REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
UNIVERSITE LARBI BEN M'HIDI OUM EL BOUAGHI
Faculté des Sciences Exactes et des Sciences de la nature et de la vie
Département de Mathématiques et Informatiques

N° d'ordre: ...
Série: ...

MÉMOIRE DE MAGISTER
de Mathématiques Appliquée

Intitulé:

PROBLÈMES INVERSES DE L'ÉQUATION DE TRANSFERT DE LA CHALEUR

Présenté par: DERBAZI Amar

Soutenu publiquement le 23/06/2011 à l'Université de Oum El Bouaghi devant le jury composé de

Mr MOHAMED Bouzit	M.C.A	Univ. Larbi.B.M'hidi O.E.B	Président
Mr MERZOUK Djebarni	M.C.A	Univ. Larbi.B.M'hidi O.E.B	Rapporteur
Mr ABDELKARIM Aliouche	M.C.A	Univ. Larbi.B.M'hidi O.E.B	Examinateur
Mr MOHAMED Dalah	M.C.A	Univ. de Canstantine	Examinateur

Table des matières

Notations **6**

Introduction Générale **7**

0.1 Introduction 7
0.2 Problèmes bien et mal posés 9
0.3 Plan du travail 10

1 Rappels et compléments d'analyse fonctionnelle **11**

1.1 Espaces de Hilbert 11
1.1.1 Définitions et exemples 11
1.1.2 Propriétés des espace de Hilbert 14
1.1.3 Bases hilbertiennes 15
1.2 Opérateurs linéaires dans les espaces de Hilbert 16
1.2.1 propriétés générales 16
1.2.2 Adjoint d'un opérateur 17
1.2.3 Opérateurs compacts 18

2 Problème inverse en Thermique **21**

3 Problèmes inverses linéaires **26**

3.1 Problèmes de moindres carrés linéaires 26
3.1.1 Propriétés mathématiques des problèmes de moindres carrés 26
3.2 La Méthode de Tikhonov 29

4 Problèmes inverses non-linéaires **38**

4.1 Les trois espaces fondamentaux 39

4.2 Formulation par moindres carrés 41

4.2.1 Difficultés des problèmes inverses 43

4.2.2 Optimisation, paramétrisation, discrétisation 44

4.3 Calcul du gradient – La méthode de l'état adjoint 45

4.3.1 Méthodes de calcul du gradient 46

4.3.2 Les différences finies 46

4.3.3 Les fonctions de sensibilité 47

4.3.4 La méthode de l'état adjoint 49

4.3.5 Calcul de l'état adjoint par le lagrangien 50

4.4 Exemples de calcul de gradient 52

4.4.1 Équation de la chaleur 52

Conclusion générale **58**

Bibliographie **59**

Remerciements

Au début et avant tous, je rends grâce à dieu tous puissant qui m'a aidé à terminer ce travail.

✓ Je tiens à souligner l'excellent remerciement à Mr DJBARNI Merzouk pour l'honneur qu'il me fait en proposant ce thème. M'encourager et me diriger pour sa bonne réalisation, et en encadrant ce mémoire et pour toute l'aide et les conseils et encore plus pour tous le temps qu'il m'a consacré pour me suivre pendant la rédaction de ce travail, malgré ces nombreuses obligations.

✓ Je tiens aussi à exprimer mes profonds remerciements à Mr BOUZIT Mohamed maître de conférence, au département de mathématiques, Université O.E.B, pour accepter de présider le jury de soutenance de mon travail.

✓ Je tiens à remercier Monsieur ALIOUCHE Abdelkarim et Monsieur DALAH Mohamed, d'avoir examiner mon travail et d'être membre de mon jury sans oublier tous mes enseignants sans exception, et se sont intéressés à mon travail et ont bien montré leur pleine disponibilité.

Enfin, je remercie vivement toute personne qui a de prés ou de loin contribué à l'élaboration de ce travail.

Résumé

Nous présentons dans ce mémoire les problèmes invreses linéaire et non linéaire, de l'un nous introduirons des méthodes de discrétisation, conduisant à des problèmes de moindres carrés. Nous aborderons l'étude des techniques pour les problèmes mal posés, tout particulièrement la méthode de régularisation de Tikhonov. Mais l'autre nous $^{2}lesproblmesnon-linaires, essentiellementlesproblmesd'estimationdeparamtresdanslesquationsd$

Mots clés: Problème inverse, Opérateur, Système évolutif, Système distribue, Conductivité thermique.

Abstract

We present in this dissertationthe problems inverses linear and nonlinear, we will introduce methods of discretization leading to problems of least squares. We will approach the study of the technique for the badly posed problems, particularly the method of regularisation of Tikhonov. The other puts (que voulez vous dire par : other puts???) we will tackle the non-linear problems, primarily the problems of estimate of parameters in the differential equations, and with the derivative partial, we will see how to pose the problem of identification in terms of minimization, which are principal difficulties which can be expected. And the significant technique of the state associated will approach to calculate the gradient of the functional claculuses which intervene in the problems of least squares. We will see examples of how to conclude this calculation in an effective way.

Key words: invert problem, Operator, evaluative System; distributed Système, thermal Conductivity.

Notations

E, F, M, U, D, Z	Espaces de Hilbert;
Ω	Domaine ouvert connexe de $\mathbb{R}^2$ ou $\mathbb{R}^3$;
E'	Le dual topologique de E;
$L(E, F)$	L'espace vectoriel des applications linéaires continues de E dans F;
A	Opérateur;
$D(A)$	Domaine de définition de A;
$\overline{D(A)}$	L'adhérence de $D(A)$;
$\operatorname{Im} A$	L'image de A;
$\ker A$	Le noyau de A;
A^{-1}	Opérateur inverse de A;
$\parallel A \parallel = sup_{\|x\|_E \leq 1 \|Ax\|_F}$	La norme de A;
A^*	Opérateur adjoint de A;
Γ	La frontière;
$H^1(0, 1)$	C'est un espace de Sobolev;

Introduction Générale

0.1 Introduction

D'après **J.B.Keller [14]**, deux problèmes sont dits inverses l'un de l'autre si la formulation de l'un met l'autre en cause. Cette définition comporte une part d'arbitraire, et fait jouer un rôle symétrique aux deux problèmes considérés. Une définition plus opérationnelle est qu'un problème inverse consiste à déterminer des causes connaissant des effets. Ainsi, ce problème est l'inverse de celui appel problème direct, consistant à déduire les effets, les causes étant connues. Cette seconde définition montre que nous sommes plus habitués à étudier des problèmes «directs». En effet, depuis Newton la notion de causalité est ancrée dans notre subconscient scientifique, et à un niveau plus prosaïque, nous avons appris à poser, puis résoudre des problèmes pour lesquels les causes sont données, et l'on cherche les effets. Cette définition montre aussi que les problème inverses risquent déposer des difficultés particulières. Nous verrons plus loin qu'il est possible de donner un contenu mathématique à la phrase «les mêmes causes produisent les mêmes effets», autrement dit, qu'il est raisonnable d'exiger que le problème direct soit «bien posé». Par contre, il est facile d'imaginer, et verrons de nombreux exemples, que les mêmes effets puissent provenir de causes différentes. Cette idée contient en germe la principale difficulté de l'étude des problème inverses: ils peuvent avoir plusieurs solutions, et il est nécessaire de disposer d'informations supplémentaires pour discriminer entre elles.

La prédiction de l'état futur d'un système physique, connaissant son état actuel, est l'exemple type du problème direct. On peut envisager divers problèmes inverses: par exemple, reconstituer l'état passé du système connaissant son état actuel (si ce système est irréversible), ou la détermination de paramètres du système, connaissant (une partie de) son évolution. Ce dernier problème est celui de l'identification de paramètres, qui sera notre principale préoccupation dans la suite.

Une difficulté pratique de l'étude des problèmes inverses est qu'elle demande souvent une bonne connaissance du problème direct, ce qui se traduit par le recours à une grande variété de notions tant physiques que mathématiques. Le succès dans la résolution d'un problème inverse repose en général sur des éléments spécifiques à ce problème. Il existe toutefois quelques techniques qui possèdent un domaine d'applicabilité étendu, et ce cours est une introduction aux principales d'entre elles : la régularisation des problèmes mal posés, et la méthode des moindres carrés.

La plus importante est la reformulation d'un problème inverse sous la forme de la minimisation d'une fonctionnelle d'erreur entre les mesures réelles et les « mesures synthétiques » (c'est-à-dire la solution du problème direct). Il sera commode de distinguer entre les problèmes linéaires et non linéaires. Précisons ici que la non linéarité dont il s'agit ici fait référence au problème inverse lui-même, et non pas au problème direct (en considérant les paramètres comme connus).

Dans le cas des problèmes linéaires, le recours à l'algèbre linéaire et à l'analyse fonctionnelle permet d'obtenir des résultats précis. L'outil fondamental est ici la décomposition en valeurs singulières de l'opérateur considéré. Nous étudierons en détail la méthode de régularisation, qui consiste à « modifier » légèrement le problème étudié en un autre qui possède de « meilleures » propriétés.

Les problèmes non linéaires sont plus difficiles, et il existe moins de résultats généraux. Un ingrédient technique essentiel (du point de vue numérique) est le calcul du gradient de la fonctionnelle à minimiser. Nous étudierons la méthode de l'état adjoint au chapitre 4. Elle permet ce calcul pour un coût qui est un (petit) multiple de celui de la résolution du problème direct.

0.2 Problèmes bien et mal posés

Dans un livre célèbre, Hadamard [15] a introduit dès 1923 la notion de problème bien posé. Il s'agit d'un problème dont :

– la solution existe ;

– elle est unique ;

– elle dépend continûment des données.

Bien entendu, ces notions doivent être précisées par le choix des espace (et des topologies) dans lesquels les données et la solution évoluent.

Dans ce même livre Hadamard laissait entendre (et c'était une opinion répandue jusqu'à récemment) que seul un problème bien posé pouvait modéliser correctement un phénomène physique. Après tout, ces trois conditions semblent très naturelles. En fait, nous verrons que les problèmes inverses ne vérifient souvent pas l'une ou l'autre de ces conditions, voire les trois ensembles. Après réflexion, cela n'est pas si surprenant :

– Un modèle physique étant fixé, les données expérimentales dont on dispose sont en général bruitées, et rien ne garantit que de telles données proviennent de ce modèle, même pour un autre jeu de paramètres.

– Si une solution existe, il est parfaitement concevable (et nous le verrons sur des exemples) que des paramètres différents conduisent aux mêmes observations.

– Le fait que la solution d'un problème inverse puisse ne pas exister n'est pas une difficulté sérieuse. Il est habituellement possible de rétablir l'existence en relaxant la notion de solution (procédé classique en mathématique).

– La non unicité est un problème plus sérieux. Si un problème a plusieurs solutions, il faut un moyen de choisir entre elles. Pour cela, il faut disposer d'informations supplémentaires (une information a priori).

– Le manque de continuité est sans doute le plus problématique, en particulier en vue d'une résolution approchée ou numérique. Cela veut dire qu'il ne sera pas possible (indépendamment de la méthode numérique) d'approcher de façon satisfaisante la solution du problème inverse, puisque les données disponibles seront bruitées donc proches, mais différentes, des données « réelles ».

Un problème qui n'est pas bien posé au sens de la définition ci-dessus est dit « mal posé » (ill-posed en anglais). Nous allons en voir un exemple qui, bien que très simple, illustre les difficultés que l'on peut rencontrer dans des situations plus générales.

0.3 Plan du travail

Dans le premier chapitre, on rappelle quelques résultats d'analyse fonctionnelle qui nous serons utile, ainsi que des compléments sur les opérateurs linéaires dans un appendice.

Dans le deuxième chapitre nous donnerons un exemple de problème inverse en thermique. Nous introduirons la notion fondamentale de problème mal posé, qui est caractéristique des problèmes inverses.

Au chapitre 3, nous introduirons une source importante de problèmes inverses linéaires: nous expliquerons en quoi ils sont mal posés. Enfin nous introduirons des méthodes de discrétisation, conduisant à des problèmes de moindres carrés. Nous aborderons l'étude des techniques pour les problèmes mal posés, tout particulièrement la méthode de régularisation de Tikhonov.

Au quatrième chapitre nous aborderons les problèmes non-linéaires, essentiellement les problèmes d'estimation de paramètres dans les équations différentielles, et aux dérivées partielles, nous verrons comment poser les problèmes d'identification en terme de minimisation, quelles sont les principales difficultés auxquelles on peut s'attendre. Et abordera la technique importante de l'état adjoint pour calculer le gradient des fonctionnelles qui interviennent dans les problèmes de moindres carrés. Nous verrons des exemples comment mener à bien ce calcul de façon efficace.

Chapitre 1

Rappels et compléments d'analyse fonctionnelle

Nous rappelons dans cet appendice les principaux résultats d'analyse fonctionelle dont nous aurons besoin, ainsi que des compléments concernant les opérateurs dans les espaces de Hilbert.

Nous ne donnerons que peu de démonstrations, renvoyant le lecteur aux (nombreux) ouvrages sur le sujet. Citons particulièrement [10]. Les ouvrages [1], [12], [19] et [4] contiennent également des introductions orientées vers les applications aux problèmes inverses.

Pour simplifier, nous ne considérerons que des espaces vectoriels sur $\mathbb{R}$. De plus nous ne considérerons que des espaces séparables, c'est-à-dire contenant une partie dénombrable dense.

1.1 Espaces de Hilbert

Commençons par rappeler quelques définitions.

1.1.1 Définitions et exemples

Définition 1. Soit E un espace vectoriel sur $\mathbb{R}$. Une norme sur E est une application de E dans $\mathbb{R}^+$, possédant les propriétés suivantes :

– $\forall x \in E, \|x\|_E \geq 0$ et $\|x\|_E = 0 \Rightarrow x = 0$.

– $\forall x \in E, \forall \alpha \in \mathbb{R}, \|\alpha x\|_E = |\alpha| \, \|x\|_E$

– $\forall \, (x, y) \in E^2, \|x + y\|_E \leq \|x\|_E + \|y\|_E$

Exemple 1. Dans le cas où E est de dimension n (nous l'identifiions alors à $\mathbb{R}^n$), les normes suivantes sont les plus utilisées :

- $\|x\|_1 = \sum_{i=1}^n |x_i|$;
- $\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{\frac{1}{2}}$
- $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$;

Définition 2. Soit E un espace vectoriel sur $\mathbb{R}$. Un produit scalaire sur E est une application de $E \times E$ dans $\mathbb{R}$, notée $(.,.)$, possédant les propriétés suivantes :

– $\forall (x; y; z) \in E^3; \forall (\alpha; \beta) \in R^2; (\alpha x + \beta y; z) = \alpha(x; z) + \beta(y; z)$;

– $\forall (x; y) \in E^2; (x; y) = (y; x)$;

– $\forall x \in E; (x; x) \geq 0$,

– $(x; x) = 0 \Longrightarrow x = 0$.

Un espace vectoriel muni d'un produit scalaire est appelé un espace préhilbertien.

Exemple 2. Sur $\mathbb{R}^n$, le produit scalaire euclidien usuel est :

$$(x; y) = \sum_{i=1}^n x_i y_i \tag{1.1}$$

Exemple 3. Soit Ω un ouvert de $\mathbb{R}^n$. L'espace vectoriel des fonctions de carré intégrable sur Ω est :

$$L^2(\Omega) = \left\{ f : \Omega \longrightarrow \mathbb{R}; \int_\Omega |f(x)|^2 \, dx < \infty \right\} \tag{1.2}$$

est un espace préhilbertien si on le munit du produit scalaire :

$$(f, g) = \int_\Omega f(x) \, g(x) \, dx \tag{1.3}$$

Un produit scalaire sur E définit une norme sur E par la formule suivante :

$$\|x\|_E = \sqrt{(x; x)}. \tag{1.4}$$

Parmi les trois normes de l'exemple 1, seule la seconde provient d'un produit scalaire (celui de l'exemple 2).

Définition 3. Un espace de Hilbert est un espace vectoriel muni d'un produit scalaire, et qui est complet pour la norme associée à ce produit scalaire.

Exemple 4. L'espace vectoriel $\mathbb{R}^n$, muni du produit scalaire euclidien usuel, est un espace de Hilbert.

Le résultat suivant est fondamental.

Proposition 1. L'espace vectoriel $L^2(\Omega)$, muni du produit scalaire défini en (1.3), est un espace de Hilbert.

Exemple .5 (Espace de Sobolev). Plaçons nous pour simplifier en dimension 1, sur l'intervalle $[0;1]$. L'espace de Sobolev d'ordre 1 est

$$H^1(0,1) = \left\{ u \in L^2(0,1), \exists u_1 \in L^2(0,1)\, tq : \forall \varphi \in C_c^1(0,1), \int_0^1 u(t)\,\varphi'(t)\,dt = -\int_0^1 u_1(t)\,\varphi(t)\,dt \right\} \tag{1.5}$$

(où $C_c^1(0;1)$ désigne l'espace des fonctions continûment dérivables, à support compact dans $[0;1]$).

Cette définition est équivalente à celle, plus usuelle utilisant la théorie des distributions. Pour $u \in H^1(0;1)$, on note $u' = u_1$.

On démontre (voir [10]) que $H^1(0;1)$ est un espace de Hilbert si on le munit du produit scalaire :

$$(u,v)_{H^1} = \int_0^1 u(t)\,v(t)\,dt + \int_0^1 u'(t)\,v'(t)\,dt \tag{1.6}$$

Dans les applications, on a souvent besoin du sous-espace de H^1 correspondant aux fonctions «nulles au bord » (au sens de la trace). Ce sous-espace est noté $H_0^1(0;1)$, et on peut le munir du produit scalaire suivant :

$$(u,v)_{H_0^1} = \int_0^1 u'(t)\,v'(t)\,dt \tag{1.7}$$

On démontre (c'est une conséquence de l'inégalité de Poincaré, voir toujours [10]) que la norme correspondante est équivalente à la norme induite par celle de l'espace H^1.

1.1.2 Propriétés des espace de Hilbert

Proposition 2. (Inégalité de Cauchy-Schwarz). Pour tous $(x;y) \in E^2$, on a l'inégalité :

$$|(x;y)|^2 \leq \|x\|_E^2 \|y\|_E^2 \tag{1.8}$$

L'égalité n'a lieu que si x et y sont proportionnels.

Proposition 3. (Identité du parallélogramme). Pour tous $(x;y) \in E^2$, on a l'identité :

$$\|x+y\|_E^2 + \|x-y\|_E^2 = 2\left(\|x\|_E^2 + \|y\|_E^2\right) \tag{1.9}$$

Le résultat suivant est l'un des plus importants de la théorie.

Théorème 1. (de projection). Soit F un sous-ensemble fermé, convexe de E, et $z \in E$ donné. Il existe un unique élément de $x_0 \in F$ tel que:

$$\|z - x_0\|_E = \inf |y - x|, x \in F \tag{1.10}$$

Le point x_0 est caractérisé par la condition:

$$x_0 \in F \quad \text{et} \quad (z - x_0, x - x_0) \leq 0, \forall x \in F \tag{1.11}$$

Le point x_0 mis en évidence au théorème 1 s'appelle la projection de z sur F. Dans le cas où F est un sous-espace vectoriel, on peut préciser ce résultat :

Corollaire 1. Soit F un sous-espace vectoriel fermé de E, et soit $z \in E$. La projection de z sur F est caracterisée par :

$$x_0 \in F \quad \text{et} \quad (z - x_0, x) = 0, \forall x \in F \tag{1.12}$$

Dans un espace de Hilbert, on dit que deux vecteurs sont orthogonaux si leur produit scalaire est nul. L'orthogonal d'un sous-espace vectoriel F est :

$$F^{\perp} = \{x \in E; (x;y) = 0; \forall y \in F\}$$

Une conséquence des résultats précédents est :

Corollaire 2. Soit F un sous-espace vectoriel de E (non nécessairement fermé). On a

$$F^{\perp} + \overline{F} = E \tag{1.13}$$

1.1.3 Bases hilbertiennes

Définition 4. Une base hilbertienne d'un espace de Hilbert E est une suite $(e_n)_{n\in\mathbb{N}^*}$ telle que:

– $\|e_n\| = 1; \forall n$; et $(e_n; e_m) = 0; \forall n \neq m$;

– l'espace vectoriel engendré par les (e_n) est dense dans E.

Précisons la deuxième condition : soit $F_n = vect\{e_1, ..., e_n\}$. Les sous-espaces F_n sont emboîtés ($F_n \subset F_m$ *pour* $n \leq m$), donc $F = \cup_{n\in\mathbb{N}^*} F_n$ est un sous-espace vectoriel. La seconde condition de la définition exprime que ce sous-espace est dense dans E, c'est-à-dire que tout élément de E peut être approché arbitrairement par un élément de F.

On démontre alors que tout espace vectoriel (séparable) admet une base hilbertienne. Étant donné une base hilbertienne $(e_n)_{n\in\mathbb{N}}$ de E, tout élément de E s'écrit :

$$x = \sum_{n=1}^{\infty} (x, e_n)\, e_n \tag{1.14}$$

avec (c'est l'égalité de Bessel–Parseval) :

$$\|x\|_E^2 = \sum_{n=1}^{\infty} |(x, e_n)|^2 \tag{1.15}$$

Un tel développement est unique, c'est-à-dire que si on a un développement

$$x = \sum_{n=1}^{\infty} x_n e_n$$

avec $\sum_{n=1}^{\infty} |x_n|^2 < \infty$, alors $x_n = (x; e_n)$. Notons toutefois qu'une base hilbertienne n'est pas une base algébrique, puisque ce développement n'est pas une combinaison linéaire finie.

On sait construire explicitement des bases hilbertiennes pour certains espaces L^2. Bien évidemment, une base orthogonale d'un espace vectoriel de dimension finie est une base hilbertienne.

Exemple 6. Les deux suites de fonctions

$$\left(\sqrt{\frac{2}{\pi}\sin nx}\right)_{n\geq 1}, \text{ ou } \left(\sqrt{\frac{2}{\pi}\cos nx}\right)_{n\geq 0}$$

sont des bases hilbertiennes de $L^2(0;\pi)$. Dans ce cas, le développement obtenu en (1.14) s'identifie à un développement en série de Fourier (après prolongement par imparité est périodicité).

1.2 Opérateurs linéaires dans les espaces de Hilbert

L'analyse fonctionnelle fait interagir la topologie et l'algèbre linéaire. Ainsi, sur un espace de Hilbert, il sera naturel d'étudier les applications qui respectent à la fois la structure d'espace vectoriel (les applications linéaires) et la structure hilbertienne (les applications continues).

1.2.1 propriétés générales

Définition 5. Un opérateur (linéaire, continu) A d'un espace de Hilbert E dans un espace de Hilbert F est une application linéaire continue de E dans F, c'est-à-dire qui vérifie :

– $\forall u \in E; Au \in F$;

– $\forall (u;v) \in E \times E; \forall (\alpha;\beta) \in R^2; A(\alpha u + \beta v) = \alpha Au + \beta Av$;

– $\exists M > 0; \forall u \in E; \|Au\|_F \leq M \|u\|_E$.

Le plus petit nombre M qui vérifie le $3^{éme}$ point ci-dessus s'appelle la norme de l'opérateur A :

$$\|A\| = \sup_{u \in E} \frac{\|Au\|_F}{\|u\|_E} \tag{1.16}$$

Il s'agit de la même notion que la norme matricielle introduite en (1.16)

Rappelons les deux espaces fondamentaux associés à un opérateur linéaire :

– Le noyau de A est le sous-espace de E : $KerA = \{u \in E; Au = 0\}$;

– L'image de A est le sous-espace de F : $ImA = \{v \in F; \exists u \in E; Au = v\}$;

Le théorème suivant est l'un des résultats fondamentaux de la théorie des opérateurs linéaires.

Théorème 2. **(de l'application ouverte).** Soit A un opérateur linéaire de E dans F. L'image par A d'un ouvert de E est un ouvert de F.

En particulier l'inverse d'un opérateur linéaire continu et bijectif est continu.

Dans le cas où l'espace d'arrivée F est le corps des scalaires, on parle de forme linéaire. L'espace vectoriel des formes linéaires continues s'appelle l'espace dual de E, et se note E'. Dans le cas d'un espace de Hilbert, le dual s'identifie de façon canonique à l'espace lui-même.

Théorème 3. (de Riesz). Soit L une forme linéaire continue sur E. Il existe un unique vecteur $x_0 \in E$ tel que

$$L(x) = (x_0; x); \forall x \in E \tag{1.17}$$

1.2.2 Adjoint d'un opérateur

Théorème 4. Soit A un opérateur linéaire continu de E dans F. Il existe un unique opérateur de F dans E, noté A^*, tel que :

$$\forall u \in E; \forall v \in F; (Au; v) = (u; A^*v) \tag{1.18}$$

Cet opérateur est appelé l'adjoint de A. Il vérifie de plus : $(A^*)^* = A \;\; et \; \|A^*\| = \|A\|$.

La proposition suivante rassemble quelques propriétés simples de l'adjoint.

Proposition 4. Soient A et B sont deux opérateurs linéaires, α et β deux scalaires.

– Linéarité : $(\alpha A + \beta B)^* = \alpha A^* + \beta B^*$.

– Composition : $(AB)^* = B^* A^*$.

Il existe des relations remarquables entre le noyau et l'image d'un opérateur et ceux de son adjoint.

Proposition 5. On a les relations suivantes (ou $\overline{X}$ indique l'adhérence de l'ensemble X) :

– $Ker A^* = (Im A)^{\perp}$;

– $(Ker A)^{\perp} = \overline{Im A^*}$;

Définition 6. Un opérateur dans E est dit auto-adjoint si et seulement si :

$$\forall (u; v) \in E \times E; (Au; v) = (u; Av) \tag{1.19}$$

Remarque 2. En dimension finie, les opérateurs auto-adjoints sont ceux qui ont une matrice symétrique.

1.2.3 Opérateurs compacts

Définition7. Soit $A \in L(E;F)$. On dit que A est un opérateur compact si et seulement si l'image de toute partie bornée de E est relativement compacte dans F.

Remarque 3. Cette condition veut dire que si $B \subset E$ est borné, $\overline{A(B)}$ (l'adhérence de $A(B)$) est compacte dans F.

Exemple 7. – Tout opérateur de rang fini est compact.

– L'injection canonique de $H_0^1(0;1) \longrightarrow L^2(0;1)$ est compacte (plus généralement, on peut remplacer $]0;1[$ par un ouvert borné régulier de $\mathbb{R}^n$).

Commençons par quelques propriétés de base des opérateurs compacts.

Proposition 6. Soit $E; F; G$ trois espaces de Hilbert.

i) L'ensemble des opérateurs compacts de E dans F est un sous-espace vectoriel de $\mathcal{L}(E;F)$.

ii) Si $A_1 \in \mathcal{L}(E;F)$ est compact et $A_2 \in \mathcal{L}(F;G)$, alors $A_2A_1 \in \mathcal{L}(E;G)$ est compact.

iii) Si $A_1 \in \mathcal{L}(E;F)$ et $A_2 \in \mathcal{L}(F;G)$ est compact, alors $A_2A_1 \in \mathcal{L}(E;G)$ est compact.

iv) Si $A \in \mathcal{L}(E;F)$ est compact, $A^* \in \mathcal{L}(F;E)$ est aussi compact.

v) Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'opérateurs compacts de E dans F. Si A_n converge vers A dans $\mathcal{L}(E;F)$, c'est-à-dire si

$$\|A_n - A\| = \sup_{u\neq 0} \frac{\|A_n u - Au\|_F}{\|u\|_E} \underset{n\to+\infty}{\longrightarrow} 0$$

alors A est compact.

En d'autres termes, les opérateurs compacts forment un sous-espace vectoriel fermé de $\mathcal{L}(E;F)$.

Le théorème suivant fournit (dans le cas des espaces de Hilbert) une caractérisation à la fois utile et plus proche de l'intuition.

Théorème 5. Un opérateur de E dans F est compact si et seulement si il est limite d'une suite d'opérateurs de rang fini.

Ce théorème signifie que les opérateurs compacts sont ceux qui « ressemblent » le plus aux opérateurs de dimension finie usuels. Signalons que ce résultat n'est plus valable si E et F sont des espaces de Banach. Il existe par contre une différence, qui sera fondamentale pour l'étude des problèmes mal

posés.

Proposition 7. Si E n'est pas de dimension finie, alors l'identité $E \longrightarrow E$ n'est jamais compacte.

Remarque 4. Nous ne donnons pas la démonstration (voir par exemple [10]), qui est une conséquence de ce que la boule unité d'un espace vectoriel normé de dimension finie n'est pas compacte.

Corollaire 3. Soit A un opérateur compact de E dans F, où E et F sont deux espaces de Hilbert qui ne sont pas de dimension finie. Alors A n'est jamais inversible dans $\mathcal{L}(E; F)$.

Remarque 5. Dans le corollaire précédent, l'inverse (algébrique) de A peut ou non exister (A peut ou non être injectif), mais s'il existe, il ne sera pas continu. Ceci est lié au caractère non fermé de l'image de A.

Pour conclure, citons sans démonstration une version « abstraite » de l'alternative de Fredholm. Ce résultat concerne les équations du type

$$(I - A)\, u = f \tag{1.20}$$

où A es tun opérateur compact dans E.

Théorème 6. Soit A un opérateur compact dans un espace de Hilbert E.

– Le noyau $Ker(I - A)$ est de dimension finie, et l'image $Im(I - A)$ est fermée dans E.

– Si $I - A$ est injectif, il est aussi surjectif, et alors l'inverse $(I - A)^{-1}$ est continue, Si $I - A$ n'est pas injectif, l'équation (1.20) a une solution si et seulement si $f \in Ker(I - A)^{\perp}$.

Remarque 6. – Ce théorème étend aux opérateurs de la forme « I +compact» le résultat analogue bien connu en dimension finie (pour tous les systèmes d'équations linéaires).

– Le second point du théorème veut dire que l'équation (1.20) n'a de solution que si le second membre satisfait des conditions d'orthogonalité au noyau (de dimension finie).

Chapitre 2

Problème inverse en Thermique

Nous présentons dans ce chapitre quelques exemples « concrets » de problèmes inverses en Thermique.

Pour déterminer la répartition de la température dans un matériau inhomogène occupant un domaine (ouvert connexe) Ω de $\mathbb{R}^3$ on écrit tout d'abord la conservation de l'énergie :

$$\rho c \frac{\partial T}{\partial t} + div(\overrightarrow{q}) = f(x;y;z) \quad dans \quad \Omega \tag{2.1}$$

où T est la température, ρ la densité du fluide, c la chaleur spécifique, $\overrightarrow{q}$ représente un flux de chaleur et f une source volumique.

La loi de Fourier relie ensuite le flux de chaleur au gradient de température :

$$\overrightarrow{q} = -K \operatorname{grad} T \tag{2.2}$$

où K est la conductivité thermique (qui peut être un tenseur, et dépend de la position).

En éliminant $\overrightarrow{q}$, on obtient l'équation de la chaleur en milieu hétérogène :

$$\rho c \frac{\partial T}{\partial t} - div(K \operatorname{grad} T) = f \quad dans \quad \Omega \tag{2.3}$$

Cette équation doit être complétée par des conditions aux limites sur le bord de l'ouvert Ω, et une condition initiale.

Le problème direct est de déterminer T connaissant les coefficients physiques ρ, c et K, ainsi que la source de chaleur f. Ce problème est bien connu, tant du point de vue théorique (existence et unicité de la solution) que du point de vue numérique. Plusieurs problèmes inverses peuvent être posés :

– étant donné une mesure de la température à un instant t_f, déterminer la température initiale. Nous l'aborderons à l'exemple1;

– étant donné une mesure (partielle) de la température, déterminer certains des coefficients de l'équation.

Notons que le premier de ces problèmes est linéaire, alors que le second est non-linéaire : en effet l'application $(\rho; c; K) \to T$ est non-linéaire.

Exemple1 (Équation de la chaleur rétrograde).

Nous prenons le cas idéal d'un matériau infini et homogène (en une dimension d'espace pour simplifier). La température est solution de l'équation de la chaleur :

$$\frac{\partial T}{\partial t} + \frac{\partial^2 T}{\partial x^2} = 0 \tag{2.4}$$

(il n'y a pas de source). On suppose connue la température à un certain instant t_f, soit $T_f(x) = T(x; t_f)$, et l'on cherche à retrouver la température initiale $T_0(x) = T(x; 0)$.

Le problème de déterminer T_f connaissant T_0 est le problème de Cauchy pour l'équation de la chaleur. Il a une solution unique, qui dépend continûment de la donnée initiale. Comme nous allons le voir, il n'en est rien pour le problème inverse que nous considérons ici. Physiquement, cela est du au caractère irréversible de la diffusion thermique. Il est bien connu que la température a tendance à s'homogénéiser au cours du temps, et cela entraîne qu'il n'est pas possible de revenir en arrière, c'est-à-dire de retrouver l'état antérieur qui peut être plus hétérogène que l'état actuel.

Grâce à la situation très simplifiée que nous avons choisie, nous pouvons calculer à la main la solution de l'équation de la chaleur (2.4). En prenant la transformée de Fourier en espace de l'équation (2.4) (nous notons $\check{T}(k; t)$ la transformée de Fourier de $T(x; t)$ en gardant t fixé), nous obtenons une équation différentielle ordinaire (où cette fois c'est k qui joue le rôle d'un paramètre) dont la solution est

$$\check{T}_f\left(k\right) = e^{-|k|^2 t_f} \check{T}_0\left(k\right) \tag{2.5}$$

En prenant la transformée de Fourier inverse, nous voyons que la solution à l'instant T_f est reliée à la condition initiale par une convolution avec la solution élémentaire de la chaleur:

$$T_f(x) = \frac{1}{2\sqrt{\pi t_f}} \int_{-\infty}^{+\infty} e^{-(x-y)^2/4t_f} T_0(y)\, dy \tag{2.6}$$

Il est bien connu [13] que, pour toute fonction T_0 « raisonnable » (continue, bornée), la fonction T_f est indéfiniment dérivable, ce qui traduit mathématiquement l'irréversibilité mentionnée ci-dessus.

En restant dans le domaine de Fourier, nous pouvons inverser ponctuellement l'équation (2.5), mais la fonction

$$k \to e^{|k|^2 t} \check{T}_f(k)$$

ne sera dans $L^2(\mathbb{R})$ que pour des fonctions T_f décroissant très rapidement à l'infini, ce qui est une restriction très sévère. Une température mesurée expérimentalement a peu de chances de la satisfaire, et c'est ce qui entraîne l'instabilité du problème inverse.

Passons maintenant à des exemples d'estimation de paramètres.

Exemple 2. (Identification d'un coefficient dans un modèle stationnaire).

Pour simplifier, on ne considère que le régime permanent, et on suppose que le bord du domaine est maintenue à une température de 0. L'équation de la chaleur (2.3), et la condition au bord, donnent alors:

$$\begin{cases} -\operatorname{div}(K \operatorname{grad} T) &= f(x,y,z) \quad \text{dans } \Omega \\ T &= 0 \qquad\qquad\;\; \text{sur } \partial\Omega \end{cases} \tag{2.7}$$

Le problème direct suppose connues la conductivité thermique K et la source de chaleur f , et cherche à déterminer la répartition de la température T en tout point du matériau. Il s'agit là du prototype des équations elliptiques du second ordre, et sa résolution est des plus classiques : sous des hypothèses raisonnables sur K ($K \in L^\infty(\Omega)$ avec $0 < K < \infty$) et sur f ($f \in L^2(\Omega)$, il admet une solution unique, d'après le théorème de Lax-Milgram. De plus, un calcul numérique avec une méthode d'éléments finis est tout-à-fait standard (voir [[20], chap. V]).

Il en va différemment du problème inverse. Pour pouvoir le spécifier, il est tout d'abord nécessaire de préciser de quelle observation l'on dispose. Cela dépend bien évidemment du dispositif expérimental utilisé, mais en tout état de cause, il ne sera généralement pas réaliste de supposer que l'on connait la température en tout point. Dans notre cas, ces observations pourraient être, par exemple, des mesures de la température à l'intérieur du matériau, ou bien des mesures du flux de chaleur $-K\frac{\partial T}{\partial \eta}$ sur le bord du domaine (on parle dans ce cas d'observation frontière). Le problème inverse est alors de chercher la (ou une) fonction de conductivité, telle qu'il existe une fonction T solution de (2.7) qui coïncide avec les observations. On voit immédiatement plusieurs difficultés possibles :

– Tout d'abord celle d'obtenir les observations ! Une expérience n'est jamais facile à réaliser. Dans notre exemple, il n'est pas réaliste de supposer que l'on puisse mesurer la température en tout point du domaine (penser à une pièce dans un appartement).

– Mais alors on risque de ne pas disposer de suffisamment d'observations par rapport au nombre de paramètres que l'on cherche. Ici, si l'on dispose d'une observation frontière, c'est-à-dire d'une fonction de deux variables, il sera difficile de retrouver une fonction de trois variables, indépendamment de la méthode utilisée.

– En particulier, on voit tout de suite que si la température est constante dans un sous-domaine de Ω, la conductivité n'y est pas déterminée. Il faudrait donc disposer d'informations supplémentaires permettant de combler ce manque de mesure.

– Enfin, toute mesure est entachée d'erreur, et d'ailleurs le modèle mathématique (2.7) ne reflète pas exactement la réalité. Ainsi, il n'y a en fait aucune raison pour que le problème inverse possède une solution.

Ce problème difficile a fait l'objet de nombreuses études, tant théoriques que numériques (voir [21] pour une introduction). Signalons qu'il intervient dans d'autres domaines d'application (médical, prospection géophysique par des méthodes électriques ou magnétiques, ...)

Exemple 3 (Identification en une dimension d'espace).

Nous pouvons mieux comprendre cet exemple en le réduisant à une seule dimension. Il s'agit de déterminer la fonction $K(x)$ à partir de l'équation

$$-\left(K(x)T'(x)\right)' = f(x) \quad pour \quad x \in]0;1[\tag{2.8}$$

(avec des conditions aux limites convenables), et de la connaissance de T. Pour simplifier, nous supposerons que l'on connaît T en tout point de l'intervalle $]0;1[$. Sous l'hypothèse que $T^{'}$ ne s'annule en aucun point de l'intervalle (qui n'est pas nécessairement vérifiée), on peut intégrer l'équation (2.8), pour obtenir:

$$K(x) = \frac{1}{T'(x)}\left(K(0)\,T'(0) - \int_0^x f(t)dt\right) \tag{2.9}$$

Cette équation montre, sous l'hypothèse que nous avons faite ci-dessus, que K est uniquement déterminée dès que sa valeur en un point est connue. Par contre, la formule donnant K fait intervenir la dérivée de l'observation T, cette opération n'est pas continue. Ainsi l'application (non-linéaire) $T \to K$ ne le sera pas non plus. Une petite erreur sur la mesure de T pourra se traduire par une erreur arbitrairement grande sur la conductivité K.

De plus, une seconde instabilité provient de la division par $T'(x)$, ce qui est un effet essentiellement non-linéaire. Si $T'(x)$ est petit, la division par $T'(x)$ risque d'amplifier les erreurs déjà présentes dans la mesure de T. Ceci est bien sur lié au fait que si T' s'annule, K n'est pas déterminé du tout.

Pour plus de renseignements sur cet exemple, on pourra consulter l'article [24] ou le livre [7].

Chapitre 3

Problèmes inverses linéaires

3.1 Problèmes de moindres carrés linéaires

Nous étudions dans cette9 section les principales propriétés des problèmes inverses linéaires. Nous nous placerons dans le cadre des espaces de Hilbert, pour que les résultats s'appliquent (par exemple) aux équations intégrales de première espèce, mais nous indiquerons les simplifications qui interviennent en dimension finie.

Nous désignerons par A un opérateur linéaire continu d'un espace de Hilbert E dans un espace de Hilbert F : $A \in \mathcal{L}(E; F)$.

3.1.1 Propriétés mathématiques des problèmes de moindres carrés

Étant donné $z \in F$, nous cherchons $x \in E$ solution de :

$$Ax = z \tag{3.1}$$

Revenons dans ce cas particulier sur la discussion du chapitre 1 concernant les problèmes bien et mal posés

– l'opérateur A peut ne pas être surjectif ;

– il peut ne pas être injectif ;

– si un inverse existe, il peut ne pas être continu.

Comme nous l'avons déjà dit, la première difficulté n'est pas sérieuse : il « suffit » de se restreindre à ImA. La seconde est plus gênante : il faut pouvoir sélectionner, parmi plusieurs solutions, celle qui est appropriée au problème. La dernière, fondamentale pour les applications, est lié au caractère fermé ou non de ImA :

Théorème1. Soit $A \in L(E; F)$; E et F deux espaces de Hilbert. Supposons que A soit injectif, et notons $A^{-1} : \mathrm{Im}\, A \to E$ l'inverse de A. On a :

$$ImA \text{ fermé} \Longleftrightarrow A^{-1} \text{ est continu}$$

Preuve. voir [16]

Pour les situations que nous considérons dans ce cours, l'opérateur pourra ou non être injectif, mais la situation générale sera que ImA n'est pas fermée (le corollaire 3 [chap.1] montre que si A est compact, A n'a pas d'inverse continu, et dans ce cas $\mathrm{Im}\, A$ ne sera pas fermé).

Le problème (3.1) n'a de solution que pour les seconds membres dans l'image de A. Comme nous venons de le voir, cette condition peut-être trop restrictive. Nous cherchons donc une autre formulation du problème original, qui permette d'étendre la notion de solution à un sous-espace plus grand.

Nous proposons une formulation comme un problème de moindres carrés : nous remplaçons donc (3.1) par :

$$\min_{x \in E} \frac{1}{2} \|Ax - z\|_F^2 \tag{3.2}$$

Nous allons voir que ce problème est équivalent à une équation linéaire, mais pour un opérateur différent de A.

Théorème 2. Soit $A \in L(E; F)$; $E; F$ deux espaces de Hilbert, et soit $z \in F$. Un élément $x \in E$ est une solution de (3.2) si et seulement si

$$A^* Ax = A^* z \tag{3.3}$$

Preuve.voir [16]

Remarque 1. L'équation normale (3.3) se réécrit : $A^*(Ax - z) = 0$.

ce qui exprime simplement que le résidu $Ax - z$ est dans le noyau de A^*, c'est-à-dire orthogonal à (la fermeture de) l'image de A (voir proposition 5 chap. 1). Ceci conduit à l'illustration géométrique bien connue :

La solution du problème de moindres carrés est telle que Ax est la projection de z sur l'image de A.

Notons que nous n'avons pour l'instant évoqué ni l'existence, ni l'unicité, pour les solutions de (3.2) (ou de (3.3)). Nous avons simplement montré l'équivalence des deux problèmes. L'unicité est évidemment liée à l'injectivité de A, comme le précise le résultat suivant.

Lemme 1. La solution du problème (3.2) est unique si, et seulement si, l'opérateur A est injectif

Preuve.voir [16]

En ce qui concerne l'existence, on a le résultat suivant :

Proposition 1. i) L'équation (3.3) admet une solution si et seulement si $z \in ImA \oplus ImA^{\perp}$.

ii) Si $z \in ImA \oplus ImA^{\perp}$, l'ensemble S des solutions de (3.3) est un convexe fermé non vide de E.

Preuve.voir [16]

Lemme 2. Le sous-espace $ImA \oplus ImA^{\perp}$ est dense dans F.

Preuve.voir [16]

On a donc bien réussi a étendre la notion de solution à un sous-espace dense dans F, ce qui est « presqu'aussi bien » que de l'étendre à F tout entier.

Remarque 2. Notons que, dans le cas général, la condition $z \in ImA \oplus ImA^{\perp}$ est non-triviale. En effet, le théorème de projection dit seulement $F = \overline{ImA} \oplus ImA^{\perp}$ ce qui est différent si ImA n'est pas fermée.

On retrouve encore ici l'importance de cette condition. Dans ce cas, et seulement dans ce cas, le problème (3.3), et donc (3.2), admet toujours une solution.

En dimension finie, cette condition est bien entendu automatiquement vérifiée (tout sous-espace est fermé). Nous retrouverons ce résultat à la proposition 3.2.

Corollaire1. Si $z \in ImA \oplus ImA^{\perp}$, le problème (3.2) admet une unique solution de norme minimale.

Preuve.voir [16]

3.2 La Méthode de Tikhonov

Nous abordons dans ce section les méthodes de régularisation pour les problèmes inverses linéaires. Régulariser un problème mal posé, c'est le « remplacer » par un autre, bien posé, de sorte que l'erreur commise soit compensée par le gain de stabilité. Bien entendu, ceci demande a être quantifié, ce que nous ferons. Ce chapitre présente une introduction aux méthodes de régularisation les plus courantes, à savoir la méthode de Tikhonov. Bien entendu, il n'est pas possible de reconstituer une information manquante, et la régularisation va conduire à une perte de précision sur la solution, que nous essayerons de quantifier en fonction de l'erreur sur les données. Nous verrons qu'il est possible d'analyser la méthode de Tikhonov dans un cadre (variationnel) très général.

La principale difficulté dans l'application d'une méthode de régularisation à un problème particulier est la détermination du paramètre de régularisation lui-même. Nous dirons quelques mots sur des solutions possibles.

Dans ce section, A désigne un opérateur linéaire continu d'un espace de Hilbert E dans un espace de Hilbert F, et nous supposerons que ce problème est mal posé, c'est-à-dire que A n'est pas inversible dans $\mathcal{L}(E;F)$. Comme nous l'avons vu au chapitre **4**, cela peut être du à ce que A n'est pas injectif, mais le cas le plus intéressant est celui ou l'image de A n'est pas fermée, ce qui sera toujours le cas si A est compact. Dans ce cas A peut où non être bijectif, mais, s'il existe, l'inverse ne sera pas continu.

Pour résoudre l'instabilité évoquée au paragraphe précédent, nous allons introduire une information a priori.

Nous nous donnons donc un estimé à priori $x_0 \in E$. Pour un nombre $\epsilon > 0$ (le coefficient de régularisation), nous remplaçons (3.2) par le problème régularisé :

$$\min_{x\in E}\left\{\frac{1}{2}\left\|Ax-z\right\|_F^2+\frac{\epsilon^2}{2}\left\|x-x_0\right\|_E^2\right\} \tag{3.4}$$

Nous allons voir que ce problème admet une solution unique, qui dépend continûment de z, et qui converge, lorsque $\epsilon \to 0$, vers la solution la plus proche de x_0 de (3.2). Évidemment, si ϵ est choisi trop petit, (3.4) sera proche de (3.2), donc mal posé, alors que si ϵ est trop grand (3.4) ne sert qu'à forcer x à être proche de x_0. Le choix « optimal » de ϵ est donc délicat.

Commençons par le résultat suivant, qui montre que (3.4) est encore un problème de moindre carrés :

Lemme 1. Posons

$$\widetilde{A} = \begin{pmatrix} A \\ \epsilon I \end{pmatrix} ; \widetilde{z} = \begin{pmatrix} z \\ \epsilon x_0 \end{pmatrix} ; \widetilde{A} \in \mathcal{L}(E, F \times E)$$

Alors (3.4) est équivalent à

$$\min_{x \in E} \frac{1}{2} \left\| \widetilde{A}x - \widetilde{z} \right\|^2_{F \times E} \tag{3.5}$$

Preuve. En effet, on calcule simplement :

$$\widetilde{A}x - \widetilde{z} = \begin{pmatrix} Ax - z \\ \epsilon(x - x_0) \end{pmatrix}$$

Il suffit ensuite de calculer le carré de la norme.

Ce résultat va nous permettre facilement de montrer que (3.4) possède une solution unique, mais il sert également de base aux méthodes numériques pour résoudre (3.4) (voir [[7], [17]]). Une conséquence du lemme (1) et du théorème (2) est la formulation suivante de (3.4).

Proposition 1. Le problème (3.4) est équivalent à :

$$(A^*A + \epsilon^2 I)x = A^*z + \epsilon^2 x_0 \tag{3.6}$$

Ce problème (et donc (3.4)) admet une solution unique, qui dépend continûment de $\widetilde{z}$.

Preuve. En effet, (3.6) n'est autre que l'équation normale pour (3.4). On l'obtient en remarquant que :

$$\widetilde{A}^* = (A^*; \epsilon I)$$

En ce qui concerne l'existence et l'unicité de solutions à (3.6), notons que

$$((A^*A + \epsilon^2 I)x; x) = \|Ax\|^2 + \epsilon^2 \|x\|^2 \geq \epsilon^2 \|x\|^2$$

On applique alors le lemme de Lax-Milgram. D'après le théorème de l'application ouverte 2, l'opérateur $A^*A + \epsilon^2 I$, continu et bijectif, a un inverse continu. En fait, nous pouvons obtenir un estimation explicite en prenant le produit scalaire de l'équation (3.6) avec x, il vient :

$$\|Ax\|^2 + \epsilon^2 \|x\|^2 \leq \|A^*\widetilde{z}\| \|x\| + \epsilon^2 \|x\| \|x_0\|$$

c'est-à-dire :

$$\|x_\epsilon\| \leq \frac{1}{\epsilon^2} \|A^*\widetilde{z}\| + \|x_0\| \tag{3.7}$$

Remarque 1. L'estimation (3.7) « explose » quand $\epsilon \to 0$. Ceci est normal, puisque la solution x de (3.2) ne dépend pas de façon continue de z.

Nous voulons maintenant aborder la question de savoir dans quelle mesure la méthode de Tikhonov a bien régularisé le problème de départ. Pour cela, il sera naturel de se placer dans le cas d'une donnée bruitée, puisqu'alors nous n'avons pas d'estimation sur l'erreur commise sur la solution. Nous allons montrer que la méthode de Tikhonov donne une telle estimation, même si elle est « non optimale», c'est-à-dire d'une ordre plus faible que l'erreur sur la donnée.

Nous supposons connue une observation « idéale » $z \in ImA$, et également une suite de mesures bruitées $z_n \in F; z_n \notin ImA$, avec $\delta_n = \|z_n - z\|_F \underset{n\to\infty}{\longrightarrow} 0$. Le nombre $\|z_n - z\|_F / \|z\|_F$ est le rapport signal sur bruit. L'hypothèse sous-jacente dans ce paragraphe est que ce rapport tend vers 0, c'est-à-dire que l'on est capable de le réduire arbitrairement, ce qui est évidemment irréaliste.

Considérons tout d'abord la suite de problèmes :

Trouver $x_\epsilon^n \in E$ réalisant le minimum de

$$\frac{1}{2} \|Ax - z_n\|_F^2 + \epsilon^2 \|x - x_0\|_E^2 \tag{3.8}$$

(x_ϵ^n existe et est unique d'après la proposition 1). Remarquer que nous ne cherchons pas, pour l'instant, à adapter le paramètre de régularisation au niveau de bruit.

Pour comprendre comment « fonctionne » la méthode de régularisation, cherchons à estimer l'erreur entre la solution du problème bruité et la solution exacte. Pour simplifier, nous ferons l'hypothèse (de régularité) que $x \in ImA^*$. Le cas général est traité par Baumeister [12].

Proposition 2. Supposons que $x \in ImA^*$, soit $x = A^*w$. Alors,

$$\forall n, \|x_\epsilon^n - x\|_E \leq \|A^*\| \frac{\delta_n}{\epsilon^2} + \frac{\epsilon}{\sqrt{2}} \|w\|_F$$

Preuve. Introduisons la quantité intermédiaire x_ϵ solution du problème

$$A^*Ax_\epsilon + \epsilon^2 x_\epsilon = A^*z + \epsilon^2 x_0 \tag{3.9}$$

L'inégalité triangulaire donne :

$$\|x_n - x\| \leq \|x_n - x_\epsilon\| + \|x_\epsilon - x\| \tag{3.10}$$

Nous allons estimer séparément chaque terme du second membre ci-dessus. La première partie correspond à l'erreur sur les données, et est amplifié par le caractère mal posé du problème sous-jacent, alors que le second est du à l'approximation de la solution exacte, et tend vers 0 avec ϵ.

En soustrayant (3.9) de l'équation normale associée à (3.8), il vient :

$$\epsilon^2 \left(x_\epsilon^n - x_\epsilon\right) + A^*A\left(x_\epsilon^n - x_\epsilon\right) = A^*\left(z_n - z\right)$$

puis en prenant le produit scalaire avec $x_n^\epsilon - x_\epsilon$

$$\epsilon^2 \left\|x_\epsilon^n - x_\epsilon\right\|^2 + \left\|A\left(x_\epsilon^n - x_\epsilon\right)\right\|^2 \leq \|A^*\| \|z_n - z\| \|x_\epsilon^n - x_\epsilon\|$$

et en particulier

$$\epsilon^2 \|x_\epsilon^n - x_\epsilon\| \leq \|A^*\| \delta_n \tag{3.11}$$

c'est une majoration de la première partie de (3..10).

Pour la second partie, écrivons :

$$\begin{aligned}
\|Ax_\epsilon - Ax\|^2 &= (Ax_\epsilon - Ax, Ax_\epsilon - Ax) \\
&= (A^* Ax_\epsilon - A^* Ax, x_\epsilon - x) && \text{par définition de } A^* \\
&= (-\epsilon^2 x_\epsilon + \epsilon^2 x - \epsilon^2 x, x_\epsilon - x) && \text{en utilisant les équations normales} \\
&= -\epsilon^2 \|x_\epsilon - x\|^2 + \epsilon^2 (x, x_\epsilon - x) \\
&= -\epsilon^2 \|x_\epsilon - x\|^2 + \epsilon^2 (w, A(x_\epsilon - x)) && \text{d'après la définition de } w \\
&\leq -\epsilon^2 \|x_\epsilon - x\|^2 + \|\epsilon^2\| \|w\| \|A(x_\epsilon - x)\| && \text{par Cauchy-Schwarz} \\
&\leq -\epsilon^2 \|x_\epsilon - x\|^2 + \tfrac{\epsilon^4}{2} \|w\|^2 + \tfrac{1}{2} \|A(x_\epsilon - x)\|^2 && \text{par l'inégalité élémentaire } ab \leq \tfrac{a^2+b^2}{2}
\end{aligned}$$

d'où nous tirons:

$$\frac{1}{2} \|A(x_\epsilon - x)\|^2 + \epsilon^2 \|x_\epsilon - x\|^2 \leq \frac{\epsilon^4}{2} \|w\|^2$$

et en particulier

$$\|x_\epsilon - x\| \leq \frac{\epsilon}{\sqrt{2}} \|w\| \tag{3.12}$$

En regroupant (3.11) et (3.12), nous obtenons l'estimation du théorème.

Ce résultat montre que, comme nous l'avons signalé, l'erreur se compose de deux termes :

– Un premier terme du aux erreurs sur les données, multiplié par un « nombre de conditionnement», qui tend vers l'infini lorsque $\epsilon \to 0$.

– Un second terme du à l'approximation de la solution exacte, et qui tend vers 0 avec ϵ ;

Nous voyons donc bien la nécessité d'adapter le paramètre de régularisation au niveau de bruit présent dans les données. Une telle stratégie de régularisation peut se concevoir de deux façons :

– Si l'on possède une estimation du niveau de bruit, on peut en déduire comment il faut choisir ϵ pour obtenir la convergence de x_ϵ^n vers x. C'est ce que nous faisons au théorème suivant. Une telle stratégie s'appelle une stratégie de régularisation à priori. Elle suppose que l'on sache estimer le bruit présent dans les données, ce qui n'est pas forcément possible ;

– L'autre stratégie, appelée a posteriori, consiste à estimer au cours du calcul la valeur convenable du paramètre, en utilisant uniquement les données disponibles. De telles stratégies existent (voir [[7], [17]]).

Dans la proposition 2, nous avons laissé ϵ tendre vers 0 indépendamment de δ, et nous avons vu qu'une telle stratégie (ou plutôt absence de stratégie) ne permettait pas la convergence de la solution régularisée vers la « vraie » solution. Nous allons donc adapter le paramètre de régularisation au niveau de bruit. Pour cela, nous modifions le problème (3.8) en

Trouver $x_n \in E$ réalisant le minimum de

$$\frac{1}{2}\|Ax - z_n\|_F^2 + \epsilon_n^2 \|x - x_0\|_F^2 \tag{3.13}$$

où nous cherchons comment choisir la suite ϵ_n est pour assurer la convergence de x_n vers une solution de (3.2). Il se trouve que cette solution sera toujours la solution la plus proche de x_0. Ce résultat la conséquence du lemme suivant.

Lemme 2. Pour $x \in E$, notons la décomposition (unique) :

$$x = x^K + x^I; x^K \in KerA; x^I \in KerA^{\perp} = \overline{ImA^*}$$

La solution du problème (3.13) vérifie :

$$\forall n; x_n^K = x_0^K$$

Preuve. Projetons orthogonalement l'équation (3.6) sur$KerA$:

– $A^*Ax_n \in ImA^* \subset (KerA)^{\perp}$

– $A^*z_n \in ImA^* \subset (KerA)^{\perp}$

d'où $\epsilon^2 x_n^K = \epsilon^2 x_0^K$

Nous pouvons maintenant énoncer le théorème de convergence, dont la démonstration élémentaire est empruntée à G. Chavent [6].

Théorème 1. Soit $z \in ImA$. Supposons $\delta_n = \|z_n - z\| \underset{n\to\infty}{\to} 0; \epsilon_n \underset{n\to\infty}{\to} 0$. Alors:

i) $\|Ax_n - z\|_F \underset{n\to\infty}{\to} 0$

ii) Si $\delta_n/\epsilon_n \underset{n\to\infty}{\to} 0$ Alors $\|Ax_n - z\|_F = o(\epsilon_n)$ et $x_n \underset{n\to\infty}{\to} x$ où x est la solution de (3.2) la plus proche.

iii) Si de plus $x - x_0 \in \operatorname{Im} A^*$ (hypothèse de régularité), et si $\delta_n/\epsilon_n^2 \underset{n\to\infty}{\to} 0$ Alors $\|Ax_n - z\|_F = o(\epsilon_n^2)$ et $\|x_n - x\|_E = o(\epsilon_n)$.

Preuve. Notons x la solution de (3.3) qui minimise $\|x - x_0\|_E$ (x est bien défini, par une variante du corollaire 3.1).

i) Par définition de x_n, nous avons:

$$\|Ax_n - z_n\|_F^2 + \epsilon_n^2 \|x_n - x_0\|_E^2 \leq \|Ax - z_n\|_F^2 + \epsilon_n^2 \|x - x_0\|_E^2$$

d'où (en ajoutant et retranchant une quantité positive):

$$\begin{aligned}
\|Ax_n - z\|_F^2 + \epsilon_n^2 \|x_n - x\|_E^2 \quad &\leq \quad \|Ax - z_n\|_F^2 + \epsilon_n^2 \|x - x_0\|_E^2 \\
&\quad + \|Ax - z_n\|_F^2 + \epsilon_n^2 \|x - x_0\|_E^2 \\
&\quad - \|Ax_n - z_n\|_F^2 - \epsilon_n^2 \|x_n - x_0\|_E^2
\end{aligned}$$

En utilisant les identités

$$Ax_n - z_n = (Ax_n - z) + (z - z_n) \quad \text{et} \quad x_n - x_0 = (x_n - x) + (x - x_0)$$

nous obtenons:

$$\begin{aligned}
\|Ax_n - z\|_F^2 + \epsilon_n^2 \|x_n - x\|_E^2 \quad &\leq \quad -2\left(Ax_n - z, z - z_n\right) - 2\epsilon_n^2 \left(x_n - x, x - x_0\right) \qquad (3.14) \\
&\leq \quad 2\delta_n \|Ax_n - z\|_F + 2\epsilon_n^2 \|x_n - x\|_E \|x - x_0\|_E
\end{aligned}$$

par l'inégalité de Cauchy-Schwarz. Si nous notons, dans l'inégalité ci-dessus:

$$a = \|Ax_n - z\|_F, b = \epsilon_n \|x_n - x\|_E, \alpha = \delta_n, \beta = \epsilon_n \|x - x_0\|_E$$

cette inégalité se réécrit

$$a^2 + b^2 \leq 2a\alpha + 2b\beta$$

et nous en déduisons

$$(a - \alpha)^2 + (b - \beta)^2 \leq \alpha^2 + \beta^2 \leq (\alpha + \beta)^2$$

puis:

$$\begin{aligned}
a - \alpha \leq \alpha + \beta \quad &\Rightarrow \quad a \leq 2\alpha + \beta \\
b - \beta \leq \alpha + \beta \quad &\Rightarrow \quad b \leq \alpha + 2\beta
\end{aligned}$$

c'est-à-dire finalement:

$$\begin{aligned} \|Ax_n - z\|_F &\leq 2\delta_n + \epsilon_n \|x - x_0\|_E \\ \|x_n - x\|_E &\leq \tfrac{\delta_n}{\epsilon_n} + 2\|x - x_0\|_E \end{aligned} \tag{3.15}$$

Nous obtenons donc la convergence des observations sous la seule hypothèse que $z \in ImA$. Par contre, et c'est normal puisque à ce stade nous n'avons pas encore lié ϵ_n à δ_n, nous ne pouvons pas conclure quant à la convergence de x_n vers x.

ii) Nous supposons de plus que $\delta_n/\epsilon_n \underset{n\to\infty}{\to} 0$ Soit alors $\eta > 0$ fixé. La définition de x entraîne que $x - x_0 \in \ker A^\perp = \overline{\operatorname{Im} A^*}$, donc:

$$\exists w \in F, \|x - x_0 - A^* w\| \leq \eta$$

Reprenons (3.14), en notant que

$$(x_n - x, A^* w) = (A(x_n - x), w) = (Ax_n - z, w)$$

$$\begin{aligned} \|Ax_n - z\|_F^2 + \epsilon_n^2 \|x_n - x\|_E^2 &\leq -2(Ax_n - z, z - z_n) - 2\epsilon_n^2 (x_n - x, x - x_0) \\ &\quad +2\epsilon_n^2 (Ax_n - z, w) - 2\epsilon_n^2 (x_n - x, A^* w) \\ &\leq 2\|Ax_n - z\|_F \left(\delta_n + \epsilon_n^2 \|w\|_F\right) + 2\epsilon_n^2 \|x_n - x\|_E \eta \end{aligned}$$

Comme précédemment, nous en déduisons:

$$\begin{cases} \bullet \|Ax_n - z\|_F \leq 2\delta_n + 2\epsilon_n^2 \|w\|_F + \epsilon_n \eta = \epsilon_n \left(\frac{2\delta_n}{\epsilon_n} + 2\epsilon_n \|w\|_F + \eta\right) \\ \bullet \|x_n - x\|_E \leq \frac{\delta_n}{\epsilon_n} + \epsilon_n \|w\|_F + 2\eta \end{cases} \tag{3.16}$$

et, puisque nous avons supposé que $\frac{\delta_n}{\epsilon_n} \to 0$, les deux quantités ci-dessus peuvent être majorées par $3\epsilon_n \eta$ (resp. 3η) pour n assez grand, et comme η est arbitraire, cela prouve la convergence de x_n vers x.

iii) Enfin, si nous supposons que $x - x_0 \in \operatorname{Im} A^*$, nous pouvons prendre $\eta = 0$ dans les inégalités (3.16).

En supposant de plus que $\frac{\delta_n}{\epsilon_n^2} \to 0$, nous obtenons:

$$\begin{cases} \bullet \|Ax_n - z\|_F \leq \epsilon_n^2 \left(\frac{\delta_n}{\epsilon_n^2} + 2\|w\|_F\right) = o(\epsilon_n^2) \\ \bullet \|x_n - x\|_E \leq \epsilon_n \left(\frac{\delta_n}{\epsilon_n^2} + \|w\|_F\right) = o(\epsilon_n^2) \end{cases}$$

Ce théorème met une fois de plus en évidence le compromis stabilité–précision inhérent aux problèmes mal-posés. Il exprime que la suite ϵ_n doit tendre vers 0 moins vite que le niveau du bruit si l'on veut obtenir la convergence des solutions régularisées. De plus, cette suite doit tendre d'autant moins vite vers 0 que la solution est plus régulière. En ce qui concerne l'erreur sur la solution, elle est d'ordre ϵ_n, qui est donc plus grand que le niveau du bruit. On a donc une perte de précision, due bien évidemment à l'instabilité. Comme nous l'avons annoncé plus haut, la méthode de Tikhonov n'est pas une méthode de régularisation optimale. Pour une discussion plus approfondie de point, et pour l'analyse d'autre méthode, voir [7].

Remarque 2. i) L'hypothèse $x - x_0 \in \operatorname{Im} A^*$correspond typiquement à un résultat de régularité pour x, solution de (3.3).

ii) Ce théorème a un intérêt essentiellement théorique. En effet, pour pouvoir l'appliquer, il faudrait connaître la suite δ_n, c'est-à-dire le niveau de bruit contenu dans les données, ce qui est difficile en pratique. Le choix de ϵ_n donné par le théorème 1 est ce que nous avons appelé plus haut un choix a priori.

Chapitre 4

Problèmes inverses non-linéaires

Nous abordons maintenant les problèmes inverses non-linéaires, et nous nous concentrerons sur l'identification de paramètres dans les équations différentielles ou aux dérivées partielles.

Dans la situation générale, nous serons en présence d'un phénomène physique dont la structure est connue, mais dont les paramètres précis de fonctionnement ne le sont pas. Il est possible de mesurer certaines propriétés de ce système, correspondant à des entrées connue. Le système fonctionne comme une boite noire, et nous voudrions en connaître le contenu, sans « ouvrir » la boite.

Nous avons vu des exemples au chapitre 2. Reprenons l'exemple 2 : on sait (plus précisément, on fait l'hypothèse) que la conduction de la chaleur obéit à l'équation (2.7), mais nous ne connaissons pas le coefficient c (qui peut être un scalaire, ou une fonction de la position). Par contre, nous supposons que nous avons accès à une mesure de la température T, dans une partie du domaine (ou de la frontière). Avec ces informations, nous voulons déterminer le coefficient c qui permette de reproduire ces mesures.

La première différence avec les chapitres précédents est que l'application entre le paramètre (c dans notre exemple) et la mesure est non seulement non-linéaire, mais est exprimé par l'intermédiaire d'une équation comme (2.7), et la mesure est alors une « partie » de la solution de cette équation.

Une équation jouant le rôle de (2.7) s'appelle une équation d'état, et la variable

T, solution de cette équation, s'appelle l'état du système. Il sera en général irréaliste de supposer que l'état tout entier du système est connu: toujours dans notre exemple thermique, il ne sera pas possible de mesurer la température en tout point du domaine.

Une seconde différence, plus pratique, et qu'il est plus difficile d'obtenir des résultats théoriques que dans le cas linéaire. Les résultats qui ont été obtenus sont souvent liés à un problème particulier. Nous pouvons citer le livre [9] pour des éléments d'une théorie générale, ainsi que les articles de G. Chavent (par exemple [5]). En tout état de cause, dans le cadre de ce cours nous n'aborderons pas ces questions (ce qui ne veut pas dire qu'elles soient moins importantes), et nous nous concentrerons sur les méthodes numériques, et tout particulièrement sur la formulation aux moindres carrés.

4.1 Les trois espaces fondamentaux

Pour donner une formulation abstraite des problèmes que nous considérerons dans la suite de ce travail, nous allons introduire 3 espaces de Hilbert, ainsi que des applications entre ces espaces. Dans toutes les applications que nous considérerons par la suite, ces espaces seront tous de dimension finie. Il nous paraît toutefois utile de donner les définitions dans ce cadre plus général.

– l'espace des modèles (ou paramètres) M;
– l'espace d'état U;
– l'espace des données (ou observations) D;

Comme nous l'avons signalé, l'introduction de l'espace U permet de rendre explicite la dépendance entre le paramètre et les données. Par contre, l'existence de l'état ne dispense pas d'introduire l'observation, puisqu'il est en général non mesurable.

Deux application mettent en évidence les relations entre ces 3 espaces :

L'équation d'état relie de façon implicite le paramètre et l'état (tous deux peuvent évidemment être des vecteurs). Nous l'écrirons

$$F(a;u) = 0; a \in M; u \in U; F(a;u) \in Z \tag{4.1}$$

où Z est un autre espace de Hilbert. Nous supposons qu'il existe un sous-espace $M_{ad} \subset M$ tel que pour tout $a \in M_{ad}$, F définit localement un état unique $u = u_a$. Cela était le cas pour tous les exemples du chapitre 3 avec $M_{ad} = M$, puisque l'application F était linéaire par rapport à u.

Il sera pratique de noter

$$u = S(a) = u_a \tag{4.2}$$

la solution de l'équation d'état (4.1). La première égalité sera commode quand nous voudrons opérer sur cette solution (la dériver par exemple), alors que la seconde est un abus de notation suggestif.

L'équation d'observation extrait de l'état la partie correspondant aux mesures. Cela sera souvent une injection, rarement l'identité (sauf dans des exemples purement pédagogiques). Elle s'écrit

$$d = Hu; u \in U \tag{4.3}$$

Nous avons fait l'hypothèse simplificatrice que l'observation est un opérateur linéaire, indépendant du paramètre. L'extension à une situation plus générale n'est pas difficile, et est laissée au lecteur.

Si nous injectons la solution de (4.1) dans (4.3), nous obtenons l'application qui relie le paramètre à l'observation. Nous la noterons

$$d = \Phi(a) = H\left(S(a)\right) = H\left(u_a\right) \tag{4.4}$$

Le problème inverse est alors, étant donnée une observation d_{obs}, de résoudre l'équation :

$$\Phi(a) = d_{obs} \tag{4.5}$$

Donnons maintenant quelques exemples pour illustrer les concepts précédents. Nous reviendrons en détail sur ces exemples plus tard dans ce chapitre puis au chapitre suivant, pour déterminer les fonctionnelles correspondantes, puis calculer leurs gradients.

Nous commençons par un exemple simple, sans réelle signification.

Exemple 1. L'équation de la chaleur est le modèle de base qui régit les phénomènes de diffusion et intervient dans un grand nombre de domaines de la physique. Étant donné un ouvert $\Omega \subset \mathbb{R}^2$ (pour fixer les idées), dont nous notons Γ la frontière, et un réel $T > 0$, nous considérons le problème:

$$\left\{ \begin{array}{rcll} \frac{\partial}{\partial t}u - \operatorname{div}(a \operatorname{grad} u) & = & f & \text{dans} \;\; \Omega\times]0;T[\\ u(x,t) & = & 0 & \text{sur} \;\; \Gamma_D\times]0;T[\\ a\frac{\partial u}{\partial \eta} & = & g & \text{sur} \;\; \Gamma_N\times]0;T[\\ u(x;0) & = & u_0(x) & \text{sur} \;\; \Omega \end{array} \right. \tag{4.6}$$

où $f \in L^2(0;T;L^2(\Omega))$; $g \in L^2(0;T;L^2(\Gamma_N))$ et $u_0 \in L^2(\Omega)$ sont des fonctions données et supposées connues (Γ_N et Γ_D forment une partition de Γ), et nous cherchons à identifier la fonction a. Avec les réserves maintenant habituelles sur son caractère non-hilbertien, le choix naturel pour l'espace M est

$$M = L^\infty(\Omega), \text{ et } M_{ad} = \{a \in M, a(x) \geq a_* > 0\}$$

Nous ferons l'hypothèse que u est mesurée sur la partie Γ_N du bord (et que cette observation est continue en temps), et également que $u(x;T)$ est connue sur tout Ω à l'instant final. Dans ces conditions, les données consistent en deux fonctions: $d_N \in L^2(0,T;\Gamma_N)$ et $d_T \in L^2(\Omega)$, et l'espace $D = L^2(0,T;\Gamma_N) \times L^2(\Omega)$.

4.2 Formulation par moindres carrés

Dans les exemples que nous venons d'examiner, l'application Φ est définie implicitement. Elle est non-linéaire, même si l'équation d'état et l'équation d'observation sont linéaires. Cela rend évidemment plus difficile la résolution du problème inverse.

Ce que nous avons dit précédemment laisse à penser que l'équation (4.4) peut ne pas avoir de solution, et que même si elle en a, l'application inverse n'est pas nécessairement continue. Nous allons donc introduire une formulation, à priori plus faible, qui a fait la preuve de son utilité. Nous remplaçons l'équation (4.4) par le problème de minimisation suivant,

$$\text{minimiser} \;\; J(m) = \frac{1}{2}\left\|\Phi(m) - d_{obs}\right\|_D^2 \;\; \text{pour} \;\; m \in M_{ad} \tag{4.7}$$

Cette formulation s'appelle une méthode de moindres carrés, et J est la fonction coût, ou fonctionnelle d'erreur (la littérature anglo-saxonne dira : output least squares method, for the cost function, or functional, J). Il est important de comprendre comment « fonctionne » cette fonctionnelle. L'observation étant donnée une fois pour toute, pour évaluer la fonctionnelle J en un paramètre p, on commence par résoudre l'équation d'état (4.1), puis l'équation d'observation (4.3), et l'on compare l'observation simulée à celle mesurée.

Remarque 1. (L'erreur d'équation). Il existe une autre façon de reformuler un problème inverse comme un problème d'optimisation. Il consiste à remplacer l'état par l'observation (quitte à interpoler cette dernière). Cela conduit à une fonctionnelle

$$J(m,d) = \frac{1}{2}\left\| F\left(p, \widetilde{d}\right)\right\|^2 \tag{4.8}$$

où $\widetilde{d}$ est un interpolé de d. Comme elle est quadratique par rapport aux paramètres, cette méthode est très populaire auprès des physiciens et des ingénieurs. Son principal désavantage est de nécessiter d'interpoler l'observation.

Nous allons maintenant revenir sur les exemples 1, et proposer pour chacun d'eux une formulation en terme de minimisation de fonctionnelle.

Exemple 2 (suite de l'exemple 1). Comme dans les cas précédents, nous agrégeons les erreurs de mesure en une fonction coût unique:

$$J(a) = \int_0^T \int_{\Gamma_N} |u - d_N|^2 \, dxdt + \frac{1}{2}\int_{\Omega} |u(x,T) - d_T|^2 \, dx \tag{4.9}$$

Qu'apporte une telle reformulation ? Il est clair qu'elle ne peut pas changer comme par magie un problème mal posé en problème bien posé. Par contre, elle permet de rétablir l'existence. En effet, si il n'existe pas de solution à l'équation (4.4), le problème de minimisation aura forcément une solution (la fonction coût J est positive). Par contre, rien ne garantit que le minimum soit atteint en un point $p \in M_{ad}$. Il existe des contre-exemples, que l'on pourra trouver dans [9]. Une autre question essentielle est celle de l'unicité. On voit facilement qu'elle est liée à la convexité de la fonctionnelle J. Encore une fois, rien ne garantit cette propriété. La formulation (4.7) présente toutefois des avantages:

– elle donne une façon systématique pour poser les problèmes inverses ;

– dans certains cas, on peut démontrer des propriétés de la fonctionnelle J ;

– Comme nous le verrons au chapitre suivant, cette formulation permet de régulariser le problème, c'est-à-dire de l'approcher par une famille de problèmes bien posés, dont la solution converge vers la solution du problème original ;

– il existe de méthodes numériques robustes et bien étudiées pour résoudre les problèmes d'optimisation;

– sous des hypothèses raisonnables sur les données, la fonctionnelle J est différentiable, et se prête à l'attaque par une méthode d'optimisation locale de type gradient.

Remarque 2 (Sur le choix des normes). Nous avons travaillé dès le départ avec des espaces de Hilbert, et donc des normes hilbertiennes. En pratique, dans les espaces de fonction cela se traduit par des normes L^2 ou de Sobolev. Il n'y a rien de sacré dans ce choix, qui est essentiellement fait par commodité. En un certain sens, il n'est pas naturel, puisque les paramètres varient souvent dans (un sous-espace de) L^∞. Le principal avantage des normes hilbertiennes est qu'elles conduisent à des situations pour lesquelles on sait faire les calculs. On peut également justifier le choix de normes L^2 par des considérations statistiques dans lesquelles nous n'entrerons pas.

D'un autre coté, une norme hilbertienne présente le désavantage de donner plus de poids aux points aberrants, ce qui ne serait pas le cas pour une norme de type L^1. Cela a conduit à des méthodes d'inversions dites « robustes ».

4.2.1 Difficultés des problèmes inverses

La difficulté des problèmes inverses provient d'une combinaison de facteurs.

– Comme nous l'avons mentionné au paragraphe 4.2, la fonction coût est en général non convexe. Cela conduit à l'existence de minima locaux, et la méthode d'optimisation peut converger vers n'importe lequel de ces minima.

– Le problème inverse peut-être sous-determiné, du fait d'un manque de données (qui est intrinsèque au problème). Cela conduit à l'existence de plusieurs solutions, autrement dit de plusieurs paramètres produisant les mêmes observations.

– Le manque de continuité produit une instabilité. Même si l'on peut (en théorie) résoudre le problème pour des observations exactes, cela ne veut pas dire que l'on pourra le résoudre pour des données bruitées, même si le niveau de bruit est faible.

– Une difficulté de nature différente est liée au coût de la résolution, en supposant que l'on puisse s'affranchir des obstacles précédents. En effet, la simple évaluation de la fonction coût demande la résolution de l'équation d'état, c'est-à-dire en général d'une (ou de plusieurs) équation aux dérivées partielles.

4.2.2 Optimisation, paramétrisation, discrétisation

Le problème d'optimisation (4.7) ainsi que l'équation d'état (4.1) sont en général posés « en dimension infinie », c'est-à-dire que les espaces de Hilbert M, U et D sont de dimension infinie. Ce sont en général des espaces de fonctions, ainsi que nous l'avons vu dans les exemples du chapitre 2. pour la résolution numérique il est bien entendu nécessaire de se ramener à un problème en dimension finie. Comme on ne dispose habituellement que d'un nombre fini d'observations, l'espace D est souvent de dimension finie dès le départ.

Il n'en est pas de même pour l'espace des paramètres. Le processus qui consiste à remplacer M par un espace de dimension finie est la paramétrisation. Des exemples couramment utilisés sont des fonctions constantes par morceaux (une valeur par cellule, dans le cas d'une grille), des approximations polynômiales par morceaux (fonctions spleenes), mais d'autres choix sont possibles. Il est souhaitable de conserver le paramètre sous forme « fonctionnelle » le plus longtemps possible, de façon à pouvoir changer facilement de paramétrisation. Bien entendu, cela implique que l'algorithme d'optimisation devra fonctionner en dimension infinie.

Il est par contre nécessaire de discrétiser l'équation d'état. On remplace donc U par un espace de dimension finie. Ce choix a évidemment une influence capitale, car en changer implique d'avoir à recommencer tout le processus d'analyse. On obtient donc un problème d'optimisation pour lequel l'inconnue est toujours de dimension infinie, mais l'équation d'état est posée en dimension finie. Ceci permet d'obtenir le gradient exact de la fonction coût effectivement utilisée par le programme. Une alternative serait de discrétiser l'équation d'état au dernier moment, et de discrétiser également le gradient continu. L'expérience prouve que cette seconde manière de procéder dégrade la convergence des méthodes d'optimisation, et nous procéderons en calculant le gradient exacte de la fonctionnelle approchée.

Une fois ces choix faits, il faut choisir une discrétisation de a adaptée à la simulation, qui pourra être différente de la paramétrisation évoquée plus haut. Cette discrétisation sera d'une finesse compatible avec celle utilisée pour l'équation d'état, mais ne sera pas nécessairement adaptée à l'optimisation. On calculera tout de même le gradient de la fonction coût par rapport à ces paramètres. Une autre paramétrisation sera simplement une étape supplémentaire bien distincte de la simulation, dont le gradient s'obtient par application de la règle de dérivation composée. Cette méthode permet de séparer les différentes composantes du logiciel : simulation, calcul du gradient, paramétrisation et optimisation sont ainsi des modules séparés avec des interfaces bien spécifiées.

4.3 Calcul du gradient – La méthode de l'état adjoint

Dans cette section, nous allons exposer une méthode de calcul du gradient d'une fonction du type

$$J(a) = \frac{1}{2} \left\| \Phi(a) - d_{obs} \right\|_D^2 \tag{4.10}$$

où l'application non-linéaire Φ est définie par la résolution d'une équation d'état:

$$F(a, u) = 0 \tag{4.11}$$

dont la solution est u_a, puis en extrayant de u_a une observation

$$\Phi(a) = H u_a \tag{4.12}$$

La difficulté est clairement dans le calcul de la dérivée de l'application (implicite) $a \to u_a$. Nous allons commencer par passer en revue plusieurs méthodes pour calculer ce gradient. Nous verrons plus en détail la méthode de l'état adjoint qui permet ce calcul pour un coût indépendant du nombre de paramètres. Après avoir exposé la méthode dans un cadre général, abstrait, nous donnerons plusieurs exemples explicites pour voir comment mener à bien ce calcul dans une situation concrète.

4.3.1 Méthodes de calcul du gradient

Nous présentons dans ce paragraphe plusieurs façons de calculer le gradient de (4.10).

Il s'agit tout d'abord de la méthode des différences finies (qui n'est pas recommandée, mais qui présente tout de même une utilité pour fournir des valeurs de références), de la méthode des sensibilités, et enfin de la méthode de l'état adjoint. Laquelle des deux dernières est la plus adaptée à une situation dépend du nombre de paramètres par rapport au nombres d'observations.

4.3.2 Les différences finies

Cette méthode est, en apparence, très simple à mettre en œuvre, ce qui peut expliquer sa popularité, mais elle n'est pas recommandable, puisque non seulement son coût est proportionnel au nombre de paramètres à identifier, mais elle donne un résultat approché, avec une précision qu'il est difficile d'évaluer. Dans certaines circonstances, elle peut servir à valider un calcul de gradient par l'une des autres méthodes (sensibilités ou état adjoint).

On remplace le calcul d'une dérivée partielle par le quotient aux différences

$$\frac{\partial J}{\partial a} \approx \frac{J\left(a+h_j\right)-J\left(a\right)}{h_j} \tag{4.13}$$

On constate immédiatement que le nombre d'évaluations de J est égal au nombre de paramètres à identifier (plus un, mais ce calcul est toujours nécessaire). Par rapport à la méthode précédente, le coût est équivalent, mais l'on n'obtient que le gradient, pas le jacobien. De plus, le résultat n'est pas exact. D'ailleurs, en précision finie, l'erreur commise se compose de deux termes : l'erreur d'approximation et l'erreur d'arrondi. Nous allons analyser précisément comment se combinent ces deux effets. Pour simplifier, nous nous plaçons dans le cas d'une fonction f d'une variable réelle x.

Dans ce cas, le développement de Taylor de f à l'ordre un donne:

$$f'\left(x\right)=\frac{f(x+h)-f(x)}{h}+h/2f''\left(x\right)+o\left(h^2\right) \tag{4.14}$$

Supposons par ailleurs que f soit calculée avec une précision relative ϵ_f. Ce peut être simplement l'erreur d'arrondi, auquel cas ϵ_f est la précision de l'arithmétique de l'ordinateur (de l'ordre de 10^{-16} en double précision), ou bien une valeur bien plus grande si f est le résultat d'un calcul complexe.

Dans ce cas, ce qui est effectivement calculé est $\widetilde{f} = f(1+\epsilon_f)$, et la différence entre le quotient calculé et la vraie dérivée vaut, en négligeant l'erreur d'arrondi due à la division):

$$f'(x) = \frac{\widetilde{f}(x+h) - \widetilde{f}(x)}{h} + 2\epsilon_f/hf(x) + h/2f''(x) + o\left(h^2\right) \tag{4.15}$$

La somme des deux premiers termes est minimisée par le choix

$$h = 2\sqrt{\epsilon_f \frac{f(x)}{f''(x)}} \tag{4.16}$$

et l'erreur totale est alors proportionnelle à $\sqrt{\epsilon_f}$. En double précision, cela veut dire que la dérivée aura environ 8 chiffres exacts. Notons que la valeur de la fonction à dériver, et de sa dérivée seconde, influent sur le choix du pas optimal, comme le montre l'équation (4.16). Plus $f(x)$ sera grand, plus on pourra choisir ϵ grand. De même, plus $f''(x)$ sera grand, c'est-à-dire plus $f'(x)$ varie rapidement, plus le choix de ϵ devra être petit. Ainsi, le choix effectif du pas reste délicat, même si $\sqrt{\epsilon_f}$ est une première estimation raisonnable.

Le livre [11] contient un algorithme de choix du pas, dans le cas de plusieurs variables, qui prend en comte les divers facteurs d'échelle qui peuvent intervenir. Toutefois, comme nous déconseillons de calculer le gradient par cette méthode, nous ne donnerons pas de détails.

Il est tout de même des cas où le calcul du gradient par différences finies peut-être utile : en particulier pour vérifier un calcul par état adjoint. Dans ce cas, on choisit un paramètre « au hasard», et l'on calcule le gradient en ce point par différences finies, pour plusieurs valeurs du pas (disons de 10^{-2} à 10^{-15}). Si les deux calculs sont justes, l'erreur doit passer par un minimum au voisinage de 10^{-7}).

4.3.3 Les fonctions de sensibilité

Il s'agit de la méthode la plus naturelle pour calculer le gradient de J. Elle consiste à dériver l'équation d'état explicitement par rapport au paramètre a, puis à utiliser la règle dérivation d'une fonction composée. Insistons sur le fait que cette méthode donne un résultat exact.

Donc:

$$\nabla J(a) = \Phi'(a)^*(\Phi(a) - d_{obs}) \tag{4.17}$$

Une première idée est donc de calculer le Jacobien de Φ. Comme Φ est définie de façon implicite par la résolution de l'équation d'état (4.1) et l'équation d'observation (4.3), nous devons faire appel au théorème des fonctions implicites (plus exactement à son corollaire qui permet le calcul de la différentielle de l'application implicite une fois que l'on sait que celle-ci est différentiable). Ce résultat dit que l'on obtient le jacobien de Φ en dérivant l'équation d'état:

$$\partial_u F(a; u)\delta u + \partial_a F(a; u)\delta a = 0; \tag{4.18}$$

puis en résolvant l'équation (linéaire) précédente, et en composant avec (la dérivée de) l'observation (que nous avons supposé linéaire)

$$\Phi'(a) = -H\left(\partial_u F(a; u)\right)^{-1} \partial_a F(a; u) \tag{4.19}$$

En regroupant les équations (4.17) et (4.19), on obtient finalement le gradient de J:

$$\nabla J(a) = \left(\Phi'(a)\right)^* \left(Hu\,(a) - d_{obs}\right) \tag{4.20}$$

Le principal désavantage de cette méthode réside dans le fait que le calcul de δu demande la résolution d'une équation d'état (linéarisée) pour chaque valeur de δa. Après passage en dimension finie, cela veut dire que le calcul de chaque dérivée partielle $\partial J/\partial a_j$ demande la résolution d'une équation comme (4.18). Le coût du calcul du gradient est donc proportionnel au nombre de paramètres. Or, dans une grande partie des situations d'intérêt, ce nombre peut être très grand: plusieurs centaines, voire plusieurs milliers. Nous verrons plus bas, c'est le principal avantage de la méthode de l'état adjoint, qu'il est possible de réaliser ce calcul à un coût proportionnel à celui d'une seule équation linéarisée, et en particulier, indépendant du nombre de paramètres.

En contrepartie, cette méthode fournit plus que le gradient, puisque nous avons vu qu'elle calcule le jacobien de Φ. Une fois ce jacobien disponible, il est possible de l'exploiter en calculant, par exemple, ses valeurs singulières. De plus, la méthode de Gauss-Newton nécessite la connaissance de ce jacobien. Si le nombre de paramètres n'est pas trop élevé, la méthode de Gauss-Newton en calculant le gradient comme en (4.20) peut être plus économique qu'une méthode de quasi-Newton avec calcul du gradient par l'état adjoint.

4.3.4 La méthode de l'état adjoint

Nous avons déjà noté que la méthode des fonctions de sensibilité fournissait plus que le gradient de J. Si nous n'avons besoin que du gradient, nous pouvons réarranger le calcul menant à (4.20) pour éviter le calcul du jacobien complet. En reportant (4.19) dans (4.20), et en transposant le produit, nous obtenons (la notation - * représente l'inverse de l'adjoint):

$$\nabla J(a) = -\left(\partial_a F(a;u)^* \left(\partial_u F(a;u)^*\right)^{-1} H^*\right)\left(Hu\left(a\right) - d_{obs}\right) \tag{4.22}$$

La remarque d'apparence triviale qui va nous permettre de simplifier le calcul est qu'il est possible de parenthéser différemment cette expression:

$$\nabla J(a) = -\left(\partial_a F(a;u)\right)^* \left(\left(\partial_u F(a;u)^*\right)^{-1}\left(H^*\left(Hu\left(a\right) - d_{obs}\right)\right)\right) \tag{4.23}$$

Il est commode de donner un nom à la quantité à l'intérieur de la seconde parenthèse, et d'introduire le vecteur p solution de

$$\partial_u F(a;u)^* p = -H^*\left(Hu\left(a\right) - d_{obs}\right) \tag{4.24}$$

On appelle cette équation l'équation adjointe, et p est l'état adjoint. Une fois cette équation résolue, le gradient se calcule alors par:

$$\nabla J(a) = \partial_u F(a;u)^* p \tag{4.25}$$

Résumons le résultat de ces calculs dans le

Théorème 1. Si p est la solution de l'équation adjointe (4.23), le gradient de J au point a est donné par (4.24), où $u = u(a)$ est la solution de l'équation d'état (4.1) correspondant à a.

Remarque 1. Le théorème 1 est d'une grande importance. Il fournit le cadre général sur lequel s'appuie la méthode de l'état adjoint. Toutefois, il est difficile à appliquer tel quel en pratique, et c'est pourquoi nous proposerons une méthode plus simple pour aboutir au même résultat. En effet, il peut être difficile, dans un cas concret d'identifier les différents adjoints concernés, voire même l'opérateur F lui-même. Ce sera en particulier vrai pour les problèmes d'évolution.

Remarque 2. Nous voyons, d'après l'équation (4.24) que nous obtenons le gradient complet de J en résolvant la seule équation adjointe (4.23). Ceci implique que la méthode de l'état adjoint permet de calculer le gradient à un coût proportionnel à celui du calcul de la fonction elle-même. Dans la plupart des cas, ce coût est un petit (entre 3 et 5) multiple de celui de la fonction (ce résultat est démontré dans [2]). Ceci justifie, à posteriori, la méthode de l'état adjoint, et l'abandon de la méthode « naturelle » vue au paragraphe 4.3.3.

Remarque 3. L'opérateur qui intervient dans l'équation adjointe est donc l'adjoint de $H\partial_u F(a;u)^{-1}$, qui n'est autre que la dérivée de l'application Φ qui associe le paramètre a données. C'est cette propriété qui à donné son nom à la méthode. La différence avec le calcul par les fonctions de sensibilité est qu'ici cette transposition est implicite, « cachée » dans la définition de l'équation adjointe.

Remarque 4. L'équation adjointe (4.23) est toujours linéaire, et ce même si l'équation d'état était non-linéaire. Les sources de l'équation adjointe sont formées par le résidu de l'équation d'observation.

En pratique, nous verrons qu'il est parfois difficile de réaliser concrètement les transpositions indiquées dans le théorème 1. Nous allons introduire une autre façon de mener à bien ce calcul qui se révèle d'utilisation plus simple en pratique.

4.3.5 Calcul de l'état adjoint par le lagrangien

Le paragraphe précédent a montré comment calculer le gradient de notre fonction coût en résolvant seulement deux équations : l'équation d'état, suivie de l'équation adjointe (les opérations de la formule (4.24) sont typiquement très simples). Comme nous l'avons déjà signalé, la mise en oeuvre de ce résultat n'est pas aussi simple. Nous allons donner une technique qui conduit au même résultat, mais se révèle plus souple d'emploi, comme nous le verrons dans les exemples qui suivent (voir paragraphe dans le chapitre 4).

La méthode repose sur ce qui peut être vu comme une astuce de calcul. On commence par prétendre que les variables a et u varient indépendamment, et l'on considère l'équation d'état comme une contrainte. Dans ces conditions, il est naturel d'introduire un lagrangien. Dans notre cas, celui-ci s'écrit (au moins formellement, puisqu'ici nous n'avons pas un nombre fini de contraintes):

$$L\left(a,u,p\right)=\frac{1}{2}\left\|Hu\left(a\right)-d_{obs}\right\|^{2}+\left(p,F(a;u)\right) \tag{4.26}$$

La remarque (encore une fois en apparence triviale) fondamentale est que, si u vérifie l'équation d'état correspondant au paramètre a, on a l'identité:

$$L(a;u(a);p)=J(a);\forall p\in Z \tag{4.27}$$

En dérivant cette relation, il vient:

$$J'(a)\delta a=\partial_{a}L(a;u)\delta a+\partial_{u}L(a;u)\partial_{a}u(a)\delta a \tag{4.28}$$

La partie difficile à calculer est $\partial_{a}u(a)$. Si nous pouvons choisir $p\in Z$ de façon que ce terme disparaisse, nous aurons une expression simple pour la dérivée de J. Pour cela, nous considérerons que $\delta u=\partial_{a}u(a)\delta a$ est une quantité indépendante, et nous demandons que l'opérateur $\delta u\rightarrow\partial_{u}L(a;u)$ s'annule. Définissons alors l'équation adjointe abstraite par:

$$\partial_{u}L(a;u(a))\delta u=0;\forall\delta u\in U; \tag{4.29}$$

soit, en explicitant l'expression de $\partial_{u}L$

$$(H\delta u;Hu(a)-d_{obs})+(p;\partial_{u}F(a;u(a))\delta u)=0;\forall\delta u\in U \tag{4.30}$$

Nous retrouvons précisément l'équation adjointe (4.23), puis la différentielle de J se calcule par la formule:

$$J'(a)\delta a=(p;\partial_{a}F(a;u(a))\delta a) \tag{4.31}$$

et nous voyons qu'en fait nous obtenons le gradient de J:

$$\nabla J(a) = (\partial_a F(a; u(a)))^* p \tag{4.32}$$

identique à (4.24).

Remarque 5. En pratique, la forme la plus utile de l'équation adjointe est la formulation « variationnell» (4.28). En effet, comme nous l'avons déjà signalé, il n'est pas toujours commode de calculer les opérateurs adjoints. Par contre, il est toujours simple (nous le constaterons) de partir de la forme (4.28), et de la manipuler (par des intégrations ou des sommations par parties), pour aboutir à une équation adjointe explicite.

De même, il est souvent plus commode de partir de l'équation (4.29) et de la manipuler pour identifier le gradient que d'utiliser la formule (4.30) telle quelle.

Cette méthode est suffisamment importante pour que nous en résumions les étapes principales.

Théorème 2. Le calcul du gradient de la fonctionnelle (4.7) passe par les étapes suivantes :

i) Définir le lagrangien par (4.25) ;

ii) Résoudre l'équation adjointe (variationnelle) (4.28) qui détermine l'état adjoint p;

iii) La différentielle de J est donnée par (4.29) qui permet d'identifier le gradient de J.

4.4 Exemples de calcul de gradient

Nous allons appliquer la théorie développée au paragraphe précédent aux exemples que nous avons commencé à examiner: le modèle des équation aux dérivées partielles (l'équation de la chaleur) peut se traiter en utilisant l'état adjoint.

4.4.1 Équation de la chaleur

L'équation de la chaleur est le modèle de base qui régit les phénomènes de diffusion et intervient dans un grand nombre de domaines de la physique. Étant donné un ouvert

$\Omega \subset \mathbb{R}^2$ (pour fixer les idées) et un réel $T > 0$, nous considérons le problème:

$$\begin{cases} \frac{\partial u}{\partial t} - \operatorname{div}(a \operatorname{grad} u) & = & f & \text{dans} & \Omega \times]0, T[\\ u(x;t) & = & 0 & \text{sur} & \Gamma_D \times]0, T[\\ a\frac{\partial u}{\partial \eta} & = & g & \text{sur} & \Gamma_N \times]0, T[\\ u(x;0) & = & u_0(x) & \text{sur} & \Omega \end{cases} \tag{4.33}$$

où $f \in L^2(0;T;L^2(\Omega))$; $g \in L^2(0;T;L^2(\Gamma_N))$; et $u_0 \in L^2(\Omega)$ sont des fonctions données (et supposées connues), et nous cherchons à identifier la fonction a. nous ferons l'hypothèse que u est mesurée sur la partie Γ_N du bord, mais également que $u(x;T)$ est connue sur tout Ω à l'instant final. Dans ces conditions, les données consistent en deux fonctions : $d_N \in L^2(0;T;\Gamma_N)$ et $d_T \in L^2(\Omega)$. Avec les réserves maintenant habituelles sur son caractère non-hilbertien, le choix naturel pour l'espace M est $M = L^\infty(\Omega)$, et $M_{ad} = \{a \in M; a(x) \geq a_* > 0\}$ La fonction coût est

$$J(a) = \frac{1}{2}\int_0^T \int_{\Gamma_N} |u - d_N|^2 \, dxdt + \frac{1}{2}\int_\Omega |u(x,T) - d_T|^2 \, dx \tag{4.34}$$

Il est standard de mettre (4.37) sous forme variationnelle, et d'en déduire à la fois un théorème d'existence et la méthode d'approximation. Il est usuel de poser

$$H = L^2(\Omega); V = \left\{v \in H^1(\Omega); v = 0 \ \ sur \ \ \Gamma_D\right\} \tag{4.35}$$

La formulation faible est alors

$$\begin{cases} \text{Chercher} \ \ u \in L^2(0;T;V) \cap C^0(0;T;H) \ \ \text{telque} \\ \frac{d}{dt}\int_\Omega u(t)v + \int_\Omega a \operatorname{grad} u(t) \operatorname{grad} v = \int_\Omega f(t)v + \int_{\Gamma_N} g(t)v, \forall v \in V, \quad \text{p.p. sur} \quad]0, T[\\ u(0) = u_0 \end{cases} \tag{4.36}$$

On peut énoncer un résultat d'existence pour le problème (4.36). Nous renvoyons le lecteur à [[20],vol. 8] pour plus de détails et nous passerons directement au schéma d'approximation. En ce qui concerne la discrétisation en espace, nous utiliserons une méthode d'éléments finis, et en ce qui concerne la discrétisation en temps, nous prendrons le schéma de Crank-Nicolson (voir [[20], vol. 9], [[18], [3]] pour plus de détails tant sur les éléments finis que les schémas pour les problèmes d'évolution). Nous ne supposerons pas que la discrétisation en espace est nécessairement uniforme, et nous poserons donc, pour une partition $0 = t^0 < t^1 < ... < t^K = T$, $\Delta t^{k+1/2} = t^{k+1} - t^k$. Ce choix est justifié par la physique du problème : on sait, en effet, que les solution de (4.33) deviennent de plus en plus régulières au cours du temps, et il sera naturel de vouloir augmenter le pas de temps au cours de la simulation. De plus nous verrons que dans ce cas, le problème adjoint discret ne sera pas nécessairement celui qui aurait été obtenu par discrétisation d'un état adjoint continu. Il sera commode, dans la suite de poser:

$$u_h^{k+1/2} = \frac{u_h^k + u_h^{k+1}}{2}; k = 0, ..., K-1$$

Avec ces notations, le problème direct est

$$\left\{ \begin{array}{l} \text{Chercher } u_h = (u_h^0, u_h^1, ..., u_h^K) \in U_h^{K+1} \text{ tel que, pour tout } \ v_h \in U_h \\ I_\Omega^h\left(\frac{u_h^{k+1} - u_h^k}{\Delta t^{k+1/2}} v_h\right) + I_\Omega^h\left(a_h \operatorname{grad} u_h^{k+1/2} \operatorname{grad} v_h\right) = I_\Omega^h\left(f_h^{k+1/2} v_h\right) + I_{\Gamma_N}^h\left(g_h^{k+1/2} v_h\right) \\ I_\Omega^h\left(u_h^0 v_h\right) = I_\Omega^h\left(u_0 v_h\right); \forall v_h \in U_h \end{array} \right. \tag{4.37}$$

où $f_h^{k+1/2}$ et $g_h^{k+1/2}$ sont des approximations de f_h et g_h à l'instant $t^{k+1/2}$. La dernière équation de (4.37) définit la condition initiale discrète par projection sur $L^2(\Omega)$, ce qui se signifie simplement que u_h^0 est la fonction de u_h qui prend les mêmes valeurs que u_0 aux points du maillage situés hors de Γ_D.

Le dernier point à préciser concerne l'approximation des intégrales en temps. Par analogie avec le schéma numérique, nous poserons, pour θ définie sur $[0; T]$ (c'est la règle du trapèze):

$$I_{\Delta t}(\theta) \approx \int_0^T \theta(t)\, dt = \sum_{k=0}^{K-1} \frac{\theta(t^k) + \theta(t^{k+1})}{2} \Delta t^{k+1/2} \tag{4.38}$$

Nous pouvons alors définir le lagrangien, par:

$$\begin{aligned} L_h(a_h,u_h,p_h) &= \frac{1}{4}\sum_{k=0}^{K-1} I_\Omega^h \left|u_h^k - d_h^k\right|^2 \Delta t^{k+1/2} \\ &+\frac{1}{4}\sum_{k=0}^{K-1} I_\Omega^h \left|u_h^{k+1} - d_h^{k+1}\right|^2 \Delta t^{k+1/2} \\ &+\frac{1}{2} I_{\Gamma_N}^h \left(\left|u_h^K - d_T\right|^2\right) \\ &+\sum_{k=0}^{K-1} I_\Omega^h \left(\frac{u_h^{k+1}-u_h^k}{\Delta t^{k+1/2}} p_h^{k+1/2}\right) \Delta t^{k+1/2} \\ &+\sum_{k=0}^{K-1} I_\Omega^h \left(a_h \operatorname{grad} u_h^{k+1/2} \operatorname{grad} p_h^{k+1/2}\right) \Delta t^{k+1/2} \\ &-\sum_{k=0}^{K-1} \left\{ I_\Omega^h \left(f_h^{k+1/2} p_h^{k+1/2}\right) + I_{\Gamma_N}^h \left(g_h^{k+1/2} p_h^{k+1/2}\right)\right\} \Delta t^{k+1/2} \end{aligned} \tag{4.39}$$

Suivant la procédure habituelle, nous commençons par écrire l'équation adjointe. Nous noterons $p_h = (p_h^{1/2}, ..., p_h^{K-1/2})$ le multiplicateur. L'équation adjointe « abstraite » est

$$\frac{\partial L_h}{\partial u_h}\delta u_h = 0, \forall \delta u_h = \left(0, \delta u_h^1, ..., \delta u_h^K\right) \in U_h^{K+1}$$

Nous obtenons donc, pour $\delta u_h \in U_h^{K+1}$:

$$\begin{aligned} &\sum_{k=1}^{K} \frac{1}{2} I_\Omega^h \left\{\left(u_h^k - d_h^k\right)\delta u_h^k + \left(u_h^{k+1} - d_h^{k+1}\right)\delta u_h^{k+1}\right\} \Delta t^{k+1/2} \\ &+I_{\Gamma_N}^h \left(\left(u_h^K - d_T\right)\delta u_h^K\right) \\ &+\sum_{k=0}^{K-1} \left\{ I_\Omega^h \left(\frac{\delta u_h^{k+1} - \delta u_h^k}{\Delta t^{k+1/2}} p_h^{k+1/2}\right) + I_\Omega^h \left(a_h \operatorname{grad} \delta u_h^{k+1/2} \operatorname{grad} p_h^{k+1/2}\right)\right\} \Delta t^{k+1/2} \\ = \;& 0 \end{aligned} \tag{4.40}$$

Nous pratiquons ensuite une intégration par parties discrète, c'est-à-dire un décalage d'indices, de façon à faire apparaître δu_h^k en facteur. L'équation précédente devient:

$$\begin{aligned}
&\sum_{k=1}^{K} \frac{1}{2} I_\Omega^h \left\{ \left(u_h^k - d_h^k\right) \delta u_h^k + \left(u_h^{k+1} - d_h^{k+1}\right) \delta u_h^{k+1} \right\} \Delta t^{k+1/2} \\
&+ I_{\Gamma_N}^h \left(\left(u_h^K - d_T\right) \delta u_h^K \right) + \sum_{k=1}^{K} I_\Omega^h \left(\delta u_h^k p_h^{k-1/2} \right) \\
&+ \frac{1}{2} \sum_{k=1}^{K} I_\Omega^h \left(a_h \operatorname{grad} \delta u_h^k \operatorname{grad} p_h^{k-1/2} \right) \Delta t^{k-1/2} \\
&- \sum_{k=0}^{K-1} I_\Omega^h \left(\delta u_h^k p_h^{k+1/2} \right) + \frac{1}{2} \sum_{k=0}^{K-1} I_\Omega^h \left(a_h \operatorname{grad} \delta u_h^k \operatorname{grad} p_h^{k+1/2} \right) \Delta t^{k+1/2} \\
= \;& 0
\end{aligned} \tag{4.41}$$

Rappelons que nous avons $\delta u_h^0 = 0$, ce qui élimine les termes avec $k = 0$ dans l'égalité précédente.

Pour interpréter l'équation adjointe, nous procédons en deux étapes: nous allons choisir des perturbations δu_h d'un type particulier, n'agissant que sur un seul instant. Nous traiterons d'abord le cas général $k < K$, puis l'instant final $k = K$.

Choisissons donc $\delta u_h = (0, ..., v_h, ..., 0)$, où v_h est en position k, pour $k = 1, ..., K-1$. En posant $\Delta t^k = 1/2 \left(\Delta t^{k-1/2} + \Delta t^{k+1/2}\right)$ nous obtenons, pour tout $k = 1, ..., K-1$:

$$\begin{aligned}
&\Delta t^k I_{\Gamma_N}^h \left(u_h^k - d_h^k\right) + I_\Omega^h \left(v_h p_h^{k+1/2} \right) - I_\Omega^h \left(v_h p_h^{k-1/2} \right) \\
&+ \frac{\Delta t^{k+1/2}}{2} I_\Omega^h \left(a_h \operatorname{grad} v_h \operatorname{grad} p_h^{k+1/2} \right) \\
&+ \frac{\Delta t^{k-1/2}}{2} I_\Omega^h \left(a_h \operatorname{grad} v_h \operatorname{grad} p_h^{k-1/2} \right) \\
= \;& 0
\end{aligned} \tag{4.42}$$

que nous pouvons réécrire sous une forme analogue à (4.37):

$$\begin{aligned}
&I_\Omega^h \left(\frac{p_h^{k-1/2} - p_h^{k+1/2}}{\Delta t^k} v_h \right) + I_\Omega^h a_h \frac{\Delta t^{k-1/2}}{2\Delta t^k} \operatorname{grad} p_h^{k-1/2} \operatorname{grad} v_h \\
&+ I_\Omega^h a_h \frac{\Delta t^{k+1/2}}{2\Delta t^k} \operatorname{grad} p_h^{k+1/2} \operatorname{grad} v_h \\
= \;& -I_{\Gamma_N}^h \left(\left(u_h^k - d_h^k\right) v_h \right); \forall v_h \in U_h, \forall k = 1, ..., K-1.
\end{aligned} \tag{4.43}$$

Il reste un degré de liberté: choisissons $\delta u_h = (0, ..., v_h)$ dans (4.45). Il vient:

$$\begin{aligned} & \frac{\Delta t^{K-1/2}}{2} I^h_{\Gamma_N} \left(\left(u^K_h - d^K_h \right) v_h \right) + I^h_\Omega \left(\left(u^K_h - d_T \right) v_h \right) \\ & + I^h_\Omega \left(v_h p^{K-1/2}_h \right) + \frac{1}{2} \left(a_h \operatorname{grad} v_h \operatorname{grad} p^{K-1/2}_h \right) \\ = \; & 0; \forall v_h \in U_h. \end{aligned}$$

Cette équation définit $p_h^{K-1/2}$. Nous pouvons la mettre sous la même forme que pour $k < K$, en introduisant un multiplicateur fictif. Posons

$$p_h^{K+1/2} = -\left(u^K_h - d_T \right), \quad \text{et} \quad \Delta t^K = 1/2 \Delta t^{K-1/2}$$

Remarque 6. L'équation (4.43), compris la condition finale ci-dessus, correspond à une discrétisation (consistante) de l'équation de la chaleur rétrograde:

$$\left\{ \begin{array}{rcll} -\frac{\partial p}{\partial t} - \operatorname{div}(a \operatorname{grad} p) & = & 0 & \text{dans} \quad \Omega \times]0, T[\\ p & = & 0 & \text{sur} \quad \Gamma_D \times]0, T[\\ a \frac{\partial p}{\partial \eta} & = & -(u - d_N) & \text{sur} \quad \Gamma_N \times]0, T[\\ p(x; T) & = & -(u(x, T) - d_T) & \text{sur} \quad \Omega \end{array} \right.$$

mais les différents facteurs Δt rendent cette discrétisation non-standard. Il est peu probable que nous ayons pu obtenir ce schéma particulier en discrétisant le schéma (4.41).

Remarque 7. Comme l'observation avait lieu à l'instant final, la condition finale sur l'état adjoint n'est pas homogène.

Nous pouvons enfin calculer le gradient de J_h:

$$\delta J_h = \frac{\partial L}{\partial a_h} \sum_{k=0}^{K-1} I^h_\Omega \left(\delta a_h \operatorname{grad} u^{k+1/2}_h \operatorname{grad} p^{k+1/2}_h \right) \Delta t^{k+1/2} \tag{4.44}$$

ce qui donne, en revenant à la définition de I^h_Ω, la dérivée partielle par rapport aux éléments de a_h:

$$\frac{\partial J_h}{\partial a_T} = |K| \sum_{k=0}^{K-1} \Delta t^{k+1/2} \left(\operatorname{grad} u^k_{h|T} + \operatorname{grad} u^{k+1}_{h|T} \right) \operatorname{grad} p^{k+1/2}_{h|T} \tag{4.45}$$

Conclusion générale

En ce thème nous avons présenté un exemple de problème inverse en thermique. Nous introduirons la notion fondamentale de problème mal posé, qui est caractéristique des problèmes inverses. Nous introduirons une source importante de problèmes inverses linéaires et nous introduirons des méthodes de discrétisation, conduisant à des problèmes de moindres carrés. Nous aborderons l'étude des techniques pour les problèmes mal posés, tout particulièrement la méthode de régularisation de Tikhonov. Nous aborderons les problèmes non-linéaires, essentiellement les problèmes d'estimation de paramètres dans les équations différentielles, et aux dérivées partielles, nous verrons comment poser les problèmes d'identification en terme de minimisation, quelles sont les principales difficultés auxquelles on peut s'attendre. Et abordera la technique importante de l'état adjoint pour calculer le gradient des fonctionnelles qui interviennent dans les problèmes de moindres carrés.

Bibliographie

[1] A. Kirsch. An Introduction to the Mathematical Theory of Inverse Problems. Number 120 in Applied Mathematical Sciences. Springer-Verlag, New-York, 1996.

[2] A. Griewank. Some bounds on the complexity of gradients, jacobians, and hessians. Technical Report MCS-P355-0393, Mathematics and Computer Science Division, Argonne National Laboratory, 1993.

[3] B. Lucquin and O. Pironneau. Introduction au calcul scientifique. Masson, Paris, 1996.

[4] C. W. Groetsch. Inverse Problems in the Mathematical Sciences. Vieweg, Wiesbaden, 1993.

[5] G. Chavent. On the theory and practice of non-linear least squares. Advances inWater Resources, 14(2) :55–63, 1991.

[6] G. Chavent. Problèmes inverses. notes de cours de DEA, Université Paris Dauphine.

[7] H. W. Engl, M. Hanke, and A. Neubauer. Regularization of Inverse Problems. Kluwer, Dordrecht, 1996.

[8] H. W. Engl. Regularization methods for the stable solution of inverse problems. Surveys Math. Indust., 3 :71–143, 1993.

[9] H. T. Banks and K. Kunisch. Estimation Techniques for Distributed Parameter Systems. Birkhäuser-Verlag, Zürich, 1989.

[10] H. Brézis. Analyse fonctionnelle – Théorie et applications. Masson, Paris, 1983.

[11] J. E. Dennis and R. B. Schnabel. Numerical Methods for Unconstrained Optimization and Nonlinear Equations. SIAM, Philadelphia, 1996.

[12] J. Baumeister. Stable Solution of Inverse Problems. Vieweg, Braunschweig, 1987.

[13] J. R. Cannon. The one-dimensional heat equation. Addison-Wesley Publishing Company Advanced Book Program, Reading, MA, 1984. With a foreword by Felix E. Browder.

[14] J. B. Keller. Inverse problems. Amer. Math. Monthly, 83 :107–118, 1976.

[15] J. Hadamard. Lectures on Cauchy's Problem in Linear Partial Differential Equations. Yale University Press, 1923.

[16] Michel KERN. Problèmes Inverses. École Superieure D'Ingénieures LÉONARD DE VINCI 2002–2003

[17] P. C. Hansen. Rank-deficient and discrete ill-posed problems : numerical aspects of linear inversion. SIAM, Philadelphia, 1998.

[18] P.-A. Raviart and J.-M. Thomas. Introduction à l'analyse numérique des équations aux dérivées partielles. Collection Mathématiques appliquées pour la maîtrise. Masson, Paris, 1983.

[19] R. Kress. Linear Integral Equations, volume 82 of Applied Mathematical Sciences. Springer, 1989.

[20] R. Dautray and J. L. Lions, editors. Analyse mathématique et calcul numérique pour les sciences et les techniques. Masson, 1982.

[21] V. Isakov. Inverse Problems for Partial Differential Equations. Number 127 in Applied Mathematical Sciences. Springer, New-York, 1998.

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
UNIVERSITE LARBI BEN M'HIDI OUM EL BOUAGHI
Faculté des Sciences Exactes et des Sciences de la nature et de la vie
Département de Mathématiques et Informatiques

N° d'ordre: ...
Série: ...

MÉMOIRE DE MAGISTER
de Mathématiques Appliquée

Intitulé:
PROBLÈMES INVERSES DE L'ÉQUATION DE TRANSFERT DE LA CHALEUR

Présenté par: DERBAZI Amar

Soutenu publiquement le 23/06/2011 à l'Université de Oum El Bouaghi devant le jury composé de

Mr MOHAMED Bouzit	M.C.A	Univ. Larbi.B.M'hidi O.E.B	Président
Mr MERZOUK Djebarni	M.C.A	Univ. Larbi.B.M'hidi O.E.B	Rapporteur
Mr ABDELKARIM Aliouche	M.C.A	Univ. Larbi.B.M'hidi O.E.B	Examinateur
Mr MOHAMED Dalah	M.C.A	Univ. de Canstantine	Examinateur