

Méthodes numériques pour un problème inverse

PAR MAXIME MOGÉ

Projet de Master supervisé par Philippe Caussignac

Janvier 2011



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Section de Mathématiques

EPFL

SUISSE

Table des matières

1	Introduction	3
2	Eléments finis discontinus	3
2.1	Notations	3
2.2	Formulation du problème	3
2.2.1	Définitions et introduction du formalisme de Friedrichs	3
2.2.2	Approximation par éléments finis (discontinus)	5
2.2.3	Application à un problème inverse 1D	6
2.3	Etude de 3 situations en 1D	7
2.3.1	Cas général	7
2.3.2	Problème 1	10
2.3.3	Problème 2	12
2.3.4	Problème 3	16
2.4	Données avec bruit	20
2.4.1	Tentatives d'application des mêmes méthodes	21
2.4.2	Problème 1	21
2.4.3	Problème 2	22
2.4.4	Problème 3	24
2.4.5	Idée de solution : Interpolation par splines cubiques avec contraintes de convexité	25
3	Méthode OLS	30
3.1	Formulation du problème	30
3.2	Tentative de résolution du problème approché	31
3.3	Modification de la méthode de résolution du problème approché	32
3.3.1	Méthode de la plus grande pente hybride	32
3.3.2	Calcul du gradient	33
3.3.3	Résolution du problème approché	34
3.4	Application à un problème 1D	35
3.4.1	Application au problème 1	36
3.4.2	Application au problème 2bis	38
4	MOLS	41
4.1	Formulation du problème	41
4.2	Discrétisation	42
4.3	Résolution du problème discrétisé	42
4.4	Application à un problème 1D	44
4.4.1	Application au problème 4	45
5	Conclusion	48
	Bibliographie	49

1 Introduction

Un problème inverse est une situation dans laquelle on tente de déterminer les paramètres d'un modèle à partir de mesures expérimentales.

Ce projet a pour objet l'estimation d'un coefficient dans une équation aux dérivées partielles elliptique à partir de mesures de la solution d'un problème aux limites associé.

Par exemple pour une équation

$$\begin{cases} -\nabla \cdot (a\nabla u) = f \text{ sur } \Omega \\ + \text{Conditions aux Limites sur } \partial\Omega, \end{cases} \quad (1)$$

le problème direct consiste à trouver u avec a et f donnés, et le problème inverse consiste à trouver a à partir de mesures de u et f .

Afin de résoudre le problème inverse, nous nous intéressons à trois méthodes numériques : pour la première on réduit le problème à un système de Friedrichs positif, la deuxième est une méthode de type OLS (Output Least Squares), et la troisième une méthode MOLS (Modified OLS).

Les méthodes étudiées le sont pour un problème 1D, mais se généralisent pour des problèmes 2D, voire 3D.

2 Eléments finis discontinus

Dans une équation aux dérivées partielles de la forme (1), on peut voir le problème inverse comme une équation différentielle pour a . On considère alors l'équation obtenue comme un système différentiel du premier ordre symétrique au sens de Friedrichs. Ceci permet d'assurer l'existence et l'unicité d'une solution sous certaines conditions, et on peut alors résoudre numériquement le problème à l'aide de méthodes d'éléments finis. La méthode pour les systèmes différentiels du premier ordre symétriques au sens de Friedrichs décrite dans cette partie est tirée de P.Lesaint [6].

2.1 Notations

Soit Ω un ouvert de \mathbb{R}^p de frontière $\partial\Omega$, C^1 par morceaux.

On considère l'espace $(L^2(\Omega))^p$ (noté L^2) muni du produit scalaire

$$(u, v)_{L^2} = \int_{\Omega} (u(x), v(x)) dx,$$

où (\cdot, \cdot) désigne le produit scalaire dans \mathbb{R}^p .

L'opérateur A est donné par

$$Au(x) = \sum_{i=1}^n A_i(x) \frac{\partial u}{\partial x_i}(x) + A_0(x)u(x).$$

Les matrices $A_i \in \mathcal{L}(\mathbb{R}^p)$ sont symétriques à coefficients lipschitziens en x .

La matrice $A_0 \in \mathcal{L}(\mathbb{R}^p)$ n'est pas nécessairement symétrique et a ses coefficients dans $L^\infty(\Omega)$.

2.2 Formulation du problème

2.2.1 Définitions et introduction du formalisme de Friedrichs

Le problème (P) consiste à trouver une fonction u solution de

$$Au = f. \quad (2)$$

Définition 1. u est une solution faible de (2) si $u \in L^2$ et u satisfait le système au sens des distributions.

Définition 2. On dit que l'opérateur A défini sur un sous-espace W de $(C^1(\bar{\Omega}))^p$ est coercif s'il existe une constante $\alpha > 0$ telle que

$$(Av, v)_{L^2} \geq \alpha \|v\|_{L^2}^2, \forall v \in W.$$

On a

$$\begin{aligned} 2(Av, v)_{L^2} &= 2 \int_{\Omega} (A(x)v(x), v(x)) dx \\ &= 2 \int_{\Omega} \left(\sum_{i=1}^n \left(A_i(x) \frac{\partial v}{\partial x_i}(x), v(x) \right) + (A_0(x)v(x), v(x)) \right) dx \\ &= \int_{\Omega} (Cv, v) dx + \int_{\partial\Omega} (Bv, v) ds, \end{aligned} \quad (3)$$

où

$$\begin{aligned} C(x) &= A_0(x) + A_0(x)^T - \sum_{i=1}^n \frac{\partial A_i}{\partial x_i}(x) \quad \forall x \in \Omega, \\ B(x) &= \sum_{i=1}^n \nu_i(x) \frac{\partial A_i}{\partial x_i}(x) \quad \forall x \in \partial\Omega, \end{aligned}$$

$\nu_i(x)$ désignant les composantes de la normale extérieure unitaire à $\partial\Omega$.

Définition 3. On dit que l'opérateur A est positif au sens de Friedrichs s'il existe $\gamma > 0$ tel que

$$x^T C(x) x \geq 2\gamma x^T x, \forall x \in \mathbb{R}^p.$$

Proposition 4. Si l'opérateur A est positif au sens de Friedrichs, on a

$$\int_{\Omega} (Cv, v) dx \geq 2\gamma \|v\|_{L^2}^2, \forall v \in C^1(\bar{\Omega}).$$

Donc si l'opérateur A est positif au sens de Friedrichs avec constante de positivité γ , il suffit d'avoir $\int_{\partial\Omega} (Bv, v) ds \geq 0, \forall v \in W$ sous – espace de $C^1(\bar{\Omega})$ pour avoir A coercif sur W . De plus, γ est alors une constante de coercivité de A .

Proposition 5. Pour

$$W = \{v \in C^1(\bar{\Omega}); B(x)v(x) = M(x)v(x), \text{ pour } x \in \partial\Omega\} \quad (4)$$

avec $M \in \mathcal{L}(\mathbb{R}^p)$ à coefficients continus pour $x \in \partial\Omega$ et telle que $M + M^T$ est semi-définie positive, on a

$$\int_{\partial\Omega} (Bv, v) ds \geq 0, \forall v \in W.$$

Définition 6. Les conditions aux limites exprimées dans (4) sont dites semi-admissibles si la matrice $M + M^T$ est semi-définie positive.

Donc si l'opérateur A est positif au sens de Friedrichs et que le sous-espace W est défini comme dans la proposition 5, on peut déduire de (3) que l'opérateur A est coercif.

On pose alors le problème

$$\begin{cases} \text{Trouver } u \text{ tel que} \\ Au = f \text{ dans } \Omega \\ (B - M)u = 0 \text{ sur } \partial\Omega \end{cases} \quad (5)$$

Définition 7. u est appelée solution faible du problème (5) si $u \in L^2$ et si on a

$$(u, A^* \varphi)_{L^2} = (f, \varphi)_{L^2}, \quad \forall \varphi \in C^1(\bar{\Omega}) \text{ avec } (B + M^T)\varphi = 0 \text{ sur } \partial\Omega.$$

Théorème 8. Si l'opérateur A est positif au sens de Friedrichs et que les conditions aux limites exprimées dans (4) sont semi-admissibles, alors pour toute fonction $f \in L^2$, le problème (5) a au moins une solution faible qui vérifie

$$\|u\|_{L^2} \leq \frac{1}{\gamma} \|f\|_{L^2},$$

où γ est la constante de positivité de l'opérateur A .

Définition 9. u est appelée solution forte du problème (5) si $u \in (L^2(\bar{\Omega}))^p$ et si il existe une suite $\{u_j\}$ de fonctions de $(C^1(\bar{\Omega}))^p$, telles que $(B - M)u_j = 0$ sur $\partial\Omega$ et

$$\lim_{j \rightarrow +\infty} (\|u_j - u\|_{L^2} + \|Au_j - f\|_{L^2}) = 0.$$

Lemme 10. Toute solution forte est aussi solution faible

Définition 11. Si la condition aux limites $(B - M)u = 0$ est semi-admissibles et que la matrice M vérifie

$$\text{Im}(B - M) \cap \text{Im}(B + M) = \{0\}$$

ou

$$\text{Ker}(B + M^T) + \text{Ker}(B - M^T) = \mathbb{R}^p.$$

Alors la condition aux limites est dite admissible.

Théorème 12. Soit A un opérateur positif au sens de Friedrichs. Supposons que $\partial\Omega$ est de classe C^2 , que les conditions aux limites sont admissibles et que $\text{Ker}(B - M)$ varie continuellement sur $\partial\Omega$, alors pour toute fonction $f \in L^2$, le problème (5) a au moins une unique solution forte. Cette solution forte vérifie

$$\|u\|_{L^2} \leq \frac{1}{\gamma} \|f\|_{L^2},$$

où γ est la constante de positivité de l'opérateur A .

2.2.2 Approximation par éléments finis (discontinus)

Méthode de Rayleigh Ritz Galerkin

Soit V_h un sous-espace de dimension finie inclus dans $(H^1(\Omega))^p$, où h est un paramètre destiné à tendre vers zéro.

On définit une solution approchée u_h du problème (5), avec A positif au sens de Friedrichs et W défini comme dans la proposition 5, comme une solution du problème approché

$$\begin{cases} \text{Trouver } u_h \in V_h \text{ tel que} \\ (Au_h, v_h)_{L^2} - \frac{1}{2} \int_{\partial\Omega} ((B - M)u_h, v_h) ds = (f, v_h)_{L^2}, \forall v_h \in V_h. \end{cases} \quad (6)$$

Soit T_h un maillage de $\bar{\Omega}$ de type éléments finis.

$\forall K \in T_h$, on pose $h_K = \text{diam}(K)$. $h = \max \{h_K, K \subset \bar{\Omega}\}$.

On considère l'espace d'éléments finis $X_h^r = \{v_h \in C^0(\bar{\Omega}); v_h|_K \in \mathbb{P}_r \forall K \in T_h\}$, où \mathbb{P}_r désigne l'espace des polynôme de degré $\leq r$.

On définit $V_h = X_h^r$.

Alors on a le résultat suivant concernant la convergence

Théorème 13. *On suppose que la solution du problème (5) appartient à $H^{r+1}(\Omega)$.*

Alors si la solution du problème (6) est construite à partir de l'espace V_h décrit ci-dessus, on a pour $r \geq 1$

$$\|u - u_h\|_{L^2} = O(h^r)$$

Methodes d'éléments finis discontinues

Pour les méthodes discontinues, on généralise à chaque maille K le traitement des conditions aux limites décrit par la formule (6).

Soit $W_h = \{v_h \in (L^2(\Omega))^p; v_h|_K \in \mathbb{P}_r \forall K \in T_h\}$

On considère le problème

$$\left\{ \begin{array}{l} \text{Trouver } u_h \in W_h \text{ tel que } \forall v_h \in W_h \\ \sum_{K \subset \Omega} \left\{ (Au_h - f, v_h)_{L^2(K)} - \int_{\partial K} \left(\frac{B_K - M_K}{2} (u_h^{\text{int}} - u_h^{\text{ext}}), v_h^{\text{int}} \right) ds \right\} = 0 \end{array} \right. \quad (7)$$

où $B_K(x) = \sum_{i=1}^n A_i(x) \nu_{i,K}(x)$ pour $x \in \partial K$, avec $\vec{\nu}_K = (\nu_{1,K}, \dots, \nu_{n,K})$ la normale extérieure unitaire à ∂K .

Les fonctions u_h^{int} et u_h^{ext} désignent les valeurs de u_h le long de ∂K à l'intérieur et à l'extérieur respectivement de K . Pour une face $S \subset \partial\Omega$ on impose $u_h^{\text{ext}} = 0$ s'il s'agit d'un problème homogène.

Pour les faces incluses dans $\partial\Omega$, on a $M_K = M$ donné par les conditions aux limites.

Les autres matrices M_K sont à choisir en respectant pour toutes les faces S intérieures au domaine Ω

$$M_{K_1,S} = M_{K_2,S} = M_S, \text{ avec } M_S + M_S^T \text{ semi-définie positive}$$

si K_1 et K_2 sont deux éléments adjacents le long d'une face S .

Théorème 14. *Si les matrices M_K sont définies en satisfaisant les conditions décrites ci-dessus, alors le problème (7) admet une solution unique $u_h \in W_h$ et on a*

$$\alpha \|u_h\|_{L^2} \leq \|f\|_{L^2}$$

où α est la constante de positivité de l'opérateur A .

Théorème 15. *Si $M_S + M_S^T$ est définie positive (uniformément en h) pour tout S face d'un élément K , $S \not\subset \partial\Omega$, et que W_h est construit avec des éléments simpliciaux et des polynômes de degré $\leq k$, on a*

$$\|u - u_h\|_{L^2(\Omega)} = O(h^{k+1/2})$$

2.2.3 Application à un problème inverse 1D

Pour une équation

$$\left\{ \begin{array}{l} -(au')' = f \text{ sur } \Omega \\ + \text{Conditions aux Limites sur } \partial\Omega \end{array} \right.$$

avec $\Omega = (x_0, x_1)$, en supposant connues les fonctions u et f . En pratique, on a des mesures de u et f , et on obtient des approximations de ces fonctions par interpolation (par exemple avec des splines, afin de pouvoir imposer une certaine régularité).

Le problème inverse peut donc s'écrire

$$\begin{cases} \text{Trouver } a \text{ tel que} \\ Aa = f \text{ sur } \Omega \\ + \text{ conditions aux limites sur } \partial\Omega \end{cases}$$

avec

$$\begin{aligned} Aa(x) &= A_1(x)a'(x) + A_0(x)a(x), \forall x \in \Omega, \\ A_1(x) &= u'(x), \quad A_0(x) = u''(x), \forall x \in \Omega, \\ C(x) &= 2A_0(x) - A_1'(x) = u''(x), \forall x \in \Omega, \\ B(x) &= \nu(x)A_1(x), \forall x \in \partial\Omega, \text{ avec } \nu(0) = -1 \text{ et } \nu(1) = 1. \end{aligned}$$

On définit

$$\begin{aligned} M(x) &= |A_1(x)| = |u'(x)|, \forall x \in \Omega, \\ W &= \{v \in C^1(\bar{\Omega}); B(x)v(x) = M(x)v(x), \text{ pour } x \in \partial\Omega\}. \end{aligned}$$

On remarque que M est semi-définie positive, donc si $\exists \alpha > 0; C(x) = u''(x) \geq \alpha, \forall x \in \Omega$, alors l'opérateur A est positif au sens de Friedrichs, et donc A est coercif sur W d'après la Proposition 5.

L'existence d'une solution faible au problème inverse est alors donnée par le Théorème 7.

On applique alors une méthode d'éléments finis discontinus afin d'obtenir une solution approchée.

On choisit un maillage T_h de Ω , et $M_K(x) = |A_1(x)| = |u'(x)|, \forall x \in K, \forall K \in T_h$, ce qui garantit l'existence et l'unicité de la solution au problème approché (7).

2.3 Etude de 3 situations en 1D

2.3.1 Cas général

Pour les trois problèmes suivants, on définit

$$\begin{aligned} \Omega &= (0, 1), \\ h &= \frac{1}{N}, \text{ avec } N \in \mathbb{N} \text{ et } N \geq 2, \\ \text{Pour } i &= 0, \dots, N, x_i = ih, \\ T_h &= \{K_i, i = 0, \dots, N-1\}, \text{ avec } K_i = [x_i, x_{i+1}], \\ W_h &= \{v_h \in L^2(\Omega); v_h|_K \in \mathbb{P}_r \forall K \in T_h\}, \text{ avec } r = 0 \text{ ou } r = 1. \end{aligned}$$

On a alors le problème approché

$$\left\{ \begin{array}{l} \text{Trouver } a_h \in W_h \text{ tel que } \forall v_h \in W_h \\ \sum_{i=0}^{N-1} \left\{ \int_{x_i}^{x_{i+1}} (Aa_h(x) - f(x))v_h(x)dx - \left[\frac{\nu_{K_i}(x)A_1(x) - |A_1(x)|}{2} (a_h^{\text{int}}(x) - a_h^{\text{ext}}(x)), v_h^{\text{int}}(x) \right]_{\partial K_i} \right\} = 0, \end{array} \right.$$

avec $\nu_{K_i}(x_i) = -1$ et $\nu_{K_i}(x_{i+1}) = 1$.

En prenant $\{\varphi_j, j = 0, \dots, M\}$ une base de W_h , le problème approché s'écrit

$$\left\{ \begin{array}{l} \text{Trouver } a_h \in W_h \text{ tel que } \forall j = 0, \dots, M \\ \sum_{i=0}^{N-1} \left\{ \int_{x_i}^{x_{i+1}} (Aa_h(x) - f(x))\varphi_j(x)dx - \left[\frac{\nu_{K_i}(x)A_1(x) - |A_1(x)|}{2} (a_h^{\text{int}}(x) - a_h^{\text{ext}}(x)), \varphi_j^{\text{int}}(x) \right]_{\partial K_i} \right\} = 0, \end{array} \right.$$

et a_h s'écrit $a_h = \sum_{j=0}^M a_j \varphi_j(x)$.

Pour $r=0$

On prend comme base de W_h l'ensemble

$$\{\varphi_j = \mathbb{1}_{K_j}, j = 0, \dots, N-1\}$$

où

$$\mathbb{1}_{K_j}(x) = \begin{cases} 1 & \text{si } x \in K_j \\ 0 & \text{sinon} \end{cases}.$$

Le problème revient alors à

$$\begin{aligned} & \begin{cases} \text{Trouver } a_h \in W_h \text{ tel que } \forall j = 0, \dots, N-1 \\ \sum_{i=0}^{N-1} \left\{ \int_{x_i}^{x_{i+1}} (Aa_h(x) - f(x)) \varphi_j(x) dx - \left(\frac{-A_1(x_i) - |A_1(x_i)|}{2} (a_{i+\frac{1}{2}} - a_{i-\frac{1}{2}}) \varphi_j^{\text{int}}(x_i) \right) \right. \\ \quad \left. - \left(\frac{A_1(x_{i+1}) - |A_1(x_{i+1})|}{2} (a_{i+\frac{1}{2}} - a_{i+\frac{3}{2}}) \right) \varphi_j^{\text{int}}(x_{i+1}) \right\} = 0 \end{cases} \\ \Leftrightarrow & \begin{cases} \text{Trouver } (a_{i+\frac{1}{2}})_{i=0, \dots, N-1} \subset \mathbb{R} \text{ tel que } \forall i = 0, \dots, N-1 \\ a_{i+\frac{1}{2}} \int_{x_i}^{x_{i+1}} u''(x) dx - \left(\frac{-A_1(x_i) - |A_1(x_i)|}{2} (a_{i+\frac{1}{2}} - a_{i-\frac{1}{2}}) \right) - \left(\frac{A_1(x_{i+1}) - |A_1(x_{i+1})|}{2} (a_{i+\frac{1}{2}} - a_{i+\frac{3}{2}}) \right) \\ \quad = \int_{x_i}^{x_{i+1}} f(x) dx \end{cases} \\ \Leftrightarrow & \text{Mat}_0 \cdot \vec{a} = \vec{f}, \end{aligned}$$

avec

$$\vec{a} = \begin{pmatrix} a_{\frac{1}{2}} \\ a_{\frac{3}{2}} \\ \vdots \\ a_{N-\frac{1}{2}} \end{pmatrix},$$

$$\vec{f} = \begin{pmatrix} \int_{x_0}^{x_1} f(x) dx - \frac{A_1(x_0) + |A_1(x_0)|}{2} a_{-\frac{1}{2}} \\ \int_{x_1}^{x_2} f(x) dx \\ \vdots \\ \int_{x_{N-1}}^{x_N} f(x) dx - \frac{A_1(x_N) - |A_1(x_N)|}{2} a_{N+\frac{1}{2}} \end{pmatrix},$$

où $a_{i+\frac{1}{2}}$ désigne la valeur de la fonction a_h sur l'intervalle K_i pour $i = 0, \dots, N-1$.

$a_{-\frac{1}{2}}$ et $a_{N+\frac{1}{2}}$ sont donnés (si nécessaire, c'est-à-dire si $\frac{A_1(x_0) + |A_1(x_0)|}{2} \neq 0$, $\frac{A_1(x_N) - |A_1(x_N)|}{2} \neq 0$) par les conditions aux limites en 0 et en 1.

$\text{MAT}_0 \in \mathbb{R}^{N \times N}$ est la matrice tridiagonale correspondant à l'équation.

Il suffit alors de résoudre le système linéaire pour obtenir la solution approchée du problème inverse.

Pour calculer les coefficients de la matrice MAT_0 et du vecteur \vec{f} , on approche les intégrales à l'aide de la méthode de Simpson. Pour cela on a besoin des valeurs de u et u' aux points x_i et $\frac{x_i + x_{i+1}}{2}$. On obtient des valeurs approchées en interpolant la fonction u avec des splines cubiques à partir des mesures données. On dérive et on évalue ensuite le polynôme par morceaux obtenu pour trouver les valeurs désirées.

Pour $r=1$

On prend comme base de W_h l'ensemble

$$\{\varphi_{j,1}, \varphi_{j,2}, j = 0, \dots, N-1\},$$

où

$$\begin{aligned}\varphi_{j,1}(x) &= \begin{cases} \frac{x-x_j}{h} \text{ si } x \in K_j, \\ 0 \text{ sinon} \end{cases}, \\ \varphi_{j,2}(x) &= \begin{cases} 1 - \frac{x-x_j}{h} \text{ si } x \in K_j. \\ 0 \text{ sinon} \end{cases}.\end{aligned}$$

On écrit alors

$$a_h = \sum_{j=0}^{N-1} (a_{j,1}\varphi_{j,1}(x) + a_{j,2}\varphi_{j,2}(x)),$$

et le problème revient à

$$\begin{aligned}& \left\{ \begin{array}{l} \text{Trouver } (a_{j,k})_{j=0,\dots,N-1;k=1,2} \subset \mathbb{R} \text{ tel que } \forall i=0, \dots, N-1 \\ \int_{x_i}^{x_{i+1}} Aa_h(x)\varphi_{i,1}(x)dx - \frac{A_1(x_{i+1}) - |A_1(x_{i+1})|}{2}(a_{i,1} - a_{i+1,2}) \Big\} = \int_{x_i}^{x_{i+1}} f(x)\varphi_{i,1}(x)dx \\ \int_{x_i}^{x_{i+1}} Aa_h(x)\varphi_{i,2}(x)dx + \frac{A_1(x_i) + |A_1(x_i)|}{2}(a_{i,2} - a_{i-1,1}) \Big\} = \int_{x_i}^{x_{i+1}} f(x)\varphi_{i,2}(x)dx \end{array} \right. \\ \Leftrightarrow & \left\{ \begin{array}{l} \text{Trouver } (a_{j,k})_{j=0,\dots,N-1;k=1,2} \subset \mathbb{R} \text{ tel que } \forall i=0, \dots, N-1 \\ \left(\int_{x_i}^{x_{i+1}} A\varphi_{i,1}(x)\varphi_{i,1}(x)dx \right) a_{i,1} + \left(\int_{x_i}^{x_{i+1}} A\varphi_{i,2}(x)\varphi_{i,1}(x)dx \right) a_{i,2} \\ \quad - \frac{A_1(x_{i+1}) - |A_1(x_{i+1})|}{2}(a_{i,1} - a_{i+1,2}) \Big\} = \int_{x_i}^{x_{i+1}} f(x)\varphi_{i,1}(x)dx \\ \left(\int_{x_i}^{x_{i+1}} A\varphi_{i,1}(x)\varphi_{i,2}(x)dx \right) a_{i,1} + \left(\int_{x_i}^{x_{i+1}} A\varphi_{i,2}(x)\varphi_{i,2}(x)dx \right) a_{i,2} \\ \quad + \frac{A_1(x_i) + |A_1(x_i)|}{2}(a_{i,2} - a_{i-1,1}) \Big\} = \int_{x_i}^{x_{i+1}} f(x)\varphi_{i,2}(x)dx \end{array} \right. \\ \Leftrightarrow & \text{MAT}_1 \cdot \vec{a} = \vec{f},\end{aligned}$$

avec

$$\vec{a} = \begin{pmatrix} a_{0,1} \\ a_{0,2} \\ a_{1,1} \\ a_{1,2} \\ \vdots \\ a_{N-2,1} \\ a_{N-2,2} \\ a_{N-1,1} \\ a_{N-2,2} \end{pmatrix}, \quad \vec{f} = \begin{pmatrix} \int_{x_0}^{x_1} f(x)\varphi_{i,1}(x)dx \\ \int_{x_0}^{x_1} f(x)\varphi_{i,2}(x)dx + \frac{A_1(x_0) + |A_1(x_0)|}{2}a_{-1,1} \\ \int_{x_1}^{x_2} f(x)\varphi_{i,1}(x)dx \\ \int_{x_1}^{x_2} f(x)\varphi_{i,2}(x)dx \\ \vdots \\ \int_{x_{N-2}}^{x_{N-1}} f(x)\varphi_{i,1}(x)dx \\ \int_{x_{N-2}}^{x_{N-1}} f(x)\varphi_{i,2}(x)dx \\ \int_{x_{N-1}}^{x_N} f(x)\varphi_{i,1}(x)dx - \frac{A_1(x_N) - |A_1(x_N)|}{2}a_{N,2} \\ \int_{x_{N-1}}^{x_N} f(x)\varphi_{i,2}(x)dx \end{pmatrix},$$

où $a_{-1,1}$ et $a_{N,2}$ sont donnés (si nécessaire, c'est-à-dire si $\frac{A_1(x_0) + |A_1(x_0)|}{2} \neq 0$, $\frac{A_1(x_N) - |A_1(x_N)|}{2} \neq 0$) par les conditions aux limites en 0 et en 1.

$\text{MAT}_1 \in \mathbb{R}^{2N \times 2N}$ est la matrice correspondant à l'équation, et à une structure du type

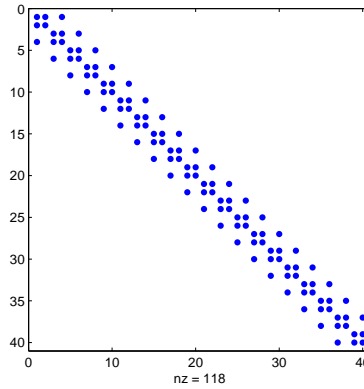


Figure 1. Structure de la matrice MAT pour $N = 20$

Il ne reste alors plus qu'à résoudre le système linéaire $\text{MAT}_1 \cdot \vec{a} = \vec{f}$ pour obtenir la solution approchée du problème inverse.

Pour calculer les coefficients de la matrice MAT_1 et du vecteur \vec{f} , on approche les intégrales à l'aide de la méthode de Simpson. Pour cela on a besoin des valeurs de u , u' et u'' aux points x_i et $\frac{x_i + x_{i+1}}{2}$. On obtient des valeurs approchées en interpolant la fonction u avec des splines cubiques à partir des mesures données. On dérive et on évalue ensuite le polynôme par morceaux obtenu pour trouver les valeurs désirées.

Nous allons maintenant étudier et résoudre numériquement trois exemples de problèmes inverses, et tenter de résoudre les problèmes qui surviennent.

Les programmes ont été écrit en MATLAB.

2.3.2 Problème 1

Posons

$$\begin{aligned} u(x) &= x^2 - x, \\ f(x) &= 1. \end{aligned}$$

Alors le problème inverse défini par (1) se résout facilement, et la solution exacte est donnée par $a(x) = \frac{1}{2}$. Cela nous permettra de calculer l'erreur de la méthode numérique.

On a

$$C(x) = 2A_0(x) - A_1'(x) = u''(x) = 2, \forall x \in (0, 1),$$

donc

$$\exists \alpha > 0; C(x) \geq \alpha, \forall x \in \Omega.$$

Donc d'après la partie 2.2.3, le problème approché admet une unique solution. On applique donc la méthode d'éléments finis discontinus \mathbb{P}_0 décrites ci-dessus pour la résolution numérique de ce problème avec $N = 2, \dots, 50$ et M points de mesure $(X_{\text{mes}_i}, Y_{\text{mes}_i})$ avec pour $i = 0, \dots, M - 1$,

$$\begin{aligned} X_{\text{mes}_i} &= \frac{i}{M-1}, \\ Y_{\text{mes}_i} &= u(X_{\text{mes}_i}). \end{aligned}$$

On évaluera ensuite l'erreur par rapport à la solution exacte en prenant

$$\text{err} = \|a - a_h\|_{L^2}$$

où a_h est la solution approchée.

Dans la suite, on utilisera la norme L^2 pour tous les calculs d'erreur.

On obtient des résultat du type

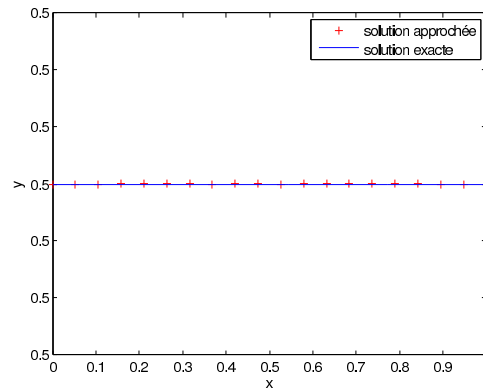


Figure 2. Solution approchée pour $N = 20$ et $M = 4$ avec des éléments finis \mathbb{P}_0

avec erreur = $1.6799e - 016$.

En faisant varier M et N , on obtient les erreurs (en norme L^2) suivantes pour différentes valeurs de M :

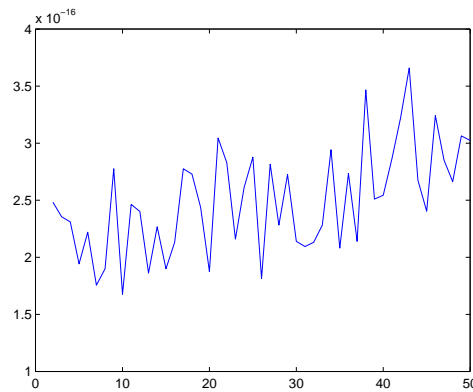


Figure 3. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 4$ avec des éléments finis \mathbb{P}_0

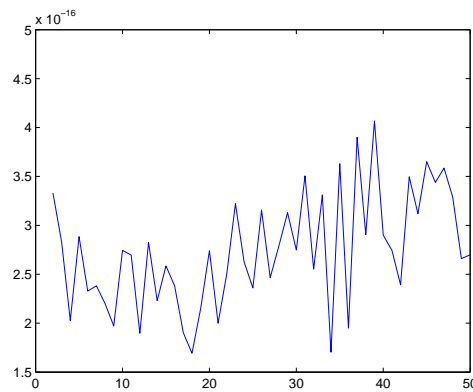


Figure 4. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 7$ avec des éléments finis \mathbb{P}_0

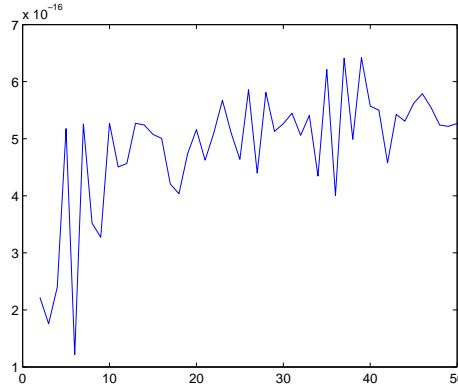


Figure 5. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 10$ avec des éléments finis P_0

On voit que l'erreur est négligeable pour toutes les valeurs de N et de M observées.

En fait, pour ce problème, la méthode d'éléments finis discontinus donne la solution exacte. En effet :

- la solution exacte a appartient à l'espace d'approximation W_h
- l'interpolation de u et f à partir des points de mesure est faite par splines cubiques, ce qui donne un résultat exact car u et f sont des polynômes de degré ≤ 2 .

De plus, les méthodes numériques utilisées pour l'intégration (méthode de Simpson) et la dérivation (dérivation formelle des polynômes par morceaux) sont exacte pour les fonctions intervenant dans ce problème.

Donc l'erreur obtenue est due à la précision de la machine.

2.3.3 Problème 2

Posons

$$\begin{aligned} u(x) &= \frac{x^4}{6} - \frac{x^3}{6} + x, \\ f(x) &= -8x^3 + \frac{13x^2}{2} - x - 3. \end{aligned}$$

La solution exacte du problème inverse défini par (1) est donnée par $a(x) = 3x - 1$.

On a

$$C(x) = 2A_0(x) - A_1'(x) = u''(x) = 2x^2 - x, \forall x \in (0, 1),$$

donc

$$\exists x \in \Omega; C(x) < 0.$$

Donc les conditions des théorèmes de convergence ne sont pas vérifiées.

On introduit alors la fonction

$$\psi(x) = u'(x)e^{-\alpha x}, \forall x \in \Omega, \text{ avec } \alpha > 0,$$

et le problème modifié

$$\begin{aligned} &\psi Aa = \psi f \text{ dans } \Omega \\ \Leftrightarrow &\psi(x)(A_1(x)a'(x) + A_0(x)a(x)) = \psi(x)f(x) \forall x \in \Omega \\ \Leftrightarrow &\overline{A}_1(x)a'(x) + \overline{A}_0(x)a(x) = \overline{f}(x) \forall x \in \Omega, \end{aligned}$$

avec

$$\begin{aligned} \overline{A}_1(x) &= \psi(x)A_1(x) = \psi(x)u'(x), \\ \overline{A}_0(x) &= \psi(x)A_0(x) = \psi(x)u''(x), \\ \overline{f}(x) &= \psi(x)f(x). \end{aligned}$$

On a alors

$$\begin{aligned}\bar{C}(x) &= 2\bar{A}_0(x) - \bar{A}_1'(x) \\ &= u''(x)\psi(x) - u'(x)\psi'(x) \\ &= \alpha(u'(x))^2 e^{-\alpha x} \quad \forall x \in \Omega,\end{aligned}$$

On remarque que si

$$u'(x) \neq 0, \forall x \in \Omega,$$

alors

$$\exists \varepsilon > 0; \bar{C}(x) \geq \varepsilon, \forall x \in \Omega,$$

et donc d'après la partie 2.2.3, le problème approché correspondant au problème modifié admet une unique solution.

Remarque 16. La condition $\exists \alpha > 0; C(x) = u''(x) \geq \alpha, \forall x \in \Omega$ du cas général a été remplacée par la condition $u'(x) \neq 0, \forall x \in \Omega$ pour le problème modifié.

On applique les méthodes d'éléments finis discontinus \mathbb{P}_0 et \mathbb{P}_1 décrites en 2.3.1 pour la résolution numérique de ce problème avec $N = 2, \dots, 50$ et M points de mesure (X_{mes_i}, Y_{mes_i}) avec pour $i = 0, \dots, M - 1$,

$$\begin{aligned}X_{mes_i} &= \frac{i}{M-1}, \\ Y_{mes_i} &= u(X_{mes_i}).\end{aligned}$$

On obtient des résultat du type

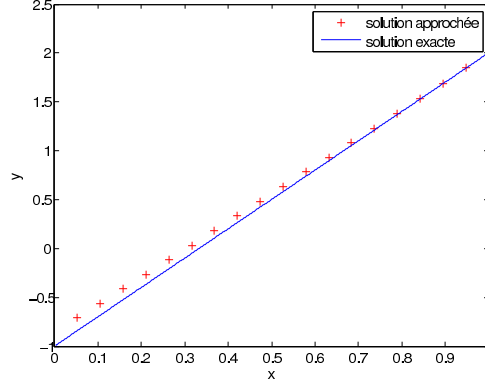


Figure 6. Solution approchée pour $N = 20$ et $M = 4$ avec des éléments finis \mathbb{P}_0

avec erreur = 0.0732.

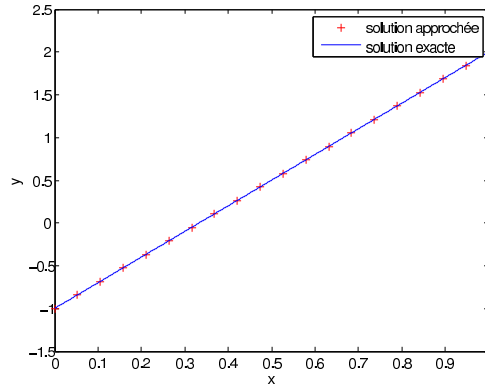


Figure 7. Solution approchée pour $N = 20$ et $M = 4$ avec des éléments finis \mathbb{P}_1

avec erreur = $3.9867e - 016$.

En faisant varier M et N , on obtient les erreurs (en norme L^2) suivantes pour différentes valeurs de M :

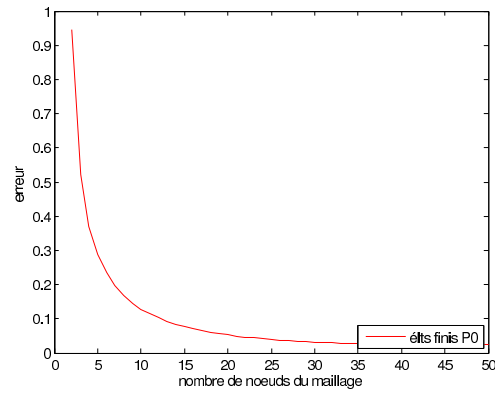


Figure 8. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 4$ avec des éléments finis \mathbb{P}_0

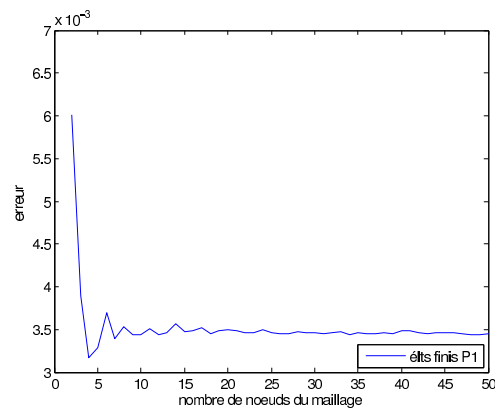


Figure 9. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 7$ avec des éléments finis \mathbb{P}_0

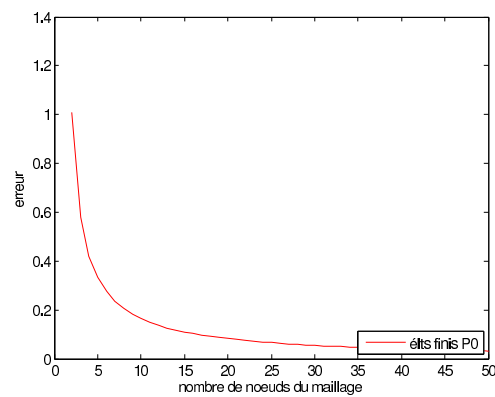


Figure 10. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 10$ avec des éléments finis \mathbb{P}_0

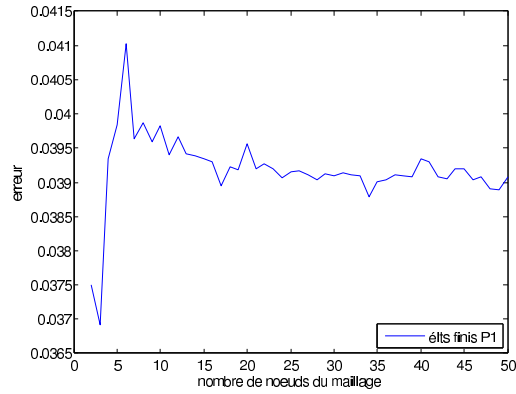


Figure 11. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 4$ avec des éléments finis \mathbb{P}_1

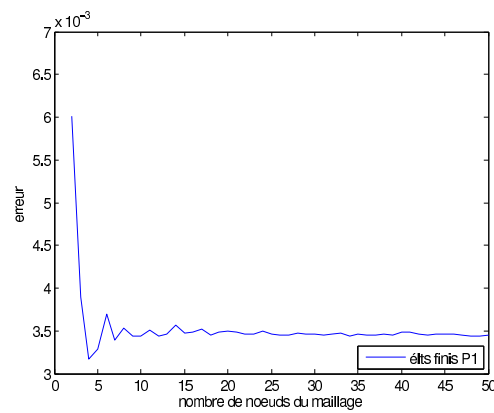


Figure 12. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 7$ avec des éléments finis \mathbb{P}_1



Figure 13. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 10$ avec des éléments finis \mathbb{P}_1

La solution approchée est plus proche de la solution exacte (en norme L^2) avec des éléments finis \mathbb{P}_1 qu'avec des éléments finis \mathbb{P}_0 .

En revanche l'erreur ne se comporte pas comme prévu par le théorème 15.

Cela est dû aux erreurs d'interpolation de u et f . En effet, en prenant les vraies valeurs de u et f au lieu d'une interpolation réalisée à partir des valeurs mesurées, on obtient la solution exacte du problème (aux erreurs d'arrondi près) pour les éléments finis \mathbb{P}_1 . Cela est dû au fait qu'avec

l'interpolation, on modifie le problème, et donc la solution qu'on n'obtient ne converge pas forcément vers la solution exacte.

Remarque 17. Si on applique la méthode au problème non modifié, on obtient des résultats similaires. En effet, les conditions pour assurer la convergence ne sont alors pas vérifiées, mais ces conditions sont suffisantes, et pas nécessaires.

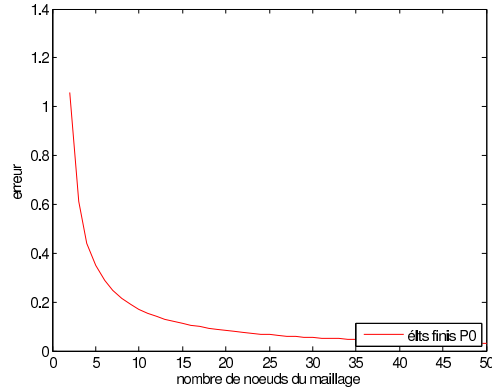


Figure 14. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 7$ avec des éléments finis \mathbb{P}_0 - problème non modifié

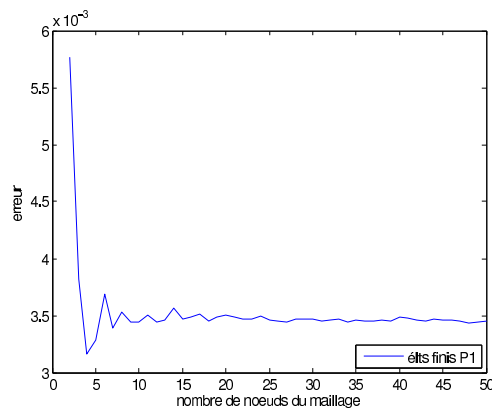


Figure 15. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 7$ avec des éléments finis \mathbb{P}_1 - problème non modifié

2.3.4 Problème 3

Posons

$$u(x) = -\frac{1}{\sigma\sqrt{2\pi}}e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)},$$

$$f(x) = (2x-1)\left(\frac{(x-\mu)}{\sigma^3\sqrt{2\pi}}e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}\right) + x(x-1)\left(-\frac{(x^2-2x\mu+\mu^2-\sigma^2)}{\sigma^5\sqrt{2\pi}}e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}\right),$$

avec $\mu = \sigma = \frac{1}{2}$.

On a alors

$$u'(x) = \frac{x-\mu}{\sigma^3\sqrt{2\pi}}e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)},$$

$$u''(x) = -\frac{(x^2-2x\mu+\mu^2-\sigma^2)}{\sigma^5\sqrt{2\pi}}e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}.$$

La solution exacte du problème inverse défini par (1) est donnée par

$$a(x) = x(x - 1).$$

On a

$$C(x) = 2A_0(x) - A_1'(x) = u''(x) = -\frac{(x^2 - 2x\mu + \mu^2 - \sigma^2)}{\sigma^5\sqrt{2\pi}}e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}, \forall x \in (0, 1),$$

donc

$$\exists x \in \Omega; C(x) < 0.$$

Donc les conditions des théorèmes de convergence ne sont pas vérifiées.

On introduit alors les fonctions ψ et b telles que

$$\psi(x) = e^{u(x)}, \forall x \in \Omega,$$

$$b(x) = \psi^{-1}(x)a(x), \forall x \in \Omega.$$

Alors on a

$$\begin{aligned} a &= b\psi, \\ a' &= b'\psi + b\psi', \end{aligned}$$

et on introduit le problème modifié

$$\begin{aligned} \bar{A}b &= f \text{ dans } \Omega \\ \Leftrightarrow (\psi(x)u'(x))b'(x) + (\psi'(x)u'(x) + \psi(x)u''(x))b(x) &= f(x) \forall x \in \Omega \\ \Leftrightarrow \bar{A}_1(x)b'(x) + \bar{A}_0(x)b(x) &= f(x) \forall x \in \Omega, \end{aligned}$$

avec

$$\begin{aligned} \bar{A}_1(x) &= u'(x)\psi(x), \\ \bar{A}_0(x) &= u'(x)\psi'(x) + u''(x)\psi(x), \end{aligned}$$

donc on a

$$\begin{aligned} \bar{B}(x) &= \nu(x)\bar{A}_1(x), \\ \bar{M}(x) &= |\bar{A}_1(x)|, \\ \bar{C}(x) &= 2\bar{A}_0(x) - \bar{A}_1'(x) \\ &= u'(x)\psi'(x) + u''(x)\psi(x) \\ &= e^{u(x)}(u''(x) + (u'(x))^2). \end{aligned}$$

On remarque que si

$$\inf_{x \in \Omega} (u''(x) + (u'(x))^2) > 0,$$

alors

$$\exists \varepsilon > 0; \bar{C}(x) \geq \varepsilon, \forall x \in \Omega,$$

et donc d'après la partie 2.2.3, le problème approché correspondant au problème modifié admet une unique solution.

On obtient alors une solution approchée b_h du problème modifié.

Pour avoir une solution approchée a_h du problème de départ, il suffit alors de multiplier b_h par ψ .

Remarque 18. La condition $\exists \alpha > 0; C(x) = u''(x) \geq \alpha, \forall x \in \Omega$ du cas général a été remplacée par la condition $\inf_{x \in \Omega} (u''(x) + (u'(x))^2) > 0$ pour le problème modifié.

On applique les méthodes d'éléments finis discontinus \mathbb{P}_0 et \mathbb{P}_1 décrites en 2.3.1 pour la résolution numérique de ce problème avec $N = 2, \dots, 50$ et M points de mesure (X_{mes_i}, Y_{mes_i}) avec pour $i = 0, \dots, M - 1$,

$$\begin{aligned} X_{mes_i} &= \frac{i}{M-1}, \\ Y_{mes_i} &= u(X_{mes_i}). \end{aligned}$$

On obtient des résultat du type

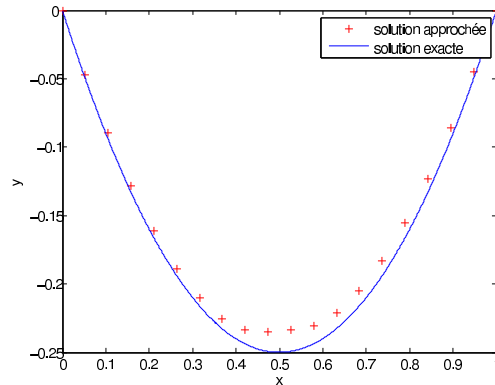


Figure 16. Solution approchée pour $N = 20$ et $M = 4$ avec des éléments finis \mathbb{P}_0

avec erreur = 0.0092.

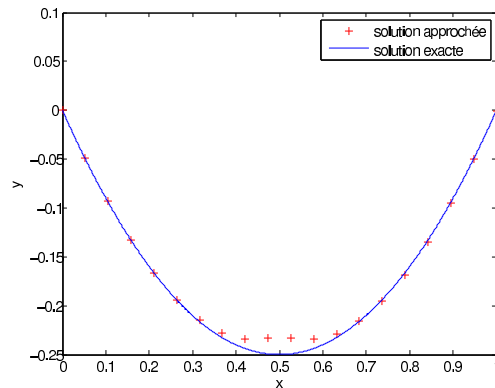


Figure 17. Solution approchée pour $N = 20$ et $M = 4$ avec des éléments finis \mathbb{P}_1

avec erreur = 0.0064.

En faisant varier M et N , on obtient les erreurs (en norme L^2) suivantes pour différentes valeurs de M :

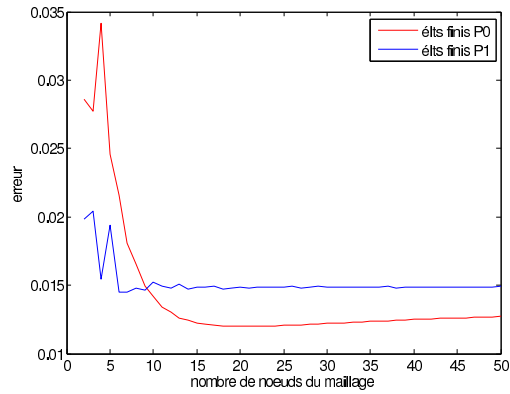


Figure 18. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 4$

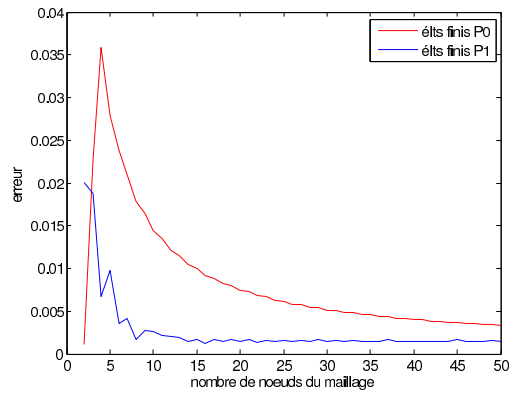


Figure 19. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 7$

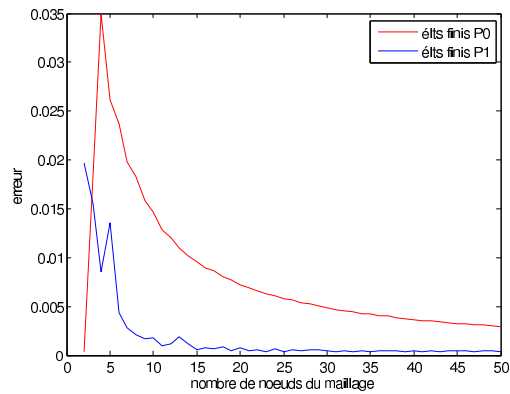


Figure 20. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 10$

La solution approchée est plus proche de la solution exacte (en norme L^2) avec des éléments finis \mathbb{P}_1 qu'avec des éléments finis \mathbb{P}_0 .

De même que pour le problème 2, l'erreur ne se comporte pas comme prévu par le théorème 15. Cela est à nouveau dû aux erreurs d'interpolation de u et f .

Remarque 19. Si on applique la méthode au problème non modifié, on obtient des résultats similaires. En effet, les conditions pour assurer la convergence ne sont alors pas vérifiées, mais

ces conditions sont suffisantes, et pas nécessaires.

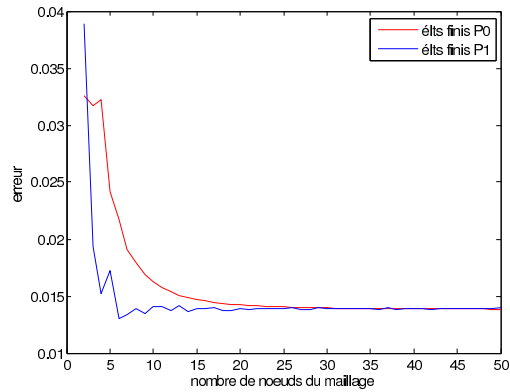


Figure 21. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 4$ - pb non modifié

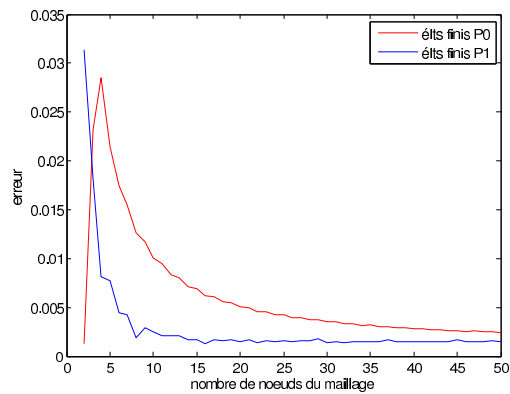


Figure 22. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 7$ - pb non modifié

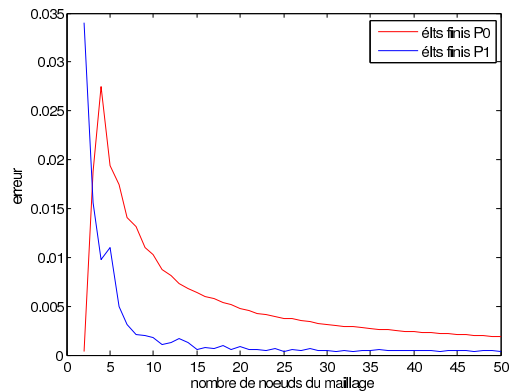


Figure 23. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 10$ - pb non modifié

2.4 Données avec bruit

En pratique, les mesures de u et f ne sont pas exactes.

En effet, on observe généralement un certain bruit. Pour prendre cela en compte, on a ajouté un bruit aléatoire inférieur à 5% de la valeur réelle aux mesures de u et f .

En prenant en compte le bruit, on ne peut plus nécessairement garantir la convergence de la méthode. En effet, les conditions sur u' et u'' pour les problèmes modifiés ou non ne sont plus forcément vérifiées.

2.4.1 Tentatives d'application des mêmes méthodes

Dans un premier temps, on observe le comportement de la méthode numérique en présence de bruit en regardant l'allure et l'erreur de la solution approchée lorsqu'on fait varier les paramètres N et M avec une fonction u 'bruitée'.

2.4.2 Problème 1

On obtient des résultat du type

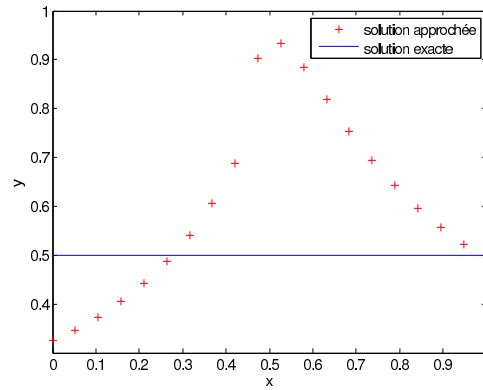


Figure 24. Solution approchée pour $N = 20$ et $M = 4$ avec des éléments finis \mathbb{P}_0

avec erreur = 0.2115.

En faisant varier M et N , on obtient les erreurs (en norme L^2) suivantes pour différentes valeurs de M :

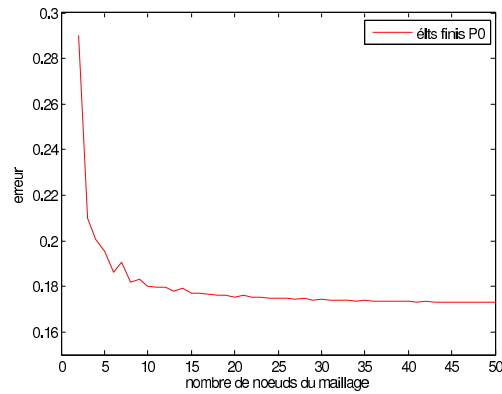


Figure 25. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 4$ avec des éléments finis \mathbb{P}_0 - avec bruit

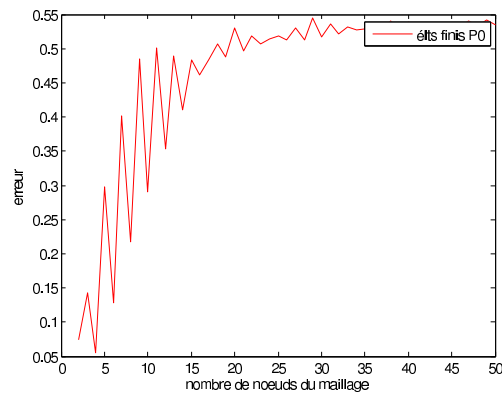


Figure 26. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 7$ avec des éléments finis \mathbb{P}_0 - avec bruit

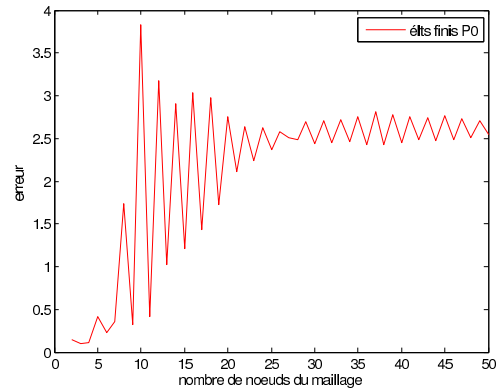


Figure 27. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 10$ avec des éléments finis \mathbb{P}_0 - avec bruit

2.4.3 Problème 2

On obtient des résultat du type

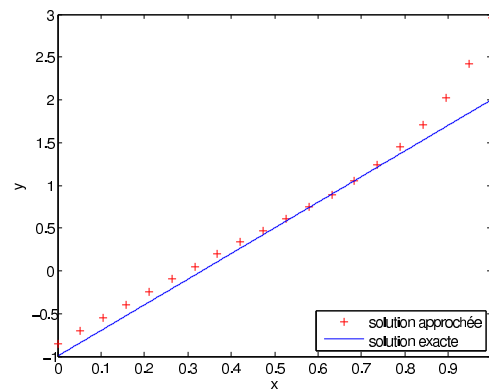


Figure 28. Solution approchée pour $N = 20$ et $M = 4$ avec éltis finis \mathbb{P}_0 - avec bruit

avec erreur = 0.2374.

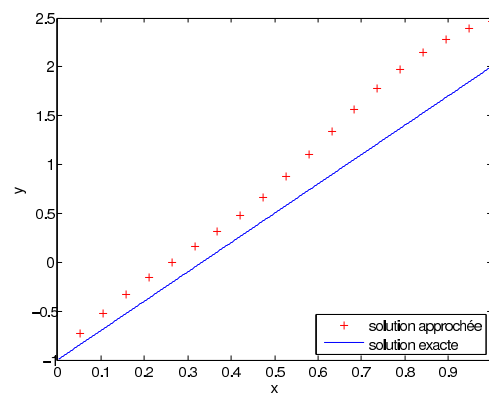


Figure 29. Solution approchée pour $N = 20$ et $M = 4$ avec éltis finis \mathbb{P}_1 - avec bruit

avec erreur = 0.3898.

En faisant varier M et N , on obtient les erreurs (en norme L^2) suivantes pour différentes valeurs de M :

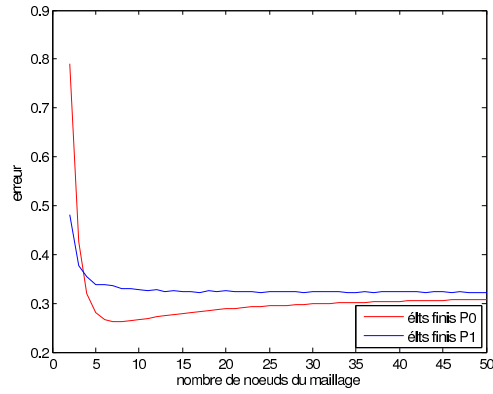


Figure 30. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 4$ - avec bruit

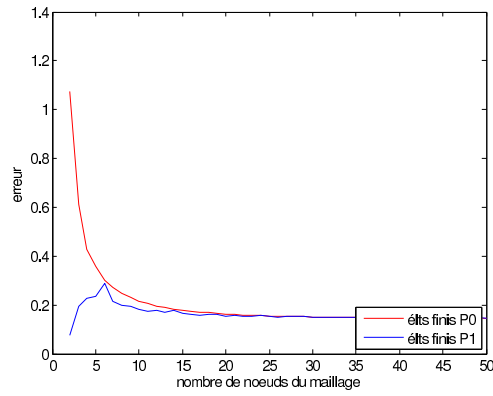


Figure 31. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 7$ - avec bruit

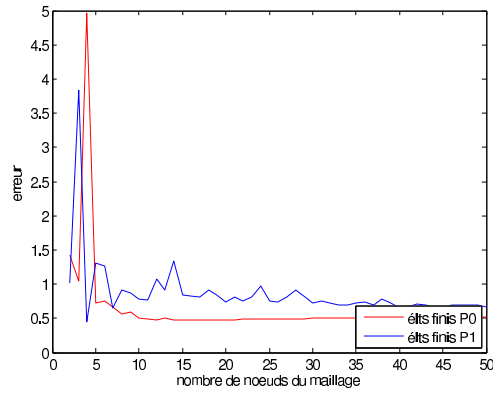


Figure 32. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 10$ - avec bruit

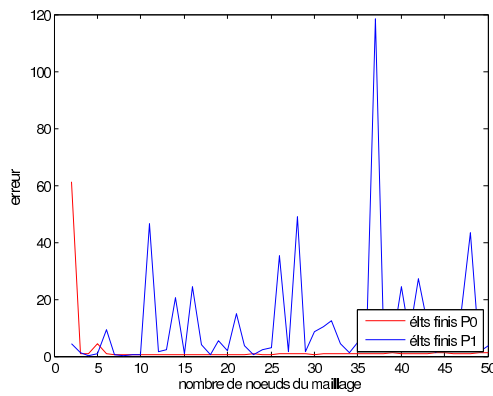


Figure 33. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 15$ - avec bruit

2.4.4 Problème 3

On obtient des résultat du type

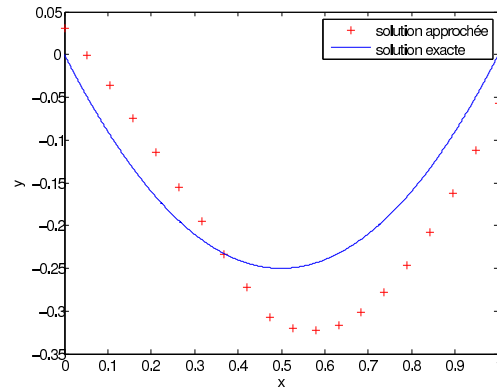


Figure 34. Solution approchée pour $N = 20$ et $M = 4$ avec éltis finis \mathbb{P}_0

avec erreur = 0.0621.

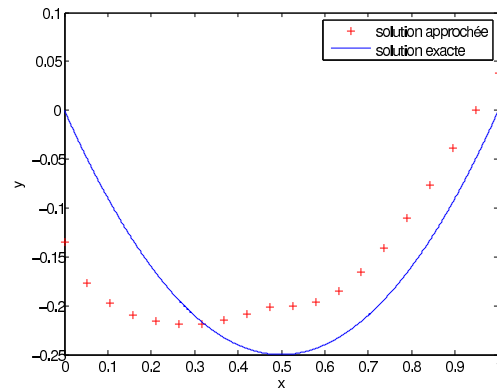


Figure 35. Solution approchée pour $N = 20$ et $M = 4$ avec éltis finis \mathbb{P}_1

avec erreur = 0.0622.

En faisant varier M et N , on obtient les erreurs (en norme L^2) suivantes pour différentes valeurs de M :

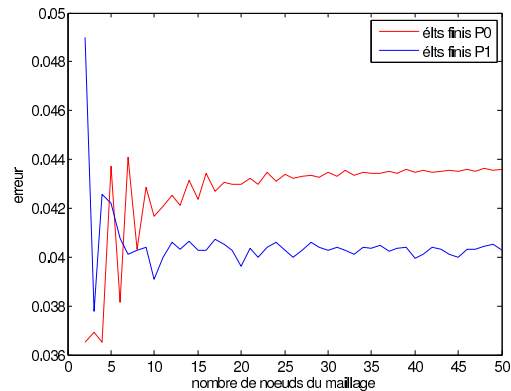


Figure 36. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 4$ - avec bruit

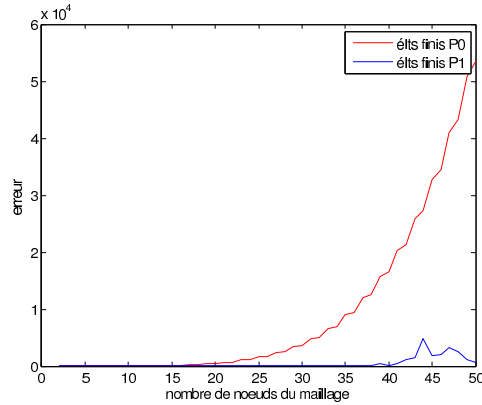


Figure 37. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 7$ - avec bruit

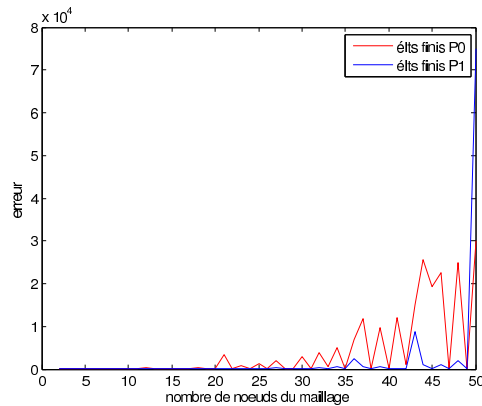


Figure 38. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 10$ - avec bruit

On remarque que la méthode ne converge pas, ou s'éloigne énormément de la solution, lorsque le nombre de points de mesures augmente. En effet, u n'est alors plus assez régulière pour que la méthode converge.

2.4.5 Idée de solution : Interpolation par splines cubiques avec contraintes de convexité

Idée

On essaye de résoudre ce problème dans le cas 1

$$\begin{aligned} u(x) &= x^2 - x, \\ f(x) &= 1. \end{aligned}$$

Les problèmes observés plus haut proviennent de perturbations dans les données, qui entraînent une interpolation faussée des fonctions u et f . En effet, de faibles variations des valeurs de la fonction obtenue par interpolation u_{interp} entraînent une erreur conséquente sur la dérivée et la dérivée seconde de cette fonction, et la condition

$$\exists \alpha > 0; u''(x) > \alpha \quad \forall x \in \Omega$$

n'est alors plus forcément satisfaite, et donc les conditions des théorèmes assurant l'existence d'une solution du problème approché ne sont plus.

Une solution consiste à réaliser une interpolation avec des contraintes sur le signe de la dérivée seconde. On s'intéresse donc à une méthode d'interpolation à l'aide de splines avec des contraintes de convexité.

Néanmoins, cela suppose qu'on connaît à priori la convexité de la fonction mesurée u , ce qui n'est pas forcément le cas en pratique.

Généralités à propos des splines

Définition 20. *Étant donnés $n + 2$ points t_i dans un intervalle $[a, b]$ avec $a = t_0 < t_1 < t_2 < \dots < t_n < t_{n+1} = b$, la fonction $s: [a, b] \rightarrow \mathbb{R}$ est appelée spline cubique si*

- i. $s \in C^2([a, b])$
- ii. la restriction de s sur chaque intervalle $[t_{i-1}, t_i]$, $i = 1, 2, \dots, n + 1$ est un polynôme de degré 3 ou moins.

Définition 21. *Étant donnés $n + 2$ points t_i dans un intervalle $[a, b]$ avec $a = t_0 < t_1 < t_2 < \dots < t_n < t_{n+1} = b$, on appelle fonctions B-splines normalisées les fonctions données par*

$$N_j(x) = (t_{j+4} - t_j) \Delta_t^4(t_j, t_{j+1}, \dots, t_{j+4}) g(t; x),$$

où

$$g(t; x) = (t - x)_+^3 = \begin{cases} (t - x)^3 & \text{si } t \geq x \\ 0 & \text{si } t < x \end{cases},$$

et $\Delta_t^4(z_j, z_{j+1}, \dots, z_{j+4}) f(t)$ désigne la différence divisée d'ordre 4 de la fonction $f(t)$ aux points $z_j, z_{j+1}, \dots, z_{j+4}$.

Remarque 22. Pour évaluer les fonctions B-splines normalisées, nous utilisons pour les besoins de ce projet l'algorithme donné par De Boor [1].

On introduit des noeuds

$$t_{-3} \leq t_{-2} \leq t_{-1} \leq t_0 = a \text{ et } t_{n+4} \geq t_{n+3} \geq t_{n+2} \geq t_{n+1} = b,$$

afin d'avoir une base de B-splines.

Ces noeuds sont arbitraires, et on choisit pour ce projet

$$\begin{aligned} t_{-3} &= t_{-2} = t_{-1} = t_0 = a, \\ t_{n+4} &= t_{n+3} = t_{n+2} = t_{n+1} = b. \end{aligned}$$

On a alors une unique représentation de la fonction s , spline cubique avec pour noeuds $t_0, t_1, t_2, \dots, t_n, t_{n+1}$, sous la forme

$$s(x) = \sum_{j=-3}^n c_j N_j(x), a \leq x \leq b, \quad (8)$$

avec $c_j \in \mathbb{R}$, $j = -3, \dots, n$.

Description de la méthode d'interpolation

L'algorithme décrit dans ce paragraphe est celui présenté par Dierckx [3].

Étant donnés des points (x_i, y_i) et des poids w_i , $i = 1, \dots, m$ ainsi que des noeuds $a = t_0 < t_1 < t_2 < \dots < t_n < t_{n+1} = b$, on cherche à obtenir une spline cubique s avec pour noeuds les t_j telle que

- i. $\sum_{i=1}^m w_i (y_i - s(x_i))^2$ est minimale
- ii. à chaque noeud t_j , $j = 0, 1, \dots, n + 1$ $s(x)$ est localement convexe, concave ou sans contrainte,

où on choisit les contraintes sur le signe de s'' aux noeuds (car s'' est un polynôme de degré ≤ 1).

Dans les applications, on prendra $w_i = 1$ pour $i = 1, \dots, m$.

Avec la représentation (8), on trouve

$$s''(t_l) = \sum_{j=-3}^n d_{j,l} c_j, \tag{9}$$

avec

$$\begin{aligned} d_{j,l} &= 0 \text{ si } j \leq l-4 \text{ ou } j \geq l, \\ d_{l-3,l} &= \frac{6}{(t_{l+1}-t_{l-2})(t_{l+1}-t_{l-1})}, \\ d_{l-1,l} &= \frac{6}{(t_{l+2}-t_{l-1})(t_{l+1}-t_{l-1})}, \\ d_{l-2,l} &= -(d_{l-3,l} + d_{l-1,l}). \end{aligned}$$

Donc pour trouver s vérifiant les conditions énoncées plus haut, il suffit de calculer les coefficient $c_j, j = -3, -2, \dots, n$ en résolvant le problème de programmation quadratique

$$\begin{aligned} &\text{Minimiser } \sum_{i=1}^m \left(y_i - \sum_{j=-3}^n c_j N_j(x_i) \right)^2 \\ &\text{tel que } e_l \sum_{j=-3}^n d_{j,l} c_j \leq 0, l = 0, 1, \dots, n+1, \end{aligned} \tag{10}$$

où $e_l = -1, 1$ ou 0 selon les contraintes de convexité que l'on veut imposer à s .

Il est possible de choisir les noeuds t_j automatiquement de façon à obtenir une erreur d'interpolation $\sum_{i=1}^m w_i (y_i - s(x_i))^2$ aussi faible qu'on le désire en utilisant l'algorithme décrit par Dierckx [3]. Il consiste à commencer avec $n = 0$ et à ajouter d'autres noeuds en fonction des contraintes imposées sur le signe de s'' jusqu'à ce qu'on arrive à l'erreur désirée.

Application au Problème 1

En appliquant cette méthode d'interpolation, en imposant que l'erreur d'interpolation soit inférieure à 0.01, au lieu de l'interpolation par splines cubiques classique, on obtient les résultats suivant :

Remarque 23. Pour chaque valeur de M , on s'intéresse à l'erreur de la solution approchée en utilisant l'interpolation par splines cubiques classique et la méthode d'interpolation par splines avec contraintes de convexité en utilisant les mêmes mesures bruitées de u et f . On peut donc comparer directement les courbes.

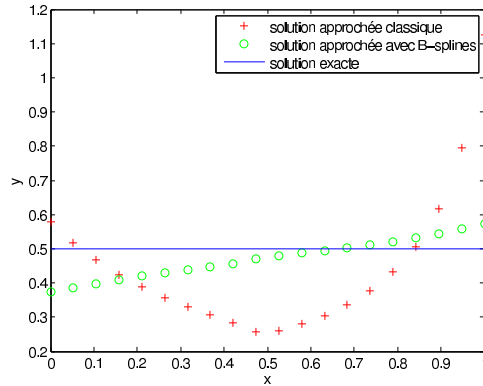


Figure 39. Solution approchée pour $N = 20$ et $M = 4$ avec élts finis \mathbb{P}_0

On a ici erreur = 0.1923 pour la solution approchée classique, et erreur = 0.0609 pour la solution approchée obtenue avec les B-splines.

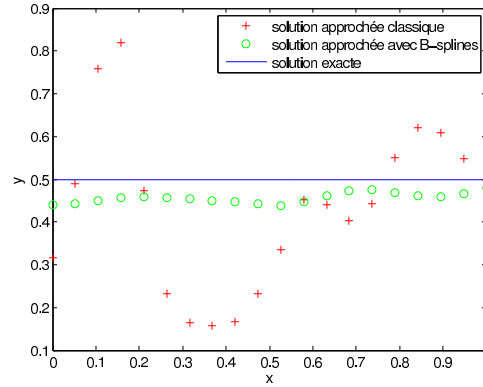


Figure 40. Solution approchée pour $N = 20$ et $M = 9$ avec élts finis \mathbb{P}_0

On a ici erreur = 0.1987 pour la solution approchée classique, et erreur = 0.0452 pour la solution approchée obtenue avec les B-splines.

En faisant varier M et N , on obtient les erreurs (en norme L^2) suivantes pour différentes valeurs de M :

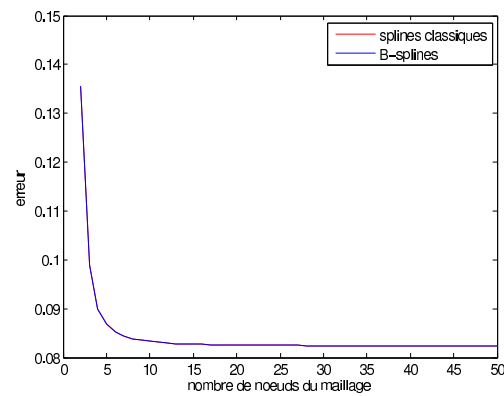


Figure 41. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 4$ avec élts finis \mathbb{P}_0 - avec bruit

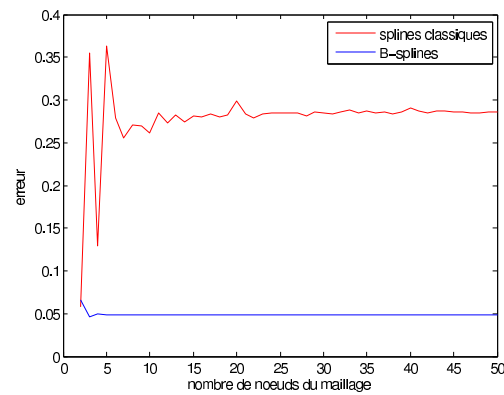


Figure 42. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 7$ avec élts finis \mathbb{P}_0 - avec bruit

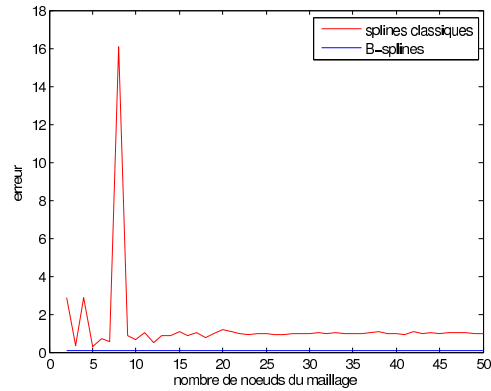


Figure 43. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 10$ avec éls finis \mathbb{P}_0 - avec bruit

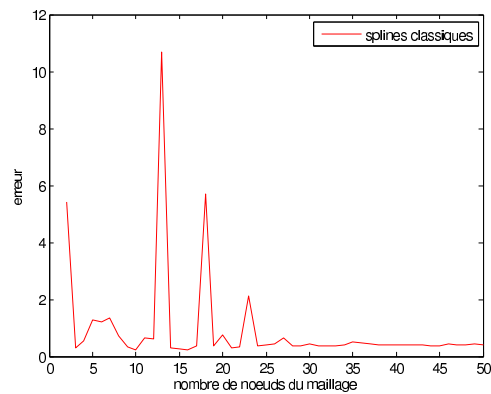


Figure 44. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 15$ avec éls finis \mathbb{P}_0 - interpolation par splines classique - avec bruit

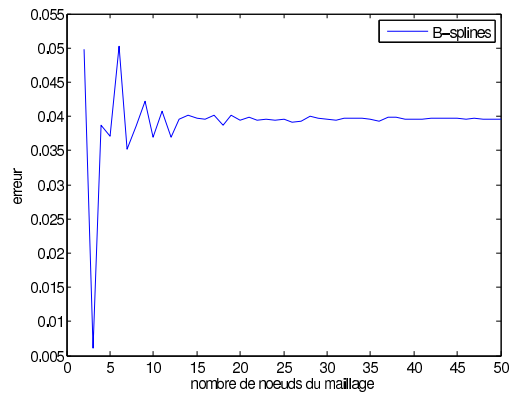


Figure 45. Erreur en norme L^2 pour $N = 2, \dots, 50$ et $M = 15$ avec éls finis \mathbb{P}_0 - interpolation avec contraintes de convexité - avec bruit

On voit que l'erreur est nettement inférieure avec la méthode d'interpolation par splines avec contraintes de convexité qu'avec l'interpolation par splines classique, et on n'a plus de problèmes de convergence.

Cependant l'inconvénient de cette méthode est qu'il est nécessaire de connaître a priori le signe de la fonction u'' .

On pourrait utiliser une méthode similaire pour les problèmes 2 et 3, mais il faudrait pour cela

faire une étude préliminaire des conditions à imposer sur u pour assurer l'existence d'une solution au problème approché.

3 Méthode OLS

Nous allons maintenant nous intéresser à une autre approche, basée sur la minimisation de fonctionnelles, pour résoudre le même type de problèmes inverses : la méthode Output Least Squares (OLS).

3.1 Formulation du problème

Dans une équation aux dérivées partielles de la forme (1)

$$\begin{cases} -\nabla \cdot (a \nabla u) = f \text{ sur } \Omega \\ + \text{Conditions aux Limites sur } \partial\Omega \end{cases}$$

on suppose que f est donnée, que a ne peut pas être mesuré directement, mais que l'on peut obtenir des valeurs de u en certains points de mesure de manière à obtenir une approximation z . On suppose de plus qu'on peut simuler les données observables (c'est-à-dire u) pour toute estimation raisonnable \bar{a} des paramètres a . On obtient alors une estimation $u(\bar{a})$ des données u . Cette estimation peut être obtenue par la méthode des éléments finis, et on a alors $u_h(a_h)$, où u_h et a_h sont des polynômes par morceaux.

On peut alors approcher le coefficient a par la fonction a_h telle que les données simulées $u_h(a_h)$ soient aussi proches que possible des données mesurées z , c'est-à-dire en résolvant le problème approché

$$\text{Trouver } a_h \in K_h \text{ tel que } J(a_h) = \inf_{\bar{b} \in K_h} J(\bar{b}), \quad (11)$$

où

$$\begin{aligned} J(\bar{b}) &= \|u_h(\bar{b}) - z\|_{L^2(\Omega)}^2, \\ K_h &= \{\bar{b} \in T_h; 0 < \alpha \leq \bar{b} \leq \beta\}, \quad \alpha, \beta \in \mathbb{R}_{>0}, \end{aligned}$$

et $u_h(\bar{b}) \in S_h$ est donné par

$$-\int_{\Omega} \bar{b} \nabla u_h(\bar{b}) \cdot \nabla v_h \, dx = \int_{\Omega} f v_h \, dx, \quad \forall v_h \in \overset{\circ}{S}_h = \{w \in S_h; w(x) = 0 \forall x \in \partial\Omega\},$$

avec $T_h \subset L^2(\Omega)$ et $S_h \subset H^1(\Omega)$ sous-espaces de dimension finie.

u_h est polynomiale par morceaux, donc le gradient de $u_h(\bar{b})$ est bien défini (par morceaux).

On définit l'espace K_h ainsi afin de pouvoir calculer $u_h(\bar{b})$ à partir de cette équation. En effet, si on avait par exemple $\bar{b} = 0$, on ne pourrait pas trouver $u_h(\bar{b})$ de cette façon.

Le théorème suivant [5] nous permet d'obtenir une estimation de l'erreur :

Théorème 24. *Supposons que a et u sont représentés par des polynômes par morceaux de degrés r et $r+1$ respectivement.*

Supposons de plus que $a \in H^{r+1}(\Omega)$ et qu'il existe $\gamma > 0$ et un vecteur unitaire constant ν tel que $\nabla u \cdot \nu \geq \gamma$ dans Ω .

Alors il existe une constante $C > 0$ telle que $\forall a_h$ qui minimise J ,

$$\|a_h - a\|_{L^2(\Omega)} \leq C \left(h^r + \frac{\|z - u\|_{L^2(\Omega)}}{h^2} \right),$$

pour h suffisamment petit.

Remarque 25. Si z est un interpolant de u par polynômes de degré $r + 1$, alors $\|z - u\|_{L^2(\Omega)} = O(h^{r+2})$

Corollaire 26. Si les hypothèses du théorème sont vérifiées et que z est un interpolant de u par polynômes de degré $r + 1$, alors

$$\|a_h - a\|_{L^2(\Omega)} = O(h^r).$$

3.2 Tentative de résolution du problème approché

Soit Δ_h une triangulation régulière et quasi-uniforme de $\bar{\Omega}$.

On choisit

$$\begin{aligned} T_h &= \{v \in C^0(\bar{\Omega}); v|_T \in \mathbb{P}_r \forall T \in \Delta_h\}, \\ S_h &= \{v \in C^0(\bar{\Omega}); v|_T \in \mathbb{P}_{r+1} \forall T \in \Delta_h\}. \end{aligned}$$

Soient $\{\varphi_i\}_{i=1,\dots,m}$ et $\{\psi_i\}_{i=1,\dots,n}$ des bases de S_h et T_h respectivement.

On prend $z \in S_h$ approximation de u .

Calcul de u_h

On peut alors écrire $u_h(\bar{b}) = \sum_{i=1}^m u_i \varphi_i$.

Alors on a $\forall v_h \in S_h$

$$\begin{aligned} \int_{\Omega} \bar{b} \nabla u_h(\bar{b}) \cdot \nabla v_h \, dx &= \int_{\Omega} f v_h \, dx \\ \Leftrightarrow \sum_{i=1}^m \int_{\Omega} \bar{b} u_i \nabla \varphi_i \cdot \nabla \varphi_j \, dx &= \int_{\Omega} f \varphi_j \, dx, \quad \forall j = 1, \dots, m \\ \Leftrightarrow A(\bar{b}) \vec{u}_h(\bar{b}) &= \vec{f}, \end{aligned} \tag{12}$$

avec

$$\begin{aligned} A_{i,j}(\bar{b}) &= \int_{\Omega} \bar{b} \nabla \varphi_i \cdot \nabla \varphi_j \, dx, \\ \vec{u}_h(\bar{b}) &= \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}, \\ \vec{f} &= \begin{pmatrix} \int_{\Omega} f \varphi_1 \, dx \\ \vdots \\ \int_{\Omega} f \varphi_m \, dx \end{pmatrix}. \end{aligned}$$

On peut alors obtenir $\vec{u}_h(\bar{b})$, pour $\bar{b} \in K_h = \{\bar{b} \in T_h; 0 < \alpha \leq \bar{b} \leq \beta\}$ donné, en résolvant le système 12.

Minimisation de $J(\bar{b})$

On cherche maintenant à résoudre le problème (11).

Pour cela, on veut utiliser la méthode de la plus grande pente pour minimiser la fonctionnelle J :

- On choisit $\bar{b}_0 \in K_h$, $k := 0$
- On calcule $\nabla J(\bar{b}_k)$
- Tant que $\|\nabla J(\bar{b}_k)\| > \text{tol}$,
 1. On minimise la fonction $s: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ définie par $s(x) = J(\bar{b}_k - x \nabla J(\bar{b}_k))$. Le minimum est atteint en x_k
 2. On définit $\bar{b}_{k+1} = \bar{b}_k - x_k \nabla J(\bar{b}_k)$

3. $k = k + 1$

4. On calcule $\nabla J(\bar{b}_k)$

- $\bar{a} = \bar{b}_k$

Soit $\bar{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ la solution obtenue.

Alors une solution de (11) sera donnée par

$$a_h = \sum_{i=1}^n a_i \psi_i.$$

Cependant la méthode de la plus grande pente présentée ci-dessus ne garantit pas que la solution a_h soit dans K_h . Il nous faut donc adapter la méthode de manière à prendre en compte les contraintes, nous nous intéressons alors à la méthode de la plus grande pente hybride décrite par M.Gerard, B.De Schutter et M.Verhaegen [7].

De plus le calcul du gradient $\nabla J(\bar{b}_k)$ peut être réalisé efficacement en suivant la méthode décrite par R.Falk [4] et G.Chavent [2].

3.3 Modification de la méthode de résolution du problème approché

3.3.1 Méthode de la plus grande pente hybride

Considérons un problème d'optimisation convexe

$$\begin{aligned} \min_{x \in \mathbb{R}^m} J(x) \\ \text{tel que } g(x) \leq 0, \end{aligned} \quad (13)$$

où $J: \mathbb{R}^m \rightarrow \mathbb{R}$ est une fonction différentiable convexe, et $g: \mathbb{R}^m \rightarrow \mathbb{R}^n$, $g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}$, est telle que chaque g_i est une fonction différentiable convexe.

On suppose

1. $\exists x_f$ tel que $g(x_f) \leq 0$, et on définit $c^* = \min_x \{J(x); g(x) \leq 0\}$.
2. c^* est fini.

On cherche alors une fonction $f: \mathbb{R}^m \rightarrow \mathbb{R}^m$ telle qu'une solution du système

$$\dot{x}(t) = f(x(t)) \quad (14)$$

ait les propriétés suivantes :

- Si $\exists \bar{t}; g(x(\bar{t})) > 0$, alors $\exists t_f > \bar{t}; g(x(t_f)) \leq 0$.
- Si $g(x(t_f)) \leq 0$, alors $\forall t > t_f, g(x(t)) \leq 0$.
- $\forall t_2 > t_1 \geq t_f, J(x(t_1)) > J(x(t_2)) > c^*$ et $\lim_{t \rightarrow \infty} J(x(t)) = c^*$.

M.Gerard, B.De Schutter et M.Verhaegen [7] ont montré le résultat suivant

Proposition 27. *Si l'on choisit*

$$f(x) = \begin{cases} -\nabla J(x) & \text{si } g_j(x) \leq 0 \forall j \\ -\sum_{i \in L(x)} \nabla g_i(x) & \text{si } \exists j; g_j(x) > 0 \end{cases}, \quad (15)$$

où $L(x) = \{i; g_i(x) \geq 0\}$, une solution $x(t)$ du système (14) est telle que $\lim_{t \rightarrow \infty} J(x(t)) = c^*$, où c^* est solution du problème (13), si ce problème admet une solution.

Pour obtenir une solution du problème (13), on peut utiliser une méthode d'Euler explicite : on définit la suite $\{x_i\}_{i \geq 0}$ telle que

$$\begin{aligned} x_0 &= \lambda \text{ fixé,} \\ x_{k+1} &= x_k + \Delta t f(x_k), \end{aligned}$$

avec Δt suffisamment petit pour avoir convergence.

On a alors

$$\lim_{k \rightarrow +\infty} J(x_k) = c^*,$$

où c^* est solution du problème (13).

Remarque 28. Si la fonction J n'est pas convexe, le minimum obtenu par la méthode de la plus grande pente hybride peut n'être qu'un minimum local.

3.3.2 Calcul du gradient

Nous nous intéressons maintenant à une méthode efficace pour le calcul du gradient $\nabla J(b_k)$, qui est décrite par R.Falk [4] et G.Chavent [2].

On conserve les mêmes notations que dans la partie 3.2.

On a

$$J(b) = \|u_h(b) - z\|_{L^2(\Omega)}^2 = (u_h(b) - z, u_h(b) - z)_{L^2}.$$

Notons $U(b)$ la dérivée de $u_h(b)$.

Alors on a

$$u_h(b + \delta b) = u_h(b) + U(b)\delta b + O(\|\delta b\|^2),$$

et

$$J(b + \delta b) = J(b) + 2(U(b)\delta b, u_h(b) - z)_{L^2} + O(\|\delta b\|^2),$$

donc

$$\nabla J(b)\delta b = 2(U(b)\delta b, u_h(b) - z)_{L^2}. \quad (16)$$

Tout $v_h \in S_h$ peut s'écrire $\sum_{i=1}^m v_i \varphi_i$, où $\{\varphi_i\}_{i=1, \dots, m}$ est une base de S_h .

On définit formellement l'opérateur \tilde{A} qui à $b \in T_h$ associe $\tilde{A}(b)$ tel que pour tout $v_h, w_h \in S_h$,

$$\left(\tilde{A}(b)v_h, w_h \right) = \sum_{j=1}^m \left(\sum_{i=1}^m w_i A_{i,j}(b) v_i \right)$$

où $A(b)$ est la matrice définie dans la partie 3.2.

On a alors en particulier pour tout $b \in T_h$,

$$\tilde{A}(b)u_h(b) = f$$

et donc

$$\begin{aligned} \tilde{A}(b + \delta b) &= \tilde{A}(b) + \tilde{A}(\delta b), \\ \tilde{A}(b + \delta b)u_h(b + \delta b) &= \tilde{A}(b)u_h(b) + \tilde{A}(b)U(b)\delta b + \tilde{A}(\delta b)u_h(b) + O(\|\delta b\|^2). \end{aligned}$$

Donc

$$\begin{aligned} \tilde{A}(b)U(b)\delta b + \tilde{A}(\delta b)u_h(b) &= \left(\nabla \tilde{A}(b)u_h(b) \right) \delta b \\ &= (\nabla f)\delta b \\ &= 0. \end{aligned}$$

Donc

$$U(b)\delta b = -\tilde{A}(b)^{-1}\tilde{A}(\delta b)u_h(b). \quad (17)$$

On peut alors d eduire de (16) et (17) que

$$\begin{aligned}\nabla J(a)\delta a &= -2\left(\tilde{A}(b)^{-1}\tilde{A}(\delta b)u_h(b), u_h(b) - z\right)_{L^2} \\ &= -2\left(\tilde{A}(\delta b)u_h(b), \left(\tilde{A}(b)^{-1}\right)^T(u_h(b) - z)\right)_{L^2}.\end{aligned}$$

Notons $p(b)$ la solution de

$$\tilde{A}(b)^T = 2(u_h(b) - z). \quad (18)$$

On obtient alors

$$\nabla J(a)\delta a = -\left(\tilde{A}(\delta b)u_h(b), p(b)\right)_{L^2}. \quad (19)$$

$u_h(b)$ et $p(b)$ sont dans S_h , donc on peut  crire

$$\begin{aligned}u_h(b) &= \sum_{i=1}^m u_i\varphi_i, \\ p(b) &= \sum_{i=1}^m p_i\varphi_i.\end{aligned}$$

$b \in T_h$, donc on peut  crire

$$b = \sum_{i=1}^n b_i\psi_i.$$

On peut alors obtenir $\nabla J(b)$ pour tout $b \in K_h$ en calculant pour $k = 1 \dots n$

$$\begin{aligned}\frac{\partial J(b)}{\partial b_k} &= \nabla J(b) \cdot \psi_k \\ &= -(A(\psi_k)u_h(b), p(b))_{L^2} \\ &= -\sum_{i,j=1}^m u_i p_j \int_{\Omega} (A(\psi_k)\varphi_i) \cdot \varphi_j \, dx,\end{aligned} \quad (20)$$

et

$$\nabla J(b) = \begin{pmatrix} \frac{\partial J(b)}{\partial b_1} \\ \vdots \\ \frac{\partial J(b)}{\partial b_n} \end{pmatrix}. \quad (21)$$

Remarque 29. On peut d emontrer ce r esultat rigoureusement en supposant que $z \in S_h$, alors on peut  crire $z = \sum_{i=1}^m z_i\varphi_i$ et $u_h(b) = \sum_{i=1}^m u_i(b)\varphi_i$, et en utilisant la norme l^2 pour J

$$J(b) = \|\vec{u}_h(b) - \vec{z}\|_{l^2}^2 = (\vec{u}_h(b) - \vec{z})^T (\vec{u}_h(b) - \vec{z}),$$

o u

$$\begin{aligned}\vec{u}_h(b) &= \begin{pmatrix} u_1(b) \\ \vdots \\ u_m(b) \end{pmatrix}, \\ \vec{z} &= \begin{pmatrix} z_1 \\ \vdots \\ z_m \end{pmatrix}.\end{aligned}$$

3.3.3 R esolution du probl eme approch e

On a le probl eme d'optimisation (11). On introduit donc le probl eme de minimisation suivant :

$$\begin{aligned}\min_{\vec{b} \in \mathbb{R}^n} J(\vec{b}) \\ \text{tel que } 0 < \alpha \leq b_i \leq \beta, \forall i = 1, \dots, n\end{aligned} \quad (22)$$

Soit $g: \mathbb{R}^n \rightarrow \mathbb{R}^{2n}$ définie par $g(\bar{b}) = \begin{pmatrix} \alpha - a_1 \\ \vdots \\ \alpha - a_m \\ a_1 - \beta \\ \vdots \\ a_m - \beta \end{pmatrix}$.

On a

$$0 < \alpha \leq b_i \leq \beta, \forall i = 1, \dots, m \Leftrightarrow g(\bar{b}) \leq 0$$

Pour trouver une solution du problème (22) on applique la méthode de la plus grande pente hybride, en calculant le gradient de J comme décrit plus haut.

On obtient alors l'algorithme suivant :

- On choisit $\bar{b}_0 \in K_h, k := 0$
- On calcule $\nabla J(\bar{b}_k)$ en utilisant la méthode décrite en 3.3.2.
- Tant que $\|\nabla J(\bar{b}_k)\| > \text{tol}$,
 1. On définit $\bar{b}_{k+1} = \bar{b}_k + \Delta t f(\bar{b}_k)$, où f est la fonction définie en (15).
 2. $k = k + 1$
 3. On calcule $\nabla J(\bar{b}_k)$ en utilisant la méthode décrite en 3.3.2.
- $\bar{a} = \bar{b}_k$

Soit $\bar{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ la solution obtenue, alors une solution de (11) sera donnée par $a_h = \sum_{i=1}^n a_i \psi_i$.

3.4 Application à un problème 1D

Par souci de simplicité, on choisit $\Omega = (0, 1)$, et on considère l'équation

$$\begin{cases} -(au')' = f \text{ sur } (0, 1) \\ u(0) = \mu_0 \\ u(1) = \mu_1 \end{cases}, \quad (23)$$

on suppose connue la fonction f , et on définit le maillage

$$\Delta_h = \{K_i = [X_i, X_{i+1}], i = 0, \dots, N-1\},$$

avec $X_i = (i-1)h$ pour $i = 1, \dots, N$, $h = \frac{1}{N}$.

On choisit

$$\begin{aligned} T_h &= \{v \in C^0(\bar{\Omega}); v|_T \in \mathbb{P}_r \forall T \in \Delta_h\}, \\ S_h &= \{v \in C^0(\bar{\Omega}); v|_T \in \mathbb{P}_{r+1} \forall T \in \Delta_h\}. \end{aligned}$$

Soient $\{\varphi_i\}_{i=1, \dots, m}$ et $\{\psi_i\}_{i=1, \dots, n}$ des bases de S_h et T_h respectivement.

Pour $r = 1$, on définit

$$\begin{aligned} x_{2i} &= X_i \text{ pour } i = 0, \dots, N, \\ x_{2i+1} &= \frac{X_i + X_{i+1}}{2} \text{ pour } i = 0, \dots, N-1, \end{aligned}$$

et les bases de S_h et T_h sont données par $\{\varphi_i\}_{i=0,\dots,2N-1}$ et $\{\psi_i\}_{i=0,\dots,N}$

$$\varphi_{2i}(x) = \begin{cases} \frac{(x-x_{2i-2})(x-x_{2i-1})}{(x_{2i}-x_{2i-2})(x_{2i}-x_{2i-1})} & \text{si } i > 0 \text{ et } x \in [x_{2i-2}, x_{2i}] \\ \frac{(x-x_{2i+1})(x-x_{2i+2})}{(x_{2i}-x_{2i+1})(x_{2i}-x_{2i+2})} & \text{si } i < N \text{ et } x \in [x_{2i}, x_{2i+2}] \\ 0 & \text{sinon} \end{cases},$$

$$\varphi_{2i+1}(x) = \begin{cases} \frac{(x-x_{2i})(x-x_{2i+2})}{(x_{2i+1}-x_{2i})(x_{2i+1}-x_{2i+2})} & \text{si } x \in [x_{2i}, x_{2i+2}] \\ 0 & \text{sinon} \end{cases},$$

$$\psi_i(x) = \begin{cases} \frac{x-X_i}{h} & \text{si } i > 0 \text{ et } x \in [X_{i-1}, X_i] \\ 1 - \frac{x-X_i}{h} & \text{si } i < N \text{ et } x \in [X_i, X_{i+1}] \\ 0 & \text{sinon} \end{cases}.$$

On obtient une approximation $z \in S_h$ de u à partir de M valeurs de u mesurées (par interpolation).

On définit $K_h = \{\bar{b} \in T_h; 0 < \alpha \leq \bar{b} \leq \beta\}$

Pour trouver la solution du problème inverse en utilisant la méthode OLS, il faut alors résoudre le problème (11).

Pour tout $b \in K_h$, on peut calculer $u_h(b)$ en résolvant le système (12).

On obtient alors $\nabla J(b)$ en résolvant le système (18), ce qui nous donne $p(b)$, puis en calculant pour $k = 1 \dots n$

$$\frac{\partial J(b)}{\partial b_k} = - \sum_{i,j=0}^m u_i p_j \int_{\Omega} (A(\psi_k) \varphi_i) \cdot \varphi_j \, dx,$$

avec $u_h(b) = \sum_{i=0}^m u_i \varphi_i$ et $p(b) = \sum_{i=0}^m p_i \varphi_i$.

On applique ensuite l'algorithme décrit dans la partie 3.3.3, ce qui nous donne $\bar{a} = \begin{pmatrix} a_0 \\ \vdots \\ a_n \end{pmatrix}$, et une solution approchée du problème inverse est alors donnée par

$$a_h = \sum_{i=0}^n a_i \psi_i.$$

Remarque 30. Si J n'est pas convexe, on peut n'obtenir qu'un minimum local, qui peut être éloigné de la solution exacte.

Il faut donc prendre des bornes α et β aussi précises que possible pour obtenir une solution acceptable.

Remarque 31. Pour la méthode OLS, faire varier N pour voir le comportement de l'erreur n'a pas beaucoup d'intérêt.

En effet, on n'arrête l'algorithme que lorsque que la norme du gradient est inférieure à une tolérance fixée tol , et ce quel que soit le nombre de noeuds du maillage. La tolérance tol permet de définir la précision à partir de laquelle on juge que la solution approchée est suffisamment proche de la solution du problème d'optimisation, ce qui est assez similaire à l'erreur en norme L^2 de la solution approchée a_h .

Alors pour une tolérance fixée, pour tout N la solution approchée sera à la même 'distance' de la solution exacte (si la méthode converge), et dnc l'erreur sera du même ordre.

Par conséquent, dans les applications suivantes, on résout pour des valeurs de M et N fixées.

3.4.1 Application au problème 1

On choisit

$$\begin{aligned} u(x) &= x^2 - x, \\ f(x) &= 1. \end{aligned}$$

Alors la solution exacte du problème inverse est donnée par $a(x) = \frac{1}{2}$.
On prend

$$\begin{aligned} M &= 4, \\ N &= 10, \\ K_h &= \{\bar{b} \in T_h; \alpha = 0.4 \leq \bar{b} \leq 0.6 = \beta\}, \end{aligned}$$

et on obtient les résultats suivants en prenant un $\bar{b}_0 \in K_h$ aléatoire pour la méthode de la plus grande pente hybride :

Sans bruit :

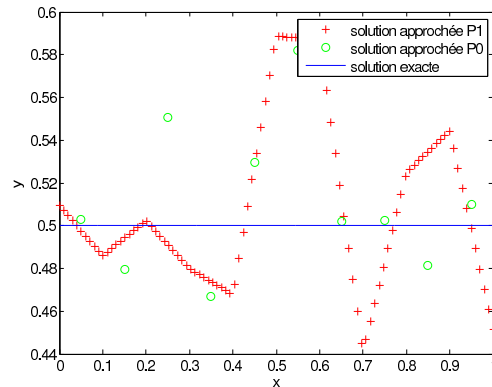


Figure 46. Solution approchée du pb 1 avec OLS - $\text{tol} = 1e - 5$

On a ici erreur = 0.0367 pour la solution approchée P0, et erreur = 0.0393 pour la solution approchée P1.

La solution approchée \bar{a}_h du problème 1 est bien dans les bornes α et β , mais paraît assez éloignée de la solution exacte.

Afin de vérifier qu'on a bien obtenu une solution approchée \bar{a}_h correcte du problème d'optimisation, on regarde l'allure de $u_h(\bar{a}_h)$.

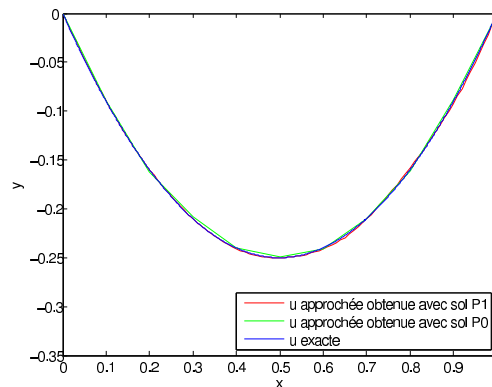


Figure 47. $u_h(\bar{a}_h)$, avec \bar{a}_h solution approchée du pb 1 - $\text{tol} = 1e - 5$

On voit que $u_h(\bar{a}_h)$ est proche de u exacte (pour des éléments finis P1 et P0), donc on peut raisonnablement supposer que la méthode a convergé correctement, mais seulement vers un minimum local.

Afin d'obtenir une meilleure solution, on pourrait affiner les bornes α et β .

Avec bruit :

Pour un bruit trop élevé, on remarque que la méthode ne converge pas forcément. On prend donc un critère de convergence plus large que précédemment ($\text{tol} = 1e - 4$), et on ajoute un bruit aléatoire inférieur à 2.5% de leur valeur réelle aux mesures de u et f .

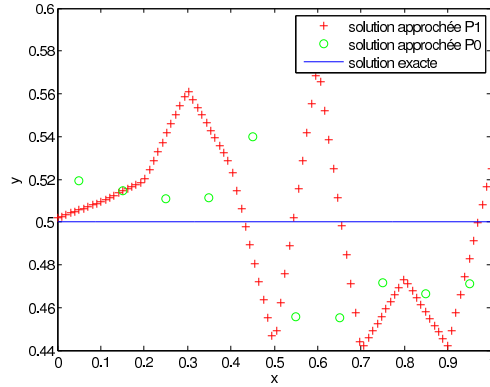


Figure 48. Solution approchée du pb 1 avec OLS - tol = $1e-4$ - avec bruit

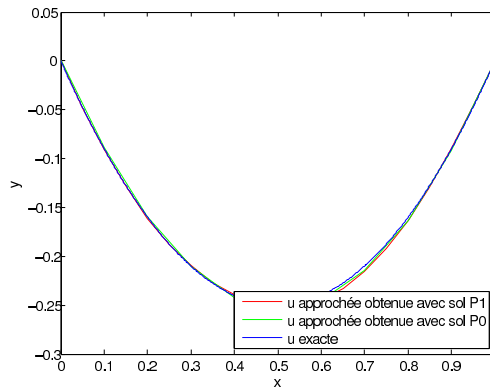


Figure 49. $u_h(\bar{a}_h)$, avec \bar{a}_h solution approchée du pb 1 - tol = $1e-4$ - avec bruit

On erreur = 0.0309 pour la solution approchée P0, et erreur = 0.0356 pour la solution approchée P1.

Les résultats sont similaires à ceux obtenus sans bruit, et l'approximation $u_h(\bar{a}_h)$ est acceptable.

3.4.2 Application au problème 2bis

On choisit

$$u(x) = \frac{x^4}{6} - \frac{x^3}{6} + x,$$

$$f(x) = -8x^3 + \frac{5x^2}{2} + x - 3.$$

Alors la solution exacte du problème inverse est donnée par $a(x) = 3x + 1$.

On prend

$$M = 4,$$

$$N = 10,$$

$$K_h = \{\bar{b} \in T_h; \alpha \leq \bar{b} \leq \beta\},$$

et on obtient les résultats suivants en prenant un $\bar{b}_0 \in K_h$ aléatoire pour la méthode de la plus grande pente hybride :

Sans bruit :

On prend dans un premier temps $\alpha = 0.5$, $\beta = 5$.

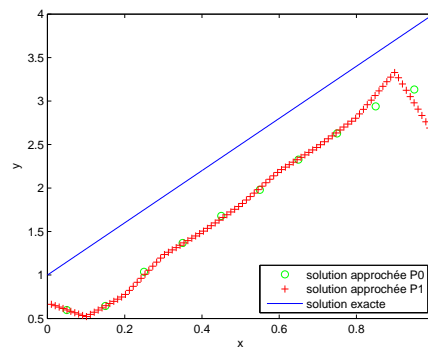


Figure 50. Sol approchée du pb 2bis avec OLS -
 $\text{tol} = 1e - 5$, $\alpha = 0.5$, $\beta = 5$

On a ici erreur = 0.8972 pour la solution approchée avec des éléments finis P0, et erreur = 0.9752 pour la solution approchée avec des éléments finis P1.

La solution approchée \bar{a}_h du problème 1 est bien dans les bornes α et β , mais est très éloignée de la solution exacte .

On regarde l'allure de $u_h(\bar{a}_h)$

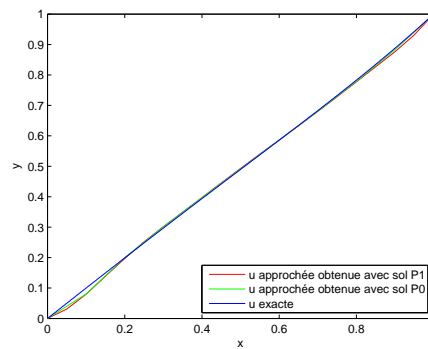


Figure 51. $u_h(\bar{a}_h)$, avec \bar{a}_h solution approchée du pb 2bis -
 $\text{tol} = 1e - 5$, $\alpha = 0.5$, $\beta = 5$

On voit que $u_h(\bar{a}_h)$ est assez proche de u exacte (pour P1 et P0), on peut considérer que la méthode a convergé correctement, mais vers un minimum local.

Afin d'obtenir une meilleure solution, on affine les bornes α et β .

On prend alors $\alpha'(x) = 3x + 0.5$, $\beta'(x) = 3x + 1.5$.

C'est-à-dire que chaque point de la solution approchée \bar{a}_h doit se trouver à une distance inférieure ou égale à 0.5 de la solution exacte.

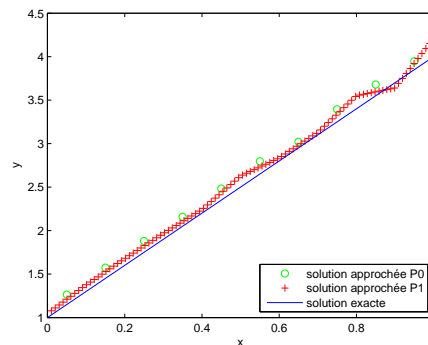


Figure 52. Sol approchée du pb 2bis avec OLS - $\text{tol} = 1e - 6$, α' , β'

On a erreur = 0.1237 pour la solution approchée avec des éléments finis P0, et erreur = 0.0773 pour la solution approchée avec des éléments finis P1.

La solution approchée \bar{a}_h du problème 1 est désormais acceptable.

On vérifie l'allure de $u_h(\bar{a}_h)$.

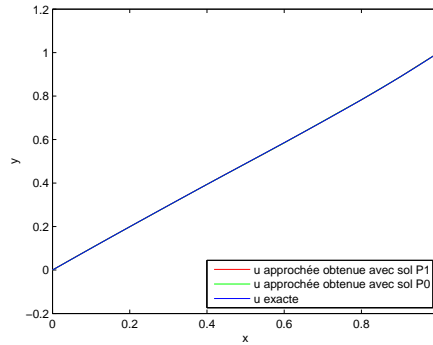


Figure 53. $u_h(\bar{a}_h)$, avec \bar{a}_h sol approchée du pb 2bis - tol = $1e-6$, α' , β'

$u_h(\bar{a}_h)$ est proche de u exacte.

Avec bruit :

Pour un bruit trop élevé, on remarque que la méthode ne converge pas forcément. On prend donc un critère de convergence plus large que précédemment (tol = $1e-4$), et on ajoute un bruit aléatoire inférieur à 2.5% de leur valeur réelle aux mesures de u et f .

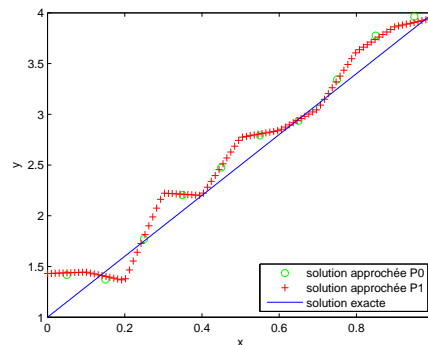


Figure 54. Sol approchée du pb 2 avec OLS - tol = $1e-4$, α' , β' - avec bruit

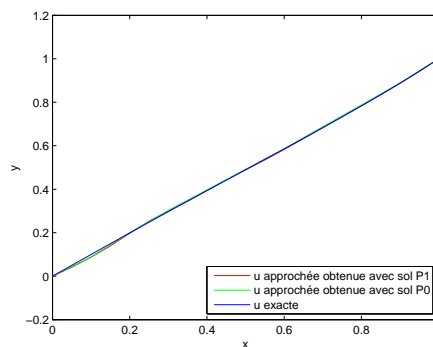


Figure 55. $u_h(\bar{a}_h)$, avec \bar{a}_h sol approchée du pb 2 - tol = $1e-4$, α' , β' - avec bruit

On a erreur = 0.1364 pour la solution approchée avec des éléments finis P0, et erreur = 0.1665 pour la solution approchée avec des éléments finis P1.

L'erreur est assez faible, et l'allure de la courbe de la solution approchée \bar{a}_h est similaire à celle de la solution exacte. Donc ces résultats sont acceptables.

On a constaté que la méthode OLS permet d'obtenir une erreur faible lorsque le problème d'optimisation est convexe ou que les bornes α et β sont précises. Cependant les propriétés de la solution, notamment la régularité ou la convexité, ne sont pas forcément respectées.

En présence d'un bruit faible ($\leq 2.5\%$) sur u et f , la méthode donne encore une solution acceptable.

On s'intéresse désormais à une troisième méthode, la méthode MOLS (Modified Output Least Square).

4 MOLS

La méthode MOLS proposée par Zou [8] est similaire à la méthode OLS. En effet, il s'agit de résoudre le problème inverse en résolvant un problème de minimisation, mais en prenant une fonctionnelle modifiée de manière à obtenir un problème convexe.

4.1 Formulation du problème

Dans une équation aux dérivées partielles de la forme (1)

$$\begin{cases} -\nabla \cdot (a\nabla u) = f \text{ sur } \Omega \\ u(x) = 0 \text{ sur } \Gamma \subset \partial\Omega \end{cases},$$

on suppose que f est donnée, que a ne peut pas être mesuré directement, mais que l'on peut obtenir des valeurs de u en certains points de mesure de manière à obtenir une approximation z . Comme précédemment, on suppose qu'on peut simuler les données observables (c'est-à-dire u) pour toute estimation raisonnable \bar{a} des paramètres a . On obtient alors une estimation $u(\bar{a})$ des données u . Cette estimation peut être obtenue par la méthode des éléments finis, et on a alors $u_h(a_h)$, où u_h et a_h sont des polynômes par morceaux.

On définit le problème d'identification des paramètres

$$\text{Trouver } a \in K \text{ tel que } J_{\text{MOLS}}(a) = \inf_{b \in K} J_{\text{MOLS}}(b), \quad (24)$$

où

$$J_{\text{MOLS}}(b) = \int_{\Omega} b |\nabla v - \nabla z|^2 dx + \gamma N(b),$$

avec $v = v(b) \in H_0^1(\Omega)$ tel que

$$\int_{\Omega} b \nabla v \nabla \phi dx = \int_{\Omega} f \phi dx \quad \forall \phi \in H_0^1(\Omega),$$

$\gamma > 0$ paramètre de régularisation,

$$N(b) = \int_{\Omega} |\nabla b|^2 dx,$$

$$K = \left\{ b \in L_1(\Omega); \|b\|_{H^1(\Omega)} < \infty \text{ et } \alpha \leq b(x) \leq \beta \text{ presque partout dans } \Omega \right\}, \quad \alpha, \beta \in \mathbb{R}_{\geq 0}.$$

Le terme $\gamma N(b)$ permet de rendre le problème convexe.

4.2 Discrétisation

On définit

T_h une triangulation régulière de Ω par éléments finis simplectiques
 V_h espace d'éléments finis continus et linéaires par morceaux sur T_h
 $\overset{\circ}{V}_h$ sous-espace de V_h tel que $\forall v_h \in \overset{\circ}{V}_h, v_h|_{\partial\Omega} = 0$,
 $K_h = \{v_h \in V_h; \alpha \leq v_h(x) \leq \beta, \forall x \in \Omega\}$.

On considère $\{x_i\}_{i=1, \dots, N}$ l'ensemble des noeuds de la triangulation T_h .
 On a alors

$$K_h = \{v_h \in V_h; \alpha \leq v_h(x_i) \leq \beta, \forall i = 1, \dots, N\}.$$

On définit le problème discrétisé :

$$\text{Trouver } a_h \in K_h \text{ tel que } J_h(a_h) = \inf_{\bar{b} \in K_h} J_h(\bar{b}), \quad (25)$$

où

$$J_h(\bar{b}) = \int_{\Omega} \bar{b} |\nabla v_h - \nabla z|^2 dx + \gamma N_h(\bar{b}),$$

avec $v_h = v_h(\bar{b}) \in \overset{\circ}{V}_h$ tel que $\int_{\Omega} \bar{b} \nabla v_h \nabla \phi dx = \int_{\Omega} f \phi dx \forall \phi \in \overset{\circ}{V}_h$,

$$\gamma > 0,$$

$$N_h(\bar{b}) = \int_{\Omega} |\nabla \bar{b}|^2 dx.$$

Zou [8] a montré les résultats suivants :

Proposition 32. *Il existe au moins un minimiseur du problème (24).*

Théorème 33. *Soit $\{b_h\}_{h>0}$ une suite de minimiseurs du problème discrétisé (25).*

Alors chaque sous-suite de $\{b_h\}_{h>0}$ a une sous-suite convergeant vers un minimiseur du problème continu (24).

4.3 Résolution du problème discrétisé

On résout le problème discrétisé à l'aide de la méthode de la plus grande pente hybride 3.3.1.

Pour cela on a besoin de calculer le gradient de J_h .

On remplace z par son interpolant z_h dans la base des éléments finis.

On définit

$$J_1(\bar{b}) = \int_{\Omega} \bar{b} |\nabla v_h - \nabla z_h|^2 dx,$$

$$J_2(\bar{b}) = N_h(\bar{b}) = \int_{\Omega} |\nabla \bar{b}|^2 dx.$$

On a $J_h = J_1 + \gamma J_2$.

On cherche maintenant à calculer le gradient de J_1 et J_2 .

Lemme 34. $v_h(\bar{b}): V_h \rightarrow \overset{\circ}{V}_h$ satisfait $v_h'(\bar{b}) p_h \in \overset{\circ}{V}_h$ et

$$\int_{\Omega} \bar{b} \nabla(v_h'(\bar{b}) p_h) \nabla \phi_h dx = - \int_{\Omega} p_h \nabla v_h(\bar{b}) \nabla \phi_h dx \forall \phi_h \in \overset{\circ}{V}_h. \quad (26)$$

Démonstration. $v_h(\bar{b}) \in \overset{\circ}{V}_h$ est défini comme solution de

$$\int_{\Omega} \bar{b} \nabla v_h(\bar{b}) \nabla \phi \, dx = \int_{\Omega} f \phi \, dx \quad \forall \phi \in \overset{\circ}{V}_h.$$

Donc

$$\int_{\Omega} \bar{b} \nabla (v_h(\bar{b} + tp_h) - v_h(\bar{b})) \nabla \phi_h \, dx + t \int_{\Omega} p_h \nabla v_h(\bar{b} + tp_h) \nabla \phi_h \, dx = 0 \quad \forall \phi \in \overset{\circ}{V}_h.$$

Donc en divisant par t prenant la limite quand $t \rightarrow 0$, on obtient le résultat du lemme. \square

Lemme 35. $J'_1(\bar{b})p_h = - \int_{\Omega} p_h (|\nabla v_h(\bar{b})|^2 - |\nabla z_h|^2) dx.$

Démonstration. $\forall 0 < t < 1, \bar{b}, p_h \in V_h$, on a

$$\begin{aligned} J_1(\bar{b} + tp_h) - J_1(\bar{b}) &= \int_{\Omega} (\bar{b} + tp_h) |\nabla v_h(\bar{b} + tp_h) - \nabla z_h|^2 dx - \int_{\Omega} \bar{b} |\nabla v_h(\bar{b}) - \nabla z_h|^2 dx \\ &= \int_{\Omega} \bar{b} \nabla (v_h(\bar{b} + tp_h) - v_h(\bar{b})) \cdot \nabla (v_h(\bar{b} + tp_h) + v_h(\bar{b}) - 2z_h) dx + \\ &\quad t \int_{\Omega} p_h |\nabla v_h(\bar{b} + tp_h) - \nabla z_h|^2 dx. \end{aligned}$$

Donc

$$\begin{aligned} J'_1(\bar{b})p_h &= \int_{\Omega} \bar{b} \nabla (v'_h(\bar{b})p_h) \nabla (2v_h(\bar{b}) - 2z_h) dx + \int_{\Omega} p_h |\nabla v_h(\bar{b}) - \nabla z_h|^2 dx \\ &= - \int_{\Omega} p_h (|\nabla v_h(\bar{b})|^2 - |\nabla z_h|^2) dx. \end{aligned}$$

\square

Lemme 36. $J'_2(\bar{b})p_h = 2 \int_{\Omega} \nabla \bar{b} \nabla p_h dx.$

Démonstration. $\forall 0 < t < 1, \bar{b}, p_h \in V_h$, on a

$$\begin{aligned} J_2(\bar{b} + tp_h) - J_2(\bar{b}) &= \int_{\Omega} |\nabla \bar{b} + t \nabla p_h|^2 dx - \int_{\Omega} |\nabla \bar{b}|^2 dx \\ &= 2 \int_{\Omega} \nabla \bar{b} \nabla p_h dx + t^2 \int_{\Omega} |\nabla p_h|^2 dx. \end{aligned}$$

Donc

$$\begin{aligned} J'_1(\bar{b})p_h &= \lim_{t \rightarrow 0} \frac{J_2(\bar{b} + tp_h) - J_2(\bar{b})}{t} \\ &= 2 \int_{\Omega} \nabla \bar{b} \nabla p_h dx. \end{aligned}$$

\square

Et donc si on prend $\{\varphi_i\}_{i=0, \dots, N}$ une base de V_h , on peut écrire

$$\bar{b} = \sum_{i=0}^N b_i \varphi_i,$$

et donc

$$\frac{\partial J_h(\bar{b})}{\partial b_i} = J'_1(\bar{b})\varphi_i + \gamma J'_2(\bar{b})\varphi_i \quad (27)$$

et donc obtenir le gradient $\nabla J_h(\bar{b})$ pour tout $\bar{b} \in V_h$.

4.4 Application à un problème 1D

On prend $\Omega = (0, 1)$, et on définit

$$\begin{aligned} 0 &= x_0 < \dots < x_N = 1, \\ h &= \frac{1}{N}, \\ T_h &= \{[x_i, x_{i+1}]\} \text{ triangulation régulière,} \\ V_h &= \{v_h \in C^0(\Omega); v_h|_{[x_i, x_{i+1}]} \in \mathbb{P}_1([x_i, x_{i+1}]) \forall i\} \text{ de base } \{\varphi_i\}_{i=0, \dots, N}, \\ \overset{\circ}{V}_h &= \{v_h \in V_h; v_h(0) = v_h(1) = 0\} \text{ de base } \{\varphi_i\}_{i=1, \dots, N-1}, \\ K_h &= \{v_h \in V_h; \alpha \leq v_h(x) \leq \beta, \forall x \in \Omega\}. \end{aligned}$$

On obtient $z_h \in V_h$ à partir de M mesures de u .

Le problème approché (25) s'écrit

$$\text{Trouver } a_h \in K_h \text{ tel que } J_h(a_h) = \inf_{\bar{b} \in K_h} J_h(\bar{b}),$$

où

$$J_h(\bar{b}) = \int_{\Omega} \bar{b} |v_h'(x) - z_h'(x)|^2 dx + \gamma N_h(\bar{b}),$$

avec $\gamma > 0$, $v_h = v_h(\bar{b}) = \sum_{i=1}^{N-1} v_i \varphi_i \in \overset{\circ}{V}_h$ tel que

$$\begin{aligned} \int_{\Omega} \bar{b} v_h' \varphi_j' dx &= \int_{\Omega} f \varphi_j' dx \quad \forall j = 1, \dots, N-1 \\ \Leftrightarrow \sum_{i=1}^{N-1} v_i \int_{\Omega} \bar{b} \varphi_i' \varphi_j' dx &= \int_{\Omega} f \varphi_j' dx \quad \forall j = 1, \dots, N-1 \\ \Leftrightarrow A(\bar{b}) \vec{v} &= \vec{f}. \end{aligned}$$

où

$$\begin{aligned} \vec{v} &= \begin{pmatrix} v_1 \\ \vdots \\ v_{N-1} \end{pmatrix}, \\ \vec{f} &= \begin{pmatrix} \int_{\Omega} f \varphi_1' dx \\ \vdots \\ \int_{\Omega} f \varphi_{N-1}' dx \end{pmatrix}, \\ A_{i,j}(\bar{b}) &= \int_{\Omega} \bar{b} \varphi_i' \varphi_j' dx \text{ pour } i, j = 1, \dots, N-1, \end{aligned}$$

et

$$N_h(\bar{b}) = \int_{\Omega} |\bar{b}'|^2 dx = \vec{b}^T B \vec{b},$$

où $\bar{b} \in V_h$ peut s'écrire $\bar{b} = \sum_{i=0}^N b_i \varphi_i$, et

$$\begin{aligned} \vec{b} &= \begin{pmatrix} b_0 \\ \vdots \\ b_N \end{pmatrix}, \\ B_{i,j} &= \int_{\Omega} \varphi_i' \varphi_j' dx \text{ pour } i, j = 0, \dots, N. \end{aligned}$$

Calcul du gradient :

on calcule le gradient $\nabla J_h(\bar{b})$ pour tout $\bar{b} \in V_h$ en utilisant (27) pour $i=0, \dots, N$:

$$\begin{aligned} \frac{\partial J_h(\bar{b})}{\partial \beta_i} &= J'_1(\bar{b})\varphi_i + \gamma J'_2(\bar{b})\varphi_i \\ &= -\sum_{k,l=1}^{N-1} v_k v_l \int_{\Omega} \varphi_i \varphi'_k \varphi'_l dx + \sum_{k,l=0}^N z_k z_l \int_{\Omega} \varphi_i \varphi'_k \varphi'_l dx + 2\gamma \sum_{j=0}^N b_j \int_{\Omega} \varphi'_i \varphi'_j dx \\ &= \vec{v}^T \text{Mat}_i \vec{v} + \vec{z}^T \widetilde{\text{Mat}}_i \vec{z} + 2\gamma (B\vec{b})_i, \end{aligned}$$

avec

$$\begin{aligned} (\text{Mat}_i)_{k,l} &= \int_{\Omega} \varphi_i \varphi'_{k+1} \varphi'_{l+1} dx \text{ pour } k, l = 1, \dots, N-1, \\ (\widetilde{\text{Mat}}_i)_{k,l} &= \int_{\Omega} \varphi_i \varphi'_k \varphi'_l dx \text{ pour } k, l = 0, \dots, N, \end{aligned}$$

et

$$\nabla J_h(\bar{b}) = \begin{pmatrix} \frac{\partial J_h(\bar{b})}{\partial \beta_0} \\ \vdots \\ \frac{\partial J_h(\bar{b})}{\partial \beta_N} \end{pmatrix}.$$

Calcul de $J_h(\bar{b})$:

On a pour tout $\bar{b} \in V_h$,

$$\begin{aligned} N_h(\bar{b}) &= \int_{\Omega} |\bar{b}'|^2 dx = \vec{b}^T B\vec{b}, \\ \int_{\Omega} \bar{b} |v'_h(x) - z'_h(x)|^2 dx &= \sum_{k=0}^N b_k \left(\sum_{i,j=0}^N (v_i - z_i)(v_j - z_j) \int_{\Omega} \varphi_k \varphi'_i \varphi'_j dx \right) \\ &= (\vec{v} - \vec{z})^T A(\bar{b})(\vec{v} - \vec{z}), \end{aligned}$$

donc

$$J_h(\bar{b}) = (\vec{v} - \vec{z})^T A(\bar{b})(\vec{v} - \vec{z}) + \gamma \vec{b}^T B\vec{b}.$$

4.4.1 Application au problème 4

On définit le problème 4 par

Trouver a tel que

$$\begin{cases} -(au')' = f \text{ sur } (0, 1) \\ u(0) = u(1) = 0 \end{cases},$$

avec

$$\begin{aligned} u(x) &= \sin(2\pi x), \\ f(x) &= -(4x - 4\pi \cos(2\pi x))(2\pi \cos(2\pi x)) - (3 + 2x^2 - 2\sin(2\pi x))(-4\pi^2 \sin(2\pi x)), \end{aligned}$$

et la solution exacte est donnée par

$$a(x) = 3 + 2x^2 - 2\sin(2\pi x).$$

Les fonctions u et f ne sont pas polynomiales, donc il est nécessaire de prendre plus de points de mesures que précédemment afin d'obtenir une interpolation polynomiale acceptable.

On prend

$$\begin{aligned} M &= 16, \\ N &= 19, \\ K_h &= \{\bar{b} \in T_h; \alpha \leq \bar{b} \leq \beta\}, \end{aligned}$$

et on obtient les résultats suivants en prenant un $\bar{b}_0 \in K_h$ aléatoire pour la méthode de la plus grande pente hybride :

Sans bruit :

On prend $\alpha = 0.5$, $\beta = 7.5$ et on regarde les résultats pour différents valeurs de γ .

En effet, en modifiant la valeur du paramètre de régularisation γ , on modifie le problème, et on veut donc observer les effets sur la solution approchée.

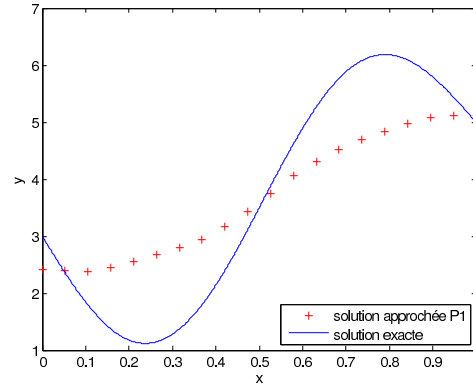


Figure 56. Sol approchée du pb 4 avec MOLS - $\text{tol} = 1e - 5$, $\gamma = 1e - 1$

On a erreur = 0.9863. La solution approchée est très éloignée de la solution exacte, on réduit donc la valeur de γ .

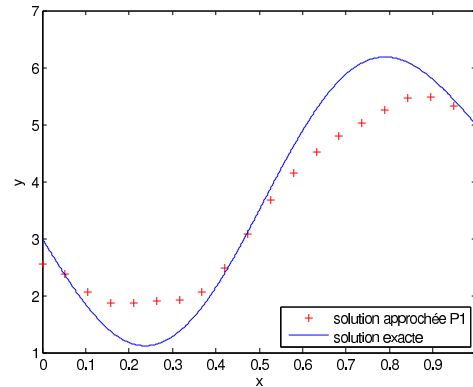


Figure 57. Sol approchée du pb 4 avec MOLS - $\text{tol} = 1e - 5$, $\gamma = 1e - 2$

On a erreur = 0.5710.

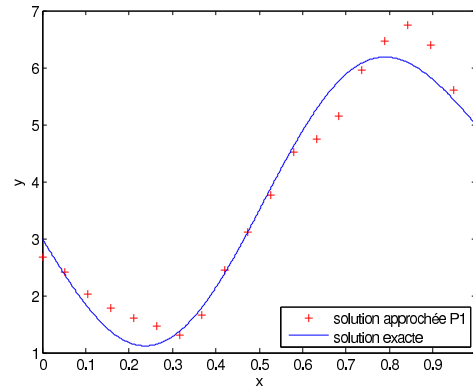


Figure 58. Sol approchée du pb 4 avec MOLS - $\text{tol} = 1e - 5$, $\gamma = 1e - 3$

On a erreur = 0.3361.

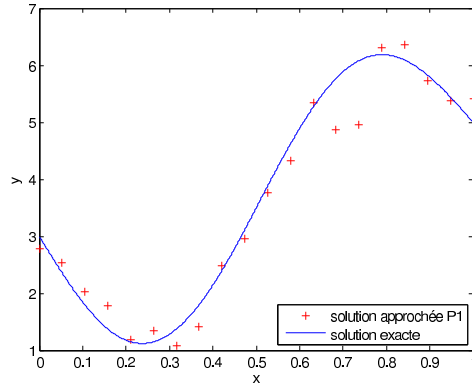


Figure 59. Sol approchée du pb 4 avec MOLS - $\text{tol} = 1e - 5, \gamma = 1e - 4$

On a erreur = 0.3937.

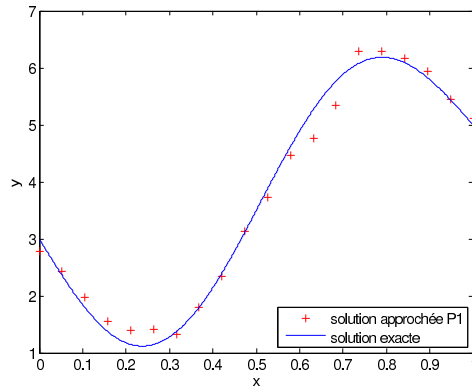


Figure 60. Sol approchée du pb 4 avec MOLS - $\text{tol} = 1e - 5, \gamma = 1e - 5$

On a erreur = 0.2036.

La solution approchée semble acceptable, donc on regarde l'allure de $u_h(a_h)$

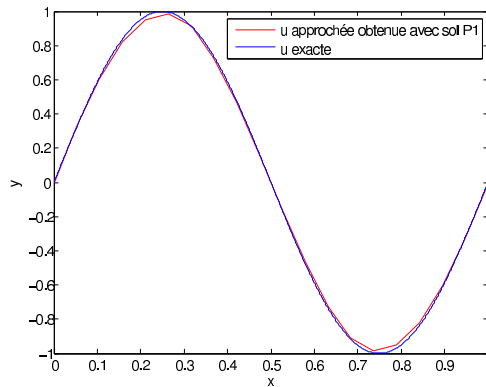


Figure 61. $u_h(a_h)$, avec a_h solution approchée du pb 4 - $\text{tol} = 1e - 5, \gamma = 1e - 5$

On voit que $u_h(a_h)$ est proche de u exacte (pour P1 et P0), donc la méthode a convergé correctement, et la valeur de γ est suffisamment petite pour rendre le problème convexe sans avoir trop d'influence sur la solution.

Avec bruit :

On ajoute un bruit aléatoire inférieur ou égal à 2.5% de la valeur réelle aux mesures de u et f .

On prend $\gamma = 1e - 5$, puisque c'est la valeur pour laquelle on a obtenu les meilleurs résultats.

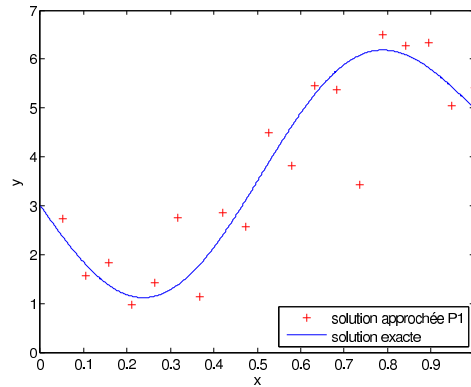


Figure 62. Sol approchée du pb 4 avec MOLS -
tol = $1e-5$, $\gamma = 1e-5$ - avec bruit

On a erreur = 0.8037.

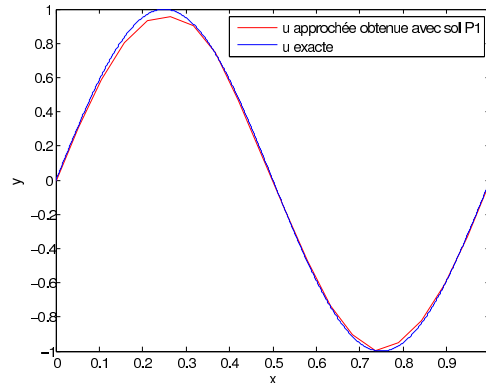


Figure 63. $u_h(a_h)$, avec a_h sol approchée du pb 4 -
tol = $1e-5$, $\gamma = 1e-5$ - avec bruit

La solution approchée a_h n'est pas très bonne, mais la fonction $u_h(a_h)$ est proche de u .

La méthode MOLS semble sensible au bruit.

5 Conclusion

Parmi les méthodes de résolution de problèmes inverses étudiés dans ce projet, la meilleure semble être la méthode de résolution du problème inverse par éléments finis discontinus en utilisant les systèmes de Friedrichs. En effet, elle est plus précise que les méthodes OLS et MOLS, et il est possible d'améliorer les résultats en présence de bruit en utilisant des splines avec contraintes de convexité.

Pour la méthode OLS, si le problème d'optimisation n'est pas convexe il est nécessaire d'avoir beaucoup d'informations à propos de la solution exacte pour obtenir une solution approchée acceptable, faute de quoi la méthode converge vers un minimum local. Pour la méthode MOLS, le paramètre de régularisation modifie le problème, ce qui peut entraîner une erreur non négligeable, et la méthode semble sensible au bruit.

Cependant, en ce qui concerne la méthode de résolution du problème inverse par éléments finis discontinus en utilisant les systèmes de Friedrichs, afin de vérifier les hypothèses des théorèmes de convergence il est nécessaire d'étudier chaque problème et de le modifier de manière adéquate. Il semble néanmoins qu'en appliquant la méthode sur les problèmes non modifiés on obtienne également une solution convenable.

Pour les trois méthodes, il est nécessaire d'avoir des informations à propos de la solution exacte du problème inverse afin d'obtenir une solution approchée acceptable. Par exemple pour la première méthode étudiée, on a besoin des valeurs de la solution exacte sur $\partial\Omega$, et pour la méthode OLS on a besoin d'un maximum d'informations afin d'éviter que la méthode ne converge vers un minimum local du problème d'optimisation.

Bibliographie

- [1] C. De Boor. On calculating with b-splines. *Journal of Approximation Theory*, 1972.
- [2] G. Chavent. Identification of distributed parameters. In *Proc. 3rd IFAC Sympos. (P.Eykhoff,Ed.)*, pp.649-660.
- [3] P. Dierckx. An algorithm for cubic spline fitting with convexity constraints. *Computing*, 1980.
- [4] R. Falk. Error estimates for the numerical identification of a variable coefficient. *Mathematics of Computation*, 1983.
- [5] M. S. Gockenbach. Numerical analysis of elliptic inverse problems with interior data. *Journal of Physics: Conference Series*, 2008.
- [6] P. Lesaint. *Méthodes d'éléments finis pour les systèmes différentiels du premier ordre symétriques au sens de Friedrichs*. EPFL, 1995.
- [7] M. Verhaegen M. Gerard, B. De Schutter. A hybrid steepest descent method for constrained convex optimization. *Automatica*, 2009.
- [8] Jun Zou. Numerical methods for elliptic inverse problems. *International Journal of Computer Mathematics*, 1998.