Sur l'approche variationnelle de la mollification dans la théorie des problèmes mal posés et applications

Walter Cédric Simo Tao Lee

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *10/12/2020* **par :**

Walter Cedric SIMO TAO LEE

On the variational approach to mollification in the theory of ill-posed problems and applications

JURY

Samir ADLY	Professeur des Universités	Université de Limoges
Fabrice GAMBOA	Professeur des Universités	Université Paul Sabatier
Jan JOHANNES	Professeur des Universités	Heidelberg University
Pierre MARÉCHAL	Professeur des Universités	Université Paul Sabatier
Alberto SEEGER	Professeur des Universités	Université d'Avignon
Faouzi TRIKI	Maitre de Conférences	Université de Grenoble Alpes
Anne VANHEMS	Assimilés Professeur des Universités	Toulouse Business School

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
Institut de Mathématiques de Toulouse
Directeur(s) de Thèse :
Pierre MARÉCHAL et *Anne VANHEMS*
Rapporteurs :
Faouzi TRIKI et *Jan JOHANNES*

Résumé

Les problèmes inverses constituent un domaine en pleine expansion en mathématiques appliquées qui a suscité une grande attention au cours des dernières décennies en raison de son omniprésence dans plusieurs domaines des sciences et technologies. Le plus souvent, les problèmes inverses donnent lieu à des équations mathématiques instables. Autrement dit, les solutions ne dépendent pas continument des données. En effet, de très petites perturbations sur les données peuvent causer des erreurs arbitrairement grandes sur les solutions. Étant donné que le bruit est généralement inévitable, inverser l'équation mal-posée échoue à résoudre le problème. Il est alors nécessaire d'appliquer une méthode de régularisation afin de récupérer des approximations stables des solutions. À cet égard, plusieurs techniques de régularisation ont été développées dans la littérature.

Globalement, ces méthodes de régularisation peuvent être divisées en deux classes : Une classe de méthodes qui tentent de reconstruire les solutions inconnues initiales et une classe de méthodes qui tentent de réconstruire des versions lisses des solutions inconnues. L'objectif de cette thèse est de contribuer à la promotion de la deuxième classe de méthode de régularisation à travers l'étude et l'application de la formulation variationnelle de la mollification.

Dans ce manuscrit, nous montrons que l'approche variationnelle de la mollification peut être étendue à la régularisation de problèmes mal-posés impliquant des opérateurs non compacts. À cet égard, nous étudions et appliquons avec succès la méthode à la régression instrumentale non-paramétrique.

Une contribution supplémentaire de cette thèse est la conception et l'étude d'une nouvelle méthode de régularisation adaptée aux problèmes linéaires exponentiellement mal-posés. Une comparaison numérique de cette nouvelle méthode aux méthodes classiques de régularisation telles que Tikhonov, la *spectral cut-off*, la régularisation asymptotique et la méthode des gradients conjugués est effectuée sur trois problèmes test tirés de la littérature. L'aspect pratique de la sélection du paramètre de régularisation avec un niveau de bruit inconnu est également considéré.

Outre l'étude et l'application des méthodes de régularisation, cette thèse traite également de l'application d'une règle de sélection de paramètres de régularisation très populaire connue sous le nom du principe de Morozov. En utilisant la dualité de Lagrange, nous fournissons un algorithme simple et rapide pour le calcul du paramètre de régularisation correspondant à cette règle pour les méthodes de régularisation du type Tikhonov.

L'intérêt de cette étude est qu'elle met en avant une méthode de régularisation mal connue qui pourtant a un grand potentiel et est capable de fournir des solutions approchées comparativement meilleures que certaines techniques de régularisation classiques bien connues. Un autre apport de cette thèse est la con-

ception d'une nouvelle méthode de régularisation qui, selon nous, est prometteuse dans la régularisation de problèmes exponentiellement mal-posés, en particulier pour les problèmes inverses de conduction thermique.

Abstract

Inverse problems is a fast growing area in applied mathematics which has gained a great attention in the last decades due to its ubiquity in several fields of sciences and technology. Yet, most often, inverse problems result in mathematical equation which are unstable. That is, the solutions do not continuously depend on the data. As a matter of fact, very little perturbations on the data might cause arbitrary large errors on the solutions. Therefore, given that the noise is generally unavoidable in the data, direct attempts to solve the problem fail and one needs to apply a regularization method in order to recover stable approximates of the unknown solutions. In this respect, several regularization techniques have been developed in the literature.

Globally, all these regularization methods can be split into two classes: A class of methods which attempt to reconstruct the unknown solutions and a class of methods which try to recover smooth versions of the unknown solutions. The aim of this thesis is to contribute to the promotion of the second class of regularization method via the study and application of the variational formulation of mollification.

In this work, we show that the variational approach can be extended to the regularization of ill-posed problems involving non-compact operators. In this respect, we study and successfully apply the method to a problem coming from statistics namely the nonparametric instrumental regression.

An additional contribution of this thesis is the design and study of a novel regularization method suitable for linear exponentially ill-posed problems. A numerical comparison of the new method to classical regularization methods such as Tikhonov, spectral cut-off, asymptotic regularization and conjugate gradient is carried out on three test problems from literature. The practical aspect of selection of the regularization parameter without knowledge of the noise level is also considered.

Apart from the study and application of regularization methods, this thesis also focuses on the application of a very popular parameter selection rule known as the Morozov principle. Using Lagrange duality, we provide a simple and rapid algorithm for the computation of the regularization parameter corresponding to this rule for Tikhonov-like regularization methods.

A relevance of this study is that it highlights a poorly known regularization method which yet has a great potential and is able to provide comparatively better approximate solutions compared to well-known classical regularization techniques. Another benefit of this thesis is the design of a new regularization method which, we believe, is promising in the regularization of exponentially ill-posed problems, especially for inverse heat conduction problems.

Contents

Notations

Ensemble de nombres

$\mathbb{R}$: L'ensemble des nombres réels

$\mathbb{R}_+$: L'ensemble des nombres réels positifs

$\mathbb{N}$: L'ensemble des nombres entiers naturels

$\mathbb{Z}$: L'ensemble des nombres entiers

$\mathbb{C}$: L'ensemble des nombres complexes

Espaces fonctionnels

$L^p(\mathbb{R}^p)$: Ensemble des fonctions f tel que $|f|^p$ est intégrable sur $\mathbb{R}^p$

$H^s(\mathbb{R}^p)$: L'espace de Sobolev des fonctions de $L^2(\mathbb{R}^p)$ dont toutes les dérivées au sens des distributions d'ordre inférieur à s sont dans $L^2(\mathbb{R}^p)$

$L^p(V)$: Sous espace des fonctions de $L^2(\mathbb{R}^p)$ à support dans le sous ensemble borné V de $\mathbb{R}^p$

$C^p(\Omega)$: Ensemble des fonctions p fois continument dérivables sur Ω

Opérateurs linéaires

$\mathcal{D}(T)$: Domaine de l'opérateur T

$\mathcal{R}(T)$: Image de l'opérateur T

$\mathcal{N}(T)$: Noyau de l'opérateur T

$T^\dagger$: Inverse généralisé de Moore-Penrose de l'opérateur T

$\sigma(T)$: Spectre d'un opérateur compact T

Ensemble des opérateurs linéaires

$\mathcal{B}(E, F)$: Ensemble des opérateurs linéaires continus entres les espaces vectoriels normés E et F

$\mathcal{B}(E) := \mathcal{B}(E, F)$

Transformations

$\mathcal{F}(f)$: Transformée de Fourier de la fonction f

$\hat{f}$: Transformée de Fourier de la fonction f

$f \star g$: Convolée des fonctions f et g

Chapter 1

Généralités sur les Problèmes Inverses et Mollification

Un problème inverse peut être considéré comme un problème dont la formulation implique implicitement (ou explicitement) celle d'un autre problème, appelé le problème direct. En général, étant donné deux problèmes dont la formulation de l'un implique l'autre, il n'est pas toujours évident de savoir lequel est considéré comme le problème direct (ou le problème inverse). Dans certains cas, le problème direct est le premier problème apparu dans la littérature, et le problème inverse est celui qui est apparu plus tard. Pourtant, ce n'est pas une règle générale. Si nous nous limitons aux problèmes du monde réel, une distinction naturelle entre les problèmes directs et inverses s'établit généralement. Par exemple, si l'on souhaite prédire les caractéristiques futures d'un système physique sur la base de l'état actuel (ou de l'état passé), on considérera qu'il s'agit du problème direct. Alors que le problème inverse consistera à déterminer l'état actuel (ou paramètre physique) à partir d'observations futures (par exemple, problèmes inverses de conduction thermique, identification de paramètres dans des équations aux dérivées partielles). Dans ce dernier cas, les applications comprennent la récupération d'informations passées à partir d'informations présentes, ainsi que la conception de contrôles qui permettent de diriger un système vers un état souhaité dans le futur.

Dans la plupart des cas, les problèmes inverses ne répondent pas aux trois critères de Hadamard qui définissent un problème bien-posé à savoir: l'existence et l'unicité des solutions pour toute donnée admissible et la continuité des solutions par rapport aux données. Dans ces cas, le problème inverse est dit mal-posé. Rappelons que parmi les trois postulats de Hadamard, c'est l'échec du troisième, c'est-à-dire le manque de continuité des solutions par rapport aux données, qui caractérise généralement la mal-position. En effet, contrairement au troisième postulat de continuité des solutions par rapport aux données, l'existence et l'unicité des solutions peuvent être renforcées par des hypothèses supplémentaires sur l'espace des solutions.

Ce chapitre se décompose en deux parties majeures: Une première partie dans laquelle on aborde les généralités sur les problèmes inverses et leur régularisation et une autre qui traite exclusivement de l'état de l'art de la mollification au début de ma thèse. Enfin dans un dernier temps, nous présentons le plan du manuscrit.

1.1 Généralités sur les Problèmes Inverses et leur Régularisation

Cette partie est organisée comme suit. Dans la section 1.1.1, nous décrivons quelques exemples de problèmes linéaires mal-posés. On illustre la discontinuité des solutions par rapport aux données et enfin on fournit quelques domaines d'application. La section 1.1.2 est dédiée à un rappel général concernant les équations linéaires mal-posés, la caractérisation de la mal-position et un bref résumé sur la théorie spectrale des opérateurs auto-adjoints. La section 1.1.3 traite des méthodes de régularisation. Dans cette section, nous passons en revue la définition d'une méthode de régularisation, illustrons une façon de construire des schémas de régularisation et décrivons quelques méthodes de régularisation classiques. La section 1.1.4 est consacrée à un résumé sur les règles de sélection de paramètres de régularisation. Nous y présentons une caractérisation de la convergence des règles de choix de paramètres et nous décrivons certaines règles de sélection de paramètres très répandues.

1.1.1 Exemples de problèmes mal-posés

Les problèmes inverses mal-posés sont omniprésents dans les sciences et l'ingénierie et le spectre de leurs applications couvre les domaines du traitement de signaux et d'images, de la médecine, de la théorie du potentiel, de l'hydraulique, des statistiques, de l'économétrie, de l'astronomie et de l'archéologie (voir, e.g. [4, 31, 32, 33, 43]). Rappelons que le plus souvent, les problèmes mal-posés sont liés à une équation intégrale de prémière espèce. Par exemple, dans la prospection géologique, la stéréographie stellaire, l'immunologie, la spectroscopie de Fourier, la radiothérapie, le profilage atmosphérique, le modèle physique conduit à une équation intégrale de première espèce mal-posée (voir, e.g. [43, Section 2]).

Dans cette section, nous passons brièvement en revue quatre problèmes inverses linéaires mal-posés: la différenciation numérique, l'équation rétrograde de la chaleur, la synthèse de Fourier importante et la régression instrumentale non-paramétrique.

Outre ces quatre problèmes que nous décrivons ci-dessous, il existe plusieurs autres exemples intéressants de problèmes linéaires mal-posés parmi lesquels la tomographie informatisée [42, 81, 101], le traitement des images [69], la gradiométrie par satellite [36, 106]. Par ailleurs, il existe de nombreux autres problèmes inverses mal-posés qui ne sont pas linéaires. On peut citer entre autre le problème de la gravimétrie [96], les problèmes d'identification de paramètres [8, 92], le problème de diffusion inverse [65, 73] et la tomographie à impédance électrique non linéaire [58]. Pour une discussion générale sur les problèmes mal-posés, voir par exemple les livres [31, 32, 43, 68] et les références qui y figurent.

1.1.1.1 Différenciation numérique

La différenciation numérique (voir, e.g. [3, 57, 89, 90]) est un problème omniprésent en analyse numérique qui consiste à calculer la dérivé g' d'une fonction $g \in C^1(I)$ où I est un intervalle de la droite réelle. Une caractéristique particulière de la différenciation est l'amplification des composantes de hautes fréquences qui induit la mal-position du problème. En effet, considérons la suite de fonctions $(g_n)_n$ définies par

$$g_n(x) = \epsilon \cos(nx/\epsilon).$$

Il est facile de constater que la suite $(g_n)_n$ est uniformément bornée par ϵ. Cependant,

$$g_n'(x) = -n \sin(nx/\epsilon),$$

ce qui implique que la norme infini $\|g_n'\|_\infty$ de g_n' explose lorsque n tend vers l'infini. Par conséquent, si nous perturbons une fonction $g \in C^1(I)$ par une bruit contenant des fréquences élevées, alors l'erreur commise dans le calcul de la dérivé g' devient arbitrairement importante. C'est-à-dire qu'une petite perturbation dans les données g peut induire une erreur très importante dans le calcul de la solution g'. Ceci est l'expression de la mal-position de la différenciation. En notant que la dérivé g' de g n'est rien d'autre que la solution f de l'équation

$$\int_{-\infty}^{x} f(t)\,\mathrm{d}t = g(x), \tag{1.1}$$

nous pouvons voir, que le problème direct correspondant à la différenciation est l'intégration. De plus, la mal-position de la différenciation est en fait liée aux propriétés de lissage de l'opérateur d'intégration. En effet, l'opérateur d'intégration amortit les composantes de hautes fréquences (par exemple, transforme e^{inx} en $\frac{1}{n}e^{inx}$) de sorte que les effets des composantes à haute fréquence sont annihilés par l'intégration. Cette propriété de lissage est en fait responsable de la mal-position de la différenciation.

Par ailleurs, notons qu'au regard de l'équation (1.1), nous pouvons également considérer la dérivé g' de g comme la solution d'une équation intégrale de première espèce de noyau $k(x,t) = 1_{(-\infty,x]}(t)$ où 1_S dénote la fonction indicatrice de l'ensemble S.

1.1.1.2 Équation rétrograde de la chaleur

L'équation rétrograde de la chaleur (voir, e.g. [97]) vise à récupérer la température initiale d'un corps étant donné la température à un moment ultérieur. Plus précisément, soit Ω un domaine régulier de $\mathbb{R}^p$ avec $p \leq 3$ et $u : \Omega \times (0,\bar{t}) \to \mathbb{R}$ la solution au problème à valeur initiale et aux conditions limites

$$\begin{cases} \frac{\partial u}{\partial t} = \Delta u, & \Omega \times (0,\bar{t}) \\ u(\cdot,0) = f, & \Omega \\ u = 0 \quad \text{or} \quad \frac{\partial u}{\partial \nu} = 0, & \text{on} \quad \partial\Omega \times (0,\bar{t}] \end{cases} \tag{1.2}$$

où tous les paramètres physiques sont normalisés à 1 afin de simplifier le modèle. On rappelle que dans (1.2), Δ est l'opérateur de Laplace qui va de $H^2(\Omega)$ dans $L^2(\Omega)$. Supposons que nous voulions récupérer la température initiale $f \in L^2(\Omega)$ à partir d'une température ultérieure $u(\cdot,\bar{t})$. En interprétant l'équation (1.2) comme une équation différentielle ordinaire pour la fonction $U : [0,\bar{t}] \to \mathcal{D}(\Delta) \subset L^2(\Omega)$, $t \to U(t) := u(\cdot,t)$, avec la valeur initiale $U(0) = f$ où

$$\mathcal{D}(\Delta) = H^2(\Omega) \cap H_0^1(\Omega) \quad \text{ou} \quad \mathcal{D}(\Delta) = \left\{ f \in H^2(\Omega), \quad \frac{\partial f}{\partial \nu} = 0 \ \text{ on } \ \partial\Omega \right\},$$

(selon que l'on considère les conditions limites de Dirichlet ou de Neumann), on obtient que $U(t) = \exp(t\Delta)f$ pour $t \in (0,\bar{t}]$, où $(\exp(t\Delta))_{t>0}$ est le semi-groupe fortement continu généré par l'opérateur linéaire auto-adjoint non-borné Δ (voir, e.g. [20, Chapitre 7]). Cela implique que l'équation satisfaite par la température initiale f n'est rien d'autre que

$$\exp(\bar{t}\Delta)f = u(\cdot,\bar{t}), \tag{1.3}$$

ce qui implique que $f = \exp(-\bar{t}\Delta)u(\cdot,\bar{t})$. En remarquant que le laplacien est dissipatif (et non-borné) sur $\mathcal{D}(\Delta)$ c'est-à-dire

$$\forall f \in \mathcal{D}(\Delta), \quad \langle \Delta f, f \rangle \leq 0,$$

on déduit que l'opérateur $\exp(-\bar{t}\Delta)$ est non-borné d'où la mal-position du problème d'identification de f à partir de $u(\cdot,\bar{t})$.

Une fois de plus, la mal-position du problème de calcul de f à partir de $u(\cdot,\bar{t})$ provient des fortes propriétés de lissage de l'opérateur direct $\exp(\bar{t}\Delta)$ dans (1.3). Là encore, l'opérateur direct $\exp(\bar{t}\Delta)$ annihile les composantes de hautes fréquences. En effet, soient $(\lambda_k)_k$ et $(\phi_k)_k$ les valeurs propres et fonctions propres normalisés du problème de Dirichlet ou de Neumann, c'est-à-dire que pour tout $k \in \mathbb{N}$, $\|\phi_k\|_{L^2(\Omega)} = 1$ et

$$\begin{cases} \Delta\phi_k + \lambda_k\phi_k = 0 \quad \text{dans} \quad \Omega \\ \phi_k = 0 \quad \text{ou} \quad \frac{\partial\phi_k}{\partial\nu} = 0 \quad \text{sur} \quad \partial\Omega. \end{cases}$$

En désignant par f_k les composantes de la température initiale f, c'est-à-dire $f = \sum_k f_k\phi_k$, on voit que l'expression de la température finale est donnée par

$$u(\cdot,\bar{t}) = \sum_k f_k e^{-\lambda_k\bar{t}}\,\phi_k. \tag{1.4}$$

De (1.4), depuis $\lambda_k \to +\infty$, on déduit que l'opérateur direct $\exp(\bar{t}\Delta)$ annihile assez rapidement les composantes de hautes fréquences de la température initiale f. Ceci induit la mal-position sévère de l'équation de la chaleur rétrograde.

Dans des cas spécifiques, l'équation de chaleur rétrograde peut être formulée comme une équation intégrale de première espèce en utilisant les fonctions de Green. En outre, il convient de noter que l'équation de la chaleur rétrograde a une application dans l'archéologie thermique (voir, e.g. [43, Section 2]).

1.1.1.3 Synthèse de Fourier

Le problème de la synthèse de Fourier (voir, e.g. [1]) consiste à récupérer une fonction f à partir d'une connaissance inexacte et incomplète de sa transformée de Fourier $\mathcal{F}(f)$ sur un domaine borné. Plus précisément, soit W un domaine borné de $\mathbb{R}^p$ et T_W l'opérateur défini par $T_W : L^2(\mathbb{R}^p) \to L^2(\mathbb{R}^p), f \mapsto 1_W\mathcal{F}(f)$. Nous cherchons à identifier la solution f de l'équation $T_W f = g$ où g est une fonction de carré intégrable à support dans W.

La définition de l'opérateur T_W illustre sa propriété de lissage qui consiste à annuler les composantes de hautes fréquences en dehors du domaine borné W. Cette propriété de lissage induit la mal-position du problème qui consiste à récupérer f à partir de la connaissance de $\mathcal{F}(f)$ sur W. En effet, supposons que $p = 1$ et $W = [-R, R]$ avec $R > 0$, définissez la suite

$$g_n(\xi) = 1_{[-R,R]}\ \mathrm{sinc}\left(\frac{\xi}{n}\right).$$

Étant donné que $\mathcal{F}(1_{[-\frac{1}{2n},\frac{1}{2n}]})(\xi) = \frac{1}{n}\,\mathrm{sinc}\left(\frac{\xi}{n}\right)$, on en déduit que

$$f_n(x) = n1_{[-\frac{1}{2n},\frac{1}{2n}]}(x),$$

est la solution de l'équation $T_W f = g_n$. Cependant, la suite $(g_n)_n$ est bornée alors que la suite $(f_n)_n$ diverge. En effet

$$\|g_n\|^2_{L^2(\mathbb{R})} = \int_{-R}^{R} \mathrm{sinc}^2\left(\frac{\xi}{n}\right)\,\mathrm{d}\xi = n\int_{-R/n}^{R/n} \mathrm{sinc}^2(\xi)\,\mathrm{d}\xi \le n\int_{-R/n}^{R/n}\mathrm{d}\xi = 2R,$$

tandis que

$$\|f_n\|^2_{L^2(\mathbb{R})} = n^2 \int_{-1/2n}^{1/2n} \mathrm{d}x = n \to \infty.$$

Par ailleurs il convient de noter que le problème de synthèse de Fourier peut également s'écrire sous la forme de l'équation intégrale de première espèce :

$$\int_{\mathbb{R}^p} 1_W(\xi) e^{-2\pi i \langle \xi, x \rangle} f(x)\, \mathrm{d}x = g(\xi).$$

Enfin, notons que la synthèse de Fourier a plusieurs applications parmi lesquelles, *aperture synthesis* [70], l'analyse du signal [18], la tomographie [42, 81, 84] et plus généralement a une grande importance dans le domaine du traitement de signal et d'images.

1.1.1.4 Régréssion instrumentale non paramétrique

En statistique, on est souvent confronté au problème de régression qui consiste à évaluer la relation entre une variable d'intérêt et un ensemble de variables explicatives. Plus précisément, soient $Z\colon \Omega \to \mathbb{R}^p$ et $Y, \varepsilon\colon \Omega \to \mathbb{R}$ des variables aléatoires continues satisfaisant la relation

$$Y = h(Z) + \varepsilon.$$

Ici, Z est la variable explicative, Y est la réponse, ε est l'erreur et $h\colon \mathbb{R}^p \to \mathbb{R}$ est une fonction inconnue appelée fonction de régression. Ce problème (voir, e.g. [21, 24, 34, 35]) est standard en économétrie et en statistique, et de nombreuses applications peuvent être trouvées dans divers domaines comme l'économie du travail, la bio-statistique, la microéconomie et la demande des consommateurs.

Nous souhaitons identifier h, en partant du principe que Z et ε sont liés: $E(\varepsilon|Z) \neq 0$, où $E(\varepsilon|Z)$ dénote l'espérance conditionnelle de ε suivant Z. Notons que $E(\varepsilon|Z) = 0$ si et seulement si $E(Y|Z) = E(h(Z)|Z) = h(Z)$, dans un tel cas, h est caractérisé par $h(Z) = E(Y|Z)$. Dans le cas où $E(\varepsilon|Z) \neq 0$, la fonction h ne peut pas être identifiée comme précédemment et il est alors d'usage d'introduire une variable instrumentale $W\colon \Omega \to \mathbb{R}^q$, qui est liée à Z et qui est indépendante de ε. Le modèle s'écrit alors $Y = h(Z) + \varepsilon,\ E(\varepsilon|W) = 0$ et l'équation $E(\varepsilon|W) = 0$ implique que

$$E(Y|W) = E(h(Z)|W). \tag{1.5}$$

Si toutes les variables Y, Z, W sont absolument continues par rapport à la mesure de Lebesgue (notée λ quelle que soit la dimension), on peut définir les densités respectives f_Y, f_Z, f_W. Soient f_{YW} (respectivement f_{ZW}) la densité conjointe de (Y, W) (respectivement (Z, W)) et $f_{Z|W}$ est la densité conditionnelle de Z étant donné W. L'équation (1.5) se réduit à l'équation fonctionnelle intégrale

$$\int \frac{f_{YW}(y,w)}{f_W(w)} y\, \mathrm{d}y = \int \frac{f_{ZW}(z,w)}{f_W(w)} h(z)\, \mathrm{d}z, \quad \text{pour presque tout } w \in \{\omega \in \mathbb{R}^q | f_W(\omega) \neq 0\},$$

qui est équivalent à

$$\int f_{YW}(y,w) y\, \mathrm{d}y = \int f_{ZW}(z,w) h(z)\, \mathrm{d}z, \quad \text{pour presque tout } w \in \{\omega \in \mathbb{R}^q | f_W(\omega) \neq 0\}.$$

Définissons l'opérateur intégral linéaire:

$$\begin{aligned} T\colon \quad L^2(\mathbb{R}^p) &\longrightarrow L^2(\mathbb{R}^q) \\ h &\longmapsto \int f_{ZW}(z,\cdot) h(z)\, \mathrm{d}z = \int k(z,\cdot) h(z)\, \mathrm{d}z, \end{aligned}$$

où la fonction k désigne le noyau de T et la fonction $r := \int f_{YW}(y,\cdot)y\,\mathrm{d}y$. En supposant que r appartient à $L^2(\mathbb{R}^q)$ (ce qui est vrai si $E[Y^2] < \infty$ et f_W est borné), l'équation (1.5) prend la forme concise :

$$Th = r. \tag{1.6}$$

Dans l'hypothèse où la densité f_{ZW} (équivalent au noyau k) est de carré $(\lambda \otimes \lambda)$-intégrable, l'opérateur T est un opérateur de Hilbert-Schmidt. Il s'ensuit que T est compact, et que son pseudo-inverse généralisé $T^\dagger$ est non-borné, ce qui implique la mal-position du problème.

1.1.2 Rappel sur les problèmes linéaires mal-posés

Dans cette section, nous décrivons le cadre des équations à opérateur linéaire, donnons une caractérisation de la mal-position et rappelons quelques notions importantes de la théorie spectrale qui seront utiles dans la suite.

Considérons l'équation

$$Tf = g, \tag{1.7}$$

où $T : H_1 \to H_2$ est un opérateur linéaire borné entre deux espaces de Hilbert de dimension infinie H_1 et H_2, $f \in H_1$ est l'inconnu et $g \in H_2$ est la donnée. Pour simplifier les notations, les normes (resp. les produits scalaires) sur les espaces de Hilbert H_1 et H_2 seront désignées par $||\cdot||$ (resp. $\langle\cdot,\cdot\rangle$) sans les indices H_1 et H_2.

Au sens strict, l'équation (1.7) admet une solution $f \in H_1$ si et seulement si les données g sont atteignables, c'est-à-dire si g appartient à l'image $\mathcal{R}(T)$ de T. Étant donné que la condition $g \in \mathcal{R}(T)$ peut être très restrictive dans de nombreux cas, ci-après, la notion de solution de l'équation (1.7) sera assimilée à une solution (si elle existe) du problème des moindres carrés

$$\min_{f\in H_1} \left\|Tf - g\right\|^2. \tag{1.8}$$

Definition 1.1. *Soit g une donnée dans H_2.*

- *Un élément $\bar{f}$ est appelé une solution (de moindres carrés) de l'équation (1.7) si*

$$\left\|T\bar{f} - g\right\| = \min_{f\in H_1} \left\|Tf - g\right\|.$$

- *Un élément $f^* \in H_1$ est appelé meilleure solution approchée de l'équation (1.7) si f^* est la solution (de moins carrée) de norme minimale, c'est-à-dire*

$$\begin{cases} \left\|Tf^* - g\right\| = \min_{f\in H_1} \left\|Tf - g\right\| \\ \left\|f^*\right\| = \min\left\{\left\|\bar{f}\right\|, \quad \bar{f} \in H_1, \quad \left\|T\bar{f} - g\right\| = \min_{f\in H_1} \left\|Tf - g\right\|\right\}. \end{cases}$$

Dans la suite, $f^\dagger$ désigne la meilleure solution approchée de l'équation (1.7). Étant donné $f^\dagger$, l'ensemble des solutions (de moins carrées) de l'équation (1.7) est défini par $f^\dagger + \mathcal{N}(T)$, où $\mathcal{N}(T)$ est le noyau de l'opérateur T. Notons que, en utilisant la condition d'optimalité de premier ordre, le problème des moindres carrés (1.8) est équivalent à l'équation normale

$$T^*Tf = T^*g, \tag{1.9}$$

où T^* désigne l'adjoint de l'opérateur T. Contrairement à l'équation (1.7) qui n'admet de solution f que si $g \in \mathcal{R}(T)$, l'équation des moindres carrés (1.8) (ou l'équation normale (1.9)) admet toujours une solution f pour une donnée g appartenant au sous-espace dense $\mathcal{R}(T) + \mathcal{R}(T)^\perp$ de H_2, où $\mathcal{R}(T)^\perp$ dénote l'orthogonale de l'image de T. Nous allons maintenant voir que la solution de moindre norme $f^\dagger$ peut être définie via un opérateur dénommée inverse (généralisé) de Moore-Penrose de T.

Definition 1.2. *Soit $\bar{T} := T|_{\mathcal{N}(T)^\perp} : \mathcal{N}(T)^\perp \to \mathcal{R}(T)$ la restriction de T à l'orthogonale du noyau de T. Alors $\bar{T}$ est une bijection, de sorte que $\bar{T}^{-1}$ est bien défini. L'inverse (généralisé) de Moore-Penrose $T^\dagger$ de T est le prolongement de $\bar{T}^{-1}$ à $\mathcal{R}(T) + \mathcal{R}(T)^\perp$ défini par*

$$\begin{aligned} T^\dagger : \quad \mathcal{R}(T) + \mathcal{R}(T)^\perp \subseteq H_2 \quad &\longrightarrow \quad H_1 \\ Tf + \tilde{g} \quad &\longmapsto \quad \bar{T}^{-1} T f = f, \end{aligned} \tag{1.10}$$

où $\tilde{g}$ désigne un élément de $\mathcal{R}(T)^\perp$.

Avec la définition 1.2, la solution de moindre norme $f^\dagger$ de l'équation 1.7 peut être définie comme $f^\dagger = T^\dagger g$. De plus, $f^\dagger$ existe et est unique si et seulement si g appartient au domaine de $T^\dagger$ qui est $\mathcal{D}(T^\dagger) = \mathcal{R}(T) + \mathcal{R}(T)^\perp$.

1.1.2.1 Caractérisation de la mal-position

Après avoir donné des détails sur l'existence et la définition des solutions de l'équation (1.7), abordons à présent l'aspect le plus important dans la résolution de l'équation (1.7) : la continuité de f par rapport à g.

Proposition 1.1. *L'inverse (généralisé) de Moore-Penrose $T^\dagger$ de T est borné (i.e. continu) si et seulement si l'image $\mathcal{R}(T)$ de T est fermée.*

La preuve que la Proposition 1.1 repose sur le théorème du graphe fermé et le fait que $T^\dagger$ a un graphe fermé (voir, e.g. [32, Proposition 2.4]).

La Proposition 1.1 implique que l'équation (1.7) est mal-posée si $\mathcal{R}(T)$ n'est pas fermé. Outre la non-fermeture de l'image de T, la mal-position de l'équation (1.7) peut aussi être caractérisée par une accumulation du spectre de T en 0 comme l'énonce la proposition suivante dont la preuve est presentée en Annexe.

Proposition 1.2. *L'équation (1.7) est mal-posée si et seulement si*

$$\inf_{\substack{\|f\|=1 \\ f \in \mathcal{N}(T)^\perp}} \|Tf\|^2 = 0. \tag{1.11}$$

Notons que dans la plupart des cas, la caractérisation (1.11) est plus facile à vérifier par rapport à l'inspection si l'image de T est fermée ou non.

Une classe spéciale de problèmes inverses mal-posés est constituée des problèmes inverses associés à des opérateurs compacts T. En effet, supposons que T soit compact, alors il existe une décomposition en valeurs singulières (σ_k, u_k, v_k) de T (voir, e.g. [20, Chapitre 6]), c'est-à-dire :

- la suite $(f_k)_k$ forme une base de Hilbert de H_1,
- la suite $(g_k)_k$ forme une base de Hilbert de la fermeture de l'image de T,

- la suite $(\sigma)_k$ est une suite positive décroissante convergeant vers 0 (ou s'accumulant à 0) et satisfait $Tf_k = \sigma_k g_k$ et $T^* g_k = \sigma_k f_k$.

Ainsi on déduit (1.11) d'où la mal-position. Notons que, étant donné la décomposition en valeurs singulières de T, nous pouvons caractériser la condition d'admissibilité $g \in \mathcal{D}(T^\dagger)$ et définir les opérateurs T, T^*T et l'inverse généralisé $T^\dagger$ de T comme suit (voir, e.g. [32, Theoreme 2.8]):

Proposition 1.3. *Soit (σ_k, f_k, g_k) la décomposition en valeurs singulières d'un opérateur linéaire compact T et $g \in H_2$, alors*

- *$g \in \mathcal{D}(T^\dagger)$ si et seulement si*
$$\sum_{k=1}^{\infty} \frac{|\langle g, g_k \rangle|^2}{\sigma_k^2} < \infty,$$
- *Étant donné $f \in H_1$,*
$$Tf = \sum_{k=1}^{\infty} \sigma_k \langle f, f_k \rangle g_k \quad et \quad T^*Tf = \sum_{k=1}^{\infty} \sigma_k^2 \langle f, f_k \rangle f_k,$$
- *Étant donné $g \in \mathcal{D}(T^\dagger)$,*
$$T^\dagger g = \sum_{k=1}^{\infty} \frac{\langle g, g_k \rangle}{\sigma_k} f_k. \tag{1.12}$$

De (1.12), on peut facilement voir pourquoi l'accumulation du spectre autour de 0 induit la mal-position de l'équation (1.7). En effet, à partir de (1.12), on obtient que $T^\dagger g_k$ est f_k/σ_k qui explose lorsque k va à l'infini bien que $||g_k||$ reste égal à 1. Rappelons qu'une classe importante d'opérateurs compacts sont les opérateurs de Hilbert-Schmidt dont les exemples bien connues sont les opérateurs intégraux avec noyau à carré intégrable.

1.1.2.2 Rappel sur la théorie spectrale

Dans cette section, nous rappelons quelques définitions et propriétés importantes de la théorie spectrale des opérateurs auto-adjoints, voir, e.g. [27, Section 7],[25, 60],[32, Section 2.3].

Soit $h : (0, \sigma_1] \to \mathbb{R}$ une fonction continue à valeur réelle. En utilisant la décomposition en valeurs singulières $(\sigma_k, f_k, g_k)_{k\in\mathbb{N}}$ d'un opérateur compact T, nous pouvons appliquer la fonction h à l'opérateur auto-adjoint T^*T et TT^* et ainsi obtenir de nouveaux opérateurs $h(T^*T)$ et $h(TT^*)$ définis par

$$\begin{cases} \forall f \in \mathcal{D}(h(T^*T)) \subseteq H_1, & h(T^*T)f = \sum_{k=1}^{\infty} h(\sigma_k^2)\langle f, f_k \rangle f_k, \\ \forall g \in \mathcal{D}(h(TT^*)) \subseteq H_2, & h(TT^*)g = \sum_{k=1}^{\infty} h(\sigma_k^2)\langle g, g_k \rangle g_k, \end{cases} \tag{1.13}$$

où la fonction h est appliquée uniquement aux valeurs propres de T^*T (ou TT^*). Rappelons que les domaines $\mathcal{D}(h(T^*T))$ (resp. $\mathcal{D}(h(TT^*))$) de l'opérateur $h(T^*T)$ (resp. $h(TT^*)$) correspondent en fait au sous-espace des fonctions f (resp. g) tel que la première (resp. la deuxième) série de (1.13) converge dans H_1 (resp. H_2). À titre d'illustration, en considérant h comme la fonction identité sur H_1 (resp. H_2), on retrouve bien dans les séries de (1.13) des résultats classiques de caractérisation d'opérateurs compacts auto-adjoints (voir, e.g. [20, Chapitre 6]). En outre, en utilisant la densité des polynômes dans l'espace

des fonctions continues sur un intervalle borné, on peut aisément vérifier que les propriétés suivantes sont valides

$$h(T^*T)T^* = T^*h(TT^*), \quad \text{et} \quad Th(T^*T) = h(TT^*)T^*. \tag{1.14}$$

Il est important de noter que la définition (1.13) est un cas particulier d'une définition générale qui est valable pour les opérateurs auto-adjoints. Dans le cas général, la décomposition en valeurs singulières de T est remplacée par la notion de famille spectrale qui existe pour tout opérateur auto-adjoint.

Definition 1.3. *Une famille de projecteurs orthogonaux $\{E_\lambda\}_{\lambda\in\mathbb{R}}$ sur un espace de Hilbert H_1 est appelée famille spectrale si elle satisfait les conditions suivantes:*

- $E_{\lambda_1}E_{\lambda_2} = E_{\min(\lambda_1,\lambda_2)}$, $\forall\lambda_1,\lambda_2 \in \mathbb{R}$,
- $\forall f \in H_1$, $\lim_{\lambda\to-\infty} E_\lambda f = 0$ *et* $\lim_{\lambda\to+\infty} E_\lambda f = f$
- $\forall f \in H_1$, $\lim_{\epsilon\to0} E_{\lambda-\epsilon} f = E_\lambda f$

Étant donnée une famille spectrale $\{E_\lambda\}_{\lambda\in\mathbb{R}}$ et une fonction continue $h : \mathbb{R} \to \mathbb{R}$ et $f \in H_1$, on définit l'intégrale $\int_a^b h(\lambda)\,\mathrm{d}\,E_\lambda f$ comme la limite de la somme de Riemann :

$$\sum_{k=1}^{n} h(\xi_k)(E_{\lambda_{k+1}} - E_{\lambda_k})f, \tag{1.15}$$

quand $\max_{1\le k\le n} |\lambda_{k+1} - \lambda_k| \to 0$ où $-\infty < a = \lambda_1 < \cdots < \lambda_{n+1} = b < \infty$, et $\xi_k \in (\lambda_k, \lambda_{k+1}]$. La Proposition 2.11 de [32] prouve l'existence de la limite de la somme (1.15).

Ayant défini l'intégrale $\int_a^b h(\lambda)\,\mathrm{d}E_\lambda f$, on peut alors définir l'intégrale

$$\int_{-\infty}^{+\infty} h(\lambda)\,\mathrm{d}\,E_\lambda f$$

comme la limite (si elle existe) de $\int_a^b h(\lambda)\,\mathrm{d}E_\lambda f$ quand a tends vers $-\infty$ et b tends vers $+\infty$. Pour plus de détails concernant la construction d'intégrale suivant une famille spectrale, voir, e.g. [27, Section 7.1b]

Proposition 1.4. *Soit A un opérateur linéaire auto-adjoint défini sur un espace de Hilbert H_1, alors il existe une famille spectrale $\{E_\lambda\}_{\lambda\in\mathbb{R}}$ telle que*

$$\forall f \in H_1, \quad Af = \int_{-\infty}^{+\infty} \lambda\,\mathrm{d}\,E_\lambda f \quad \textit{et} \quad ||Af||^2 = \int_{-\infty}^{+\infty} \lambda^2\,\mathrm{d}\,||E_\lambda f||^2.$$

Pour plus de détails sur la Proposition 1.4 et sa preuve voir, e.g. [27, Theorem 7.2.1]. Rappelons que l'intégrale suivant $||E_\lambda f||$ est définit de façon similaire à l'intégrale suivant $E_\lambda f$ en remplaçant $(E_{\lambda_{k+1}} - E_{\lambda_k})f$ par $(||E_{\lambda_{k+1}}f|| - ||E_{\lambda_k}f||)$ dans (1.15).

À titre illustratif, dans ce cas où $A = T^*T$ où $T : H_1 \to H_2$ est un opérateur compact et injectif, en notant par $(\sigma_k, f_k, g_k)_{k\in\mathbb{N}}$ la décomposition en valeurs singulières de T, alors on peut vérifier que E_λ est le projecteur orthogonal sur l'espace X_λ défini par

$$X_\lambda = \left\{ f_k \,\middle|\, k \in \mathbb{N}, \sigma_k^2 < \lambda \right\}, \quad i.e. \quad E_\lambda f = \sum_{\substack{k\in\mathbb{N} \\ \sigma_k^2<\lambda}} \langle f, f_k\rangle\, f_k.$$

On peut aussi vérifier que la famille $\{E_\lambda\}_{\lambda\in\mathbb{R}}$ respecte les propriétés suivantes:

- Pour tout $\lambda \leq 0$, $E_\lambda = 0$
- Pour tout $\lambda > \sigma_1^2$, $E_\lambda = I_{H_1}$ où I_{H_1} dénotes l'application identité sur H_1,
- E_λ est constant par morceaux et présente un saut en $\lambda = \sigma_k^2$ égale à $\sum_{\substack{k\in\mathbb{N} \\ \sigma_k^2=\lambda}} \langle \cdot, f_k \rangle f_k$
- $\forall \lambda_1 \leq \lambda_2$, et $\forall f \in H_1$, $\langle E_{\lambda_1} f, f \rangle \leq \langle E_{\lambda_2} f, f \rangle$,

Avec ces propriétés, pour $f \in H_1$, on peut alors définir l'intégrale suivant la mesure $E_\lambda f$ comme intégrale suivant une fonction constante par morceaux, et retrouver les définitions (1.13).

Dans la suite, $||T||_+^2$ désigne tout nombre strictement supérieur à $||T||^2$. Une fois défini la notion de famille spectrale et les intégrales associés, on peut alors définir une fonction continue appliqué à l'opérateur auto-adjoint T^*T (resp. TT^*) comme suit.

Soit $\{E_\lambda\}_\lambda$ (resp. $\{F_\lambda\}_\lambda$) la famille spectrale associée à l'opérateur auto-adjoint borné T^*T (resp. TT^*) et $h : (0, \sigma_1] \to \mathbb{R}$ une fonction continue à valeur réelle. Les opérateurs $h(T^*T)$ et $h(TT^*)$ se définissent par

$$\begin{cases} \forall f \in \mathcal{D}(h(T^*T)) \subseteq H_1, \quad h(T^*T)f = \int_0^{||T||_+^2} h(\lambda)\, \mathrm{d}\, E_\lambda f \\ \forall g \in \mathcal{D}(h(TT^*)) \subseteq H_2, \quad h(TT^*)g = \int_0^{||T||_+^2} h(\lambda)\, \mathrm{d}\, F_\lambda g, \end{cases} \tag{1.16}$$

où $\mathcal{D}(h(T^*T))$ et $\mathcal{D}(h(TT^*))$ sont définis par

$$\begin{cases} \mathcal{D}(h(T^*T)) = \left\{ f \in H_1, \quad \int_0^{||T||_+^2} h^2(\lambda)\, \mathrm{d}\, ||E_\lambda f||^2 < \infty \right\} \\ \\ \mathcal{D}(h(TT^*)) = \left\{ g \in H_2, \quad \int_0^{||T||_+^2} h^2(\lambda)\, \mathrm{d}\, ||F_\lambda g||^2 < \infty \right\}. \end{cases} \tag{1.17}$$

Comme illustration de (1.16), pour tout $k \in \mathbb{N}$,

$$(T^*T)^k f = \int_0^{||T||_+^2} \lambda^k\, \mathrm{d}\, E_\lambda f, \quad \text{et} \quad (TT^*)^k g = \int_0^{||T||_+^2} \lambda^k\, \mathrm{d}\, F_\lambda g.$$

Étant donné un élément $f_1 \in H_1$ (resp. $g_1 \in H_2$), le produit scalaire $\langle h(T^*T)f, f_1 \rangle$ (resp. $\langle h(TT^*)g, g_1 \rangle$) est défini par

$$\int_0^{||T||_+^2} h^2(\lambda)\, \mathrm{d}\, \langle E_\lambda f, f_1 \rangle \quad \left(\text{resp.} \quad \int_0^{||T||_+^2} h^2(\lambda)\, \mathrm{d}\, \langle F_\lambda g, g_1 \rangle \right),$$

où l'intégrale suivant $\langle E_\lambda f, f_1 \rangle$ (resp. $\langle F_\lambda g, g_1 \rangle$) est défini de façon similaire à l'intégrale suivant $E_\lambda f$ en remplaçant $(E_{\lambda_{k+1}} - E_{\lambda_k})f$ par $\langle E_{\lambda_{k+1}} f, f_1 \rangle - \langle E_{\lambda_k} f, f_1 \rangle$ (resp. $\langle F_{\lambda_{k+1}} g, g_1 \rangle - \langle F_{\lambda_k} g, g_1 \rangle$) dans (1.15). De façon similaire, étant donné $f \in \mathcal{D}(h(T^*T))$ et $g \in \mathcal{D}(h(TT^*))$, les normes $||h(T^*T)f||$ et $||h(TT^*)g||$ sont définies par

$$||h(T^*T)f|| = \int_0^{||T||_+^2} h^2(\lambda)\, \mathrm{d}||E_\lambda f||^2, \quad ||h(TT^*)g|| = \int_0^{||T||_+^2} h^2(\lambda)\, \mathrm{d}||F_\lambda g||^2. \tag{1.18}$$

Pour un opérateur non-compact T, étant donné que les propriétés (1.14) sont triviales pour les fonctions polynomiales h, et que les fonctions continues sont uniformément approchées par des fonctions polynomiales sur des domaines bornés, alors les propriétés (1.14) sont également valables pour les fonctions

continues h sur $(0, ||T||^2]$. À partir de (1.18), nous pouvons déduire une borne supérieure de la norme des opérateurs $h(T^*T)$ et $h(TT^*)$ comme suit

$$||h(T^*T)|| \leq \sup_{\lambda\in[0,||T||^2]} |h(\lambda)|, \quad \text{et} \quad ||h(TT^*)|| \leq \sup_{\lambda\in[0,||T||^2]} |h(\lambda)|. \tag{1.19}$$

En outre, en utilisant (1.14) et l'équation de l'adjoint, on a

$$||h(T^*T)T^*g||^2 = \langle h(TT^*)g, TT^*h(TT^*)g\rangle \ \text{ et } \ ||h(TT^*)Tf||^2 = \langle h(T^*T)f, T^*Th(T^*T)f\rangle\,. \tag{1.20}$$

À partir de (1.20), et (1.18), nous pouvons établir que

$$||h(T^*T)T^*|| \leq \sup_{\lambda\in[0,||T||^2]} \sqrt{\lambda}|h(\lambda)| \quad \text{et} \quad ||T^*h(TT^*)|| \leq \sup_{\lambda\in[0,||T||^2]} \sqrt{\lambda}|h(\lambda)|. \tag{1.21}$$

Une estimation intéressante basée sur l'équation de l'adjoint, les propriétés (1.14) et le fait que $T^*T = (T^*T)^{\frac{1}{2}}(T^*T)^{\frac{1}{2}}$ (resp. $TT^* = (TT^*)^{\frac{1}{2}}(TT^*)^{\frac{1}{2}}$) est la suivante :

$$||Th(T^*T)f|| = ||(T^*T)^{\frac{1}{2}}h(T^*T)f|| \quad \text{et} \quad ||h(TT^*)Tf|| = ||(T^*T)^{\frac{1}{2}}h(T^*T)f||. \tag{1.22}$$

Avant de terminer cette exposé sur les fonctions à valeur réelle appliquées à des opérateurs auto-adjoints, énonçons l'inégalité d'interpolation suivante (basée sur l'inégalité de Hölder)

$$\forall f \in \mathcal{D}((T^*T)^p), \ \text{ et } \ \forall\ q > p \geq 0, \quad ||(T^*T)^p f|| \leq ||(T^*T)^q f||^{\frac{p}{q}} ||f||^{1-\frac{p}{q}}. \tag{1.23}$$

1.1.3 Méthodes de régularisation classiques

Dans la pratique, afin d'approximer la solution $f^\dagger$ de l'équation (1.7), on ne dispose que d'une donnée bruité g^δ qui est lié à la donnée exacte g par la condition

$$||g - g^\delta|| \leq \delta. \tag{1.24}$$

Quand l'équation (1.7) est mal-posée, l'erreur dans les données g (c'est-à-dire $g - g^\delta$), aussi petite soit-elle, peut conduire à une erreur incontrôlable dans l'approximation de $f^\dagger$. Pour cette raison, une méthode de régularisation est essentielle pour récupérer une approximation raisonnable de $f^\dagger$ à partir de g^δ. Une méthode de régularisation est une famille d'opérateurs continus $(R_\alpha)_{\alpha>0}$ qui converge ponctuellement vers $T^\dagger$ quand α tend vers 0. Plus rigoureusement, on peut aussi donner la définition suivante.

Definition 1.4. *Soit $T : H_1 \to H_2$ un opérateur linéaire borné entre deux espaces de Hilbert H_1 et H_2. Une famille $(R_\alpha)_{\alpha>0}$ d'opérateur continu $R_\alpha : H_2 \to H_1$ est appelé une méthode de régularisation pour $T^\dagger$ si pour tout $g \in \mathcal{D}(T^\dagger)$, il existe une règle de choix de paramètre $\Lambda : \mathbb{R}_+ \times H_2 \to \mathbb{R}$, $(\delta, g^\delta) \mapsto \Lambda(\delta, g^\delta)$ telle que*

$$\lim_{\delta\to 0} \sup \left\{ \Lambda(\delta, g^\delta), \quad g^\delta \in H_2, \ ||g - g^\delta|| \leq \delta \right\} = 0, \tag{1.25}$$

et

$$\lim_{\delta\to 0} \sup \left\{ ||R_{\Lambda(\delta,g^\delta)}g^\delta - T^\dagger g||, \quad g^\delta \in H_2, \ ||g - g^\delta|| \leq \delta \right\} = 0. \tag{1.26}$$

La proposition suivante permet de réduire les conditions (1.25) et (1.26) à la simple condition de convergence ponctuelle de la famille $(R_\alpha)_{\alpha>0}$ vers $T^\dagger$ quand α tend vers 0.

Proposition 1.5. *Soit $(R_\alpha)_{\alpha>0}$ une famille d'opérateur continu de H_2 vers H_1. Si $(R_\alpha)_{\alpha>0}$ converge ponctuellement vers $T^\dagger$ quand α tend vers 0, c'est-à-dire*

$$\forall g \in \mathcal{D}(T^\dagger), \quad R_\alpha g \to T^\dagger g, \quad \textit{quand} \quad \alpha \to 0, \tag{1.27}$$

alors la famille $(R_\alpha)_{\alpha>0}$ est une méthode de régularisation pour $T^\dagger$. C'est-à-dire qu'il existe une règle de sélection des paramètres $\Lambda(\delta, y^\delta)$ satisfaisant (1.25) et (1.26).

Pour une preuve de la Proposition 1.5 voir, e.g. [32, Proposition 3.4]. Dans la suite, à chaque fois que l'on considérera l'équation (1.7), la famille $(R_\alpha)_{\alpha>0}$ sera simplement appelée méthode de régularisation et l'opérateur R_α sera appelé opérateur de régularisation. Il est important de noter que quand $T^\dagger$ est non-borné, lorsque α s'approche de 0, la norme de l'opérateur de régularisation R_α augmente et finit par exploser lorsque α tend vers 0, c'est-à-dire,

$$||R_\alpha|| \to +\infty, \quad \text{quand} \quad \alpha \to 0. \tag{1.28}$$

Notons que (1.28) et (1.27) impliquent l'existence d'une donnée $g \in H_2 \setminus \mathcal{D}(T^\dagger)$ telle que $||R_\alpha g|| \to \infty$ quand α tends vers 0. Si l'opérateur de régularisation R_α satisfait $\sup_{\alpha>0} ||TR_\alpha|| < \infty$, alors l'ensemble des g tel que $||R_\alpha g|| \to +\infty$ est exactement le complément du domaine de $T^\dagger$ (voir [32, Proposition 3.6]), c'est-à-dire

$$\sup_{\alpha>0} ||TR_\alpha|| < \infty \quad \Rightarrow \quad \forall g \in H_2 \setminus \mathcal{D}(T^\dagger), \quad ||R_\alpha g|| \to +\infty. \tag{1.29}$$

En utilisant la Proposition 1.1, on peut vérifier que l'ensemble $H_2 \setminus \mathcal{D}(T^\dagger)$ est non vide pour tout opérateur T tel que $T^\dagger$ est non-borné.

Grâce à Bakushinskii [6], en utilisant la théorie spectrale des opérateurs linéaires auto-adjoints rappelée dans la section 1.1.2.2, on peut définir une méthode générale de construction d'opérateurs de régularisation basée sur une famille de fonctions à valeur réelle $(\upsilon_\alpha)_{\alpha>0}$.

Proposition 1.6. *Considérons une famille de fonctions à valeur réelle continue par morceaux $(\upsilon_\alpha)_{\alpha>0}$ telle que $\upsilon_\alpha : [0, \|T\|^2] \to \mathbb{R}$. Supposons que*

$$\forall \lambda \in (0, ||T||^2], \quad \upsilon_\alpha(\lambda) \to 1/\lambda \quad \textit{quand} \quad \alpha \to 0, \tag{1.30}$$

et qu'il existe une constante positive C telle que

$$\forall \lambda \in (0, ||T||^2], \ \forall \alpha > 0, \quad |\lambda \upsilon_\alpha(\lambda)| \leq C. \tag{1.31}$$

Alors

$$\forall g \in \mathcal{D}(T^\dagger), \quad \upsilon_\alpha(T^*T)T^*g \to T^\dagger g, \quad \textit{quand} \quad \alpha \to 0. \tag{1.32}$$

De plus, pour tout $g \notin \mathcal{D}(T^\dagger)$, $||\upsilon_\alpha(T^*T)T^*g|| \to +\infty$ quand α tend vers 0.

Pour une preuve de la Proposition 1.6, voir, e.g. [32, Theoreme 4.1]. À partir de (1.32) et la Proposition 1.5, nous déduisons que si la famille de fonctions continues par morceaux à valeur réelle $(\upsilon_\alpha)_{\alpha>0}$ satisfait (1.30) et (1.31), alors cette famille définit un opérateur de régularisation $R_\alpha : H_2 \to H_1$ donné par

$$R_\alpha := \upsilon_\alpha(T^*T)T^*, \tag{1.33}$$

où l'opérateur $\upsilon_\alpha(T^*T)$ dans (1.33) est défini via la famille spectrale $(E_\lambda)_\lambda$ de l'opérateur auto-adjoint T^*T.

Dans la suite, le terme *fonction génératrice* sera utilisé pour désigner une fonction υ_α continue par morceaux à valeur réelle vérifiant (1.30) et (1.31). Une fonction très importante dans l'étude d'une méthode de régularisation définie via une *fonction génératrice* υ_α est la fonction *résiduelle* $r_\alpha : [0, ||T||^2] \to \mathbb{R}$ définie par

$$r_\alpha(\lambda) = 1 - \lambda\upsilon_\alpha(\lambda). \tag{1.34}$$

En effet, soit R_α un opérateur de régularisation défini par (1.33), à l'aide de la fonction r_α, on peut évaluer l'erreur de régularisation $T^\dagger g - R_\alpha g$ et aussi $g - TR_\alpha g$ comme suit

$$\begin{cases} f^\dagger - f_\alpha = (I - \upsilon_\alpha(T^*T)T^*T)f^\dagger = r_\alpha(T^*T)f^\dagger \\ g - Tf_\alpha = (I - T\upsilon_\alpha(T^*T)T^*)g = r_\alpha(TT^*)g, \end{cases} \tag{1.35}$$

où $f^\dagger = T^\dagger g$ et $f_\alpha = R_\alpha g$. Enfin, notons que de nombreuses méthodes de régularisation classiques peuvent être reformulées via une *fonction génératrice* υ_α.

Dans les sections suivantes, nous passons en revue certaines méthodes de régularisation classiques, à savoir la méthode Tikhonov, de Landweber, des gradients conjugués et de Showalter. Outre ces méthodes, il existe d'autres méthodes de régularisation telles que la *spectral cut-off*, la méthode ν [48, 50] et la méthode itérative régularisée de Gauss-Newton [5, 15, 64].

1.1.3.1 Méthode de Tikhonov

La méthode de Tikhonov (voir, e.g. [79, 112, 110, 111]) est probablement la méthode la plus répandue pour la régularisation de problèmes mal-posés. La version ordinaire de la méthode de Tikhonov approxime la solution $f^\dagger$ de l'équation (1.7) ou (1.8) par le minimiseur f_α de la fonctionnelle

$$\mathcal{T}_\alpha(f) = \|Tf - g\|^2 + \alpha\|f\|^2. \tag{1.36}$$

D'après la définition de f_α, on voit que la stabilité est introduite en pénalisant simplement la norme de f et la continuité de f_α par rapport aux données g découle trivialement de l'inégalité $\|f_\alpha\| \leq (1/\sqrt{\alpha})\|g\|$ (vue que $\mathcal{T}_\alpha(f_\alpha) \leq \mathcal{T}_\alpha(0)$).

Au regard de la régularité de la fonctionnelle $\mathcal{T}_\alpha$, il est facile de vérifier que le minimiseur f_α de $\mathcal{T}_\alpha$ est la solution de l'équation

$$(T^*T + \alpha I)\, f = T^*g, \tag{1.37}$$

où I désigne l'opérateur d'identité sur H_1. De l'équation (1.37), on peut voir que le spectre de l'opérateur T^*T a été décalé de α, ce qui permet d'éviter l'instabilité due à l'accumulation du spectre de T^*T en 0. Il existe également une version itérative de la méthode de Tikhonov (voir, e.g. [32, Section 5.1]) où la solution approximative $f_{\alpha,n}$ est définie par la récurrence suivante

$$f_{\alpha,0} = 0, \quad f_{\alpha,n+1} = \underset{f\in H_1}{\operatorname{argmin}} \|Tf - g\|^2 + \alpha\|f - f_{\alpha,n}\|^2,$$

qui est équivalent à

$$f_{\alpha,0} = 0, \quad (T^*T + \alpha I)\, f_{\alpha,n+1} = T^*g + \alpha f_{\alpha,n}. \tag{1.38}$$

En outre, il existe une version généralisée de la méthode Tikhonov qui ajoute un opérateur non borné L dans le terme de pénalité. C'est-à-dire, avec un opérateur non borné $L : \mathcal{D}(L) \subset H_1 \to H_2$, la solution

approximative f_α est définie comme le minimiseur sur $\mathcal{D}(L)$ de la fonctionnelle

$$\mathcal{T}_\alpha(f, L) = \|Tf - g\|^2 + \alpha\|Lf\|^2. \tag{1.39}$$

Dans cette généralisation, L est un opérateur qui prend en compte des informations a-priori sur la solution $f^\dagger$ et qui remplit la condition complémentaire

$$\forall f \in \mathcal{D}(L), \quad \|Tf\|^2 + \|Lf\|^2 \geq \gamma\|f\|^2,$$

pour une certaine constante positive γ. Enfin, notez que la méthode de Tikhonov ordinaire (resp. itérative) peut également être interprétée comme une méthode de régularisation définie via une fonction génératrice

$$v_\alpha(\lambda) = \frac{1}{\lambda + \alpha} \qquad \left(\text{resp.} \qquad \varphi_{\alpha,n}(\lambda) = \frac{(\lambda+\alpha)^n - \alpha^n}{\lambda(\lambda+\alpha)^n}\right). \tag{1.40}$$

Évidemment, à partir de (1.40), en utilisant la Proposition 1.6, on peut montrer que les solutions approximatives f_α (de Tikhonov ordinaire) et $f_{\alpha,n}$ (de Tikhonov itéré) convergent vers la solution $T^\dagger g$ de l'équation (1.7) quand α tend vers 0. Pour une présentation plus exhaustive de la méthode de Tikhonov, voir, e.g. [32, Section 5,8,10].

1.1.3.2 Méthode de Landweber

Afin de régulariser l'équation (1.7), Landweber et Fridman [37, 72] ont proposé de réécrire l'équation (1.7) sous la forme d'une équation à point fixe et d'appliquer un algorithme du point fixe. En effet, en multipliant l'équation normale (1.9) associée à (1.7) par un constante positive $\theta < 1/\|T\|^2$ et en ajoutant f des deux côtés on obtient l'équation à point fixe

$$f = (I - \theta T^*T)f + \theta T^*g. \tag{1.41}$$

La majorant de θ garantit que l'opérateur à droite de l'équation(1.41) est une application contractante. Nous pouvons donc appliquer un algorithme de point fixe pour obtenir une approximation de la solution f de l'équation (1.41). Cela conduit à la suite de solution approximative $(f_m)_m$ définie par

$$f_0 \in H_1, \quad f_{m+1} = (I - \theta T^*T)f_m + \theta T^*g. \tag{1.42}$$

D'après le théorème du point fixe de Banach, on peut facilement voir que la suite f_m converge vers la solution de l'équation normale (1.9). De plus, en utilisant la proposition 1.6, avec α remplacé par $1/m$, on peut établir que $\lim_m f_m = T^\dagger g$.

A partir de (1.42), nous pouvons déduire l'expression de f_m en fonction de m et f_0 comme suit :

$$f_m = \theta \sum_{k=0}^{m-1} (I - \theta T^*T)^k T^*g + (I - \theta T^*T)^m f_0. \tag{1.43}$$

Si nous considérons $f_0 = 0$, alors à partir de (1.43), nous déduisons que $\|f_m\| \leq \theta m\|Tg\|$ ce qui implique que l'application $g \mapsto f_m$ est continue.

En réécrivant la suite $(f_m)_m$ définie dans (1.42) comme

$$f_{m+1} = f_m - \theta T^*(Tf_m - g), \tag{1.44}$$

on peut remarquer que la suite $(f_m)_m$ correspond en fait à l'algorithme de la plus forte descente avec un pas θ appliqué au problème des moindres carrés (1.8). Comme la méthode Tikhonov, la méthode de Landweber avec $f_0 = 0$ peut également être interprétée comme une méthode de régularisation définie via la fonction génératrice

$$g_m(\lambda) = \sum_{k=0}^{m-1} (1-\lambda)^k$$

où le paramètre de régularisation α est discret et égal à $1/k$.

1.1.3.3 Méthode des Gradients conjugués

La méthode des gradients conjugués (voir, e.g. [23, 49, 62, 95]) est une méthode de régularisation itérative non-linéaire initialement conçue pour la résolution de systèmes d'équations linéaires. La méthode des gradients conjugués régularise le problème (1.8) en approchant itérativement $f^\dagger$ par le minimiseur f_k du résidu $\|Tf - g\|^2$ sur les sous-espaces de Krylov de dimension finie

$$V_k = \text{span}\left\{T^*g, (T^*T)T^*g, ..., (T^*T)^{k-1}T^*g\right\},$$

où $k \geq 1$ et $k \in \mathbb{N}$. L'algorithme 1.1 à la page 26, décrit un algorithme de gradient conjugué qui permet de calculer les itérations f_k qui représentent la solution approximative du problème des moindres carrés (1.8).

Par définition du sous-espace de Krylov V_k, on peut voir que la solution approximative f_k de la méthode des gradients conjugués peut être écrite sous la forme

$$f_k = p_k(T^*T)T^*g, \tag{1.45}$$

où p_k est un polynôme de degré inférieur ou égal à k. Au regard de l'expression (1.45), on pourrait être tenté d'assimiler la méthode du gradient conjugué à une méthode de régularisation linéaire. Cependant, en inspectant l'algorithme 1.1, on peut voir que les coefficients du polynôme p_k dépendent de la donnée g, d'où la non-linéarité de la méthode du gradient conjugué.

On peut montrer que, pour tout $g \in \mathcal{D}(T^\dagger)$, l'itéré f_k de l'algorithme du gradient conjugué converge vers $T^\dagger g$ quand k tend vers l'infini (voir, e.g. [32, Theorem 7.9]). Cependant, il est important de noter que la méthode du gradient conjugué n'est pas une méthode de régularisation continue (voir, e.g. [32, Theorem 7.6]). En effet, en considérant par $f^\delta_{k(\delta)}$ la solution approximative correspondant à une donnée bruité g^δ satisfaisant (1.24), alors il n'existe pas de règles de sélection a-priori de paramètre $k(\delta)$ telle que $f^\delta_{k(\delta)}$ converge vers $f^\dagger$ quand $\delta \to 0$ (voir, e. g. [26] et [32, Corollary 7.7]). Néanmoins, l'ensemble des points de discontinuité de l'opérateur de gradient conjugué $R_k = p_k(T^*T)T^*$ est bien défini pour un opérateur compact comme l'ensemble des combinaisons linéaires d'au plus k fonctions propres de T^*T (voir, e.g. [32, Theorem 7.8]).

Un avantage majeur du gradient conjugué est le calcul facile de la solution régularisée f_k et la convergence numérique rapide, contrairement aux autres méthodes de régularisation itérative (par exemple la méthode Landweber, méthode ν).

Algorithm 1.1 Gradient conjugué

$f_0 = 0$
$k = 0$
if $T^*g = 0$ **then**
    return f_0
else
    $v_0 = T^*(Tf_0 - g)$
    $t_0 = \langle v_0, p_0 \rangle / \|Tp_0\|^2$
    $f_1 = f_0 - t_0 p_0$
    while $T^*(Tf_{k+1} - g) \neq 0$ **do**
        $\gamma_k = \|T^*(Tf_{k+1} - g)\|^2 / \|T^*(Tf_k - g)\|^2$
        $p_{k+1} = T^*(Tf_{k+1} - g) + \gamma_k p_k$
        $k = k + 1$
        $v_k = T^*(Tf_k - g)$
        $t_k = \langle v_k, p_k \rangle / \|Tp_k\|^2$
        $f_{k+1} = f_k - t_k p_k$
    end while
end if

1.1.3.4 Méthode de Showalter

La méthode Showalter (voir, e.g. [32, Section 4.1] ou la régularisation asymptotique approxime la solution $f^\dagger$ de l'équation (1.7) par f_α qui est la solution $u : \mathbb{R}_+ \to X$ du problème à valeur initiale :

$$\begin{cases} u'(t) + T^*Tu(t) = T^*g, & t \in \mathbb{R}_+ \\ u(0) = 0, \end{cases} \tag{1.46}$$

évalué à $t = 1/\alpha$. C'est-à-dire, $f_\alpha = u(1/\alpha)$. On peut vérifier (voir [32, Example 4.7]) que la fonction $u(t)$ définie par

$$u(t) = v_t(T^*T)T^*g, \quad \text{avec} \quad v_t(\lambda) = \frac{1 - \exp(\lambda t)}{\lambda},$$

est bien solution de (1.46). Ainsi on déduit que la solution régularisée f_α suivant la méthode de Showalter est définie par

$$f_\alpha = v_\alpha(T^*T)T^*g, \quad \text{avec} \quad v_\alpha(\lambda) = \frac{1 - e^{-\lambda/\alpha}}{\lambda}. \tag{1.47}$$

A partir de (1.47), nous pouvons vérifier que les conditions (1.30) et (1.31) de la proposition 1.6 sont remplies, ce qui permet de déduire la convergence de f_α vers $T^\dagger g$ quand α tend vers 0.

En appliquant un schéma numérique de différence finie avant de pas θ au problème à valeur initiale (1.46), nous obtenons l'approximation suivante de u:

$$u(t + h) \approx u(t) + \theta T^* \left(g - Tu(t)\right), \quad \text{avec} \quad u(0) = 0, \tag{1.48}$$

Ainsi, à partir de (1.48), si nous définissons $\alpha_m = m\theta$, alors l'itération f_{α_m} de la méthode de Showalter en utilisant la différence finie avant coïncide en fait avec l'itéré f_m de la méthode de Landweber définie par (1.44) avec $f_0 = 0$.

1.1.4 Revue des méthodes de sélection de paramètre

Une étape très cruciale dans l'application d'une méthode de régularisation est le choix du paramètre de régularisation. Étant donné un opérateur de régularisation $R_\alpha : H_2 \to H_1$ et une donnée bruitée g^δ satisfaisant la condition de niveau de bruit (1.24), il faut trouver une règle de sélection des paramètres $\alpha(\delta, g^\delta)$ telle que

$$||R_{\alpha(\delta,g^\delta)}g^\delta - T^\dagger g|| \to 0 \quad \text{quand} \quad \delta \to 0. \tag{1.49}$$

En d'autres termes, étant donné une donnée bruitée g^δ satisfaisant un niveau de bruit δ (c'est-à-dire satisfaisant (1.24)), on cherche la bonne quantité de régularisation (mesurée par le paramètre de régularisation) nécessaire pour obtenir une approximation stable et précise de la solution $f^\dagger$ de l'équation (1.7). En effet, d'une part, les grandes valeurs du paramètre de régularisation α correspondent à une grande quantité de régularisation, c'est-à-dire une grande stabilité, mais une faible fidélité au modèle initial $Tf = g$. Alors que, d'autre part, de petites valeurs du paramètre de régularisation correspondent à un faible niveau de régularisation, c'est-à-dire une faible stabilité mais une plus grande fidélité au modèle initial $Tf = g$. Dans la suite, étant donné un opérateur de régularisation R_α et une donnée bruitée g^δ, on désigne par f_α^δ l'opérateur de régularisation R_α appliqué à g^δ, i.e. $f_\alpha^\delta = R_\alpha g^\delta$.

Une caractérisation importante des règles de sélection de paramètres convergents $\alpha(\delta)$ pour un opérateur de régularisation R_α pour $T^\dagger$ est la suivante (voir, e.g. [32, Proposition 3.7]).

Proposition 1.7. *Soit $R_\alpha : H_2 \to H_1$ un opérateur de régularisation pour $T^\dagger$. Une règle de sélection de paramètres $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$, $\delta \to \alpha(\delta)$ est convergente, c'est-à-dire vérifie (1.25) et (1.26) si et seulement si*

$$\begin{cases} \alpha(\delta) \to 0 \\ \delta||R_{\alpha(\delta)}|| \to 0. \end{cases} \quad \textit{quand} \quad \delta \to 0. \tag{1.50}$$

Notons que la règle (1.26) implique (1.49). Outre l'exigence de convergence (1.49), on s'intéresse également à l'optimalité par rapport aux a-priori sur la solution inconnue $f^\dagger$. En effet, supposons que $\mathcal{M}$ soit un sous-espace de H_1, étant donné un opérateur de régularisation R_α et une règle de sélection des paramètres $\Lambda(\delta, g^\delta) := \alpha(\delta, g^\delta)$, l'erreur la plus grande suivant un niveau de bruit (1.24) et l'a-priori $f^\dagger \in \mathcal{M}$ est définie par

$$\Delta(\delta, \mathcal{M}, \Lambda) := \sup \left\{ ||R_{\Lambda(\delta,g^\delta)}g^\delta - f||, \quad f \in \mathcal{M},\ g^\delta \in H_2,\ ||Tf - g^\delta|| \leq \delta \right\}. \tag{1.51}$$

Étant donné l'a-priori $f \in \mathcal{M}$, on aimerait trouver une règle de sélection de paramètres $\Lambda(\delta, g^\delta) := \alpha(\delta, g^\delta)$ qui minimise la pire erreur $\Delta(\delta, \mathcal{M}, \Lambda)$ parmi toutes les règles de sélection de paramètres Λ.

Il existe trois types de règles de sélection de paramètres :

- **Les règles a-priori de sélection de paramètres:** Ici, le paramètre $\alpha(\delta)$ est défini avant la procédure de régularisation. Plus précisément, le paramètre $\alpha(\delta)$ est choisi et la solution approximative est définie par application de l'opérateur de régularisation R_α avec le $\alpha(\delta)$ prescrit. Notons que dans ce cas, le paramètre de régularisation α dépend uniquement du niveau de bruit δ et des a-priori disponibles sur la solution inconnue $f^\dagger$. Les règles a-priori de sélection sont les moins couteuse à tout égard. En effet, avec une règle a-priori de sélection de paramètre, l'étape de régularisation s'effectue une seule fois (contrairement aux méthodes de sélection a-posteriori) avec le paramètre de régularisation défini a-priori . Cependant, elles sont rarement utilisées car il est extrêmement

difficile en pratique d'obtenir des informations précises à la fois sur le niveau de bruit δ et sur la régularité de la solution inconnue $f^\dagger$.

- **Les règles a-posteriori de sélection de paramètres:** Contrairement aux règles a-priori de sélection de paramètres, le paramètre α est calculé a posteriori de la procédure de régularisation. Dans ce cas, le paramètre de régularisation α dépend à la fois du niveau de bruit δ et des données bruitées g^δ c'est-à-dire $\alpha := \alpha(\delta, g^\delta)$. Les règles a-posteriori sont coûteuses par rapport aux règles a-priori, car l'opérateur de régularisation R_α est appliqué plusieurs fois avant de choisir le paramètre $\alpha(\delta, g^\delta)$. Cependant, les règles a-posteriori de choix de paramètres sont moins restrictives car elles ne requièrent que le niveau de bruit δ (les données approximatives g^δ étant toujours disponibles).

- **Les règles heuristiques de sélection de paramètres:** Les règles heuristiques (ou empirique) de choix de paramètre, contrairement aux règles a-priori et a-posteriori, sont entièrement basées sur les données bruitées g^δ et n'utilisent aucune autre information telle que le niveau de bruit δ ou les a-priori sur la solution inconnue $f^\dagger$. Comme on pouvait s'y attendre, cet avantage est compensé par un défaut. En fait, en raison du véto de Bakushinskii [7], aucune règle heuristique de choix de paramètres converge. C'est-à-dire que nous ne pouvons pas obtenir (1.49) pour tout $f^\dagger \in H_1$ et $g^\delta \in H_2$ satisfaisant (1.24). Cependant, les règles heuristiques peuvent donner de meilleures approximations que les règles a-posteriori sophistiquées (voir, e.g. [51]) et cela ne devrait pas être surprenant car le résultat de Bakushinskii est basé sur le pire des scénarios. En outre, le plus souvent, les règles heuristiques de choix de paramètres sont les seules règles applicables, car dans la plupart des circonstances, le niveau de bruit δ dans la pratique est rarement disponible ou estimé avec précision.

De la section 1.1.4.1 à 1.1.4.4, nous passons en revue certaines règles de choix de paramètres bien connues qui pourraient être étendues à une méthode de régularisation générale, à savoir le principe de Morozov, la validation croisée généralisée, la méthode *L-curve* et la règle de quasi-optimalité. Rappelons que la plupart de ces règles ont été étudiées intensément pour la régularisation de Tikhonov.

Outre les règles de choix des paramètres que nous décrivons dans les sections suivantes, il existe de nombreuses autres règles de choix des paramètres parmi lesquelles la règle de Hanke-Raus [52, 93], la règle de Engl-Gfrerer [29] et la règle de Neubauer [94]. Pour une comparaison des règles heuristiques de choix de paramètres, voir, e.g. [40, 46, 78].

1.1.4.1 Principe de Morozov

Le principe de Morozov (voir, e.g. [28, 85, 88, 98, 99, 107]) est l'une des règles a-posteriori de choix de paramètres les plus populaires. Cette règle de sélection a été largement étudiée dans la littérature et possède plusieurs variantes (voir, e.g. [113, 114]). Cependant, l'idée directrice de toutes les variantes est la même : la méthode consiste à sélectionner le paramètre $\alpha(\delta, g^\delta)$ de telle sorte que le résidu $||Tf^\delta_{\alpha(\delta,g^\delta)} - g^\delta||$ soit du même ordre que le niveau de bruit δ. Nous considérons la variante suivante du principe de Morozov :

$$\alpha(\delta, g^\delta) := \sup\left\{\alpha > 0, \quad ||Tf^\delta_\alpha - g^\delta|| \leq C\delta\right\} \tag{1.52}$$

où C est une constante positive satisfaisante

$$C > \sup_{\alpha>0} \left\|I - TR_\alpha\right\|. \tag{1.53}$$

Notons que si l'on considère les méthodes de régularisation définies via une fonction génératrice, alors (1.53) est équivalent à

$$C > \sup\left\{|r_\alpha(\lambda)|,\ \alpha > 0,\ \lambda \in [0, ||T||^2]\right\}. \tag{1.54}$$

À partir de (1.54), nous pouvons déduire que $C > 1$ étant donné que $r_\alpha(0) = 1$. Rappelons que dans la plupart des cas, le résidu $||Tf_\alpha^\delta - g^\delta|| = ||(TR_\alpha - I)g^\delta||$ est une fonction continue et monotone de α. Il en résulte que $\alpha(\delta, g^\delta)$ défini dans (1.52) peut également être caractérisé par l'équation

$$||Tf^\delta_{\alpha(\delta,g^\delta)} - g^\delta|| = C\delta. \tag{1.55}$$

Étant donné une donnée bruitée g^δ satisfaisant (1.24), il n'est pas raisonnable de demander que $||Tf_\alpha^\delta - g^\delta|| < \delta$ vu que l'erreur dans la solution ne saurait être inférieur à l'erreur dans la donnée. Cela explique pourquoi la constante C doit être supérieure ou égale à 1. Ensuite, étant donné que les petites valeurs de α induisent moins de régularisation et donc plus d'instabilité, il est donc préférable de prendre la plus grande valeur de α qui satisfait l'inégalité dans (1.52).

Avant d'entrer dans une analyse de convergence de la règle de sélection de Morozov, définissons la notion de qualification de type Hölder pour une méthode de régularisation $(R_\alpha)_{\alpha>0}$.

Definition 1.5. *Soit $(R_\alpha)_{\alpha>0}$ une méthode de régularisation pour $T^\dagger$. La qualification de type Hölder de la méthode de régularisation $(R_\alpha)_{\alpha>0}$ est la plus grande valeur μ_0 du paramètre μ telle que la relation suivante est satisfaite :*

$$\sup\left\{||(R_\alpha T - I)f^\dagger||,\quad f^\dagger \in \mathcal{R}((T^*T)^\mu), ||f|| \le 1\right\} = \mathcal{O}\left(\alpha^\mu\right) \quad \textit{quand} \quad \alpha \to 0. \tag{1.56}$$

De (1.56), on déduit que si la méthode de régularisation $(R_\alpha)_{\alpha>0}$ est définie via une fonction génératrice υ_α, alors (1.56) est équivalent à

$$\sup_{\lambda\in(0,||T||^2]} |\lambda^\mu r_\alpha(\lambda)| = \mathcal{O}\left(\alpha^\mu\right) \quad \textit{quand} \quad \alpha \to 0. \tag{1.57}$$

Voyons maintenant comment est généralement abordée l'analyse de la convergence d'une méthode de régularisation couplée au principe de Morozov. Premièrement, il est important de noter que le terme d'erreur totale $f^\dagger - f_\alpha^\delta$ se compose de deux parties : l'erreur de régularisation (ou d'approximation) $f^\dagger - f_\alpha$ et l'erreur dû au bruit dans les données $f_\alpha - f_\alpha^\delta$. Ainsi, une procédure classique d'analyse de convergence de la méthode de régularisation (voir, e.g. [32, Section 4.3]) consiste à borner chacun de ces deux termes et à déduire finalement une borne de l'erreur totale via l'inégalité triangulaire

$$||f^\dagger - f_\alpha^\delta|| \le ||f^\dagger - f_\alpha|| + ||f_\alpha - f_\alpha^\delta||. \tag{1.58}$$

Supposons que $g = Tf^\dagger$ avec $f^\dagger$ satisfaisant la condition source générale

$$f^\dagger = \phi(T^*T)w, \quad w \in H_1 \quad \text{tel que} \quad ||w|| \le \rho, \tag{1.59}$$

où ϕ est une fonction index, c'est-à-dire une fonction continue et croissante satisfaisant $\lim_{t\downarrow 0}\phi(t) = 0$. Décrivons une procédure classique pour borner les deux termes d'erreur dans (1.58) lorsque $\alpha(\delta, g^\delta)$ est défini par (1.52).

Pour simplifier la notation, dans le reste de cette section, nous indiquerons par α (au lieu de $\alpha(\delta, g^\delta)$) le paramètre de régularisation défini par (1.52).

- **erreur de régularisation** : En général, en utilisant la condition source satisfaite par $f^\dagger$, une première étape consiste à prouver l'existence d'une fonction index ψ telle que

$$||f^\dagger - f_\alpha|| \leq C_1 \psi \left(||T(f^\dagger - f_\alpha)|| \right). \tag{1.60}$$

Comme illustration, si l'on considère les conditions source de type Hölder, c'est-à-dire

$$f^\dagger = (T^*T)^\mu w, \quad w \in H_1, \quad \text{tel que} \quad ||w|| \leq \rho, \tag{1.61}$$

avec $\mu > 0$ et une méthode de régularisation définie par une fonction υ_α, alors l'estimation (1.60) peut être obtenue comme suit. Soit r_α la fonction *résiduelle* définie par (1.34) correspondant à la fonction génératrice υ_α, alors

$$||f^\dagger - f_\alpha|| = ||r_\alpha(T^*T) f^\dagger|| = \left\|(T^*T)^\mu r_\alpha(T^*T) w\right\|. \tag{1.62}$$

Notons que dans (1.62), on a utilisé la commutation $r_\alpha(T^*T)(T^*T)^\mu = (T^*T)^\mu r_\alpha(T^*T)$, qui se justifie par le fait que la commutation est trivial si r_α est polynôme et μ entier naturel et le fait que les polynômes sont denses dans l'ensemble des fonctions continues sur un domaine borné. En utilisant l'inégalité d'interpolation (1.23) avec $f = r_\alpha(T^*T)w$, $p = \mu$ et $q = \mu + 1/2$, on obtient

$$\begin{aligned} \left\|(T^*T)^\mu r_\alpha(T^*T) w\right\| &\leq \left\|r_\alpha(T^*T) w\right\|^{\frac{1}{2\mu+1}} ||(T^*T)^{\frac{1}{2}} r_\alpha(T^*T)(T^*T)^\mu w||^{\frac{2\mu}{2\mu+1}} \\ &\leq (\gamma_0 \rho)^{\frac{1}{2\mu+1}} ||(T^*T)^{\frac{1}{2}} r_\alpha(T^*T) f^\dagger||^{\frac{2\mu}{2\mu+1}}, \end{aligned} \tag{1.63}$$

où $\gamma_0 := \sup \left\{ r_\alpha(\lambda),\ \alpha > 0,\ \lambda \in (0, \left\|T\right\|^2] \right\}$. Mais

$$||T(f_\alpha - f^\dagger)|| = ||T r_\alpha(T^*T) f^\dagger|| = ||(T^*T)^{\frac{1}{2}} r_\alpha(T^*T) f^\dagger||. \tag{1.64}$$

Ainsi, d'après (1.61), (1.63) et (1.64), on obtient que

$$||f^\dagger - f_\alpha|| \leq (\gamma_0 \rho)^{\frac{1}{2\mu+1}} \psi \left(||T(f_\alpha - f^\dagger)|| \right), \quad \text{où} \quad \psi(t) = t^{\frac{2\mu}{2\mu+1}}. \tag{1.65}$$

Ensuite, après avoir obtenu l'estimation (1.60), la second étape consiste à utiliser l'inégalité triangulaire pour lier le terme $||g - Tf_\alpha||$ à δ comme suit :

$$\begin{aligned} ||g - Tf_\alpha|| &\leq ||g^\delta - Tf_\alpha^\delta|| + ||g - g^\delta - T(f_\alpha - f_\alpha^\delta)|| \\ &= ||g^\delta - Tf_\alpha^\delta|| + ||(TR_\alpha - I)(g - g^\delta)|| \\ &\leq C\delta + \gamma\delta \\ &= (C + \gamma)\delta \end{aligned} \tag{1.66}$$

où la constante γ est une borne supérieure de $||TR_\alpha - I||$ suivant α, c'est-à-dire

$$\gamma := \sup_{\alpha > 0} ||TR_\alpha - I||. \tag{1.67}$$

Notons que dans le cas où R_α est défini à partir d'une famille de fonctions $\{\upsilon_\alpha\}_{\alpha>0}$, alors la constante γ défini en (1.67) est bien égale à la constante $\gamma_0 := \sup \left\{ r_\alpha(\lambda),\ \alpha > 0,\ \lambda \in (0, ||T||^2] \right\}$ avec r_α défini en (1.34).

Ainsi, les estimations (1.58) et (1.66) implique la borne suivante sur l'erreur de régularisation :

$$||f^\dagger - f_\alpha|| \leq C_1 \psi((C + \gamma)\delta). \tag{1.68}$$

- **erreur dûe au bruit dans les données** : En utilisant les résultats classiques et la définition de l'opérateur de régularisation R_α, on obtient en général l'inégalité suivante

$$||f_\alpha - f_\alpha^\delta|| = ||R_\alpha(g - g^\delta)|| \leq \delta||R_\alpha|| \leq C_2\frac{\delta}{\sqrt{\alpha}}. \tag{1.69}$$

En effet, l'estimation (1.69) est simplement basée sur $\left\|R_\alpha\right\| \leq 1/\sqrt{\alpha}$ et cela dépend uniquement de la méthode de régularisation.

Par exemple, si l'on considère une méthode de régularisation définie à l'aide d'une fonction génératrice υ_α, alors l'exigence suivante qui implique (1.69)

$$\sup_{\lambda\in(0,\left\|T\right\|^2]} \sqrt{\lambda}\upsilon_\alpha(\lambda) \leq C_2\frac{1}{\sqrt{\alpha}}, \quad \text{pour une constante positive} \quad C_2 \tag{1.70}$$

est inclus par certains auteurs dans la définition d'une méthode de régularisation définie via une fonction génératrice (voir, e.g. [86]). Au fait, on notera que la condition (1.70) est remplie le plus souvent (e.g. méthodes Tikhonov, landweber et Showalter).

Une fois l'estimation (1.69) établie, l'étape suivante consiste à trouver une borne inférieure de α en fonction de δ en utilisant (1.52), la condition source sur $f^\dagger$ et la qualification de la méthode de régularisation. En suivant les mêmes lignes que dans (1.66), on obtient

$$||g^\delta - Tf_{2\alpha}^\delta|| \leq ||g - Tf_{2\alpha}|| + ||g^\delta - g - T(f_{2\alpha}^\delta - f_{2\alpha})|| \leq ||g - Tf_{2\alpha}|| + \gamma\delta.$$

ce qui implique en utilisant (1.52) que

$$||g - Tf_{2\alpha}|| \geq ||g^\delta - Tf_{2\alpha}^\delta|| - \gamma\delta \geq (C - \gamma)\delta, \tag{1.71}$$

où $C - \gamma > 0$ au regard de (1.53) et (1.67). En utilisant la condition source sur $f^\dagger$ et la qualification de la méthode de régularisation, on peut trouver une borne supérieure de $||g - Tf_{2\alpha}||$ et ainsi obtenir à partir de (1.71) l'estimation suivante :

$$(C - \gamma)\delta \leq \gamma_*\phi(2\alpha)\sqrt{2\alpha}, \tag{1.72}$$

où ϕ est une fonction strictement croissante sur $\mathbb{R}_+$ qui s'annule en 0.

À titre d'illustration, si nous considérons à nouveau une condition source de type Hölder (1.61) et une méthode de régularisation définie à l'aide d'une fonction génératrice υ_α, alors à partir de (1.71) nous obtenons

$$\begin{aligned}
(C - \gamma)\delta \leq ||g - Tf_{2\alpha}|| &= ||(T(f^\dagger - f_{2\alpha})|| \\
&= ||Tr_{2\alpha}(T^*T)(T^*T)^\mu w|| \quad \text{vue (1.35) et (1.61)} \\
&= ||(T^*T)^{1/2}r_{2\alpha}(T^*T)(T^*T)^\mu w|| \quad \text{vue (1.22)} \\
&\leq \sup_{\lambda\in(0,||T||^2]} \lambda^{\mu+1/2}|r_{2\alpha}(\lambda)|.
\end{aligned} \tag{1.73}$$

À condition que la méthode de régularisation ait une qualification de type Hölder $\mu_0 > 1/2$ et que $\mu + 1/2 < \mu_0$, on obtient alors

$$\sup_{\lambda\in(0,||T||^2]} \lambda^{\mu+1/2}|r_{2\alpha}(\lambda)| \leq \gamma_*(2\alpha)^{\mu+1/2}, \tag{1.74}$$

où γ_* est une constante positive. Nous en déduisons donc la borne supérieure suivante de $(C-\gamma)\delta$:

$$(C-\gamma)\delta \leq \gamma_*(2\alpha)^{\mu+1/2}, \tag{1.75}$$

qui est équivalent à (1.72) avec $\phi(t) = t^\mu$.

Une fois que (1.72) est établi, nous pouvons déduire à l'aide de la monotonicité de la fonction ϕ que

$$\alpha \geq \frac{1}{2}\Theta^{-1}\left(\frac{C-\gamma}{\gamma_*}\delta\right) \tag{1.76}$$

où $\Theta(t) = \sqrt{t}\phi(t)$. En utilisant les estimations (1.76) et(1.69), on déduit la borne suivante sur l'erreur dû au bruit dans les données

$$||f_\alpha - f_\alpha^\delta|| \leq \frac{\delta}{\sqrt{\frac{1}{2}\Theta^{-1}\left(\frac{C-\gamma}{\gamma_*}\delta\right)}}. \tag{1.77}$$

Enfin à partir des estimations (1.68) et (1.77), on déduit l'estimation suivante sur l'erreur totale :

$$||f^\dagger - f_\alpha^\delta|| \leq C_1\psi((C+\gamma)\delta) + \frac{\delta}{\sqrt{\frac{1}{2}\Theta^{-1}\left(\frac{C-\gamma}{\gamma_*}\delta\right)}}. \tag{1.78}$$

Il est important de noter que la borne de l'erreur dû au bruit dans les données est obtenue sous la condition que la qualification de type Hölder μ_0 de la méthode de régularisation soit strictement supérieure à $1/2$. Cette condition est en fait une restriction générale pour le principe de Morozov.

Si l'on considère une méthode de régularisation $(R_\alpha)_{\alpha>0}$ définie via une fonction génératrice υ_α satisfaisant (1.70) et que l'on suppose que la qualification de type Hölder μ_0 de la méthode est strictement supérieure à $1/2$, i.e. $\mu_0 > 1/2$, alors sous la condition source (1.61) avec $\mu \leq \mu_0 - 1/2$, les estimations (1.65), (1.66), (1.69) et (1.75) donnent le taux de convergence

$$||f^\dagger - f_{\alpha(\delta,g^\delta)}^\delta|| = \mathcal{O}\left(\delta^{\frac{2\mu}{2\mu+1}}\right), \quad \text{quand} \quad \delta \to 0, \tag{1.79}$$

où $\alpha(\delta, g^\delta)$ est défini par (1.52) et g^δ satisfait $||Tf^\dagger - g^\delta|| \leq \delta$. Étant donné que le taux $\delta^{\frac{2\mu}{2\mu+1}}$ est optimal sous la condition source de type Hölder (1.61) (voir, e.g. [32, Section 4.1]), l'estimation (1.79) implique que la règle de sélection de paramètres (1.52) est convergente et d'ordre optimal pour les conditions sources (1.61) pour $\mu \in (0, \mu_0 - 1/2]$.

Le principe de Morozov donne également un taux de convergence d'ordre optimal pour les méthodes de régularisation itératives (e.g. Landweber [32, Section 6.1], gradient conjugué [93]). L'optimalité du principe de Morozov sous des conditions sources générales (1.59), pour les méthodes de régularisation continue définies via les fonctions génératrices, est étudiée dans [85, 91].

En raison de la limitation de l'optimalité du principe de Morozov aux conditions sources (1.59) avec $\mu \in (0, \mu_0 - 1/2]$ où μ_0 est la qualification de type Hölder de la méthode de régularisation, de nombreux auteurs (voir, e. g. [29, 38, 52]) ont développé d'autres règles a-posteriori de sélection de paramètres qui sont d'ordre optimal pour tout $\mu \in (0, \mu_0]$. Cependant, ces règles sont plus compliquées et nécessitent des conditions supplémentaires (comparé au principe de Morozov) et ne conviennent pour la plupart qu'à la méthode régularisation de Tikhonov.

1.1.4.2 Validation croisée généralisée

La validation croisée généralisée (GCV) introduite par Wahba et al. (voir, e.g. [41, 77, 117]) est une règle heuristique de sélection de paramètres dans la régularisation d'équations linéaires avec un opérateur dont l'image est de dimension finie. Dans cette section, nous considérons l'équation

$$Kx = y \tag{1.80}$$

où K est un opérateur linéaire d'un espace Hilbert X vers $\mathbb{R}^m$, $y \in \mathbb{R}^m$ est la donnée et $x \in X$ est la solution. Au départ, la méthode a été développée pour la sélection de paramètre dans le modèle de régression linéaire [41]. Notons qu'en utilisant une méthode de discrétisation (par exemple collocation, projection), l'équation (1.80) peut être dérivée de l'équation (1.7). En effet, si nous supposons que l'espace de Hilbert H_2 est un espace de fonction, alors nous pouvons définir le vecteur y comme une discrétisation de la fonction g sur une grille finie $\{t_i\}_{i=1,..,m}$ et de même Kx est défini comme une discrétisation de Tf à la grille $\{t_i\}_{i=1,..,m}$. C'est-à-dire,

$$y_i = g(t_i), \quad x = f, \quad \text{et} \quad (Kx)_i = (Tx)(t_i). \tag{1.81}$$

En pratique, au lieu des données exactes y, on ne connaît qu'une donnée bruitée y^δ. La GCV exige que le bruit dans les données soit blanc et non-corrélé. C'est-à-dire que le vecteur d'erreur $y^\delta - y$ satisfait

$$E[y^\delta - y] = 0, \quad \text{et} \quad E[(y^\delta - y)(y^\delta - y)^\top] = \sigma^2 I_m, \tag{1.82}$$

pour une certaine constante positive σ, où I_m est la matrice carré identité d'ordre m, $E[y^\delta - y]$ dénote l'espérance du vecteur aléatoire $y^\delta - y$ et $E[(y^\delta - y)(y^\delta - y)^\top]$ dénote l'espérance de la matrice $(y^\delta - y)(y^\delta - y)^\top$. Notons que (1.82) implique que

$$\mathbb{E}[||y^\delta - y||^2] = m\sigma^2,$$

à partir duquel nous identifions le niveau de bruit dans ce cas comme $\delta = \sigma\sqrt{m}$.

La GCV est une généralisation de la validation croisée ordinaire (voir, e.g. [2, 41, 109]) dont l'idée est la suivante : Soit $y_k^\delta \in \mathbb{R}^{m-1}$ obtenu à partir de y^δ en supprimant sa k-ième composante. C'est-à-dire, en considérant la discrétisation (1.81), y_k^δ n'est rien d'autre que la discrétisation de g^δ sur la grille $\{t_1, ..., t_{k-1}, t_{k+1}, ..., t_m\}$ de taille $m-1$. Soit $K_k : X \to \mathbb{R}^{m-1}$ la discrétisation de (1.7) correspondant à la grille $\{t_1, ..., t_{k-1}, t_{k+1}, ..., t_m\}$ définie de façon similaire à (1.81). Désignons par $R_\alpha(K_k)$ un opérateur de régularisation pour $(K_k)^\dagger$. Soit $x_{\alpha,k}^\delta$ l'estimateur de la solution $x^\dagger$ de l'équation (1.80) correspondant à la grille de discrétisation $\{t_1, ..., t_{k-1}, t_{k+1}, ..., t_m\}$, c'est-à-dire $x_{\alpha,k}^\delta = R_\alpha(K_k)y_k^\delta$. Si α est un bon paramètre de régularisation, alors la k-ième composante de $Kx_{\alpha,k}^\delta$ devrait être un bon estimateur de la $k-$ième composante $(y^\delta)_k$ du vecteur y^δ. De cette façon, en prenant la moyenne sur tous les k de 1 à m, un bon choix du paramètre α peut être défini comme le minimiseur de la fonction

$$P(\alpha) = \frac{1}{m}\sum_{k=1}^{m}\left[\left(Kx_{\alpha,k}^\delta\right)_k - (y^\delta)_k\right]^2. \tag{1.83}$$

La règle de validation croisée ordinaire (OCV) consiste en fait à choisir le paramètre de régularisation α_* comme minimiseur de la fonction P définie en (1.83), c'est-à-dire,

$$\alpha_* = \underset{\alpha>0}{\operatorname{argmin}} \frac{1}{m}\sum_{k=1}^{m}\left[\left(Kx_{\alpha,k}^\delta\right)_k - (y^\delta)_k\right]^2. \tag{1.84}$$

Bien que l'idée d'OCV soit intuitivement rationnelle, il a été démontré dans [41] que, lorsqu'on considère le modèle de régression linéaire, si K est presque diagonal, la règle de sélection (1.84) révèle une faiblesse due au fait que la fonction P peut avoir plusieurs minimiseurs (possiblement en nombre infinis). Par ailleurs, comme indiqué dans [41], "[...] divers arguments peuvent être avancés pour justifier que toute bonne estimation de α devrait être invariante par rotation du système de coordonnées (de mesure)". La validation croisée généralisée est en effet une forme invariante par rotation de la validation croisée ordinaire (voir, e.g. [41, Section 2]) qui définit le paramètre de régularisation α comme le minimiseur de la fonction

$$V(\alpha) = \frac{||Kx_\alpha^\delta - y^\delta||^2}{\text{trace}(I - KR_\alpha)^2}, \tag{1.85}$$

où R_α est l'opérateur de régularisation pour $K^\dagger$. Si nous considérons une méthode de régularisation définie à l'aide d'une fonction génératrice υ_α, alors la fonction $V(\alpha)$ peut être réécrite comme

$$V(\alpha) = \frac{||Kx_\alpha^\delta - y^\delta||^2}{\text{trace}(r_\alpha(KK^*))^2}, \tag{1.86}$$

où r_α est la fonction *résiduelle* définie par (1.34).

Remark 1.1. *Étant donné un vecteur $y \in \mathbb{R}^m$, le vecteur $r_\alpha(KK^*)y$ est en fait le résidu $Kx_\alpha - y$ où $x_\alpha = R_\alpha y$. Par conséquent, une évaluation de la matrice $r_\alpha(KK^*)$ peut être effectuée en une seule étape de régularisation.*

Une étape importante de la GCV est le calcul du terme de trace présent dans la définition de la fonction V. En effet, la matrice $r_\alpha(KK^*)$ n'est guère explicitement calculable en général et il pourrait donc être compliqué d'évaluer sa trace. Dans [39], les auteurs ont suggéré une procédure pour approximer ce terme de trace. En effet, si $e \in \mathbb{R}^m$ est un vecteur de bruit blanc dont la matrice de covariance est égale à la matrice identité I_m, alors pour toute matrice symétrique $A \in \mathbb{R}^{m\times m}$, on obtient

$$\text{trace}(A) = \mathbb{E}\left[e^\top A\, e\right]. \tag{1.87}$$

Nous pouvons donc estimer $\text{trace}\left(r_\alpha(KK^*)\right)$ comme

$$\text{trace}\left(r_\alpha(KK^*)\right) \approx e^\top r_\alpha(KK^*)\, e, \tag{1.88}$$

où e est un vecteur de bruit blanc dont la matrice de covariance est égale à l'identité dans $\mathbb{R}^m$. À partir de la remarque 1.1, l'évaluation de la partie droite de (1.88) peut être effectuée par une étape de régularisation. Cependant, il convient de noter que l'approximation faite dans (1.88) pourrait être améliorée en augmentant la taille de l'échantillon des vecteurs e_k utilisé pour approximer l'espérance dans le côté droit de l'équation (1.87). L'optimalité asymptotique de la GCV a été étudiée par Lukas [77] pour la régularisation de Tikhonov. Enfin, rappelons que contrairement à la plupart des règles heuristiques de choix de paramètres, la fonction $V(\alpha)$ de la GCV et la fonction $P(\alpha)$ de la OCV ne sont pas des estimations de l'erreur $||x^\dagger - x_\alpha^\delta||$ mais sont plutôt des estimations de la *predictive mean square error*

$$Q(\alpha) = ||y - Kx_\alpha^\delta||^2. \tag{1.89}$$

1.1.4.3 La méthode L-curve

La méthode *L-curve* introduite par Hansen [53, 56] est une règle heuristique de choix de paramètres populaire basée sur la courbe de la norme de solution approximative $||f_\alpha^\delta||$ versus la norme du résidu $||g^\delta - Tf_\alpha^\delta||$. Considérons la courbe de $||f_\alpha^\delta||$ versus $||g^\delta - Tf_\alpha^\delta||$ dans une double échelle logarithmique, le paramètre de régularisation α étant le paramètre implicite du graphe. À partir de cette courbe, on peut obtenir une sorte de compromis entre la fidélité au modèle et la stabilité via le compromis entre la minimisation de la norme du résidu $||g^\delta - Tf_\alpha^\delta||$ et la norme de solution approximative $||f_\alpha^\delta||$.

En effet, la courbe $(\log(||g^\delta - Tf_\alpha^\delta||), \log(||f_\alpha^\delta||))$ présente généralement une forme en "L" (voir, e.g. la Figure 1.1). En effet, pour de très petits paramètres α, la solution approximative f_α^δ est dominée par le bruit et donc $||f_\alpha^\delta||$ explose tandis que la norme résiduelle $||g^\delta - Tf_\alpha^\delta||$ diminue lentement et sature à $||g^\delta - Qg^\delta||$, où Q est le projecteur orthogonal sur la fermeture de l'image de T. Ainsi, pour de tels α, la partie correspondante de la courbe est approximativement verticale. D'autre part, lorsque α est grand, $||f_\alpha^\delta||$ sature autour d'une constante généralement égale à 0 tandis que $||g^\delta - Tf_\alpha^\delta||$ croit. Cela donne une planéité à la partie correspondante du graphe. Ainsi, étant donné la partie verticale du graphe correspondant aux petits paramètres de régularisation et la partie plate correspondant au grands paramètres de régularisation, il en résulte que l'ensemble de la courbe présente souvent une forme en "L".

A titre d'illustration, considérons la régularisation de Tikhonov. Si nous supposons que T est compact et que $(\sigma_k, f_k, g_k)_k$ est la décomposition en valeurs singulières de T, alors la norme de solution approchée au carré et la norme du résidu au carré sont respectivement :

$$\eta(\alpha) = ||f_\alpha^\delta||^2 = \sum_{k=1}^{\infty} \frac{\sigma_k^2}{(\sigma_k^2 + \alpha)^2} (g_k^\delta)^2 \tag{1.90}$$

$$\rho(\alpha) = ||g^\delta - Tf_\alpha^\delta||^2 = \sum_{k=1}^{\infty} \frac{\alpha^2}{(\sigma_k^2 + \alpha)^2} (g_k^\delta)^2 + ||g_\perp^\delta||^2 \tag{1.91}$$

où $g_k^\delta = \langle g^\delta, g_k \rangle$ sont les coefficients de Fourier de g^δ et $g_\perp^\delta = g^\delta - Qg^\delta$ la projection de g^δ sur l'orthogonale d'image de T. De (1.90) et (1.91), on déduit que

$$\frac{\partial \eta}{\partial \rho} = \frac{\partial \eta(\alpha)}{\partial \alpha} \left(\frac{\partial \rho(\alpha)}{\partial \alpha} \right)^{-1} = -\frac{1}{\alpha}. \tag{1.92}$$

Ainsi, à partir de (1.92), on voit que lorsque α tend vers 0, $1/\alpha$ tend vers ∞ et donc la pente de la courbe $(\rho(\alpha), \eta(\alpha))$ est verticale, alors que lorsque α tend vers ∞, $1/\alpha$ tend vers0 et donc la pente de la courbe $(\rho(\alpha), \eta(\alpha))$ est horizontale. Entre les deux (α ni petit ni grand), il y a une région de croisement qui définit la forme en "L".

Lorsque la L-curve présente un *corner*, le paramètre de régularisation correspondant au *corner* permet intuitivement d'obtenir un compromis équitable entre une faible norme du résidu et une norme de solution raisonnable. Par conséquent, ce paramètre peut également permettre d'obtenir un bon compromis entre l'erreur de régularisation et l'erreur dû au bruit dans les données. Une difficulté pratique de la méthode *L-curve* est le calcul du paramètre α_* correspondant au *corner* de la courbe. Hansen et O'Leary [56] ont proposé de calculer ce α_* comme le paramètre qui maximise la courbure

$$C(\alpha) = \frac{|X'(\alpha)Y''(\alpha) - X''(\alpha)Y'(\alpha)|}{[X'(\alpha)^2 + Y'(\alpha)^2]^{3/2}}, \tag{1.93}$$

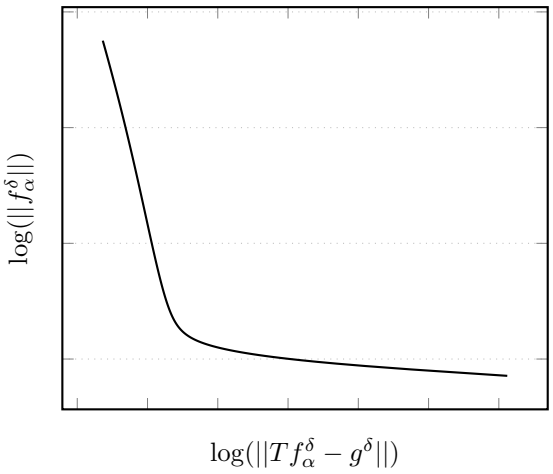

Figure 1.1: L-curve correspondant à la méthode de Tikhonov appliquée à une équation de la chaleur rétrograde en une dimension.

où $X(\alpha) = \log(||Tf_\alpha^\delta - g^\delta||)$ et $Y(\alpha) = \log(||f_\alpha^\delta||)$. Toutefois, pour définir la courbure $C(\alpha)$ dans (1.93), les termes $X(\alpha)$ et $Y(\alpha)$ doivent pouvoir être différentiables deux fois, ce qui n'est pas toujours possible et accessible.

Si la différentiabilité des composantes $X(\alpha)$ et $Y(\alpha)$ ne se vérifie pas, il faut alors trouver une alternative. À cet égard, plusieurs algorithmes ont été développés afin de calculer le paramètre de la méthode *L-curve*.

Dans l'article original [56] de Hansen, l'auteur a suggéré un algorithme pour approximer le paramètre de régularisation correspondant au *corner* de la L-curve. Dans [105], un algorithme basé sur la L-curve est décrit pour approximer l'indice k de la *spectral cut-off* est donné. D'autre part, un autre algorithme basé sur la L-curve a été développé pour la méthode des gradients conjugués dans [22]. Un algorithme plus général pour l'approximation du paramètre correspondant au *corner* de la courbe L pour les méthodes de régularisation discrètes est présenté dans [55]. Engl et Gfrerer [30] ont montré que pour la régularisation de Tikhonov, le paramètre de régularisation α qui maximise la courbure de la L-curve $(||g^\delta - Tf_\alpha^\delta||^2, ||f_\alpha^\delta||^2)$ reste supérieur ou égal à $1/||T||^2$ quelle que soit la valeur de δ, ce qui entraîne la non-convergence de la méthode dans ce cas. En outre, ils ont proposé une variante de la méthode *L-curve* qui utilise le niveau de bruit δ et permet de calculer le paramètre de régularisation correspondant au principe de Morozov. Dans [47], l'auteur a construit un problème pour lequel la méthode *L-curve* ne converge pas. D'autres résultats de la non-convergence de la méthode *L-curve* dans un cadre semi-discret, semi stochastique pour certaines classes de problèmes sont présentés dans [116].

Remark 1.2. *La non-convergence théorique de la méthode L-curve dans le cadre général (due à la saturation du paramètre de régularisation) n'a été prouvée qu'à l'échelle lin-lin, c'est-à-dire pour la courbe* $(||g^\delta - Tf_\alpha^\delta||^2, ||f_\alpha^\delta||^2)$ *(voir [30]) et* $(||g^\delta - Tf_\alpha^\delta||, ||f_\alpha^\delta||)$ *(voir [100] et ses références). Aucun résultat de ce type n'a été établi pour la double échelle logarithmique* $(\log(||g^\delta - Tf_\alpha^\delta||^2), \log(||f_\alpha^\delta||^2))$.

Remark 1.3. *Une remarque faite par Hanke [47] est que la méthode L-curve peut être peu performante pour une solution très lisse. Hansen et O'Leary [56] ont illustré via des simulation numériques que la méthode L-curve est plus robuste que la validation croisée généralisée (GCV). De plus, de leurs*

simulations, la méthode L-curve est capable de détecter les erreurs corrélées (à la fois sur les données et sur l'opérateur) contrairement à la GCV dans le cas particulier de la régularisation de Tikhonov.

Notons qu'une variante intéressante de la méthode *L-curve* basé sur la minimisation de la fonction

$$\Psi_\mu(\alpha) = ||g^\delta - Tf_\alpha^\delta||^2 ||f_\alpha^\delta||^{2\mu}, \quad \mu > 0. \tag{1.94}$$

proposé par Reginska [100] a été étudié dans plusieurs articles avec des applications, voir, e.g. [13, 14, 40, 61].

1.1.4.4 La méthode de Quasi-optimalité

La méthode de quasi-optimalité ([9, 10, 11, 12, 66, 67, 74]) est une règle heuristique de sélection de paramètre qui a été largement étudiée dans le cadre de la régularisation de Tikhonov. Le cœur de la règle de quasi-optimalité consiste à choisir le paramètre de régularisation α par :

$$\alpha_* = \underset{(0,||T||^2]}{\operatorname{argmin}} \ \alpha \Big\| \frac{\partial f_\alpha^\delta}{\partial \alpha} \Big\|. \tag{1.95}$$

En échantillonnant le paramètre de régularisation α comme suit

$$\alpha_n := \alpha_0 q^n, \quad \alpha_0 \in (0, ||T||^2], \quad 0 < q < 1, \tag{1.96}$$

à partir de (1.95), on peut définir une version discrète de la méthode de quasi-optimalité en considérant le paramètre de régularisation α_{n_*} avec n_* défini par

$$n_* = \underset{n \geq 0}{\operatorname{argmin}} \Big\| f_{\alpha_n}^\delta - f_{\alpha_{n+1}}^\delta \Big\|. \tag{1.97}$$

La Figure 1.2 illustre une application de la règle de quasi-optimalité discrète (1.96)-(1.97) à la régularisation de Tikhonov appliquée à une équation de chaleur rétrograde en une dimension.

Si l'on considère la méthode de Tikhonov itéré d'ordre n, alors on peut constater (voir, e.g. [10]) que

$$\alpha \Big\| \frac{\partial f_{\alpha,n}^\delta}{\partial \alpha} \Big\| = n \Big\| f_{\alpha,n+1}^\delta - f_{\alpha,n}^\delta \Big\|. \tag{1.98}$$

Ainsi le paramètre de régularisation α_* correspondant à la règle de quasi-optimalité est défini comme

$$\alpha_* = \underset{(0,||T||^2]}{\operatorname{argmin}} \, n \Big\| f_{\alpha,n+1}^\delta - f_{\alpha,n}^\delta \Big\|. \tag{1.99}$$

Nous rappelons que les solutions approximatives $f_{\alpha,n}$ de la méthode itérative de Tikhonov sont définies par (1.38). Les résultats de convergence de la règle de quasi-optimalité pour la méthode Tikhonov itéré ont été donnés dans [10] à condition que les données g et g^δ satisfassent la condition :

$$t^2 \int_t^\infty \frac{1}{\lambda} \, \mathrm{d} \Big\| F_\lambda Q(g - g^\delta) \Big\|^2 \leq C \int_0^t \lambda \, \mathrm{d} \Big\| F_\lambda Q(g - g^\delta) \Big\|^2 \quad \forall \, t > 0, \tag{1.100}$$

où C est une constante positive indépendante de t, $(F_\lambda)_\lambda$ la famille spectrale de TT^* et Q le projecteur orthogonal sur la fermeture de l'image de T. Une condition suffisante pour obtenir (1.100) est donnée par le lemme suivant [10, Lemme 6.3].

Lemma 1.1. *Soit*

$$G_\delta(t) := \int_0^t \lambda \, \mathrm{d}F_\lambda ||Qg^\delta - g||^2. \tag{1.101}$$

S'il existe $a > 1$ *et une constante* $d_a < 1$ *telle que*

$$\frac{G_\delta(a\,t)}{G_\delta(t)} \leq d_a\, a^2 \quad \forall\, t, \tag{1.102}$$

alors la condition (1.100) *est satisfaite.*

Si la condition (1.100) est satisfaite, alors la règle de quasi-optimalité converge pour la méthode de Tikhonov itéré et les taux de convergence peuvent être établis (voir [10, Theoreme 7.1-7.3]).

Theorem 1.1. *Soit* $\{g^\delta\}_{\delta>0}$ *une famille de données satisfaisant* (1.100) *et* $\alpha(g^\delta)$ *défini par* (1.99)*. Supposons que la solution* $f^\dagger$ *de l'équation* (1.7) *satisfasse la condition source* $f^\dagger \in \mathcal{R}(T^*T)^\mu$ *avec* $\mu \leq n$*, alors*

$$\left\| f^\dagger - f^\delta_{\alpha(g^\delta),n} \right\| = \mathcal{O}\left(\delta^{\frac{2\mu}{2\mu+1}\frac{\mu}{n}}\right)$$

En outre, nous pouvons obtenir un taux d'ordre optimal $\mathcal{O}\left(\delta^{\frac{2\mu}{2\mu+1}}\right)$ *si nous supposons en outre qu'il existe des constantes* c *et* $\bar{t}$ *telles que la solution* $f^\dagger$ *satisfait :*

$$t^{2n} \int_t^\infty \lambda^{-2n} \, \mathrm{d}\left\|E_\lambda f^\dagger\right\|^2 \geq c \int_0^t \mathrm{d}\left\|E_\lambda f^\dagger\right\|^2 = c\left\|E_t f^\dagger\right\|^2 \quad \forall\, 0 < t < \bar{t}. \tag{1.103}$$

Dans [12], la règle de quasi-optimalité a été étudiée pour la *spectral cut-off* et les résultats de convergence dans le cadre stochastique ont été donnés. Dans [66], la règle a été étudiée pour la méthode de Tikhonov itéré. La convergence de la version discrète de la règle de quasi-optimalité définie par (1.96) et (1.97) pour la méthode de Tikhonov est étudiée dans [11]. Une étude sur les résultats de convergence de la quasi-optimalité est menée dans [10] et des résultats de convergence sous-optimaux ont été donnés à la fois dans un cadre stochastique et déterministe pour la méthode Tikhonov ordinaire et itéré.

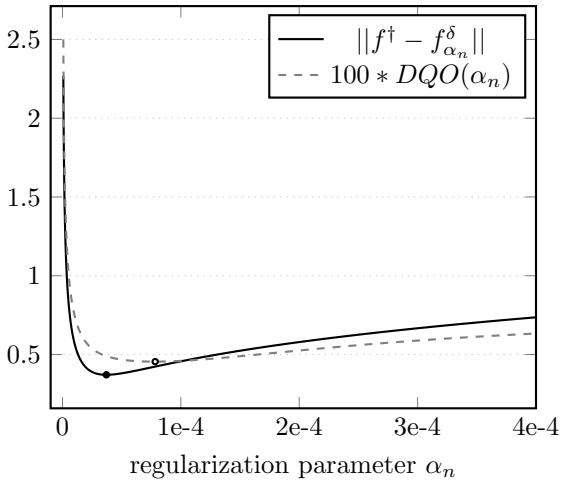

Figure 1.2: Courbe de comparaison de l'erreur $||f^\dagger - f^\delta_{\alpha_n}||$ et la courbe de la fonction DQO $DQO(\alpha_n) = ||f^\delta_{\alpha_{n+1}} - f^\delta_{\alpha_n}||$ et leurs minimiseurs correspondants. Ici, la suite $(\alpha_n)_n$ est définie par (1.96).

1.2 Régularisation par mollification

Dans cette section, nous présentons un état de l'art des méthodes de régularisation par mollification existante au début de ma thèse.

Étant donné que, par définition, la mollification atténue les composantes à haute fréquence qui sont généralement responsables de l'instabilité des problèmes mal-posés, il est donc naturel d'utiliser la mollification dans la régularisation des problèmes mal-posés. À cet égard, plusieurs auteurs ont développé diverses méthodes de régularisation basées sur la mollification. Dans toutes ces méthodes, malgré les différentes formulations, un aspect essentiel est préservé : l'idée de reconstruire, non pas l'objet original $f^\dagger$, qui est inaccessible étant donné la mal-position du problème, mais une version lisse de $f^\dagger$ définie via un *mollifier*. En fonction de leur formulation et de la procédure de régularisation, ces méthodes de régularisation peuvent être regroupées en trois classes principales.

Une première classe développée par Vasin [115], Murio [89, 90], Manselli et Miller [80] et Hào et al. [57]. Une deuxième classe référé connue sous le nom des *approximation inverse* développée par Louis et Mass [75, 76]. Et enfin une troisième classe basée sur une formulation variationnelle qui, à notre connaissance, a été introduite pour la première fois à la fin des années 80 par Lannes et al. [71] dans le traitement des signaux et des images. À présent, nous allons décrire le processus de régularisation dans chacune de ses trois classes.

1.2.1 Formulation de Vasin, Murio, Hegland, Anderssen, Hào et al.

Cette formulation a été conçue et appliquée à quelques problèmes, à savoir la différenciation numérique (entière et fractionnaire), l'équation intégrale d'Abel et les problèmes inverses de conduction thermique.

Ici, la régularisation de l'équation (1.7) consiste à pré-lisser la donnée g avant la résolution de l'équation. Plus précisément, la procédure de régularisation est la suivante: Tout d'abord, on définit une famille d'opérateurs continus $(M_\alpha)_{\alpha>0}$ définis sur H_2 par

$$M_\alpha g = m_\alpha \star g \quad \text{avec} \quad m_\alpha \quad \text{tel que} \quad \forall\, g \in H_2 \quad (m_\alpha \star g \to g \quad \text{quand} \quad \alpha \to 0)\,. \tag{1.104}$$

Ensuite, à l'aide de l'opérateur M_α, on définit la nouvelle équation

$$Tf = M_\alpha g \tag{1.105}$$

où l'opérateur M_α est pré-appliqué à la donnée g. À ce stade, la prochaine étape consiste à montrer que l'équation (1.105) est stable par rapport à la donnée g contrairement à l'équation initiale (1.7).

Comme illustration, considérons le problème de différenciation numérique mentionné dans la section 1.1.1.1. Soit $T : C^0(I) \to C^1(I)$ l'opérateur d'intégration sur I, où I désigne un intervalle de $\mathbb{R}$ et $g \in C^1(I)$. On sait que le pseudo inverse $T^\dagger$ de T n'est rien d'autre que l'opérateur de dérivation qui est discontinue de $C^1(I)$ vers $C^0(I)$ comme illustré en section 1.1.1.1.

L'approche décrite ci-dessus consiste à approximer la dérivée g' de g par la dérivée de $m_\alpha \star g$. En effet, on peut montrer que la dérivation de $m_\alpha \star g$ est une transformation continue par rapport à la donnée g vue que

$$\begin{aligned} \left\|(m_\alpha \star g_1)' - (m_\alpha \star g_2)'\right\|_\infty &\leq \left\| \int_{\mathbb{R}} \left| \frac{\partial}{\partial x} m_\alpha(x-s) \right| |g_1(s) - g_2(s)|\, \mathrm{d}s \right\|_\infty \\ &\leq \left\|g_1 - g_2\right\|_\infty \times \int_{\mathbb{R}} \left|m'_\alpha(s)\right|\, \mathrm{d}s. \end{aligned}$$

Ainsi si m_α est dérivable sur $\mathbb{R}$ et de dérivée intégrable sur $\mathbb{R}$, alors on a

$$\big\|(m_\alpha \star g_1)' - (m_\alpha \star g_2)'\big\|_\infty \leq C_\alpha \big\|g_1 - g_2\big\|_\infty \quad \text{avec} \quad C_\alpha = \int_{\mathbb{R}} \big|m'_\alpha(x)\big| \, \mathrm{d}x. \tag{1.106}$$

Ainsi pour $\alpha > 0$ fixé, en notant par $f_{\alpha,1}$ et $f_{\alpha,2}$ les solutions de l'équation (1.105) correspondant à g_1 et g_2, (1.106) équivaut à

$$\big\|f_{\alpha,1} - f_{\alpha,2}\big\|_\infty \leq C_\alpha \big\|g_1 - g_2\big\|_\infty$$

ce qui implique la stabilité de l'équation (1.105). Enfin la consistance de l'approximation de g' par $(m_\alpha \star g)'$ vient du fait que

$$(m_\alpha \star g)' = m_\alpha \star g' \to g' \quad \text{quand} \quad \alpha \to 0.$$

Notons que la méthode décrite ci dessus équivaut à introduire une régularisation par pré-multiplication du pseudo inverse $T^\dagger$ de T par M_α avant l'application de la donnée g. Plus précisément, la solution approchée de (1.7) n'est rien d'autre que

$$f_\alpha = T^\dagger M_\alpha g.$$

Il est important de noter que Murio [90] a également conçu une procédure de régularisation très similaire pour la résolution des problèmes inverses de conduction thermique où l'opérateur *mollifier* M_α, au lieu d'être appliqué aux données g est appliqué sur la solution inconnue $f^\dagger$. Ici, la procédure de régularisation consiste d'abord à calculer une équation $T_\alpha f = g$ satisfaite par la version mollifiée $M_\alpha f^\dagger$ de la solution $f^\dagger$ de l'équation (1.7) et ensuite à montrer que cette nouvelle équation est effectivement stable. Enfin, la solution approximative est définie comme la solution de l'équation stable $T_\alpha f = g$ c'est à dire

$$f_\alpha = T_\alpha^\dagger g.$$

Comme exemple, considérons l'équation de la chaleur unidimensionnelle

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{\partial^2 u}{(\partial x)^2} & \forall x \in (0,1), t > 0 \\ u(x,0) = 0 & \forall x \in (0,1) \end{cases} \tag{1.107}$$

associées aux conditions limites $u(1,t) = g(t)$ et $u(0,t) = f(t)$ pour tout $t > 0$, avec f et g appartenant à $L^2(\mathbb{R}_+)$. Considérons le problème qui consiste à estimer la température $f(t)$ à l'extrémité gauche de la tige étant donné $g(t)$.

En appliquant la transformé de Fourier suivant la variable t à l'équation (1.107), on obtient

$$\frac{\partial \hat{u}}{(\partial x)^2}(x,w) = 2\pi i w \hat{u}(x,w) \quad x > 0, \quad w \in \mathbb{R}$$

dont la solution générale prend la forme

$$u(x,w) = A(w) \exp\left(\sqrt{\pi w}(1 + \mathrm{sgn}(w)i)x\right) + B(w) \exp\left(-\sqrt{\pi w}(1 + \mathrm{sgn}(w)i)x\right),$$

ou $\mathrm{sgn}(w)$ désigne le signe w. Si on suppose que la température reste uniformément bornée quand x tend vers l'infini, alors on obtient

$$u(x,w) = A(w) \exp\left(-\sqrt{\pi w}(1 + \mathrm{sgn}(w)i)x\right). \tag{1.108}$$

En appliquant (1.108) pour $x = 0$ et $x = 1$, on obtient que

$$\hat{g}(w) = \hat{f}(w) \exp\left(-\sqrt{\pi w}(1 + \mathrm{sgn}(w)i)\right) \tag{1.109}$$

ce qui implique que

$$\hat{f}(w) = \hat{g}(w) \exp\left(\sqrt{\pi w}(1 + \mathrm{sgn}(w)i)\right). \tag{1.110}$$

L'équation (1.110) illustre la mal-position du problème qui consiste à calculer f à partir de g. À présent, au lieu de chercher à reconstruire f, intéressons nous plutôt à une version mollifiée de f i.e. $f \star m_\alpha$. Considérons par exemple un noyau de convolution gaussien, i.e.

$$m_\alpha(t) = \frac{1}{\alpha\sqrt{2\pi}} \exp(-\frac{t^2}{2\alpha^2})$$

Ainsi à partir de (1.110) et la transformation de la convolution en produit par la transformée de Fourier, on déduit que l'équation satisfaite par $f \star m_\alpha$ est

$$\mathcal{F}(f \star m_\alpha)(w) = \mathcal{F}(g)(w) \exp\left(\sqrt{\pi w}(1 + \mathrm{sgn}(w)i) - 2\pi^2\alpha^2 w^2\right). \tag{1.111}$$

L'équation (1.111) démontre bien que le problème de reconstruire l'objet mollifié $f \star m_\alpha$ à partir de g est un problème stable contrairement au problème initial de reconstruction de f à partir de g. Ainsi, la seconde procédure de régularisation proposé par Murio [90] consiste à approximer f suivant (1.111) par

$$f_\alpha = \mathcal{F}^{-1}\left(\mathcal{F}(g)(w) \exp\left(\sqrt{\pi w}(1 + \mathrm{sgn}(w)i) - 2\pi^2\alpha^2 w^2\right)\right).$$

La consistance de l'approximation découle une fois de plus du fait que $m_\alpha \star f \to f$ quand $\alpha \to 0$. Notons que à partir de l'équation de la chaleur (1.107), et des conditions au limites données par f et g, on peut déduire que l'équation au dérivées partielles satisfaite par la température mollifiée $u(x, \cdot) \star m_\alpha$ est

$$\begin{cases} \frac{\partial u}{\partial t}(x,t) = \frac{\partial^2 u}{(\partial x)^2}(x,t) & \forall x \in (0,1), t > 0 \\ u(x,0) = 0 & \forall x \in (0,1) \\ u(1,t) = (m_\alpha \star g)(t) & \forall t > 0, \\ u(0,t) = (m_\alpha \star f)(t) & \forall t > 0. \end{cases} \tag{1.112}$$

Enfin, il est important de noter que les procédures de regularizations que nous venons d'illustrer peuvent aussi s'appliquer pour le cas multidimensionnelle.

Dans les articles [57, 89, 90, 115], nous pouvons voir les applications numériques de ces schémas de régularisation et observer leur efficacité pour les problèmes de type différenciation numérique (entière ou fractionnaire), équation intégral d'Abel et problèmes inverses de conduction thermique. Cependant, un inconvénient majeur de cette méthode est sa faible généralisation à une classe plus large de problèmes inverses au-delà des exemples cités précédemment.

Hegland et Anderssen [59] ont proposé une généralisation en utilisant la théorie des semi-groupes de Hille et Phillips [63] dans le cas où $H_1 = H_2 = L^2(\mathbb{R})$. Ils ont ainsi définis deux types de façons de régulariser à l'aide d'un opérateur de mollification M_α: La *range mollification* où l'opérateur de régularisation R_α est défini par

$$R_\alpha = T^\dagger M_\alpha \tag{1.113}$$

et la *domain mollification* où R_α est défini par

$$R_\alpha = \overline{M_\alpha T^\dagger}, \tag{1.114}$$

où $\overline{M_\alpha T^\dagger}$ désigne une extension continue de $M_\alpha T^\dagger$ à H_2. Cependant, la définition (1.113) (resp. (1.114)) n'est valable que si l'opérateur $T^\dagger M_\alpha$ (resp. $M_\alpha T^\dagger$) est borné, ce que les auteurs ont démontré pour les opérateurs T invariant par translation. Ainsi, la généralisation de Hegland et Anderssen [59] est encore faible car elle n'intègre que les equations $Tf = g$ où T est invariant par translation. De plus, un autre inconvénient important de cette généralisation est le calcul de la solution approximative f_α, qui nécessite l'inversion directe de l'opérateur T (i.e. la connaissance du pseudo inverse $T^\dagger$), couplé au fait qu'une expression explicite de l'inverse de T n'est généralement pas disponible.

1.2.2 Les approximate inverses

Une deuxième classe de méthode de mollification se réfère à la méthode dite des *approximate inverse* développée par Louis et Mass [75, 76]. La philosophie de la régularisation de l'équation (1.7) est la suivante. En raison de la mal-position de l'équation (1.7), il est difficile de déterminer la solution $f^\dagger$ avec précision. Étant donné cela, on pourrait plutôt essayer de récupérer une version lisse $E_\gamma f^\dagger$ de $f^\dagger$, où E_γ est un opérateur de mollification approprié. De toute évidence, l'opérateur E_γ doit satisfaire à certaines conditions qui garantissent que $E_\gamma f^\dagger$ est une approximation raisonnable de $f^\dagger$. Dans [76], cette condition est simplement la faible convergence de $E_\gamma f$ vers f pour tout $f \in H_1$ quand γ tend vers 0. C'est-à-dire,

$$E_\gamma f \rightharpoonup f, \quad \forall f \in H_1, \quad \text{quand} \quad \gamma \to 0.$$

À condition que H_1 soit un espace de fonction approprié, l'opérateur E_γ est défini via un *mollifier* e_γ comme suit :

$$\left(E_\gamma f\right)(x) := \left\langle e_\gamma(x,\cdot), f \right\rangle.$$

Ensuite, si e_γ se trouve dans l'image de l'adjoint T^* de T, c'est-à-dire qu'il existe $v_{x,\gamma} \in H_1$ tel que

$$T^* v_{x,\gamma} = e_\gamma(x,\cdot), \tag{1.115}$$

nous pouvons facilement calculer la solution approximative $E_\gamma f^\dagger$ comme

$$\left(E_\gamma f^\dagger\right)(x) := \left\langle e_\gamma(x,\cdot), f^\dagger \right\rangle = \left\langle T^* v_{x,\gamma}, f^\dagger \right\rangle = \left\langle v_{x,\gamma}, T f^\dagger \right\rangle = \left\langle v_{x,\gamma}, g \right\rangle.$$

Ainsi, la solution approximative f_γ des *approximate inverses* n'est rien d'autre que

$$f_\gamma = \left\langle v_{x,\gamma}, g \right\rangle, \tag{1.116}$$

où $v_{x,\gamma}$ est la solution de l'équation (1.115). Au vue de (1.116), il en découle que la résolution de l'équation (1.7) se réduit à celle de l'équation (1.115) qui est également mal-posée. Cependant, contrairement à l'équation (1.7), les données $e_\gamma(x,\cdot)$ dans l'équation (1.115) peuvent être considérées comme exactes étant donné que la fonction e_γ choisie est généralement définie analytiquement. Dans le cas où l'équation (1.115) n'a pas de solution, la fonction $v_{x,\gamma}$ est calculée comme la solution des moindres carrés

$$\min_{v \in H_2} \quad \left\| T^* v - e_\gamma(x,\cdot) \right\|^2,$$

qui est équivalent à l'équation normale

$$TT^* v_{x,\gamma} = Te_\gamma(x,\cdot).$$

Dans [76], les auteurs ont souligné certains avantages de la méthode, parmi lesquels l'absence de discrétisation de f, le fait que le paramètre de régularisation n'apparaisse qu'à droite de l'équation (1.115) et l'adaptabilité de la méthode au *parallel processing*.

Cependant, le calcul du noyau de reconstruction $v_{x,\gamma}$ reste une étape critique étant donné la mal-position de l'équation (1.115). En effet, même si $e_\gamma(x,\cdot)$ peut avoir une expression analytique, dans la résolution numérique, nous avons toujours des erreurs d'arrondi et aussi des erreurs de troncature lorsque $e_\gamma(x,\cdot)$ est exprimé sous forme de série. Ainsi, en pratique, sauf lorsque la paire *noyau de reconstruction-mollifier* ($v_{x,\gamma}$, e_γ) est explicitement connu, la méthode des *approximate inverses* a un inconvénient principal qui est le calcul du noyau de reconstruction $v_{x,\gamma}$. Rieder et Schuster [101, 102, 103] ont étudié cette méthode en profondeur et l'ont appliquée à la tomographie en utilisant une paire explicite connue *noyau de reconstruction-mollifier* ($v_{x,\gamma}$, e_γ). Pour une étude approfondie de la méthode des *approximate inverses* et plusieurs applications, voir par exemple [108].

1.2.3 Formulation variationnelle

La troisième classe est la formulation variationnelle qui, à notre connaissance, a été introduite pour la première fois à la fin des années 80 par Lannes et al. [71] dans le traitement des signaux et d'images. Par la suite, Alibaud et al. [1] ont défini une formulation variationnelle de la mollification pour résoudre le problème de la synthèse de Fourier décrit dans la section (1.1.1.3). Enfin, Bonnefond et Maréchal [16] généralise la formulation variationnelle à l'inversion d'opérateur linéaire compact.

Dans cette formulation, $H_1 = L^2(\mathbb{R}^p)$ et l'objet cible est $C_\beta f^\dagger$ où C_β est un opérateur de convolution. Étant donné que l'objet cible $C_\beta f^\dagger$ doit rester proche de $f^\dagger$, la famille d'opérateur $(C_\beta)_{\beta>0}$ est définie comme une approximation de l'unité, c'est-à-dire

$$\forall f \in H_1, \quad C_\beta f \to f \quad \text{as } \beta \downarrow 0. \tag{1.117}$$

Dans les articles de Alibaud et al. [1] et Bonnefond et Maréchal [16], l'opérateur C_β est defini par

$$\forall f \in L^2(\mathbb{R}^p), \quad C_\beta f = \phi_\beta \star f, \quad \text{où} \quad \phi_\beta(x) = \frac{1}{\beta^p}\phi\left(x/\beta\right) \tag{1.118}$$

avec $\phi \in L^1(\mathbb{R}^p)$ satisfaisant $\int_{\mathbb{R}^p} \phi(x)\,\mathrm{d}x = 1$. Après avoir défini l'objet cible $C_\beta f^\dagger$, ils définissent la nouvelle équation satisfaite par l'objet cible. À condition qu'il existe un opérateur continu Φ_β solution de l'équation dite d'entrelacement

$$TC_\beta = \Phi_\beta T, \tag{1.119}$$

l'équation satisfaite par l'objet cible $C_\beta f^\dagger$ est $Tf = \Phi_\beta g$. D'autre part, étant donné que nous visons à ce que $C_\beta f^\dagger$ reste aussi proche que possible de $f^\dagger$, un terme de pénalité naturel est $\left\|(I - C_\beta)f\right\|^2$. Ainsi, si un opérateur continu Φ_β solution de (1.119) existe, la formulation variationnelle

$$\min_f \quad \left\|\Phi_\beta g - Tf\right\|^2_{H_2} + \left\|(I - C_\beta)f\right\|^2_{L^2(\mathbb{R}^p)}, \tag{1.120}$$

est légitime. La formulation (1.120) est celle considérée par Alibaud et al. dans [1] où T est un opérateur de Fourier tronqué. À l'exception que dans la formulation étudiée dans [1], un paramètre α est ajouté

comme poids du terme de pénalisation $\left\|(I - C_\beta)f\right\|^2$. Cependant, dans [1], il est démontré qu'un tel paramètre n'est pas vraiment significatif et peut être omis.

Étant donné la présence de l'opérateur Φ_β dans la formulation variationnelle (1.120), il est nécessaire de s'intéresser à l'équation d'entrelacement (1.119) dont la solution défini l'opérateur Φ_β. De prime à bord, il est important de noter que dans un cadre général d'opérateur linéaire T, un opérateur Φ_β satisfaisant l'équation (1.119) peut ne pas exister. Pour faire face à cet éventualité, Bonnefond et Maréchal [16] ont proposé de remplacer l'équation (1.119) par le problème d'optimisation

$$(\mathscr{L}_\beta) \qquad \Phi_\beta = \operatorname{argmin}\left\{ \left\|TC_\beta - XT_{|E}\right\|^2_{\mathcal{B}(E,L^2(\mathbb{R}^p))} \,\middle|\, X \in \mathcal{B}(L^2(\mathbb{R}^p)), \quad X = 0 \text{ on } \left(\mathcal{R}(T_{|E})\right)^\perp \right\},$$

où $T_{|E}$ désigne la restriction de l'opérateur T au sous espace E défini par

$$E = L^2(V) \cap H^s(\mathbb{R}^p) \quad \text{avec} \quad L^2(V) = \left\{ f \in L^2(\mathbb{R}^p) \,\middle|\, \operatorname{supp}(f) \subset V \right\},$$

V étant un sous espace de $\mathbb{R}^p$. Par ailleurs, notons qu'il existe de nombreux cas où l'opérateur Φ_β est connu et défini explicitement.

Example 1.1. *Pour le problème de deconvolution [83], on a $Tf = \gamma \star f$ et donc $TC_\beta f = \gamma \star (\phi_\beta \star f) = \phi_\beta \star (\gamma \star f)$ d'où la solution Φ_β de l'équation (1.119) est tout simplement $\Phi_\beta = C_\beta$.*

Example 1.2. *Considérons le problème de synthèse de Fourier, i.e. $Tf = 1_W \mathcal{F}(f)$. Alors $TC_\beta f = 1_W \mathcal{F}(\phi_\beta \star f) = 1_W \mathcal{F}(\phi_\beta)\mathcal{F}(f) = \mathcal{F}(\phi_\beta)Tf$ d'où la solution Φ_β de l'équation (1.119) est tout simplement le produit par la transformée de Fourier de ϕ_β, i.e. $\Phi_\beta f = \mathcal{F}(\phi_\beta)f$.*

Allant plus loin, on peut aussi définir explicitement Φ_β pour le problème de différenciation numérique, et d'inversion de la transformée de Radon.

À présent, intéressons nous au problème $(\mathscr{L}_\beta)$ qui définit l'opérateur Φ_β dans le cadre général. À cet effet, rappelons que les auteurs de [16] ont précisé que dans la plupart des cas, le problème $(\mathscr{L}_\beta)$ est un problème d'optimisation mal-posé (voir [16, p. 5]). Néanmoins, comme mentionnés par les auteurs de [16], sous certaines conditions raisonnable la solution du problème $(\mathscr{L}_\beta)$ peut être définit de façon explicite.

Proposition 1.8. *Soit $T : L^2(\mathbb{R}^p) \to H_2$ un opérateur linéaire continue. Si $TC_\beta T^\dagger$ est borné, alors $TC_\beta T^\dagger$ peut s'étendre à un opérateur continu sur H_2 qui est la solution du problème $(\mathscr{L}_\beta)$.*

La proposition précédente qui correspond à la proposition 3.1 de [16] illustre le fait que la difficulté du problème $(\mathscr{L}_\beta)$ réside essentiellement dans la discontinuité de $T^\dagger$. En effet, la discontinuité de $T^\dagger$ pourrait induire celle de l'opérateur $TC_\beta T^\dagger$. Dans le cas où T est un opérateur intégral, Bonnefond et Maréchal ont donné des conditions suffisantes pour garantir que $TC_\beta T^\dagger$ est borné. Ces conditions sont résumées dans la proposition suivante (voir, [16, Proposition 3.3]).

Proposition 1.9. *Soit T un opérateur intégral de noyau k. Supposons que*

- *$\int_{\mathbb{R}^p}\int_{\mathbb{R}^p} |k(x,y)|^2 \,\mathrm{d}x\,\mathrm{d}y < \infty$, c'est à dire que T est un opérateur de Hilbert-Schmidt;*
- *pour tout $x, y, z \in \mathbb{R}^p$, $k(x, y+z) = k(x,y)g(x,z)$;*

- *il existe une constante positive M_φ dépendant de φ uniquement tel que*

$$\forall x \in \mathbb{R}^p, \quad \left| \int_{\mathbb{R}^p} \varphi(z) g(x,z)\, \mathrm{d}z \right| \leq M_\varphi.$$

Alors, $TC_\beta T^\dagger$ est borné sur son domaine.

Avant d'arriver au résultats de consistance de l'approximation définie à travers la formulation variationnelle (1.120), Bonnefond et Maréchal assume l'hypothèse suivante.

Hypothèse 1.1. *V est un domaine borné de $\mathbb{R}^p$ contenant le support de la solution inconnue $f^\dagger$, $T : L^2(\mathbb{R}^p) \to H_2$ est un opérateur linéaire injectif et borné dont la restriction $T|_{L^2(V)}$ à $L^2(V)$ est compact.*

Avec l'hypothèse ci-dessus, on obtient que l'objet inconnu $f^\dagger$ a un support borné. Cependant il est à noté que l'objet cible $C_\beta f^\dagger = \phi_\beta \star f^\dagger$ a un support plus large voir même non borné. Néanmoins, les auteurs de [16] reconstruisent l'objet dans $L^2(V_1)$ où V_1 est un domaine borné de $\mathbb{R}^p$ contenant V, i.e. $V \subset V_1$. Ainsi, à partir de la formulation variationnelle (1.120), la solution approximative f_β de $f^\dagger$ est définie pas

$$f_\beta = \underset{f \in L^2(V_1)}{\operatorname{argmin}} \left\| \Phi_\beta g - Tf \right\|_{H_2}^2 + \left\| (I - C_\beta) f \right\|_{L^2(\mathbb{R}^p)}^2, \tag{1.121}$$

où Φ_β est défini soit par (1.119) ou $(\mathscr{L}_\beta)$. Une fois défini l'approximation f_β ci dessus, le résultat suivant de consistance est établi dans [16, Théorème 4.1].

Théorème 1.1. *Considérons l'hypothèse 1.1. Pour tout $\beta \in (0,1]$, considérons l'opérateur C_β défini par (1.118) et supposons que le problème $(\mathscr{L}_\beta)$ admette une solution Φ_β. Alors*

- *Pour tout $\beta \in (0,1]$, f_β donné par (1.121) est bien défini et dépend de façon continue à $g \in H_2$.*
- *Supposons que $\mathcal{F}(\phi)(\xi) \neq 1$ pour tout $\xi \in \mathbb{R}^p \setminus \{0\}$ et qu'il existe deux constantes K et s tel que*

$$|1 - \mathcal{F}(\phi)| \sim_{\xi \to 0} K|\xi|^s.$$

Pour tout $g \in \mathcal{D}(T^\dagger)$ tel que $T^\dagger g \in L^2(V) \cap H^s(\mathbb{R}^p)$, la solution approximative f_β défini par (1.121) converges fortement vers $T^\dagger g$ dans $L^2(\mathbb{R}^p)$.

1.3 Plan de la thèse

La section 1.2 présenté ci dessus décrit l'état de l'art de la régularisation par mollification au début de ma thèse. Ma thèse s'est focalisé sur la formulation variationnelle de la mollification avec pour objectif l'application au problème de régression non-paramétrique instrumentale. Durant mes trois années de thèses, j'ai effectué les travaux suivant:

- J'ai collaboré à la finalisation d'un article débuté par mon directeur de thèse Pierre Maréchal et Xavier Bonnefond portant sur la mise en œuvre d'un algorithme pour l'implémentation du principe de Morozov pour les méthodes de régularisation du type Tikhonov.
- J'ai développé une nouvelle méthode de régularisation qui est particulièrement approprié pour la régularisation des problèmes exponentiellement mal-posés.

- J'ai collaboré avec mon directeur de thèse Pierre Maréchal et ma co-directrice de thèse Anne Vanhems à l'application de la formulation variationnelle de la mollification au problème de régression non-paramétrique instrumentale. Pour ce problème de régression instrumentale, il n'y a pas de garanti que l'opérateur Φ_β solution du problème $(\mathscr{L}_\beta)$ existe. Autrement dit, il n'y a pas de preuve que $TC_\beta T^\dagger$ est borné dans ce cas.

  Nous avons dans un premier temps défini Φ_β par $TC_\beta T^\dagger$. Étant donné que $T^\dagger$ est non-borné, nous avons utilisé les algorithmes itératifs (régularisant) de type proximal [44, 45, 104] et les méthodes de Krylov [118] pour l'évaluation de $T^\dagger$. Cependant, les résultats obtenus n'étaient pas satisfaisant ce qui pourrait indiquer que l'opérateur $TC_\beta T^\dagger$ n'est pas borné dans le cas sous-jacent. Ainsi nous avons décidé d'omettre l'opérateur Φ_β dans la formulation variationnelle qui définit la solution approximative f_β. Il est important de rappeler que l'opérateur Φ_β a pour unique objectif d'adapter la donnée g au nouvel objet cible $C_\beta f$. Ainsi en supposant que $C_\beta f$ est assez proche de f (i.e. $\beta \ll 1$), on peut omettre l'opérateur Φ_β qui est alors remplacé par l'identité de H_2. D'autres part, d'un point de vue théorique, l'omission de Φ_β dans la formulation (1.121) ne change pas les résultats de consistance mais simplifie la preuve de la consistance.

La suite du manuscrit s'articule autour de 3 articles (publiés ou soumis) dans l'ordre suivant:

Le chapitre 2 est consacré à la conception d'un algorithme simple et efficace pour le calcul du paramètre de régularisation correspondant au principe de Morozov. Dans ce chapitre, nous explorons la dualité découlant de l'exigence selon laquelle le résidu $\left\|Tf_\alpha^\delta - g^\delta\right\|$ doit prendre une valeur donnée basée sur l'estimation du niveau de bruit. Nous prouvons que, dans des hypothèses raisonnables, la fonction duale est lisse, et que sa maximisation indique la valeur appropriée du paramètre de régularisation de Tikhonov. En outre, nous montrons que la fonction duale et sa dérivé peuvent être facilement évaluées. Enfin, nous décrivons un algorithme de Quasi-Newton combiné à la recherche linéaire de Wolfe-Lemarechal pour la résolution du problème dual. La pertinence numérique de notre approche est établie au moyen d'un exemple illustratif sur le problème de la régression instrumentale non paramétrique.

Le chapitre 3 présente une nouvelle méthode de régularisation adaptée aux problèmes linéaires exponentiellement mal-posés. Sous des conditions source de type logarithmique (qui ont une interprétation naturelle en termes d'espaces de Sobolev dans le contexte mentionné précédemment), les concepts de qualification ainsi que les taux de convergence d'ordre optimales sont présentés. L'optimalité sous des conditions source générales définie à laide de fonctions index est également étudiée. Enfin, nous comparons la nouvelle méthode de régularisation aux méthodes classiques tel que Tikhonov, la *spectral cut-off*, la régularisation asymptotique et la méthode des gradients conjugués pour la régularisation de trois problèmes test mal-posés.

Dans le chapitre 4, nous appliquons l'approche variationnelle de la mollification sans l'opérateur Φ_β à la régularisation du problème de la régression instrumentale non-paramétrique. Nous étudions les propriétés asymptotiques de notre estimateur dans le cadre stochastique classique où l'opérateur T et la donnée g sont approximés. Pour une applicabilité pratique de notre approche, nous appliquons une règle de sélection de paramètres empirique qui donne de bons résultats. Enfin, nous effectuons une simulation de Monté-Carlo afin de confirmer l'efficacité de notre estimateur couplé à la règle de sélection empirique

de paramètre utilisée.

Annexe

Preuve de la proposition 1.2. Prouvons que (1.11) est une condition nécessaire et suffisante pour la mal-position de l'équation (1.7).

Si (1.11) est satisfaite, alors il existe une suite $(f_k)_k$ de fonction dans $(\mathcal{N}(T))^\perp$ telle que $\|f_k\| = 1$ et $\|Tf_k\| \to 0$. Soit $\tilde{f}_k = f_k/\|Tf_k\|$ alors par définition, $||\tilde{f}_k||$ diverge alors que $||T\tilde{f}_k||$ reste égal à 1.

Inversement, supposons qu'il existe une suite $(f_k)_k$ de fonctions dans $(\mathcal{N}(T))^\perp$ telle que $\|f_k\|$ diverge alors que $\|Tf_k\|$ est borné. Soit $\tilde{f}_k = f_k/\|f_k\|$ alors par définition, $||\tilde{f}_k|| = 1$ et $||T\tilde{f}_k|| = \|Tf_k\|/\|f_k\| \to 0$. ■

Preuve de la Proposition 1.6. Soit $f^\dagger = T^\dagger g$, en utilisant (1.31) et le théorème de convergence dominée, on a

$$\begin{aligned} \left\|\upsilon_\alpha(T^*T)T^*g - T^\dagger g\right\|^2 &= \left\|\upsilon_\alpha(T^*T)T^*Tf^\dagger - f^\dagger\right\|^2 \\ &= \int_0^{||T||_+^2} |1-\lambda\upsilon_\alpha(\lambda)|^2 \,\mathrm{d}||E\lambda f^\dagger||^2 \\ &\to \int_0^{||T||_+^2} \lim_{\alpha\to 0} |1-\lambda\upsilon_\alpha(\lambda)|^2 \,\mathrm{d}||E\lambda f^\dagger||^2 \quad \text{quand} \quad \alpha \to 0. \end{aligned} \tag{1.122}$$

De (1.30), on obtient que pour tous les $\lambda > 0$, $1-\lambda\upsilon_\alpha(\lambda) \to 0$ quand $\alpha \to 0$, ce qui implique grâce à (1.122) que

$$\left\|\upsilon_\alpha(T^*T)T^*g - T^\dagger g\right\|^2 \to \lim_{\lambda\to 0} \left\|E_\lambda f^\dagger\right\|^2 - ||E_0 f^\dagger||^2 = \left\|P_{\mathcal{N}(T)} f^\dagger\right\|^2, \quad \text{quand} \quad \alpha \to 0, \tag{1.123}$$

où $P_{\mathcal{N}(T)}$ est le projecteur orthogonal sur le noyau de T. Ainsi de (1.123), étant donné que $f^\dagger \in \mathcal{N}(T)^\perp$, on obtient que $\left\|\upsilon_\alpha(T^*T)T^*g - T^\dagger g\right\|^2 \to 0$ quand $\alpha \to 0$. ■

Chapter 2

Note sur le Principe de Morozov

Dans ce chapitre, on s'intéresse à la sélection du paramètre de régularisation pour les méthodes de régularisation du type Tikhonov. En particulier, on considère le principe de Morozov qui est une règle de sélection de paramètre assez répandue. En utilisant la dualité à la contrainte suivant laquelle le résidu doit être de l'ordre du niveau de bruit dans les données, on définit un algorithme simple et rapide pour le calcul du paramètre de régularisation selon le principe de Morozov pour la méthode de Tikhonov. On démontre que le calcul du paramètre de régularisation se résume à la maximisation d'une fonction lisse et concave en dimension 1. Un apport secondaire de l'algorithme que nous proposons est un moyen de juger de la pertinence de l'estimation du niveau de bruit. En effet, selon le comportement de la fonction objective à maximiser, on peut déduire si le niveau de bruit estimé est trop petit ou trop grand. Enfin, l'efficacité et la rapidité de notre approche est confirmé par une application numérique au problème de régression instrumentale non-paramétrique.

L'article qui suit a été coécrit avec Pierre Maréchal et Xavier Bonnefond et a été publié dans le journal *Set-Valued and Variational Analysis Theory and Applications* [17]. Ma contribution dans cet article (dont l' écriture initiale est antérieure au début de mes travaux de thèse) se résume principalement à la section numérique où l'algorithme est appliqué au problème de régression instrumentale ainsi qu'à une révision partielle du Théorème 6.1 et sa preuve.

A note on the Morozov principle via Lagrange duality

Xavier Bonnefond, Pierre Maréchal and Walter Cedric Simo Tao Lee

Université de Toulouse, France

Abstract

Considering a general linear ill-posed equation, we explore the duality arising from the requirement that the discrepancy should take a given value based on the estimation of the noise level, as is notably the case when using the Morozov principle. We show that, under reasonable assumptions, the dual function is smooth, and that its maximization points out the appropriate value of Tikhonov's regularization parameter. The numerical relevance of our approach is established by means of an illustrative example from nonparametric instrumental regression, a standard problem in statistics.

Keywords: Ill-posed problems, Morozov principle, Lagrange duality, Variational analysis.

Mathematics Subject Classifications (2010) 34K29, 65K05, 65K10

1 Introduction

Let us consider the ill-posed linear inverse problem

$$\mathscr{A} f = g, \tag{1}$$

in which $\mathscr{A}$ maps the Hilbert space F into the Hilbert space G linearly, $g \in G$ is the data and $f \in F$ is the unknown. As usual, we assume that

$$g = g_0 + \delta g$$

where $g_0 := \mathscr{A} f_0$ for some $f_0 \in F$ and δg is the unknown noise. In many applications, the ill-posedness of (1) is the consequence of the compactness of $\mathscr{A}$. Numerous strategies have been developed in the last decades to regularize such problems. The variational approach consists in defining the reconstructed object as the minimizer of some functional, which usually is the sum of a *fit term* and of a *regularization term*. More precisely, in this approach, a family of such functionals is considered, which

depends on one parameter α (or more). A natural requirement is that, for positive values of α, the corresponding variational problem is well-posed, while letting $\alpha \downarrow 0$ yields, at the limit, a least square solution of Problem (1).

In practice, the choice of α is a crucial step. As a matter of fact, large values of α correspond to coarse approximations of the original model, while small values cause high sensitivity of the solution to perturbations on the data side (which we may call *hypersensitivity*).

In [9], it was shown that the Golub-Kahan bidiagonalization algorithm enables the estimation of the noise level. This was achieved by observing the corruption by noise of the iterates produced by the algorithm. An estimation of the noise level in the data may also be obtained from the modelling of the data acquisition process. At all events, it then seems reasonable to find a value of α such that the corresponding solution f_α of Problem $(\mathscr{P}_\alpha)$ satisfies

$$\|\mathscr{A} f_\alpha - g\|^2 \simeq \tau^2, \tag{2}$$

in which τ is an estimation of $\|\delta g\|$. The Morozov Principle states that α should be chosen in such a way that Eq. (2) is satisfied exactly with τ replaced by $c\tau$, where c is a constant strictly greater than 1, and in fact close to 1. This principle is often used as a stopping criterion for iterative regularization schemes (see, e.g, [4]).

In this note, we explore the duality arising from the application of the Morozov principle. Our analysis is based on the simple observation that the nonconvex constraint $\|\mathscr{A} f - g\| = c\tau$ can be dealt with via the convex constraint $\|\mathscr{A} f - g\| \leq c\tau$. Smoothness properties of the dual problem will then be obtained, as well as the explicit link between the shape of the dual function and the relevance of the estimation of the noise level.

In Section 2, we fix the context and make comments on the interplay between the various manners to relax the initial constraint equation (1). In Section 3, we explore the duality arising from a constraint of the form (2). Our reference books on inverse problems are [14, 2, 10], and our reference books on variational and convex analysis are [1, 7, 8, 13, 16].

2 Notation and preliminary remarks

The spaces F and G are endowed with the norms $\|\cdot\|_F$ and $\|\cdot\|_G$ associated with the inner products $\langle\cdot,\cdot\rangle_F$ and $\langle\cdot,\cdot\rangle_G$, respectively. We shall frequently omit the subscripts F and G, since most of the time the context leaves no ambiguity.

Most variational regularization techniques consist in defining the reconstructed object as the solution of

$$(\mathscr{P}_\alpha) \quad \left| \begin{array}{rl} \text{Minimize} & \|\mathscr{A} f - g\|^2 + \alpha J(f) \\ \text{s.t.} & f \in F, \end{array} \right.$$

in which J is the so called *regularizer*. It is customary to make the following assumption:

Assumption 1. The function J is proper convex, lower semi-continuous and coercive.

Recall that a function $J(f)$ is said to be coercive if $J(f) \to \infty$ as $\|f\| \to \infty$. In this paper, we also make throughout the additional (reasonable) assumption:

Assumption 2. The function J is strictly convex along $\ker \mathscr{A}$, nonnegative and vanishes on $\ker \mathscr{A}$.

It is well known that the solution f_α to Problem $(\mathscr{P}_\alpha)$ is also solution to the following constrained problem:

$$(\mathscr{Q}) \quad \left| \begin{array}{ll} \text{Minimize} & J(f) \\ \text{s.t.} & \|\mathscr{A}f - g\|^2 = \varepsilon \end{array} \right.$$

with $\varepsilon = \varepsilon(\alpha) := \|\mathscr{A}f_\alpha - g\|^2$. This is a particular case of Everett's lemma (see [8], for example). Lagrange duality makes it possible to go the other way around: starting from a problem such as $(\mathscr{Q})$ with $\varepsilon = \tau^2$ (since we wish to prescribe the tolerance τ), we may compute the value of α ensuring that f_α is also solution to $(\mathscr{P}_\alpha)$.

Notice that Problem $(\mathscr{Q})$ is not convex, but that whenever

$$\|g\|^2 \geq \varepsilon \tag{3}$$

(which is a natural assumption since the noise is usually smaller than the data), Problem $(\mathscr{Q})$ is equivalent to

$$(\mathscr{Q}^*) \quad \left| \begin{array}{ll} \text{Minimize} & J(f) \\ \text{s.t.} & \|\mathscr{A}f - g\|^2 \leq \varepsilon \end{array} \right.$$

Indeed the only case in which the solutions of these problems are different occurs when the solution f^* of Problem $(\mathscr{Q}^*)$ satisfies $\|\mathscr{A}f^* - g\|^2 < \varepsilon$. This means that the constraint is not active and that the optimality condition reads

$$0 \in \partial J(f^*).$$

According to Assumption 2, this yields $\mathscr{A}f^* = 0$, so that $\|g\|^2 < \varepsilon$, in contradiction with (3).

For convenience, we shall speak of *tolerance* or *penalized* formulations in order to refer to problems such as $(\mathscr{Q}^*)$ or $(\mathscr{P}_\alpha)$, respectively.

3 Duality for the Morozov principle

From now on, we fix $\varepsilon = \tau^2$ in Problem $(\mathscr{Q}^*)$. The Lagrangian of $(\mathscr{Q}^*)$ is given by

$$L(f, \lambda) := J(f) + \lambda \left(\|\mathscr{A}f - g\|^2 - \varepsilon \right), \quad f \in F,\ \lambda \in \mathbb{R}_+,$$

and the Lagrange problem associated to $(\mathscr{Q})$ reads

$$(\mathscr{L}_\lambda) \left| \begin{array}{ll} \text{Minimize} & L(f, \lambda) \\ \text{s.t.} & f \in F. \end{array} \right.$$

We see right away that the above Lagrange problem is equivalent, for $\lambda > 0$, to the Tikhonov problem $(\mathscr{P}_\alpha)$ with $\alpha = 1/\lambda$. It is convex, and the first order optimality condition reads:

$$0 \in \partial J(f) + 2\lambda \left(\mathscr{A}^* \mathscr{A} f - \mathscr{A}^* g \right). \tag{4}$$

From Assumptions 1 and 2, it is readily seen that, for every $\lambda > 0$, Problem $(\mathscr{L}_\lambda)$ has a unique solution f_λ satisfying (4).

The *dual function* is defined as the optimal value in the Lagrange problem:

$$D(\lambda) := \inf\{L(f,\lambda) | f \in F\} = L(f_\lambda, \lambda), \quad \lambda \in \mathbb{R},$$

and the dual problem associated with $(\mathscr{Q})$ is:

$$\left| \begin{array}{rl} \text{Maximize} & D(\lambda) \\ \text{s.t.} & \lambda \in \mathbb{R}, \end{array} \right.$$

which is obviously equivalent to

$$(\mathscr{D}) \left| \begin{array}{rl} \text{Maximize} & D(\lambda) \\ \text{s.t.} & \lambda > 0. \end{array} \right.$$

Before stating the main result of this section, we recall basic facts about the subdifferential of a supremum of convex functions.

Proposition 3. Let Y be any set, and let $\varphi\colon Y \times \mathbb{R}^n \to \mathbb{R}$ be a mapping such that, for every $y \in Y$, the mapping $\varphi(y,\cdot)$ is convex. Let

$$\Phi(x) := \sup_{y \in Y} \varphi(y,x).$$

Then, for every $x \in \mathbb{R}^n$,

$$\partial\Phi(x) \supset \bigcup_{y \in Y(x)} \partial\varphi(y,\cdot)(x)$$

in which $Y(x) := \{y \in Y | \varphi(y,x) = \Phi(x)\}$.

Notice that if, for every $y \in Y$, $\varphi(y,\cdot)$ is differentiable and attains its maximum at a unique point $y(x) \in Y$, then the inclusion reads

$$\partial\Phi(x) \supset \nabla\varphi\big(y(x),\cdot\big)(x).$$

Finally, if we can ensure differentiability of Φ, we clearly obtain the equality

$$\nabla\Phi(x) = \nabla\varphi\big(y(x),\cdot\big)(x).$$

More results on the subdifferential of a supremum of convex functions may be obtained is more specific setting. The interested reader may consult [6] and the references therein.

Now, getting back to the context of our study, it is rather straightforward to check that the function

$$\varphi(f,\lambda) := -L(f,\lambda)$$

satisfies all the previous requirements provided that J and the mapping $\lambda \mapsto f_\lambda$ are Fréchet differentiable. From now on, we make these extra assumptions, from which we easily infer that the dual function D is differentiable on $(0,\infty)$, and that its derivative is given by

$$D'(\lambda) = \|\mathscr{A} f_\lambda - g\|^2 - \varepsilon.$$

We are ready to state our main result:

Theorem 4. Suppose that the noise estimation τ and the data g satisfy:

$$\operatorname{dist}(g, \overline{\mathscr{A}F}) < \tau < \|g\|. \tag{5}$$

Then Problem $(\mathscr{D})$ has at least one solution $\bar{\lambda} > 0$, and the unique solution $f_{\bar{\lambda}}$ of Problem $(\mathscr{L}_{\bar{\lambda}})$ satisfies $\|\mathscr{A}f_{\bar{\lambda}} - g\|^2 = \tau^2$, and is consequently a solution of Problem $(\mathscr{Q})$ too. Moreover, any other solution $\bar{\lambda}' > 0$ of Problem $(\mathscr{D})$ leads to the same $f_{\bar{\lambda}'} = f_{\bar{\lambda}}$.

Proof. Clearly, $D(0) = 0$ and $D'(0) = \|g\|^2 - \varepsilon > 0$, in which $D'(0)$ denotes the right derivative of D at 0. In addition, we have $\operatorname{dist}(g, \overline{\mathscr{A}F}) < \tau$, so that there exists some $f_0 \in F$ such that $\|\mathscr{A}f_0 - g\|^2 - \varepsilon < 0$. Then,

$$D(\lambda) \leq L(f_0, \lambda) = J(f_0) + \lambda(\|\mathscr{A}f_0 - g\|^2 - \varepsilon) \longrightarrow -\infty \quad \text{as} \quad \lambda \to \infty,$$

and this is sufficient to prove that Problem $(\mathscr{D})$ has a solution $\bar{\lambda}$. One has

$$D'(\bar{\lambda}) = 0 = \|\mathscr{A}f_{\bar{\lambda}} - g\|^2 - \varepsilon,$$

so that $f_{\bar{\lambda}}$ is actually solution of Problem $(\mathscr{Q})$. Now, let $\bar{\lambda}$ and $\bar{\lambda}'$ be two solutions of Problem $(\mathscr{D})$ and let $\bar{f} := f_{\bar{\lambda}} \neq \bar{f}' := f_{\bar{\lambda}'}$. Note that, since $L(\bar{f}, \bar{\lambda}) = L(\bar{f}', \bar{\lambda}')$, we have $J(\bar{f}) = J(\bar{f}')$. Using the optimality condition (4) at $\bar{f}$ and $\bar{f}'$ consecutively, one gets:

$$J(\bar{f}') \geq J(\bar{f}) - \langle 2\bar{\lambda}\mathscr{A}^*(\mathscr{A}\bar{f} - g), \bar{f}' - \bar{f}\rangle$$

and

$$J(\bar{f}) \geq J(\bar{f}') - \langle 2\bar{\lambda}'\mathscr{A}^*(\mathscr{A}\bar{f}' - g), \bar{f} - \bar{f}'\rangle.$$

This yields:

$$0 \leq \langle 2\mathscr{A}^*(\mathscr{A}\bar{f} - g), \bar{f}' - \bar{f}\rangle \quad \text{and} \quad 0 \leq \langle 2\mathscr{A}^*(\mathscr{A}\bar{f}' - g), \bar{f} - \bar{f}'\rangle.$$

Adding the last two inequalities, we get

$$\left\|\mathscr{A}(\bar{f} - \bar{f}')\right\|^2 \leq 0,$$

so that $\bar{f} - \bar{f}' \in \ker \mathscr{A}$. Since J satisfies Assumption 2, the element

$$\bar{f}'' := \frac{\bar{f} + \bar{f}'}{2}$$

satisfies $J(\bar{f}'') < J(\bar{f})$ and $\mathscr{A}\bar{f}'' = \mathscr{A}\bar{f}$. Finally, we get $L(\bar{f}'', \bar{\lambda}) < L(\bar{f}, \bar{\lambda})$, which contradicts the optimality of $\bar{f}$. In conclusion, $\bar{f} = \bar{f}'$. ∎

Remark 5. Condition (5) is rather natural. On the one hand, feasibility in Problem $(\mathscr{Q})$ requires $\operatorname{dist}(g, \overline{\mathscr{A}F}) \leq \tau$, where the equality case yields the unstable least square problem. On the other hand, one expects the noise to be small compared to the data, and this is always the case in practice. ∎

Let us recall that for the Tikhonov regularization, $J(f) = \|f\|^2$ and the first order optimality condition in Problem $(\mathscr{P}_\alpha)$ yields $f_\lambda = \left(\mathscr{A}^*\mathscr{A} + \lambda^{-1}\mathscr{I}\right)^{-1}\mathscr{A}^*g$, where $\mathscr{I}$ denotes the identity operator on F. Hence the extra assumptions of differentiability of the mappings J and $\lambda \mapsto f_\lambda$ are obviously fulfilled at every positive λ, as a consequence of the boundedness of the operator $\left(\mathscr{A}^*\mathscr{A} + \lambda^{-1}\mathscr{I}\right)^{-1}$. Then, Theorem 4 allows to define the following simple procedure for computing the Tikhonov parameter corresponding to the Morozov condition:

1. Compute $\bar{\lambda} = \operatorname{argmax}_{\lambda>0} D(\lambda)$;
2. Set $\alpha = 1/\bar{\lambda}$.

In Step 1, the evaluation of D and D' can be performed by solving a Lagrange problem. Obviously, such problems are well behaved for small values of λ, and their condition deteriorates as λ increases. We emphasize that the problem of finding a maximizer $\bar{\lambda}$ of D is a one dimensional problem and then is quite easy to solve. Indeed, any ascent method or Quasi-Newton method with steps satisfying the first and second Wolfe conditions (see [3, 11, 15]) will work well enough for finding $\bar{\lambda}$.

In figure 1 below, we sketch the behavior of D in several situations:

1. The solid line corresponds to the assumption (5) in the above theorem, which ensures that D has a maximum on $(0, \infty)$.
2. The dotted line corresponds to the case where the data is dominated by the noise. In this case, one can easily see that $D'(\lambda) < 0$ for all $\lambda > 0$, so that there is no positive maximizer of D.
3. In the case of the dashed line, The dual function does not attain a maximum because the estimation of the noise is too optimistic. In this case, $D'(\lambda) > 0$ for every $\lambda > 0$. Such a behavior would occur for example if one takes $\varepsilon = 0$, in which case the constraint in Problem $(\mathscr{Q})$ is equivalent to the equality constraint (1).

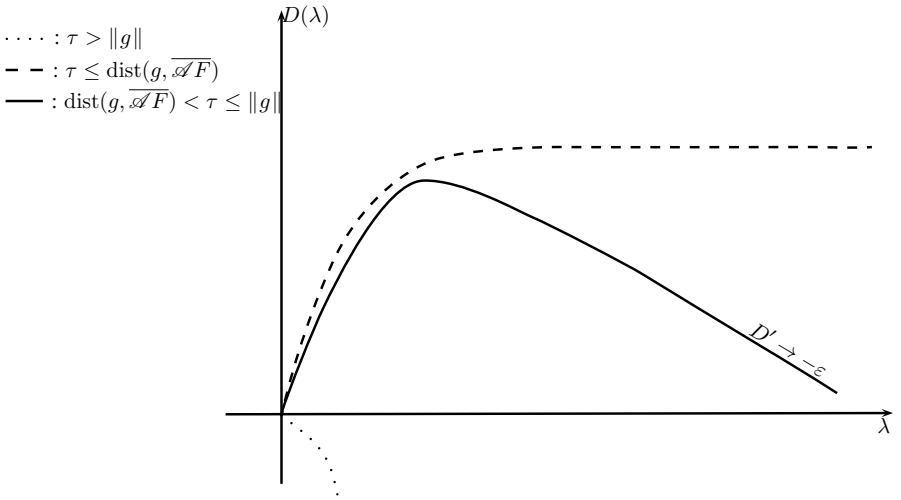

Fig. 1: Aspect of the dual function D in several situations

4 Numerical illustration

We consider here an ill-posed inverse problem of wide importance in statistics: the problem of nonparametric instrumental regression, as described in [5].

Let $Z\colon \Omega \to \mathbb{R}^p$ and $Y, \varepsilon\colon \Omega \to \mathbb{R}$ be random variables satisfying the relationship

$$Y = f(Z) + \varepsilon.$$

Here, Z is the *explanatory variable*, Y is the *response* and ε is an *error*, while $f\colon \mathbb{R}^p \to \mathbb{R}$ is an unknown function referred to as a *regression function*. The problem of nonparametric instrumental regression is that of identifying f, under the assumption that Z and ε are linked: $E(\varepsilon|Z) \neq 0$.

In order to cope with the dependence of ε and Z, it is customary to introduce an *instrumental variable* W, that is linked to Z and not to ε. The model can then be written as

$$Y = f(Z) + \varepsilon, \quad E(\varepsilon|W) = 0,$$

which implies that

$$E(Y|W) = E(f(Z)|W). \tag{6}$$

If all variables are absolutely continuous with respect to the Lebesgue measure, the problem can be written in terms of densities. Let P_Z and P_W be the laws of Z and W respectively, let V_Z and V_W be the supports of the variables Z and W. We define the spaces

$$L^2(V_Z, P_Z) = \left\{ \phi \in \mathbb{R}^{\mathbb{R}} \;\middle|\; \int_{V_Z} \phi(z)^2 \,\mathrm{d}P_Z(z) < \infty \right\}$$

and

$$L^2(V_W, P_W) = \left\{ \psi \in \mathbb{R}^{\mathbb{R}} \;\middle|\; \int_{V_W} \psi(w)^2 \,\mathrm{d}P_W(w) < \infty \right\}$$

which are Hilbert spaces endowed with their natural inner products. Next, we define the operator

$$\begin{array}{rccl} \mathscr{A}\colon & L^2(V_Z, P_Z) & \longrightarrow & L^2(V_W, P_W) \\ & \phi & \longmapsto & \mathbb{E}(\phi(Z)|W = \cdot) := \displaystyle\int_{V_Z} \phi(z) \frac{f_{ZW}(z, \cdot)}{f_Z(z) f_W(\cdot)} \,\mathrm{d}P_Z(z) \end{array}$$

where f_{ZW} is the joint probability density function of (Z, W) and f_W, f_Z are the probability density functions of W and Z respectively. We recall that, under the assumption that the kernel

$$K(z, w) = \frac{f_{ZW}(z, w)}{f_Z(z) f_W(w)}$$

belongs to $L^2(V_Z \times V_W, P_Z \otimes P_W)$, $\mathscr{A}$ is a Hilbert-Schmidt, thus compact. Whence the ill-posedness of the problem of identifying the regression function f.

For approximating the unknown function $f \in L^2(V_Z, P_Z)$, we consider its projection onto a finite dimensional subspace S_I^Z of $L^2(V_Z, P_Z)$, with $S_I^Z = \mathrm{span}\{\phi_1, ..., \phi_I\}$ where the family $\{\phi_i\}_{i=1,...,I}$ is orthonormal:

$$\forall i, j \in \{1, ..., I\}, \quad \langle \phi_i, \phi_j \rangle_Z = \mathbb{E}(\phi_i(Z)\phi_j(Z)) = \int_{V_Z} \phi_i(z)\phi_j(z) \,\mathrm{d}P_Z(z) = \delta_{ij}.$$

We also consider the projection of g onto some finite dimensional subspace S_J^W of $L^2(V_W, P_W)$, with $S_J^W = \mathrm{span}\{\psi_1, ..., \psi_J\}$ where the family $\{\psi_j\}_{j=1,...,J}$ is orthonormal:

$$\forall i, j \in \{1, ..., J\}, \quad \langle \phi_i, \phi_j \rangle_W = \mathbb{E}(\psi_i(W)\psi_j(W)) = \int_{V_W} \psi_i(w)\psi_j(w) \,\mathrm{d}P_W(w) = \delta_{ij}.$$

Now, restricting f to lie in S_I^Z and projecting the model equation $\mathscr{A}f = g$ onto S_I^W yields the linear system $Ax = y$, in which $A = (A_{ji}) \in \mathbb{R}^{J\times I}$ is given by

$$A_{ji} = \langle \mathscr{A}\phi_i, \psi_j \rangle_W = \mathbb{E}[\mathscr{A}\phi_i(W)\psi_j(W)], \quad i = 1, ..., I,\ j = 1, ..., J, \tag{7}$$

and $y = (y_j) \in \mathbb{R}^J$ is given by

$$y_j = \mathbb{E}[g(W)\psi_j(W)], \quad j = 1, ..., J. \tag{8}$$

The unknown is now $x = (x_1, \ldots, x_I)^\top \in \mathbb{R}^J$, the vector of the components of f in the basis $\{\phi_i\}$. The expectations occurring in the latter formulæ can be approximated by using empirical means through the statistical sample $\{Y_k, W_k, Z_k\}_{k=1,...,N}$ of the variables $\{Y, W, Z\}$ as follows:

$$A_{ji} \approx \hat{A}_{ji} = \frac{1}{N}\sum_{k=1}^{N} \phi_i(Z_k)\psi_j(W_k)$$

and

$$y_j \approx \hat{y}_j = \frac{1}{N}\sum_{k=1}^{N} Y_k\psi_j(W_k).$$

In our simulation, the functions $\{\phi_i\}_{i=1,...,I}$ and $\{\psi_j\}_{j=1,...,J}$ are defined to be gate functions on V_Z and V_W, and $I = J = 700$. The variable Z, the noise ϵ and the instrument W are linked by

$$Z = 0.45 + 0.1W + 0.9\epsilon$$

where W and ϵ are drawn from the uniform distribution on $[0, 1]$ and $[-0.5, 0.5]$, respectively. The variable Y satisfies $Y = f(Z) + \epsilon$, where

$$f(x) = 1_{[0,1]}(x)\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

with $\mu = 0.5$ and $\sigma = 0.1$.

In order to maximize the dual function D, we used the algorithms 1 and 2 given in the appendix. We used as stopping criterion $|D'(\lambda)| < 10^{-11}$. We recall that, for maximizing the function D, we could also use the steepest ascent method, but we choose a Quasi-Newton algorithm. The main reason is that Quasi-Newton algorithms have better convergence speed. Notice also that, in dimension one, its implementation is particularly simple. In the figures 2, 3 and 4, we use the notation $\mathscr{A}, f$ and g to mean in fact A, x and y, respectively.

Figure 2 clearly illustrates the results presented in the previous section. Figure 3 shows the evolution of $D'(\lambda_k)$ (left picture) and the residual $\|\mathscr{A}f_{\lambda_k} - g\|$ (right picture) as a function of k and the table of Figure 4 shows their values for a few iterations, in the favorable case where Condition (5) is satisfied (with $\tau = 0.9$). The above stopping criterion is reached after only 16 iterations. The corresponding value of α is found to be equal to 0.0050249, and we have that f_λ indeed satisfies the Morozov condition with an error of magnitude 10^{-14}.

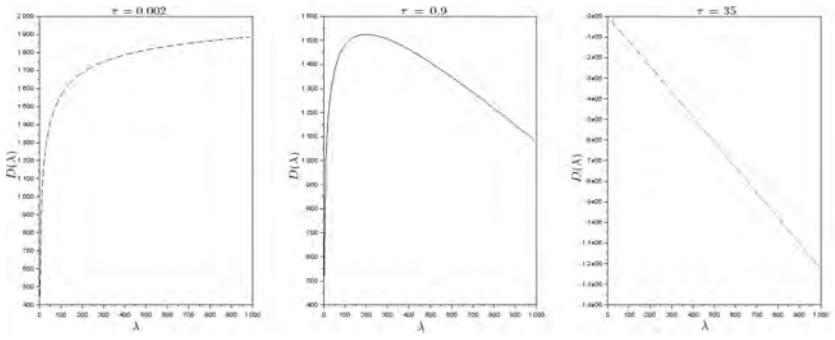

Fig. 2: Evolution of $D(\lambda)$ for various values of τ. In this study, $\|g\| = 30.0044$. In the left picture, $\tau = 0.002$, which corresponds to a too optimistic estimation of the noise level; in the middle picture, $\tau = 0.9$ for which Condition (5) is satisfied; in the right picture, $\tau = 35$ which corresponds to the case where the noise dominates the data.

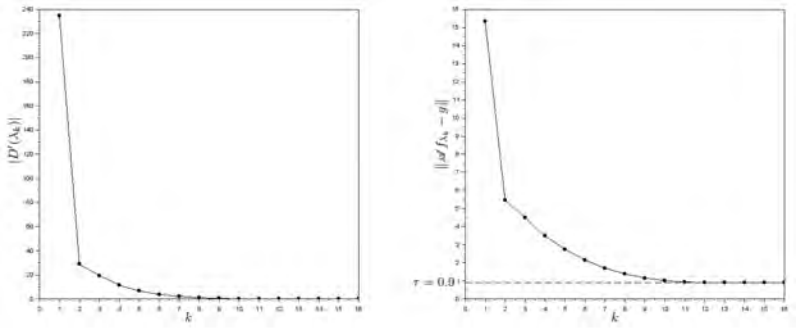

Fig. 3: Evolution of $|D'(\lambda_k)|$ and the residual $\|\mathscr{A} f_{\lambda_k} - g\|$ in function of iterations k for $\tau = 0.9$ which satisfies Condition (5). The threshold $\tau = 0.9$ of the residual to reach is represented by the horizontal line.

k	$D'(\lambda_k)$	$\|\|\mathscr{A} f_{\lambda_k} - g\|\|$
1	234.602	15.343
2	28.877	5.448
3	19.546	4.511
4	11.328	3.484
8	1.126	1.391
12	$1.271e-02$	0.907
14	$1.156e-05$	$0.9+0.642e-05$
15	$1.066e-08$	$0.9+0.59e-08$
16	$1.301e-13$	$0.9+0.72e-13$

Fig. 4: Values of $D'(\lambda_k)$ and the residual $\|\mathscr{A} f_{\lambda_k} - g\|$ in function of the iterations k, for $\tau = 0.9$ which satisfies Condition (5).

5 Conclusion

In this note, we have explored some aspects of the dualization of the constraint arising from the implementation of the Morozov principle. The shape of the dual function is directly related to the magnitude of the noise and the quality of its estimation. In the favorable cases, the desired Tikhonov parameter can be easily obtained via the maximization of the dual function, which is one-dimensional and smooth. We have illustrated our analysis by means of the problem of *nonparametric instrumental regression*, an important problem in statistics. The numerical results corroborate our analysis. The dualization of the Morozov constraint yields a simple, stable and numerically efficient algorithm for computing the corresponding Tikhonov parameter and, at the same time, the solution to the initial Morozov problem.

6 Appendix: algorithmic details

In order to maximize D, which we reformulate here as the minimization of $\bar{D} := -D$, we combine a Quasi-Newton method with Wolfe-Lemaréchal line-search (see the algorithms 1 and 2 below).

Recall that at each iteration k, we have an update of the form $\lambda_{k+1} = \lambda_k + \alpha_k d_k$, where d_k denotes the direction of descent computed via $\bar{D}'$ and an approximation of $\bar{D}''$. The stepsize α_k is chosen so as to satisfy the two Wolfe conditions $(C1)$ and $(C2)$ (see [11]) in order to guarantee the monotonicity of the sequence $\bar{D}(\lambda_k)$:

$$\begin{aligned} (C1): &\quad \bar{D}(\lambda_k + \alpha_k d_k) - \bar{D}(\lambda_k) - \alpha_k \beta_1 \bar{D}'(\lambda_k) d_k < 0 \\ (C2): &\quad \bar{D}'(\lambda_k + \alpha_k d_k) d_k - \beta_2 \bar{D}'(\lambda_k) d_k \geq 0. \end{aligned}$$

In $(C1)$ and $(C2)$, the parameters β_1 and β_2 are taken in $(0,1)$ (see [3, 15]). In Algorithm 1, the stopping criterion is $|\bar{D}'(\lambda_k)| < \epsilon$ with $\epsilon > 0$ defining the tolerance. For computing α_k at line 6, we propose Algorithm 2, which is based on the *line-search* algorithm (see [11]).

In Algorithm 2, (α_g, α_d) is the interval in which the step α_k will be chosen. Here, M is a large number which emulates ∞. For example, we may set $M = 10^{10}$. Recall that the failure of Condition $(C1)$ means that α_k is too large, while the failure of Condition $(C2)$ means that α_k is too small. If it happens that $|\alpha_d - \alpha_g| \approx 0$, indicating that β_1 is too big or β_2 is too small, one may then merely adjust these parameters. This situation occurs rarely. For instance, $\beta_1 = 0.25$ and $\beta_2 = 0.75$ worked perfectly well for all our simulations.

Algorithm 1.

1. Set $H_0 = 1$
2. Set λ_0 (initial guess)
3. set $k = 0$
4. **While** $|\bar{D}'(\lambda_k)| > \epsilon$ **do**:
5. $\quad d_k = -\bar{D}'(\lambda_k)/H_k$
6. $\quad$ Compute α_k satisfying $(C1), (C2)$.
7. $\quad \lambda_{k+1} = \lambda_k + \alpha_k d_k$
8. $\quad v_k = \bar{D}'(\lambda_{k+1}) - \bar{D}'(\lambda_k)$
9. $\quad H_{k+1} = v_k / (\lambda_{k+1} - \lambda_k)$
10. $\quad k = k + 1$
11. **End While**
12. set $\bar{\lambda} = \lambda_k$

Algorithm 2.

1. set β_1 and β_2
2. set $\alpha_g = 0$ and $\alpha_d = M \gg 1$
3. Set $\alpha_k = \alpha_0$ in (α_g, α_d)
4. **While** $(C1)$ or $(C2)$ is false **do**
5. $\quad$ **If** $(C1)$ is false **then**
6. $\quad\quad \alpha_d = \alpha_k$
7. $\quad$ **End If**
8. $\quad$ **If** $(C2)$ is false **then**
9. $\quad\quad \alpha_g = \alpha_k$
10. $\quad$ **End If**
11. $\quad \alpha_k = (\alpha_g + \alpha_d)/2$
12. $\quad$ **If** $|\alpha_d - \alpha_g| \approx 0$ **then**
13. $\quad\quad$ **Break**
14. $\quad$ **End If**
15. **End While**

Acknowledgements. We thank the anonymous referees for their helpful comments and suggestions which have enabled us to improve the manuscript.

References

[1] J. Borwein and A. Lewis, *Convex Analysis and Nonlinear Optimization*, CMS Books in Mathematics, Springer, 2nd edition, 2005.

[2] H.W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Springer, 1996.

[3] R. Fletcher, *Practical Methods of Optimization: Unconstrained Optimization*, J. Wiley and Sons, New York, 1980.

[4] K. Frick and M. Grasmair, *Regularization of linear ill-posed problems by the augmented lagrangian method and variational inequalities*, Inverse Problems, 28, 2012.

[5] P. Hall and J. Horowitz, *Nonparametric methods for inference in the presence of instrumental variables*, Annals of Statistics, 33(6), 2005.

[6] A. Hantoute, M.A. López and C. Zălinescu, *Subdifferential calculus rules in convex analysis: A unifying approach via pointwise supremum functions*, SIAM Journal on Optimization, 2008.

[7] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I. A Series of Comprehensive Studies in Mathematics*, Springer, 1993.

[8] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms II. A Series of Comprehensive Studies in Mathematics*, Springer-Verlag, 1993.

[9] I. Hnětynková and M. Plešinger and Z. Strakoš, *The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level in the data*, BIT Numerical Mathematics, 49, 2009.

[10] A. Kirsch, *An introduction to the mathematical theory of inverse problems*, Springer, 2011.

[11] C. Lemaréchal, *A view of line-searches*, in A. Auslender, W. Oettli and J. Stoer, Editors, *Optimization and Optimal Control*, Lecture Notes in Control and Information Sciences Vol. 30, Springer, 1981.

[12] V.A. Morozov, *Choice of parameter for the solution of functional equations by the regularization method*, Sov. Math. Doklady Vol. 8, 1967.

[13] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[14] A.N. Tikhonov and V.Y. Arsenin, *Solutions of Ill-Posed Problems*, Wiley, New York, 1977.

[15] P. Wolfe, *Convergence conditions for ascent methods*, SIAM Review, 11, 1969.

[16] C. Zălinescu, *Convex analysis in general vector spaces*, World Scientific, 2002.

Chapter 3

Nouvelle Méthode de Régularisation

3.1 Présentation

Ce chapitre introduit une nouvelle méthode de régularisation pour la résolution des problèmes linéaires mal-posés, et tout particulièrement des problèmes linéaires exponentiellement mal-posés. La méthode est étudiée sous des conditions sources de type logarithmique (qui correspondent aux problèmes exponentiellement mal-posés) et aussi sous les conditions sources générales définies à l'aide de fonctions index.

L'optimalité de la méthode ainsi que les notions de qualification et taux de convergence sont établis sous des conditions sources logarithmiques et une quasi-optimalité est obtenue sous des conditions sources générales. Une procédure particulière pour le calcul des solutions régularisées dans le cas des problèmes de conduction thermiques est décrite. Enfin, une comparaison du nouveau schéma de régularisation aux méthodes classiques (Tikhonov, *spectral cut-off*, régularisation asymptotique et gradients conjugués) est établie avec des simulations sur trois problèmes test venant de la littérature.

L'article suivant a été rédigé seul et est une de mes contributions principales dans cette thèse. Il a été soumis pour publication dans la special issue d'optimisation correspondant à la conférence franco-indienne d'optimisation et analyse variationnelle tenue en février 2020 à Varanasi.

3.2 Discussion sur les applications numériques

Dans l'article qui suit, une importante partie est consacrée aux applications numériques où la nouvelle méthode de régularisation est comparée empiriquement à d'autres méthodes de régularisation classiques pour la résolution de trois problèmes tests.

Tout d'abord, rappelons que chacun des trois problèmes tests considérés dans l'article peut se formuler comme une équation intégrale de première espèce. C'est à dire que chaque problème test traité peut s'écrire sous de la forme

$$\int_a^b K(s,t)f(t)\,\mathrm{d}t = g(s), \quad s \in [c,d] \tag{3.1}$$

où $K(s,t)$ est une fonction connue défini sur $[c,d] \times [a,b]$, f est la fonction inconnue à déterminer et g est la donnée. La discrétisation de l'équation (3.1) est effectuée soit par collocation (en utilisant les méthodes de quadrature pour approximer l'intégrale) soit par la méthode de Galerkin (en projetant les fonctions f et g dans des bases orthogonales de dimensions finis).

En utilisant la méthode de collocation avec n points de collocation sur l'intervalle $[a, b]$ et m points sur l'intervalle $[c, d]$, on a

$$\int_a^b K(s,t)f(t)\,\mathrm{d}t \approx \sum_{j=1}^{n} w_j K(s,t_j)f(t_j), \quad s \in [c,d].$$

Ainsi l'équation (3.1) est discrétisée en le système linéaire $Ax = b$ où la matrice A et le vecteur b sont définis par

$$A_{ij} = w_j K(s_i,t_j)f(t_j), \quad b_i = g(s_i) \quad i = 1,...,m,\ j = 1,...,n$$

où $t_j \in [a, b]$ et $s_i \in [c, d]$ sont des points de collocation, w_j sont les poids de la formule de quadrature utilisée et $x_j = f(t_j)$, $j = 1, ..., n$ sont les inconnus de l'équation.

Pour la méthode de Galerkin, on choisit une base orthonormée de dimension finie $\{\phi_j\}_{j=1,...,n}$ définie sur l'intervalle $[a, b]$ (resp. $\{\psi_i\}_{i=1,...,m}$ définie sur l'intervalle $[c, d]$) dans laquelle on projette f (resp. g). Ainsi, f et g sont approximés par

$$f \approx \sum_{j=1}^{n} f_j\phi_j \quad \text{et} \quad g \approx \sum_{i=1}^{m} g_i\psi_i \quad \text{avec} \quad g_i = \int_c^d g(s)\psi_i(s)\,\mathrm{d}s,$$

et l'équation (3.1) est discrétisée en le système linéaire $Ax = b$ où la matrice A et le vecteur b sont définis par

$$A_{ij} = \int_a^b \int_c^d K(s,t)\phi_j(t)\psi_i(s)\,\mathrm{d}s\,\mathrm{d}t, \quad b_i = \int_c^d g(s)\psi_i(s)\,\mathrm{d}s, \quad i = 1,...,m,\ j = 1,...,n$$

et $x_j \approx \int_a^b f(t)\phi_j(t)\,\mathrm{d}t$, $j = 1, ..., n$ sont les coefficients indéterminée de l'approximation de f dans la base $\{\phi_j\}_{j=1,...,n}$.

Notons que les problèmes tests `shaw` et `heat` sont discrétisées suivant la méthode de collocation et le problème test `baart` est discrétisée selon la méthode de Galerkin avec projection dans les bases de fonctions portes, voir [54, Section 4] pour plus de détails.

Une fois chaque problème test discrétisé en un système linéaire $Ax = b$, un important challenge est le calcul et l'évaluation de la matrice $(A^*A)^{\sqrt{\alpha}}$. Dans les problèmes tests considérés, hors mis le problème de chaleur `heat`, la seule alternative possible pour le calcul de $(A^*A)^{\sqrt{\alpha}}$ est la décomposition en valeurs singulières qui est un processus délicat vu que la matrice A présente un conditionnement assez grand hérité de la mal-position du problème test. Néanmoins, dans les trois problèmes tests considérés qui sont tous unidimensionnel, nous avons pu contourné ce problème d'une part en choisissant des paramètres de discrétisations n et m juste assez grand pour avoir des reconstructions lisses et raisonnables et d'autres part en utilisant l'état de l'art des routines de décomposition en valeurs singulières tel que `dgesvd()`.

Il convient de préciser que dans le cas des problèmes inverses à grandes échelles, la nouvelle méthode s'accommode (pour le moment) uniquement aux problèmes de conductions thermiques, auxquels cas, le calcul de $(T^*T)^{\sqrt{\alpha}}$ peut se faire suivant l'alternative exposée en Section 4.1 et illustré dans le cas de l'équation rétrograde de la chaleur.

Une fois passée l'étape de la décomposition en valeurs singulières de matrice A, le calcul de la solution régularisée (pour les méthodes spectrales) découle des formules données en Section 4. Par ailleurs, le calcul de la solution régularisée des gradients conjugués se calcule à partir de l'algorithme 1.1.

Après le calcul des solutions régularisées, l'étape suivante est le calcul du paramètre de régularisation. Le calcul de paramètre de régularisation pour chaque méthode heuristique utilisée consiste à calculer les

minimums locaux d'une fonction Ψ sur un intervalle bornée. Ainsi le paramètre de régularisation α_{mh} (resp. n_{mh}) suivant une méthode heuristique `mh` est

$$\alpha_{mh} = \underset{\alpha\in(\alpha_0,\alpha_1)}{\text{argmin loc}}\ \Psi(\alpha) \quad \left(\text{resp.} \quad n_{mh} = \underset{1\leq n\leq N}{\text{argmin loc}}\ \Psi(n)\right),$$

dépendant du cas où la méthode de régularisation est continue ou itérative. Pour plus de détails sur la fonction Ψ dépendant de la méthode heuristique de choix de paramètre, voir Section 6. Ainsi une première étape est le choix de l'intervalle dans lequel rechercher le paramètre de régularisation.

Pour les méthodes de régularisation continues (la nouvelle méthode, Tikhonov et Showalter), on recherche le paramètre de régularisation dans l'intervalle $(\alpha_0, ||A||^2)$ où α_0 est choisi de sorte que le terme de régularisation soit de norme du même ordre que l'erreur d'arrondi machine généralement appelé *epsilon machine* $\epsilon_M \approx 10^{-16}$ dans notre cas. Ainsi pour la nouvelle méthode, α_0 est choisi de sorte que

$$\left\|(I-(A^*A)^{\sqrt{\alpha_0}})^*(I-(A^*A)^{\sqrt{\alpha_0}})\right\| \sim \epsilon_M.$$

Pour la méthode de Tikhonov, étant donné que $\left\|\alpha I\right\| \sim \alpha$, alors α_0 est choisi tel que $\alpha_0 \sim \epsilon_M$. Pour la méthode de Showalter qui n'est pas une méthode variationnelle, α_0 est choisi de sorte que

$$\exp(-||A||^2/\alpha_0) \sim \epsilon_M.$$

La motivation d'un tel choix est que avec la méthode de Showalter, (1.47) entraine que l'erreur de régularisation est bornée comme suit

$$||f^\dagger - f_\alpha|| = ||r_\alpha(T^*T)f^\dagger|| \leq ||f^\dagger|| \sup_{\lambda\in(0,||T||^2]} |r_\alpha(\lambda)| \quad \text{avec} \quad r_\alpha(\lambda) = \exp(-\lambda/\alpha).$$

Pour les méthodes de régularisation itératives (*truncated singular value decomposition* et gradients conjugués), la plage où choisir le paramètre de régularisation est plus facile. Ainsi pour la *truncated singular value decomposition*, le paramètre de régularisation est choisi entre 1 et n ou n est le nombre de colonne de la matrice A. Pour la méthode des gradients conjugués, étant donnée la convergence assez rapide de cette méthode, le paramètre de régularisation optimale est généralement atteint pour des itérations relativement faibles. Ainsi, pour la méthode des gradients conjugués, dans nos simulations, le paramètre de régularisation est choisi entre 1 et 100.

Une fois que l'intervalle où rechercher le paramètre est fixé, il peut arriver qu'on ait plusieurs minimum locaux de la fonction Ψ, auquel cas il faut définir des critères pour la sélection parmi les minimums locaux observés. Nous utilisons trois critères de sélection à savoir: le choix du minimiseur local le plus grand (qui induit une plus grande stabilité), le choix du minimiseur local le plus petit (qui induit une plus grande fidélité au modèle) ou alors le minimiseur local qui réalise le plus petit minimum parmi les minimums locaux. Afin de trancher lequel de ces critères utiliser pour un problème précis, on passe par deux phases: une première phase d'étude où la morphologie et la régularité des solutions régularisées correspondant à chaque minimum local est comparé aux a-priori de la solution inconnue, et une deuxième phase où le critère dont le paramètre est plus fidèle aux a-priori sur la solution inconnue est sélectionné.

A new regularization method for linear exponentially ill-posed problems

Walter Cedric Simo Tao Lee*

Abstract

This paper provides a new regularization method which is particularly suitable for linear exponentially ill-posed problems. Under logarithmic source conditions (which have a natural interpretation in terms of Sobolev spaces in the aforementioned context), concepts of qualifications as well as order optimal rates of convergence are presented. Optimality results under general source conditions expressed in terms of index functions are also studied.

Finally, numerical experiments on three test problems attest the better performance of the new method compared to the well known Tikhonov method in instances of exponentially ill-posed problems.

Keywords: Ill-posed problems, Regularization, logarithmic source conditions, qualifications, order-optimal rates.

1 Introduction

In this paper, we are interested in the solution to the equation

$$Tx = y \tag{1}$$

where $T : X \to Y$ is a linear bounded operator between two infinite dimensional Hilbert spaces X and Y with non-closed range. The data y belongs to the range of T and we assume that we only have approximated data y^δ satisfying

$$||y^\delta - y|| \leq \delta. \tag{2}$$

In such a setting, Equation (1) is ill-posed in the sense that the Moore Penrose generalized inverse $T^\dagger$ of T which maps y to the best-approximate solution $x^\dagger$ of (1) is not continuous. Consequently a little perturbation on the data y may induce an arbitrarily large error in the solution $x^\dagger$. Instances of such ill-posed inverse problems are encountered in several fields in applied sciences among which: signal and image processing, computer tomography, immunology, satellite gradiometry, heat conduction problems, inverse scattering problems, statistics and econometrics to name just a few (see, e.g. [11, 14, 20, 21, 31]). As a result of the ill-posedness of Equation (1), a regularization method needs to be applied in order to recover from the noisy data y^δ a stable approximation x^δ of the solution $x^\dagger$. A regularization method can be regarded as a family of continuous operators $R_\alpha : Y \to X$ such that there exists a function $\Lambda : \mathbb{R}_+ \times Y \to \mathbb{R}_+$ satisfying the following: for every $y \in \mathcal{D}(T^\dagger) \subset Y$ and $y^\delta \in Y$ satisfying (2)

$$R_{\Lambda(\delta, y^\delta)} y^\delta \to x^\dagger \quad \text{as} \quad \delta \downarrow 0. \tag{3}$$

*Institut de Mathématiques de Toulouse, Université Paul Sabatier, Toulouse, France
Email: wsimotao@math.univ-toulouse.fr

Some examples of regularizations methods are Tikhonov, Landweber, spectral cut-off, asymptotic regularization, approximate inverse and mollification (see, e.g. [1, 7, 11, 21, 23, 24]). As a matter of fact, we would like to get estimates on the error committed while approximating $x^\dagger$ by $x^\delta = R_{\Lambda(\delta,y^\delta)}y^\delta$.

It is well known that for arbitrary $x^\dagger \in X$, the convergence of x^δ towards $x^\dagger$ is arbitrarily slow (see, e.g. [11, 35]). But still, by allowing smoothness of the solution $x^\dagger$, convergence rates could be established. Standard smoothness conditions known as Hölder type source condition take the form

$$x^\dagger \in X_\mu(\rho) = \{(T^*T)^\mu w, \quad w \in X \quad \text{s.t.} \quad ||w|| \leq \rho\}, \tag{4}$$

where μ and ρ are two positive constants. However such source conditions have shown their limitations as they are too restrictive in many problems and do not yield a natural interpretation. For this reason, general source conditions have been introduced in the following form:

$$x^\dagger \in X_\varphi(\rho) = \{\varphi(T^*T)w, \quad w \in X \quad \text{s.t.} \quad ||w|| \leq \rho\}, \tag{5}$$

where ρ is a positive constant and $\varphi : [0, ||T^*T||] \to \mathbb{R}_+$ is an index function, i.e. a non-negative monotonically increasing continuous function satisfying $\varphi(\lambda) \to 0$ as $\lambda \downarrow 0$. An interesting discussion on these source conditions can be found in [29] where the author explores how general source conditions of the form (5) are. Once the solution $x^\dagger$ satisfies a smoothness condition i.e. $x^\dagger$ belongs to a proper subspace M of X, it is possible to derive convergence rates and the next challenge is about optimality. More precisely, for a regularization method $R : Y \to X$, we are interested in the worst case error:

$$\Delta(\delta, R, M) := \sup\left\{||Ry^\delta - x^\dagger||, \ x^\dagger \in M, \ y^\delta \in Y, \ \text{s.t.} \ ||y^\delta - Tx^\dagger|| \leq \delta\right\}, \tag{6}$$

and we would like a regularization which minimizes this worst case error. In this respect, a regularization method $\bar{R} : Y \to X$ is said to be optimal if it achieves the minimum worst case error over all regularization methods, i.e. if

$$\Delta(\delta, \bar{R}, M) = \Delta(\delta, M) := \inf_R \Delta(\delta, R, M).$$

Similarly, a regularization is said to be order optimal if it achieves the minimum worst case error up to a constant greater than one, i.e. if

$$\Delta(\delta, \bar{R}, M) \leq C\Delta(\delta, M).$$

for some constant $C > 1$. When the subset M is convex and balanced, it is shown in [30] that

$$\omega(\delta, M) \leq \Delta(\delta, M) \leq 2\omega(\delta, M), \tag{7}$$

where $\omega(\delta, M)$ is the modulus of continuity of the operator T over M i.e.

$$\omega(\delta, M) = \sup\{||x||, \ x \in M, \ \text{s.t.} \ ||Tx|| \leq \delta\}. \tag{8}$$

In other words, we get the following:

$$\Delta(\delta, X_\varphi(\rho)) = \mathcal{O}\left(\omega(\delta, X_\varphi(\rho))\right). \tag{9}$$

Recall that, under mild assumptions on the index function φ, the supremum defining the modulus of continuity is achieved and a simple expression of $\omega(\delta, X_\varphi(\rho))$ in term of function φ is available (see, e.g. [20, 28, 37]). Let us remind that a relevant notion in the study of optimality of a regularization method is qualification. In fact, the qualification of a regularization measures the capability of the method to take into account smoothness assumptions on the solution $x^\dagger$, i.e. the higher the qualification, the more the method is able to provide best rates for very smooth solutions.

Besides optimality, converse results and saturation results are also important aspects of regularization algorithms (see, [11, 27, 33, 34]). For converse results, we are interested in the following: given a particular convergence rate of $||x^\delta - x^\dagger||$ towards 0, which smoothness condition does the solution $x^\dagger$ needs to satisfy? Saturation results are about the maximal smoothness on the solution $x^\dagger$ for which a regularization method can still deliver the best rates of convergence. Finally, another significant aspect of regularization is the selection of the regularization parameter i.e. finding a function $\Lambda(\delta, y^\delta)$ which guarantees convergence and possibly order-optimality.

Coming back to (5), notice that a very interesting subclass of general source conditions are logarithmic source conditions expressed as:

$$x^\dagger \in X_{f_p}(\rho) = \left\{(-\ln(T^*T))^{-p}w, \quad w \in X \quad \text{s.t.} \quad ||w|| \leq \rho\right\}, \tag{10}$$

where p and ρ are positive constants and T satisfies $\|T^*T\| < 1$. Such smoothness conditions have clear interpretations in term of Sobolev spaces in exponentially ill-posed problems (see, e.g. [20, 37]). The latter class includes several problems of great importance such as backward heat equation, sideways heat equation, inverse problem in satellite gradiometry, control problem in heat equation, inverse scattering problems and many others (see, [20]). Because of the importance of exponentially ill-posed problems, it is desirable to design regularization methods particularly suitable for this class of problems. It is precisely the aim of this paper to provide such a regularization scheme.

In the next section, we define the new regularization method using both the variational formulation and the definition in terms of the so called *generator* function g_α. A brief comparison with the Tikhonov method is done. Moreover basic estimates on the *generator* function g_α and its corresponding *residual* function r_α are also carried out.

Section 3 is devoted to optimality of the new method. Here we recall well known optimality results under general source conditions of the form (5) (see, [19, 20, 28, 32, 37]). For the specific case of logarithmic source conditions, qualification of the method is given and order optimality is shown. Next we study optimality under general source conditions.

In Section 4, we present a comparative analysis of the new method with Tikhonov method, spectral cut-off, asymptotic regularization and conjugate gradient.

Section 5 is about numerical illustrations. In this section, in order to confirm our prediction of better performance of the new method compared to Tikhonov and spectral cut-off in instance of exponentially ill-posed problems, we numerically compare the efficiency of the five regularization methods on three test problems coming from literature: A problem of image reconstruction taken from [36], a Fredholm integral equation of the first kind found in [2] and an inverse heat equation problem.

Finally in Section 6, for a fully applicability of the new method, we exhibit heuristic selection rules which fit with the new regularization technique. Moreover, we also compare the five regularization methods for each heuristic parameter choice rule under consideration.

2 The new regularization method

For the sake of simplicity, we assume henceforth that the operator T is injective. Hereafter, we set a positive number a such that the operator norm of T^*T is less than a i.e. $||T^*T|| \leq a$. In the sequel, we assume that $a < 1$ which is always possible by scaling Equation (1).

Let us consider the general variational formulation of a regularization method

$$x_\alpha = \underset{x \in X}{\operatorname{argmin}}\, \mathscr{F}(Tx, y) + \mathscr{P}(x, \alpha) \tag{11}$$

where $\mathscr{F}(Tx, y)$ is the fit term, $\mathscr{P}(x, \alpha)$ is the penalty term and $\alpha > 0$ is the regularization parameter. We recall that the fit term aims at fitting the model, the penalty term aims at introducing stability in the initial model $Tx = y$ and the regularization parameter α controls the level of regularization.

In most cases, the fit term $\mathscr{F}(Tx, y)$ is nothing but

$$\mathscr{F}(Tx, y) = ||Tx - y||^2 \tag{12}$$

and the penalty term depends on the regularization method. For instance, for Tikhonov regularization, $\mathscr{P}(x, \alpha)$ is given by

$$\mathscr{P}(x, \alpha) = \alpha||x||^2. \tag{13}$$

This penalization can sometimes compromises the quality of the resulting approximate solution x_α. Indeed, let $X = L^2(\mathbb{R}^n)$, then by Parseval identity, we see that

$$\mathscr{P}(x, \alpha) = \alpha||\hat{x}||^2_{L^2(\mathbb{R}^n)} \tag{14}$$

where $\hat{x}$ is the Fourier transform of x. Equation (14) implies that the stability is introduced by uniformly penalizing all frequency components irrespective of the magnitude of frequencies. Yet, it is well known that the instability of the initial problem comes from high frequency components on the contrary to low frequency components.

Let us introduce the following penalty term where the regularization parameter α is no more defined as a weight but as an exponent:

$$\mathscr{P}(x, \alpha) = \left\| \left[I - (T^*T)^{\sqrt{\alpha}} \right] x \right\|^2. \tag{15}$$

In (15), $(T^*T)^{\sqrt{\alpha}}$ is defined via the spectral family $(E_\lambda)_\lambda$ associated to the self-adjoint operator T^*T i.e.

$$(T^*T)^{\sqrt{\alpha}} x = \int_{\lambda=0}^{||T^*T||_+} \lambda^{\sqrt{\alpha}} \, \mathrm{d}E_\lambda x.$$

We keep the fit term defined in (12) and then the variational formulation of our new regularization method is given by

$$x_\alpha = \underset{x \in X}{\operatorname{argmin}} \, ||Tx - y||^2 + \left\| \left[I - (T^*T)^{\sqrt{\alpha}} \right] x \right\|^2. \tag{16}$$

From the first order optimality condition, we get that x_α is the solution to the linear equation :

$$\left[T^*T + \left(I - (T^*T)^{\sqrt{\alpha}} \right)^2 \right] x = T^*y,$$

that is,

$$x_\alpha = \left[T^*T + \left(I - (T^*T)^{\sqrt{\alpha}} \right)^2 \right]^{-1} T^*y. \tag{17}$$

From (17), we see that the new method can also be defined via the so called *generator* function g_α, i.e.

$$x_\alpha = g_\alpha(T^*T)T^*y, \tag{18}$$

with the function g_α defined by

$$g_\alpha(\lambda) = \frac{1}{\lambda + (1 - \lambda^{\sqrt{\alpha}})^2}, \quad \lambda \in (0, ||T^*T||]. \tag{19}$$

Let us also define the *residual* function r_α corresponding to g_α as follows

$$r_\alpha(\lambda) := 1 - \lambda g_\alpha(\lambda) = \frac{(1 - \lambda^{\sqrt{\alpha}})^2}{\lambda + (1 - \lambda^{\sqrt{\alpha}})^2}, \quad \lambda \in (0, ||T^*T||]. \tag{20}$$

The functions g_α and r_α defined in (19) and (20) are important since they will be repeatedly used in the convergence analysis of the regularization method. In fact, the regularization error $x^\dagger - x_\alpha$ and the propagated error $x_\alpha - x_\alpha^\delta$ are expressed via the functions r_α and g_α as follows:

$$x^\dagger - x_\alpha = r_\alpha(T^*T)x^\dagger, \quad x_\alpha - x_\alpha^\delta = g_\alpha(T^*T)T^*(y - y^\delta).$$

Finally, notice that the function g_α defined in (19) indeed satisfies the basic requirements for defining a regularization method i.e.

a) g_α is continuous,

b) $\forall\, \alpha > 0, \quad \sup_{\lambda\in(0,||T^*T||]} \lambda g_\alpha(\lambda) \leq 1 < \infty$,

c) $\lim_{\alpha\downarrow 0}\, g_\alpha(\lambda) = 1/\lambda$.

From b) and c), we deduce the convergence of the new regularization method by application of [11, Theorem 4.1]. Before going to optimality results, let us state some basic estimates (proven in the appendix) about the functions g_α and r_α.

Proposition 1. *Let the function g_α be defined by (19). Then for all $a < 1$ and $\alpha < 1$,*

$$\sup_{\lambda\in(0,a]} \sqrt{\lambda} g_\alpha(\lambda) = \mathcal{O}\left(\frac{1}{\sqrt{\alpha}}\right). \tag{21}$$

Lemma 1. *For all α and λ satisfying $0 < \alpha \leq \lambda < 1$, the following estimates hold for the function r_α defined in (20):*

$$r_\alpha(\lambda) \leq \frac{9}{4}\left(\frac{\alpha|\ln(\lambda)|^2}{\lambda + \alpha|\ln(\lambda)|^2}\right). \tag{22}$$

3 Optimality results

Before studying the optimality of the method presented in Section 2, we need first to recall general optimality results under source condition of the form (5). For doing so, let us specify assumptions on the function φ which defines the source set $X_\varphi(\rho)$.

Assumption 1. *The function $\varphi : (0, a] \to \mathbb{R}_+$ is continuous, monotonically increasing and satisfies:*

(i) $\lim_{\lambda\downarrow 0} \varphi(\lambda) = 0$,

(ii) the function $\phi : (0, \varphi^2(a)] \to (0, a\varphi^2(a)]$ defined by

$$\phi(\lambda) = \lambda(\varphi^2)^{-1}(\lambda) \tag{23}$$

is convex.

Under Asummption 1 on the function φ, the following result from [37] holds and we can then defines optimality under source condition (5).

Theorem 1. *Let $X_\varphi(\rho)$ be as in (5) and let Assumption 1 be fulfilled. Let the function ϕ be defined by (23). Then*

$$\omega(\delta, X_\varphi(\rho)) \leq \rho\sqrt{\phi^{-1}\left(\frac{\delta^2}{\rho^2}\right)}. \tag{24}$$

Moreover, if $\delta^2/\rho^2 \in \sigma\left(T^*T\varphi^2(T^*T)\right)$, then equality holds in (24).

A similar result to this theorem can be found in [20, Section 2], and [28, Section 3].

Remark 1. *In [28], the results corresponding to Theorem 1 are given in term of the function* $\Theta : (0, a] \to (0, a\varphi(a)]$ *defined by:*

$$\Theta(\lambda) = \sqrt{\lambda}\varphi(\lambda). \tag{25}$$

Then, by simple computations, we can find that

$$\rho\sqrt{\phi^{-1}\left(\frac{\delta^2}{\rho^2}\right)} = \rho\,\varphi\left(\Theta^{-1}(\delta/\rho)\right). \tag{26}$$

In such a case, the convexity of the function ϕ *defined in (23) is equivalent to the convexity of the function* $\chi(\lambda) = \Theta^2\left((\varphi^2)^{-1}(\lambda)\right)$ *and the condition* $\delta^2/\rho^2 \in \sigma\left(T^*T\varphi^2(T^*T)\right)$ *which allows to get the equality in (24) is equivalent to* $\delta/\rho \in \sigma\left(\Theta(T^*T)\right)$.

From Theorem 1 and Remark 1, we can deduce that under the source condition (5) and Assumption 1, the best possible worst case error is $\rho\,\varphi\left(\Theta^{-1}(\delta/\rho)\right)$ whence the following definition.

Definition 1 (Optimality under general source conditions)**.** *Let Assumption 1 be satisfied and consider the source condition* $x^\dagger \in X_\varphi(\rho)$. *A regularization method* $R(\delta) : Y \to X$ *is said to be:*

- *optimal if* $\Delta(\delta, R(\delta), X_\varphi(\rho)) \leq \rho\,\varphi\left(\Theta^{-1}(\delta/\rho)\right)$;
- *order optimal if* $\Delta(\delta, R(\delta), X_\varphi(\rho)) \leq C\rho\,\varphi\left(\Theta^{-1}(\delta/\rho)\right)$ *for some constant* $C \geq 1$;
- *quasi-order optimal if for all* $\epsilon > 0$, $\Delta(\delta, R(\delta), X_\varphi(\rho)) = \mathcal{O}\left(f_\epsilon(\delta)\right)$ *where the function* $f_\epsilon : \mathbb{R}_+ \to \mathbb{R}_+$ *converges to* $\varphi\left(\Theta^{-1}(\delta/\rho)\right)$ *as* ϵ *decreases to* 0 *i.e. for all* $\delta > 0$, $f_\epsilon(\delta) \to \varphi\left(\Theta^{-1}(\delta/\rho)\right)$ *as* ϵ *decreases to* 0.

Having defined the optimality under general source conditions, let us now consider the particular case of logarithmic source conditions. For logarithmic source conditions, the function φ equals the function $f_p : (0, a] \to \mathbb{R}_+$ defined by:

$$f_p(\lambda) = (-\ln(\lambda))^{-p}. \tag{27}$$

Next it is easy to see that the only point to check in Assumption 1 is the convexity of the function ϕ defined in (23). Precisely, for the index function f_p, this function is $\phi_p : (0, \ln(1/a)^{-2p}] \to (0, a\ln(1/a)^{-2p}]$ defined by

$$\phi_p(\lambda) = \lambda\exp(-\lambda^{-1/2p})$$

which was proven to be convex on the interval $[0, 1]$ in [26]. In order to fulfill Assumption 1 and avoid the singularity of the function f_p at $\lambda = 1$, we assume that $a \leq \exp(-1) < 1$, i.e. $||T^*T|| \leq \exp(-1)$. Notice that this is not actually a restriction, since Equation (1) can always be rescaled in order to meet this criterion.

Due to (24) it suffices to compute $\sqrt{\phi_p^{-1}\left(\delta^2/\rho^2\right)}$ in order to define the optimality in logarithmic source conditions. Thanks again to [26], we have that

$$\sqrt{\phi_p^{-1}(s)} = f_p(s)(1 + o(1)) \text{ as } s \to 0. \tag{28}$$

Hence, we deduce the following definition of optimality in case of logarithmic source condition.

Definition 2 (Optimality under logarithmic source condition)**.** *Consider logarithmic source condition (10), on defining* f_p *as in (27), a regularization method* $R(\delta) : Y \to X$ *is said to be:*

- *optimal if* $\Delta(\delta, R(\delta), X_{f_p}(\rho)) \leq \rho f_p(\delta^2/\rho^2)(1 + o(1))$ *as* $\delta \to 0$,

- *order optimal if* $\Delta(\delta, R(\delta), X_{f_p}(\rho)) \leq C\rho f_p(\delta^2/\rho^2)(1+o(1))$ *as* $\delta \to 0$.

In the sequel, we are interested to optimality with respect to the noise level δ. In this respect, we can characterize the order-optimality under logarithmic source conditions as follows.

Remark 2. *By definition of the function* f_p*, we get that* $\mathcal{O}\left(f_p(\delta^2/\rho^2)\right) = \mathcal{O}\left(f_p(\delta)\right)$ *as* $\delta \to 0$*. Hence, equivalently to Definition 2, a regularization method* $R(\delta) : Y \to X$ *is said to be order optimal under logarithmic source condition if*

$$\Delta(\delta, R(\delta), X_{f_p}(\rho)) = \mathcal{O}\left(f_p(\delta)\right) \quad as \quad \delta \to 0.$$

3.1 Optimality under logarithmic source conditions

Having given all the necessary definitions, let us now study the optimality of the method proposed in Section 2.

Proposition 2. *The regularization* g_α *defined by (19) has qualification* f_p*. That is:*

$$\sup_{0<\lambda\leq a} |r_\alpha(\lambda)|\, f_p(\lambda) = \mathcal{O}\left(f_p(\alpha)\right). \tag{29}$$

The proof the Proposition 2 heavily relies on the following lemma which is proven in the appendix.

Lemma 2. *Let* p *and* α *be two positive numbers with* $\alpha \leq \bar{\alpha} < 1$*, let* $a \in (0,1)$ *and* $\Psi_{p,\alpha} : (0,a] \to \mathbb{R}_+$ *be the function defined by*

$$\Psi_{p,\alpha}(\lambda) = \frac{|\ln(\lambda)|^{2-p}}{\lambda + \alpha|\ln(\lambda)|^2}. \tag{30}$$

Then, the following hold:

(i) The function $\Psi_{p,\alpha}$ *is well defined and differentiable on* $(0,a]$*, and its derivative is given by*

$$\Psi'_{p,\alpha}(\lambda) = \frac{\lambda^{-1}|\ln(\lambda)|^{1-p}}{(\lambda + \alpha|\ln(\lambda)|^2)^2} h(\lambda), \tag{31}$$

where

$$h(\lambda) = \alpha p|\ln(\lambda)|^2 - \lambda\left(2 - p + |\ln(\lambda)|\right). \tag{32}$$

(ii) If $p \leq 2$*, there exists at least one* $\lambda(\alpha, p)$ *where* h *vanishes. Moreover for every such* $\lambda(\alpha,p)$*, the following holds*

$$\lambda(\alpha, p) \simeq \alpha|\ln(\alpha)|, \tag{33}$$

that is, there exists two constants c_1 *and* c_2 *depending on* p *only such that*

$$c_1\alpha|\ln(\alpha)| \leq \lambda(\alpha,p) \leq c_2\alpha|\ln(\alpha)|.$$

Moreover, this result still holds if $p > 2$*,* $\lambda < c \leq \exp(2-p)$ *and* α *is small.*

(iii) The supremum of the function $\Psi_{p,\alpha}$ *on* $(0,a]$ *satisfies*

$$\sup_{0<\lambda\leq a} \Psi_{p,\alpha}(\lambda) = \mathcal{O}\left(\alpha^{-1}|\ln(\alpha)|^{-p}\right). \tag{34}$$

Having stated the above lemma, the proof of Proposition 2 easily follows:

PROOF. If $\lambda \leq \alpha$ then the monotonicity of the function f_p and the fact that the residual function r_α is bounded by 1 on $(0, a]$ yields (29). If $\lambda \geq \alpha$ then from Lemma 2, we deduce that

$$\sup_{0<\lambda\leq a} \frac{\alpha|\ln(\lambda)|^2}{\lambda + \alpha|\ln(\lambda)|^2} f_p(\lambda) = \mathcal{O}\left(f_p(\alpha)\right)$$

which together with Lemma 1 yields (29). □

From Proposition 2, we deduce the following optimality result.

Theorem 2. *Let $p > 0$, $x^\dagger \in X_{f_p}(\rho)$, and $y^\delta \in Y$ satisfying (2) with $y = Tx^\dagger$. Assume that $||T^*T|| \leq \exp(-1)$ and let $x(\delta) = g_{\alpha(\delta)}(T^*T)T^*y^\delta$ with the function g_α being defined by (19) and let $\alpha(\delta) = \Theta_p^{-1}(\delta)$ with Θ_p defined by*

$$\Theta_p(\lambda) = \sqrt{\lambda}(\ln(1/\lambda))^{-p}. \tag{35}$$

Then the order optimal estimate

$$||x^\dagger - x(\delta)|| = \mathcal{O}\left(f_p(\delta)\right) \quad as \quad \delta \to 0 \tag{36}$$

holds. Thus the regularization g_α defined by (19) is order optimal under logarithmic source conditions.

PROOF. As usual, we start with the following splitting

$$||x^\dagger - x_\alpha^\delta|| \leq ||x^\dagger - x_\alpha|| + ||x_\alpha - x_\alpha^\delta||. \tag{37}$$

Using that $x^\dagger - x_\alpha = r_\alpha(T^*T)x^\dagger$, $x_\alpha - x_\alpha^\delta = g_\alpha(T^*T)T^*(y - y^\delta)$ together with the source condition $x^\dagger \in X_{f_p(\rho)}$, we deduce that:

$$||x^\dagger - x_\alpha|| \leq C_1 \sup_{\lambda\in(0,a]} r_\alpha(\lambda)\, f_p(\lambda) \tag{38}$$

and

$$||x_\alpha - x_\alpha^\delta|| \leq \delta\ C_2 \sup_{\lambda\in(0,a]} \sqrt{\lambda} g_\alpha(\lambda). \tag{39}$$

By applying the propositions 1 and 2 to (38), (39) and using (37), we get that

$$||x^\dagger - x_\alpha^\delta|| \leq C_1' f_p(\alpha) + C_2' \frac{\delta}{\sqrt{\alpha}}, \tag{40}$$

where C_1' and C_2' are constants independent of α and λ. Hence, by taking $\alpha := \Theta_p^{-1}(\delta)$, the estimate in (36) follows from

$$||x^\dagger - x(\delta)|| = \mathcal{O}\left(f_p(\Theta_p^{-1}(\delta))\right) = \mathcal{O}\left(f_p(\delta^2)\right) = \mathcal{O}\left(f_p(\delta)\right).$$

□

Corollary 1. *Let $p > 0$, $x^\dagger \in X_{f_p}(\rho)$, and $y^\delta \in Y$ satisfying (2) with $y = Tx^\dagger$. Assume that $||T^*T|| \leq \exp(-1)$ and let $x(\delta) = g_{\alpha(\delta)}(T^*T)T^*y^\delta$ with the function g_α being defined by (19) and $\alpha(\delta) = \delta$. Then the order optimal estimate*

$$||x^\dagger - x(\delta)|| = \mathcal{O}\left(f_p(\delta)\right) \quad as \quad \delta \to 0$$

holds. Thus the regularization g_α defined by (19) is order optimal under logarithmic source conditions with an a-priori parameter choice rule independent of the smoothness of the solution $x^\dagger$.

PROOF. By considering $\alpha(\delta) = \delta$ in (40), we get

$$||x^\dagger - x_\alpha^\delta|| \leq C_1' f_p(\delta) + C_2' \sqrt{\delta} = \mathcal{O}\left(f_p(\delta)\right) \quad \text{as} \quad \delta \to 0,$$

since $\sqrt{\delta} = \mathcal{O}\left(f_p(\delta)\right)$ as $\delta \to 0$. □

The next proposition describes a Morozov-like discrepancy rule which leads to order-optimal convergence rates under logarithmic source conditions.

Proposition 3. *Let $p > 0$, $x^\dagger \in X_{f_p}(\rho)$, and $y^\delta \in Y$ satisfying (2) with $y = Tx^\dagger$. Assume that $||T^*T|| \leq \exp(-1)$ and consider the a-posteriori parameter choice rule*

$$\alpha(\delta, y^\delta) = \sup\left\{\alpha > 0, \quad ||Tx_\alpha^\delta - y^\delta|| \leq \delta + \sqrt{\delta}\right\}. \tag{41}$$

Let $x(\delta) = g_{\alpha(\delta,y^\delta)}(T^*T)T^*y^\delta$ with the function g_α defined by (19), then the order optimal estimate

$$||x^\dagger - x(\delta)|| = \mathcal{O}\left(f_p(\delta)\right) \quad \text{as} \quad \delta \to 0 \tag{42}$$

holds. Thus the regularization g_α defined by (19) is order optimal under logarithmic source conditions with the a-posteriori parameter choice rule defined by (41).

The proof of Proposition 3 is deferred to Appendix.

A converse result

Theorem 2 establishes that the logarithmic source condition (10) is sufficient to imply the rate $f_p(\delta)$ in (36). Now we are going to prove that the logarithmic source condition (10) is not only sufficient but also almost necessary. The following result based on [20, Theorem 8] establishes a converse result in the noise free case for the new regularization method.

Theorem 3. *Let $x_\alpha = g_\alpha(T^*T)Ty$ with $y = Tx^\dagger$ and let the function g_α be defined in (19). Then the estimate*

$$||x^\dagger - x_\alpha|| = \mathcal{O}\left(f_p(\alpha)\right) \tag{43}$$

implies that $x^\dagger \in X_{f_q}(\rho)$ for some $\rho > 0$ for all $0 < q < p$.

The proof consists in checking that the function g_α defined in (19) satisfies all the conditions stated in Theorem 8 of [20]. More precisely, we just need to check that there exists a constant $C_g > 0$ such that

$$\sup_{\lambda \in (0, ||T^*T||]} g_\alpha(\lambda) \leq \frac{C_g}{\alpha}.$$

But, from (62), we see that the latter condition is obviously fulfilled.

3.2 Optimality under general source conditions

Let us state the following quasi-optimal result under general source conditions.

Theorem 4. *Let $p > 0$, $x^\dagger \in X_\varphi(\rho)$, where φ is a concave index function satisfying Assumption 1 and $y^\delta \in Y$ satisfying $||y - y^\delta|| \leq \delta$ with $y = Tx^\dagger$ and $\delta \leq \Theta(a)$. Assume that $||T^*T|| \leq a \leq \exp(-1)$ and let $x(\delta) = g_{\alpha(\delta)}(T^*T)T^*y^\delta$ with the function g_α defined in (19). For small positive ϵ, let $\alpha(\delta) = \Theta_\epsilon^{-1}(\delta)$ where the function Θ_ϵ is defined by $\Theta_\epsilon(\lambda) = \lambda^{-\epsilon}\Theta(\lambda)$ with Θ given in (25).*

Then the estimate

$$||x^\dagger - x(\delta)|| = \mathcal{O}\left((\Theta_\epsilon^{-1}(\delta))^{-\epsilon}\varphi(\Theta_\epsilon^{-1}(\delta))\right) \quad \text{as} \quad \delta \to 0$$

holds. Moreover, as $\epsilon \downarrow 0$, $(\Theta_\epsilon^{-1}(\delta))^{-\epsilon}\varphi(\Theta_\epsilon^{-1}(\delta)) \to \varphi(\Theta^{-1}(\delta))$. Thus the regularization method defined via the function g_α given in (19) is quasi-order optimal under general source conditions.

PROOF. We study two cases: $\alpha \geq \lambda$ and $\alpha < \lambda$. In the first case, $\sup_{(0,\exp(-1)]} r_\alpha(\lambda)\varphi(\lambda) \leq \varphi(\alpha)$ by monotonicity of the function φ and the order-optimality follows trivially. Let us study the main case when $\alpha < \lambda$. From Lemma 1, we get, for $\lambda \in (0, a]$,

$$\begin{aligned}
r_\alpha(\lambda)\varphi(\lambda) &\leq \frac{9}{4}|\ln(\lambda)|^2 \frac{\alpha}{\lambda + \alpha|\ln(\lambda)|^2}\varphi(\lambda) \\
&\leq \frac{9}{4}|\ln(\alpha)|^2 \frac{\alpha}{\lambda + \alpha|\ln(a)|^2}\varphi(\lambda) \\
&\leq \frac{9}{4}\alpha^{-\epsilon}(\alpha^{\epsilon/2}|\ln(\alpha)|)^2 \frac{\alpha}{\lambda + \alpha|\ln(a)|^2}\varphi(\lambda) \\
&\leq \frac{9}{4}\frac{4}{\epsilon^2}\alpha^{-\epsilon} \frac{\alpha\lambda}{\lambda + \alpha|\ln(a)|^2}\frac{\varphi(\lambda)}{\lambda} \\
&\leq \frac{9}{4}\frac{4}{\epsilon^2}\alpha^{-\epsilon} \frac{\alpha\lambda}{\lambda + \alpha|\ln(a)|^2}\frac{\varphi(\alpha)}{\alpha} \quad \text{by concavity of } \varphi \\
&\leq C_\epsilon \alpha^{-\epsilon}\varphi(\alpha). \qquad (44)
\end{aligned}$$

Hence $\sup_{(0,a]} r_\alpha(\lambda)\varphi(\lambda) \leq C_\epsilon\alpha^{-\epsilon}\varphi(\alpha)$. From (38) and (39), and (21) we get

$$||x^\dagger - x_\alpha^\delta|| \leq C_\epsilon\alpha^{-\epsilon}\varphi(\alpha) + \frac{\delta}{\sqrt{\alpha}}.$$

By taking $\alpha(\delta) = \Theta_\epsilon^{-1}(\delta)$ with $\Theta_\epsilon(\lambda) = \lambda^{1/2-\epsilon}\varphi(\lambda)$, we get

$$||x^\dagger - x(\delta)|| = \mathcal{O}\left((\Theta_\epsilon^{-1}(\delta))^{-\epsilon}\varphi(\Theta_\epsilon^{-1}(\delta))\right).$$

Now, it remains to show that $(\Theta_\epsilon^{-1}(\delta))^{-\epsilon}\varphi(\Theta_\epsilon^{-1}(\delta))$ converges to the optimal rate $\varphi(\Theta^{-1}(\delta))$ as ϵ goes to 0. Let $\alpha_* = \Theta^{-1}(\delta)$ and $\alpha_\epsilon = \Theta_\epsilon^{-1}(\delta)$, let us show that α_ϵ converges to α_* as ϵ goes to 0. By the monotonicity of Θ_ϵ for $\epsilon \in (0, 1/2)$ and the fact that $\delta \leq \Theta(a)$ and $a < 1$, we get that, for all $\epsilon \in (0, 1/2)$,

$$\frac{\delta}{\Theta(a)} \leq 1 < a^{-\epsilon} \quad \Rightarrow \quad \delta \leq a^{-\epsilon}\Theta(a) = \Theta_\epsilon(a) \quad \Rightarrow \quad \alpha_\epsilon = \Theta_\epsilon^{-1}(\delta) \leq a.$$

Hence $\alpha_\epsilon \in (0, a]$ and the sequence $(\alpha_\epsilon)_\epsilon$ is bounded and thus it admits a converging subsequence. Let $(\alpha_{\epsilon_n})_n$ a converging subsequence of $(\alpha_\epsilon)_\epsilon$, and let $\tilde{\alpha}$ be its limit. Let us show that $\tilde{\alpha} = \alpha_*$.

Since $\alpha_{\epsilon_n} \to \tilde{\alpha}$ and Θ is continuous, $\Theta(\alpha_{\epsilon_n}) \to \Theta(\tilde{\alpha})$. But $\Theta(\alpha_{\epsilon_n}) = \alpha_{\epsilon_n}^{\epsilon_n}\Theta(\alpha*)$ since $\delta = \Theta(\alpha_*)$ and $\delta = \Theta_\epsilon(\alpha_\epsilon)$ for all small positive ϵ. So we get

$$\alpha_{\epsilon_n}^{\epsilon_n}\Theta(\alpha*) \to \Theta(\tilde{\alpha}) \quad \text{i.e.} \quad \alpha_{\epsilon_n}^{\epsilon_n} \to \frac{\Theta(\tilde{\alpha})}{\Theta(\alpha*)}. \qquad (45)$$

By the convergence of the sequence $(\alpha_{\epsilon_n})_n$, we get that $\alpha_{\epsilon_n}^{\epsilon_n} = \exp(\epsilon_n \ln(\alpha_{\epsilon_n}))$ converges to 1, (45) proves that $\Theta(\tilde{\alpha}) = \Theta(\alpha*)$ and by bijectivity of the function Θ, we deduce that $\tilde{\alpha} = \alpha_*$. Since the sequence $(\epsilon_n)_n$ was arbitrarily chosen, we deduce that the whole sequence $(\alpha_\epsilon)_\epsilon$ converges to α_* as $\epsilon \downarrow 0$. Thus we deduce that $\alpha_\epsilon^{-\epsilon} \to 1$ and $\varphi(\alpha_\epsilon) \to \varphi(\alpha_*)$ which implies that

$$(\Theta_\epsilon^{-1}(\delta))^{-\epsilon}\varphi(\Theta_\epsilon^{-1}(\delta)) \to \varphi(\Theta^{-1}(\delta)).$$

□

For Holder type source conditions, Theorem 4 reduces to the following theorem.

Theorem 5. *Consider the setting of Theorem 4 with the function $\varphi(t) = t^{\mu}$ i.e. $x^{\dagger} \in \mathrm{Ran}\,(T^*T)^{\mu}$, then there exists an a priori selection rule $\alpha(\delta)$ such that the following holds:*

$$\left\| x^{\dagger} - x^{\delta}_{\alpha(\delta)} \right\| = \begin{cases} \mathcal{O}\left(\delta^{\frac{2\sigma}{2\sigma+1}}\right) & \forall\, \sigma < \mu, \ \text{if } \mu \leq 1 \\ \mathcal{O}\left(\delta^{\frac{2}{3}}\right) & , \ \text{if } \mu > 1. \end{cases} \tag{46}$$

Remark 3. *By defining a variant of the new regularization method where the approximate solution x^{δ}_{α} is defined as the solution of the optimization problem*

$$x^{\delta}_{\alpha} = \underset{x \in X}{\mathrm{argmin}}\, ||(T^*T)^{\sqrt{\alpha}} y^{\delta} - Tx||^2 + || \left[I - (T^*T)^{\sqrt{\alpha}} \right] x||^2,$$

we can prove order optimal rate under Holder type source condition but with a lower qualification index $\mu_0 = 1/2$. This variant is motivated by the mollification regularization method, where a target object defined as a smooth version of $x^{\dagger}$ is fixed prior to the regularization (see e.g. [1, 7]). In this respect, the target object here is given as $(T^*T)^{\sqrt{\alpha}} x^{\dagger}$. This choice is legitimated by the smoothness property of the operator T and the fact that as α goes to 0, this target object converges to the solution $x^{\dagger}$. The study of this variant and the corresponding optimality results is beyond the scope of this paper.

4 A framework for comparison

In the sequel, we are going to compare the new method with three continuous regularization methods: Tikhonov [38], spectral cut-off [11], Showalter [11] and one iterative regularization method: conjugate gradient [11, 21]. We recall that the first three methods (Tikhonov, spectral cut-off and Showalter) are linear methods on the contrary to conjugate gradient which is an iterative non-linear regularization method. Obviously the new method, Tikhonov, spectral cut-off and Showalter are members of the family of general regularization methods defined via a *generator* function. Roughly speaking, each regularization method is defined via a so-called *generator* function $g^{reg}_{\alpha}(\lambda)$ which converges pointwise to $1/\lambda$ as α goes to 0 and the regularized solution $x^{\delta}_{\alpha,reg}$ is defined by :

$$x^{\delta}_{\alpha,reg} = g^{reg}_{\alpha}(T^*T)T^*y^{\delta}. \tag{47}$$

In this respect, the functions $g^{reg}_{\alpha}(\lambda)$ associated to Tikhonov, spectral cut-off, Showalter and the new method are defined as follows:

$$g^{tik}_{\alpha}(\lambda) = \frac{1}{\lambda + \alpha}, \quad g^{sc}_{\alpha}(\lambda) = \frac{1}{\lambda} 1_{\{\lambda \geq \alpha\}}, \quad g^{sw}_{\alpha}(\lambda) = \frac{1 - e^{-\lambda/\alpha}}{\lambda}, \quad g^{nrm}_{\alpha}(\lambda) = \frac{1}{\lambda + (1 - \lambda\sqrt{\alpha})^2} \tag{48}$$

where $\lambda \in (0, a]$ with $||T^*T|| \leq a < 1$.

Before getting into comparison of the new method to other regularization techniques, let us first point out a way of computing the regularized solution $x^{\delta}_{\alpha,nrm}$ of the new method.

4.1 Computation of the regularized solution $x^{\delta}_{\alpha,nrm}$

One way of computing the regularized solution $x^{\delta}_{\alpha,nrm}$ of the new method is by computing the singular value decomposition of operator T. That is to find a system (u_k, σ_k, v_k) such that:

- the sequence $(u_k)_k$ forms a Hilbert basis of X,
- the sequence $(v_k)_k$ forms a Hilbert basis of the closure of the range of T,
- the sequence $(\sigma)_k$ is positive, decreasing and satisfies $Tu_k = \sigma_k v_k$ and $T^*v_k = \sigma_k u_k$.

Given that decomposition of T, it is trivial to see that the operator T^*T is diagonal in the Hilbert basis $(u_k)_k$. Therefore, given a function g defined on the interval $(0, \sigma_1^2)$, the operator $g(T^*T)$ is nothing but the diagonal operator defined on the Hilbert basis $(u_k)_k$ by $g(T^*T)u_k = g(\sigma_k^2)u_k$. Hence given the singular value decomposition (u_k, σ_k, v_k) of T, from (47) (with $reg = nrm$), the regularized solution $x^\delta_{\alpha,nrm}$ can be computed as

$$x^\delta_{\alpha,nrm} = \sum_k g_\alpha^{nrm}(\sigma_k^2)\,\langle T^*y^\delta, u_k\rangle\, u_k = \sum_k \frac{\sigma_k}{\sigma_k^2 + \left(1 - \sigma_k^{2\sqrt{\alpha}}\right)^2}\,\langle y^\delta, v_k\rangle\, u_k. \tag{49}$$

Remark 4. *The above singular value decomposition of operator T is only possible if T is a compact operator. However, it is important to notice that the new method does not apply only to compact operator. Indeed, the new method is based on the spectral family $(E_\lambda)_\lambda$ associated to the self adjoint operator T^*T, and spectral family exists even for non-compact operator as pointed out in [11, Proposition 2.14]. This allows for the definition of a function applied to a self-adjoint non compact operator. Of course, one might ask how we can compute the regularized solution $x^\delta_{\alpha,nrm}$ in such a case. By noticing that in practice, we always discretize Equation (1) into matrix formulation, we can compute the singular value decomposition of the matrix representing the discretization of operator T and then apply (49) to compute $x^\delta_{\alpha,nrm}$.*

It is important to notice that a crucial step in the computation of the regularized solution $x^\delta_{\alpha,nrm}$ is the singular value decomposition step which should be done rigorously especially for exponentially ill-posed problems. That is why we propose a state of the art algorithm as LAPACK's `dgesvd()` routine for SVD computation (see e.g. [12, Section 8.6] for description of method). For an easy application, it is to be noted that this routine is implemented in the function `svd()` in `Matlab`. In Section 5, we will see that even for a very ill-conditionned matrix, we can still compute the regularized solution $x^\delta_{\alpha,nrm}$ very efficiently using the function `svd()` in `Matlab`.

Above, we saw that the new approximate solution $x^\delta_{\alpha,nrm}$ is computable using the singular value decomposition of operator T which might be delicate to compute. However, in some cases, there is an alternative for computing $x^\delta_{\alpha,nrm}$ when the operator $\log(T^*T)$ is explicitly known. Indeed, if the operator $\log(T^*T)$ is explicitly known, then the solution $u : \mathbb{R}_+ \to X$ to the initial value problem:

$$\begin{cases} u'(t) - \log(T^*T)u(t) = 0, & t \in \mathbb{R}_+ \\ u(0) = x, \end{cases} \tag{50}$$

evaluated at $t = \sqrt{\alpha}$ is nothing but $(T^*T)^{\sqrt{\alpha}}x$, i.e. $(T^*T)^{\sqrt{\alpha}}x = u(\sqrt{\alpha})$. Hence, through the resolution of the ordinary differential equation (50), the penalty term $\left\|\left[I - (T^*T)^{\sqrt{\alpha}}\right]x\right\|^2$ can be computed and this allows to compute the approximate solution $x^\delta_{\alpha,nrm}$.

An example of exponentially ill-posed problems for which the operator $\log(T^*T)$ is known is the backward heat equation. More precisely, let Ω be a smooth subset of $\mathbb{R}^n$ with $n \leq 3$ and $u : \Omega \times (0, \bar{t}] \to \mathbb{R}$ be the solution to the initial boundary value problem

$$\begin{cases} \frac{\partial u}{\partial t} = \Delta u, & \Omega \times (0, \bar{t}) \\ u(\cdot, 0) = f, & \Omega \\ u = 0 \quad \text{or} \quad \frac{\partial u}{\partial \nu} = 0, & \text{on} \quad \partial\Omega \times (0, \bar{t}]. \end{cases} \tag{51}$$

Assume we want to recover the initial temperature $f \in L^2(\Omega)$ given the final temperature $u(\cdot, \bar{t})$. By interpreting the heat equation (51) as an ordinary differential equation for the function $U : [0, \bar{t}] \to \mathcal{D}(\Delta) \subset L^2(\Omega)$, $t \to U(t) = u(\cdot, t)$, with the initial value $U(0) = f$ where

$$\mathcal{D}(\Delta) = H^2(\Omega) \cap H_0^1(\Omega) \quad \text{or} \quad \mathcal{D}(\Delta) = \left\{ f \in H^2(\Omega), \quad \frac{\partial f}{\partial \nu} = 0 \text{ on } \partial\Omega \right\},$$

we get that $U(t) = \exp(t\Delta)f$ for $t \in (0, \bar{t}]$, where $(\exp(t\Delta))_{t>0}$ is the strongly continuous semi-group generated by the unbounded self-adjoint linear operator Δ. This implies that the equation satisfied by the initial temperature f is nothing but

$$\exp(\bar{t}\Delta)f = u(\cdot, \bar{t}). \tag{52}$$

From (52), we deduce that $T^*T = \exp(2\bar{t}\Delta)$ and $\log(T^*T) = 2\bar{t}\Delta$ and thus operator $(T^*T)^{\sqrt{\alpha}}$ can be evaluated at a function $x \in L^2(\Omega)$ as the solution to the initial value problem

$$\begin{cases} u'(t) - 2\bar{t}\Delta u(t) = 0, & t \in \mathbb{R}_+ \\ u(0) = x, \\ u(t) \in \mathcal{D}(\Delta), & \text{for} \quad t \in \mathbb{R}_+, \end{cases} \tag{53}$$

evaluated at $t = \sqrt{\alpha}$.

In addition to the backward heat equation, there are other exponentially ill-posed problems for which $\log(T^*T)$ is known. This includes sideways heat equation (see [20, Section 8.3]) and more generally inverse heat conduction problems (see, e.g. [31, Section 3 & 4]).

4.2 Tikhonov versus new method

From the variational formulation of Tikhonov and the new method, we can see that both methods differ by the penalty term. For Tikhonov method, the penalty term is $\alpha \|x\|^2$ whereas for the new method, the penalty term is $\left\| \left[I - (T^*T)^{\sqrt{\alpha}} \right] x \right\|^2$. By considering $X = L^2(\mathbb{R}^n)$ for instance, by using the Parseval identity, we see that the penalty term is equal to $\alpha \|\hat{x}\|_{L^2(\mathbb{R}^n)}$. Therefore the weight α equally penalizes all frequency components irrespective of the magnitude of frequencies even though instability mainly comes from high frequency components. This is actually a drawback of the Tikhonov method which may induce an unfavorable trade-off between stability and fidelity to the model (see e.g. [1], Figure 4). On the contrary, for the new regularization method, high frequency components are much more regularized compared to low frequency components which are less and less regularized as the singular values increase to 1. In this way, we expect the new method to achieve a better trade-off between stability and fidelity to the model. Moreover, for exponentially ill-posed problems, the ill-posedness is accentuated due to the magnitude of singular values, the instability introduced by high frequency components are more pronounced and we expect the new regularization method to yield better approximations of $x^\dagger$.

4.3 Spectral cut-off versus new method

On the contrary to Tikhonov method, both spectral cut-off and the new method treat high frequency components and low frequency components differently. However, spectral cut-off regularized high frequency components by a mere cut-off and this may be too violent in several situations. Indeed even though high frequency components induce instability, they also carry some information which should not completely left out. For instance, for mildly ill-posed problems, this truncation will be very damaging on the quality of the approximation while for exponentially ill-posed problem, this truncation will be less damaging. A smooth transition (in term of regularization) from small singular values to other singular values would be more meaningful and desirable. This is actually what is done for the new method. Another advantage of the new method compared to spectral cut-off is the variational formulation of the new method which allows to add to the problem a-priori constraint on the solution (e.g. positivity, geometrical constraints, etc...).

4.4 Showalter versus new method

A major difference between Showalter method and the new method is that Showalter method does not have a variational formulation. Given that, for the Showalter method, it is not clear what is actually penalized in order to stabilize the problem. Moreover it would be difficult if not impossible to add a-priori constraints on the solution. Given a data y^δ, by inspecting the Showalter regularized solution which is given by $x_\alpha^\delta = \int_0^{1/\alpha} e^{-sT^*T}\,\mathrm{d}s\,T^*y^\delta$, we see that the method introduces stability by truncating the integral $\int_0^{+\infty} e^{-sT^*T}\,\mathrm{d}s\,T^*y^\delta = (T^*T)^{-1}T^*y^\delta$ on the interval $(0, 1/\alpha)$. On the other hand, we can see that, as the Tikhonov method, for all regularization parameter $\alpha > 0$, the *generator* function g_α^{sw} of Showalter method is strictly decreasing on the contrary to the generator function g_α^{nrm} of the new method which always exhibits a maximum close to $\lambda = 0$. This implies that the Showalter method cannot be seen as a smooth version of spectral cut-off which yields a smooth transition (in term of regularization) from high frequency components to low frequency components, on the contrary to the new method. Concerning the computation of the regularized solution $x_{\alpha,sw}^\delta$ for the Showalter method, it is important to notice that $x_{\alpha,sw}^\delta$ is the solution $u_\delta : \mathbb{R}_+ \to X$ of the initial value problem:

$$\begin{cases} u_\delta'(t) + T^*Tu_\delta(t) = T^*y^\delta, \quad t \in \mathbb{R}_+ \\ u_\delta(0) = 0, \end{cases} \tag{54}$$

evaluated at $t = 1/\alpha$, i.e. $x_{\alpha,sw}^\delta = u_\delta(1/\alpha)$. By solving (54) using the forward finite difference of step size h, we get that u_δ can be approximated as:

$$u_\delta(t+h) \approx u_\delta(t) + h\left[T^*y^\delta - T^*Tu_\delta(t)\right], \quad \text{with} \quad u_\delta(0) = 0. \tag{55}$$

4.5 Conjugate gradient versus new method

Unlike all the other regularization methods under consideration (Tikhonov, spectral cut-oof, Showalter and the new method), the conjugate gradient method is an iterative non-linear regularization method. The conjugate gradient method regularizes Problem (1) by iteratively approximating $x^\dagger$ by the minimizer x_k of the functional $f(x) = ||Tx - y||^2$ on finite dimensional Krylov subspaces

$$V_k = \text{span}\left\{T^*y, (T^*T)T^*y, ..., (T^*T)^{k-1}T^*y\right\},$$

where $k \geq 1$ and $k \in \mathbb{N}$. A major advantage of the conjugate gradient is the easy computation of regularized solution x_k (see e.g. algorithm given in [21, Figure 2.2]) and the fast convergence on the contrary to Landweber. However, as pointed out in [11, Theorem 7.6], the operator R_k which maps the data y to the regularized solution x_k is not always continuous contrarily to the new method. Moreover, compared to other regularization methods, there is no a-priori rules $k(\delta)$ such that $x_{k(\delta)}^\delta$ converges to $x^\dagger$ as $\delta \to 0$ (see, e.g. [9]).

A comparative plot of the *generator* functions g_α^{reg} associated to Tikhonov, spectral cut-off, Showalter and the new method is given in Figure 1.

Remark 5. *On the contrary to generator functions of Tikhonov and Showalter, the generator function g_α^{nrm} associated to the new regularization always exhibits a maximum close to $\lambda = 0$ and the function always equals 1 at $\lambda = 0$. Indeed, it is trivial to check that both functions g_α^{tik} and g_α^{sw} are strictly decreasing for all $\alpha > 0$. Hence, the function g_α^{nrm} is the only one which can be seen as a smooth version of the function g_α^{sc} associated to spectral cut-off which has a very crude transition at $\lambda = \alpha$.*

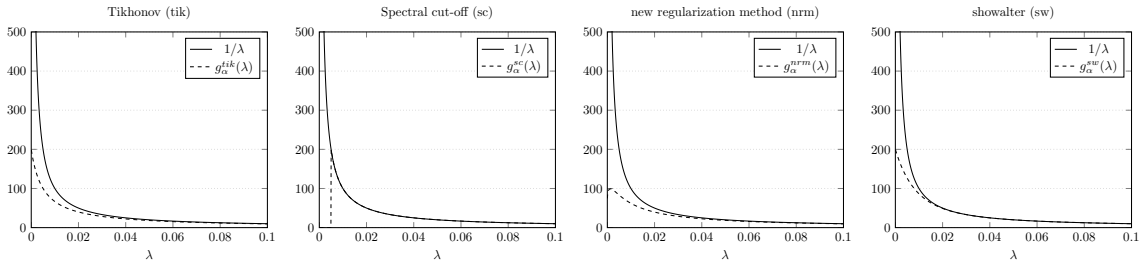

Fig. 1: Comparison generator function g_α^{reg} to function $\lambda \mapsto 1/\lambda$ for the four regularization methods (reg = tik,sc,nrm,sw).

5 Numerical illustration

The aim here is to compare the performance of our new regularization method (`nrm`) to the classical Tikhonov method (`tik`), spectral cut-off (`tsvd`), Showalter (`sw`) and conjugate gradient (`cg`) for some (ill-posed) test problems. We consider three test problems. The first one is a problem of image reconstruction found in [36]. The second problem is a Fredholm integral equation of the first kind taken from [2] and the last one is an inverse heat problem. For the discretization of these problems, we use the functions `shaw()`, `baart()` and `heat()` of the `matlab` regularization tool package (see [18]). For the `heat()` and `shaw()` test problems, the discretization is done by collocation with approximation of integrals by quadrature rules. For the `baart()` test problem, the discretization is done by Galerkin methods with orthonormal box functions as basis functions. In the `matlab` regularization tool package, each of the functions `shaw()`, `baart()` and `heat()` takes as input a discretization level n representing either the number of collocations points or the number of box functions considered on the interval $[0, 1]$. Given the input n, each function returns three outputs: a matrix A, a vector $x^\dagger$ and the vector y obtained by discretization without noise added. In this section, we considered the following discretization level for the `shaw()`, `baart()` and `heat()` test problem respectively: $n_{shaw} = 160$, $n_{baart} = 150$ and $n_{heat} = 150$. For the simulations, we define noisy data $y_\xi = y + \xi$ where ξ is a random white noise vector. In order to compute the regularized solution $x^\delta_{\alpha,nrm}$ for the new method, we compute the SVD with the function `svd()` in `Matlab` and applied (49).

We consider a 4% noise level, the noise level being defined here by the ratio of the noise to the exact data. More precisely, given a noisy data $y_\xi = y + \xi$, the noise level is defined by $\sqrt{E(||\xi||^2)}/||y||$. In order to illustrate the ill-posedness of each test problem, we give on Figure 2 the conditioning associated to each matrix A_{shaw}, A_{baart}, and A_{heat} obtained from the discretization of each problem.

	shaw	baart	heat
cond(A)	2.3283×10^{19}	2.4561×10^{17}	1.2706×10^{49}

Fig. 2: Conditioning of the matrices A_{shaw}, A_{baart} and A_{heat} for $n_{shaw} = 160$, $n_{baart} = 150$ and $n_{heat} = 150$.

We perform a Monte Carlo experiment of 3000 replications. In each replication, we compute the best relative error for each regularization method. Next we compute the minimum, maximum, average and standard deviation errors (denoted by e_{min}, e_{max}, $\bar{e}$, $\sigma(e)$ over the 3000 replications for each schemes (`nrm` and `tik`, `tsvd`, `sw` and `cg`). Figure 3 summarizes the results of the overall simulations.

In order to assess and compare the trade-off between stability and fidelity to the model for Tikhonov and the new method, we plot the curve of the conditioning versus relative error. The conditioning

here is the condition number of the reconstructed operator $g_\alpha^{reg}(T^*T)$ associated to the regularization method. For instance, using the invariance of conditioning by inversion, for the new method, the conditioning corresponds to the condition number of the operator $T^*T + \left[I - (T^*T)^{\sqrt{\alpha}}\right]^2$ while for Tikhonov method, it corresponds to the condition number of $T^*T + \alpha I$. In this respect, for two regularization methods, the best one is the one whose curve is below the other one as it achieves the same relative errors with smaller conditioning. On Figure 4, for each test problem, we compare the curve of conditioning versus relative error of the new method and Tikhonov method.

Notice that the first two problems (`shaw` and `baart`) are mildly ill-posed while the third problem (`heat`) is exponentially ill-posed.

baart

	nrm	tik	tsvd	sw	cg
e_{max}	0.34774	0.356887	0.346026	0.348538	0.346263
e_{min}	0.053773	0.055549	0.114955	0.051623	0.059753
$\bar{e}$	**0.16575**	**0.17078**	**0.19223**	**0.16581**	**0.19237**
$\sigma(e)$	**0.03993**	**0.04547**	**0.03466**	**0.04016**	**0.03488**

shaw

	nrm	tik	tsvd	sw	cg
e_{max}	0.186337	0.200609	0.181682	0.187038	0.186546
e_{min}	0.049132	0.048754	0.052527	0.047292	0.050999
$\bar{e}$	**0.13684**	**0.13857**	**0.15665**	**0.13746**	**0.15884**
$\sigma(e)$	**0.03307**	**0.03507**	**0.01955**	**0.03332**	**0.02269**

heat

	nrm	tik	tsvd	sw	cg
e_{max}	0.274714	0.283154	0.271837	0.270857	0.271285
e_{min}	0.100167	0.10736	0.115237	0.097847	0.106211
$\bar{e}$	**0.18857**	**0.19861**	**0.19007**	**0.18862**	**0.19365**
$\sigma(e)$	**0.02788**	**0.02756**	**0.02889**	**0.02725**	**0.02514**

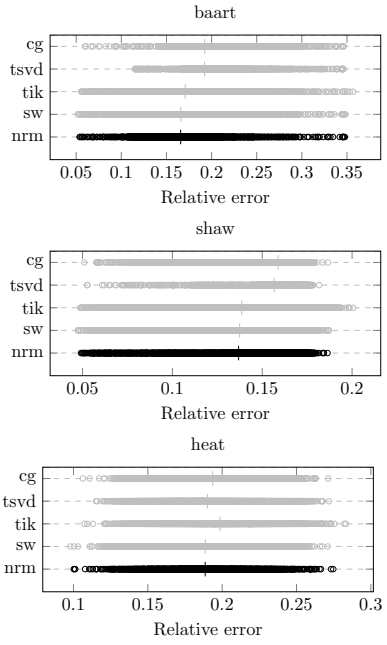

Fig. 3: Summary of the Monte Carlo experiment. On the right figure, the average relative error for each method is represented by the vertical stick.

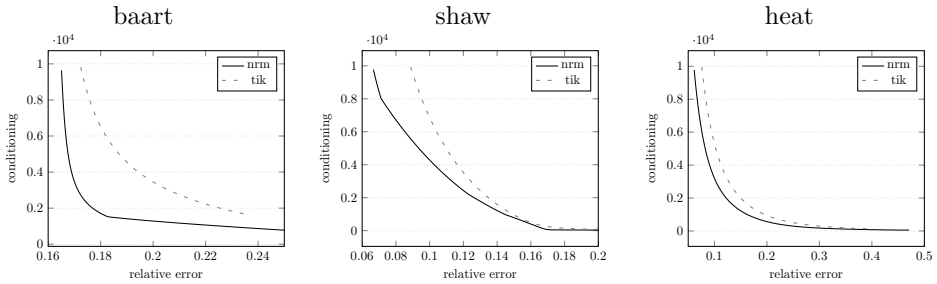

Fig. 4: Comparison of the trade-off between stability and accuracy of the new method (`nrm`) to Tikhonov (`tik`) for the three test problems: `shaw`, `baart` and `heat`.

Comments:

From Figure 3 and 4, we can do the following comments:

- The new method always yields the smallest average relative errors among the five methods.
- From Figure 3, we can see that both spectral cut-off and conjugate gradient yield the worst average relative errors except for the `heat` test problem where their average relative errors are smaller than the one of Tikhonov.
- For the two mildly ill-posed problem `shaw` and `baart`, Tikhonov method yields average relative errors close to the smallest one. On the contrary, for the exponentially ill-posed problem `heat`, Tikhonov method yields the worst average relative error among all the five methods.
- For the two mildly ill-posed problems (`shaw` and `baart`), the errors of the new method are not significantly smaller than those of Tikhonov on the contrary to the exponentially ill-posed problem (`heat`) where the new method produces smaller error than Tikhonov (about 5% smaller). This confirms our prediction about the better performance of the new method in instance of exponentially ill-posed problems compared to Tikhonov.
- For all three test problems, the new method performs better than spectral cut-off as could be expected. Moreover, the gap between the error is larger for the first two test problems which are mildly ill-posed. This also confirms the prediction about the poor performance of spectral cut-off for mildly ill-posed problems.
- On the contrary to the two mildly ill-posed problems (`shaw` and `baart`), spectral cut-off performs better than Tikhonov on the last test problem (`heat`), which is exponentially ill-posed. This emphasizes, especially in exponentially ill-posed problems, the drawback of Tikhonov method which regularizes all frequency component in the same way.
- From Figure 4, we can see that the new method achieves a better trade-off between stability and fidelity to the model compared to the Tikhonov method. Indeed, for the three test problems the curve associated to the new method lies below the one of Tikhonov. This means that given a stability level κ (measured in term of conditioning), the new method provided a smaller error than Tikhonov. Conversely, for a given error level ϵ, the new method provides a lower conditioning of the reconstructed operator compared to Tikhonov. This also validates the prediction stated earlier.

6 Parameter selection rules

In this section, we are interested in the choice of the regularization parameter α. For practical purposes, we assume that we don't know the smoothness conditions satisfied by the unknown solution $x^\dagger$. Consequently, we are left with two types of parameter choice rules: A-posteriori rules which use information on the noise level δ and heuristic rules which depend only on the noisy data y^δ. However a huge default of a-posteriori parameter choice rules is their dependence on the noise level δ which, in practice, is hardly available or well estimated in most circumstances. In [8], it is shown how an underestimation or overestimation of the noise level δ may induce serious computation issues for the Morozov principle. Moreover, in [15], it is illustrated how heuristic rules may outperform sophisticated a-posteriori rules. Given those reasons, we turn to heuristic (or data driven) selection rules. We recall that, due to Bakushinskii véto [3], such rules are not convergent. But still, as mentioned earlier, heuristic rules may yields better approximations compared to sophisticated a-posteriori rules (see e.g. [15]) and this is not surprising as the Bakushinskii result is based on worst case scenario.

We applied five noise-free parameter choice rules to the new method and the four regularization methods on the three test problems defined in Section 5: the generalized cross validation (GCV), the

discrete quasi-optimality rule (DQO), two heuristic rules (H1 and H2) and a variant of the L-curve method (LCV) each described in [11, Section 4.5]. Roughly speaking, the parameter α chosen by each of those selection rules is as follows:

- The GCV rule consists in choosing $\hat{\alpha}$ as

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \frac{||Tx_\alpha^\delta - y^\delta||}{\operatorname{tr}(r_\alpha(T^*T))},$$

  where r_α is the *residual* function associated to the regularization method under consideration. For the new method, r_α is defined in (20).

- The DQO method consists in discretizing the regularization parameter α as

$$\alpha_n = \alpha_0 q^n, \quad \alpha_0 \in (0, ||T^*T||], \quad \text{and} \quad 0 < q < 1.$$

  Next, the parameter $\hat{\alpha}$ is chosen as

$$\hat{\alpha} = \alpha_{\hat{n}} \quad \text{with} \quad \hat{n} = \underset{n\in\mathbb{N}}{\operatorname{argmin}} ||x_{\alpha_{n+1}}^\delta - x_{\alpha_n}^\delta||. \tag{56}$$

  Recall that this rule defined by (56) is actually one of the variants of the continuous quasi-optimality rule defined by

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} ||\alpha \frac{\partial x_\alpha^\delta}{\partial \alpha}||.$$

- The third rule H1 taken in [11, Section 4.5] consists in choosing the parameter $\hat{\alpha}$ as

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \frac{1}{\sqrt{\alpha}} ||Tx_\alpha^\delta - y^\delta||. \tag{57}$$

- The fourth rule H2 which is a variant of the third rule H1 consists in choosing the parameter $\hat{\alpha}$ as

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \frac{1}{\alpha} ||T^*(Tx_\alpha^\delta - y^\delta)||. \tag{58}$$

- The variant of the L-curve (LCV) method considered here (see [11, Proposition 4.37]) consists in choosing the regularization parameter $\hat{\alpha}$ as

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} ||x_\alpha^\delta||\, ||Tx_\alpha^\delta - y^\delta||. \tag{59}$$

  Recall that this rule actually tries to locate the parameter $\hat{\alpha}$ corresponding to the corner of the L-curve plot $||x_\alpha^\delta||$ versus $||Tx_\alpha^\delta - y^\delta||$ in a log-log scale. For more details about the L-curve method, see e.g. [10, 16, 17].

For a comprehensive discussion of the above heuristic rules and conditions under which convergence is established, see [13, 25, 39] for GCV, [4, 5, 6, 22]) for Quasi-optimality and [11, Section 4.5] for the rules H1, H2 and LCV.

For assessing the performance of each selection rule, we perform a Monte Carlo experiment of 3000 replications. For each replication, each test problem (`baart`, `shaw`, `heat`), and each regularization method (`nrm`, `tik`, `tsvd`, `sw` and `cg`), we compute the optimal regularization parameter α_{OPT}, the one chosen by each selection rule ($\alpha_{GCV}, \alpha_{DQO}, \alpha_{H1}, \alpha_{H2}, \alpha_{LCV}$). We also compute the corresponding relative errors:

$$\frac{||x^\dagger - x_{\alpha_{OPT}}^\delta||}{||x^\dagger||}, \quad \frac{||x^\dagger - x_{\alpha_{GCV}}^\delta||}{||x^\dagger||}, \quad \frac{||x^\dagger - x_{\alpha_{DQO}}^\delta||}{||x^\dagger||}, \quad \frac{||x^\dagger - x_{\alpha_{H1}}^\delta||}{||x^\dagger||}, \quad \frac{||x^\dagger - x_{\alpha_{H2}}^\delta||}{||x^\dagger||}, \quad \text{and} \quad \frac{||x^\dagger - x_{\alpha_{LCV}}^\delta||}{||x^\dagger||}.$$

In order to analyse the convergence behavior of the selection rules, we consider two noise levels: 2% and 4%. The results are shown in Figure 5 and Tables 1 to 9.

From Tables 1, 2 and Figure 5, we can see the following concerning the new regularization method:

		shaw						baart					
		OPT	GCV	DQO	H1	H2	LCV	OPT	GCV	DQO	H1	H2	LCV
4% nl	e_{max}	0.18634	12.5834	x	0.20946	x	0.237316	0.347742	73.7316	x	0.349441	x	0.348385
	e_{min}	0.049132	0.055259	x	0.171478	x	0.096168	0.053773	0.143293	x	0.337422	x	0.181454
	$\bar{e}$	**0.13684**	**0.22309**	**x**	**0.18391**	**x**	**0.16075**	**0.16575**	**0.49844**	**x**	**0.34279**	**x**	**0.26562**
	$\sigma(e)$	**0.03307**	**0.44215**	**x**	**0.00512**	**x**	**0.01818**	**0.03993**	**3.10514**	**x**	**0.00165**	**x**	**0.03412**
	$\overline{reg.par.}$	**0.02095**	**0.02113**	**x**	**0.16317**	**x**	**0.03472**	**4.221e-3**	**6.81e-3**	**x**	**0.16309**	**x**	**0.02471**
2% nl	e_{max}	0.17458	6.32855	x	0.18501	x	0.245324	0.25839	42.4191	x	0.33493	x	0.273046
	e_{min}	0.03759	0.052238	x	0.16929	x	0.052099	0.05213	0.114199	x	0.307401	x	0.162974
	$\bar{e}$	**0.11391**	**0.17803**	**x**	**0.17507**	**x**	**0.12994**	**0.14712**	**0.42564**	**x**	**0.32104**	**x**	**0.19394**
	$\sigma(e)$	**0.03420**	**0.29453**	**x**	**0.00212**	**x**	**0.02828**	**0.03134**	**1.84369**	**x**	**0.00414**	**x**	**0.01828**
	$\overline{reg.par.}$	**7.58e-3**	**0.01144**	**x**	**0.11727**	**x**	**7.769e-3**	**2.627e-3**	**2.25e-3**	**x**	**0.0483**	**x**	**9.38e-3**

Tab. 1: Summary of the Monte carlo experiment with the five heuristic rules GCV, DQO, H1, H2 and LCV applied to the new method for the test problems `shaw` and `baart`. The x indicates columns where the average relative error is greater than 1.

		heat					
		OPT	GCV	DQO	H1	H2	LCV
4% nl	e_{max}	0.274714	2.44329	0.279108	0.962294	7.22502	0.306221
	e_{min}	0.100167	0.109427	0.130933	0.267733	0.267816	0.101507
	$\bar{e}$	**0.18857**	**0.23329**	**0.205499**	**0.73711**	**0.647173**	**0.19349**
	$\sigma(e)$	**0.02788**	**0.13091**	**0.0209**	**0.30401**	**0.47841**	**0.02709**
	$\overline{reg.par.}$	**8.14e-4**	**6.235e-4**	**1.145e-3**	**0.64709**	**1.677e-4**	**8.842e-4**
2% nl	e_{max}	0.207777	2.34338	0.25523	0.261773	7.98943	0.289426
	e_{min}	0.073866	0.082679	0.081314	0.187784	0.228114	0.081314
	$\bar{e}$	**0.13947**	**0.17187**	**0.15295**	**0.2261**	**0.60093**	**0.16643**
	$\sigma(e)$	**0.01988**	**0.09237**	**0.02109**	**0.01094**	**0.51234**	**0.02929**
	$\overline{reg.par.}$	**5.204e-4**	**3.823e-4**	**6.909e-4**	**1.642e-3**	**8.736e-5**	**3.245e-4**

Tab. 2: Summary of the Monte carlo experiment with the five heuristic rules GCV, DQO, H1, H2 and LCV applied to the new method for the test problem `heat`.

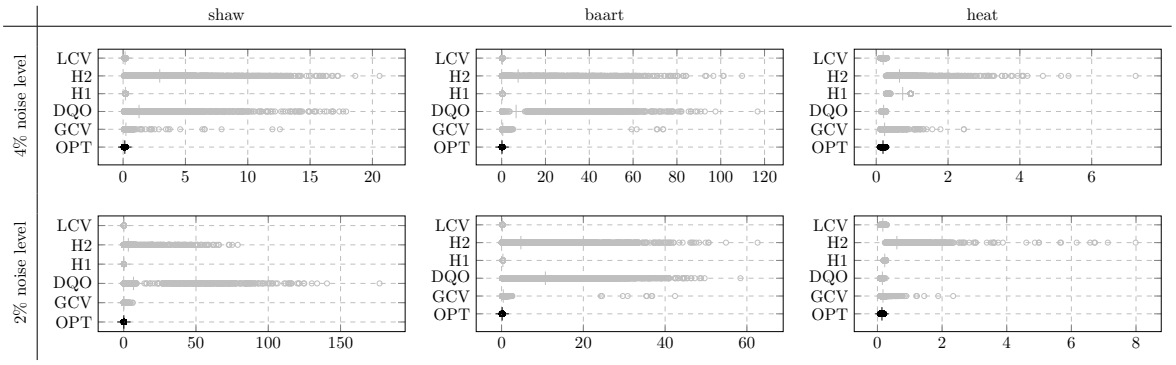

Fig. 5: Comparison of the relative error obtained by each selection rules (GCV, DQO, H1, H2 and LCV) for the two noise levels with the new method for the three tests problems `shaw`, `baart` and `heat`. On each plot, the x-axis corresponds to relative error and the vertical stick indicates the average relative error.

- For the exponentially ill-posed problem `heat`, from Table 2 and the last column of Figure 5, we can see that the discrete quasi-optimality rule and the variant of the L-curve are very efficient parameter choice rules for the new method. Indeed both the average relative errors and the

GCV		shaw				baart				heat			
		nrm	tik	tsvd	sw	nrm	tik	tsvd	sw	nrm	tik	tsvd	sw
4% nl	e_{max}	12.5834	5.5176	x	6.67616	73.7316	x	x	9.58465	2.44329	2.37791	x	2.69057
	e_{min}	0.055259	0.057251	x	0.052697	0.143293	x	x	0.152739	0.109427	0.111222	x	0.106857
	$\bar{e}$	**0.22309**	**0.38167**	**x**	**0.37981**	**0.49844**	**x**	**x**	**0.61592**	**0.23329**	**0.27623**	**x**	**0.23077**
	$\sigma(e)$	**0.44215**	**0.77228**	**x**	**0.81026**	**3.10514**	**x**	**x**	**1.08024**	**0.13091**	**0.18167**	**x**	**0.14511**
	$\overline{reg.par.}$	**0.02113**	**3.425e-3**	**x**	**0.02126**	**6.81e-3**	**x**	**x**	**4.376e-3**	**6.235e-4**	**2.816e-5**	**x**	**9.504e-5**
2% nl	e_{max}	6.32855	6.88993	x	8.85682	42.4191	x	x	4.83936	2.34338	3.20621	x	1.4366
	e_{min}	0.052238	0.047132	x	0.048763	0.114199	x	x	0.114434	0.082679	0.087569	x	0.080352
	$\bar{e}$	**0.17803**	**0.40427**	**x**	**0.42985**	**0.42564**	**x**	**x**	**0.37425**	**0.17187**	**0.21141**	**x**	**0.16403**
	$\sigma(e)$	**0.29453**	**0.94332**	**x**	**1.05855**	**1.84369**	**x**	**x**	**0.52374**	**0.09237**	**0.1747**	**x**	**0.09442**
	$\overline{reg.par.}$	**0.01144**	**1.487e-3**	**x**	**0.01023**	**2.25e-3**	**x**	**x**	**1.113e-3**	**3.823e-4**	**1.224e-5**	**x**	**3.608e-5**

Tab. 3: Summary of the Monte Carlo experiment with GCV rule applied to nrm,tik,tsvd and sw for the two noise levels on the three tests problems `shaw`, `baart` and `heat`.The x indicates columns where the average relative error is greater than 1.

DQO		shaw					baart					heat				
		nrm	tik	tsvd	sw	cg	nrm	tik	tsvd	sw	cg	nrm	tik	tsvd	sw	cg
4% nl	e_{max}	x	1.0787	x	1.06752	x	x	5.18865	x	5.01141	x	0.279108	0.283843	x	0.330631	x
	e_{min}	x	0.06105	x	0.128984	x	x	0.128386	x	0.12797	x	0.130933	0.124984	x	0.14209	x
	$\bar{e}$	**x**	**0.24255**	**x**	**0.1676**	**x**	**x**	**0.30292**	**x**	**0.26028**	**x**	**0.2055**	**0.20955**	**x**	**0.19973**	**x**
	$\sigma(e)$	**x**	**0.12996**	**x**	**0.03473**	**x**	**x**	**0.56905**	**x**	**0.4043**	**x**	**0.02090**	**0.02349**	**x**	**0.02085**	**x**
	$\overline{reg.par.}$	**x**	**8.16e-3**	**x**	**0.03038**	**x**	**x**	**8.146e-4**	**x**	**1.052e-3**	**x**	**1.145e-3**	**1.064e-4**	**x**	**1.486e-4**	**x**
2% nl	e_{max}	x	3.14131	x	2.83056	x	x	x	x	2.80503	x	0.25523	0.215776	x	0.312175	x
	e_{min}	x	0.043051	x	0.066213	x	x	x	x	0.081915	x	0.081314	0.085678	x	0.081303	x
	$\bar{e}$	**x**	**0.25222**	**x**	**0.23781**	**x**	**x**	**x**	**x**	**0.63955**	**x**	**0.15295**	**0.15585**	**x**	**0.16007**	**x**
	$\sigma(e)$	**x**	**0.31277**	**x**	**0.33941**	**x**	**x**	**x**	**x**	**0.47203**	**x**	**0.02109**	**0.01896**	**x**	**0.02309**	**x**
	$\overline{reg.par.}$	**x**	**3.589e-3**	**x**	**0.02565**	**x**	**x**	**x**	**x**	**1.486e-4**	**x**	**6.909e-4**	**4.773e-5**	**x**	**7.363e-5**	**x**

Tab. 4: Summary of the Monte Carlo experiment with DQO rule applied to nrm,tik,tsvd,sw and cg for the two noise levels on the three tests problems `shaw`, `baart` and `heat`.The x indicates columns where the average relative error is greater than 1.

H1		shaw					baart				
		nrm	tik	tsvd	sw	cg	nrm	tik	tsvd	sw	cg
4% nl	e_{max}	0.20946	0.256012	0.186581	0.248679	0.258152	0.349441	0.571788	0.348214	0.370115	0.346263
	e_{min}	0.171478	0.20863	0.169889	0.222574	0.159991	0.337422	0.372977	0.345054	0.338217	0.337497
	$\bar{e}$	**0.18391**	**0.23142**	**0.17112**	**0.23484**	**0.16948**	**0.34279**	**0.56375**	**0.34527**	**0.35003**	**0.34174**
	$\sigma(e)$	**5.12e-3**	**7.41e-3**	**1.44e-3**	**4.26e-3**	**5.95e-3**	**1.65e-3**	**4.39e-3**	**2.8e-4**	**4.19e-3**	**1.26e-3**
	$\overline{reg.par.}$	**0.16317**	**0.20483**	**0.25**	**0.42888**	**0.25025**	**0.16309**	**0.99972**	**0.5**	**0.20382**	**0.5**
2% nl	e_{max}	0.18501	0.201602	0.17421	0.226267	0.173244	0.334929	0.299022	0.345847	0.334085	0.34373
	e_{min}	0.16929	0.176076	0.169887	0.20998	0.163789	0.307401	0.187861	0.192709	0.30379	0.192243
	$\bar{e}$	**0.17507**	**0.1876**	**0.1702**	**0.21769**	**0.16824**	**0.32104**	**0.23703**	**0.34423**	**0.31878**	**0.341**
	$\sigma(e)$	**2.12e-3**	**3.77e-3**	**3.64e-4**	**2.61e-3**	**1.18e-3**	**4.14e-3**	**0.01658**	**0.01082**	**4.52e-3**	**8.65e-3**
	$\overline{reg.par.}$	**0.11728**	**0.07912**	**0.25**	**0.31221**	**0.25**	**0.0483**	**6.405e-3**	**0.49834**	**0.0483**	**0.49884**

Tab. 5: Summary of the Monte Carlo experiment with rule H1 applied to nrm,tik,tsvd,sw and cg for the two noise levels on the two tests problems `shaw` and `baart`

average regularization parameters produced by the DQO and LCV rules are very near the optimal ones and decrease as the noise level decreases. Moreover, by looking at the standard deviation of the relative error $\sigma(e)$, we see that those rules are very stable with respect to variations of the error term in y. Next, the GCV rule exhibit good average relative error, however the GCV is not stable with respect to the noise in y and this is shown by the spreading of dots along the x-axis or the corresponding large standard deviation $\sigma(e)$. Finally, the rule H2 is unstable and produces large relative errors norm whereas the rule H1 is more stable but do not yield

H2		shaw					baart					heat				
		nrm	tik	tsvd	sw	cg	nrm	tik	tsvd	sw	cg	nrm	tik	tsvd	sw	cg
4% nl	e_{max}	x	0.398139	x	1.06734	x	x	0.571788	x	3.89167	x	7.22502	0.968345	x	0.966995	x
	e_{min}	x	0.377665	x	0.128937	x	x	0.553372	x	0.151206	x	0.267816	0.967581	x	0.248926	x
	$\bar{e}$	**x**	**0.38798**	**x**	**0.16766**	**x**	**x**	**0.56381**	**x**	**0.28348**	**x**	**0.64717**	**0.968**	**x**	**0.96617**	**x**
	$\sigma(e)$	**x**	**2.96e-3**	**x**	**0.0347**	**x**	**x**	**2.67e-3**	**x**	**0.18975**	**x**	**0.47841**	**1.025e-4**	**x**	**0.01839**	**x**
	$\overline{reg.par.}$	**x**	**1**	**x**	**0.0306806**	**x**	**x**	**1**	**x**	**0.0356374**	**x**	**1.677e-4**	**1**	**x**	**0.999333**	**x**
2% nl	e_{max}	x	0.218418	x	1.2439	x	x	0.299022	x	2.05708	x	7.98943	0.413507	x	0.228358	x
	e_{min}	x	0.198588	x	0.066161	x	x	0.187861	x	0.118871	x	0.228114	0.389721	x	0.088958	x
	$\bar{e}$	**x**	**0.20743**	**x**	**0.16357**	**x**	**x**	**0.23703**	**x**	**0.21752**	**x**	**0.60093**	**0.40162**	**x**	**0.16603**	**x**
	$\sigma(e)$	**x**	**2.79e-3**	**x**	**0.02574**	**x**	**x**	**0.01658**	**x**	**0.17003**	**x**	**0.51234**	**3.941e-3**	**x**	**0.01742**	**x**
	$\overline{reg.par.}$	**x**	**0.13262**	**x**	**0.0273778**	**x**	**x**	**6.405e-3**	**x**	**6.403e-3**	**x**	**8.736e-5**	**1.04e-3**	**x**	**1.008e-4**	**x**

Tab. 6: Summary of the Monte Carlo experiment with rule H2 applied to nrm,tik,tsvd,sw and cg for the two noise levels on the three tests problems `shaw`, `baart` and `heat`.The x indicates columns where the average relative error is greater than 1.

H1		heat				
		nrm	tik	tsvd	sw	cg
4% nl	e_{max}	0.962294	0.968345	x	0.966995	0.314004
	e_{min}	0.267733	0.967581	x	0.966207	0.196006
	$\bar{e}$	**0.73711**	**0.968**	**x**	**0.96665**	**0.2548**
	$\sigma(e)$	**0.3040**	**1.025e-4**	**x**	**1.065e-4**	**0.02256**
	$\overline{reg.par.}$	**0.64709**	**1**	**x**	**1**	**0.1522**
2% nl	e_{max}	0.261773	0.413507	x	0.576622	0.235608
	e_{min}	0.187784	0.25434	x	0.226524	0.118365
	$\bar{e}$	**0.2261**	**0.36496**	**x**	**0.52891**	**0.20326**
	$\sigma(e)$	**0.01094**	**0.04828**	**x**	**0.10989**	**0.01554**
	$\overline{reg.par.}$	**1.642e-3**	**8.135e-4**	**x**	**5.594e-3**	**0.1229**

Tab. 7: Summary of the Monte Carlo experiment with rule H1 applied to nrm,tik,tsvd,sw and cg for the two noise levels on the test problem `heat`.The x indicates columns where the average relative error is greater than 1.

LCV		shaw					baart				
		nrm	tik	tsvd	sw	cg	nrm	tik	tsvd	sw	cg
4% nl	e_{max}	0.237316	0.238876	0.186581	0.239964	0.186546	0.348385	0.362202	0.348214	0.348893	0.346263
	e_{min}	0.096168	0.082638	0.169889	0.087736	0.159991	0.181454	0.180919	0.345054	0.18049	0.337497
	$\bar{e}$	**0.16075**	**0.15554**	**0.17112**	**0.16065**	**0.16917**	**0.26562**	**0.26142**	**0.34527**	**0.27845**	**0.34174**
	$\sigma(e)$	**0.01818**	**0.02355**	**1.437e-3**	**0.01931**	**2.664e-3**	**0.03412**	**0.02758**	**2.804e-4**	**0.04829**	**1.26e-3**
	$\overline{reg.par.}$	**0.03472**	**8.899e-3**	**0.25**	**0.03802**	**0.25**	**0.02471**	**9.998e-3**	**0.5**	**0.04848**	**0.5**
2% nl	e_{max}	0.245324	0.243518	0.281998	0.245056	0.276412	0.273046	0.27934	0.240446	0.274307	0.243326
	e_{min}	0.052097	0.048994	0.146842	0.051787	0.060724	0.162974	0.14908	0.166265	0.15763	0.158944
	$\bar{e}$	**0.12994**	**0.12848**	**0.1598**	**0.12998**	**0.14911**	**0.19394**	**0.19351**	**0.17416**	**0.19351**	**0.17377**
	$\sigma(e)$	**0.02828**	**0.03026**	**0.01637**	**0.0282**	**0.0297**	**0.01828**	**0.01885**	**0.01036**	**0.01832**	**0.01068**
	$\overline{reg.par.}$	**7.769e-3**	**2.173e-3**	**0.2**	**3.491e-3**	**0.21322**	**9.38e-3**	**2.34e-3**	**0.33333**	**4.392e-3**	**0.33333**

Tab. 8: Summary of the Monte Carlo experiment with LCV rule applied to nrm,tik,tsvd,sw and cg for the two noise levels on the two tests problems `shaw` and `baart`

satisfactory errors.

- For the mildly ill-posed test problems `shaw` and `baart`, the best heuristic rule for the new method is the variant of the L-curve method. Indeed, from Table 1 and two first columns of Figure 5, we notice that the relative errors produced by the LCV rule are near the optimal ones. Moreover, the LCV rule is very stable with respect to the noise in y and both the relatives errors and the

LCV		heat				
		nrm	tik	tsvd	sw	cg
4% nl	e_{max}	0.306221	0.328995	0.364771	0.334528	0.345675
	e_{min}	0.101507	0.113791	0.134865	0.103206	0.120651
	$\bar{e}$	**0.19349**	**0.20276**	**0.2104**	**0.19778**	**0.19995**
	$\sigma(e)$	**0.02709**	**0.02928**	**0.03006**	**0.02904**	**0.02629**
	$\overline{reg.par.}$	**8.842e-4**	**5.708e-5**	**0.06756**	**1.253e-4**	**0.10016**
2% nl	e_{max}	0.289426	0.29641	0.345262	0.295178	0.30907
	e_{min}	0.081314	0.089006	0.100711	0.083142	0.086875
	$\bar{e}$	**0.16643**	**0.18146**	**0.17867**	**0.16141**	**0.16292**
	$\sigma(e)$	**0.02929**	**0.02828**	**0.03907**	**0.02992**	**0.03034**
	$\overline{reg.par.}$	**3.245e-4**	**1.245e-5**	**0.05105**	**2.233e-5**	**0.06903**

Tab. 9: Summary of the Monte Carlo experiment with LCV rule applied to nrm,tik,tsvd,sw and cg for the two noise levels on the test problem `heat`.

regularization parameters decrease as the noise level decreases. The second best rule is rule H1 which is also stable and convergent but produces relative errors larger than the one of LCV rule. Finally the rules DQO, GCV and H2 are unstable and produce large relative error norm.

From Tables 3 to 9, we apply the five selection rules GCV, DQO, H1, H2 and LCV to each regularization method. Obviously the GCV rule cannot be applied to conjugate gradient method due to its non-linear character. Although, the DQO is originally designed for continuous regularization methods, notice that the rule defined in (56) can be applied to regularization methods with discrete regularization parameter such as truncated singular value decomposition and conjugate gradient. Indeed, we can applied the DQO rule to `tsvd` and `cg` by replacing $x_{\alpha_n}^{\delta}$ by x_k^{δ} in (56). Similarly the rules H1 and H2 originally designed for continuous regularization methods may be applicable to discrete regularization by defining the regularization parameter α as the inverse of the discrete parameter k. Following that idea, we applied the rules H1 and H2 to `tsvd` and `cg` by replacing α by $1/k$ in (57) and (58).

From Tables 3 to 9, we can do the following comments:

- The variant of the L-curve method defined through (59) is a very efficient heuristic parameter choice rule for each considered regularization method. Indeed, from Tables 8 and 9, by looking at the standard deviation $\sigma(e)$ of the relative error, we see that the LCV rule is stable for each regularization method, each test problem and each noise level. Next, the rule exhibits a convergent behavior for each test problem and each regularization method since the average relative error $\bar{e}$ and the average regularization parameter $\overline{reg.par.}$ decrease as the noise level decreases. Finally from Tables 3 to 9, we find that the LCV rule always yields the smallest average relative error $\bar{e}$ among all the heuristic rules considered except in 4 cases (out of 30 cases in total) : `baart` test problem with 4% noise level for Showalter method and `heat` test problem with 2% noise level for the new method, Tikhonov and Showalter method. Notice that in each of those four cases, LCV rule yields the second best average relative error $\bar{e}$ after the DQO rule.

- For the exponentially ill-posed test problem `heat`, Table 10 summarizes the best heuristic rules for each regularization method:

- For the mildly ill-posed test problems `shaw` and `baart`, the best heuristic rule is always the LCV rule. For the new method, Tikhonov, truncated singular value decomposition and conjugate gradient, the LCV rule is followed by rule H1 whereas for the Showalter method, the LCV rule is followed by rule H2.

	nrm	tik	tsvd	sw	cg
best heuristic rules	DQO,LCV	DQO,LCV	LCV	DQO,LCV	LCV

Tab. 10: Summary best heuristic rules for each regularization method for the exponentially ill-posed test problem `heat`.

- For the exponentially ill-posed test problem `heat`, by comparing the five regularization methods combined each with its best heuristic selection rule among GCV, DQO, H1, H2 and LCV, we see that the new method equipped with the DQO rule (resp. the LCV rule) for 4% noise level (resp. for 2% noise level) yields the smallest average relative error $\bar{e}$ (about 2% smaller than the second best average relative error). For 4% noise level, the second smallest average relative error is achieved by Showalter method equipped with LCV rule whereas for 2% noise level, the second smallest average relative error is achieved by Tikhonov method equipped with DQO rule.
- For the two mildly ill-posed problems `shaw` and `baart`, by comparing the five regularization methods combined each with its best heuristic selection rule among GCV, DQO, H1, H2 and LCV, we notice there is no regularization method which always yields the smallest average relative error. For the `shaw` test problem, Tikhonov method with LCV rule yields the smallest average relative error $\bar{e}$. For the `baart` test problem, for 4% noise level, the smallest average relative error is obtained by the Showalter method equipped with the DQO rule. However, for this test problem, the DQO rule is not converging for the Showalter method as the average relative error $\bar{e}$ increases from 0.26028 to 0.63955 as the noise level decreases from 4% to 2%. If we discard Showalter with DQO rule, then for 4% noise level, the smallest average relative error are obtained by Tikhonov method equipped with LCV rule while for 2% noise level, the smallest average relative errors are obtained from conjugate gradient method equipped with LCV rule.

Remark 6. *From Tables 1 to 9, we see that, the heuristic parameter choice rule LCV yields very satisfactory results for each considered regularization method. This reinforces the idea that the Bakushinskii véto [3] should not be seen as a limitation of heuristic parameter choice rule but rather as a safeguard to be taken into account.*

In summary, we see that for the exponentially ill-posed test problem `heat`, the new regularization method always yields the smallest average relative error among the five considered regularization methods even when we consider heuristic parameter choice rules. Hence in practical situation of exponentially ill-posed problems, we expect the new method to perform better than the other regularization methods (Tikhonov, truncated singular value decomposition, Showalter method and conjugate gradient).

7 Conclusion

In this paper, we presented a new regularization method which is particularly suitable for linear exponentially ill-posed problems. We study convergence analysis of the new method and we provided order optimal convergence rates under logarithmic source conditions which has a natural interpretation in term of Sobolev spaces for exponentially ill-posed problems. For a general source conditions expressed via index functions, we only provided quasi order optimal rates. From the simulatins performed, we saw that the new method performs better than Tikhonov method, spectral cut-off, Showalter and conjugate gradient for the considered exponentially ill-posed problem, even with heuristic parameter choice rules. For the two mildly ill-posed problems treated, we saw that the new method actually yields results quite similar to those of Tikhonov and Showalter methods. The results of Section 6, where we

applied five *error-free* selection rules to the five regularization methods suggest that the variant of the L-curve method defined in (59) and the discrete quasi-optimality rule defined in (56) are very efficient parameter choice rules for the new method in the context of exponentially ill-posed problem. In the context of mildly ill-posed problems, the results of experiments suggest that the LCV rule described in Section 6 is preferable.

Interesting perspectives would be a theoretical analysis of the LCV and DQO rules for the new regularization method in the framework of exponentially ill-posed problems in order to shed light on their good performances.

Acknowledgements: The author would like to thank Pierre Maréchal, Anne Vanhems for their helpful comments, readings and remarks.

8 Appendix

Proof of Proposition 1. Let us state the following standard inequality that we will use in the sequel:

$$\forall\, t \geq 0, \quad \exp(-t) \leq \frac{1}{1+t}. \tag{60}$$

Using (60) applied with $t = -\sqrt{\alpha}\ln(\lambda) \geq 0$, we get

$$1 - \exp(\sqrt{\alpha}\ln(\lambda)) \geq 1 - \frac{1}{1-\sqrt{\alpha}\ln(\lambda)} = \frac{-\sqrt{\alpha}\ln(\lambda)}{1-\sqrt{\alpha}\ln(\lambda)} = \sqrt{\alpha}\frac{|\ln(\lambda)|}{1+\sqrt{\alpha}|\ln(\lambda)|}. \tag{61}$$

But since $\alpha < 1$, $1 + \sqrt{\alpha}|\ln(\lambda)| < 1 + |\ln(\lambda)|$. Furthermore For all $\lambda \leq a < 1$, by the monotonicity of the function $t \longrightarrow |\ln(t)|/(1+|\ln(t)|) = -\ln(t)/(1-\ln(t))$ on $(0,1)$, we get that

$$\frac{|\ln(t)|}{1+|\ln(t)|} \geq \frac{|\ln(a)|}{1+|\ln(a)|} \quad \forall\, t \in (0,a).$$

By applying the above inequality to (61) and taking the square, we get:

$$\forall \lambda \in (0,a), \quad (1-\lambda^{\sqrt{\alpha}})^2 \geq M\alpha \quad \text{with} \quad M = \left(\frac{|\ln(a)|}{1+|\ln(a)|}\right)^2.$$

Whence the following inequality:

$$\frac{1}{\lambda + (1-\lambda^{\sqrt{\alpha}})^2} \leq \frac{1}{\lambda + M\alpha}, \tag{62}$$

which implies that

$$\sqrt{\lambda} g_\alpha(\lambda) \leq \frac{\lambda^{1/2}}{\lambda + M\alpha}. \tag{63}$$

It is rather straightforward to prove that the supremum over $\lambda \in (0,1)$ of the right hand side of (63) is of order $\alpha^{-1/2}$ from which we deduce that

$$\sup_{\lambda \in (0,a]} \sqrt{\lambda} g_\alpha(\lambda) = \mathcal{O}\left(\frac{1}{\sqrt{\alpha}}\right) \tag{64}$$

□

Proof of Lemma 1. Let $\lambda \in (0,1)$. On the one hand, by applying the estimate $(1-\exp(t)) \geq -t/(1-t)$ which holds for all $t<0$ to $t=\sqrt{\alpha}\ln(\lambda)$ and by taking squares, we have:

$$(1-\lambda^{\sqrt{\alpha}})^2 \geq \frac{\alpha|\ln(\lambda)|^2}{(1+\sqrt{\alpha}|\ln(\lambda)|)^2}. \tag{65}$$

On the other hand, using the estimate $t^2 \geq (1-\exp(t))^2$ valid for all $t<0$ to $t=\sqrt{\alpha}\ln(\lambda)$, we get

$$(1-\lambda^{\sqrt{\alpha}})^2 \leq \alpha|\ln(\lambda)|^2 \tag{66}$$

Now, for $\alpha \leq \lambda < 1$, $|\ln(\alpha)| \geq |\ln(\lambda)|$ which implies that $\sqrt{\alpha}|\ln(\lambda)| \leq \sqrt{\alpha}|\ln(\alpha)|$. Using the estimate $t^\mu \ln(1/t) \leq \mu$ which is true for all t in $(0,1)$ and every positive μ to $t=\lambda$ and $\mu=1/2$, we deduce that

$$1+\sqrt{\alpha}|\ln(\alpha)| \leq 3/2. \tag{67}$$

So, from (65) and (67), we deduce that

$$(1-\lambda^{\sqrt{\alpha}})^2 \geq \frac{4}{9}\alpha|\ln(\lambda)|^2 \tag{68}$$

which implies that

$$r_\alpha(\lambda) \leq \frac{(1-\lambda^{\sqrt{\alpha}})^2}{\lambda + (4/9)\alpha|\ln(\lambda)|^2} \tag{69}$$

Finally, applying (66) and the fact that $\lambda \geq (4/9)\lambda$ to (69) yields (22). □

Proof of Lemma 2. (i) It is straightforward to check that (31) is indeed the derivative of the function $\Psi_{p,\alpha}$.

(ii) First notice that $\lim_{\lambda\to 0} h(\lambda) = +\infty$, hence, it suffices to find a $\bar{\lambda}$ such that $h(\bar{\lambda}) < 0$ to deduce the existence of a root of the function h on $(0,\bar{\lambda}]$. If $p<2$, then $h(1)<0$. If $p=2$, then $h(\lambda) = |\ln(\lambda)|(2\alpha|\ln(\lambda)| - \lambda)$. Thus, $h(\lambda)<0$ for λ close to 1 but smaller than 1. If $p>2$, then $\lim_{\alpha\to 0} h(\lambda) = \lambda(p-2+\ln(\lambda)) < 0$ for all $\lambda < \exp(2-p)$.

Now let us show that for every $\lambda(p,\alpha)$ which vanishes h, (33) holds.

$$h(\lambda)=0 \quad \Longrightarrow \quad \alpha = \lambda|\ln(\lambda)|^{-1}\left(\frac{2-p+|\ln(\lambda)|}{p|\ln(\lambda)|}\right) \tag{70}$$

by monotonicity of the function $t \to (2-p+t)/(pt)$ (irrespective of the sign of $2-p$) and $t \to |\ln(\lambda)|$, we get that the function $l(\lambda) = \frac{2-p+|\ln(\lambda)|}{p|\ln(\lambda)|}$ is monotonic. If $p<2$, the function l is increasing and we then get that, for all $\lambda \in (0,c]$ with $c<1$,

$$\frac{1}{p} \leq l(\lambda) \leq l(c). \tag{71}$$

On the other hand, if $p \geq 2$, the function l is decreasing and for $\lambda \in (0,c]$ with $c < \exp(2-p)$, we get

$$l(c) \leq l(\lambda) \leq 1/p. \tag{72}$$

From (70), (71) and (72), we deduce that

$$h(\lambda)=0 \quad \Longrightarrow \quad \alpha \sim \lambda|\ln(\lambda)|^{-1}. \tag{73}$$

From [37, Lemma 3.3], we get that

$$\alpha \sim \lambda|\ln(\lambda)|^{-1} \Rightarrow \lambda \sim \alpha|\ln(\alpha)|(1+o(1)) \quad \text{for} \quad \alpha \to 0.$$

This shows that the maximizers $\lambda(p,\alpha)$ of the function $\Psi p,\alpha$ satisfies (33). Now let us deduce (34). We have

$$\alpha|\ln(\alpha)|^p\Psi_{p,\alpha}(\alpha|\ln(\alpha)|) = \frac{|\ln(\alpha)|^p \times |\ln(\alpha|\ln(\alpha)|)|^{2-p}}{|\ln(\alpha)| + |\ln(\alpha|\ln(\alpha)|)|^2} < |\ln(\alpha)|^p \times |\ln(\alpha|\ln(\alpha)|)|^{-p}$$

With the change of variable $\varrho = |\ln(\alpha)|$ (i.e. $\alpha = \exp(-\varrho)$), we have

$$\begin{aligned} |\ln(\alpha)|^p \times |\ln(\alpha|\ln(\alpha)|)|^{-p} &= \frac{\varrho^p}{|\ln(\varrho\exp(-\varrho))|^p} \\ &= \frac{\varrho^p}{|-\varrho+\ln(\varrho)|^p} \\ &= \frac{\varrho^p}{(\varrho-\ln(\varrho))^p} \to 1 \quad \text{as} \quad \varrho\to\infty. \end{aligned}$$

This proves that

$$\alpha|\ln(\alpha)|^p\Psi_{p,\alpha}(\alpha|\ln(\alpha)|) = \mathcal{O}(1)$$

and thus from (33), we deduce that (34) holds.

□

Proof of Proposition 3. For simplicity of notation, let $\alpha := \alpha(\delta, y^\delta)$. In order to establish (42), we are going to bound the terms $||x^\dagger - x_\alpha||$ and $||x_\alpha - x_\alpha^\delta||$ separately. Let us start with the regularization error term. Given that $x^\dagger \in X_{f_p}(\rho)$, we have $x^\dagger = f_p(T^*T)w$ and thus $x^\dagger - x_\alpha = r_\alpha(T^*T)x^\dagger = f_p(T^*T)r_\alpha(T^*T)w$. Hence by applying [20, Proposition 1] to $x^\dagger - x_\alpha$, we get

$$||x^\dagger - x_\alpha|| \le ||r_\alpha(T^*T)w||\sqrt{\phi_p^{-1}\left(||y-Tx_\alpha||^2/\rho^2\right)} \le \rho\sqrt{\phi_p^{-1}\left(||y-Tx_\alpha||^2/\rho^2\right)}. \tag{74}$$

From (28) and (74), we deduce that

$$||x^\dagger - x_\alpha|| \le \rho f_p\left(||y-Tx_\alpha||^2/\rho^2\right)(1+o(1)). \tag{75}$$

But

$$\begin{aligned} ||y-Tx_\alpha|| &\le ||y^\delta - Tx_\alpha^\delta|| + ||y - Tx_\alpha - (y^\delta - Tx_\alpha^\delta)|| \\ &\le \delta + \sqrt{\delta} + ||r_\alpha(T^*T)(y-y^\delta)|| \\ &\le 2\delta + \sqrt{\delta} \\ &= \sqrt{\delta}(2\sqrt{\delta}+1). \end{aligned} \tag{76}$$

From (75) and (76), we deduce that

$$||x^\dagger - x_\alpha|| \le \rho f_p\left(\delta(2\sqrt{\delta}+1)^2/\rho^2\right)(1+o(1)). \tag{77}$$

Using (77) and the fact that

$$\frac{f_p\left(\delta(2\sqrt{\delta}+1)^2/\rho^2\right)}{f_p(\delta)} = \left(\frac{-\ln(\delta)}{-\ln(\delta) - 2\ln(1+2\sqrt{\delta}) + 2\ln(\rho)}\right)^p \to 1 \quad \text{as} \quad \delta\to 0, \tag{78}$$

yields

$$||x^\dagger - x_\alpha|| = \mathcal{O}\left(f_p(\delta)\right) \quad \text{as} \quad \delta\to 0. \tag{79}$$

Now let us estimate the propagated data noise term. Let $\bar\alpha = q\alpha$ with $q \in (1,2)$. From (41), since $\bar\alpha > \alpha$, we get

$$||Tx_{\bar\alpha}^\delta - y^\delta|| > \delta + \sqrt{\delta}. \tag{80}$$

Therefore,

$$\begin{aligned} ||Tx_{\bar{\alpha}} - y|| &\geq ||Tx^{\delta}_{\bar{\alpha}} - y^{\delta}|| - ||T(x^{\delta}_{\bar{\alpha}} - x_{\bar{\alpha}}) - (y^{\delta} - y)|| \\ &> \delta + \sqrt{\delta} - ||r_{\bar{\alpha}}(T^*T)(y^{\delta} - y)|| \\ &\geq \delta + \sqrt{\delta} - \delta \\ &= \sqrt{\delta}. \end{aligned} \tag{81}$$

On the other hand, $||Tx_{\bar{\alpha}} - y|| = ||T(x_{\bar{\alpha}} - x^{\dagger})|| = ||(T^*T)^{1/2}(x_{\bar{\alpha}} - x^{\dagger})|| = ||(T^*T)^{1/2} r_{\bar{\alpha}}(T^*T)x^{\dagger}||$. By applying (44) with $\varphi(t) = \sqrt{t}$ and $\epsilon = 1/8$, we get that there exists a constant C such that $||(T^*T)^{1/2} r_{\bar{\alpha}}(T^*T)x^{\dagger}|| \leq C\bar{\alpha}^{3/8}$. This implies that

$$||Tx_{\bar{\alpha}} - y|| \leq C\bar{\alpha}^{3/8}. \tag{82}$$

From (81) and (82), we deduce that $\bar{\alpha}^{3/8} \geq \sqrt{\delta}/C$ which implies that $\bar{\alpha} \geq \bar{C}\delta^{4/3}$ with $\bar{C} = C^{-8/3}$. From (21), (39), the above lower bound of $\bar{\alpha}$ and the fact that $\alpha > \bar{\alpha}/2$, we get that, there exists a positive constant C' such that

$$||x_{\alpha} - x^{\delta}_{\alpha}|| \leq C' \frac{\delta}{\sqrt{\alpha}} \leq C'\sqrt{2}\frac{\delta}{\sqrt{\bar{\alpha}}} \leq C'\sqrt{2/\bar{C}}\frac{\delta}{\sqrt{\delta^{4/3}}} = \delta^{1/3} C'\sqrt{2/\bar{C}}. \tag{83}$$

Given that $\delta^{1/3} = \mathcal{O}\left(f_p(\delta)\right)$ as $\delta \to 0$, we deduce that $||x_{\alpha} - x^{\delta}_{\alpha}|| = \mathcal{O}\left(f_p(\delta)\right)$ as $\delta \to 0$ which together with (79) implies (42). □

References

[1] N. Alibaud, P. Maréchal and Y. Saesor, *A variational approach to the inversion of truncated Fourier operators*, Inverse Problems 25 (2009), no. 4 .

[2] M. L. Baart, *The use of auto-correlation for pseudorank determination in noisy ill-conditioned linear least-squares problems*, IMA J. Numer. Anal. 2 (1982), no. 2, pp 241–247.

[3] A. B. Bakušinskiĭ, *Remarks on the choice of regularization parameter from quasioptimality and relation tests*, Zh. Vychisl. Mat. i Mat. Fiz. 24 (1984), no. 8, 1258–1259.

[4] F. Bauer and S. Kindermann, *Recent results on the quasi-optimality principle*, J. Inverse Ill-Posed Probl. 17 (2009), no. 1, pp 5–18.

[5] F. Bauer and S. Kindermann, *The quasi-optimality criterion for classical inverse problems*, Inverse Problems 24 (2008), no. 3.

[6] F. Bauer and M. Reiss, *Regularization independent of the noise level: an analysis of quasi-optimality*, Inverse Problems 24 (2008), no. 5.

[7] X. Bonnefond and P. Maréchal, *A variational approach to the inversion of some compact operators*, Pac. J. Optim. 5 (2009), no. 1, pp 97–110.

[8] X. Bonnefond, P. Maréchal and W. C. Simo Tao Lee, *A note on the Morozov principle via Lagrange duality*, Set-Valued Var. Anal. 26 (2018), no. 2, 265–275.

[9] B. Eicke , A. K. Louis and R. Plato, *The instability of some gradient methods for ill-posed problems*, Numer. Math., 58:129–134, 1990.

[10] H. W. Engl and W. Grever, *Using the L-curve for determining optimal regularization parameters*, Numer. Math. 69 (1994), no. 1, 25–31.

[11] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of inverse problems*, Mathematics and its Applications, 375. Kluwer Academic Publishers Group, Dordrecht, 1996.

[12] G. H. Golub, C. F. Van Loan, *Matrix computations.* , Third edition, Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 1996.

[13] G. H. Golub, M. Heath and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics 21 (1979), no. 2, pp 215–223.

[14] C. W. Groetsch, *Inverse problems in the mathematical sciences*, Vieweg Mathematics for Scientists and Engineers. Friedr. Vieweg & Sohn, Braunschweig, 1993.

[15] M. Hanke and P. C. Hansen, *Regularization methods for large-scale problems*, Surveys Math. Indust. 3 (1993), no. 4, pp 253–315.

[16] P. C. Hansen , *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev. 34 (1992), no. 4, 561–580.

[17] P. C. Hansen and D. P. O'Leary , *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput. 14 (1993), no. 6, 1487–1503.

[18] P. C. Hansen, *Regularization Tools version 4.0 for Matlab 7.3*, Numer. Algorithms 46 (2007), no. 2, pp 189–194.

[19] B. Hofmann and P. Mathé, *Analysis of profile functions for general linear regularization methods*, SIAM J. Numer. Anal. 45 (2007), no. 3, pp 1122–1141.

[20] T. Hohage, *Regularization of exponentially ill-posed problems*, Numer. Funct. Anal. Optim. 21(2000), no. 3-4, pp 439-464.

[21] A. Kirsch , *An introduction to the mathematical theory of inverse problems*, Applied Mathematical Sciences, 120. Springer-Verlag, New York, 1996.

[22] A. S. Leonov, *On the choice of regularization parameters by means of quasi-optimality and ratio criteria*, Soviet. Math. Dokl. 19 (1978), no. 3.

[23] A. K. Louis, *A unified approach to regularization methods for linear ill-posed problems*, Inverse Problems 15 (1999), no. 2, pp 489–498.

[24] A. K. Louis and P. Mass, *A mollifier method for linear operator equations of the first kind*, Inverse Problems 6 (1990), no. 3,pp 427–440.

[25] M. A. Lukas, *Asymptotic optimality of generalized cross-validation for choosing the regularization parameter*, Numer. Math. 66 (1993), no. 1, pp 41–66.

[26] B. A. Mair, *Tikhonov regularization for finitely and infinitely smoothing operators*, SIAM J. Math. Anal. 25 (1994), no. 1, pp 135–147.

[27] P. Mathé, *Saturation of regularization methods for linear ill-posed problems in Hilbert spaces*, SIAM J. Numer. Anal. 42 (2004), no. 3, pp 968–973.

[28] P. Mathé and S. V. Pereverzev, *Geometry of linear ill-posed problems in variable Hilbert scales*, Inverse Problems 19(2003), no. 3, pp 789–803.

[29] P. Mathé and B. Hofmann, *How general are general source conditions?*, Inverse Problems 24 (2008), no. 1.

[30] C. A. Micchelli and T. J. Rivlin, *A survey of optimal recovery. Optimal estimation in approximation theory*, Proc. Internat. Sympos., Freudenstadt, 1976, Plenum Press (1977), pp 1–54.

[31] D. A. Murio, *The mollification method and the numerical solution of ill-posed problems*, A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1993.

[32] M. T. Nair, E. Schock and U. Tautenhahn, *Morozov's discrepancy principle under general source conditions*, Z. Anal. Anwendungen 22 (2003), no. 1, pp 199–214.

[33] A. Neubauer, *On converse and saturation results for regularization methods*, Beiträge zur angewandten Analysis und Informatik, Shaker, Aachen, (1994) pp 262–270.

[34] A. Neubauer, *On converse and saturation results for Tikhonov regularization of linear ill-posed problems*, SIAM J. Numer. Anal. 34 (1997), no. 2, pp 517–527.

[35] E. Schock, *Approximate solution of ill-posed equations: arbitrarily slow convergence vs. superconvergence*, Constructive methods for the practical treatment of integral equations (1984), pp 234–243.

[36] C. B. Jr. Shaw, *Improvement of the resolution of an instrument by numerical solution of an integral equation*, J. Math. Anal. Appl. 37 (1972), pp 83–112.

[37] U. Tautenhahn, *Optimality for ill-posed problems under general source conditions*, Numer. Funct. Anal. Optim. 19 (1998), no. 3-4, pp 377–398.

[38] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems.*, John Wiley & Sons, 1977.

[39] G. Wahba, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM J. Numer. Anal. 14 (1977), no. 4, pp 651–667.

Chapter 4

Une approche par mollification de la régression non paramétrique instrumentale

4.1 Présentation

Dans ce chapitre, il est question d'appliquer la formulation variationnelle de la mollification pour la régularisation du problème de régression instrumentale non-paramétrique. La consistance et la stabilité de l'estimateur obtenu par l'approche sont démontrées dans le cadre stochastique classique où l'opérateur T et la donnée g sont tous les deux inconnus mais estimés à l'aide d'échantillons statistiques. Une comparaison de notre approche à la méthode de Tikhonov et la *spectral cut-off* est aussi considéré. Enfin une section numérique est consacré à des simulations qui confirment l'efficacité de cette nouvelle approche dans le cadre de la régularisation du problème de régression instrumentale.

L'article suivant a été coécrit avec Pierre Maréchal et Annes Vanhems et a été soumis pour publication dans le journal *Business & Economic Statistics*. Mes contributions principales dans cet article ont été l'analyse des hypothèses liées à la mal-position du problème et l'écriture des codes et de la section Simulations. J'ai aussi participé activement à la preuve de la consistance des estimateurs. Le reste de l'article résulte d'une collaboration avec mes co-auteurs.

4.2 Discussion sur les applications numériques

Dans l'article qui suit, une partie importante est consacrée à des simulations numériques où la forme variationnelle de la mollification est appliquée au problème de régression instrumentale. Nous décrivons ici certains détails concernant les simulations.

Tout d'abord, le schéma d'approximation numérique d'une fonction continue h sur un intervalle $[a, b]$ est celui des fonctions étagées. Plus précisément, après discrétisation de l'intervalle $[a, b]$ en n sous-intervalles $[x_{i-1}, x_i)$, $i = 1, ..., n$, la fonction h est approximé par la fonction $\bar{h}_n$ définie par

$$\forall x \in (a, b), \quad \bar{h}_n(x) = \sum_{i=1}^{n} h(t_i) 1_{[x_{i-1}, x_i[}(x), \quad \text{avec} \quad t_i \in [x_{i-1}, x_i).$$

En optant pour une discrétisation uniforme, pour i allant de 1 à n, on a $x_i = a + i\delta$ avec δ étant le pas de discrétisation définie par $\delta = (b-a)/n$.

Dans l'application de la mollification, un opérateur clé est l'opérateur de convolution C_β défini dans le cas unidimensionnelle par

$$C_\beta h = \varphi_\beta \star h, \quad \text{avec} \quad \varphi_\beta(x) = \frac{1}{\beta}\varphi\left(\frac{x}{\beta}\right),$$

où la fonction φ satisfait $\int_{\mathbb{R}} \varphi(t)\,\mathrm{d}t = 1$. Nous rappelons qu'une propriété clé de l'opérateur C_β est la convergence vers l'opérateur identité quand β tends vers 0. Dans le souci de préserver cette propriété essentielle dans la version discrétisée de C_β et de rester au mieux fidèle à la convolution continue, nous optons pour une méthode non pas basée sur la convolution discrète (qui présente généralement des défauts aux extrémités de l'intervalle considéré) mais sur une approche différente que nous décrivons ci-après.

Tout d'abord, appelons par ϕ_i la fonction porte sur le sous-intervalle $1_{[x_{i-1},x_i)}(x)$, i.e. $\phi_i(x) = 1_{[x_{i-1},x_i)}(x)$. Étant donnée une fonction h_n définie sur un intervalle $[a,b]$ par

$$h_n(x) = \sum_{i=1}^{n} a_i 1_{[x_{i-1},x_i)}(x) = \sum_{i=1}^{n} a_i \phi_i(x),$$

on a

$$C_\beta h_n(x) = \sum_{i=1}^{n} a_i (C_\beta \phi_i)(x) = \sum_{i=1}^{n} a_i\, \varphi_\beta \star \phi_i(x).$$

Ainsi la fonction $h_{n,\beta}$ définie par

$$h_{n,\beta} = \sum_{j=1}^{n} \left(C_\beta h_n\right)(t_j)\phi_j(x) = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} a_i\, \varphi_\beta \star \phi_i(t_j)\right) \phi_j(x), \quad \text{avec} \quad t_j \in [x_{j-1}, x_j) \tag{4.1}$$

est une approximation plausible de $C_\beta h_n$ sur l'intervalle (a,b). Noter que les fonctions $h_{n,\beta}$ et h_n sont définis dans le même base $\{\phi_i\}_{i=1,\ldots,n}$. Pour déterminer $h_{n,\beta}$, il ne reste plus qu'à calculer $\varphi_\beta \star \phi_i(t_j)$.

$$\begin{aligned}
\varphi_\beta * \phi_i(t_j) = \int_{\mathbb{R}} \varphi_\beta(t_j - y)\phi_i(y)\,\mathrm{d}y &= \int_{x_{i-1}}^{x_i} \varphi_\beta(t_j - y)\,\mathrm{d}y \\
&= \int_{x_{i-1}}^{x_i} \frac{1}{\beta}\varphi\left(\frac{t_j - y}{\beta}\right)\,\mathrm{d}y \\
&= \int_{\frac{t_j - x_i}{\beta}}^{\frac{t_j - x_{i-1}}{\beta}} \varphi(u)\,\mathrm{d}u \\
&= cdf(\frac{t_j - x_{i-1}}{\beta}) - cdf(\frac{t_j - x_i}{\beta}),
\end{aligned} \tag{4.2}$$

où cdf est la fonction définie par $cdf(x) = \int_{-\infty}^{x} \varphi(t)\,\mathrm{d}t$. Notons que si la fonction φ est positive alors, cdf est la fonction de répartition associé à la densité φ.

Soit $\bar{C}_{\beta,n}$ la matrice carré d'ordre n définie par

$$(\bar{C}_{\beta,n})_{ji} = \varphi_\beta * \phi_i(t_j) = cdf(\frac{t_j - x_{i-1}}{\beta}) - cdf(\frac{t_j - x_i}{\beta}). \tag{4.3}$$

L'équation (4.1) entraine que les coordonnées $c_{n,\beta}$ de $h_{n,\beta}$ (approximation de la fonction $C_\beta h_n$) est la matrice $\bar{C}_{\beta,n}$ multiplié par le vecteur c_n (coordonnées de la fonction h_n).

Ainsi on approxime l'opérateur de convolution C_β par la matrice $\bar{C}_{\beta,n}$ défini en (4.3). Il ne reste plus qu'à définir les t_j de façon que la matrice $\bar{C}_{\beta,n}$ converge bien vers la matrice identité quand β tends vers 0.

En prenant $t_j = x_{j-1} + \delta/2 = x_{j-1/2}$, on obtient

$$\begin{aligned}(\bar{C}_{\beta,n})_{ji} = \varphi_\beta * \phi_i(t_j) &= cdf\left(\frac{x_{j-1/2} - x_{i-1}}{\beta}\right) - cdf\left(\frac{x_{j-1/2} - x_i}{\beta}\right) \\ &= cdf\left(\frac{(j-i+1/2)\delta}{\beta}\right) - cdf\left(\frac{(j-i-1/2)\delta}{\beta}\right). \end{aligned} \tag{4.4}$$

Ainsi de (4.4), en utilisant le fait que φ est d'intégrale égale à 1, i.e. $cdf(+\infty) = 1$, on obtient que

$$\begin{cases} (\bar{C}_{\beta,n})_{ji} \to cdf(+\infty) - cdf(+\infty) = 0 \quad \text{quand} \quad \beta \to 0 \quad \text{si} \quad i < j \\ (\bar{C}_{\beta,n})_{ji} \to cdf(-\infty) - cdf(-\infty) = 0 \quad \text{quand} \quad \beta \to 0 \quad \text{si} \quad i > j \\ (\bar{C}_{\beta,n})_{ii} \to cdf(+\infty) - cdf(-\infty) = 1 \quad \text{quand} \quad \beta \to 0, \end{cases} \tag{4.5}$$

ce qui prouve que la matrice $\bar{C}_{\beta,n}$ converge vers la matrice identité quand β tends vers 0.

Remark 4.1. *Si la fonction φ es paire, alors la fonction cdf vérifie $cdf(x) = 1 - cdf(-x)$ pour tout x dans $\mathbb{R}$ et on peut aisément vérifier que dans ce cas, $(\bar{C}_{\beta,n})_{ij} = (\bar{C}_{\beta,n})_{ji}$. Ainsi si la fonction φ est paire, alors la matrice $\bar{C}_{\beta,n}$ est symétrique.*

Remark 4.2. *Si t_j est égale à x_{j-1} ou x_j, alors on perd l'estimation (4.5) et par conséquent la matrice $\bar{C}_{\beta,n}$ ne converge plus vers la matrice identité quand β tends vers 0.*

Remark 4.3. *En choisissant φ comme des densités usuelles (e.g. noyau gaussien), l'évaluation de la fonction de répartition cdf intervenant dans le calcul des coefficients de la matrice $\bar{C}_{\beta,n}$ peut être fait directement à l'aide des fonctions cumulatives intégrés dans les packages de probabilité des langages de programmation.*

Par exemple, si φ est la densité de la Gaussienne normale, alors la fonction cdf équivaut à la fonction `normcdf` de `Matlab`. Et compte tenu de la symétrie de la matrice $\bar{C}_{\beta,n}$, l'évaluation de la fonction `normcdf` aux points $(k-1/2)\delta/\beta$, pour k allant de 0 à n suffit pour déterminer la matrice $\bar{C}_{\beta,n}$.

Dans l'idée de suivre la formulation originale de la mollification qui incorpore l'opérateur d'adéquation des données Φ_β (voir (1.120)), nous avons essayé dans un premier temps d'approximer cet opérateur par $TC_\beta T^\dagger$, expression qui est valide à condition que $TC_\beta T^\dagger$ soit borné, ce qui n'est pas en général le cas vu la composition avec l'opérateur non borné $T^\dagger$. Notons que la Proposition 1.9 qui donne une condition suffisante pour que $TC_\beta T^\dagger$ soit borné n'est pas vérifiée dans le problème de régression instrumentale. Par ailleurs, nous n'avons pas pu établir un résultat permettant d'affirmer la continuité ou bien la discontinuité de cet opérateur. Néanmoins, dans l'espoir que cet opérateur soit borné, nous avons essayé l'évaluation de cet opérateur à l'aide d'algorithmes de point proximal (voir, e.g. [44, 45]) et les méthodes de Krylov (voir, e.g. [118, Sections 6-7]) qui sont bien connus pour leurs propriétés régularisantes. Cependant nos tentatives ne furent pas couronné de succès, ce qui pourrait éventuellement indiquer qu'il n'existe pas d'opérateur borné Φ_β dans le cas de la régression instrumentale. C'est dans cet optique que nous avons opté pour la formulation sans l'opérateur Φ_β.

Suivant la description faite en Section 5.1, à partir d'un échantillon de données $(Z_l, W_l, Y_l)_{l=1,...,n}$, l'opérateur T, et la donnée r de l'équation

$$T\,h = r$$

sont discrétisés en la matrice T_n et le vecteur r_n. En plus la fonction inconnue h est discrétisée en

$$h_{proj} = \sum_{i=1}^{I} \mathrm{h}_i \phi_i$$

où $\mathrm{h} = (\mathrm{h}_1, ..., \mathrm{h}_I)$ est inconnue et $\{\phi_i\}_{i=1,...,I}$ est une base composée de fonctions portes sur l'intervalle $[0, 1]$ où est définie la fonction de régression. Noter qu'une telle discrétisation de la fonction h incorpore la contrainte lié au support, i.e. $h_{proj} \in L^2(V)$, avec $[0, 1] \subset V$. Ainsi au regard de la Section 4, on déduit que le vecteur h représentant les coefficients de la solution régularisée est la solution du problème de minimisation

$$\min_{\mathrm{h}\in\mathbb{R}^I} \left\|\mathrm{T}_n\mathrm{h} - \mathrm{r}_n\right\|_2^2 + \left\|(Id_I - \bar{C}_{\beta,I})\mathrm{h}\right\|_2^2, \tag{4.6}$$

où Id_I est la matrice carré identité d'ordre I, $\bar{C}_{\beta,I}$ est la discrétisation de l'opérateur C_β décrite plus haut, et $||\cdot||_2$ est la norme euclidienne sur $\mathbb{R}^I$. À partir de (4.6), on déduit alors que le vecteur inconnu h est solution de l'équation linéaire

$$\left[\mathrm{T}_n^*\mathrm{T}_n + (Id_I - \bar{C}_{\beta,I})^*(Id_I - \bar{C}_{\beta,I})\right]\mathrm{h} = T_n^*\mathrm{r}_n. \tag{4.7}$$

Ainsi, pour déterminer la solution régularisée, il suffit de résoudre le système linaire (4.7). Notons que pour les valeurs de β assez petite, une précaution particulière est nécessaire lors de la résolution du système (4.7) vue que dans de tel cas, la matrice $(Id_I - \bar{C}_{\beta,I})$ est de norme très petite et le conditionnement de la matrice $\left[\mathrm{T}_n^*\mathrm{T}_n + (Id_I - \bar{C}_{\beta,I})^*(Id_I - \bar{C}_{\beta,I})\right]$ approche alors celui de la matrice $T_n^*T_n$ qui est très grand.

La figure 4.1 compare la solutions non-régularisée (solution de l'équation $\mathrm{T}_n\mathrm{h} = \mathrm{r}_n$) avec la fonction de régression (dans les deux cas considérés dans l'article) et illustre la mal-position du problème et la nécessité d'appliquer des méthodes de régularisation afin d'obtenir des solutions raisonnables.

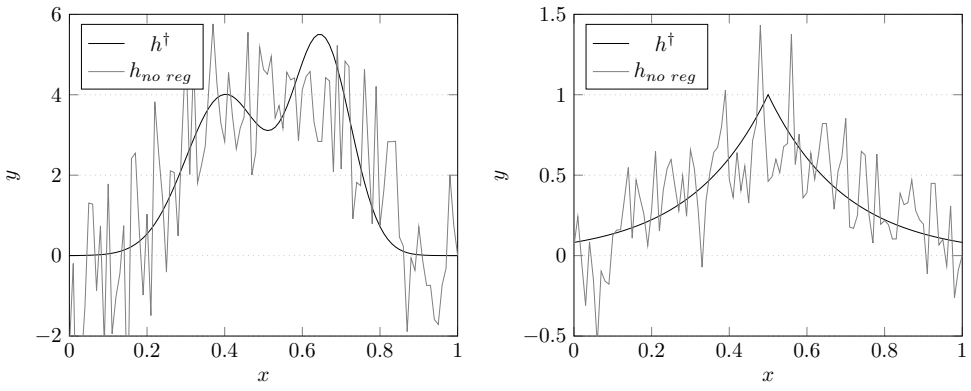

Figure 4.1: Comparaison de la fonction de régression avec la solution non-régularisée pour deux cas différents de la fonction de régression.

Une fois passé l'étape du calcul des solutions régularisées, vient l'étape du calcul du paramètre de régularisation β. Afin de déterminer une méthode de sélection de paramètre, nous avons testé plusieurs

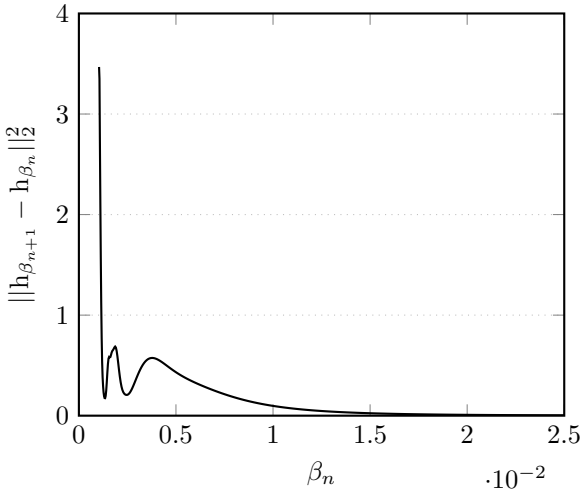

Figure 4.2: Illustration d'une courbe de $\left\|\mathrm{h}_{\beta_{n+1}} - \mathrm{h}_{\beta_n}\right\|_2^2$ qui présente plusieurs minimum locaux.

méthodes à savoir les cinq méthodes heuristiques utilisées dans le Chapitre 3. Après ces essais, la méthode de quasi-optimalité discrète s'est révélé être la meilleure, raison pour laquelle elle est celle appliquée dans l'article qui suit.

Comme décrit dans la Section 1.1.4.4 du Chapitre 1, la méthode de quasi-optimalité discrète consiste à discrétiser la paramètre de régularisation β en

$$\beta_n = \beta_0 q^n, \quad \text{avec} \quad \beta_0 \in (0, ||T||^2], \quad q \in (0,1) \quad n \in \mathbb{N}$$

et à considérer la paramètre

$$\beta^* = \beta_{n^*}, \quad \text{avec} \quad n^* = \underset{n \in \mathbb{N}}{\operatorname{argmin\,loc}} \left\|h_{\beta_{n+1}} - h_{\beta_n}\right\|_{L^2}^2, \tag{4.8}$$

où h_β représente la solution régularisée correspondant au paramètre de régularisation β.

Dans les simulations, on considère $q = 0.98$ de façon à avoir une discrétisation avec un maillage assez fin, $\beta_0 \sim ||\mathrm{T}_n||^2$ et la suite $(\beta_n)_n$ est tronqué à un rang N tel que

$$\left\|Id_I - \bar{C}_{\beta_N, I}\right\|_\infty \sim \epsilon_M$$

où ϵ_M représente l'erreur d'arrondi machine qui dans notre cas est de l'ordre de 10^{-16}. Étant donné que dans les simulations, la solution régularisée h_{β_n} est une fonction constante par intervalle, on constate alors que le carré de la norme L^2 de $h_{\beta_{n+1}} - h_{\beta_n}$ est proportionnelle au carré de la norme euclidienne du vecteur $\mathrm{h}_{\beta_{n+1}} - \mathrm{h}_{\beta_n}$, où h_β est le vecteur des coordonnées de la solution régularisée h_β dans la base des fonctions portes $\{\phi_i\}_{i=1,\dots,I}$. Ainsi dans l'approximation numérique, (4.8) est remplacé par

$$\beta^* = \beta_{n^*}, \quad \text{avec} \quad n^* = \underset{1 \le n \le N}{\operatorname{argmin\,loc}} \left\|\mathrm{h}_{\beta_{n+1}} - \mathrm{h}_{\beta_n}\right\|_2^2. \tag{4.9}$$

Il convient de noter que la fonction $\Psi(n) = \left\|\mathrm{h}_{\beta_{n+1}} - \mathrm{h}_{\beta_n}\right\|_2^2$ peut présenter plusieurs minimum locaux sur l'ensemble $\{1, ..., N\}$ comme on peut le voir sur la Figure 4.2. Dans un tel cas, nous appliquons la même procédure décrite dans la Section 3.2 du Chapitre 3.

A mollifier approach to the nonparametric instrumental regression problem

2020-11-09

Abstract

We propose a mollification method for the problem of instrumental regression. We show that our estimator is consistent, and compare its performances to that of the classical regularization methods. A finite sample study enables us to demonstrate the efficiency of mollification compared to other estimation methods.

Keywords– Nonparametric instrumental regression, ill-posed problems, regularization, mollification, parameter selection rule.

1 Introduction

The issue of instability in nonparametric estimation is now well-known in statistics and econometrics and has been analyzed in various important problems such as deconvolution or instrumental regression. For example, in the deconvolution problem, since the pioneering works by Stefanski and Carroll [1990] and Carroll and Hall [1988], a huge literature has been devoted to the instability issue: Fan [1991], Meister [2004], Johannes [2009], Comte et al. [2006, 2007, 2008], Carrasco and Florens [2011], among others.

In instrumental regression, since the work by JP Florens in 2000, many papers were published on the topic. Let us mention in particular Hall et al. [2005], Darolles et al. [2011], Carrasco et al. [2007, 2014] among others.

The issue of instability has also been studied in other various fields of econometrics as the Generalized methods of moments with a continuum of moment conditions (Carrasco and Florens [2000]), treatment models (Florens et al. [2002]), game-theoretic models (Florens and Sbaï [2010], Enache and Florens [2020]), Bayesian econometrics (Florens and Simoni [2012a,b]) and transformation models (Florens and Sokullu [2017]).

In this paper we use the variational approach to mollification. The originality and one main advantage of the mollification approach, compared to other variational methods, is that it reduces significantly the conflict between the stability and fidelity objectives. Moreover, an explicit targeted solution is defined, which is a definite conceptual advantage. One of the aims of the present work is to empirically enlighten the improvement provided by mollification on the above mentioned tradeoff.

100

The paper is organized as follows. In Section 2, we present the problem of instrumental regression and specify our functional setting. In Section 3, we review the main regularization techniques used in the literature. In Section 4, we introduce the mollification approach and analyze the asymptotic properties of the resulting estimator. Finally, in Section 5, we provide a finite sample study, discuss the issue of parameter selection, and evaluate the merits of mollification in comparison with other classical methods. The proofs of the asymptotic consistency results are gathered in the appendix.

2 Problem statement

2.1 Functional setting

In the classical regression setting, the objective is to model the relationship between a variable of interest and a set of explanatory variables. More precisely, consider the continuous random variables $Z\colon \Omega \to \mathbb{R}^p$ and $Y, \varepsilon\colon \Omega \to \mathbb{R}$ satisfying the relationship

$$Y = h(Z) + \varepsilon. \tag{1}$$

Here, Z is the *explanatory variable*, Y is the *response* and ε is an *error*, while $h\colon \mathbb{R}^p \to \mathbb{R}$ is an unknown function referred to as a *regression function*. As mentioned in the introduction, this setting is standard in econometrics and statistics, and many applications can be found in diverse fields like labor economics, bio-statistics, microeconomics and consumer demand.

The problem we address is that of identifying h, under the assumption that Z and ε are linked: $E(\varepsilon|Z) \neq 0$. Note that $E(\varepsilon|Z) = 0$ if and only if $E(Y|Z) = E(h(Z)|Z) = h(Z)$, thus in this case h is identified by $h(Z) = E(Y|Z)$. In the case where $E(\varepsilon|Z) \neq 0$, the function h cannot be identified as previously and it is then customary to introduce an *instrumental variable* $W\colon \Omega \to \mathbb{R}^q$, that is linked to Z and not to ε. The model is then written as $Y = h(Z) + \varepsilon, \quad E(\varepsilon|W) = 0$ and equation $E(\varepsilon|W) = 0$ implies that

$$E(Y|W) = E(h(Z)|W). \tag{2}$$

In order to put Equation (2) into a well defined and tractable functional framework, we shall make the following assumptions:

Assumption 1. The laws P_Z, P_W, P_Y are absolutely continuous with respect to the Lebesgue measure.

We denote by λ the Lebesgue measure on $\mathbb{R}^d$ whatever may be the dimension d. The densities of Y, Z, W are respectively denoted by f_Y, f_Z, f_W. We also let f_{YW}, f_{ZW} be the joint densities of $(Y, W), (Z, W)$, respectively. Finally, the conditional density of Z given W is denoted by $f_{Z|W}$. The equation $E(Y|W) = E(h(Z)|W)$ then reduces to the functional integral equation

$$\int \frac{f_{YW}(y,w)}{f_W(w)} y \,\mathrm{d}y = \int \frac{f_{ZW}(z,w)}{f_W(w)} h(z) \,\mathrm{d}z, \quad w \in \{\omega | f_W(\omega) \neq 0\}, \tag{3}$$

which is equivalent to

$$\int f_{YW}(y,w) y \,\mathrm{d}y = \int f_{ZW}(z,w) h(z) \,\mathrm{d}z, \quad w \in \{\omega | f_W(\omega) \neq 0\}. \tag{4}$$

It is customary to assume in addition that:

Assumption 2. The kernel $k := f_{ZW}$ is $\lambda \otimes \lambda$-square integrable.

Under this assumption, the linear integral operator

$$\begin{array}{rccl} T\colon & L^2(\mathbb{R}^p) & \longrightarrow & L^2(\mathbb{R}^q) \\ & h & \longmapsto & \displaystyle\int f_{ZW}(z,\cdot)h(z)\,\mathrm{d}z = \int k(z,\cdot)h(z)\,\mathrm{d}z, \end{array} \tag{5}$$

is a Hilbert-Schmidt, thus compact, operator. Notice that for Equation (4) to have a solution, it is necessary that $\int f_{YW}(y,\cdot)y\,\mathrm{d}y \in L^2(\mathbb{R}^q)$. Thus, we finally assume that:

Assumption 3. The function $r(w) := \int f_{YW}(y,w)y\,\mathrm{d}y$ belongs to $L^2(\mathbb{R}^q)$.

The last assumption is satisfied in particular if $E[Y^2] < \infty$ and f_W is bounded. In practice, r is estimated from observed sample, and the constraint that $r \in L^2(\mathbb{R}^q)$ may be incorporated in the estimation process.

Remark 1. Assumption 2 ensures that T is well-defined and compact. It should be noticed that the integral operator $h \mapsto \int f_{ZW}(z,\cdot)h(z)\,\mathrm{d}z$ is well-defined under the weaker assumption that $f_{ZW}(\cdot,w)$ is in $L^2(\mathbb{R}^q)$ for almost all values of w, which occurs naturally in the context of instrumental regression, as shows the following simple example. Suppose $Y = h(Z) + \varepsilon$ with $Z = W + U$, where W and U are independent. Here, necessarily, $p = q$. Then,

$$f_{Z|W=w}(z) = f_U(z - w),$$

in which f_U denotes the density of U. It follows that the corresponding kernel of T is given in this case by

$$k(z,w) = f_{ZW}(z,w) = f_W(w)f_U(z-w),$$

which does not necessarily belong to L^2. In this case, Equation (3) has the form of a deconvolution problem. The problem is, of course, ill-posed, and mollification has been explored for this type of problems. ∎

Under the assumptions 1, 2 and 3, the problem takes the form of the linear operator equation

$$Th = r. \tag{6}$$

The chosen Hilbert space setting facilitates the transfer of classical regularization methods from deterministic inverse problems to stochastic ones, as was shown in particular in Carrasco et al. [2007], Hall et al. [2005].

Notice that there is no practical downside to considering that Z takes its values in a compact set $V \subset \mathbb{R}^p$, an assumption which will be in force throughout. Our setting is then similar to that of Hall et al. [2005]. It differs significantly from that of Carrasco et al. [2007] by the following fact: the authors of the latter reference consider Equation (3) in the *weighted* Hilbert spaces $L^2(\mathbb{R}^p, \mathrm{d}P_Z)$ and $L^2(\mathbb{R}^q, \mathrm{d}P_W)$. By doing so, they ensure square integrability of the kernel of T, which places the problem in the familiar setting of compact operators. However, stabilizing the estimator of h in this topology has a different meaning than that of stabilizing it in the original L^2-topology.

2.2 Estimation

In practice, both the operator T and the function r are unknown in Equation (6), since they depend on unknown density functions. They must be replaced by estimated counterparts prior to solving the inverse problem.

Let $(Y_i, Z_i, W_i)_{i=1,\dots,n}$ be an i.i.d sample of observation of the variable (Y, Z, W). We denote by T_n and r_n be estimated versions of T and r, respectively. Various nonparametric estimation methods may be used, such as kernel based method or projection-based method. Our theoretical development, in particular our asymptotic results, will be provided with the generic notations T_n and r_n. In the simulation study, we shall use a projection based method for these estimators.

Notation We specify here the notation used in the paper. Given a set $V \subset \mathbb{R}^p$, we denote by $L^2(V)$ the space of square-integrable functions on $\mathbb{R}^p$ with essential support in V. If $T\colon L^2(V) \to L^2(\mathbb{R}^q)$ is a linear operator, we denote by $\operatorname{ran} T$ and $\ker T$ its range and kernel, respectively. The inner product in L^2-spaces is denoted by $\langle \cdot, \cdot \rangle$. The adjoint and the pseudo-inverse of T with respect to the usual inner products of $L^2(V)$ and $L^2(\mathbb{R}^q)$ are denoted by T^* and $T^\dagger$, respectively. The Fourier transform of an integrable function $\varphi\colon \mathbb{R}^p \to \mathbb{C}$ is defined as

$$\hat{\varphi}(\xi) = \int e^{-2i\pi\langle \xi, x\rangle} \varphi(x)\, \mathrm{d}x,$$

and the Fourier-Plancherel operator is defined accordingly.

3 Review of classical regularization techniques

The ill-posedness of Equation (6) is well-known. Existence, uniqueness and stability (Hadamard conditions) fail to be satisfied altogether. In our setting, $\operatorname{ran} T$ is not closed, so that a least square solution of (6) may not exist. As a matter of fact, the pseudo-inverse $T^\dagger$, defined on the dense proper subspace $\mathscr{D}(T^\dagger) = \operatorname{ran} T + \ker T^*$, is unbounded, so that the minimum norm least square solution $T^\dagger r$ does not depend continuously on the data r.

A desirable property of T is *injectivity*, which corresponds to the notion of completeness of the density f_{ZW}. The question of completeness of densities has been intensively studied (see e.g. Hu et al. [2017], Hu and Shiu [2011], Florens et al. [2011] and the references therein). In those papers, sufficient conditions for completeness were given in various settings. Nevertheless, these conditions are generally not testable in practice. Authors usually provide examples of conditional densities for which completeness holds.

Regularization methods aim at reformulating the problem in such a way that the Hadamard conditions are satisfied. This requires, of course, to somehow reduce the amount of information to be recovered. For example, *projection methods* use finite dimensional approximations of the objects under consideration, yielding linear systems with reasonable conditioning. Most of the modern methods introduce regularization constraints in the infinite dimensional problem prior to discretization.

Tikhonov-Philips regularization is historically the first regularization principle, and it remains to this day one of the most used in practice. It consists in replacing the original equation $Th = r$ by the *regularized normal equation*

$$(T^*T + \alpha Q^*Q)h = T^*r,$$

in which Q is an operator ensuring that $T^*T + \alpha Q^*Q$ admits a bounded inverse. It can equivalently be defined as the minimizer of the functional

$$h \mapsto \|Th - r\|^2 + \alpha \|Qh\|^2.$$

The weight α is referred to as the *regularization parameter*. The standard Tikhonov method correspond to the case $Q = I$. Interesting instances are obtained with the choice $Q = D$, where D is a differential operator (see Locker and Prenter [1980], Nair et al. [1997], Trummer [1984]). In this case, the domain of the regularized normal equation must be restricted to the domain of D. In the particular case where the operator D is a the second order differential operator, the function h is searched for in the Sobolev space $H^2(\mathbb{R}^p)$. Intuitively, the presence of D penalizes strong variations of h, hence encourages its smoothness.

Spectral cut-off is a regularization method which consists in cutting-off high spectral components of the operator T. Recall that, in the case of a compact operator T, the solution $h^\dagger$ is given by

$$h^\dagger = \sum_{j=1}^{\infty} \frac{1}{\sigma_j} \langle r, u_j \rangle v_j \tag{7}$$

where $(\sigma_j, u_j, v_j)_{j=1,\dots}$ is the singular value decomposition of T. The spectral cut-off consists in truncating the series:

$$h_k = \sum_{j=1}^{k} \frac{1}{\sigma_j} \langle r, u_j \rangle v_j. \tag{8}$$

The regularization parameter is here the truncation parameter k and the spectral cut-off belongs to the family of spectral methods (see Engl et al. [1996]). Unless the SVD is explicit, this method is difficult to implement since it requires the computation of the SVD.

Promoting smoothness of a solution may be achieved by means of the concept of mollification. In the next section, we introduce the variational mollification approach to the regularization of our problem.

4 Mollification

The term *mollification* is a generic term referring to the use of *mollifiers* for the purpose of regularizing ill-posed problems. The main concept of mollification consists in defining a smoothed version of the original unknown object as the *target object*, with the hope that the recovery of this target object will be well-posed.

Mollification takes three distinct forms:

1. In the earlier works (see Murio [2011] and the references therein), mollification was applied directly to the data prior to inversion. Obviously, this requires an explicit inversion formula for the operator, which limits the realm of potential applications.

2. In Louis and Maass [1990], another approach was introduced, based on Hermitian adjunction, which gave rise to the so-called *approximate inverses*. This approach requires to solve an adjoint equation, at least on a family of basis functions. The resulting methodology is referred to as the *approximate inverses* (see Schuster [2007] and the references therein).

3. A *variational* formulation of mollification appeared in Lannes et al. [1987] in the specific field of Fourier synthesis and deconvolution, and has developed independently: the analysis of the variational form has been studied mostly in the references Alibaud et al. [2009], Bonnefond and Maréchal [2009] and its application to various fields of applied science, including statistics and econometrics, is very promising.

We focus here on the third approach. Recall that we search for a solution h in the HIlbert space $H = L^2(V)$ where the set $V \subset \mathbb{R}^p$ is assumed to be compact. Working with $L^2(V)$ is a way to incorporate a priori knowledge on the support of h, which we always do in practice. We define the reconstructed object to be the solution of the optimization problem

$$(\mathscr{P}_\beta) \quad \left| \begin{array}{rl} \text{Minimize} & \|Th - r\|^2_{L^2(\mathbb{R}^q)} + \|(I - C_\beta)h\|^2_{L^2(\mathbb{R}^p)}, \\ \text{s.t.} & h \in H, \end{array} \right.$$

in which I and C_β denote respectively the natural injection of $L^2(V)$ into $L^2(\mathbb{R}^p)$ and some convolution operator on $L^2(\mathbb{R}^p)$ to be specified below.

Our variational formulation can be justified by the fact that our aim is now the reconstruction of a mollified version of the true unknown object $h^\dagger$, namely $C_\beta h^\dagger$, in which $C_\beta \colon L^2(\mathbb{R}^p) \to L^2(\mathbb{R}^p)$ is the convolution operator given by $C_\beta h := \varphi_\beta * h$. Here,

$$\varphi_\beta(x) = \frac{1}{\beta^p}\varphi\left(\frac{x}{\beta}\right), \quad x \in \mathbb{R}^p,$$

in which φ is an integrable function with unit integral. As $\beta \downarrow 0$, the function φ_β emulates the behaviour of the Dirac distribution, and the family $(C_\beta)_{\beta\in(0,1]}$ is commonly referred to as an *approximate unity*. The parameter β is our regularization parameter: decreasing β amounts to aiming at a higher resolution version of h. Recall that, by a standard approximation theorem, $C_\beta h \to h$ in L^2 as $\beta \downarrow 0$.

Remark 2. In the case where an explicit operator Φ_β such that $\Phi_\beta T = TC_\beta$ is known, the data r may be replaced in Problem $(\mathscr{P}_\beta)$ by the *mollified data* $\Phi_\beta r$ (see e.g. Alibaud et al. [2009], Maréchal et al. [2000, 2018]). We call the equation $\Phi_\beta T = TC_\beta$ an *intertwining relationship*, and Φ_β an *intertwining operator*. Most often, however, our instrumental regression operator T does not admit such an explicit operator, and we merely require fit to the *unmollified* data. ∎

We now proceed to establish mathematical stability and consistency of our regularization procedure. Recall that the operators T and $(I - C_\beta)$ are said to satisfy the *completion condition* if

$$\exists \gamma_\beta > 0 \colon \quad \forall h \in H, \quad \left\|Th\right\|^2 + \left\|(I - C_\beta)h\right\|^2 \geq \gamma_\beta \left\|h\right\|^2. \tag{9}$$

See Morozov [2012] and Engl et al. [1996, Chapter 8]. From elementary Hilbert space analysis, Condition (9) ensures that the operator $T^*T + (I - C_\beta)^*(I - C_\beta)$ admits a bounded inverse, and therefore that the unique solution to Problem $(\mathscr{P}_\beta)$, namely

$$h_\beta = R_\beta h := \left(T^*T + (I - C_\beta)^*(I - C_\beta)\right)^{-1} T^* r,$$

depends continuously on r. Here, A^* denotes as usual the adjoint of A with respect to the underlying Hilbert space structures. Notice that Condition (9) is automatically satisfied in the case where $H = L^2(V)$ with V compact. As a matter of fact, in this case, we can prove that $\|(I - C_\alpha)h\|^2_H \geq \nu_\alpha \|h\|^2_H$ for some positive constant ν_α (see, Alibaud et al. [2009, Lemma 12 and Proposition 5]), from which (9) follows immediately.

105

Recall that a *parameter choice rule* is a mapping $\beta\colon \mathbb{R}_+ \times H \to (0,1]$ such that

$$\sup\left\{ \beta(\delta, h^\delta) \;\middle|\; h^\delta \in H,\ \|h^\delta - h\| \leq \delta \right\} \longrightarrow 0 \quad \text{as} \quad \delta \downarrow 0.$$

The family of mappings $(R_\beta)_{\beta\in(0,1]}$ is called a *regularization* of $T^\dagger$ if for every $h \in \mathscr{D}(T^\dagger) = \operatorname{ran} T + \ker T^*$ there exists a *parameter choice rule* $\beta = \beta(\delta, h^\delta)$ such that

$$\sup\left\{ \|R_{\beta(\delta,h^\delta)}h^\delta - T^\dagger h\| \;\middle|\; h^\delta \in L^2(\mathbb{R}^p),\ \|h^\delta - h\| \leq \delta \right\} \longrightarrow 0 \quad \text{as} \quad \beta \downarrow 0.$$

Recall at last that, if $R_\beta h \to T^\dagger h$ for every $h \in \mathscr{D}(T^\dagger)$ then (R_β) is a regularization of $T^\dagger$. See Engl et al. [1996, Proposition 3.4].

In the next theorem, we establish our basic consistency result. As we shall see, consistency is granted only in the case where the *source* $h^\dagger$ satisfies a Sobolev smoothness condition. The proof of the next theorem mostly follows most of the arguments of Bonnefond and Maréchal [2009]. It will be provided in the appendix.

Theorem 1. Assume that $H = L^2(V)$ with V bounded and that T is injective. Assume in addition that

$$|1 - \hat{\varphi}(\xi)| \sim c|\xi|^s \quad \text{as} \quad \xi \to 0 \tag{10}$$

where c, s are positive constants, and that $\hat{\varphi}(\xi) < 1$ for every $\xi \neq 0$. Let $r \in T\big(L^2(V) \cap H^s(\mathbb{R}^p)\big)$ and let h_β be the minimizer of Problem $(\mathscr{P}_\beta)$ with $H = L^2(V)$.

Then $h_\beta \to h^\dagger$ in $L^2(\mathbb{R}^p)$ as $\beta \downarrow 0$.

It is easy to manufacture kernels φ satisfying Condition (10). In particular, the Lévy kernels, defined by

$$\hat{\varphi}(\xi) = e^{-|\xi|^s}, \quad \xi \in \mathbb{R}^d, \tag{11}$$

obviously satisfy (10). Recall that, for $s \in (0,2]$, the kernel defined by (11) is positive, of class $\mathscr{C}^\infty$ and isotropic, which are clearly desirable properties for a mollifier[1]. We now state and prove a consistency theorem that accounts for the fact that our instrumental regression operator T is itself the result of an estimation.

Theorem 2. Assume that $H = L^2(V)$ with V bounded and that T is injective. Assume in addition that Condition (10) is satisfied, and that $\hat{\varphi}(\xi) < 1$ for every $\xi \neq 0$. Let $r \in T\big(L^2(V) \cap H^s(\mathbb{R}^p)\big)$, let $h^\dagger = T^\dagger r$ and, for every $n \in \mathbb{N}$, let T_n and r_n respectively denote an injective estimate of T and an estimate of r such that

$$E\left[\|T_n h^\dagger - T h^\dagger\|^2_{L^2(\mathbb{R}^q)}\right] \to 0, \quad E\left[\|r_n - r\|^2_{L^2(\mathbb{R}^q)}\right] \to 0 \quad \text{as} \quad n \to \infty.$$

Let

$$h_{\beta,n} := \operatorname*{argmin}_{h\in L^2(V)} \|T_n h - r_n\|^2_{L^2(\mathbb{R}^q)} + \|(I - C_\beta)h\|^2_{L^2(\mathbb{R}^p)} \tag{12}$$

$$h^\dagger_{\beta,n} := \operatorname*{argmin}_{h\in L^2(V)} \|T_n h - T_n h^\dagger\|^2_{L^2(\mathbb{R}^q)} + \|(I - C_\beta)h\|^2_{L^2(\mathbb{R}^p)} \tag{13}$$

Then, the regularization error satisfies:

$$E\left[\|h^\dagger - h^\dagger_{\beta,n}\|^2_{L^2(\mathbb{R}^p)}\right] \to 0 \quad \text{as} \quad \beta \downarrow 0. \tag{14}$$

Moreover, the estimation error satisfies:

$$E\left[\|h^\dagger_{\beta,n} - h_{\beta,n}\|^2_{L^2(\mathbb{R}^p)}\right] \leq \frac{1}{C m_\beta}\Big\{E\big[\|T_n h^\dagger - T h^\dagger\|^2_{L^2(\mathbb{R}^q)}\big] + E\big[\|r - r_n\|^2_{L^2(\mathbb{R}^q)}\big]\Big\}, \tag{15}$$

where C is a constant and $m_\beta = \mathcal{O}\left(\beta^{2s}\right)$ as $\beta \to 0$ is defined in Lemma 1.

[1] Recall that $s = 2$ corresponds to a Gaussian kernel and $s = 1$ corresponds to a Cauchy kernel.

106

For statistical inverse problems, a parameter choice rule is an application $\beta(n)$ such that the regularized solution converges to the true solution as the size of the sample goes to infinity. An immediate consequence of the previous theorem is the following:

Corollary 1. Given estimates on the noise levels $E\left[\|T_n h^\dagger - T h^\dagger\|^2_{L^2(\mathbb{R}^q)}\right]$ and $E\left[\|r - r_n\|^2_{L^2(\mathbb{R}^q)}\right]$ in function of n, we can define explicit apriori parameter choice rule $\beta(n) \to 0$ as $n \to \infty$ such that

$$E\left[\|h^\dagger - h_{\beta(n),n}\|^2_{L^2(\mathbb{R}^p)}\right] \to 0, \quad \text{as} \quad n \to \infty. \tag{16}$$

5 Simulations

In this section, we describe our simulation setting, and proceed to assess the performances of mollification in comparison with other methods. In particular, mollification will be compared to Tikhonov methods and spectral cut-off. For Tikhonov, we consider two versions: the ordinary Tikhonov (i.e. $Q = I$) and a second form where Q is the second order differential operator.

5.1 Discretization

For the sake of numerical simplicity, we set $p = q = 1$, and assume that Z take its values in $[0, 1]$. We denote by $(Z_l, W_l, Y_l)_{l=1,...,n}$ the observed sample.

Following Johannes et al. [2013], we turn to finite dimensions via a Galerkin projection method: h and r are projected onto finite dimensional orthonormal subspaces

$$\text{span}\{\phi_1, ..., \phi_I\} \quad \text{and} \quad \text{span}\{\psi_1, ..., \psi_J\},$$

respectively. The projections of h and r are respectively given by

$$h_{\text{proj}} = \sum_{i=1}^{I} \mathrm{h}_i \phi_i \quad \text{and} \quad r_{\text{proj}} = \sum_{j=1}^{J} \mathrm{r}_j \psi_j,$$

in which $\mathrm{h}_i = \langle h, \phi_i \rangle$, for $i = 1, ..., I$ and $\mathrm{r}_j = \langle r, \psi_j \rangle$ for $j = 1, ..., J$. Here, $\langle \cdot, \cdot \rangle$ denotes the standard inner product of $L^2([0, 1])$. The equation $T h_{\text{proj}} = r_{\text{proj}}$ reads:

$$\sum_{i=1}^{I} \mathrm{h}_i T\phi_i = \sum_{j=1}^{J} \mathrm{r}_j \psi_j,$$

which, by the orthonormality of the family $\{\psi_j\}_{j=1,...,J}$, is equivalent to

$$\sum_{i=1}^{I} \mathrm{h}_i \langle T\phi_i, \psi_j \rangle = \mathrm{r}_j, \quad j \in \{1, ..., J\}.$$

The matrix formulation of the equation $T h_{\text{proj}} = r_{\text{proj}}$ is written as $\mathrm{Th} = \mathrm{r}$, where T is a $J \times I$ real matrix given by

$$\mathrm{T}_{ji} = \langle T\phi_i, \psi_j \rangle = \mathbb{E}[\phi_i(Z)\psi_j(W)], \quad \forall (i, j) \in \{1, ..., I\} \times \{1, ..., J\},$$

$\mathrm{r} = (\mathrm{r}_1, ..., \mathrm{r}_J)^\top$ is a J-column vector defined by

$$\mathrm{r}_j = \langle r, \psi_j \rangle = \mathbb{E}[Y\psi_j(W)], \quad \forall j \in \{1, ..., J\},$$

and $\mathrm{h} = (\mathrm{h}_1, ..., \mathrm{h}_I)^\top$ is the unknown vector. These expectations are then estimated by the empirical means

$$\mathrm{T}_{n,ji} := \frac{1}{n} \sum_{l=1}^{n} \phi_i(Z_l)\psi_j(W_l) \tag{17}$$

and

$$\mathrm{r}_{n,j} := \frac{1}{n} \sum_{l=1}^{n} Y_l \psi_j(W_l). \tag{18}$$

The problem $Th = r$ is then approximated by $\mathrm{T}_n\mathrm{h} = \mathrm{r}_n$.

Remark 3. The above projection-based estimation yields a linear system of dimension I, whereas in the kernel-based estimation proposed in Darolles et al. [2011], the linear system is of dimension n, which may be much larger in practice. This practical point is in favor for the projection-based approach. ∎

In our simulations, we use bases of gate functions: given I, J, we let

$$\phi_i := \sqrt{I} \cdot \mathbb{1}_{I_i} \quad \text{and} \quad \psi_j := \sqrt{J} \cdot \mathbb{1}_{I_j} \tag{19}$$

with $I_i = [(i-1)/I, i/I]$ and $I_j = [(j-1)/J, j/J]$.

5.2 Simulation design

The function φ that generates our approximate unity is taken to be the Gaussian function

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \tag{20}$$

Its Fourier transform is given by $\hat{\varphi}(\xi) = \exp(-2\pi^2\xi^2)$. It clearly belongs to the family of Lévy kernels with $s = 2$ (see (10)).

We also consider two regression functions on $[0, 1]$: a smooth function h_1 defined by

$$h_1(x) = f_{(0.4,0.1)}(x) + f_{(0.65,0.075)}(x),$$

where $f_{(\mu,\sigma)}$ is the density function of the Gaussian of mean μ and standard deviation σ; and a less smooth function h_2 with discontinuous derivative defined by

$$h_2(x) = \exp(-5|x - 0.5|).$$

Our objective is to check the reliability of the mollification for less regular functions.

At last, the data $(Z_l, W_l, Y_l)_{l=1,...,n}$ are generated as follows:

$$\begin{aligned} W_l &\sim \mathcal{U}(-0.5, 0.5) \\ E_l &\sim \mathcal{U}(-0.5, 0.5) \\ Z_l &= 0.25 + 0.5\, W_l + 0.5\, E_l \\ Y_l &= h_i(Z_l) + E_l, \quad i = 1, 2 \end{aligned}$$

The sample size n is 2000 and the sizes I and J of the gate function bases are both equal to 100.

5.3 Regularization parameter selection

The regularized estimated solution depends on a regularization parameter that needs to be fixed. In the following, we adopted two different strategies:

- For the comparison of the regularization methods, we first consider an optimal (and theoretical) choice of the regularization parameter. More precisely, given an approximate solution h_α^{reg} obtained from a generic regularization method reg with regularization parameter α, we compute the absolute error norm

$$\|h^\dagger - h_\alpha^{\text{reg}}\|_{L^2([0,1])},$$

  and the parameter α_{opt} is defined as

$$\alpha_{\text{opt}} := \underset{\alpha>0}{\text{argmin}} \|h^\dagger - h_\alpha^{reg}\|_{L^2([0,1])}.$$

  Indeed, in a simulation setting, we know the true regression function $h^\dagger$ and such an optimal regularization parameter is numerically computable. It allows to compare the best approximate solutions obtained from each regularization method.

- However, since in practical situation, we never know the function $h^\dagger$, we need to design a practical parameter selection rule. Several data driven methods exist among which, the L-curve method introduced by Hansen Hansen [1992], Hansen and O'Leary [1993], the Quasi-optimality method (Leonov [1978], Bauer [2007], Bauer and Kindermann [2008], Bauer and Reiß [2008], Bauer and Kindermann [2009]), the generalized cross validation (see, e.g. Wahba [1977], Lukas [1993]) the Reginska rule Regińska [1996] and other heuristic parameter selection rules (see, e.g. Engl et al. [1996, Section 4.5]). The results we present below were produced using the discrete quasi-optimality method.

5.4 Results and comments

The results of our simulations are given in the following graphs and tables:

- In Figure 1, we display an example of sample data $(Z_l, Y_l)_{l=1,\dots,n}$ considered for the functions h_1 and h_2.
- In Figures 2 and 3, we compare the best approximate solutions of each regularization method, for both h_1 and h_2.
- In Figures 4 and 5, we compare the absolute L^2-error norm versus the regularization parameter for the four regularization methods, for both h_1 and h_2.
- Finally, in Figures 6 and 7 and Tables 9 and 8, we display the results of a Monte Carlo simulation with 1000 replications. Figures 6 and 7 were obtained with a sample size equal to 2000 and the x-axis corresponds to the absolute L^2-error and each point represents the best absolute error for a single replication. For both tables 9 and 8, the sample size takes the values 1000, 2000 and 3000.

In all the figures, we use the following abbreviations: `Mol` for mollification, `Tik` for ordinary Tikhonov, `Tik df` for variant of Tikhonov with second order differential

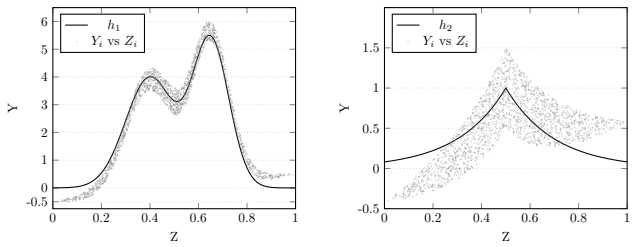

Figure 1: Sample data $(Z_l, Y_l)_{l=1,\dots,n}$ for the functions h_1 and h_2.

operator, `Tsvd` for spectral cut-off. At last, `Mol(QO)` stands for the mollification using discrete quasi optimality as parameter selection rule.

From Figures 2 and 3, we see that the mollification and the variant of Tikhonov estimators are quite close to the true function and behave very similarly. On the contrary, Tikhonov and Spectral cut-off are far from the true function and present a lot of oscillations. A similar comment can be done when looking at the absolute error graphs 4 and 5. We see again that Tikhonov and spectral cut-off perform poorly, compared to mollification and the variant of Tikhonov.

We can also note that the mollified estimator behaves well whatever the smoothness of the true solution (h_1 or h_2). In addition, unlike the variant of Tikhonov, the mollification does not need the true solution to be in the Soboloev space H^2. More precisely, for the mollification, the smoothness condition $h^\dagger \in H^s$ was needed only to derive consistency results when $H = L^2(V)$ (whereas the condition $h^\dagger \in H^2$ is necessary for the variant of Tikhonov). Note also that the implementation of the variant of Tikhonov entails applying the discrete second derivative (unbounded) operator D which may induce some extra computational difficulties.

The results of the Monte Carlo simulations yields similar conclusions with an obvious superiority of mollification and the variant of Tikhonov with respect to Tikhonov and the spectral cut-off. We can note that, for all regularization methods except the spectral cutoff, the empirical error decreases as n increases and so does the selected regularization parameter. There is also empirical evidence of the convergence of the discrete quasi-optimality selection rule for mollification.

From Figures 2, 3, 4, 5,7 and Tables 8 and 9, we can assert that the discrete quasi optimality rule provides very good approximate solutions, which indicates the practical applicability of mollification to real world problems.

6 Conclusion

In this paper, we introduced a mollification method to solve the ill-posedness of the instrumental regression. Mollification has never be used to solve the instrumental regression inverse problem. We proved the consistency of our estimator and discussed the regularization parameter selection. Using simulations, we compared the finite-sample performance of mollification with respect to the Tikhonov method, the Spectral cut-off and the generalized Tikhonov method with second order differential operator. Our simulations empirically demonstrated the outperformance of mollification with respect to the classical Tikhonov method and the spectral cutoff. Although the performance of mollification versus the generalized Tikhonov method with second order differential operator is empirically less notable, we stress that mollification is more flexible, since its definition does not require to work under

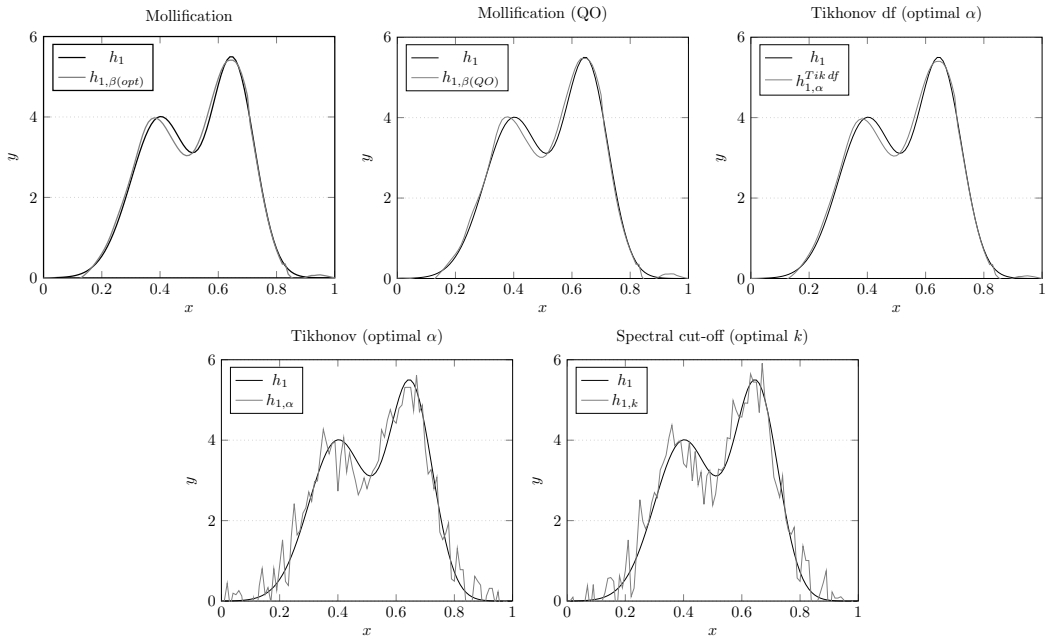

Figure 2: Best approximate solution for the four regularization methods considered with the function h_1.

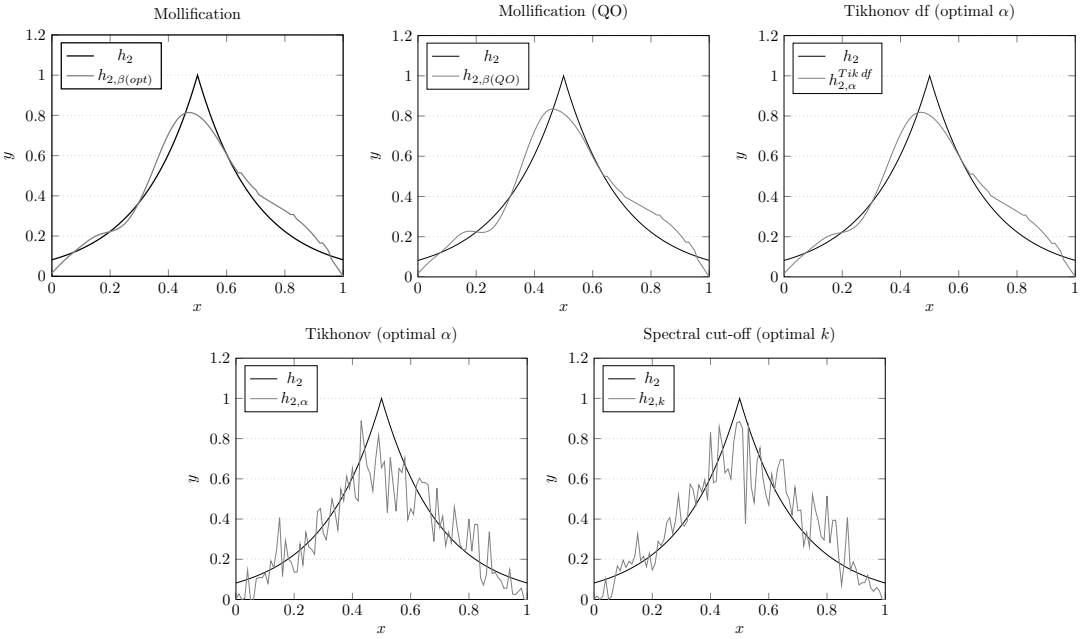

Figure 3: Best approximate solution for the four regularization methods considered with the function h_2.

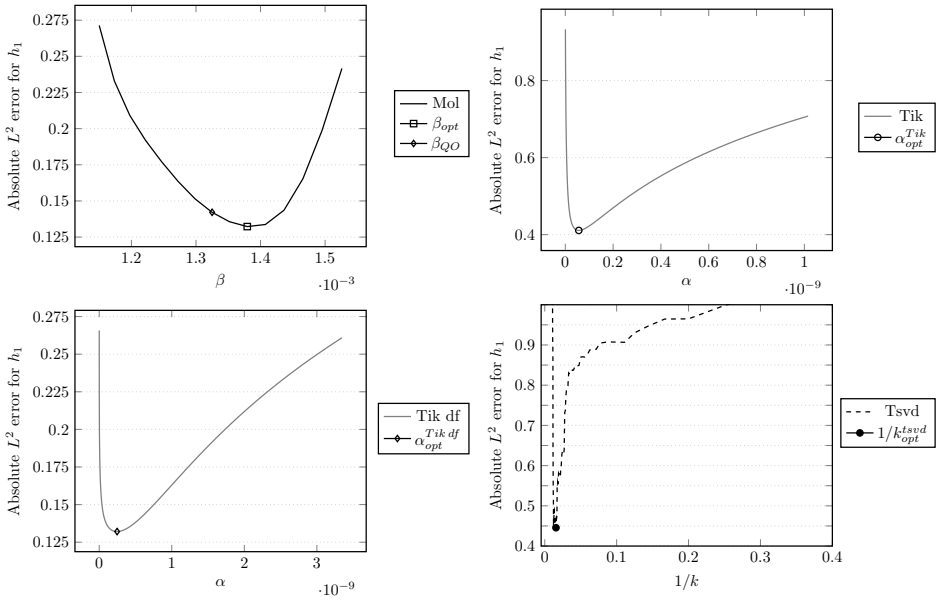

Figure 4: Absolute error with respect to regularization parameter for the four regularization methods considered with the function h_1.

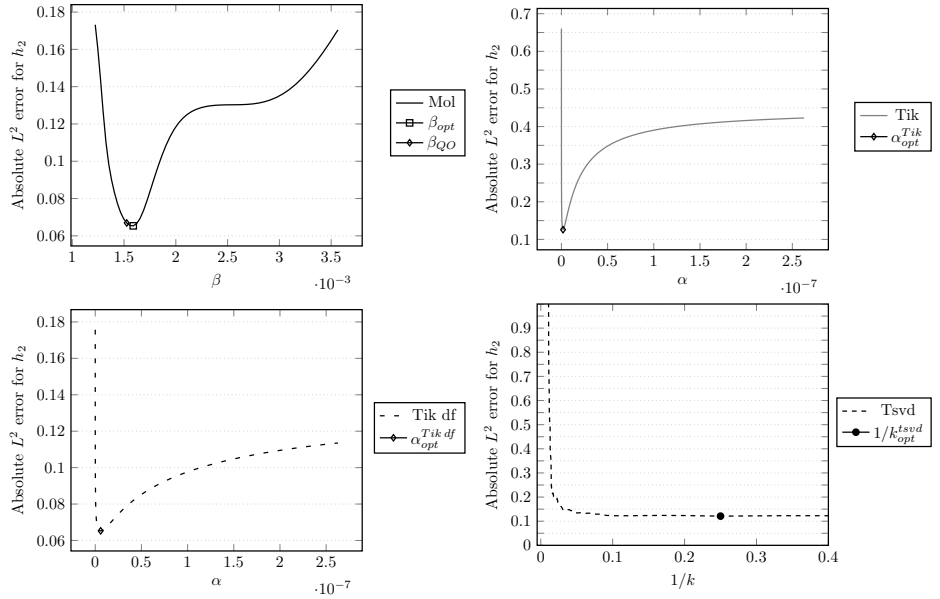

Figure 5: Absolute error with respect to regularization parameter for the four regularization methods considered with the function h_2.

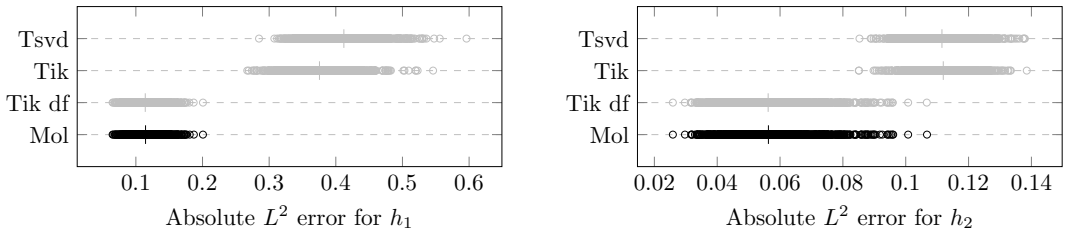

Figure 6: Results of Monte Carlo simulation for the functions h_1 and h_2: comparaison of the best absolute errors .

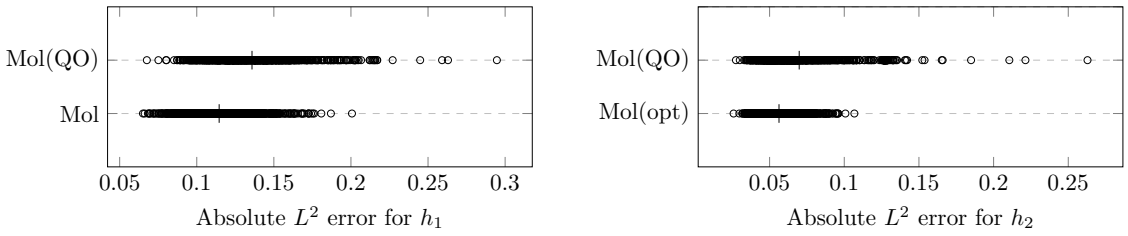

Figure 7: Results of Monte Carlo simulation for the functions h_1 and h_2: comparison of the absolute error using discrete quasi-optimality selection rule.

		Mol	Mol(QO)	Tik df	Tik	Tsvd
n=1000	$\bar{e}$	0.116694	0.138965	0.116273	0.418063	0.45311
	$\sigma(e)$	0.019953	0.026140	0.020034	0.048831	0.049612
	$\overline{\text{reg. par.}}$	0.001413	0.001337	4.44e-10	9.43e-11	0.013593
n=2000	$\bar{e}$	0.114526	0.13583	0.114116	0.37549	0.412087
	$\sigma(e)$	0.019719	0.026008	0.019742	0.041063	0.042657
	$\overline{\text{reg. par.}}$	0.001377	0.001307	2.21e-10	7.26e-11	0.014325
n=3000	$\bar{e}$	0.113155	0.133331	0.112746	0.353807	0.388567
	$\sigma(e)$	0.017537	0.022518	0.017592	0.037621	0.041197
	$\overline{\text{reg. par.}}$	0.001358	0.001291	1.48e-10	5.99e-11	0.014767

Figure 8: Summary Monte Carlo performances (for the function h_1) of each method over 1000 replications of sample size $n = 1000, 2000$ and 3000. $\bar{e}$ (resp. $\sigma(e)$) is the average (resp. standard deviation) of the best reconstruction error. $\overline{\text{reg.par.}}$ is the average of the optimal regularization parameter for each method, except Mol(QO) where $\overline{\text{reg.par.}}$ is the average of the regularization parameter chosen according to discrete quasi-optimality selection rule.

		Mol	Mol(QO)	Tik df	Tik	Tsvd
n=1000	$\bar{e}$	0.064882	0.079426	0.064814	0.134157	0.137683
	$\sigma(e)$	0.013473	0.029079	0.013490	0.009797	0.01110
	$\overline{\text{reg. par.}}$	0.001622	0.001597	1.39e-08	1.80e-09	0.061599
n=2000	$\bar{e}$	0.056298	0.069826	0.056215	0.11195	0.111528
	$\sigma(e)$	0.012569	0.026433	0.012582	0.007699	0.008742
	$\overline{\text{reg. par.}}$	0.001562	0.001542	6.06e-09	1.36e-09	0.106781
n=3000	$\bar{e}$	0.053091	0.066501	0.052991	0.101806	0.101693
	$\sigma(e)$	0.012560	0.026211	0.012580	0.007537	0.007779
	$\overline{\text{reg. par.}}$	0.001530	0.001512	3.80e-09	1.12e-09	0.126470

Figure 9: Summary Monte Carlo performances (for the function h_2) of each method over 1000 replications of sample size $n = 1000, 2000$ and 3000. $\bar{e}$ (resp. $\sigma(e)$) is the average (resp. standard deviation) of the best reconstruction error. $\overline{\text{reg.par.}}$ is the average of the optimal regularization parameter for each method, except Mol(QO) where $\overline{\text{reg.par.}}$ is the average of the regularization parameter chosen according to discrete quasi-optimality selection rule..

Sobolev smoothness.

A Proofs of the consistency results

Before tackling the proofs of Theorems 1 and 2, we state some technical results, which can be essentially found in Alibaud et al. [2009], Bonnefond and Maréchal [2009].

Lemma 1. Let φ satisfy the assumptions of Theorem 1. Let

$$m_\beta := \min_{|\xi|=1} \left|1-\hat{\varphi}(\beta\xi)\right|^2 \quad \text{and} \quad M_\beta := \max_{|\xi|=1} \left|1-\hat{\varphi}(\beta\xi)\right|^2. \tag{21}$$

Then:

(i) For all $\beta > 0$, $0 < m_\beta \leq M_\beta \leq \left(1+\|\varphi\|_{L^1(\mathbb{R}^d)}\right)^2$;

(ii) $\sup_{\beta>0}(M_\beta/m_\beta) < \infty$ and M_β tends to zero as $\beta \downarrow 0$;

(iii) there exist positive constants ν_0 and C_0 such that, for every $\beta \in (0,1]$ and every $\xi \in \mathbb{R}^d \setminus \{0\}$,

$$\nu_0 \left(|\xi|^{2s} \mathbb{1}_{B_{1/\beta}}(\xi) + \frac{1}{M_\beta} \mathbb{1}_{B^c_{1/\beta}}(\xi) \right) \leq \frac{|1-\hat{\varphi}(\beta\xi)|^2}{|1-\hat{\varphi}(\beta\xi/|\xi|)|^2} \leq C_0 |\xi|^{2s}.$$

Proof. See Alibaud et al. [2009, Lemme 12]. ∎ □

Proposition 1. Let φ be as in the previous lemma, and let $h \in L^2(V)$. Then, there exists a positive constant C_1 such that

$$\forall \beta \in (0,1], \quad C_1 m_\beta \|h\|^2_{L^2(\mathbb{R}^p)} \leq \|(I-C_\beta)h\|^2_{L^2(\mathbb{R}^p)}. \tag{22}$$

In the case where $h \in H^s(\mathbb{R}^p)$, there exists a positive constant C_0 such that

$$\forall \beta \in (0,1], \quad \|(I-C_\beta)h\|^2_{L^2(\mathbb{R}^p)} \leq C_0 M_\beta \|h\|^2_{H^s(\mathbb{R}^p)}. \tag{23}$$

Proof. The first estimate (22) is implicit in the proof of Theorem 4.1 in Bonnefond and Maréchal [2009], while the second is Proposition 4.3 in the same reference. We prove the first estimate of the sake of completeness. For $\beta > 0$, we have:

$$\begin{aligned}
\|(I - C_\beta)h\|^2_{L^2(\mathbb{R}^p)} &= \int_{\mathbb{R}^p} \left|1 - \hat{\varphi}(\beta\xi/|\xi|)\right|^2 \frac{|1-\hat{\varphi}(\beta\xi)|^2}{|1-\hat{\varphi}(\beta\xi/|\xi|)|^2} |\hat{h}(\xi)|^2 \, \mathrm{d}\xi \\
&\geq m_\beta \nu_0 \int_{\mathbb{R}^p} \left(|\xi|^{2s} \mathbb{1}_{\{|\xi|\leq 1/\beta\}}(\xi) + \frac{1}{M_\beta} \mathbb{1}_{\{|\xi|>1/\beta\}}(\xi) \right) |\hat{h}(\xi)|^2 \, \mathrm{d}\xi \\
&\geq m_\beta \nu_0 \int_{\mathbb{R}^p} \left(\mathbb{1}_{\{1<|\xi|\leq 1/\beta\}}(\xi) + \frac{1}{M_\beta} \mathbb{1}_{\{|\xi|>1/\beta\}}(\xi) \right) |\hat{h}(\xi)|^2 \, \mathrm{d}\xi \\
&\geq m_\beta \nu_0 (1 + \|\varphi\|_{L^1})^{-2} \int_{\mathbb{R}^p} \mathbb{1}_{\{|\xi|>1\}}(\xi) |\hat{h}(\xi)|^2 \, \mathrm{d}\xi.
\end{aligned}$$

In the above, the first equality is due to Parseval's theorem, the first inequality stems from Lemma 1 (iii), and the third inequality stems from Lemma 1 (i). Now, the last integral is equal to the L^2-norm of $T_{\{|\xi|>1\}}h$, in which $T_{\{|\xi|>1\}} \colon L^2(V) \to L^2(\{|\xi| > 1\})$ denotes the *Fourier truncated operator* with truncation to the complement of the unit ball in $\mathbb{R}^p$. It is well-known that such operators admit bounded inverses (see e.g. Alibaud et al. [2009], Proposition 5). It follows that

$$\|h\|^2_{L^2(\mathbb{R}^p)} = \left\| T^{-1}_{\{|\xi|>1\}} T_{\{|\xi|>1\}} h \right\|^2_{L^2(\mathbb{R}^p)} \leq \left\| T^{-1}_{\{|\xi|>1\}} \right\|^2 \|T_{\{|\xi|>1\}} h\|^2_{L^2(\mathbb{R}^p)}.$$

Putting things together, we get the desired estimate with

$$C_1 = \frac{\nu_0 (1 + \|\varphi\|_{L^1})^2}{\|T^{-1}_{\{|\xi|>1\}}\|^2}. \blacksquare$$

□

Proof of Theorem 1. This is a mere adaptation of Theorem 4.1 (II) in Bonnefond and Maréchal [2009]. In the present context, there is no interwining operator Φ_β (see Remark 2). Notice also that the compactness of T in Bonnefond and Maréchal [2009] was useful only for obtaining the latter intertwining operator. ∎ □

Proof of Theorem 2. In this proof, in order to shorten the notation, $\|(I-C_\beta)h\|_{L^2(\mathbb{R}^p)}$ will be denoted by $\|h\|_\beta$. We recall that this actually defines a norm on $L^2(V)$ equivalent to the L^2 norm (see, e.g. [Alibaud et al., 2009, Lemma 7]).

For a fixed $n \in \mathbb{N}$ and a fixed $\omega \in \Omega$[2] (T_n and r_n being random variables), from Theorem 1 with T replaced by T_n, we get:

$$\left\| h^\dagger - h^2_{\beta,n} \right\|^2_{L^2(\mathbb{R}^p)} \to 0 \quad \text{as} \quad \beta \downarrow 0. \tag{24}$$

Now, using the optimality of $h^\dagger_{\beta,n}$ and Proposition 1, we see that

$$\|h^\dagger_{\beta,n}\|^2_\beta \leq \|T_n h^\dagger_{\beta,n} - T_n h^\dagger\|^2_{L^2(\mathbb{R}^q)} + \|h^\dagger_{\beta,n}\|^2_\beta \leq \|h^\dagger\|^2_\beta \leq C_0 M_\beta \|h^\dagger\|^2_{H^s(\mathbb{R}^p)}.$$

By Proposition 1 again, we obtain the L^2-bound

$$\|h^\dagger_{\beta,n}\|^2_{L^2(\mathbb{R}^p)} \leq \frac{C_0 M_\beta}{C_1 M_\beta} \|h^\dagger\|^2_{H^s(\mathbb{R}^p)} \tag{25}$$

[2]For sake of simplicity, in order to lighten the equations, we omit the notation ω in the following

Since the ratio M_β/m_β is bounded (see Lemma 1 (ii)), we see that $\|h^\dagger - h^\dagger_{\beta,n}\|^2_{L^2(\mathbb{R}^p)}$ is bounded independently of β and n. Now, regarding $X_n := \|h^\dagger - h^\dagger_{\beta,n}\|^2_{L^2(\mathbb{R}^p)}$ as a random variable $X_n(\omega), \omega \in \Omega$ (through its dependence in n), we can apply the Dominated Convergence Theorem to obtain:

$$E\left[\|h^\dagger - h^\dagger_{\beta,n}\|^2_{L^2(\mathbb{R}^p)}\right] \to 0 \quad \text{as} \quad \beta \downarrow 0.$$

This proves (14). Let us now prove (15). By the linearity of the mapping $r \to h_\beta$ with h_β being the minimizer of Problem $(\mathscr{P}_\beta)$ with $H = L^2(V)$, we have:

$$h^\dagger_{\beta,n} - h_{\beta,n} = \underset{h \in L^2(V)}{\operatorname{argmin}} \|T_n h - (T_n h^\dagger - r_n)\|^2_{L^2(\mathbb{R}^q)} + \|(I - C_\beta)h\|^2_{L^2(\mathbb{R}^p)}.$$

Therefore, for every $h \in L^2(V)$,

$$\|T_n(h^\dagger_{\beta,n} - h_{\beta,n}) - (T_n h^\dagger - r_n)\|^2_{L^2(\mathbb{R}^q)} + \|h^\dagger_{\beta,n} - h_{\beta,n}\|^2_\beta \leq \|T_n(h) - (T_n h^\dagger - r_n)\|^2_{L^2(\mathbb{R}^q)} + \|h\|^2_\beta.$$

Taking $h = 0$ yields the inequality

$$\|h^\dagger_{\beta,n} - h_{\beta,n}\|^2_\beta \leq \|T_n h^\dagger - r_n\|^2.$$

By the triangle inequality

$$\|h^\dagger_{\beta,n} - h_{\beta,n}\|^2_\beta \leq 2\left(\|T_n h^\dagger - T h^\dagger\|^2_{L^2(\mathbb{R}^q)} + \|T h^\dagger - r_n\|^2_{L^2(\mathbb{R}^q)}\right)$$

Now, by Proposition 1, $\|h^\dagger_{\beta,n} - h_{\beta,n}\|^2_\beta \geq C_1 m_\beta \|h^\dagger_{\beta,n} - h_{\beta,n}\|^2_{L^2(\mathbb{R}^p)}$, and (15) follows. ∎

□

References

Nathaël Alibaud, Pierre Maréchal, and Yaowaluk Saesor. A variational approach to the inversion of truncated fourier operators. *Inverse Problems*, 25(4):045002, 2009.

Frank Bauer. Some considerations concerning regularization and parameter choice algorithms. *Inverse Problems*, 23(2):837, 2007.

Frank Bauer and Stefan Kindermann. The quasi-optimality criterion for classical inverse problems. *Inverse Problems*, 24(3):035002, 2008.

Frank Bauer and Stefan Kindermann. Recent results on the quasi-optimality principle. *Journal of Inverse and Ill-posed Problems*, 17(1):5–18, 2009.

Frank Bauer and Markus Reiß. Regularization independent of the noise level: an analysis of quasi-optimality. *Inverse Problems*, 24(5):055009, 2008.

Xavier Bonnefond and Pierre Maréchal. A variational approach to the inversion of some compact operators. *Pacific journal of optimization*, 5(1):97–110, 2009.

Marine Carrasco and Jean-Pierre Florens. Generalization of gmm to a continuum of moment conditions. *Econometric Theory*, pages 797–834, 2000.

Marine Carrasco and Jean-Pierre Florens. A spectral method for deconvolving a density. *Econometric Theory*, 27(3):546–581, 2011.

Marine Carrasco, Jean-Pierre Florens, and Eric Renault. Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of econometrics*, 6:5633–5751, 2007.

Marine Carrasco, Jean-Pierre Florens, and Eric Renault. Asymptotic normal inference in linear inverse problems. *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, 73(74):140, 2014.

Raymond J Carroll and Peter Hall. Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186, 1988.

Fabienne Comte, Yves Rozenholc, and Marie-Luce Taupin. Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics*, 34(3): 431–452, 2006.

Fabienne Comte, Yves Rozenholc, and M-L Taupin. Finite sample penalization in adaptive density deconvolution. *Journal of Statistical Computation and Simulation*, 77(11):977–1000, 2007.

Fabienne Comte, Jérôme Dedecker, and Marie-Luce Taupin. Adaptive density deconvolution with dependent inputs. *Mathematical methods of Statistics*, 17(2):87, 2008.

Serge Darolles, Yanqin Fan, Jean-Pierre Florens, and Eric Renault. Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565, 2011.

Andrea Enache and Jean-Pierre Florens. Identification and estimation in a third-price auction model. *Econometric Theory*, 2020.

Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

Jianqing Fan. Global behavior of deconvolution kernel estimates. *Statistica Sinica*, pages 541–551, 1991.

Jean-Paul Florens, James Heckman, Costas Meghir, and Edward Vytlacil. Instrumental variables, local instrumental variables and control functions. Technical report, cemmap working paper, 2002.

Jean-Pierre Florens and Erwann Sbaï. Local identification in empirical games of incomplete information. *Econometric Theory*, pages 1638–1662, 2010.

Jean-Pierre Florens and Anna Simoni. Nonparametric estimation of an instrumental regression: A quasi-bayesian approach based on regularized posterior. *Journal of Econometrics*, 170(2):458–475, 2012a.

Jean-Pierre Florens and Anna Simoni. Regularized posteriors in linear ill-posed inverse problems. *Scandinavian Journal of Statistics*, 39(2):214–235, 2012b.

Jean-Pierre Florens and Senay Sokullu. Nonparametric estimation of semiparametric transformation models. *Econometric Theory*, 33(4):839–873, 2017.

Jean-Pierre Florens, Jan Johannes, and Sébastien Van Bellegem. Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory*, pages 472–496, 2011.

Peter Hall, Joel L Horowitz, et al. Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, 33(6):2904–2929, 2005.

Per Christian Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM review*, 34(4):561–580, 1992.

Per Christian Hansen and Dianne Prost O'Leary. The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM journal on scientific computing*, 14(6):1487–1503, 1993.

Yingyao Hu and Ji-Liang Shiu. Nonparametric identification using instrumental variables: sufficient conditions for completeness. Technical report, Working paper, 2011.

Yingyao Hu, Susanne M Schennach, and Ji-Liang Shiu. Injectivity of a class of integral operators with compactly supported kernels. *Journal of Econometrics*, 200(1):48–58, 2017.

Jan Johannes. Deconvolution with unknown error distribution. *The Annals of Statistics*, 37(5A):2301–2323, 2009.

Jan Johannes, Sébastien Van Bellegem, and Anne Vanhems. Iterative regularisation in nonparametric instrumental regression. *Journal of Statistical Planning and Inference*, 143(1):24–39, 2013.

André Lannes, Sylvie Roques, and Marie-José Casanove. Stabilized reconstruction in signal and image processing: I. partial deconvolution and spectral extrapolation with limited field. *Journal of modern Optics*, 34(2):161–226, 1987.

Aleksandr Sergeevich Leonov. On the choice of regularization parameters by means of the quasi-optimality and ratio criteria. In *Doklady Akademii Nauk*, volume 240, pages 18–20. Russian Academy of Sciences, 1978.

John Locker and PM Prenter. Regularization with differential operators. i. general theory. *Journal of Mathematical analysis and applications*, 74(2):504–529, 1980.

Alfred K Louis and Peter Maass. A mollifier method for linear operator equations of the first kind. *Inverse problems*, 6(3):427, 1990.

Mark A Lukas. Asymptotic optimality of generalized cross-validation for choosing the regularization parameter. *Numerische Mathematik*, 66(1):41–66, 1993.

Pierre Maréchal, Dylan Togane, and Anna Celler. A new reconstruction methodology for computerized tomography: Frect (fourier regularized computed tomography). *IEEE Transactions on Nuclear Science*, 47(4):1595–1601, 2000.

Pierre Maréchal, Léopold Simar, and Anne Vanhems. A mollifier approach to the deconvolution of probability densities. part 1: the methodology and its comparison to classical methods. *TSE Working paper*, 2018.

Alexander Meister. On the effect of misspecifying the error density in a deconvolution problem. *Canadian Journal of Statistics*, 32(4):439–449, 2004.

Vladimir Alekseevich Morozov. *Methods for solving incorrectly posed problems*. Springer Science & Business Media, 2012.

Diego A Murio. *The mollification method and the numerical solution of ill-posed problems*. John Wiley & Sons, 2011.

M Nair, Markus Hegland, and Robert Anderssen. The trade-off between regularity and stability in tikhonov regularization. *Mathematics of Computation*, 66(217): 193–206, 1997.

Teresa Regińska. A regularization parameter in discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 17(3):740–749, 1996.

Thomas Schuster. *The method of approximate inverse: theory and applications*, volume 1906. Springer, 2007.

Leonard A Stefanski and Raymond J Carroll. Deconvolving kernel density estimators. *Statistics*, 21(2):169–184, 1990.

Manfred R Trummer. A method for solving ill-posed linear operator equations. *SIAM journal on numerical analysis*, 21(4):729–737, 1984.

Grace Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM Journal on Numerical Analysis*, 14(4):651–667, 1977.

Chapter 5

Conclusion

Dans cette thèse, nous nous sommes concentrés sur la régularisation de problèmes linéaires inverses mal-posés avec une attention particulière à la formulation variationnelle de la mollification introduite à la fin des années 80 par Lannes et al. [71].

Dans le chapitre 2, en utilisant la dualité de Lagrange, nous avons proposé une méthode de calcul du paramètre de régularisation correspondant au principe de Morozov pour les méthodes de régularisation de type Tikhonov. Nous avons prouvé que l'application de la méthode est assez simple et avons décrit un algorithme Quasi-Newton unidimensionnel combiné à une recherche de ligne Wolfe-Lemaréchal pour le calcul du paramètre de régularisation. Un exemple illustratif sur le problème de la régression instrumentale non-paramétrique atteste de l'efficacité de l'algorithme.

Dans le chapitre 3, nous avons fourni une nouvelle méthode de régularisation adaptée aux problèmes linéaires exponentiellement mal-posés. Contrairement à la mollification, cette nouvelle méthode fait partie de la famille des méthodes de régularisation définies via des fonctions index. Nous avons effectué une analyse des taux de convergence de la méthode et avons montré que la nouvelle méthode est d'ordre optimale sous des conditions sources logarithmiques et d'ordre quasi-optimale sous des conditions sources générales. Inspirés par le principe de Morozov, nous avons fourni une règle de choix de paramètres a-posteriori donnant des taux de convergence d'ordre optimal sous des conditions sources logarithmiques. En outre, nous avons comparé numériquement la nouvelle méthode aux méthodes de régularisation classiques telles que la méthode de Tikhonov, la régularisation asymptotique, la *spectral cut-off* et la méthode des gradients conjugués pour la régularisation de trois problèmes tests mal-posés tirés de la littérature. Nous avons appliqué cinq règles de sélection heuristique de paramètre à chacune des méthodes de régularisation et les résultats des simulations ont montré que la nouvelle méthode est au moins aussi efficace que les autres méthodes de régularisation citées précédemment. Pour le problème test exponentiellement mal-posé considéré, les résultats montrent que la nouvelle méthode produit les plus petites erreurs. Enfin, les simulations ont révélé que la variante de la méthode *L-curve* définie dans la section 6 du chapitre 3 pourrait en fait être une très bonne règle heuristique de choix de paramètres, étant donné son efficacité pour toutes les méthodes de régularisation considérées.

Au chapitre 4, nous avons appliqué l'approche variationnelle de la mollification au problème de régression instrumentale non-paramétrique qui consiste à récupérer la fonction de régression dans un modèle statistique où les variables explicatives sont liées aux erreurs de régression. Dans le cadre stochastique classique où les données et l'opérateur sont approximés, nous avons établi la cohérence

de notre estimateur. Nous avons comparé notre approche à celle de Tikhonov ordinaire et de Tikhonov généralisé avec un opérateur différentiel du second ordre sur le terme de pénalité. Nous avons effectué des simulations de Monte Carlo et les résultats obtenus attestent de l'efficacité de la nouvelle approche. Enfin, nous avons appliqué une règle heuristique de sélection de paramètres à la technique de régularisation proposée et les résultats illustrent clairement l'adéquation de notre approche combinée à la règle de choix de paramètres.

Une intéressante perspective aux travaux effectués dans cette thèse serait l'application de la mollification aux problèmes d'equations aux derivées partielles e.g. l'equation de Helmholtz, l'étude des taux de convergence de l'approche variationnelle de la mollification ainsi que l'étude de l'optimalité de cette approche de régularisation.

Acknowledgment

First of all, I am thankful the almighty Lord for giving me the health, patience and perseverance throughout this thesis.

I would like to express my gratitude to my supervisor, Prof. Pierre Maréchal and my co-supervisor Dr. Anne Vanhems for their kindness, help, support, discussion and encouragement throughout this research period.

I would like to offer my special appreciation and thanks to the CIMI committee for having granted me the PhD fellowship which allowed me to pursuit this thesis. My gratitude also goes to Université Paul Sabatier which allowed me to be a teaching assistant during the whole three years of my thesis. During this period, I enjoyed the teaching duties and I gained a good experience. I am grateful to the Institut de Mathématiques de Toulouse which provided me all the facilities needed to pursuit my research. I also thank the doctoral school EDMITT for all the trainings in teaching, entrepreneurship, career development and others that I have attended.

My great thanks are also addressed to my wife Amandine Malonguemfo. She has always been present during stressful periods of the thesis and she has always been there to cheer me up when necessary. Her advice and kindness greatly helped me repeatedly.

I am thankful to the members of my family who have always believed in me and whose support, care and love have always been unconditional.

Special thanks to my office mates Moctar Ndiaye, Michèle Romanos and my friend Kuntal Bhandari for all the happy times spent together during my thesis.

Finally, hoping that I have not forgotten someone, I want to thank all individuals who contributed in one way or another to the achievement of this thesis. Thank you very much !

Bibliography

[1] N. ALIBAUD, P. MARÉCHAL and Y. SAESOR, *A variational approach to the inversion of truncated Fourier operators*, Inverse Problems 25 (2009).

[2] D. M. ALLEN, *The relationship between variable selection and data augmentation and a method for prediction*, Technometrics 16 (1974), 125–127.

[3] R. S. ANDERSSEN and F. R. DE HOOG *Finite difference methods for the numerical differentiation of non-exact data*, Computing, 33 (1984), 259-267.

[4] G. ANGER, R GORENFLO, H. JOCHMANN, H. MORITZ and W. WEBERS, *Inverse problems: principles and applications in geophysics, technology, and medicine*, Akademie-Verlag, Berlin, 1993.

[5] A. B. BAKUSHINSKII, *The problem of the convergence of the iteratively regularized Gauss-Newton method*, Computational mathematics and mathematical physics, 32(1992), 1353-1359.

[6] A. B. BAKUŠINSKIĬ, *A general method of constructing regularizing algorithms for a linear incorrect equation in Hilbert space*, Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, 7(1967), 672-677.

[7] A. B. BAKUŠINSKIĬ, *Remarks on the choice of regularization parameter from quasioptimality and relation tests*, Zh. Vychisl. Mat. i Mat. Fiz. 24 (1984), 1258–1259.

[8] H. T. BANKS and K. KUNISCH, *Estimation techniques for distributed parameter systems*, Birkhäuser, Boston, 1989.

[9] F. BAUER, *Some considerations concerning regularization and parameter choice algorithms*, Inverse Problems 23(2007), 837–858.

[10] F. BAUER and S. KINDERMANN, *Recent results on the quasi-optimality principle*, J. Inverse Ill-Posed Probl. 17(2009), 5–18, .

[11] F. BAUER and S. KINDERMANN, *The quasi-optimality criterion for classical inverse problems*, Inverse Problems 24(2008).

[12] F. BAUER and M. REISS, *Regularization independent of the noise level: an analysis of quasi-optimality*, Inverse Problems 24(2008).

[13] F. S. V. BAZÁN, *Fixed-point iterations in determining the Tikhonov regularization parameter*, Inverse Problems 24(2008).

[14] F. S. V. BAZÁN and J. B. FRANCISCO, *An improved fixed-point algorithm for determining a Tikhonov regularization parameter*, Inverse Problems 25 (2009).

[15] B. BLASCHKE, A. NEUBAUER and O. SCHERZER, *On convergence rates for the iteratively regularized Gauss-Newton method*, IMA J. Numer. Anal. 17(1997), 421–436.

[16] X. BONNEFOND and P. MARÉCHAL, *A variational approach to the inversion of some compact operators*, Pac. J. Optim. 5 (2009), 97–110.

[17] X. BONNEFOND, P. MARÉCHAL and W. C. SIMO TAO LEE, *A note on the Morozov principle via Lagrange duality*, Set-Valued Var. Anal. 26 (2018), 265–275.

[18] J. M. BORWEIN, P. MARÉCHAL and D. NAUGLER, *A convex dual approach to the computation of NMR complex spectra*, Math. Methods Oper. Res. 51 (2000), 91–102.

[19] H. BRAKHAGE, *On ill-posed problems and the method of conjugate gradients*, In Inverse and ill-posed Problems, (1987), 165-175.

[20] H. BREZIS, *Functional analysis, Sobolev spaces and partial differential equations*, Springer Science & Business Media, 2010.

[21] M. CARRASCO, J.P. FLORENS and E. RENAULT, *Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization*, Handbook of econometrics 6(2007), 5633-5751.

[22] J. L. CASTELLANOS, S. GÓMEZ and V GUERRA, *The triangle method for finding the corner of the L-curve*, Appl. Numer. Math. 43(2002), 359–373.

[23] J. W. DANIEL, *The conjugate gradient method for linear and nonlinear operator equations* SIAM J. Numer. Anal. 4(1967), 10–26.

[24] S. DAROLLES, Y. FAN, J.P. FLORENS and E. RENAULT, *Nonparametric instrumental regression*, Econometrica, 79(2011), 1541-1565.

[25] R. DAUTRAY and J.-L. LIONS, *Mathematical analysis and numerical methods for science and technology. Vol. 3. Spectral theory and applications*, Springer-Verlag, Berlin, 1990.

[26] B. EICKE , A. K. LOUIS and R. PLATO, *The instability of some gradient methods for ill-posed problems*, Numer. Math., 58(1990), 129–134.

[27] Y. EIDELMAN, V. D. MILMAN, A. TSOLOMITIS, *Functional analysis: an introduction* American Mathematical Society, (Vol. 66), 2004.

[28] H. W. ENGL, *Discrepancy principles for Tikhonov regularization of ill-posed problems leading to optimal convergence rates*, J. Optim. Theory Appl. 52(1987), 209–215.

[29] H. W. ENGL and H. GFRERER, *A posteriori parameter choice for general regularization methods for solving linear ill-posed problems*, Appl. Numer. Math. 4 (1988), 395–417.

[30] H. W. ENGL and W. GREVER, *Using the L-curve for determining optimal regularization parameters*, Numer. Math. 69 (1994), 25–31.

[31] H. W. ENGL and C. W. GROETSCH, *Inverse and ill-posed problems*, Elsevier,(Vol. 4),2014.

[32] H. W. ENGL, M. HANKE and A. NEUBAUER, *Regularization of inverse problems*, Mathematics and its Applications, 375. Kluwer Academic Publishers Group, Dordrecht, 1996.

[33] H. W. ENGL, A. K. LOUIS and W. RUNDELL, *Inverse problems in medical imaging and nondestructive testing*, Springer-Verlag, 1996.

[34] J.P. FLORENS, J. JOHANNES and S. VAN BELLEGEM, *Identification and estimation by penalization in Nonparametric Instrumental Regression*, Econometric Theory, 27(2011), 472-496.

[35] J.P. FLORENS, J. JOHANNES and S. VAN BELLEGEM, *Instrumental regression in partially linear models*, The Econometrics Journal, 15(2012), 304-324.

[36] W. FREEDEN and F. SCHNEIDER, *Regularization wavelets and multiresolution* Inverse Problems 14 (1998), 225–243.

[37] V. FRIDMAN, *A method of successive approximations for Fredholm integral equations of the first kind*, Uspeki Mat. Nauk., 11(1956), 233–234.

[38] H. GFRERER, *An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates*, Math. Comp. 49 (1987), 507–522.

[39] D. A. GIRARD, *A fast "Monte Carlo cross-validatio" procedure for large least squares problems with noisy data*, Numer. Math. 56 (1989), 1–23.

[40] M. S. GOCKENBACH and E. GORGIN, *On the convergence of a heuristic parameter choice rule for Tikhonov regularization*, SIAM J. Sci. Comput. 40(2018), A2694–A2719.

[41] G. H. GOLUB, M. HEATH and G. WAHBA, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics 21 (1979), 215–223.

[42] D. GOURION and D. NOLL, *The inverse problem of emission tomography*, Inverse Problems 18(2002), 1435–60.

[43] C. W. GROETSCH, *Inverse problems in the mathematical sciences*, Vieweg Mathematics for Scientists and Engineers. Friedr. Vieweg & Sohn, Braunschweig, 1993.

[44] O. GÜLER, *On the convergence of the proximal point algorithm for convex minimization*, SIAM Journal on Control and Optimization, 29(1991), 403-419.

[45] O. GÜLER, *New proximal point algorithms for convex minimization*, SIAM Journal on Optimization, 2(1992), 649-664.

[46] U. HÄMARIK and T. RAUS, *On the choice of the regularization parameter in ill-posed problems with approximately given noise level of data*, J. Inverse Ill-Posed Probl. 14 (2006), 251–266.

[47] M. HANKE, *Limitations of the L-curve method in ill-posed problems*, BIT 36 (1996), 287–301.

[48] M. HANKE, *Accelerated Landweber iterations for the solution of ill-posed equations*, Numerische mathematik 60(1991), 341-373.

[49] M. HANKE, *Conjugate gradient type methods for ill-posed problems*, Routledge, 2017.

[50] M. HANKE and H ENGL, *An optimal stopping rule for the ν-method for solving ill-posed problems using Christoffel functions*, J. Approx. Theor., 79(1994), 89–108.

[51] M. HANKE and P. C. HANSEN, *Regularization methods for large-scale problems*, Surveys Math. Indust. 3 (1993), 253–315.

[52] M. HANKE and T. RAUS, *A general heuristic for choosing the regularization parameter in ill-posed problems*, SIAM J. Sci. Comput. 17 (1996), 956–972.

[53] P. C. HANSEN, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev. 34 (1992), 561–580.

[54] P. C. HANSEN, *Regularization Tools version 4.0 for Matlab 7.3*, Numer. Algorithms 46 (2007), no. 2, pp 189–194.

[55] P. C. HANSEN, T. K. JENSEN and G. RODRIGUEZ, *An adaptive pruning algorithm for the discrete L-curve criterion*, J. Comput. Appl. Math. 198 (2007), 483–492.

[56] P. C. HANSEN and D. P. O'LEARY , *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput. 14 (1993), 1487–1503.

[57] D. N. HÀO, H. J. REINHARDT and F. SEIFFARTH, *Stable numerical fractional differentiation by mollification*, Numer. Funct. Anal. Optim. 15 (1994), 635–659.

[58] B. HARRACH, *Uniqueness and Lipschitz stability in electrical impedance tomography with finitely many electrodes*, Inverse Problems 35 (2019).

[59] M. HEGLAND and R. S. ANDERSSEN *A mollification framework for improperly posed problem*, Numer. Math. 78 (1998), 549–575.

[60] G. HELMBERG, *Introduction to spectral theory in Hilbert space*, North-Holland Series, Amsterdam, 1969.

[61] Y. HENG, S. LU, A. MHAMDI and S. V. PEREVERZEV, *Model functions in the modified L-curve method-case study: the heat flux reconstruction in pool boiling*, Inverse Problems 26 (2010).

[62] M.R. HESTENES and E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Stand., 49(1952), 409–436.

[63] E. HILLE and R.S. PHILLIPS, *Functional Analysis and Semi-Groups*, American Mathematical Society, 1996.

[64] T. HOHAGE, *Logarithmic convergence rates of the iteratively regularized Gauss-Newton method for an inverse potential and an inverse scattering problem*, Inverse Problems 13 (1997), 1279–1299.

[65] V. ISAKOV, *On uniqueness in the inverse transmission scattering problem*, Comm. Partial Differential Equations 15 (1990), 1565–1587.

[66] S. KINDERMANN and A. NEUBAUER, *On the convergence of the quasi-optimality criterion for (iterated) Tikhonov regularization*, Inv. Prob. Imag. 2 (2008), 291–299.

[67] S. KINDERMANN and A. NEUBAUER, *On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization*, Inverse Problems & Imaging, 2(2008), 291-299.

[68] A. KIRSCH, *An Introduction to the Mathematical Theory of Inverse Problems*, second edition, Springer, 2011.

[69] R. L. LAGENDIJK and J. BIEMOND, *Iterative identification and restoration of images*, Springer Science & Business Media (Vol. 118), 2012.

[70] A. LANNES, E. ANTERRIEU and K. BOUYOUCEF, *Fourier interpolation and reconstruction via Shannon-type techniques: I. Regularization principle*, J. Modern Opt. 41(1994), 1537–1574.

[71] A. LANNES, S. ROQUES and M.-J. CASANOVE, *Stabilized reconstruction in signal and image processing; Part I: partial deconvolution and spectral extrapolation with limited field*, J. Mod. Opt. 34(1987), 161-226.

[72] L. LANDWEBER, *An iteration formula for Fredholm integral equations of the first kind*, Amer. J. Math 73(1951), 615-624.

[73] P. D. LAX and R. S. PHILLIPS, *Scattering theory*, Pure and Applied Mathematics, Vol. 26 Academic Press, New York-London 1967.

[74] A. S. LEONOV, *On the choice of regularization parameters by means of quasi-optimality and ratio criteria*, Soviet. Math. Dokl. 19 (1978).

[75] A.K. LOUIS, *Approximate inverse for linear and some nonlinear problems*, Inverse Problems 12(1996), 175-190.

[76] A.K. LOUIS and P. MAASS, *A mollifer method for linear operator equations of the first kind*, Inverse Problems 6(1990), 427-440.

[77] M. A. LUKAS, *Asymptotic optimality of generalized cross-validation for choosing the regularization parameter*, Numer. Math. 66(1993), 41–66.

[78] M. A. LUKAS, *Comparisons of parameter choice methods for regularization with discrete noisy data*, Inverse Problems 14 (1998), 161–184.

[79] B. A. MAIR, *Tikhonov regularization for finitely and infinitely smoothing operators*, SIAM J. Math. Anal. 25 (1994), 135–147.

[80] P. MANSELLI and K. MILLER, *Calculation of the surface temperature and heat flux on one side of a wall from measurements on the opposite side*, Ann. Mat. Pura Appl. 123 (1980), 161–183.

[81] P. MARCHAL,D. TOGANE and A. CELLERT, *A new reconstruction methodology for computerized tomography: FRECT (Fourier Regularized Computed Tomography)*, IEEE Transactions on Nuclear Science, 47(2000), 1595-1601.

[82] P. MARÉCHAL and A. LANNES *Unification of some deterministic and probabilistic methods for the solution of linear inverse problems via the principle of maximum entropy on the mean*, Inverse Problems 13 (1997), 135–151.

[83] P. MARÉCHAL, L. SIMAR and A. VANHEMS, *A mollifier approach to the deconvolution of probability densities*, **18-965**, Toulouse School of Economics (TSE), 2018.

[84] D. MARIANO-GOULART, P. MARÉCHAL, L. GIRAUD, S. GRATTON and M. FOURCADE, *A priori selection of the regularization parameters in emission tomography by Fourier synthesis*, Comput. Med. Imaging and Graph. 31(2007), 502-509.

[85] P. MATHÉ and S. V. PEREVERZEV, *The discretized discrepancy principle under general source conditions*, J. Complexity 22(2006), 371–381.

[86] P. MATHÉ and S. V. PEREVERZEV, *Discretization strategy for linear ill-posed problems in variable Hilbert scales*, Inverse Problems 19 (2003), 1263–1277

[87] V. A. MOROZOV, *Choice of parameter for the solution of functional equations by the regularization method*, Sov. Math. Doklady Vol. 8, 1967.

[88] V. A. MOROZOV, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl. 7(1966), 414–417.

[89] D. A. MURIO, *Automatic numerical differentiation by discrete mollification*, Comput. Math. Appl. 13 (1987), 381–386.

[90] D. A. MURIO, *The mollification method and the numerical solution of ill-posed problems*, A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1993.

[91] M. T. NAIR, E. SCHOCK and U. TAUTENHAHN, *Morozov's discrepancy principle under general source conditions*, Z. Anal. Anwendungen 22 (2003), 199–214.

[92] S.I. NAKAGIRI, *Review of Japanese work of the last ten years on identifiability in distributed parameter systems* Inverse Problems, 9(1993), 143–191, .

[93] A. S. NEMIROVSKĬI, *The regularizing properties of the adjoint gradient method in ill-posed problems*, USSR Computational Mathematics and Mathematical Physics, 26(1986), 7-16.

[94] A. NEUBAUER, *The convergence of a new heuristic parameter selection criterion for general regularization methods*, Inverse Problems 24 (2008).

[95] Y. NOTAY., *On the convergence rate of the conjugate gradients in the presence of rounding errors*, Numer. Math., 65(1993), 301–318.

[96] P. S. NOVIKOV, *On the uniqueness of a solution of the inverse problem of potential theory*, Dokl. Akad. Nauk SSSR , 18(1938), 165–168.

[97] L. E. PAYNE, *Improperly posed problems in partial differential equations*, SIAM, Philadelphia, 1975.

[98] S. V. PEREVERZEV and E. SCHOCK, *Morozov's discrepancy principle for Tikhonov regularization of severely ill-posed problems in finite-dimensional subspaces*, Numer. Funct. Anal. Optim. 21(2000), 901–916.

[99] J. QI-NIAN, *Applications of the modified discrepancy principle to Tikhonov regularization of nonlinear ill-posed problems*, SIAM J. Numer. Anal. 36 (1999), 475–490.

[100] T. REGIŃSKA, *A regularization parameter in discrete ill-posed problems*, SIAM J. Sci. Comput. 17(1996), 740–749.

[101] A. RIEDER and T. SCHUSTER, *The approximate inverse in action with an application to computerized tomography*, SIAM J. Numer. Anal. 37 (2000), 1909–1929.

[102] A. RIEDER and T. SCHUSTER, *The approximate inverse in action II: convergence and stability*, Mathematics of computation, 72(2003), 1399-1415.

[103] A. RIEDER and T. SCHUSTER, *The approximate inverse in action III: 3D-Doppler tomography*, Numerische Mathematik, 97(2004), 353-378.

[104] R. T. ROCKAFELLAR, *Monotone operators and the proximal point algorithm*, SIAM journal on control and optimization, 14(1976), 877-898.

[105] G. RODRIGUEZ and D. THEIS, *An algorithm for estimating the optimal regularization parameter by the L-curve*, Rend. Mat. Appl. 25 (2005), 69–84.

[106] R. RUMMEL and O. L. COLOMBO, *Gravity field determination from satellite gradiometry*, Bulletin géodésique, 59(1985), 233-246.

[107] O. SCHERZER, *The use of Morozov's discrepancy principle for Tikhonov regularization for solving nonlinear ill-posed problems*, Computing 51 (1993), 45–60

[108] T. SCHUSTER, *The method of approximate inverse: theory and applications*, Berlin, Germany: Springer, 2007.

[109] M. STONE, *Cross-validatory choice and assessment of statistical predictions*, J. Roy. Statist. Soc. Ser. B 36(1974), 111–147.

[110] A.N. TIKHONOV, *On the solution of ill-posed problems and the method of regularization*, In Doklady Akademii Nauk, Russian Academy of Sciences 151(1963), 501-504.

[111] A.N. TIKHONOV, *On the regularization of ill-posed problems*, In Doklady Akademii Nauk, Russian Academy of Sciences 153(1963), 49-52.

[112] A.N. TIKHONOV and V. ARSENIN, *Solutions to Ill-Posed Problems*, Wiley, New York, 1977.

[113] G. M. VAĬNIKKO, *The principle of the residual for a class of regularization methods*, Zh. Vychisl. Mat. i Mat. Fiz. 22(1982), 499–515,

[114] G. M. VAĬNIKKO, *The critical level of the residue in regularization methods*, Zh. Vychisl. Mat. i Mat. Fiz. 23(1983), 1283–1297.

[115] V. V. VASIN, *The stable evaluation of a derivative in space* $C(-\infty, \infty)$, USSR Comput. Math. Math. Phys. 13(1973), 16-24.

[116] C. R. VOGEL, *Non-convergence of the L-curve regularization parameter selection method*, Inverse Problems 12 (1996), 535–547.

[117] G. WAHBA, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM J. Numer. Anal. 14(1977), 651–667.

[118] Y. SAAD, *Iterative methods for sparse linear systems*, Society for Industrial and Applied Mathematics ,2003.