à la conclusion suivante, l'erreur de mesure suit approximativement une loi de Gauss.

2• On a proposé des algorithmes pour pouvoir faire une analyse fréquentielle de la tension distribuée et du courant consommé; et on peut constater que, vu l'expression de l'erreur fournie par (1.7), l'utilisation d'algorithmes non adaptés, peut générer des erreurs importantes, et ceci a été confirmé par l'étude expérimentale faite par Landis & Gyr (service des développements exploratoires).

Par ailleurs, l'étude des signaux non stationnaires, où apparaissent des évenements transitoires, que l'on ne pouvait prévoir nécessite des techniques différentes de l'analyse de Fourier classique, car elle ne permet pas d'accéder aux caractéristiques évolutives jugées pertinentes du signal. Une approche adéquate est la représentation temps-fréquence qui permet de donner sens à une notion d'analyse spectrale évolutive car elle prend en compte une possible évolution temporelle du contenu fréquentiel du signal.
L'utilisation des Ondelettes peut s'avérer fondamentale pour une étude plus approfondie.

ANNEXE

Caractérisation des transitions entre deux régimes stationnaires :

Soit le signal $S(t)$ circulant dans une ligne éléctrique , on souhaite suivre l'état du réseau sur une plage de temps de durée T.
On dispose de N échantillons de $S(t)$ sur chaque plage de temps, à savoir :
$S(t_1), S(t_2), ..., S(t_N)$ avec $t_j = j\frac{T}{N}$.

Une première approche est de faire une approximation discrète au sens des moindres carrés de $S(t)$ par une fonction $g \in E = \{f/f(t) = \sum_{k=-n}^{k=n} C_k e^{\frac{i2k\pi t}{T}}\}$.
On sait que cette approximation existe et est unique, dans notre cas, ceci revient à tronquer le signal entre $-n$ et n.

Une deuxième approche est d'identifier le signal $S(t)$ sous la forme suivante :

$$S(t) = \sum_{l=1}^{p} C_l e^{iw_l t} \tag{3.1}$$

On cherche à déterminer les fréquences composant le signal et les amplitudes correspondantes .
La donnée de $N = 2p$ échantillons nous permet de calculer C_l et w_l.

$$S(nT') = \sum_{l=1}^{p} C_l e^{i(w_l T').n} \tag{3.2}$$

on pose :
$x_l = C_l e^{iw_l T'}$ et $z_l = e^{iw_l T'}$,

d'où :

$$S(nT') = \sum_{l=1}^{p} x_l z_l^{n-1} \quad pour \quad 1 \leq n \leq N \tag{3.3}$$

Supposons , dans un premier temps que les z_l sont connus , il suffit de résoudre le systéme (3.3) pour $1 \leq n \leq p$ de type Vander-Monde , pour calculer les p inconnues x_l.

Les z_l étant connus , on peut définir le polynôme $P(z)$ suivant :

$P(z) = \prod_{l=1}^{p}(z - z_l) = \sum_{i=0}^{p} a_i z^{p-i}$ avec $a_0 = 1$
on pose $u_n = S(nT')$

considérons : $S_n = \sum_{k=0}^{p} a_k u_{n-k}$ pour $p + 1 \leq n \leq 2p$

$$S_n = \sum_{k=0}^{p} a_k \sum_{l=1}^{p} x_l z_l^{n-k-1} = \sum_{l=1}^{p} x_l z_l^{n-p-1} \sum_{k=0}^{p} a_k z_l{}^{p-k}$$

d'où $\forall n \in \{p+1, ..., 2p\}$ on a : $\sum_{k=0}^{p} a_k u_{n-k} = 0$.

et on obtient le systéme suivant :

$n = p+1, ..., 2p.$ $\qquad -u_n = \sum_{k=1}^{p} a_k u_{n-k}$

ou encore :

$$\begin{cases} a_1 u_p + a_2 u_{p-1} + ... + a_p u_1 &= -u_{p+1} \\ a_1 u_{p+1} + a_2 u_p + ... + a_p u_2 &= -u_{p+1} \\ ..... &= ..... \\ a_1 u_{2p-1} + a_2 u_{2p-2} + ... + a_p u_p &= -u_{2p} \end{cases} \tag{3.4}$$

Soit :

$$\begin{pmatrix} u_p & u_{p-1} & u_{p-2} & .... & u_1 \\ u_{p+1} & u_{p+2} & u_{p+3} & .... & u_2 \\ .. & .. & .. & .... & .. \\ u_{2p-1} & u_{2p-2} & u_{2p-3} & .... & u_p \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ .. \\ a_p \end{pmatrix} = - \begin{pmatrix} u_{p+1} \\ u_{p+2} \\ .. \\ u_{2p} \end{pmatrix}$$

La résolution de ce système fournit les coefficients a_k.

Ce qui nous permet de déterminer les z_l et par suite la résolution du système (3.3) pour le calcul des x_l.

REFERENCES

[Bar.] J. Baranger : Analyse numérique.
Hermann, (1991).

[Win.1] S. Winograd : On the computing the discrete Fourier transform.
Mathematics of computation, vol 32, 141, (1978), p. 175-199.

[Win.2] S. Winograd : On the multiplicative complexity of the discrete Fourier transform.
Advances in Math., vol 32, (1979), p. 83-117.