

ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé

UMR7104 – U964 IGBMC

THÈSE présentée par :

Afef CHERNI

Soutenue le : 20 Septembre 2018

Pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Sciences de la vie et de la Santé

**Méthodes modernes d'analyse de données en
biophysique analytique**

Résolution des problèmes inverses en RMN DOSY et SM

THÈSE dirigée par :

Marc-André Delsuc

Directeur de recherche, IGBMC, Université de Strasbourg

Co-encadrée par :

Emilie Chouzenoux

Maître de conférences, LIGM, Université Paris Est Marne-la-Vallée

RAPPORTEURS :

Charles Soussen

Professeur, L2S, CentraleSupélec Paris

Guillaume Van Der Rest

Professeur, Laboratoire de Chimie Physique, Université Paris Sud

EXAMINATEURS :

Caroline Chauv

Chargée de recherche CNRS, I2M, Aix-Marseille Université

Vincent Mazet

Maître de conférences, ICube, Université de Strasbourg

"Il faut toujours viser la lune, car même en cas d'échec, on atterrit dans les étoiles."
Oscar Wilde

"L'art d'être tantôt très audacieux et tantôt très prudent est l'art de réussir."
Napoléon Bonaparte

À mon frère Moemen,

Remerciements

"Soyons reconnaissants aux personnes qui nous donnent du bonheur, elles sont les charmants jardiniers par qui nos âmes sont fleuries."

Marcel Proust

Je tiens à adresser mes vifs remerciements aux personnes qui m'ont encadrée tout au long de mes trois années de doctorat : Marc-André Delsuc, directeur de recherche CNRS à l'IGBMC, Université de Strasbourg, et Émilie Chouzenoux, maître de conférences à l'Université Paris Est Marne-la-Vallée. Je les remercie pour leur conseil, leur compréhension et leur soutien dans les moments difficiles, leur enthousiasme qui m'a encouragée et leur confiance en moi qui m'a donnée la motivation pour finir ma thèse avec des résultats satisfaisants.

Je tiens aussi à remercier les membres de jury d'avoir accepté d'évaluer mon travail de thèse : Merci à M. Guillaume Van Der Rest, professeur à l'Université de Paris Sud et M. Charles Soussen, professeur à CentraleSupélec d'avoir accepté d'être les rapporteurs de mon manuscrit de thèse. Merci à Mme Caroline Chaux, chargée de recherche CNRS à I2M - Aix-Marseille Université, et M. Vincent Mazet, maître de conférences à l'Université de Strasbourg d'avoir accepté d'être les examinateurs de ma thèse. Et finalement merci à M. Christian Rolando, professeur à l'université de Lille et M. Laurent Duval, ingénieur de recherche à l'IFP Energies nouvelles, d'avoir accepté de faire partie de mon jury de thèse en tant que membres invités.

Mes reconnaissances s'adressent à M. Jean-Christophe Pesquet, professeur à Centrale-Supélec, pour sa collaboration sur les approches proximales, ses propositions pertinentes et ses conseils, ainsi que pour l'accueil dans son laboratoire. Merci aussi à M. Christian Rolando pour sa collaboration sur la spectrométrie de masse et les mesures qui m'ont aidée à tester mes approches théoriques sur des données réelles.

J'exprime ma gratitude à M. Bruno Kieffer, professeur à l'Université de Strasbourg et chef d'équipe à l'IGBMC, pour m'avoir accueillie dans son laboratoire où j'ai passé trois ans dans une excellente ambiance.

J'adresse mes remerciements et mes reconnaissances à Mme. Monia Turki, professeur à l'école Nationale d'Ingénieurs de Tunis, de m'avoir fait confiance et m'avoir encouragée à faire ces études de recherche.

Remerciements

Je tiens aussi à remercier Julia, ingénieure de recherche à Cascade pour sa gentillesse et son aide tout au long de ma thèse. Merci aussi aux membres d'équipe qui m'ont accompagnée tout au long de cette thèse en particulier Laura et Yasmine. Merci à ma copine Emna qui était toujours à mes côtés même à distance. Je n'oublie pas non plus mes chères amies Takoi et Lyna.

J'adresse toute mon affection à la plus chère personne dans ma vie, mon frère Moemen, l'homme qui n'a jamais cessé de me faire confiance et m'encourager, me soutenir et me permettre de me lever le cœur léger et l'esprit tranquille. Chanceuse d'être ta sœur.

Finalement, je tiens à remercier toute ma famille : une énorme pensée à ma mère et mon père qui m'ont toujours supportée. Un grand merci à ma tante et ma chère cousine Nadia qui ont été toujours présentes lorsque j'ai eu besoin. Je remercie en particulier ma chère sœur Sana qui n'a jamais cessé de m'encourager et me soutenir.

Table des matières

1	Introduction générale	1
1.1	Contexte	1
1.2	Collaboration	2
1.3	Organisation du document	2
1.4	Contributions	3
2	Problème inverse en biophysique	5
2.1	Introduction	6
2.2	Problème inverse	6
2.2.1	Modèle d'observation	6
2.2.2	Problème inverse mal posé	7
2.3	Méthodes de résolution	7
2.3.1	Formulation pénalisée	8
2.3.2	Formulation contrainte	9
2.3.3	Conclusion	10
2.4	Outils d'analyse convexe	10
2.4.1	Définitions et notations	10
2.4.2	Opérateur proximal	13
2.4.3	Algorithmes d'optimisation proximaux usuels	15
2.4.4	Récapitulatif	19
2.5	Conclusion	19
3	Résolution du problème inverse en DOSY	21
3.1	Introduction	22
3.2	La RMN DOSY	22
3.2.1	Phénomène de diffusion	22
3.2.2	Expérience RMN DOSY	23
3.3	Reconstruction du signal DOSY	24
3.3.1	Signal de diffusion	24
3.3.2	Positionnement du problème	25
3.4	État de l'art des méthodes de résolution	26
3.4.1	Méthode des moindres carrés contraints	26
3.4.2	Pénalisation entropique	26

3.4.3	Pénalisation de parcimonie	27
3.4.4	Autres approches	27
3.5	Régularisation hybride proposée	27
3.5.1	Entropie	28
3.5.2	Parcimonie	28
3.6	Opérateurs proximaux des pénalisations hybrides	30
3.6.1	Cas de l'Entropie de Shannon	30
3.6.2	Cas de l'Entropie de Burg	32
3.6.3	Algorithme d'optimisation	34
3.6.4	Version éclatée	34
3.7	Application à la RMN DOSY	35
3.7.1	Test de reconstruction	35
3.7.2	Test de robustesse	36
3.7.3	Comparaison avec l'approche éclatée	39
3.8	Conclusion	39
4	PALMA, nouvel algorithme pour la restauration des signaux DOSY	43
4.1	Introduction	44
4.2	Algorithme PALMA	44
4.3	Résultats avec des données simulées	45
4.3.1	Données de simulation	45
4.3.2	Optimisation des paramètres	45
4.3.3	Développement asymptotique	49
4.3.4	Évaluation de PALMA	49
4.4	Application de PALMA sur des données réelles	55
4.5	Serveur Web	57
4.6	Conclusion	58
5	Nouvelle approche par dictionnaire pour résoudre le problème inverse en SM	61
5.1	Introduction	62
5.2	Revue de la littérature	62
5.2.1	Principe de la SM	62
5.2.2	Le spectre de masse	64
5.2.3	Problème des massifs isotopiques	64
5.3	Méthodes de traitement de données en SM	66
5.4	Nouvelle approche par dictionnaire	67
5.4.1	Modélisation du problème	67
5.4.2	Approche par dictionnaire	68
5.4.3	Stratégie d'optimisation	68
5.4.4	Approximation par blocs	69
5.5	Résultats et validation	71
5.5.1	Tests de simulations	71
5.5.2	Tests sur des données réelles	76
5.6	Conclusion	77

6	Nouvelle pénalité pour la résolution des signaux parcimonieux en SM	79
6.1	Introduction	80
6.2	Présentation de la pénalisation ℓ_p/ℓ_q	80
6.2.1	Version exacte de ℓ_p/ℓ_q	81
6.2.2	Approximation lisse de ℓ_p/ℓ_q	81
6.2.3	Propriétés de la fonction Ψ	82
6.3	Algorithme de minimisation	87
6.3.1	Forward-Backward à métrique variable localement ajustée	87
6.3.2	Validité de l'algorithme	87
6.3.3	Calcul de l'opérateur proximal inexact	89
6.4	Analyse de convergence	89
6.5	Résultats expérimentaux	90
6.5.1	Données synthétiques	90
6.5.2	Test de reconstruction	90
6.6	Conclusion	95
7	Méthode de débruitage des données spectrales de grande dimension	96
7.1	Introduction	97
7.2	L'échantillonnage	97
7.2.1	Théorème de Shannon-Nyquist	97
7.2.2	Échantillonnage non uniforme	98
7.3	Méthodes de débruitage en biophysique	99
7.3.1	Papoulis-Gerchberg	99
7.3.2	Cadzow	100
7.3.3	urQRd	100
7.4	Nouvelle méthode PGSane	101
7.4.1	Modélisation du problème	101
7.4.2	Nouvelle proposition	103
7.5	Application sur des signaux synthétiques	103
7.6	Conclusion	109
8	Conclusion & perspectives	110
	Annexe A Calcul des opérateurs proximaux	114
A.1	Entropie de Shannon	115
A.2	Entropie de Shannon + ℓ_0	115
A.3	Entropie de Shannon + ℓ_1	117
A.4	Entropie de Shannon + $f_{log-sum}$	117
A.5	Entropie de Shannon + f_{Cauchy}	118
A.6	Entropie de Burg	119
A.7	Entropie de Burg + ℓ_0	119
A.8	Entropie de Burg + ℓ_1	120
A.9	Entropie de Burg + $f_{log-sum}$	120
A.10	Entropie de Burg + f_{Cauchy}	121

Annexe B Fonction W de Lambert	122
B.1 Définition	122
B.2 Utilisation pratique	122
Annexe C Calcul de l'opérateur proximal de PALMA	124
C.1 Opérateur proximal : $\alpha \text{Ent} + (1 - \alpha)\ell_1$	124
C.2 Mesures	125
Annexe D Serveur Web PALMA	128
Annexe E Calcul de l'opérateur proximal	132
E.1 Calcul de l'opérateur proximal avec l'algorithme PPXA+	132
E.2 Calcul de l'opérateur proximal avec la version accélérée de Chambolle-Pock	133
Annexe F Publications	134
Bibliographie	158

Table des figures

2.1	Modèle du problème linéaire inverse	6
2.2	Exemples d'espaces convexes (<i>C1</i> et <i>C2</i>) et non convexes (<i>C3</i>)	11
2.3	Présentation schématique d'une fonction f convexe ($\lambda \in]0, 1[$)	11
2.4	Exemples de deux fonctions : $f : x \rightarrow x $ convexe et $g : x \rightarrow \sqrt{ x }$ non convexe.	12
2.5	Projection d'un point x sur un ensemble convexe fermé C	14
2.6	Opérateurs proximaux de la fonction puissance pour quelques valeurs de p	15
2.7	Exemple de quelques itérations de l'algorithme Forward-Backward pour f : une fonction lisse et g la fonction indicatrice d'un espace convexe C	16
2.8	Exemple de quelques itérations de l'algorithme Douglas-Rachford pour f : une fonction indicatrice d'un espace convexe C et g la fonction distance au carré $x \rightarrow \frac{1}{2}d_D^2$	17
3.1	Mouvement brownien d'une particule microscopique en suspension dans l'eau. (D'après un dessin de Jean Perrin).	22
3.2	Spin en mouvement de précession de Larmor autour d'un champ magnétique (\vec{B}_0)	23
3.3	Séquence de la DOSY [Gilard 2008]	24
3.4	Fonctions d'entropie (Shannon et Burg) pour un scalaire $x \in \mathbb{R}$. ($\psi_1(x) \mapsto +\infty$ pour $x < 0$)	29
3.5	Fonctions de parcimonie ($\ell_0 = x ^0$ (3.17), $\ell_1 = x $ (3.18), log-sum= $\log(x + \delta)$ (3.19) et Cauchy= $\log(x^2 + \delta)$ (3.20) avec $\delta = 1$	29
3.6	Fonction \mathcal{W} de Lambert sur $[-\frac{1}{e}, +\infty[$ [Corless 1996]	31
3.7	Signaux synthétiques : A, B, C et D. Les signaux $\bar{\mathbf{x}}$ (à gauche) et leur mesures \mathbf{y} (à droite) avec $\sigma = 10^{-3}$	37
3.8	Résultats de reconstruction de signaux A, B, C et D pour $\sigma = 10^{-3}$ avec les différentes fonctions Ψ	38
3.9	Résultats de reconstruction des signaux A, B, C et D avec l'algorithme PALMA et sa version éclatée	40
4.1	Signaux simulés	46
4.2	Mesures DOSY des signaux simulés (Annexe C.2)	47
4.3	Résultats de reconstruction des signaux A, B, C et D pour différentes valeurs de α	48
4.4	Erreur relative entre $\mathcal{W}(e^{c_n})$ et l'approximation de $\text{prox}_\psi(x_n)$ définie dans l'équation (4.11). L'erreur relative devient inférieure à 10^{-3} pour $c_n > 20$	49

4.5	Résultats de reconstruction du signal A avec les différents algorithmes proposés pour les différents niveaux de bruit	51
4.6	Résultats de reconstruction du signal B avec les différents algorithmes proposés pour les différents niveaux de bruit	52
4.7	Résultats de reconstruction du signal C avec les différents algorithmes proposés pour les différents niveaux de bruit	53
4.8	Résultats de reconstruction du signal D avec les différents algorithmes proposés pour les différents niveaux de bruit	54
4.9	Spectre 2D DOSY 19F de la FTOH dans le DMSO. L'expérience a duré 15 min 22 s à une température T=289K	55
4.10	Spectre 2D DOSY 19F de de l'échantillon du sol dans le DMSO. L'expérience a durée 15 min à une température T=289K	56
4.11	Expérience DOSY sur un extrait éthanologique d'algue. Seules les composantes principales sont présentées. La ligne horizontale indique les chaînes d'acides aminés à $220 \mu\text{m}^2\text{s}^{-1}$, le glycérol à $380 \mu\text{m}^2\text{s}^{-1}$ et le méthanol à $1100 \mu\text{m}^2\text{s}^{-1}$	56
4.12	Comparaison entre deux expériences DOSY sur un extrait éthanologique d'algue : (a) résultat de l'expérience initiale (Figure 4.11), (b) même expérience avec 0.16mgmL^{-1} de chloroquine, (c) le spectre 1D de l'extrait d'algue avec la chloroquine	57
4.13	Représentation schématique des interactions avec le serveur PALMA	58
5.1	Représentation schématique d'un spectromètre de masse	62
5.2	Résolution du pic	64
5.3	Exemple du profil isotopique	65
5.4	Profil isotopique des alcanes pour différentes valeurs de n . (Valeur en bleu : masse mono-isotopique, Valeur en rouge : masse moyenne du massif)	66
5.5	Représentation schématique de la construction du dictionnaire \mathbf{D}	69
5.6	Représentation schématique de l'approximation par blocs $\overline{\mathbf{D}}_\ell$ de M/L matrices circulantes	70
5.7	Signal synthétique A mono-chargé ($M = 3000, P = 10, Z = 1$) avec $\sigma = 10^{-2}$ (à gauche : le signal cible, à droite : le spectre mesuré)	71
5.8	Signal synthétique B multi-chargé ($M = 4000, P = 50, Z = 3$) avec $\sigma = 10^{-2}$ (à gauche : le signal cible, à droite : le spectre mesuré)	72
5.9	Signal synthétique C multi-chargé ($M = 5000, P = 100, Z = 4$) avec $\sigma = 10^{-2}$ (à gauche : le signal cible, à droite : le spectre mesuré)	72
5.10	Résultat de reconstruction du signal synthétique A mono-chargé ($Z = 1$) avec \mathbf{D} (le signal original est présenté par les " - - • " et le signal reconstruit avec les " × ")	72
5.11	Résultat de reconstruction du signal synthétique B multi-chargé ($Z = 3$) avec \mathbf{D} (le signal original est présenté par les " - - • " et le signal reconstruit avec les " × ")	73
5.12	Résultat de reconstruction du signal synthétique C multi-chargé ($Z = 4$) avec \mathbf{D} (le signal original est présenté par les " - - • " et le signal reconstruit avec les " × ")	73
5.13	Résultats de reconstruction du signal A mono-chargé : Le signal original est en "•", le signal reconstruit avec l'approche par dictionnaire \mathbf{D} en ("×") et le signal reconstruit avec l'approximation par blocs $\overline{\mathbf{D}}$ ($L = 10$) en ("--").	74

5.14	Résultats de reconstruction du signal B multi-chargé ($z \in (1, 2, 3)$) : Le signal original est en "●", le signal reconstruit avec l'approche par dictionnaire \mathbf{D} en ("×") et le signal reconstruit avec l'approximation par blocs $\overline{\mathbf{D}}$ ($L = 10$) en ("--").	74
5.15	Résultats de reconstruction du signal C multi-chargé ($z \in (1, 2, 3, 4)$) : Le signal original est en "●", le signal reconstruit avec l'approche par dictionnaire \mathbf{D} en ("×") et le signal reconstruit avec l'approximation par blocs $\overline{\mathbf{D}}$ ($L = 10$) en ("--").	74
5.16	Résultats de reconstruction du signal A avec l'approximation par blocs pour différentes valeurs de L ($\sigma = 10^{-2}$)	75
5.17	Zoom sur la reconstruction du signal C avec l'approche dictionnaire exacte \mathbf{D} (présenté avec les "×") et l'approximation par blocs $\overline{\mathbf{D}}$ pour $L = 10$ (présenté avec les "--"). Le signal originale est présenté avec "●"	76
5.18	Zoom sur le résultat d'analyse du peptide EVEALEKKVAALESKVQALEKKVEALEHG-NH2. a- spectre de masse, b- signal reconstruit avec l'approximation par blocs $\overline{\mathbf{D}}$ pour $L = 10$ et $z = 5$	77
6.1	Signaux synthétiques : signal A mono-chargé de dimension $N = 1000$ et $P = 10$ (en haut). Signal B de $N = 1000$ ayant les états de charge $z \in (1, 2, 3)$ (en bas)	91
6.2	Résultat de reconstruction des signaux A (en haut) et B (en bas) (Figure 6.1) avec la pénalité : l_1 en "×", l_1/l_2 en "+", $l_{1/4}/l_2$ en "★" et $l_{3/4}/l_3$ en "--". Le signal original est en "●"	93
6.3	Zoom sur le résultat de reconstruction des signaux A (en haut) et B (en bas) (Figure 6.2) avec la pénalité : l_1 en "×", l_1/l_2 en "+", $l_{1/4}/l_2$ en "★" et $l_{3/4}/l_3$ en "--". Le signal original est en "●"	94
6.4	Erreur d'estimation d'intensités des différentes reconstructions des signaux A mono-chargé (en haut) et B multi-chargé (en bas), obtenus avec les pénalités : l_1 , l_1/l_2 , $l_{1/4}/l_2$ et $l_{3/4}/l_3$	94
7.1	Séquence incomplète d'échantillons relevés à une grille régulière	98
7.2	a) FID à 128.000 échantillons d'un spectre de 10 fréquences de différentes intensités allant de 1 à 10, b) zoom sur une partie du FID sous-échantillonné avec un degré de parcimonie égale à 1/8, c) FT du FID qui montre le bruit d'échantillonnage	105
7.3	Résultats de restauration du signal sous-échantillonné A donné dans la Figure 7.2 avec l'algorithme Papoulis-Gerchberg : a) FT du FID reconstruit avec 100 itérations, b) zoom sur le signal restauré dans le domaine temporel.	106
7.4	Résultats de restauration du signal sous-échantillonné A donné dans la Figure 7.2 avec l'Algorithme 11 : a) FT du FID reconstruit avec 100 itérations, b) zoom sur le signal restauré dans le domaine temporel.	106
7.5	Résultats de restauration du signal sous-échantillonné A donné dans la Figure 7.2 avec l'Algorithme Sane : a) FT du FID reconstruit (100 itérations et rang=10), b) zoom sur le signal restauré dans le domaine temporel	107
7.6	Résultats de restauration du signal sous-échantillonné A donné dans la Figure 7.2 avec l'Algorithme 12 : a) FT du FID reconstruit (100 itérations et rang=10), b) zoom sur le signal restauré dans le domaine temporel	107
7.7	Résultats de restauration du signal sous-échantillonné A donné dans la Figure 7.2 avec l'Algorithme 13, FT du FID reconstruit (100 itérations et rang=10)	108

TABLE DES FIGURES

7.8	a) FID à 128.000 échantillons d'un spectre de 10 fréquences de différentes intensités allant de 1 à 10 avec un bruit gaussien additif de niveau 5%, b) FT du FID qui montre le bruit d'échantillonnage, c) zoom sur une partie du FID sous-échantillonné avec un degré de parcimonie égale à $1/8$	108
7.9	Résultats de restauration du signal sous-échantillonné B donné dans la Figure 7.8 avec l'Algorithme 13 a) FT du FID reconstruit (100 itérations et rang=10), b) zoom sur un pic de faible intensité	109
C.1	Mise en évidence de la présence du bruit	125
C.2	Présentation en échelle logarithmique des mesures (C.1)	126
D.1	Page d'accueil	128
D.2	Demande d'analyse des données DOSY	129
D.3	Présentation illustrée de la théorie derrière PALMA	130
D.4	Explications du mode d'utilisation du serveur	131
D.5	Page de contacts	131

Liste des tableaux

2.1	Conditions d'utilisations de quelques algorithmes proximaux	19
3.1	RSB obtenus en reconstruisant les signaux A, B, C et D avec les différents choix de la fonction de régularisation Ψ pour différents niveaux de bruit.	39
4.1	RSB en dB obtenus par la reconstruction des signaux A, B, C et D pour $\sigma = 10^{-4}$ pour deux dimensions : $M = 64$ et $M = 32$	50
5.1	Isotopes et abondances relatives des atomes de la protéine [Audi 1993, Rosman 1999]	66
5.2	RSB (en dB), Temps de calcul (en s), mémoire de stockage de \mathbf{D} (en Mb) et nombre de pics \hat{P} détectés pour différentes valeurs de σ . L'approximation par blocs $\bar{\mathbf{D}}$ est testée pour différentes largeurs de fenêtres L (en Dalton)	75
6.1	RSB de la reconstruction des signaux synthétiques A et B pour différentes valeurs de $p \in]0, 2[$ et $q \in [2, +\infty[$	92

Liste des algorithmes

1	Forward Backward	15
2	Forward-Backward à métrique variable	16
3	Douglas-Rachford	17
4	PPXA+	18
5	Primal-Dual	18
6	PPXA+ pour le problème (3.11) avec le choix (3.12)	34
7	PPXA+ éclaté pour le problème (3.11) avec le choix (3.12)	35
8	Primal-Dual pour résoudre (5.8)	70
9	Forward-Backward à métrique variable localement ajustée	87
10	Calcul de l'opérateur proximal avec le Dual FB	89
11	Douglas-Rachford sous contrainte de parcimonie dans le domaine fréquentiel	102
12	Douglas-Rachford sous la contrainte du rang réduit	103
13	PPXA sous contraintes de parcimonie dans le domaine fréquentiel et du rang réduit de la matrice de Hankel	104
14	Calcul de l'opérateur proximal avec PPXA+	132
15	Calcul de l'opérateur proximal avec la version accélérée de Chambolle-Pock	133

Abréviations

CONTIN	CONTINuous diffusion coefficient strategy
DOSY	Diffusion Ordered SpectroscopY
DR	Douglas-Rachford
EQM	Erreur Quadratique Moyenne
FB	Forward-Backward
FID	Free Induction Decay
FTICR	Fourier Transform Ion Cyclotron Resonance
ITAMeD	Iterative Thresholding Algorithm for Multiexponentiel Decay
MaxEnt	Maximum Entropy strategy
NUS	Non Uniform Sampling
NITPICK	Non-greedy, Iterative Template-based peak PICKer
PALMA	Proximal Algorithm for ℓ_1 combined with MAXent prior
PD	Primal Dual
PG	Papoulis-Gerchberg
PPXA	Parallel ProXimal Algorithm
PPXA+	Parallel ProXimal Algorithm+
RMN	Résonance Magnétique Nucléaire
rQRd	random QR decomposition
RSB	Rapport Signal sur Bruit
Sane	Support selection Algorithm for Noise Elimination
SM	Spectométrie de Masse
SOOT	Smoothed One-Over-Two
SPOQ	Smoothed p -Over- q norm
SVD	Singular Value Decomposition
TRAI_n	Trust Region Algorithm for Inversion of molecular diffusion NMR
TOF	Time Of Flight
urQRd	uncloied random QR decomposition
VMFB	Variable Metric Forward-Backward



Notations

\mathbb{C}	L'ensemble des nombres complexes
\mathbb{R}	L'ensemble des nombres réels
\mathbb{R}_+	L'ensemble des nombres réels positifs
\mathbb{Z}	L'ensemble des nombres entiers relatifs
\mathbb{N}	L'ensemble des nombres entiers
$\mathbb{R}^{M \times N}$	L'ensemble des matrices réelles à M lignes et N colonnes
$\mathbf{x} = (x_n)_{1 \leq n \leq N}$	Le vecteur à N valeurs
x	Les scalaires sont notés par des minuscules
\mathbf{x}	Les vecteurs sont notés par des minuscules en gras
\mathbf{X}	Les matrices sont notés par des majuscules en gras
$(\cdot)^T$	L'opérateur transposé
$\text{Ker}(\mathbf{X})$	Le noyau de la matrice \mathbf{X}
$\text{Im}(\mathbf{X})$	L'image de la matrice \mathbf{X}
\mathbf{I}	La matrice identité
$\dot{f}(x)$	Le dérivé de f par rapport à x
$\nabla f(x)$	Le gradient de f par rapport à x
$\nabla^2 f(x)$	Le Hessian de f par rapport à x
$\ \cdot\ $	La norme euclidienne
$*$	Le produit de convolution
$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{argmin}} f(\mathbf{x})$	La valeur de \mathbf{x} pour laquelle f atteint son minimum
$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{Argmin}} f(\mathbf{x})$	L'ensemble des valeurs de \mathbf{x} pour lesquelles f atteint son minimum



Introduction générale

1.1 Contexte

L'analyse de grands volumes de données devient prédominant dans la science moderne, et se retrouve dans de nombreux domaines scientifiques tels que l'astrophysique, l'astronomie, les grandes expériences de physique, la biologie, ou même l'analyse de texte. Cela représente un nouveau défi en termes de manipulation et d'analyse des données, car il faut faire appel à des approches particulières qui minimisent leur manipulation, et font appel à des algorithmes adaptés et facilement parallélisables. Le développement de nouvelles approches algorithmiques de traitement de données est donc un défi majeur, en particulier en RMN (Résonance Magnétique Nucléaire) et en SM (Spectrométrie de Masse) pour l'utilisation moderne de ces spectroscopies, généralement en métabolomique et en protéomique.

Le travail de cette thèse vise à développer de nouvelles approches d'analyses de données RMN de type DOSY (Diffusion Ordered Spectroscopy) et de données en SM. La première modalité DOSY assure la mesure de coefficients de diffusion des molécules dans un mélange donné. Ces coefficients permettent de déterminer précisément le type des molécules (monodisperses ou polydisperses). Néanmoins, l'accès direct à ces coefficients de diffusion n'est pas possible et le recours à des approches d'analyses inverses est indispensable. La deuxième application qui est la spectrométrie de masse, présente un accès indirect aux informations d'intérêt telles que la masse monoisotopique et la charge des molécules via des mesures de très grande dimension. Cela nécessite des approches d'analyses avancées capables de traiter efficacement les informations de très grande dimension. Ces deux applications particulières confrontent des données de grande taille incluant la résolution d'un problème inverse. Bien qu'il existe des méthodes d'inversion généralisées qui répondent à ce besoin avec des résultats satisfaisants, une reconstruction précise, robuste au bruit et nécessitant un temps de calcul raisonnable et une occupation mémoire acceptable, reste encore un objectif à atteindre.

Dans ce manuscrit, nous proposons de nouvelles approches algorithmiques permettant la résolution efficace du problème de restauration des données DOSY et des spectres de masse, avec un contrôle d'occupation de mémoire et du temps de calcul.

1.2 Collaboration

Ce travail de thèse est supporté par le projet CNRS MASTODONS avec la subvention du 2016TABASCO et par l'Agence Nationale pour la Recherche (ANR, France), avec la subvention 2010FT-ICR2D et la subvention du Défi de tous les savoirs 2014, ONE-SHOT-FT-ICR-MS-2D.

Cette action coordonne le travail dans le laboratoire IGBMC, INSERM U596, CNRS UMR 7104, attaché à l'université de Strasbourg, en collaboration avec le Laboratoire d'Informatique Gaspard Monge (UMR CNRS 8049), UPEM, Marne-la-Vallée, France.

1.3 Organisation du document

Le document est organisé comme suit :

Dans le deuxième chapitre, nous présenterons la notion mathématique d'un problème inverse qui représente un accès indirect aux informations cibles dans les modèles de mesure physiques. Nous introduisons des stratégies d'inversion par minimisation d'une fonction de coût intégrant une régularisation adaptée. Une introduction aux approches d'optimisation en grande dimension est effectuée, avec un accent particulier porté sur les approches proximales. Ces approches seront à la base de nos contributions présentées dans la suite de ce manuscrit.

Dans le troisième chapitre, notre intérêt se tournera vers l'expérience RMN de type DOSY. Nous présenterons tout d'abord le principe de cette expérience et son utilité, et par la suite, les méthodes d'analyses de données DOSY proposées dans la littérature. Finalement, nous proposerons une nouvelle méthode d'analyse fondée sur une régularisation hybride qui combine la fonction d'entropie avec la mesure de parcimonie. Différentes pénalisations seront présentées ainsi que les opérateurs proximaux associés, que nous avons établis. La validité de la méthode est illustrée sur des exemples tirés de signaux synthétiques et réels.

Dans le quatrième chapitre, nous focaliserons notre travail sur la régularisation hybride qui combine l'entropie de Shannon à la norme ℓ_1 , menant aux meilleurs résultats de nos tests sur des données synthétiques. Cela mène à un nouvel algorithme PALMA (Proximal Algorithm for ℓ_1 combined with MAXent prior). Le choix des différents paramètres de PALMA seront discutés. De plus, nous présenterons les résultats d'analyse des mesures DOSY synthétiques et réelles avec notre algorithme PALMA. Nous présenterons à la fin de ce chapitre notre serveur web dédié à tous les utilisateurs pour faire l'analyse de leurs données DOSY.

Dans le cinquième chapitre, nous présenterons la deuxième application de cette thèse qui est la spectrométrie de masse. Nous commencerons par présenter le principe de cette application et les difficultés rencontrées lors de l'analyse d'un spectre de masse. Ensuite, après avoir étudié les méthodes d'analyse existantes, nous proposerons une nouvelle approche par dictionnaire dédiée à l'analyse de molécules protéomiques. Cette approche se base sur une minimisation d'une pénalité renforçant la parcimonie et la positivité, sous la contrainte de respecter un critère d'attache aux données. Nous détaillerons cette approche ainsi qu'une version accélérée basée sur un dictionnaire approché pour assurer une analyse efficace des spectres de masse avec un coût de calcul réduit et une mémoire d'occupation limitée.

Dans le sixième chapitre, une nouvelle approche SPOQ (Smoothed p -Over- q norm) basée sur la fonction de pénalité ℓ_p/ℓ_q , sera présentée pour améliorer l'estimation d'intensité des signaux parcimonieux présents en spectrométrie de masse. Nous établirons des propriétés mathématiques utiles pour la minimisation d'une approximation lisse de cette pénalité et nous présenterons un nouvel algorithme Forward-Backward à métrique variable localement ajustée pour résoudre le problème de minimisation non convexe associé. Nous finirons ce chapitre par la présentation des résultats d'estimation de positions et d'intensités des spectres SM avec l'approche SPOQ pour certaines valeurs de p et de q .

Dans le septième chapitre, notre intérêt sera tourné vers les méthodes de débruitage des données biophysiques en grande dimension. Nous présenterons les approches proposées pour minimiser l'erreur du sous-échantillonnage en SM et en RMN. Bien que ces approches sont d'un caractère non convexe, nous proposerons des méthodes équivalentes avec des garanties de convergence. Par ailleurs, une nouvelle méthode de débruitage sera présentée et testée avec des signaux RMN synthétiques.

1.4 Contributions

Articles acceptés (Annexe F)

- **Afef Cherni**, Emilie Chouzenoux, Marc-André Delsuc, **Fast dictionary based approach for mass spectrometry data analysis**, In Proceedings of the 43th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2018), ppx- x+5, Calgary-Alberta-Canada, 14-20 april 2018.
- Giovanni Chierchia, **Afef Cherni**, Emilie Chouzenoux, and Jean- Christophe Pesquet, **Approche de Douglas-Rachford aléatoire par blocs. Application à la régression logistique parcimonieuse**, In Actes du 26e colloque GRETSI (GRETSI 2017), Juan-les-Pins, 5-8 september 2017
- **Afef Cherni**, Emilie Chouzenoux, Marc-André Delsuc, **Improved algorithm for DOSY signal processing**, Analyst, Vol. 142, No. 5, pages 772-779, 2016.
- **Afef Cherni**, Emilie Chouzenoux, Marc-André Delsuc, **Proximity for a class of hybrid sparsity + entropy prior. Application to DOSY NMR signal reconstruction**, In Proceedings of the International Symposium on Signal, Image, Video and Communications (ISIVC 2016), Tunis-Tunisia, 21-23 november 2016.

Articles en préparation

- **Afef Cherni**, Emilie Chouzenoux, Laurent Duval, Jean-Christophe Pesquet, **SPOQ : Smoothed p -Over- q ℓ_p/ℓ_q norm ratio Regularization for Sparse Signal reconstruction**.
- Lionel Chiron, **Afef Cherni**, Jean-Philippe Starck, Cristian Rolando, Emilie Chouzenoux, Marc-André Delsuc, **Fast and robust analysis of non uniformly sampled 2D-FT-ICR-MS datasets**.

Logiciel diffusé

- PALMA : Plateforme en Python, basée sur l'algorithme proximal PALMA, proposée pour l'analyse des données expérimentales réelles DOSY. Juin 2016. License Cecill-B <http://palma.labo.igbmc.fr/>

Séminaires / Congrès

- Conférence Européenne : "European Mass Spectrometry Conference", 03-2018, Saarbrücken-Allemagne.
Poster : **Isotopic pattern analysis based on a new Fourier deconvolution approach.**
- Congrès International : "38-th British Mass Spectrometry Society Annual Meeting", 07-2017, Manchester United Kingdom.
Poster : **New dictionary approach for the isotopic deconvolution.**
- École d'hiver : "Structural Biology meets Biophysics, 12-2016, Obernai France".
Poster : **Proximal Algorithm for ℓ_1 combined with MAXent prior : a new approach for DOSY signal processing.**

Chapitre 2

Problème inverse en biophysique

Sommaire

2.1	Introduction	6
2.2	Problème inverse	6
2.2.1	Modèle d'observation	6
2.2.2	Problème inverse mal posé	7
2.3	Méthodes de résolution	7
2.3.1	Formulation pénalisée	8
2.3.2	Formulation contrainte	9
2.3.3	Conclusion	10
2.4	Outils d'analyse convexe	10
2.4.1	Définitions et notations	10
2.4.2	Opérateur proximal	13
2.4.3	Algorithmes d'optimisation proximaux usuels	15
2.4.4	Récapitulatif	19
2.5	Conclusion	19

2.1 Introduction

Avec le niveau de développement technologique actuel, l'amélioration et l'optimisation des performances restent toujours un besoin crucial. Dans différents domaines (télécommunications, physique, médecine, astronomie, etc.), récupérer toute l'information originale d'un objet cible (signal, image, vidéo) à partir d'observations incomplètes et dégradées constitue un défi majeur. Ce problème est connu sous le nom de problème inverse. L'objectif de ce travail de thèse est de proposer des solutions à cette problématique dans le contexte de la restauration de signaux en biophysique. Dans ce chapitre, nous allons commencer par définir en Section 2.2 la notion de problème inverse et nous caractériserons mathématiquement un problème mal posé.

Nous présenterons en Section 2.3 deux formulations permettant de résoudre un problème inverse, qui passent par la résolution d'un problème d'optimisation. Nous allons finir par introduire les outils mathématiques et algorithmiques d'optimisation en Section 2.4 qui seront à la base des développements de cette thèse.

2.2 Problème inverse

2.2.1 Modèle d'observation

La résolution d'un problème inverse vise à restaurer une information à laquelle l'accès direct n'est pas toujours possible mais qui possède un lien physique avec une grandeur mesurable.

Remonter vers l'information originale à partir des mesures peut être modélisé par la Figure 2.1 où \mathbf{K} représente l'opérateur de mesure qui fait le lien entre l'information originale inconnue et les observations. Ce problème est présent dans différentes applications comme l'imagerie médicale pour restaurer l'image [Reimer 2010], le traitement audio pour reconstruire les signaux [Bilen 2017], l'acoustique sous-marine pour déterminer les obstacles [Le Gall 2016], la mécanique quantique pour la détermination du potentiel [Nicoleau 2000], l'analyse chimique des compositions [Delsuc 1998], la biologie structurale [Gagneux 1995], la thermodynamique [Shell 2008], etc.

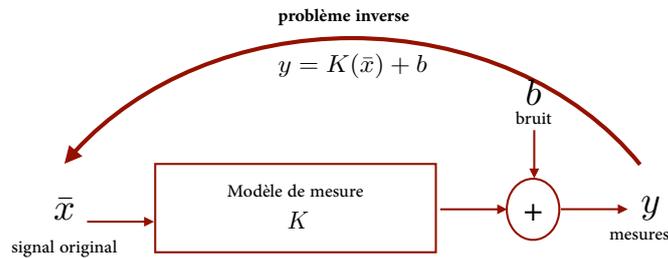


FIGURE 2.1 – Modèle du problème linéaire inverse

Dans cette thèse, nous considérons le modèle de mesure linéaire discret qui fait le lien entre le signal original qu'on cherche à reconstruire, noté par le vecteur $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N) \in \mathbb{R}^N$ et les mesures notées par le vecteur $\mathbf{y} = (y_1, y_2, \dots, y_M) \in \mathbb{R}^M$ via l'opérateur de mesure $\mathbf{K} \in \mathbb{R}^{M \times N}$ et $\mathbf{K} = (\mathbf{K}_{i,j})_{1 \leq i \leq M, 1 \leq j \leq N}$ en tenant compte du vecteur $\mathbf{b} = (b_1, b_2, \dots, b_M) \in \mathbb{R}^M$ qui représente le

bruit. Par conséquent, le problème peut être modélisé par l'équation suivante :

$$\mathbf{y} = \mathbf{K}\bar{\mathbf{x}} + \mathbf{b} \quad (2.1)$$

Pour pouvoir résoudre le problème décrit par l'équation (2.1), certaines conditions doivent être vérifiées par l'opérateur \mathbf{K} que nous précisons dans la suite.

2.2.2 Problème inverse mal posé

Un problème est dit "bien posé" si l'opérateur de mesure \mathbf{K} vérifie les trois conditions de Hadamard [Hadamard 1902] :

L'existence : $\forall \mathbf{y} \in \mathbb{R}^M, \exists \mathbf{x} \in \mathbb{R}^N$ tel que $\mathbf{y} = \mathbf{K}\mathbf{x}$.

L'unicité : Le noyau de \mathbf{K} est réduit au vecteur nul $\mathbf{x} = 0$, c'est à dire, l'unique solution de $\mathbf{K}\mathbf{x} = 0$ est $\text{Ker}(\mathbf{K}) = \{0\}$ ^a.

La continuité : L'image de \mathbf{K} est fermée, c'est à dire que toute limite de suites dans $\text{Im}(\mathbf{K})$ reste dans $\text{Im}(\mathbf{K})$ ^b.

Si la matrice de mesure \mathbf{K} est inversible (matrice carrée $M = N$ et $\text{Ker}(\mathbf{K}) = \{0\}$), la solution du problème sera envisagée par une simple inversion matricielle $\hat{\mathbf{x}} = \mathbf{K}^{-1}\mathbf{y}$. Sinon, dans le cas où au moins une de ces trois conditions n'est pas vérifiée, le problème sera considéré comme un problème inverse mal posé (au sens de Hadamard) et on aura donc besoin de méthodes de résolution plus avancées pour pouvoir estimer le signal cible $\bar{\mathbf{x}}$ à partir de la mesure \mathbf{y} .

La grande majorité des problèmes inverses rencontrés en traitement du signal, et en particulier en biophysique, sont mal posés, c'est à dire que l'une de ces conditions n'est pas satisfaite (généralement, aucune des trois conditions n'est vérifiée).

2.3 Méthodes de résolution

Toujours dans l'intérêt de reconstruire un signal, on tend à estimer l'information en exigeant un caractère de fidélité entre le modèle de mesure et les données observées.

La minimisation d'une fonction f peut répondre à cet objectif :

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} f(\mathbf{x}) \quad (2.2)$$

où typiquement, pour tout $\mathbf{x} \in \mathbb{R}^N$, $f(\mathbf{x}) = \|\mathbf{K}\mathbf{x} - \mathbf{y}\|^2$ ^c. Cela revient à trouver une estimation $\hat{\mathbf{x}}$ du signal original $\bar{\mathbf{x}}$ telle que la distance entre le modèle de mesure de cette estimation et les observations \mathbf{y} soit très faible.

Avec ce principe d'ajustement, les mesures seules ne permettent pas une exacte reconstruction de l'information originale. Un ensemble de solutions peut être obtenu. Pour trouver une solution adéquate au problème inverse, une information *a priori* sur le signal à estimer doit être introduite. Cela passe par l'ajout d'un terme qu'on appelle terme de régularisation souvent choisi selon le type de l'application physique et la nature du signal, ainsi que les propriétés qu'il possède (signe, forme, parcimonie, régularité, etc.). L'idée générale est de faire un compromis entre la fidélité du modèle de mesure avec les observations réelles et la régularité que le signal possède.

a. On désigne par $\text{Ker}(\mathbf{K})$ le noyau de l'opérateur \mathbf{K} , c'est à dire $\text{Ker}(\mathbf{K}) = \{\mathbf{x} \in \mathbb{R}^N \text{ tel que } \mathbf{K}\mathbf{x} = 0\}$

b. On désigne par $\text{Im}(\mathbf{K})$ l'image de l'opérateur \mathbf{K} , c'est à dire $\text{Im}(\mathbf{K}) = \{\mathbf{y} \text{ tel que } \mathbf{y} = \mathbf{K}\mathbf{x}\}$

c. $\|\cdot\|$ définit la norme euclidienne de \mathbb{R}^N

2.3.1 Formulation pénalisée

L'approche pénalisée est l'une des méthodes les plus utilisées pour résoudre un problème inverse mal posé. Son principe est basé sur la minimisation d'un critère de fidélité et une fonction de régularisation, et il est défini par :

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} \quad f(\mathbf{x}) + \lambda g(\mathbf{x}) \quad (2.3)$$

tel que f représente le terme de fidélité lié aux mesures, généralement choisi comme l'erreur résiduelle entre \mathbf{y} et $\mathbf{K}\mathbf{x}$, g définit le terme de régularisation introduit pour renforcer certaines propriétés souhaitables de la solution, et λ représente un coefficient positif qui ajuste le compromis entre les deux termes f et g . On trouve de nombreux choix pour la fonction de régularisation dans la littérature. Nous listons ici les plus populaires dans le domaine de la biophysique.

L'une des approches les plus connues est la régularisation de Tikhonov [Tikhonov 1976, Tikhonov 1986], basée sur un choix particulier de la fonction g telle que :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad g(\mathbf{x}) = \|\Gamma \mathbf{x}\|^2 \quad (2.4)$$

Γ est généralement choisie comme la matrice identité $\Gamma = \mathbf{I}_N$, ce qui revient à favoriser les solutions ayant une petite norme [Chiang 2005, Gavriluk 2017].

Un autre choix de la fonction de régularisation est basé sur l'entropie introduite par [Shannon 1948], défini par :

$$(\forall \mathbf{x} \in \mathbb{R}_+^N) \quad g(\mathbf{x}) = \text{ent}(\mathbf{x}) = \sum_{i=1}^N x_n \log(x_n) \quad (2.5)$$

Notons que la fonction d'entropie est définie par $-\sum_{i=1}^N x_n \log(x_n)$ pour tout $x_n \in]0, +\infty[$. La maximisation de cette mesure de la quantité d'information revient à la minimisation de son opposé, d'où vient la forme du critère de pénalité g . Cette définition a été introduite en 1982 afin de quantifier l'information contenue dans un objet (signal électrique, fichier informatique, information numérique, etc.) et mesurer la quantité d'information transmise entre un émetteur et un récepteur. Cependant, l'histoire de l'entropie a débuté en 1865 avec la thermodynamique pour définir le désordre d'un système [Landau 1970].

Ce choix de régularisation a été souvent utilisé pour résoudre différents problèmes ayant une contrainte de positivité comme la reconstruction d'image [Boccacci 2002], l'estimation spectrale en astronomie [Buck 1991, Pichon 1998], la reconstruction de signaux en spectroscopie [Skilling 1984, Delsuc 1998] et en résonance magnétique nucléaire [Delsuc 1989, Delsuc 1998], la tomographie à rayon X [Djafari 1988], etc.

En acquisition comprimée [Bruckstein 2009], la parcimonie d'un signal a été utilisée comme une fonction de pénalisation. En faite, différents signaux dans la nature possèdent des représentations avec des coefficients qui sont nuls ou approximativement nuls selon leurs domaines de traitement citons par exemple la transformée de Fourier, la transformée en cosinus et la transformée en ondelettes discrètes.

Un signal est dit parcimonieux s'il possède une représentation qui contient seulement quelques éléments significatifs (non nuls). Le degré de parcimonie \mathbf{k} est défini par :

$$\mathbf{k} = \text{card}\{n, \text{ tel que } x_n \neq 0 \quad \forall \mathbf{x} = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N\} \quad (2.6)$$

La fonction de régularisation ℓ_0 a tendance à privilégier les signaux parcimonieux [Elad 2010] :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad g(\mathbf{x}) = \ell_0(\mathbf{x}) = \sum_{n=1}^N |x_n|^0 \quad (2.7)$$

Comme sa définition le montre, cette pseudo-norme compte les éléments non nuls dans un signal $\mathbf{x} \in \mathbb{R}^N$. Elle s'avère être un outil efficace dans l'acquisition comprimée [Candès 2006, Donoho 2006, Mazet 2013, Souden 2015] afin de réduire les dimensions de l'information en gardant le nécessaire et reconstruire les signaux parcimonieux.

Il est courant d'utiliser la norme ℓ_1 à la place de la pseudo-norme ℓ_0 car c'est la norme convexe la plus proche [Chen 2001, Davenport 2011] définie par :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad g(\mathbf{x}) = \ell_1(\mathbf{x}) = \sum_{n=1}^N |x_n| \quad (2.8)$$

La norme ℓ_1 est utilisée dans différentes applications comme l'holographie acoustique [Chardon 2010], l'astronomie [Mary 2007], l'imagerie [Courtès 2015], etc.

Une généralisation de la norme de parcimonie ℓ_p avec $p \in]0, 2]$ a été proposée :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad g(\mathbf{x}) = \ell_p(\mathbf{x}) = \left(\sum_{n=1}^N |x_n|^p \right)^{\frac{1}{p}} \quad (2.9)$$

avec un ajustement de la valeur de p selon le type des données cibles [Yarlagadda 1985].

Notons que la fonction de pénalité ℓ_p (2.9) n'est autre que la pseudo-norme ℓ_0 (2.7) quand $p \rightarrow 0$. Bien que l'approche pénalisée assure la résolution du problème inverse mal posé de façon efficace, le paramètre de la régularisation λ reste difficile à régler. Une sur-régularisation (λ trop grande) ou une sous-régularisation (λ trop petite) conduit à une mauvaise reconstruction du signal. Dans les cas simulés, une estimation manuelle de λ peut être effectuée en minimisant l'erreur entre le signal original $\bar{\mathbf{x}}$ et son estimée $\hat{\mathbf{x}}$. On peut utiliser pour cela par exemple le rapport signal sur bruit noté RSB et donné par :

$$\text{RSB}(\bar{\mathbf{x}}, \hat{\mathbf{x}}) = 20 \log_{10} \left(\frac{\|\bar{\mathbf{x}}\|}{\|\bar{\mathbf{x}} - \hat{\mathbf{x}}\|} \right) \quad (2.10)$$

où aussi l'erreur quadratique moyenne (EQM) définie par :

$$\text{EQM}(\bar{\mathbf{x}}, \hat{\mathbf{x}}) = \frac{1}{N} \|\bar{\mathbf{x}} - \hat{\mathbf{x}}\|^2 \quad (2.11)$$

Cependant, dans les cas pratiques la solution $\bar{\mathbf{x}}$ est bien sûr inconnue, et le calcul de l'erreur de reconstruction n'est pas possible. Quelques travaux proposent des approches pour affiner la valeur de λ automatiquement comme par exemple la validation croisée [Golub 2012] mais cela est souvent au prix d'un accroissement du coût de calcul global de la méthode de résolution.

2.3.2 Formulation contrainte

Une stratégie alternative consiste à adopter une formulation contrainte. Avec cette approche, on tend à minimiser un critère de régularité g sous la condition que le signal cible vérifie une contrainte sur le terme d'attache aux données, quantifiée par la fonction f . Cette contrainte mesure un écart

entre les mesures bruitées et les reconstructions [Tarantola 2005]. Le problème à résoudre devient :

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} \quad g(\mathbf{x}) \quad \text{tel que} \quad f(\mathbf{x}) \leq \eta \quad (2.12)$$

avec $\eta > 0$ qui est lié à une estimation du niveau de bruit.

Avec cette approche, l'ajustement du paramètre de régularisation est remplacé par une estimation du niveau de bruit.

2.3.3 Conclusion

L'approche de régularisation et l'approche sous contrainte sont deux méthodes efficaces pour résoudre un problème inverse mal posé sous la condition de bien choisir le critère de régularité. Chacune de ces approches présente une difficulté soit pour déterminer le coefficient de régularisation, soit pour déterminer le niveau de bruit.

La résolution explicite de (2.3) où (2.12) se révèle souvent difficile et le recours à une méthode d'optimisation itérative est nécessaire. D'autre part, les problèmes à grande dimension sont souvent complexes, coûteux en temps de calculs et en occupation mémoire, ce qui augmente d'autant plus la difficulté de leur résolution. De nombreuses approches d'optimisation sont disponibles pour la résolution de ce genre de problème, fondées sur les caractéristiques mathématiques des fonctions f et g (comme la convexité, la différentiabilité, etc.).

Nous présenterons dans la section suivante les outils fondamentaux d'analyse et d'optimisation qui nous seront utiles dans le reste de ce manuscrit.

2.4 Outils d'analyse convexe

2.4.1 Définitions et notations

Nous présentons dans cette section les définitions d'analyse convexe utiles dans la suite de ce manuscrit.

Définition 2.1.

Soit C un sous-ensemble de \mathbb{R}^N . L'ensemble C est convexe si :

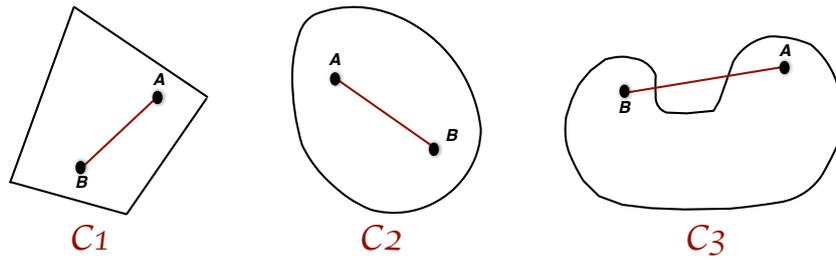
$$(\forall \lambda \in]0, 1[), (\forall (\mathbf{x}, \mathbf{y}) \in C^2) \quad \lambda \mathbf{x} + (1 - \lambda) \mathbf{y} \in C$$

Géométriquement, cela revient au fait que l'espace C joint tous les segments $[A, B]$ qui joignent deux points A et B quelconques de ses points. La Figure 2.2 présente un exemple de deux espaces convexes $C1$ et $C2$ et un espace non convexe $C3$.

Définition 2.2.

La fonction indicatrice d'un ensemble $C \subset \mathbb{R}^N$ est définie par :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \iota_C = \begin{cases} 0 & \text{si } \mathbf{x} \in C \\ +\infty & \text{sinon} \end{cases}$$

FIGURE 2.2 – Exemples d'espaces convexes ($C1$ et $C2$) et non convexes ($C3$)**Définition 2.3.**

Soit $f : \mathbb{R}^N \rightarrow]-\infty, +\infty]$. Le domaine de la fonction f est l'ensemble noté $\text{dom } f$ tel que :

$$\text{dom } f = \{\mathbf{x} \in \mathbb{R}^N \mid f(\mathbf{x}) < +\infty\}$$

Remarque 2.1.

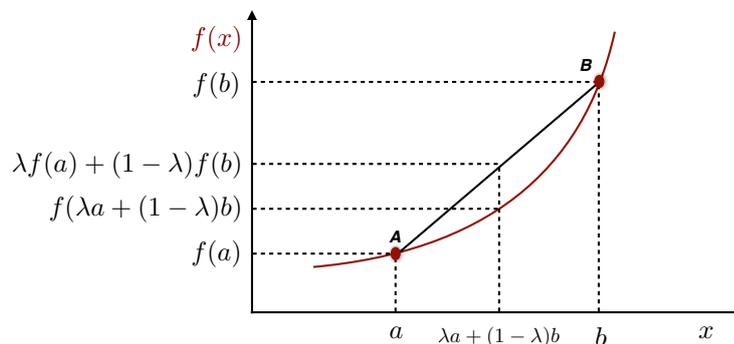
f est une fonction propre si $\text{dom } f \neq \emptyset$.

Définition 2.4.

Une fonction $f : \mathbb{R}^N \rightarrow]-\infty, +\infty]$ est dite convexe si :

$$(\forall \lambda \in]0, 1[), (\forall (\mathbf{x}, \mathbf{y}) \in (\text{dom } f)^2) \quad f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y})$$

On peut déterminer la convexité d'une fonction à valeurs scalaires à partir de son graphe. Si pour tous points A et B du graphe de f , le segment $[AB]$ est entièrement situé au-dessus du graphe de f , la fonction f est convexe (Figure 2.3).

FIGURE 2.3 – Présentation schématique d'une fonction f convexe ($\lambda \in]0, 1[$)

La Figure 2.4 présente un exemple de deux fonctions : f convexe et g non convexe :

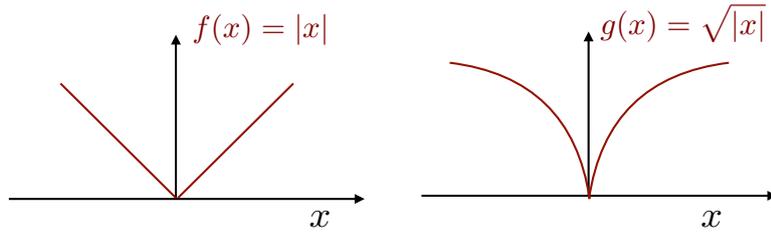


FIGURE 2.4 – Exemples de deux fonctions : $f : x \rightarrow |x|$ convexe et $g : x \rightarrow \sqrt{|x|}$ non convexe.

Définition 2.5.

Une fonction $f : \mathbb{R}^N \rightarrow]-\infty, +\infty]$ est dite strictement convexe si elle vérifie :
 $(\forall \lambda \in]0, 1[), (\forall (\mathbf{x}, \mathbf{y}) \in (\text{dom } f)^2)$

$$\mathbf{x} \neq \mathbf{y} \Rightarrow f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y})$$

Remarque 2.2.

Si f est une fonction convexe, alors $\text{dom } f$ est convexe.

Définition 2.6.

Une fonction $f : \mathbb{R}^N \rightarrow]-\infty, +\infty]$ est dite coercive si :

$$\lim_{\|\mathbf{x}\| \rightarrow +\infty} f(\mathbf{x}) = +\infty$$

Définition 2.7.

Soit une fonction $f : \mathbb{R}^N \rightarrow]-\infty, +\infty]$ et soit $\mathbf{x} \in \mathbb{R}^N$. La fonction f est semi-continue inférieurement si :

$$(\forall \mathbf{x}_0 \in \mathbb{R}^N) \quad \liminf_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) \geq f(\mathbf{x}_0)$$

Définition 2.8.

La classe $\Gamma_0(\mathbb{R}^N)$ représente la classe des fonctions $f : \mathbb{R}^N \rightarrow]-\infty, +\infty]$ convexes, semi-continues inférieurement et propres.

Théorème 2.1.

Si $f : \mathbb{R}^N \rightarrow]-\infty, +\infty]$ est une fonction propre et convexe avec $\inf(f) > -\infty$, alors tout minimum local de f est un minimum global.

Théorème 2.2.

Soit $f \in \Gamma_0(\mathbb{R}^N)$ et C un ensemble convexe fermé avec $\text{dom } f \cap C \neq \emptyset$.

(i) Si f est coercive dans C , alors il existe $\hat{\mathbf{x}} \in C$ tel que :

$$f(\hat{\mathbf{x}}) = \inf_{\mathbf{x} \in C} f(\mathbf{x})$$

(ii) Si f est strictement convexe, alors $\hat{\mathbf{x}}$ est unique.

Définition 2.9.

Soit f une fonction de $\Gamma_0(\mathbb{R}^N)$. Le sous-différentiel de Moreau de f en $\mathbf{x} \in \text{dom } f$ est défini par :

$$\partial f(\mathbf{x}) := \{\mathbf{t} \in \mathbb{R}^N \mid (\forall \mathbf{y} \in \mathbb{R}^N) \quad f(\mathbf{y}) - \langle \mathbf{y} - \mathbf{x}, \mathbf{t} \rangle \geq f(\mathbf{x})\} \quad (2.13)$$

Définition 2.10.

Une fonction $f : \mathbb{R}^N \rightarrow]-\infty, +\infty]$ est dite Gâteaux-différentiable en $\mathbf{x} \in \text{dom } f$ si et seulement si la dérivée directionnelle :

$$f'(\mathbf{x}, \mathbf{y}) = \lim_{\lambda \rightarrow 0^+} \frac{f(\mathbf{x} + \lambda \mathbf{y}) - f(\mathbf{x})}{\lambda}$$

existe pour toutes les directions $\mathbf{y} \in \mathbb{R}^N$ et la fonction $\mathbf{y} \rightarrow f'(\mathbf{x}, \mathbf{y})$ est linéaire continue.

Théorème 2.3.

Si f est convexe Gâteaux-différentiable en $\mathbf{x} \in \text{dom } f$, alors il existe un unique sous-gradient appelé gradient de f en \mathbf{x} noté $\nabla f(\mathbf{x})$ avec :

$$\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$$

Définition 2.11.

Une fonction $f : \mathbb{R}^N \rightarrow]-\infty, +\infty]$ est dite β -Lipschitz différentiable avec $\beta \in]0, +\infty[$ si :

$$(\forall (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^N \times \mathbb{R}^N) \quad \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq \beta \|\mathbf{x} - \mathbf{y}\|$$

Définition 2.12.

Une suite $(\mathbf{x}_k)_{k \in \mathbb{N}}$ d'éléments de \mathbb{R}^N converge vers un élément $\hat{\mathbf{x}} \in \mathbb{R}^N$ si :

$$\lim_{k \rightarrow +\infty} \|\mathbf{x}_k - \hat{\mathbf{x}}\| = 0$$

2.4.2 Opérateur proximal

Nous adoptons dans ce manuscrit des méthodes d'optimisation appelées "méthodes proximales". Ces approches s'appuient sur un opérateur proximal défini en un point $\mathbf{x} \in \mathbb{R}^N$ pour toute fonction $f \in \Gamma_0(\mathbb{R}^N)$ comme étant l'unique minimiseur de la fonction $f + \frac{1}{2}\|\cdot - \mathbf{x}\|^2$ [Moreau 1965]. Il est donné par :

$$\text{prox}_f : \mathbf{x} \mapsto \underset{\mathbf{y} \in \mathbb{R}^N}{\text{argmin}} \left(\frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2 + f(\mathbf{y}) \right) \quad (2.14)$$

On peut interpréter cet opérateur comme étant la minimisation de la fonction f sommée avec la fonction $\frac{1}{2}\|\cdot - \mathbf{x}\|^2$ pénalisant l'éloignement entre le point \mathbf{y} et le point \mathbf{x} . L'opérateur proximal a été largement utilisé pour proposer des algorithmes efficaces de résolution de problèmes inverses [Combettes 2007, Chambolle 2011].

Sa définition peut être généralisée pour des fonctions non nécessairement convexes [Hiriart-Urruty 1993, Sec.XV-4].

Définition 2.13.

Soit $f : \mathbb{R}^N \rightarrow]\infty, +\infty[$ une fonction propre, semi continue inférieurement, non nécessairement

convexe. L'opérateur proximal de f est donné par l'opérateur multi-valué suivant :

$$\text{prox}_f : \mathbf{x} \mapsto \underset{\mathbf{y} \in \mathbb{R}^N}{\text{Argmin}} \left(\frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 + f(\mathbf{y}) \right)$$

Parmi les propriétés importantes de l'opérateur proximal, citons l'inclusion suivante valide pour $f \in \Gamma_0(\mathbb{R}^N)$:

$$(\forall (\mathbf{x}, \mathbf{p}) \in \mathbb{R}^N \times \mathbb{R}^N) \quad \mathbf{p} = \text{prox}_f(\mathbf{x}) \Leftrightarrow \mathbf{x} - \mathbf{p} \in \partial f(\mathbf{p})$$

Cette équivalence peut être simplifiée si f est différentiable :

$$(\forall (\mathbf{x}, \mathbf{p}) \in \mathbb{R}^N \times \mathbb{R}^N) \quad \mathbf{p} = \text{prox}_f(\mathbf{x}) \Leftrightarrow \mathbf{x} - \mathbf{p} = \nabla f(\mathbf{p})$$

Il est possible d'étendre la définition 2.13 en intégrant une métrique dans le terme quadratique [Hiriart-Urruty 1993] :

Définition 2.14.

Soit $f : \mathbb{R}^N \rightarrow]-\infty, +\infty[$ une fonction propre, semi continue inférieurement, non nécessairement convexe. L'opérateur proximal de f en $\mathbf{x} \in \mathbb{R}^N$ relatif à la métrique notée $\mathbf{A} \in \mathbb{R}^{N \times N}$ symétrique définie positive, est noté par $\text{prox}_{\mathbf{A},f}$, donné par :

$$\text{prox}_{\mathbf{A},f} : \mathbf{x} \mapsto \underset{\mathbf{y} \in \mathbb{R}^N}{\text{Argmin}} \left(\frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_{\mathbf{A}}^2 + f(\mathbf{y}) \right) \quad (2.15)$$

Nous illustrerons l'intérêt de cette extension dans le Chapitre 6.

L'opérateur proximal peut être vu comme une étape de descente de (sous) gradient implicite. Cette propriété lui permet d'être bien adapté à la minimisation efficace des fonctions non lisses.

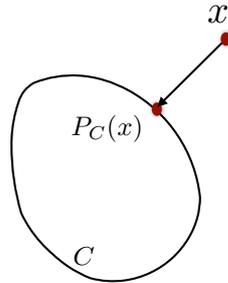
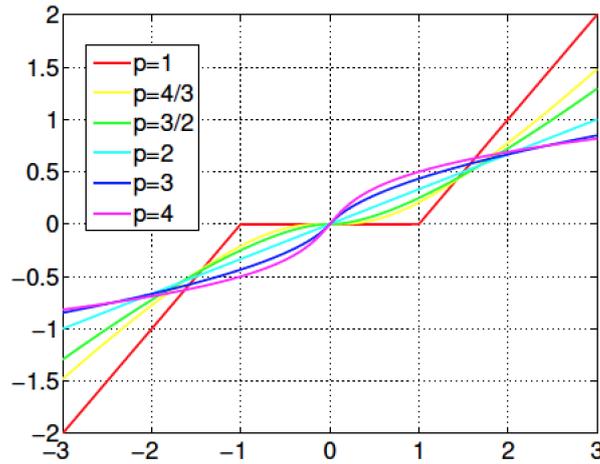


FIGURE 2.5 – Projection d'un point x sur un ensemble convexe fermé C

Nous présentons comme exemple l'opérateur proximal de la fonction indicatrice d'un ensemble C convexe fermé (Figure 2.5).

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{\iota_C}(x) = \underset{y \in C}{\text{argmin}} \frac{1}{2} \|y - x\|^2 = P_C(x) \quad (2.16)$$

Aussi, la Figure 2.6 présente quelques opérateurs proximaux de la fonction puissance définie par $\varphi : \mathbb{R} \rightarrow]-\infty, +\infty]$, $x \rightarrow |x|^p$ avec $p \geq 1$ [Chaux 2007]. En particulier, pour $p = 1$, on retrouve

FIGURE 2.6 – Opérateurs proximaux de la fonction puissance pour quelques valeurs de p

l'opérateur de seuillage doux [Combettes 2011] :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{|\cdot|^p}(x) = \text{sign}(x) \max\{|x| - 1, 0\} \quad (2.17)$$

2.4.3 Algorithmes d'optimisation proximaux usuels

Cette section est consacrée à la présentation de quelques algorithmes basés sur l'opérateur proximal, utiles pour la résolution d'un problème inverse. Notre but n'est pas de citer tous les algorithmes d'optimisation, mais plutôt de faire un rapide survol des algorithmes proximaux de base que nous allons utiliser dans la suite de cette thèse.

2.4.3.1 Forward-Backward

L'algorithme Forward-Backward (Explicite-Implicite) [Combettes 2005] est un algorithme de descente de premier ordre proposé dans l'optique de résoudre les problèmes ayant la forme générale suivante :

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} \quad f(\mathbf{x}) + g(\mathbf{x}) \quad (2.18)$$

où $g \in \Gamma_0(\mathbb{R}^N)$, $f \in \Gamma_0(\mathbb{R}^N)$ est β -Lipschitz différentiable. Cela convient à notre problème (2.3) telle que f est le terme d'attache aux données et g présente le terme de régularité. L'algorithme s'écrit comme suit :

Algorithme 1 Forward Backward

$$\left[\begin{array}{l} \mathbf{x}_0 \in \mathbb{R}^N \\ \text{Pour tout } k \in \mathbb{N}, 0 < \gamma_k < 2\beta^{-1}, \lambda_k \in]0, 1] \\ \text{Pour } k = 0, 1, \dots \\ \quad \left[\begin{array}{l} \mathbf{y}_k = \mathbf{x}_k - \gamma_k \nabla f(\mathbf{x}_k) \\ \mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k (\text{prox}_{\gamma_k g}(\mathbf{y}_k) - \mathbf{x}_k) \end{array} \right. \end{array} \right.$$

où γ_k définit le pas d'avancement de l'algorithme et λ_k est un facteur de relaxation. La suite $(\mathbf{x}_k)_{k \in \mathbb{N}}$

converge vers un minimum global de $f+g$. Comme son nom l'indique, le Forward-Backward alterne entre une étape explicite en calculant le gradient de f et une étape implicite avec l'opérateur proximal de g .

Remarquons que quand $g = 0$, on retrouve l'algorithme bien connu de descente de gradient pour minimiser f [Bertsekas 1999]. D'autre part, si $f = 0$, on obtient l'algorithme du point proximal pour minimiser g [Rockafellar 1976]. Une extension des propriétés de convergence du Forward-Backward dans le cas où f et g sont deux fonctions non convexes a été récemment proposée [Attouch 2009, Attouch 2011] pour $\lambda_k = 1$.

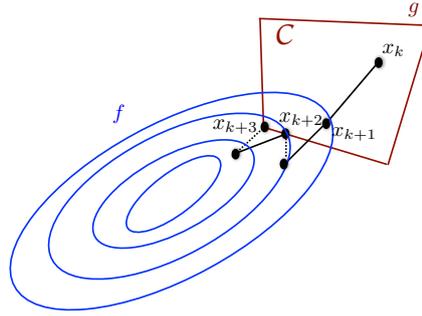


FIGURE 2.7 – Exemple de quelques itérations de l'algorithme Forward-Backward pour f : une fonction lisse et g la fonction indicatrice d'un espace convexe C

Bien que le Forward-Backward soit un algorithme efficace pour résoudre un problème d'optimisation, il peut présenter une faible vitesse de convergence. Chouzenoux *et al.* ont développé une nouvelle version accélérée de cet algorithme VMFB (Variable Metric Forward-Backward) qui consiste à utiliser une métrique variable à chaque itération et qui s'appuie sur des matrices de préconditionnement [Chouzenoux 2014]. Le Forward-Backward à métrique variable donné par l'algorithme 2, résout le problème (2.18) :

Algorithme 2 Forward-Backward à métrique variable

$$\left[\begin{array}{l} \mathbf{x}_0 \in \mathbb{R}^N \\ \text{Pour tout } k \in \mathbb{N}, \gamma_k \in]0, 2[\\ \text{Pour } k = 0, 1, \dots \\ \quad \tilde{\mathbf{x}}_k = \mathbf{x}_k - \gamma_k \mathbf{A}_k^{-1}(\mathbf{x}_k) \nabla f(\mathbf{x}_k) \\ \quad \mathbf{x}_{k+1} = \text{prox}_{\gamma_k^{-1} \mathbf{A}_k(\mathbf{x}_k), g}(\tilde{\mathbf{x}}_k) \end{array} \right.$$

où $\text{prox}_{\gamma_k^{-1} \mathbf{A}_k(\mathbf{x}_k), g}$ désigne l'opérateur proximal à métrique \mathbf{A}_k donné par (2.15) et $(\mathbf{A}_k)_{k \in \mathbb{N}}$ sont des matrices symétriques, définies positives de $\mathbb{R}^{N \times N}$ vérifiant une propriété de majoration telle que pour tout $k \in \mathbb{N}$:

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad f(\mathbf{x}) \leq f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^\top (\mathbf{x} - \mathbf{x}_k) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_k)^\top \mathbf{A}_k (\mathbf{x} - \mathbf{x}_k) \quad (2.19)$$

Nous reviendrons vers cette propriété dans le Chapitre 6.

2.4.3.2 Douglas-Rachford

L'algorithme Douglas-Rachford [Douglas 1956] vise aussi à résoudre le même problème (2.18) avec f et g dans $\Gamma_0(\mathbb{R}^N)$. Cependant, la fonction g n'est pas nécessairement différentiable ce qui rend la méthode plus souple. L'algorithme Douglas-Rachford décrit dans la suite nécessite seulement le calcul des opérateurs proximaux de f et g .

Algorithme 3 Douglas-Rachford

$$\begin{array}{l} \mathbf{x}_0 \in \mathbb{R}^N \\ \text{Pour tout } k \in \mathbb{N}, \gamma_k > 0, \lambda_k \in]0, 2] \\ \text{Pour } k = 0, 1, \dots \\ \quad \left[\begin{array}{l} \mathbf{y}_k = \text{prox}_{\gamma_k g}(\mathbf{x}_k) \\ \mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k (\text{prox}_{\gamma_k f}(2\mathbf{y}_k - \mathbf{x}_k) - \mathbf{y}_k) \end{array} \right. \end{array}$$

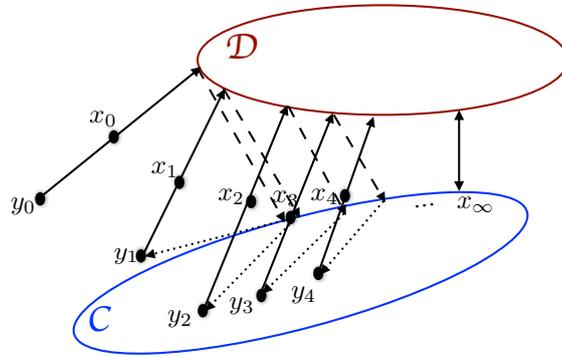


FIGURE 2.8 – Exemple de quelques itérations de l'algorithme Douglas-Rachford pour f : une fonction indicatrice d'un espace convexe C et g la fonction distance au carré $x \rightarrow \frac{1}{2}d_D^2$

La suite $(\mathbf{x}_k)_{k \in \mathbb{N}}$ converge vers le minimum global du problème (2.18). Nous présentons dans la Figure (2.8) quelques itérations de l'algorithme Douglas-Rachford pour un cas particulier des fonctions f et g où f présente la fonction indicatrice et g définie la fonction distance^d [Combettes 2011]. Cet algorithme est puissant et bien adapté pour la résolution du problème de minimisation sous contraintes (2.12) [Yu 2017].

2.4.3.3 PPXA+

Nous présenterons dans la suite une forme éclatée de Douglas-Rachford [Pesquet 2012]. Considérons maintenant le problème d'optimisation plus général suivant :

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} \quad \sum_{i=1}^m g_i(\mathbf{L}_i \mathbf{x}) \quad (2.20)$$

où, pour tout $i = 1, \dots, m$, $\mathbf{L}_i \in \mathbb{R}^{N \times M_i}$ et $g_i \in \Gamma_0(\mathbb{R}^{M_i})$.

L'algorithme correspondant à cette démarche est nommé PPXA+ [Pustelnik 2011], l'abréviation de Parallel ProXimal Algorithm, donné par :

d. On désigne par la fonction $f(x) = \frac{1}{2}d_D^2(x) = \frac{1}{2} \inf_{y \in D} \|y - x\|^2$

Algorithme 4 PPXA+

$$\left[\begin{array}{l}
 (\mathbf{x}_{0,i})_{i=1,\dots,m} \in \mathbb{R}^N, \mathbf{v}_0 = (\sum_{i=1}^m \mathbf{L}_i^* \mathbf{L}_i)^{-1} \sum_{i=1}^m \mathbf{L}_i^* \mathbf{x}_{0,i} \\
 \text{Pour tout } k \in \mathbb{N}, \gamma_k > 0, \lambda_k \in]0, 2] \\
 \text{Pour } k = 0, 1, \dots \\
 \left[\begin{array}{l}
 \mathbf{y}_{k,i} = \text{prox}_{\gamma_k g_i}(\mathbf{x}_{k,i}) \quad (i \in \{1, \dots, m\}) \\
 \mathbf{c}_k = (\sum_{i=1}^m \mathbf{L}_i^* \mathbf{L}_i)^{-1} \sum_{i=1}^m \mathbf{L}_i^* \mathbf{y}_{k,i} \\
 \mathbf{x}_{k+1,i} = \mathbf{x}_{k,i} + \lambda_k (\mathbf{L}_i (2\mathbf{c}_k - \mathbf{v}_k) - \mathbf{y}_{k,i}) \quad (i \in \{1, \dots, m\}) \\
 \mathbf{v}_{k+1} = \mathbf{v}_k + \lambda_k (\mathbf{c}_k - \mathbf{v}_k)
 \end{array} \right.
 \end{array} \right.$$

La suite $(\mathbf{v}_k)_{k \in \mathbb{N}}$ converge vers une solution de (2.20). Les opérateurs proximaux dans cet algorithme peuvent être facilement calculés d'une manière parallèle. Le PPXA+ présente une forme particulièrement intéressante de l'algorithme Douglas-Rachford et il a été utilisé dans de récents travaux consacrés à la restauration d'image par exemple [Pustelnik 2011].

2.4.3.4 Primal-Dual

Intéressons nous à une forme du problème (2.18) où la fonction g est une composition d'une fonction nécessairement lisse et d'un opérateur linéaire (2.21), Chambolle et Pock ont proposé un nouvel algorithme efficace qui vise à résoudre ce problème (2.21) ainsi que son dual [Chambolle 2011] :

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} \quad f(\mathbf{x}) + g(\mathbf{L}\mathbf{x}) \quad (2.21)$$

avec $\mathbf{L} \in \mathbb{R}^{M \times N}$, $f \in \Gamma_0(\mathbb{R}^N)$ et $g \in \Gamma_0(\mathbb{R}^M)$.

La résolution du problème (2.21) s'effectue avec l'algorithme Primal-Dual donné par :

Algorithme 5 Primal-Dual

$$\left[\begin{array}{l}
 \mathbf{x}_0 \in \mathbb{R}^N, \mathbf{y}_0 \in \mathbb{R}^M, \lambda_k > 0 \\
 \text{Pour tout } k \in \mathbb{N}, \gamma_k > 0, \kappa\nu \|\mathbf{L}\|^2 < 1 \\
 \bar{\mathbf{x}}_0 = \mathbf{x}_0 \\
 \text{Pour } k = 0, 1, \dots \\
 \left[\begin{array}{l}
 \mathbf{y}_{k+1} = \text{prox}_{\gamma_k g^*}(\mathbf{y}_k + \kappa \mathbf{L} \bar{\mathbf{x}}_k) \\
 \mathbf{x}_{k+1} = \text{prox}_f(\mathbf{x}_k - \nu \mathbf{L}^T \mathbf{y}_{k+1}) \\
 \bar{\mathbf{x}}_{k+1} = \mathbf{x}_{k+1} + \lambda_k (\mathbf{x}_{k+1} - \mathbf{x}_k)
 \end{array} \right.
 \end{array} \right.$$

Cet algorithme nécessite le calcul des opérateurs proximaux de la fonction f et g^* . On désigne par g^{*e} la conjuguée au sens de Fenchel-Rockafellar [Bauschke 2017] de g , qui vérifie la décomposition de Moreau [Moreau 1965] :

$$\text{prox}_{g^*}(\mathbf{x}) = \mathbf{x} - \text{prox}_g(\mathbf{x}) \quad (2.22)$$

Par la suite, on peut donc facilement calculer l'opérateur proximal du conjugué d'une fonction g à partir de son opérateur proximal. La convergence de la suite $(\bar{\mathbf{x}}_k)_{k \in \mathbb{N}}$ générée par l'algorithme Primal-Dual vers une solution de (2.21) est vérifiée à condition que $\kappa\nu \|\mathbf{L}\|^2 < 1$ telle qu'on désigne par $\|\mathbf{L}\|$ la plus grande valeur singulière de l'opérateur \mathbf{L} .

e. L'opérateur conjugué g^* de la fonction g est définie par $g^*(\mathbf{u}) = \sup_{\mathbf{x} \in \mathbb{R}^N} (\langle \mathbf{x}, \mathbf{u} \rangle - g(\mathbf{x}))$

2.4.4 Récapitulatif

Comme nous avons vu dans ce chapitre, un problème inverse mal posé peut être résolu en minimisant une fonction de coût tenant compte de la fidélité aux données et d'un *a priori* de régularité sur la solution. Nous avons présenté quelques algorithmes proximaux dont chacun présente des conditions que les fonctions mises en jeu doivent vérifier (dérivabilité, convexité, etc.). Un récapitulatif des conditions d'utilisations des algorithmes mentionnés dans la section précédente est donné dans le Tableau 2.1.

Algorithme	Problème	f	g
Forward-Backward	(2.18)	$\Gamma_0(\mathbb{R}^N)$ β -Lipschitz différentiable	$\Gamma_0(\mathbb{R}^N)$
Douglas-Rachford	(2.18)	$\Gamma_0(\mathbb{R}^N)$	$\Gamma_0(\mathbb{R}^N)$ non nécessairement différentiable
Primal-Dual	(2.21)	$\Gamma_0(\mathbb{R}^N)$	$\Gamma_0(\mathbb{R}^N) \circ$ opérateur linéaire
PPXA+	(2.20)	somme de fonctions convexes g_i composées avec des opérateurs linéaires L_i	

Tableau 2.1 – Conditions d'utilisations de quelques algorithmes proximaux

2.5 Conclusion

Le traitement d'information revient souvent à la résolution d'un problème inverse. La méthode de régularisation et la méthode contrainte permettent de trouver une solution efficace au problème inverse en faisant intervenir un terme d'attache aux données et un terme de régularité. Comme nous l'avons présenté dans ce chapitre, les approches proximales apportent une réponse satisfaisante pour résoudre les problèmes d'optimisation résultants de ces formulations.

Nous allons exploiter ces approches dans la suite de ce manuscrit pour résoudre le problème inverse en biophysique. Nous ciblons deux applications : la DOSY (Diffusion Ordered Spectroscopy) et la SM (Spectrométrie de Masse). Ces deux applications donnent lieu à la résolution de problèmes inverses qui possèdent le caractère mal posé au sens de Hadamard. Nous détaillerons dans la suite de ce manuscrit chacune de ces deux applications et nous présenterons nos propositions pour résoudre les problèmes de restauration de signaux associés.

Chapitre 3

Résolution du problème inverse en DOSY

Sommaire

3.1	Introduction	22
3.2	La RMN DOSY	22
3.2.1	Phénomène de diffusion	22
3.2.2	Expérience RMN DOSY	23
3.3	Reconstruction du signal DOSY	24
3.3.1	Signal de diffusion	24
3.3.2	Positionnement du problème	25
3.4	État de l'art des méthodes de résolution	26
3.4.1	Méthode des moindres carrés contraints	26
3.4.2	Pénalisation entropique	26
3.4.3	Pénalisation de parcimonie	27
3.4.4	Autres approches	27
3.5	Régularisation hybride proposée	27
3.5.1	Entropie	28
3.5.2	Parcimonie	28
3.6	Opérateurs proximaux des pénalisations hybrides	30
3.6.1	Cas de l'Entropie de Shannon	30
3.6.2	Cas de l'Entropie de Burg	32
3.6.3	Algorithme d'optimisation	34
3.6.4	Version éclatée	34
3.7	Application à la RMN DOSY	35
3.7.1	Test de reconstruction	35
3.7.2	Test de robustesse	36
3.7.3	Comparaison avec l'approche éclatée	39
3.8	Conclusion	39

3.1 Introduction

Dans ce chapitre, nous focaliserons notre intérêt sur l'application DOSY. Nous allons exploiter les outils d'optimisation et les approches proximales que nous avons présentées dans le Chapitre 2 pour résoudre le problème inverse en DOSY. Le plan de ce chapitre est le suivant : nous allons tout d'abord présenter dans la Section 3.2 l'expérience DOSY. Puis, dans la Section 3.3, nous présenterons le modèle d'acquisition et le mauvais conditionnement que l'opérateur de mesure possède. Dans la Section 3.4, nous citerons les méthodes proposées pour restaurer les signaux DOSY. Par la suite, nous proposerons dans la Section 3.5 une nouvelle régularisation hybride qui combine l'entropie et la parcimonie. La Section 3.6 sera consacrée à présenter les opérateurs proximaux des pénalisations hybrides que nous avons proposées. Finalement, une analyse comparative de nos pénalisations hybrides sur un exemple de restauration de signaux DOSY sera présentée dans la Section 3.7.

3.2 La RMN DOSY

3.2.1 Phénomène de diffusion

En 1827, le botaniste anglais Robert Brown a remarqué l'existence de différents mouvements désordonnés à l'intérieur des grains d'une espèce de fleur sauvage nord-américaine nommée "Clarkia pulchella" observés au microscope. Ne trouvant aucune explication à ce phénomène, ce mouvement aléatoire et incessant a suscité l'intérêt de plusieurs physiciens. Début 1877, une théorie complète a été établie pour confirmer que toute particule immergée dans un fluide et qui n'est soumise à aucune interaction autre que des chocs avec les molécules du fluide environnant, suit un mouvement chaotique qu'on appelle le mouvement brownien. Ce mouvement a été bien décrit par le physicien français Jean Perrin qui a tracé la trajectoire d'une particule en suspension dans l'eau observée au microscope [Van der Pas 1971, Rivet 1997] en disant : "*Ils vont et viennent en tournoyant, montent, descendent, remontent encore, sans tendre aucunement vers le repos*" [Van der Pas 1971].

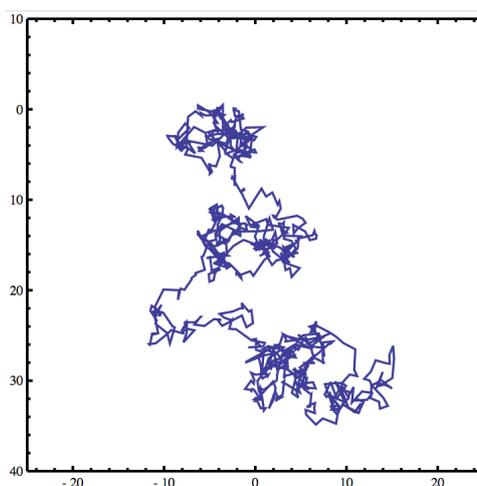


FIGURE 3.1 – Mouvement brownien d'une particule microscopique en suspension dans l'eau. (D'après un dessin de Jean Perrin).

Exclusivement lié à ce mouvement brownien, un phénomène de diffusion définit le transport de la matière et le mouvement aléatoire conduit par l'énergie interne du système. Ce phénomène est décrit par un coefficient de diffusion D qui suit l'équation de Stokes-Einstein [Einstein 1905] :

$$D = \frac{kT}{6\pi\eta r_h} \quad (3.1)$$

tels que k désigne la constante de Boltzmann, T la température, η la viscosité dynamique, et r_h le rayon hydrodynamique de la molécule supposée sphérique. L'utilisation des coefficients de diffusion est un outil efficace pour l'analyse et l'interprétation de la taille et de la forme des molécules dans un mélange complexe [Wilkins 1999, Gomez 2008] ainsi que de la polydispersité des polymères [Viéville 2011].

3.2.2 Expérience RMN DOSY

Les coefficients de diffusion peuvent être efficacement mesurés par RMN (Résonance Magnétique Nucléaire) en utilisant des gradients du champ magnétique. En fait, tout atome possède une propriété quantique qu'on appelle "spin". Le spin peut prendre différentes valeurs d'entiers (0, 1, 2, etc.) ou semi-entiers (1/2, 3/2, etc.), citons l'exemple de l'atome d'Hydrogène qui possède un spin égal à 1/2. Certains atomes possèdent des spins nuls comme le Carbone 12 et l'Oxygène 16. En présence d'un champ magnétique, un spin non nul donne naissance à un moment magnétique $\vec{\mu}$. Le principe de la RMN est de mettre le spin en mouvement autour d'un champ magnétique \vec{B}_0 en exerçant un champ magnétique résonant \vec{B}_1 . Le mouvement du moment magnétique, souvent décrit comme un mouvement de toupie (Figure 3.2), est appelé le mouvement de précession caractérisé par la fréquence de Larmor définie par : $w_0 = \gamma \vec{B}_0$, exprimée en radian s^{-1} telle que γ définit le rapport gyromagnétique (exprimé en radian $T^{-1} s^{-1}$). La perturbation du spin provoque une déviation de la trajectoire du moment $\vec{\mu}$, suivie d'une perte d'aimantation jusqu'à la disparition totale de l'aimantation et le retour à l'équilibre. Ce retour accompagné par une émission d'ondes électromagnétiques représente la mesure cible de la RMN.

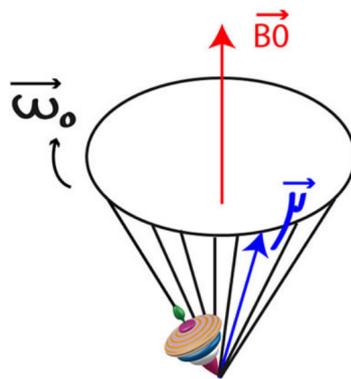


FIGURE 3.2 – Spin en mouvement de précession de Larmor autour d'un champ magnétique (\vec{B}_0)

La DOSY (Diffusion Ordered Spectroscopy), comme son nom l'indique est une mesure des coefficients de diffusion par l'effet du déplacement des molécules dans un gradient du champ magnétique. La méthode DOSY la plus classique est celle proposée par Stejskal et Tanner [Stejskal 1965]. Comme le montre la Figure 3.3, une première impulsion à 90° est appliquée, suivie par un champ

de gradient d'intensité g et de durée δ , ce qui entraîne un déphasage progressif de valeur différente selon la position du spin. Une deuxième impulsion à 180° sera ensuite appliquée pour induire un écho de spin en inversant la phase des aimantations. Finalement, un second champ de gradient similaire au premier (même intensité, même durée), sera appliqué pour refocaliser l'aimantation. Les spins résonnent pendant toute la durée Δ qui sépare les deux impulsions du gradient.

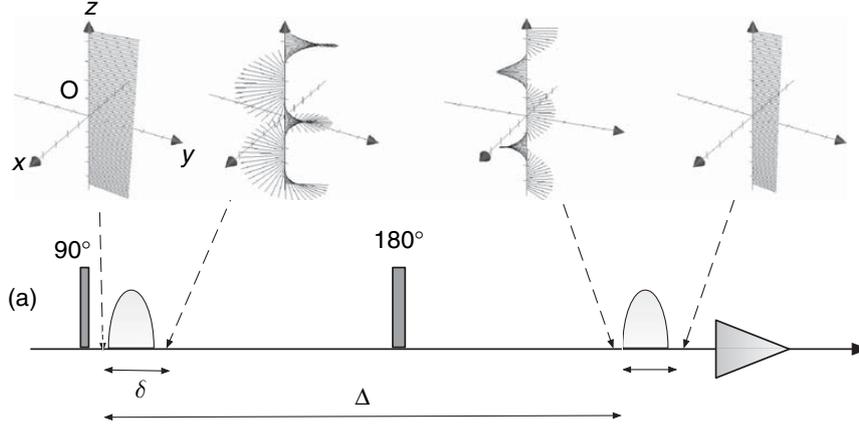


FIGURE 3.3 – Séquence de la DOSY [Gilard 2008]

L'atténuation de l'intensité I du signal mesuré est décrite par :

$$I = I_0 \exp(-D\gamma^2 g^2 \delta^2 \Delta) \quad (3.2)$$

avec D : le coefficient de diffusion, γ : le rapport gyromagnétique, g : l'intensité du gradient, et δ sa durée [Stejskal 1965, Sinnaeve 2012]. On désigne par I_0 l'intensité du signal en absence des gradients appliqués. L'intensité I s'exprime souvent en fonction de la quantité q^2 telle que :

$$I = I_0 \exp(-D\Delta q^2) \quad (3.3)$$

où $q = \gamma g \delta$.

L'expérience DOSY introduite par Johnson dans [Johnson 1999] est une présentation 2D de la mesure I telle que l'axe horizontal représente le déplacement chimique des spins et l'axe vertical présente les coefficients de diffusion D . Contrairement aux différentes expériences RMN qui utilisent la transformée de Fourier pour l'analyse spectrale, l'expérience DOSY utilise la transformée de Laplace car le signal enregistré dans le temps suit une loi exponentielle.

3.3 Reconstruction du signal DOSY

3.3.1 Signal de diffusion

L'intensité I d'une distribution de coefficients de diffusion $x(D)$ peut être définie par :

$$I(q) = \int_{D_{\min}}^{D_{\max}} x(D) \exp(-D\Delta q^2) dD \quad (3.4)$$

On peut modéliser la distribution $x(D)$ par une fonction gaussienne ou toute autre forme analytique symétrique ou asymétrique adaptée aux mesures expérimentales [Williamson 2016]. Dans le cas où la distribution contient plusieurs massifs isolés avec différentes formes, le choix de la forme paramétrique doit être bien adapté. Cette approche par modélisation est très sensible et elle dépend de la nature des données mesurées :

- Données monodisperses : Un échantillon est dit monodisperse s'il est composé de molécules ayant toutes le même coefficient de diffusion.
- Données polydisperses : Un échantillon est dit polydisperse s'il est formé de différentes molécules de différents coefficients de diffusion.

Les systèmes polydisperses les plus couramment rencontrés sont les polymères. Synthétiser des molécules polymériques ayant toutes les mêmes masses n'est pas faisable en pratique. Pour caractériser la polydispersité, on définit deux masses : M_n qui présente la masse moléculaire moyenne en nombre et M_w la masse moléculaire moyenne en poids. Le rapport de ces deux masses M_w/M_n nommé indice de polydispersité permet de caractériser la dispersion des masses. Si cet indice est égale à 1, le polymère est dit monodisperse car sa masse moléculaire moyenne en nombre est égale à sa masse moléculaire moyenne en poids. Sinon, le polymère est polydisperse [Viéville 2011].

3.3.2 Positionnement du problème

Déterminer la distribution des coefficients de diffusion $x(D)$ à partir de l'intensité I selon l'équation (3.4) revient à résoudre l'inverse de la transformée de Laplace en q^2 .

Rappelons que la transformée de Laplace d'une fonction f à variable réelle t , à support positif, est définie par :

$$F(p) = \mathcal{L}\{f(t)\} = \int_{t_{\min}}^{t_{\max}} \exp(-pt) f(t) dt \quad (3.5)$$

Une forme discrète du modèle (3.4) peut être décrite selon l'équation (3.6) telle qu'on suppose qu'une série de M mesures d'intensité y_m est effectuée pour toute valeur de q :

$$(\forall m \in \{1, \dots, M\}) \quad y_m = \sum_{n=1}^N x_n \exp(-D_n \Delta q_m^2) \quad (3.6)$$

avec D_n qui varie de D_{\min} à D_{\max} . L'expression matricielle est :

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{b} \quad (3.7)$$

telle que $\mathbf{y} = \{y_m, 1 \leq m \leq M\}$ représente la série mesurée, $\mathbf{x} = \{x_n, 1 \leq n \leq N\}$ représente la distribution de diffusion, $\mathbf{H} = (\mathbf{H}_{m,n})_{1 \leq m \leq M, 1 \leq n \leq N}$ est une matrice de taille $(M \times N)$ telle que $\mathbf{H}_{m,n} = \exp(-D_n \Delta q_m^2)$ et $\mathbf{b} = \{b_m, 1 \leq m \leq M\}$ un bruit supposé additif. Notons que \mathbf{x} présente le spectre de Laplace de la mesure \mathbf{y} via l'opérateur de Laplace \mathbf{H} .

L'opérateur de Laplace est difficile à inverser numériquement en particulier dans le cas $N > M$. L'inversion directe de \mathbf{H} ne conduit pas à une solution satisfaisante. La résolution de ce problème inverse requiert l'appel aux méthodes régularisées.

3.4 État de l’art des méthodes de résolution

Comme nous l’avons vu au début de ce chapitre, l’expérience DOSY vise à déterminer les coefficients de diffusion des molécules dans un échantillon. Le profil d’une mesure DOSY est toujours positif et il varie selon la nature des molécules. Dans le cas de monodispersité, le signal mesuré est un signal parcimonieux. Sinon, les molécules polydisperses possèdent une distribution lisse multimodale [Gostan 2004]. Par conséquent, une information *a priori* autre que la positivité peut être exploitée.

3.4.1 Méthode des moindres carrés contraints

Pour restaurer les signaux DOSY et déterminer la distribution de diffusion \mathbf{x} à partir des mesures \mathbf{y} , l’algorithme CONTIN (l’abréviation de CONTINuous diffusion coefficient) a été proposé par S.W.Provencher en 1979 [Provencher 1979, Provencher 1982a]. Cet algorithme présente une solution pour la résolution de différents problèmes inverses comme l’inversion de Laplace, l’inversion de la transformée de Fourier, etc. CONTIN a été présenté sous la forme d’un logiciel de traitement de données formé de plus de 60 sous programmes [Provencher 1982b] permettant d’assurer l’inversion des données représentées par des équations linéaires algébriques ou intégrales. Il est basé sur la méthode des moindres carrés pondérés par l’incertitude sur les données sous une contrainte de positivité. Cet outil a conduit à des résultats satisfaisants pour différentes distributions monodisperses et polydisperses. Cependant, en pratique il est très sensible au bruit. De plus, il n’est accessible que sous forme d’une boîte noire dont les sources ne sont pas disponibles ce qui nuit à la fiabilité et l’interprétabilité des résultats. Cela confirme le besoin d’une nouvelle approche de résolution plus robuste et reposant sur des fondements théoriques solides.

3.4.2 Pénalisation entropique

L’utilisation de la fonction d’entropie comme une fonction de régularisation est une stratégie très utile pour la reconstruction des signaux dans différentes applications, notamment la reconstruction d’images astronomiques [Pichon 1998], la tomographie à rayon X [Djafari 1988], l’acquisition comprimée [Foucart 2013], l’apprentissage supervisé [He 2001], etc. Aussi, le choix de cette mesure a été proposé dans [Delsuc 1989, Delsuc 1998] pour estimer une distribution de diffusion à partir des signaux DOSY. Le problème sera donné par :

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} \quad \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 + \lambda \text{ent}(\mathbf{x}) \quad (3.8)$$

où λ définit le paramètre de régularisation et la fonction ent est définie en (2.5). Cette formulation peut être résolue avec la stratégie itérative appelée MaxEnt, qui converge vers une solution stable [Skilling 1984]. En fait, la transformée de Laplace du profil de diffusion estimé permet de calculer une décroissance théorique. La distance entre cette décroissance et la décroissance des données expérimentales permet de générer l’estimation du profil de diffusion à l’itération suivante. L’algorithme réitère cette stratégie jusqu’à ce que cette distance soit très faible, généralement inférieure au niveau de bruit ce qui confirme la robustesse envers les forts niveaux de bruit. Néanmoins, le taux de convergence de cet algorithme reste lent, et une amélioration utilisant des approches modernes d’optimisation convexe peut être envisagée.

3.4.3 Pénalisation de parcimonie

Toujours avec l'approche de régularisation, mais cette fois en exploitant la notion de parcimonie, le problème peut être formulé comme :

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} \quad \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 + \lambda \ell_1(\mathbf{x}) \quad (3.9)$$

Cette norme ℓ_1 , définie en (2.8) est souvent utilisée pour la restauration des signaux parcimonieux [Donoho 2003]. En 2013, ITAMeD (l'abréviation de Iterative Thresholding Algorithm for Multiexponential Decay) [Kazmierczuk 2012], un algorithme basé sur le principe de régularisation avec la norme ℓ_1 a été proposé pour restaurer les signaux DOSY. Comparé à CONTIN et MaxEnt, cet algorithme présente de meilleurs résultats si les données sont monodisperses ce qui est tout à fait attendu puisque la norme ℓ_1 favorise la parcimonie, ce qui caractérise les signaux DOSY monodisperses. Cependant, l'application de cet algorithme reste limitée aux données monodisperses et peut ne pas être efficace dans le cas général des données polydisperses.

3.4.4 Autres approches

Partant de la régularisation avec la norme ℓ_1 , Kazmierczuk et al. ont proposé une généralisation de leur méthode [Urbańczyk 2013] avec le modèle suivant :

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} \quad \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 + \lambda \ell_p^p(\mathbf{x}) \quad (3.10)$$

La norme ℓ_p définie en (2.9) est généralement utilisée dans les applications relevant de l'estimation robuste quelque soit la valeur de p [Yarlagadda 1985], surtout le fameux cas où $p = 1$ proposé précédemment. Même si cette proposition semble améliorer la qualité de reconstruction des signaux DOSY quelques soient leurs types, le choix de la bonne valeur de p pour chaque type du signal monodisperse ou polydisperse reste un problème "ad-hoc".

Pour avoir un algorithme capable de restaurer tous les signaux DOSY (monodisperse ou polydisperse), Xu et al. ont proposé une stratégie de régularisation itérative basée sur une approche de régions de confiance [Powell 1975] nommée TRAIIn (l'abréviation de Trust Region Algorithm for Inversion of molecular diffusion NMR) [Xu 2013]. Cet algorithme dépasse les difficultés numériques de Laplace et permet d'avoir des résultats satisfaisants avec les mélanges polydisperses. Néanmoins, la délicatesse du choix du coefficient de confiance à chaque itération, la sensibilité des résultats au niveau de bruit et la difficulté à régler la stratégie d'arrêt ("early stopping") restent des points faibles de cette stratégie.

3.5 Régularisation hybride proposée

Nous proposons de résoudre le problème (3.7) en utilisant l'approche de minimisation sous contrainte avec un choix hybride du terme de régularité selon le modèle suivant :

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} \quad \Psi(\mathbf{x}) \quad \text{tel que } \|\mathbf{H}\mathbf{x} - \mathbf{y}\| \leq \eta \quad (3.11)$$

tel que le terme hybride Ψ définit la combinaison entre deux termes de régularité et η un paramètre d'estimation lié au bruit. La fonction Ψ sera donc une combinaison linéaire de deux fonctions Ψ_1

et Ψ_2 telles que :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \Psi(\mathbf{x}) = \alpha \Psi_1(\mathbf{x}) + \beta \Psi_2(\mathbf{x}) \quad (3.12)$$

Les coefficients α et β représentent les coefficients positifs de régularisation qui assurent l'équilibre entre les deux termes de régularité et qui font le compromis avec le terme de fidélité.

Le choix de ces deux fonctions Ψ_1 et Ψ_2 doit être bien adapté afin de bien reconstruire l'information. Nous nous proposons dans ce cadre de faire la combinaison entre deux pénalités efficaces du point de vue théorique et physique : la fonction d'entropie et celle de parcimonie.

La fonction Ψ sera choisie séparable, c'est à dire :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \Psi(\mathbf{x}) = \sum_{n=1}^N \psi(x_n) \quad (3.13)$$

avec

$$(\forall x_n \in \mathbb{R}) \quad \psi(x_n) = \alpha \psi_1(x_n) + \beta \psi_2(x_n) \quad (3.14)$$

La fonction $\psi_1 : \mathbb{R} \mapsto]-\infty, +\infty]$ désigne la fonction basée sur l'entropie et $\psi_2 : \mathbb{R} \mapsto \mathbb{R}$ définit la fonction de parcimonie que nous présenterons dans la suite.

3.5.1 Entropie

Toujours pour mesurer la quantité d'information, Shannon définit la fonction d'entropie à maximiser [Shannon 1948] dans le but d'avoir plus d'informations sur l'objet cible. Cela revient à utiliser (3.13) avec $\beta = 0$:

$$(\forall x \in \mathbb{R}) \quad \psi_1(x) = \begin{cases} x \log(x) & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ +\infty & \text{sinon} \end{cases} \quad (3.15)$$

Une autre mesure de la quantité d'information est l'entropie de Burg. Cette fonction présente un critère de pénalité utilisé dans différentes applications [Djafari 1994] qui revient à utiliser (3.13) avec $\beta = 0$ et :

$$(\forall x \in \mathbb{R}) \quad \psi_1(x) = \begin{cases} -\log(x) & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \quad (3.16)$$

L'entropie de Burg est utilisée comme une fonction de régularisation [Nityananda 1982] et permet d'imposer la positivité du signal restauré, comme l'entropie de Shannon.

La Figure 3.4 présente les deux fonctions d'entropie (3.15) et (3.16).

3.5.2 Parcimonie

Nous avons présenté dans le Chapitre 2, les fonctions de parcimonie les plus utilisées telles que la pseudo-norme ℓ_0 (2.7) et la norme ℓ_1 (2.8). Dans ce cas, on se retrouve à minimiser (3.13) avec $\alpha = 0$ et

$$(\forall x \in \mathbb{R}) \quad \psi_2(x) = \begin{cases} 1 & \text{si } x \neq 0 \\ 0 & \text{sinon} \end{cases} \quad (3.17)$$

dans le cas de la pénalisation ℓ_0 ou :

$$(\forall x \in \mathbb{R}) \quad \psi_2(x) = |x| \quad (3.18)$$

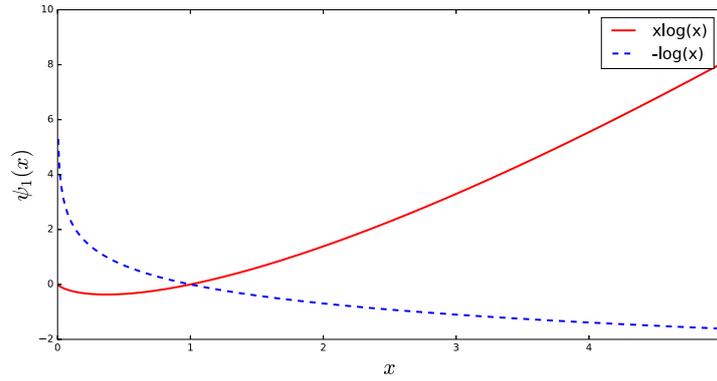


FIGURE 3.4 – Fonctions d’entropie (Shannon et Burg) pour un scalaire $x \in \mathbb{R}$. ($\psi_1(x) \mapsto +\infty$ pour $x < 0$)

dans le cas de la norme ℓ_1 .

Bien que la norme ℓ_1 et la pseudo-norme ℓ_0 soient les meilleures mesures de parcimonie, d’autres fonctions ont été proposées citons l’exemple de la pénalité log-sum [Antoniadis 2002] :

$$(\forall x \in \mathbb{R}) \quad \psi_2(x) = \log(|x| + \delta) \quad (3.19)$$

avec $\delta > 0$, et aussi la pénalité de Cauchy obtenue avec [Wipf 2010] :

$$(\forall x \in \mathbb{R}) \quad \psi_2(x) = \log(x^2 + \delta) \quad (3.20)$$

La Figure 3.5 présente les quatre pénalités de parcimonie décrites en (3.18), (3.17), (3.19), et (3.20). Il est aisé de remarquer que la pseudo-norme ℓ_0 présente la meilleure mesure de parcimonie. Cependant, cette mesure est discontinue et non convexe ce qui conduit systématiquement à des complexités numériques pour la résolution d’un problème d’optimisation. Les autres pénalités présentent des mesures de parcimonie plus simples et qui peuvent être efficaces, chacune selon ses caractéristiques (différentiabilité, continuité, convexité).

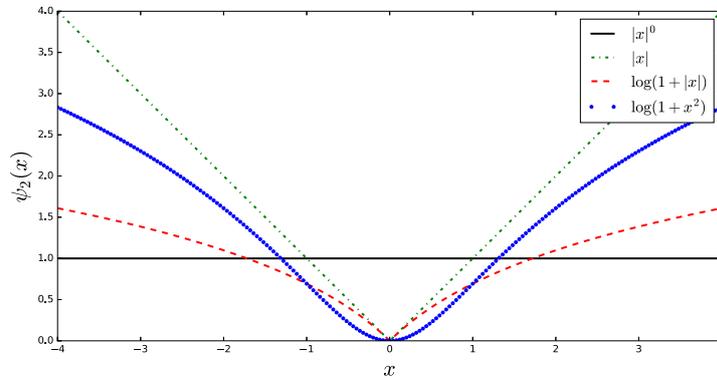


FIGURE 3.5 – Fonctions de parcimonie ($\ell_0 = |x|^0$ (3.17), $\ell_1 = |x|$ (3.18), log-sum= $\log(|x| + \delta)$ (3.19) et Cauchy= $\log(x^2 + \delta)$ (3.20) avec $\delta = 1$).

3.6 Opérateurs proximaux des pénalisations hybrides

La régularisation hybride que nous proposons dans l'équation (3.12) fait la combinaison linéaire d'une des fonctions d'entropie définies dans les équations (3.15) et (3.16) avec l'une des fonctions de parcimonie définies dans (3.18), (3.17), (3.20) et (3.19).

Pour calculer la solution de (3.11) pour chaque régularisation hybride et résoudre le problème inverse, nous utilisons la famille d'approche d'optimisation proximale nécessitant le calcul de l'opérateur proximal de Ψ . Grâce au caractère de séparabilité que la fonction de régularisation Ψ possède, l'opérateur proximal peut être exprimé par :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \text{prox}_{\Psi}(\mathbf{x}) = (\text{prox}_{\psi}(x_n))_{1 \leq n \leq N} \quad (3.21)$$

Dans la suite, nous présenterons les opérateurs proximaux calculés pour chaque fonction de régularisation hybride proposée. Tous les détails du calcul des opérateurs proximaux suivants sont dans l'annexe A.

3.6.1 Cas de l'Entropie de Shannon

3.6.1.1 Entropie de Shannon

L'opérateur proximal de la fonction (3.15) a la forme [Combettes 2011] :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{\alpha\psi_1}(x) = \alpha \mathcal{W} \left(\frac{1}{\alpha} \exp \left(\frac{x}{\alpha} - 1 \right) \right) \quad (3.22)$$

où $\alpha > 0$ et on désigne par \mathcal{W} la fonction de Lambert [Corless 1996], nommée aussi fonction Oméga (Figure 3.6) définie comme étant l'inverse de la fonction $f : z \rightarrow ze^z$ pour tout $z \in \mathbb{C}$, i.e :

$$z = we^w \Leftrightarrow w = \mathcal{W}(z) \quad (3.23)$$

Cette fonction a été très utile récemment pour résoudre différentes problématiques [Chierchia 2017, Briceno 2017] et elle possède plusieurs propriétés que nous allons exploiter dans le Chapitre 4.

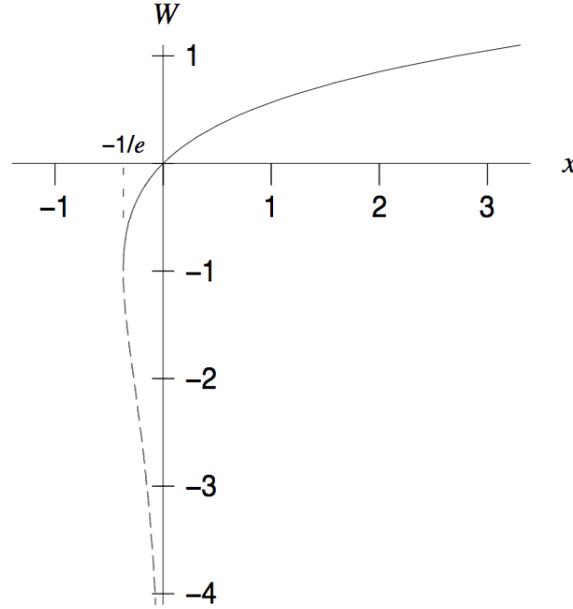
3.6.1.2 Entropie de Shannon + ℓ_0

La fonction de régularisation faisant la combinaison entre l'entropie de Shannon (3.15) et la pseudo-norme ℓ_0 (3.17) pour tout $(\alpha, \beta) \in]0, +\infty[^2$ est définie par :

$$(\forall x \in \mathbb{R}) \quad \psi(x) = \begin{cases} \alpha x \log x + \beta & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ +\infty & \text{sinon} \end{cases} \quad (3.24)$$

Cette régularisation possède un opérateur proximal de la forme :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{\psi}(x) = \begin{cases} p & \text{si } \beta < \bar{\beta} \\ \{0, p\} & \text{si } \beta = \bar{\beta} \\ 0 & \text{sinon} \end{cases} \quad (3.25)$$


 FIGURE 3.6 – Fonction \mathcal{W} de Lambert sur $[-\frac{1}{e}, +\infty[$ [Corless 1996]

avec

$$\bar{\beta} = \frac{1}{2}p^2 + \alpha p \in]0, +\infty[\quad \text{et} \quad p = \alpha \mathcal{W} \left(\frac{1}{\alpha} \exp \left(\frac{x}{\alpha} - 1 \right) \right) \quad (3.26)$$

Suite à la discontinuité et la non convexité de la pseudo-norme ℓ_0 , la fonction de régularisation présente une discontinuité en 0. De ce fait, l'opérateur proximal associé possède plusieurs valeurs quand $\beta = \bar{\beta}$.

3.6.1.3 Entropie de Shannon + ℓ_1

Pour tout $(\alpha, \beta) \in]0, +\infty[^2$, la régularisation hybride de l'entropie de Shannon (3.15) combinée à la norme de parcimonie ℓ_1 (3.18) définie par :

$$(\forall x \in \mathbb{R}) \quad \psi(x) = \begin{cases} \alpha x \log x + \beta x & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ +\infty & \text{sinon} \end{cases} \quad (3.27)$$

possède l'opérateur proximal donné par [Cherni 2017] :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{\psi}(x) = \alpha \mathcal{W} \left(\frac{1}{\alpha} \exp \left(\frac{x - \beta}{\alpha} - 1 \right) \right) \quad (3.28)$$

3.6.1.4 Entropie de Shannon + $f_{\log\text{-Sum}}$

La régularisation hybride entre l'entropie de Shannon (3.15) et la fonction log-sum (3.19) est définie pour tout $\alpha > 0$ et $\beta > 0$ par :

$$(\forall x \in \mathbb{R}) \quad \psi(x) = \begin{cases} \alpha x \log x + \beta \log(\delta + x) & \text{si } x > 0 \\ \beta \log(\delta) & \text{si } x = 0 \\ +\infty & \text{sinon} \end{cases} \quad (3.29)$$

Pour tout $(x \in \mathbb{R})$, la fonction donnée par (3.29) possède l'opérateur proximal :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{\psi}(x) = \underset{p \in]0, +\infty[\text{ tel que } \varphi(p)=0}{\text{Argmin}} \left(\frac{1}{2}(x-p)^2 + \psi(p) \right) \quad (3.30)$$

avec

$$\varphi(p) = p^2 + (\delta - x + \alpha)p + \alpha(\delta + p) \log(p) + \delta(\alpha - x) + \beta \quad (3.31)$$

3.6.1.5 Entropie de Shannon + f_{Cauchy}

En faisant la combinaison linéaire entre l'entropie de Shannon (3.15) et la fonction de Cauchy (3.20), la fonction de régularisation possède la forme suivante pour tout $(\alpha, \beta, \delta) \in]0, +\infty[^3$:

$$(\forall x \in \mathbb{R}) \quad \psi(x) = \begin{cases} \alpha x \log x + \beta \log(\delta + x^2) & \text{si } x > 0 \\ \beta \log(\delta) & \text{si } x = 0 \\ +\infty & \text{sinon} \end{cases} \quad (3.32)$$

L'opérateur proximal de cette fonction de régularisation est :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{\psi}(x) = \underset{p \in]0, +\infty[\text{ tel que } \varphi(p)=0}{\text{Argmin}} \left(\frac{1}{2}(x-p)^2 + \psi(p) \right) \quad (3.33)$$

avec

$$\varphi(p) = p^3 + (\alpha - x)p^2 + (\delta + 2\beta)p + \alpha(\delta + p^2) \log(p) + \delta(\alpha - x) \quad (3.34)$$

3.6.2 Cas de l'Entropie de Burg

Dans cette partie, nous présentons les opérateurs proximaux de la fonction de régularisation faisant intervenir l'entropie de Burg combinée aux différentes normes de parcimonie proposées dans la Section (3.5.2).

3.6.2.1 Entropie de Burg

Pour tout $(x \in \mathbb{R}, \alpha \in]0, +\infty[)$, comme il a été prouvé dans [Chaux 2007], l'opérateur proximal de la fonction (3.16) est :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{\alpha\psi_1}(x) = \frac{x + \sqrt{x^2 + 4\alpha}}{2} \quad (3.35)$$

3.6.2.2 Entropie de Burg + ℓ_0

Dans le cas où $(\alpha, \beta) \in]0, +\infty[^2$, la régularisation hybride de l'entropie de Burg (3.16) combinée à la pseudo-norme ℓ_0 (2.7) est définie par :

$$(\forall x \in \mathbb{R}) \quad \psi(x) = \begin{cases} -\alpha \log x + \beta & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \quad (3.36)$$

L'opérateur proximal de cette régularisation pour tout $x \in \mathbb{R}$, a la forme suivante :

$$\text{prox}_\psi(x) = \frac{x + \sqrt{x^2 + 4\alpha}}{2} \quad (3.37)$$

3.6.2.3 Entropie de Burg + ℓ_1

Avec la norme ℓ_1 (3.18), on définit la régularisation hybride pour tout $(\alpha, \beta) \in]0, +\infty[^2$ par :

$$(\forall x \in \mathbb{R}) \quad \psi(x) = \begin{cases} -\alpha \log x + \beta x & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \quad (3.38)$$

Cette fonction possède l'opérateur proximal donné par [Chaux 2007] :

$$\text{prox}_\psi(x) = \frac{x - \beta + \sqrt{(\beta - x)^2 + 4\alpha}}{2} \quad (3.39)$$

3.6.2.4 Entropie de Burg + $\mathbf{f}_{\log\text{-Sum}}$

Pour tout $(\alpha, \beta) \in]0, +\infty[^2$, on définit :

$$(\forall x \in \mathbb{R}) \quad \psi(x) = \begin{cases} -\alpha \log x + \beta \log(\delta + x) & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \quad (3.40)$$

L'opérateur proximal de cette fonction de régularisation hybride est :

$$\text{prox}_\psi(x) = \underset{p \in]0, +\infty[\text{ tel que } \varphi(p)=0}{\text{Argmin}} \left(\frac{1}{2}(x - p)^2 + \psi(p) \right) \quad (3.41)$$

avec

$$\varphi(p) = p^3 + (\delta - x)p^2 + p(\beta - \delta x - \alpha) - \delta\alpha \quad (3.42)$$

3.6.2.5 Entropie de Burg + $\mathbf{f}_{\text{Cauchy}}$

La dernière combinaison que nous proposons est celle entre l'entropie de Burg (3.16) et la norme de Cauchy (3.20) définie pour tout $(\alpha, \beta, \delta) \in]0, +\infty[^3$:

$$(\forall x \in \mathbb{R}) \quad \psi(x) = \begin{cases} -\alpha \log x + \beta \log(\delta + x^2) & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \quad (3.43)$$

L'ensemble des solutions peut être donné par l'opérateur proximal ayant la forme suivante :

$$\text{prox}_\psi(x) = \underset{p \in]0, +\infty[\text{ tel que } \varphi(p)=0}{\text{Argmin}} \left(\frac{1}{2}(x-p)^2 + \psi(p) \right) \quad (3.44)$$

avec

$$\varphi(p) = p^4 - xp^3 + (\delta + 2\beta - \alpha)p^2 - \delta xp - \delta\alpha \quad (3.45)$$

3.6.3 Algorithme d'optimisation

Avec ces opérateurs proximaux, le problème d'optimisation (3.11) peut être efficacement résolu en utilisant les approches proximales [Combettes 2011, Pesquet 2012, Pustelnik 2011]. Nous proposons par la suite d'utiliser l'algorithme PPXA+ décrit par l'Algorithme 4 de la Section 2.4. Dans notre contexte, il est donné par :

Algorithme 6 PPXA+ pour le problème (3.11) avec le choix (3.12)

$$\left[\begin{array}{l} \mathbf{v}^{(0,1)} \in \mathbb{R}^N, \mathbf{v}^{(0,2)} \in \mathbb{R}^M \\ \mathbf{x}^{(0)} = (\mathbf{I}_N + \mathbf{H}^\top \mathbf{H})^{-1} (\mathbf{v}^{(0,1)} + \mathbf{H}^\top \mathbf{v}^{(0,2)}) \\ \mathbf{B} = (\mathbf{I}_N + \mathbf{H}^\top \mathbf{H})^{-1} \\ \gamma \in (0, 2), \\ \text{Pour } k = 0, 1, \dots \\ \left[\begin{array}{l} \mathbf{z}^{(k,1)} = \text{prox}_\Psi(\mathbf{v}^{(k,1)}) \\ \mathbf{z}^{(k,2)} = \text{proj}_{\|\cdot - \mathbf{y}\| \leq \eta}(\mathbf{v}^{(k,2)}) \\ \mathbf{u}^{(k)} = \mathbf{B}(\mathbf{z}^{(k,1)} + \mathbf{H}^\top \mathbf{z}^{(k,2)}) \\ \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \gamma(\mathbf{u}^{(k)} - \mathbf{x}^{(k)}) \\ \mathbf{v}^{(k+1,1)} = \mathbf{v}^{(k,1)} + \gamma(2\mathbf{u}^{(k)} - \mathbf{x}^{(k)} - \mathbf{z}^{(k,1)}) \\ \mathbf{v}^{(k+1,2)} = \mathbf{v}^{(k,2)} + \gamma(\mathbf{H}(2\mathbf{u}^{(k)} - \mathbf{x}^{(k)}) - \mathbf{z}^{(k,2)}) \end{array} \right. \end{array} \right.$$

tel que \mathbf{I}_N désigne la matrice identité et $\text{proj}_{\|\cdot - \mathbf{y}\| \leq \eta}$ représente la projection sous contrainte définie par :

$$\text{proj}_{\|\cdot - \mathbf{y}\| \leq \eta}(\mathbf{z}) = \mathbf{z} + (\mathbf{z} - \mathbf{y}) \min \left(\frac{\eta}{\|\mathbf{z} - \mathbf{y}\|}, 1 \right) - \mathbf{y} \quad (\forall (\mathbf{y}, \mathbf{z}) \in (\mathbb{R}^N)^2) \quad (3.46)$$

Toute série $(\mathbf{x}^k)_{k \in \mathbb{N}}$ générée par cet algorithme converge vers une solution du problème (3.11) [Pesquet 2012] sous l'hypothèse que $\Psi \in \Gamma_0(\mathbb{R}^N)$ ^a. Cela inclut les pénalisations hybrides qui font la combinaison entre la norme ℓ_1 et l'entropie de Shannon (3.27) ainsi qu'entre la norme ℓ_1 et l'entropie de Burg (3.38). L'extension de la propriété de convergence dans le cas de Ψ non convexe reste un problème ouvert. Dans ce cas, d'autres algorithmes tels que le forward-backward (Algorithme 1), pourraient être envisagés car ils bénéficient de garanties dans le cadre non convexe [Attouch 2011].

3.6.4 Version éclatée

Comme nous l'avons mentionné dans la Section 3.6, le calcul d'un opérateur proximal est caractérisé par la séparabilité (3.21). Comme nous proposons une régularisation hybride Ψ faisant la combinaison entre deux fonctions séparables : la fonction basée sur l'entropie Ψ_1 , et la fonction de parcimonie Ψ_2 , le calcul de l'opérateur proximal peut être effectué de deux manières différentes : soit avec la fonction somme $\Psi = \alpha\Psi_1 + (1 - \alpha)\Psi_2$ comme nous l'avons fait pour l'Algorithme 6,

a. Rappelons que $\Gamma_0(\mathbb{R}^N)$ désigne la classe des fonctions $f : \mathbb{R}^N \rightarrow]-\infty, +\infty]$ convexes, semi-continues inférieurement et propres.

soit pour chacune des fonctions $\alpha\Psi_1$ et $(1 - \alpha)\Psi_2$ séparément.

Une nouvelle version éclatée, présentée dans l'algorithme suivant, peut être envisagée dans ce cas :

Algorithme 7 PPXA+ **éclaté** pour le problème (3.11) avec le choix (3.12)

$$\begin{array}{l}
 \mathbf{v}^{(0,1)} \in \mathbb{R}^N, \mathbf{v}^{(0,2)} \in \mathbb{R}^M, \mathbf{v}^{(0,3)} \in \mathbb{R}^M \\
 \mathbf{x}^0 = (2I + \mathbf{H}^T \mathbf{H})^{-1}(\mathbf{v}^{(0,1)} + \mathbf{v}^{(0,2)} + \mathbf{H}^T \mathbf{v}^{(0,3)}) \\
 \gamma \in (0, 2) \\
 \text{Pour } k = 0, 1, \dots \\
 \left[\begin{array}{l}
 \mathbf{z}^{(k,1)} = \text{prox}_{\Psi_1}(\mathbf{v}^{(k,1)}) \\
 \mathbf{z}^{(k,2)} = \text{prox}_{\Psi_2}(\mathbf{v}^{(k,2)}) \\
 \mathbf{z}^{(k,3)} = \text{proj}_{\|\cdot\|_2 \leq \eta}(\mathbf{v}^{(k,3)}) \\
 \mathbf{u}^{(k)} = (2I + \mathbf{H}^T \mathbf{H})^{-1}(\mathbf{z}^{(k,1)} + \mathbf{z}^{(k,2)} + \mathbf{H}^T \mathbf{z}^{(k,3)}) \\
 \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \gamma(\mathbf{u}^{(k)} - \mathbf{x}^{(k)}) \\
 \mathbf{v}^{(k+1,1)} = \mathbf{v}^{(k,1)} + \gamma(2\mathbf{u}^{(k)} - \mathbf{x}^{(k)} - \mathbf{z}^{(k,1)}) \\
 \mathbf{v}^{(k+1,2)} = \mathbf{v}^{(k,2)} + \gamma(2\mathbf{u}^{(k)} - \mathbf{x}^{(k)} - \mathbf{z}^{(k,2)}) \\
 \mathbf{v}^{(k+1,3)} = \mathbf{v}^{(k,3)} + \gamma(\mathbf{H}(2\mathbf{u}^{(k)} - \mathbf{x}^{(k)}) - \mathbf{z}^{(k,3)})
 \end{array} \right.
 \end{array}$$

où prox_{Ψ_1} représente l'opérateur proximal de Ψ_1 (la pénalité de l'entropie) et prox_{Ψ_2} définit l'opérateur proximal de Ψ_2 (la pénalité de parcimonie). Cette nouvelle version présente plus d'étapes dans le calcul algorithmique. Nous testerons sa complexité sur des exemples numériques dans la section suivante.

3.7 Application à la RMN DOSY

Nous avons proposé plusieurs combinaisons entre l'entropie et la parcimonie. À ce niveau, les opérateurs proximaux calculés peuvent être exploités pour résoudre le problème inverse mal posé de la RMN DOSY. Nous choisissons de nous concentrer sur le cas convexe. Par conséquent, seulement la fonction de régularisation faisant appel à l'entropie de Shannon ou de Burg combinée avec la norme ℓ_1 sera retenue. Nous utilisons l'algorithme 6 pour reconstruire les signaux DOSY.

3.7.1 Test de reconstruction

Pour simuler des signaux DOSY, nous considérons une mesure $\mathbf{y} = (y_1, y_2, \dots, y_M) \in \mathbb{R}^M$ liée au signal cible $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N) \in \mathbb{R}^N$ par la matrice $\mathbf{H} = (\exp(-d_n s_m^2))_{1 \leq m \leq M, 1 \leq n \leq N}$ via le modèle (3.7). L'échantillonnage des coefficients de diffusion $(d_n)_{1 \leq n \leq N}$ est donné par :

$$d_n = D_{\min} \exp\left(-\frac{n-1}{N} \log\left(\frac{D_{\min}}{D_{\max}}\right)\right) \quad (3.47)$$

pour tout $n \in \{1, \dots, N\}$ selon une série de mesure :

$$s_m = s_{\min} + \frac{m-1}{M} s_{\max} \quad (3.48)$$

pour $m \in \{1, \dots, M\}$.

Prenons quatre signaux synthétiques A, B, C et D tels que $M = 50$, $N = 200$, $D_{\min} = 1$, et $D_{\max} = 10^3$. Un bruit blanc \mathbf{b} additif gaussien de variance σ^2 est généré pour les quatre signaux ($\sigma = 10^{-3}$). Les trois signaux A, B et C modélisent une distribution polydisperse.

- Le signal A est formé de deux profils gaussiens localisés à $d = 8$ et $d = 62$, avec $s_{\min} = 0$ et $s_{\max} = 1$.
- Le signal B présente un profil gaussien centré en $d = 8$, tel que $s_{\min} = 0$ et $s_{\max} = 1.2$.
- Le signal C présente un profil gaussien centré en $d = 64$, avec $s_{\min} = 0$ et $s_{\max} = 0.25$.
- Le signal D présente une distribution quasi-monodisperse donnée par un profil gaussien centré en $d = 32$, tel que $s_{\min} = 0$ et $s_{\max} = 0.3$ comme le montre la Figure 3.7.

Nous appliquons l'algorithme PPXA+ (Algorithme 6) avec la régularisation hybride faisant la combinaison entre la fonction de parcimonie ℓ_1 et les deux fonctions d'entropie. Pour cela, le choix de α et β est effectué d'une manière manuelle pour minimiser l'erreur de reconstruction. En outre, nous proposons de résoudre (3.11) où $\eta = \xi\sqrt{M}\sigma$ et ξ est un coefficient positif proche de 1 [Trussell 1983, Chouzenoux 2010]. Nous discuterons de ce choix dans le Chapitre 4.

Plusieurs résultats de reconstruction sont présentés dans la Figure 3.8.

D'après les résultats de la Figure 3.8, nous remarquons que la pénalité ℓ_1 (le cas où $\alpha = 1$) n'assure pas une reconstruction des signaux A, B, C et D ce qui est attendu car ces signaux ne sont pas parcimonieux. L'entropie de Burg paraît efficace pour la reconstruction du signal B et le signal quasi-monodisperse D, mais pas pour les autres signaux polydispersés. De plus, la régularisation d'entropie de Shannon assure la reconstruction des différents signaux ce qui prouve l'efficacité de MaxEnt [Delsuc 1998]. Néanmoins, la combinaison entre l'entropie de Shannon et la norme ℓ_1 est très efficace par rapport aux autres pénalités pour les signaux polydispersés (A, B, C et D). La différence en qualité de reconstruction entre cette régularisation hybride et les autres pénalités proposées est claire visuellement. Par conséquent, nous choisissons la combinaison entre l'entropie de Shannon et la norme ℓ_1 comme la meilleure pénalisation. Pour vérifier la cohérence de cette interprétation, nous comparons les résultats de reconstruction avec les différentes pénalités en faisant varier le niveau de bruit comme il est présenté dans la section suivante.

3.7.2 Test de robustesse

Avec les mêmes signaux : A, B, C et D, pour différents niveaux de bruit (σ), nous proposons de calculer le RSB des résultats obtenus avec les différentes fonctions de pénalités.

Selon les résultats du Tableau 3.1, nous constatons que l'approche de régularisation hybride représente un outil efficace pour résoudre le problème inverse mal posé en DOSY. La combinaison entre l'entropie de Shannon et la norme de parcimonie ℓ_1 donne les valeurs de RSB les plus élevées pour différents niveaux de bruit. L'entropie de Shannon semble plus efficace pour reconstruire le signal A à $\sigma = 10^{-4}$. Cependant, elle présente un RSB équivalent à celui donné par la régularisation hybride de l'entropie de Shannon combinée à la norme ℓ_1 . De même pour l'entropie de Burg combinée à la norme ℓ_1 pour reconstruire le signal D à $\sigma = 10^{-3}$ et $\sigma = 10^{-4}$. Pour tous les autres niveaux de bruit et les autres signaux, la différence en qualité de reconstruction confirme que l'entropie de Shannon avec la norme ℓ_1 représente la meilleure fonction de pénalité. Par conséquent, nous proposons par la suite d'adopter ce choix de régularisation. Puisque cette stratégie dépend de certains paramètres comme α qui ajuste le compromis entre l'entropie de Shannon et la norme de parcimonie ℓ_1 , et le paramètre η qui définit la contrainte devant être respectée par le signal cible, le chapitre suivant s'intéresse donc à l'amélioration du choix de tous les paramètres afin de construire un nouvel algorithme robuste et assurant la restauration des données DOSY efficacement.

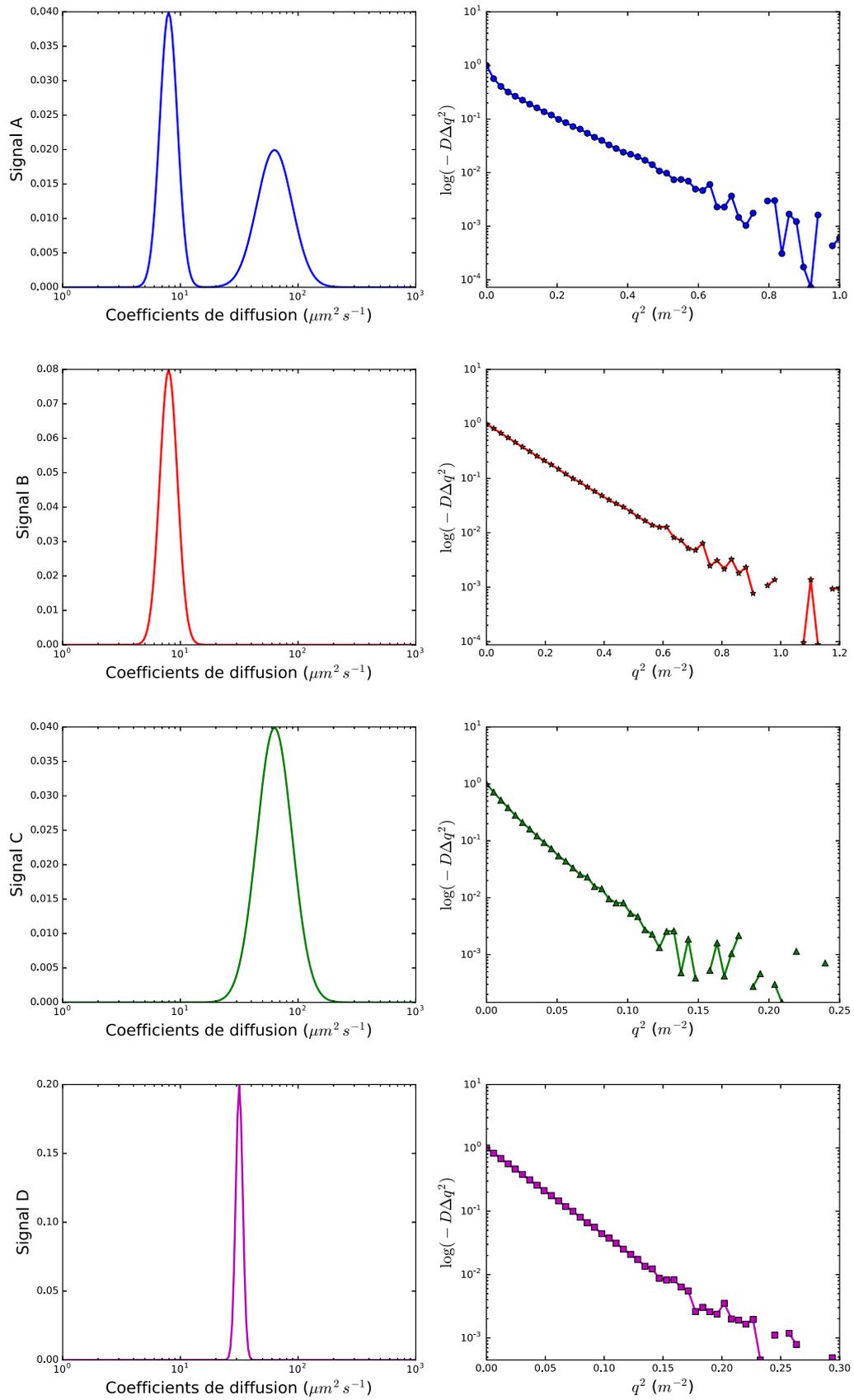


FIGURE 3.7 – Signaux synthétiques : A, B, C et D. Les signaux \bar{x} (à gauche) et leur mesures y (à droite) avec $\sigma = 10^{-3}$

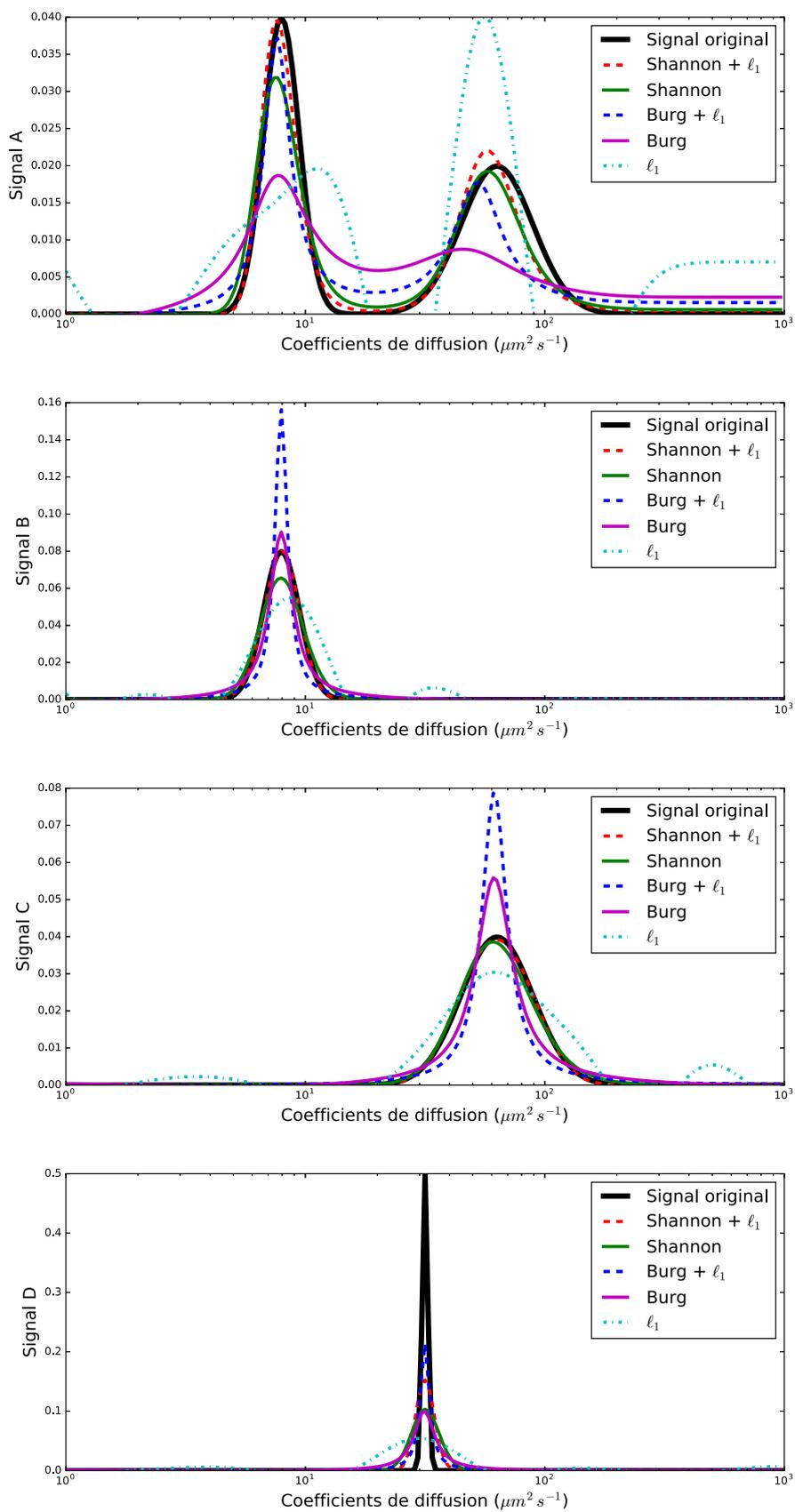


FIGURE 3.8 – Résultats de reconstruction de signaux A, B, C et D pour $\sigma = 10^{-3}$ avec les différentes fonctions Ψ .

	σ	Shannon	Shannon + ℓ_1	Burg	Burg + ℓ_1	ℓ_1
Signal A	10^{-3}	13.85	15.74	6.34	9.63	-0.74
	10^{-4}	18.70	18.35	8.22	7.48	8.03
	10^{-5}	19.42	22.19	8.47	6.84	10.17
	10^{-6}	19.38	27.62	8.49	6.79	11.98
Signal B	10^{-3}	16.69	32.60	10.51	3.49	7.14
	10^{-4}	30.36	37.63	7.60	0.86	7.66
	10^{-5}	34.92	45.16	7.34	0.51	9.61
	10^{-6}	32.99	41.35	7.30	0.45	10.41
Signal C	10^{-3}	23.76	34.29	8.17	4.65	9.16
	10^{-4}	29.45	33.60	6.21	5.70	15.29
	10^{-5}	38.55	40.76	6.04	6.42	19.59
	10^{-6}	45.42	45.28	6.02	6.50	26.01
Signal D	10^{-3}	2.13	6.26	2.94	7.35	0.71
	10^{-4}	3.35	6.96	3.36	9.60	0.82
	10^{-5}	3.80	14.71	3.42	10.03	0.88
	10^{-6}	3.86	19.87	3.43	10.09	1.09

Tableau 3.1 – RSB obtenus en reconstruisant les signaux A, B, C et D avec les différents choix de la fonction de régularisation Ψ pour différents niveaux de bruit.

3.7.3 Comparaison avec l’approche éclatée

Nous nous proposons dans cette section de comparer l’algorithme PPXA+ avec sa version éclatée. Pour cela, nous appliquons les deux algorithmes : Algorithme 6 et Algorithme 7 pour l’analyse des signaux A, B, C et D proposés dans la Section 3.7 en utilisant la meilleure fonction de pénalité : l’entropie de Shannon combinée à la norme ℓ_1 .

Comme il est présenté dans la Figure 3.9, il est clair que la version éclatée de l’algorithme PALMA assure asymptotiquement une reconstruction équivalente à celle obtenue avec l’algorithme PALMA. En outre, en traçant la courbe du RSB en fonction du temps, nous constatons que l’algorithme PALMA est plus rapide, et en particulier il atteint un RSB plus élevé pour un temps de calcul donné.

3.8 Conclusion

Nous avons présenté dans ce chapitre le problème inverse mal posé en DOSY et les méthodes proposées pour le résoudre. Compte tenu du besoin de nouvelles approches plus développées pour bien restaurer les signaux DOSY, nous proposons une régularisation hybride avec différentes fonctions d’entropie combinée à des fonctions de parcimonie. Nous utilisons les approches proximales pour créer un nouvel algorithme qui répond à ce besoin. Ce résultat a fait l’objet d’un article de conférence [Cherni 2016]. À l’occasion d’un premier test de simulation, nous avons vu que l’idée de combiner l’entropie et la parcimonie permet de bien reconstruire les signaux DOSY en particulier

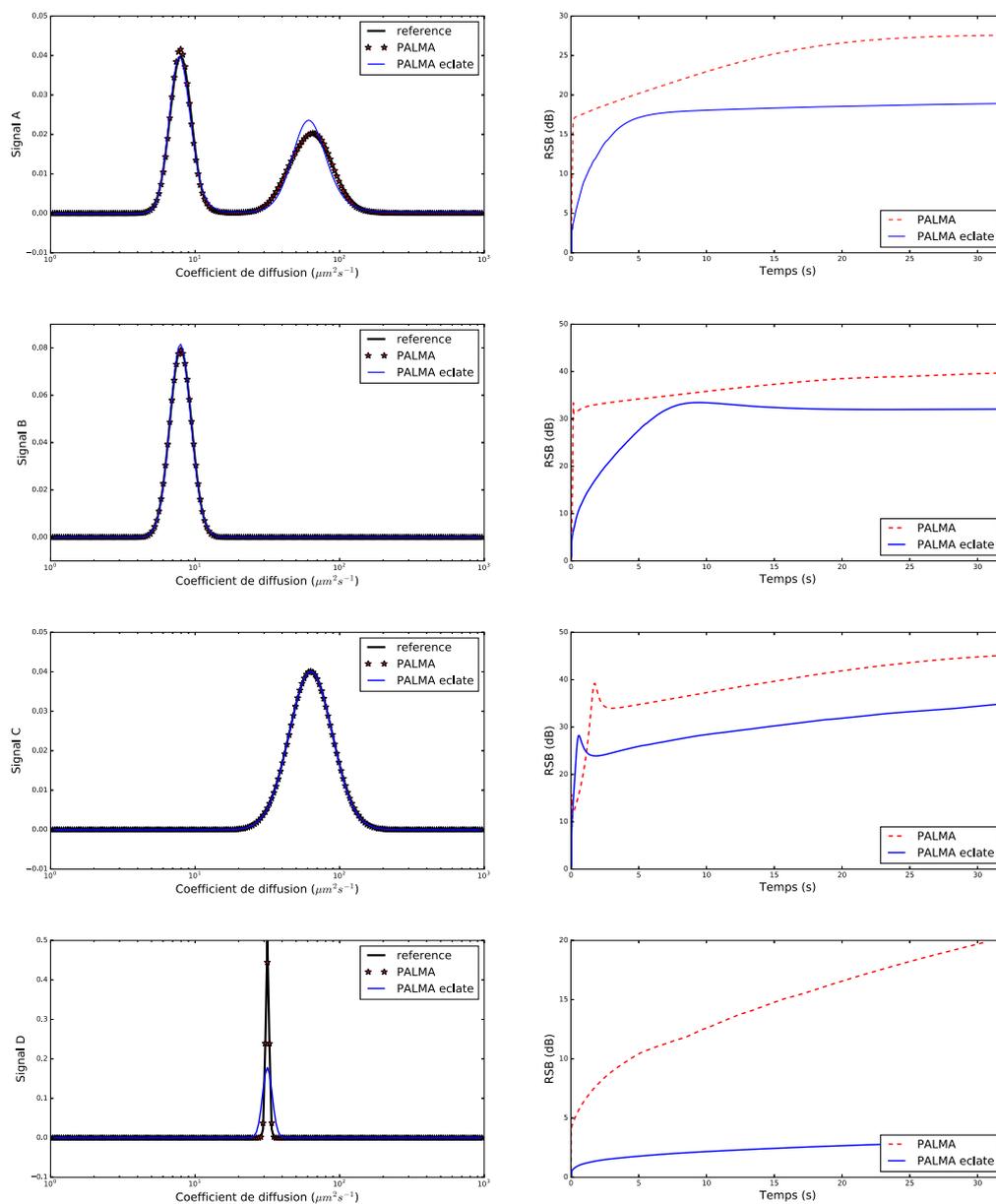


FIGURE 3.9 – Résultats de reconstruction des signaux A, B, C et D avec l’algorithme PALMA et sa version élargie

quand il s'agit de l'entropie de Shannon et de la norme ℓ_1 . Nous avons montré également l'intérêt numérique à traiter la pénalisation hybride en une seule passe (sans éclatement) dans l'algorithme, ce qui accentue l'intérêt des expressions des opérateurs proximaux hybrides que nous avons établies. Le chapitre suivant se concentrera sur l'algorithme PALMA et son application dans le contexte du traitement de données réelles de DOSY.

PALMA, nouvel algorithme pour la restauration des signaux DOSY

Sommaire

4.1	Introduction	44
4.2	Algorithme PALMA	44
4.3	Résultats avec des données simulées	45
4.3.1	Données de simulation	45
4.3.2	Optimisation des paramètres	45
4.3.3	Développement asymptotique	49
4.3.4	Évaluation de PALMA	49
4.4	Application de PALMA sur des données réelles	55
4.5	Serveur Web	57
4.6	Conclusion	58

4.1 Introduction

Nous avons présenté dans le chapitre précédent une nouvelle approche de régularisation hybride faisant la combinaison entre l'entropie et la parcimonie. Différentes combinaisons ont été proposées à travers deux fonctions d'entropie (Shannon, Burg) et l'une des quatre mesures de parcimonie (ℓ_0 , ℓ_1 , f_{Cauchy} , $f_{\text{log-Sum}}$). Les résultats de simulation ont montré que l'entropie de Shannon et la norme ℓ_1 forment la meilleure combinaison par rapport aux autres propositions. Dans ce chapitre, nous tirons parti de cette approche pour proposer un nouvel algorithme PALMA proposé dans la Section 4.2 dédiée à la reconstruction des signaux DOSY. Dans la Section 4.3, nous allons valider l'efficacité de PALMA et nous présenterons une comparaison entre cet algorithme et les méthodes proposées dans la littérature pour restaurer des signaux DOSY synthétiques. La Section 4.4 sera consacrée à l'application de PALMA pour l'analyse de données réelles. Nous finirons ce chapitre avec la Section 4.5 pour présenter notre stratégie de valorisation logicielle de l'algorithme PALMA.

4.2 Algorithme PALMA

Nous proposons de résoudre le problème DOSY par la formulation contrainte (3.11), où Ψ est la régularisation hybride introduite au chapitre précédent faisant la combinaison entre l'entropie de Shannon et la norme de parcimonie ℓ_1 . Afin de faire un compromis entre ces deux types de mesure, nous poserons $\alpha + \beta = 1$. Par conséquent, l'équation (3.12) du chapitre précédent devient :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \Psi(\mathbf{x}) = \alpha \text{Ent}(\mathbf{x}) + (1 - \alpha)\ell_1(\mathbf{x}) \quad (4.1)$$

telle que :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \text{Ent}(\mathbf{x}) = \begin{cases} \sum_{n=1}^N \frac{x_n}{a} \log\left(\frac{x_n}{a}\right) & \text{si } x_n > 0 \\ 0 & \text{si } x_n = 0 \\ +\infty & \text{sinon} \end{cases} \quad (4.2)$$

et $a > 0$ présente une information *a priori* sur le signal cible. Si $a = 1$, on retrouve l'équation (3.27) du Chapitre 3. L'opérateur proximal de Ψ sera donné par $(\forall \mathbf{x} \in \mathbb{R}^N) \quad \text{prox}_{\Psi}(\mathbf{x}) = (\text{prox}_{\psi}(x_n))_{1 \leq n \leq N}$ (Voir annexe C.1) avec :

$$\text{prox}_{\psi}(x_n) = \begin{cases} \frac{\alpha}{a} \mathcal{W} \left[\frac{a}{\alpha} \exp\left(\frac{ax_n - a(1-\alpha)}{\alpha}\right) + \log(a) - 1 \right] & \text{si } \alpha \in]0, 1[\\ \text{sign}(x_n) \max(|x_n| - 1, 0) & \text{si } \alpha = 0 \end{cases} \quad (4.3)$$

Il est clair qu'une valeur nulle de α conduit à avoir la formule du seuillage doux défini par [Combettes 2011] :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \text{prox}_{\ell_1}(\mathbf{x}) = (\text{sign}(x_n) \max(|x_n| - 1, 0))_{1 \leq n \leq N} \quad (4.4)$$

alors que si $\alpha = 1$, on retrouve la formule de l'opérateur proximal de l'entropie de Shannon [Bauschke 2011] :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \text{prox}_{\text{Ent}}(\mathbf{x}) = \left(\frac{1}{a} \mathcal{W} (a \exp(ax_n) + \log(a) - 1) \right)_{1 \leq n \leq N} \quad (4.5)$$

Avec cet opérateur proximal, notre problème d'optimisation convexe peut être efficacement résolu

avec un algorithme de minimisation convexe de type proximal [Combettes 2011, Pustelnik 2011, Pesquet 2012]. Nous proposons par la suite le nouvel algorithme PALMA, l'abréviation de "Proximal Algorithm for ℓ_1 combined with MAxent prior" basé sur l'approche PPXA+ donné par l'Algorithme 6 pour le choix particulier de Ψ (4.1).

4.3 Résultats avec des données simulées

4.3.1 Données de simulation

Nous présentons des tests de simulation de PALMA sur quatre signaux synthétiques. Un signal A qui présente un profil monodisperse, et trois signaux B, C et D qui présentent des profils polydisperses de différentes distributions comme il est présenté dans la Figure 4.1 :

- Le signal A est l'équivalent du signal simulé dans [Urbańczyk 2013]. Ce signal présente une mesure de trois espèces monodisperses ayant les coefficients de diffusion $16 \mu\text{m}^2/\text{s}$, $63 \mu\text{m}^2/\text{s}$, $230 \mu\text{m}^2/\text{s}$ d'intensités respectives 1, 0.33 et 0.66.
- Le signal B est une distribution logarithmique-normale centrée en $35 \mu\text{m}^2/\text{s}$.
- Les signaux C et D présentent des distributions asymétriques chacune formée d'un ensemble de 15 lois logarithmique-normales réparties sur un axe de diffusion de $18 \mu\text{m}^2/\text{s}$ à $85 \mu\text{m}^2/\text{s}$ ayant des intensités variant de 0.1 à 10.

Nous ajoutons à ces signaux un bruit additif gaussien de variance σ^2 .

Les mesures correspondantes aux différents signaux sont présentées dans la Figure 4.2 et l'algorithme PALMA vise donc à retrouver les signaux de la Figure 4.1.

4.3.2 Optimisation des paramètres

4.3.2.1 Paramètre α

L'algorithme PALMA est basé sur le calcul d'un opérateur proximal (4.3) qui dépend du coefficient α . Ce coefficient ajuste le compromis entre l'entropie de Shannon et le degré de parcimonie. Avec deux profils : monodisperse (signal A), et polydisperses (signaux B, C et D), nous proposons de reconstruire les distributions correspondantes en faisant varier le coefficient α de 0 à 1 (Figure 4.3). Pour $\alpha = 0$, comme nous l'avons mentionné dans l'équation (4.4), l'opérateur proximal n'est autre que le seuillage doux, l'opérateur qui est bien adapté aux signaux parcimonieux. Par conséquent, on trouve que les meilleures reconstructions du signal monodisperse A sont obtenues avec une valeur de α égale à 0. En revanche, pour une distribution gaussienne, ou lorentzienne, la valeur de $\alpha = 0$ n'assure pas une reconstruction efficace des signaux B, C ou D. La reconstruction des différents profils semble être satisfaisante pour $\alpha = 0.01$ ou $\alpha = 0.05$. Ces deux valeurs permettent de reconstruire le signal DOSY soit pour les données monodisperses (signal A), soit pour les données polydisperses (signaux B, C et D). Nous proposons de fixer l'ajustement entre l'entropie et la parcimonie avec la valeur $\alpha = 0.01$ (Figure 4.3), ce qui semble être le meilleur compromis en terme de qualité de reconstruction pour les profils mono ou polydisperses.

4.3.2.2 Paramètre η

Les résultats de la section précédente nécessitent un ajustement de la valeur de contrainte η selon le niveau de bruit. Sur nos exemples, nous avons considéré un bruit gaussien additif d'une variance σ^2 et de moyenne nulle. La valeur de σ dans le cas de données synthétiques est connue, mais notons

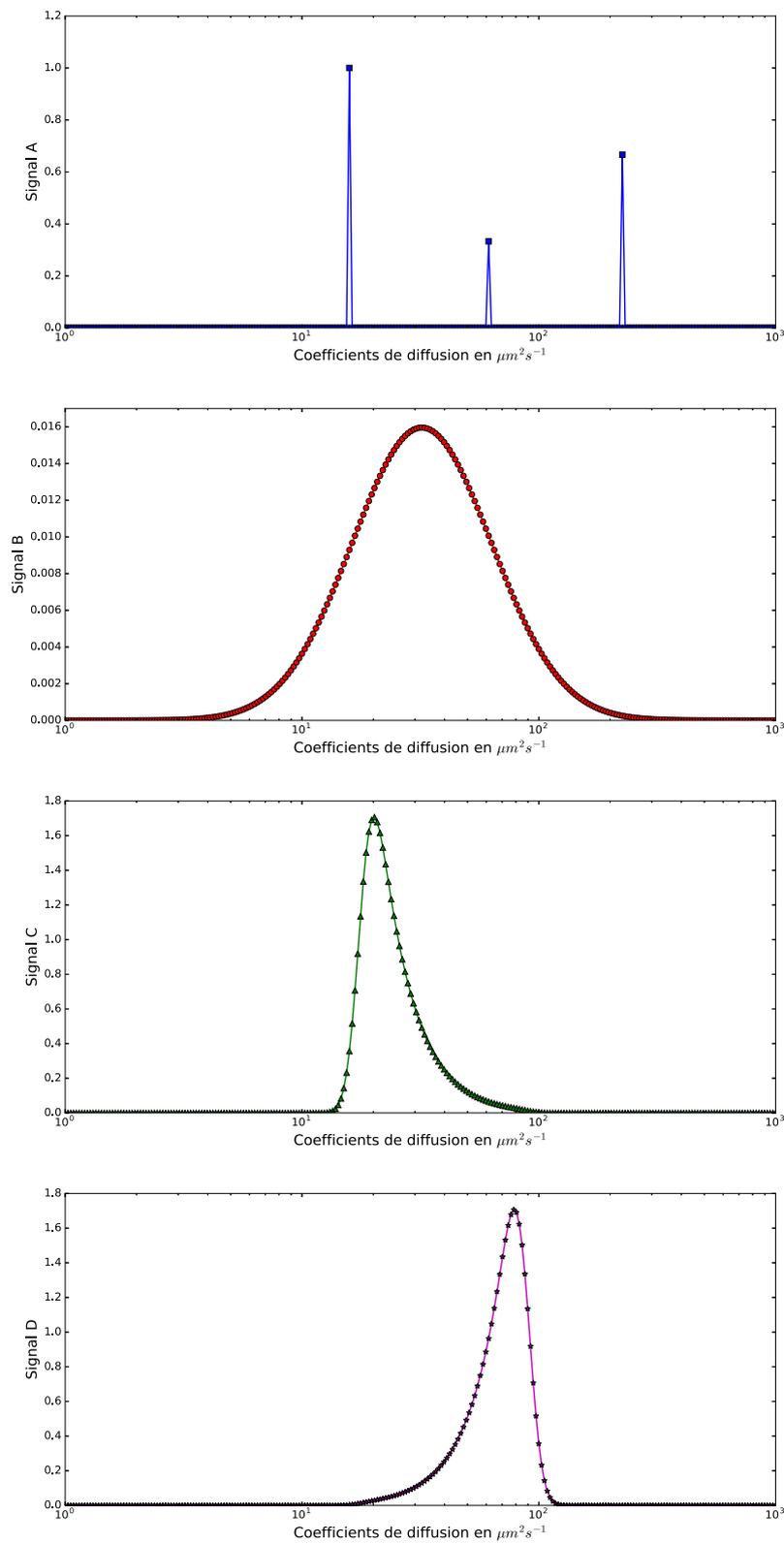


FIGURE 4.1 – Signaux simulés

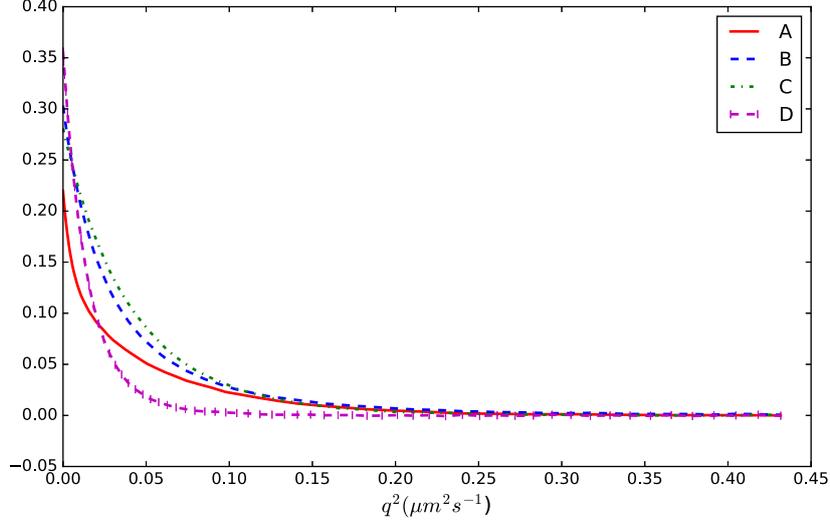


FIGURE 4.2 – Mesures DOSY des signaux simulés (Annexe C.2)

que dans le cas de données expérimentales, elle peut être estimée efficacement par un ajustement polynomial de la courbe de \mathbf{y} . Dans ce cadre, nous introduisons la loi χ^2 donnée par :

$$\chi^2(\mathbf{x}, \mathbf{y}) = \frac{\|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2}{\sigma^2} \quad (4.6)$$

La fonction χ^2 suit une loi normale de moyenne M et de variance $2M$ et vérifie l'inégalité :

$$\chi^2(\mathbf{x}, \mathbf{y}) \leq M \iff \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 \leq \sigma^2 M \quad (4.7)$$

Par la suite, la contrainte $\|\mathbf{H}\mathbf{x} - \mathbf{y}\| \leq \eta$ peut donc être utilisée pour $\eta = \sigma\sqrt{M}$ [Delsuc 1989, Chouzenoux 2010].

Supposons maintenant qu'on ait un bruit gaussien corrélé d'une matrice de covariance Σ et de moyenne nulle. La contrainte des moindres carrés $\|\mathbf{H}\mathbf{x} - \mathbf{y}\| \leq \eta$ peut être reformulée par :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad ((\mathbf{H}\mathbf{x} - \mathbf{y})^\top \Sigma^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y}))^{1/2} \leq \eta$$

avec $\mathbf{H} \leftarrow \theta \Sigma^{-1/2} \mathbf{H}$ et $\mathbf{y} \leftarrow \theta \Sigma^{-1/2} \mathbf{y}$, $\eta \leftarrow \theta \eta$, pour $\theta > 0$. Un choix simple en pratique est de fixer $\theta = \sigma_{\max}$, i.e, le maximum des valeurs singulières de Σ , ce qui donne lieu à un coefficient η de la forme $\eta = \sigma_{\max} \sqrt{M}$.

En conclusion, soulignons que nous choisirons $\eta = \xi \sigma \sqrt{M}$ avec ξ légèrement supérieur à 1, car cela semble conduire à de meilleurs résultats qualitatifs [Trussell 1983, Galatsanos 1992, Chouzenoux 2010].

4.3.2.3 Paramètre a

La fonction d'entropie que nous avons proposée dans ce chapitre exige une information *a priori* sur la dynamique du signal cible. Nous notons cette information par le paramètre a . Pour trouver une valeur de a , une approche naturelle consiste à estimer l'aire intégrée sous le signal cible. Dans notre cas, nous proposons de choisir $a = y_0 = \sum_n x_n$ [Delsuc 1989].

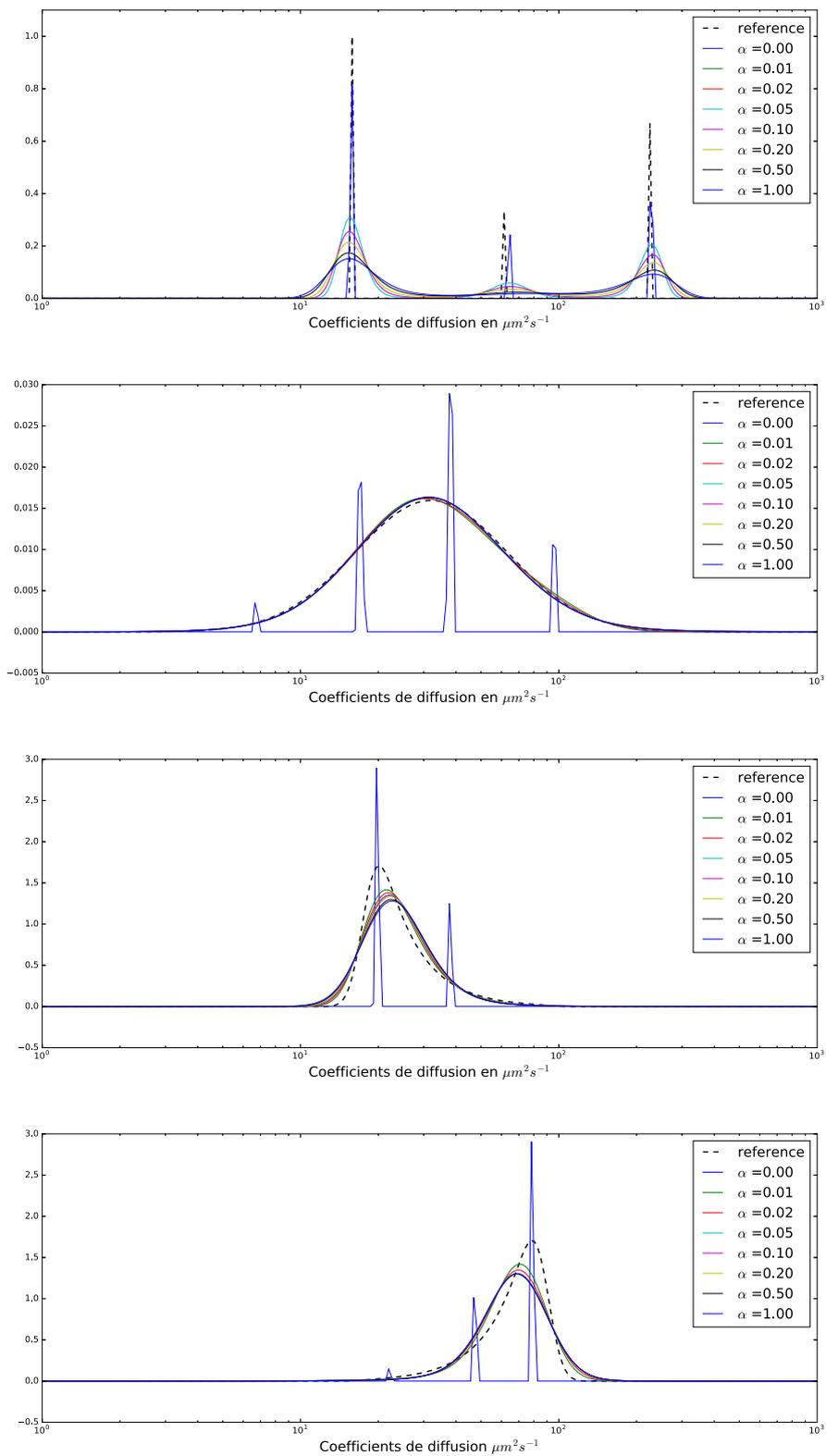


FIGURE 4.3 – Résultats de reconstruction des signaux A, B, C et D pour différentes valeurs de α

4.3.3 Développement asymptotique

L'opérateur proximal utilise la forme de la fonction de Lambert (4.3). Cette fonction possède différentes propriétés telles que :

$$\mathcal{W}(e^x) \xrightarrow{x \rightarrow +\infty} x \left(1 - \frac{\log(x)}{1+x}\right) \quad (4.8)$$

Dans ce cas seule la restriction de \mathcal{W} sur \mathbb{R}_+ est retenue.

Si $\alpha \in]0, 1]$, pour tout $n \in \{1, \dots, N\}$, on a :

$$\text{prox}_\psi(x_n) = \frac{\alpha}{a} \mathcal{W}(\exp(c_n)) \quad (4.9)$$

avec

$$c_n = \frac{ax_n - a(1-\alpha)}{\alpha} - 1 + 2\log(a) - \log(\alpha) \quad (4.10)$$

Dans le cas où la valeur de c_n est très grande (typiquement $c_n > 10^2$), l'implémentation numérique de \mathcal{W} tend vers $+\infty$. Donc, nous proposons d'utiliser l'approximation asymptotique de \mathcal{W} pour toute valeur grande de c_n :

$$\text{prox}_\psi(x_n) \approx \frac{\alpha}{a} \left(c_n - \frac{c_n}{1+c_n} \log(c_n) \right) \quad (4.11)$$

La Figure 4.4 illustre l'erreur relative de l'approximation asymptotique de $\mathcal{W}(e^{c_n})$ pour $\lambda = 1$ et $a = 1$.

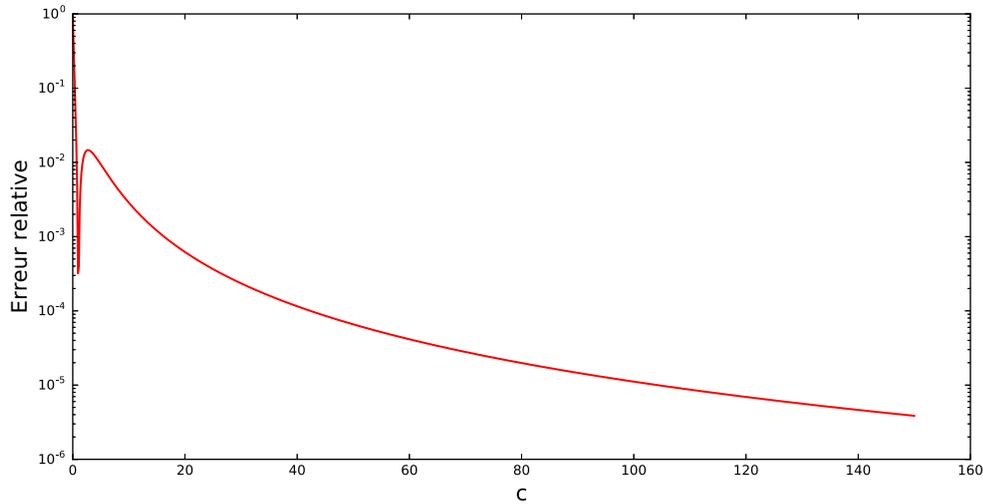


FIGURE 4.4 – Erreur relative entre $\mathcal{W}(e^{c_n})$ et l'approximation de $\text{prox}_\psi(x_n)$ définie dans l'équation (4.11). L'erreur relative devient inférieure à 10^{-3} pour $c_n > 20$.

4.3.4 Évaluation de PALMA

Pour confirmer l'efficacité de notre nouvelle approche par rapport aux autres méthodes, nous proposons de faire une comparaison entre PALMA et les récentes approches proposées dans la

littérature que nous avons citées dans le chapitre précédent. Pour différents niveaux de bruit, nous présentons dans les Figures 4.5, 4.6, 4.7 et 4.8, le résultat de reconstruction des signaux A, B, C et D avec les différents algorithmes : TRAIIn [Xu 2013], ITAMeD [Kazimierczuk 2012], ITAMeD ℓ_p [Urbańczyk 2013] et PALMA [Cherni 2017].

Selon les résultats de la Figure 4.5, avec un niveau de bruit élevé ($\sigma = 10^{-2}$), nous remarquons que la reconstruction du signal A n'est pas satisfaisante. PALMA détecte la présence des molécules monodisperses mais avec un décalage sur l'axe de diffusion. Pour les autres niveaux de bruit, PALMA présente la meilleure reconstruction. L'algorithme TRAIIn paraît aussi efficace. Cependant, l'algorithme ITAMeD ainsi que sa version avec la pénalisation ℓ_p détectent la présence des molécules mais avec une sous-estimation d'intensité.

La Figure 4.6 montre que l'algorithme ITAMeD et ITAMeD ℓ_p assurent une reconstruction avec des fluctuations remarquables surtout dans le cas où le niveau de bruit est élevé ($\sigma = 10^{-2}$ et $\sigma = 10^{-3}$). Le signal B a été bien reconstruit avec l'algorithme TRAIIn sauf pour le plus faible niveau de bruit ($\sigma = 10^{-5}$). Il est clair que pour les différents niveaux de bruit, PALMA assure une reconstruction satisfaisante du signal B.

Avec les signaux polydisperses, nous constatons que ITAMeD et ITAMeD ℓ_p sont incapables de reconstruire efficacement le profil du signal C (Figure 4.7) et le signal D (Figure 4.8). Avec TRAIIn on obtient une reconstruction satisfaisante des signaux C et D sauf pour le plus haut niveau de bruit. À nouveau, PALMA présente les meilleures qualités de reconstruction pour les différents niveaux de bruit. L'algorithme ITAMeD et sa pénalisation en norme ℓ_p présentent un très faible RSB pour un niveau de bruit élevé.

Selon tous ces résultats, nous constatons que les deux algorithmes TRAIIn et PALMA sont efficaces mais il paraît que TRAIIn est sensible au niveau de bruit. Par la suite, nous testons la robustesse de ces deux algorithmes en fonction de la dimension du problème. Le Tableau 4.1 présente les résultats en terme de qualité de reconstruction des signaux A, B, C et D pour deux dimensions M .

	$M = 64$		$M = 32$	
	PALMA	TRAIIn	PALMA	TRAIIn
Signal A	8.92	0.69	26.61	2.60
Signal B	27.78	12.59	41.69	26.53
Signal C	20.92	15.11	21.38	28.37
Signal D	14.72	6.60	25.74	23.36

Tableau 4.1 – RSB en dB obtenus par la reconstruction des signaux A, B, C et D pour $\sigma = 10^{-4}$ pour deux dimensions : $M = 64$ et $M = 32$

Bien que l'algorithme TRAIIn présente un RSB légèrement supérieur au RSB obtenu avec l'algorithme PALMA quand $M = 32$ pour reconstruire le signal C, les autres résultats de reconstruction dans les deux dimensions $M = 64$ et $M = 32$ mettent en évidence la robustesse de PALMA. En fait, pour un signal monodisperse (signal A) l'algorithme PALMA assure un gain de 24 dB par rapport à TRAIIn quand $M = 32$. Pour les autres signaux (signaux B, C et D), PALMA donne les RSB les plus élevés. Par conséquent, la qualité de reconstruction de PALMA est clairement plus satisfaisante que celle donnée par l'algorithme TRAIIn. Aussi, rappelons que TRAIIn nécessite un réglage attentif de la stratégie d'arrêt et de région de confiance, on peut donc considérer que l'algorithme PALMA se distingue des méthodes concurrentes par sa robustesse et la qualité des résultats obtenus.

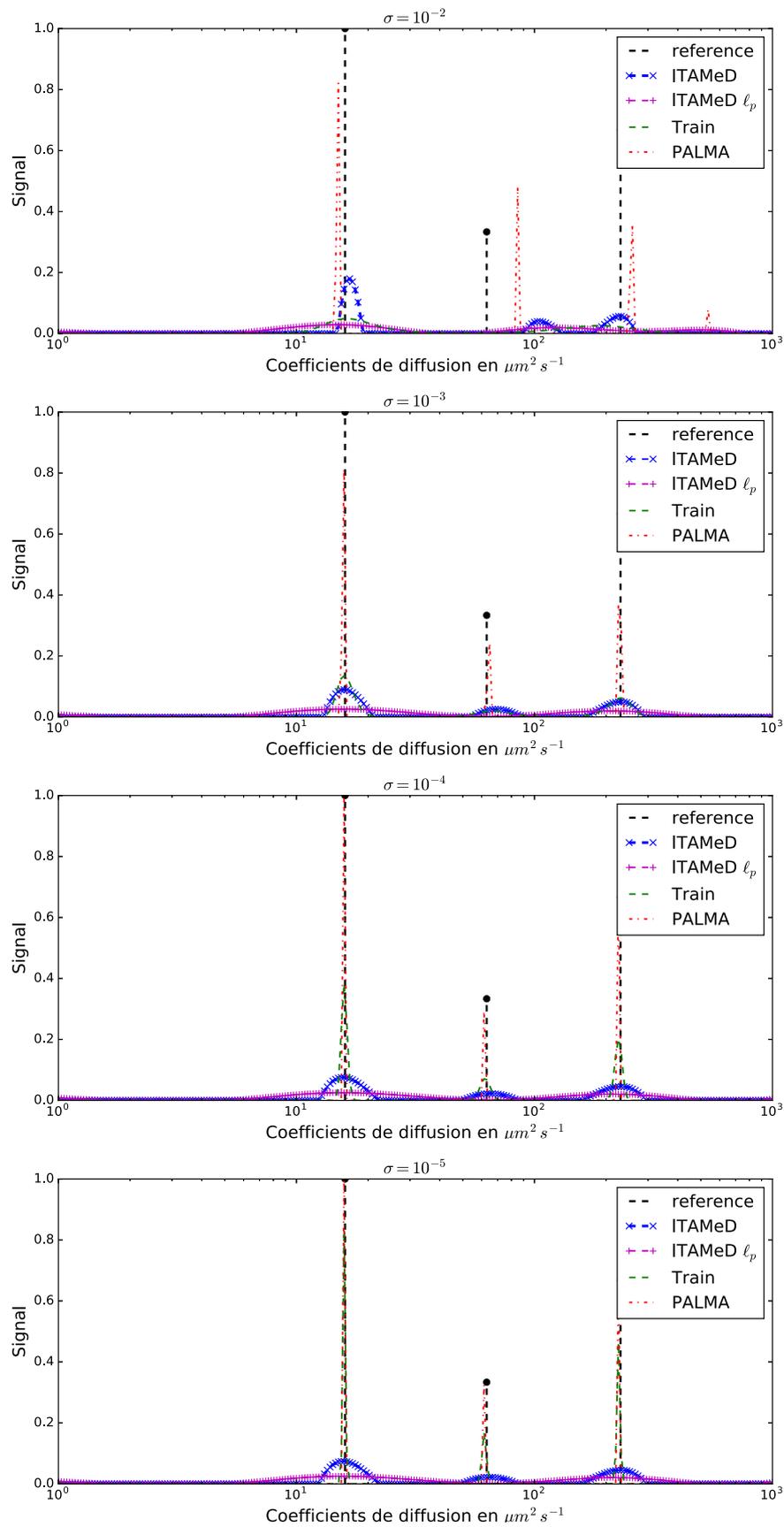


FIGURE 4.5 – Résultats de reconstruction du signal A avec les différents algorithmes proposés pour les différents niveaux de bruit

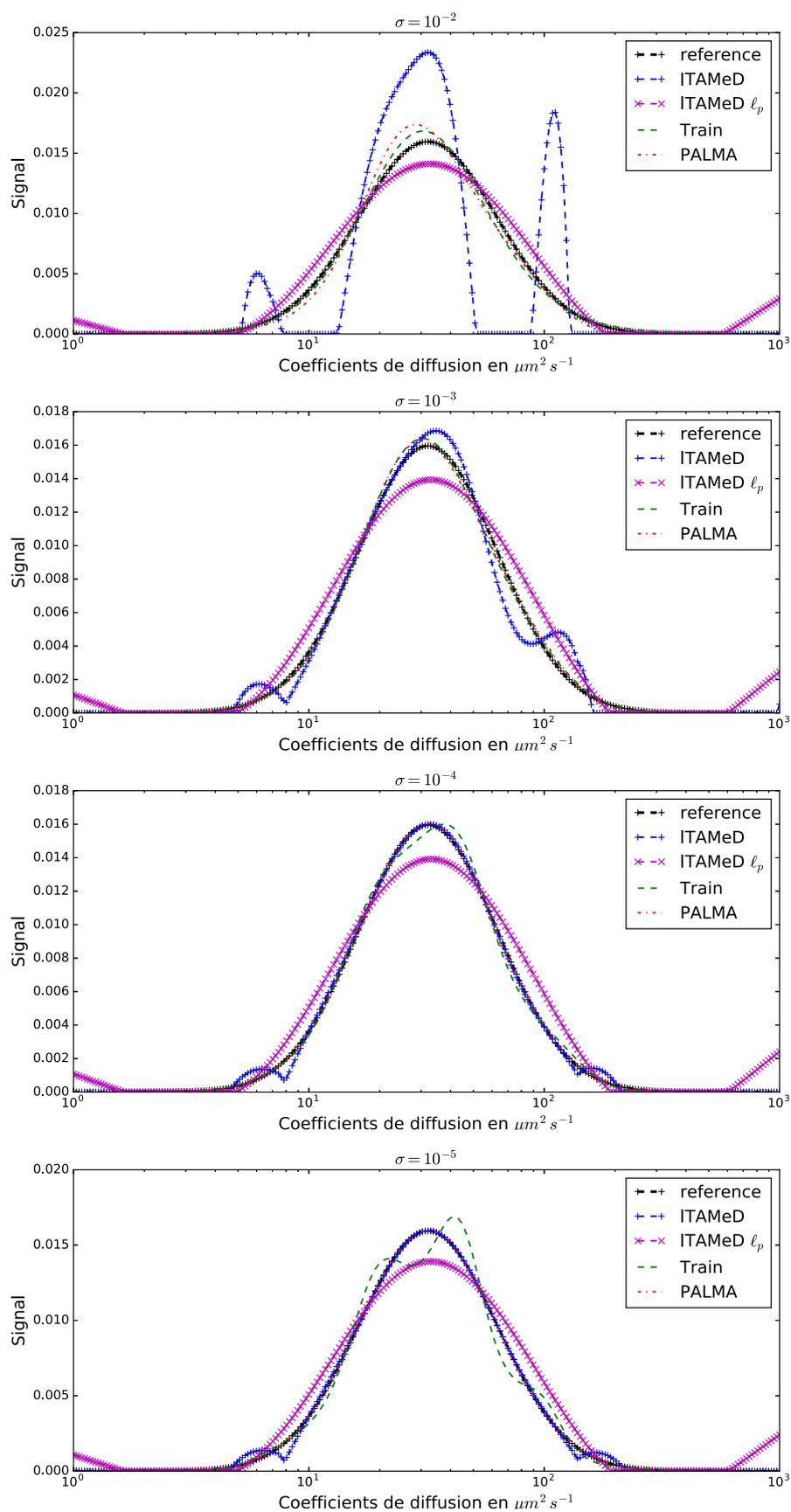


FIGURE 4.6 – Résultats de reconstruction du signal B avec les différents algorithmes proposés pour les différents niveaux de bruit

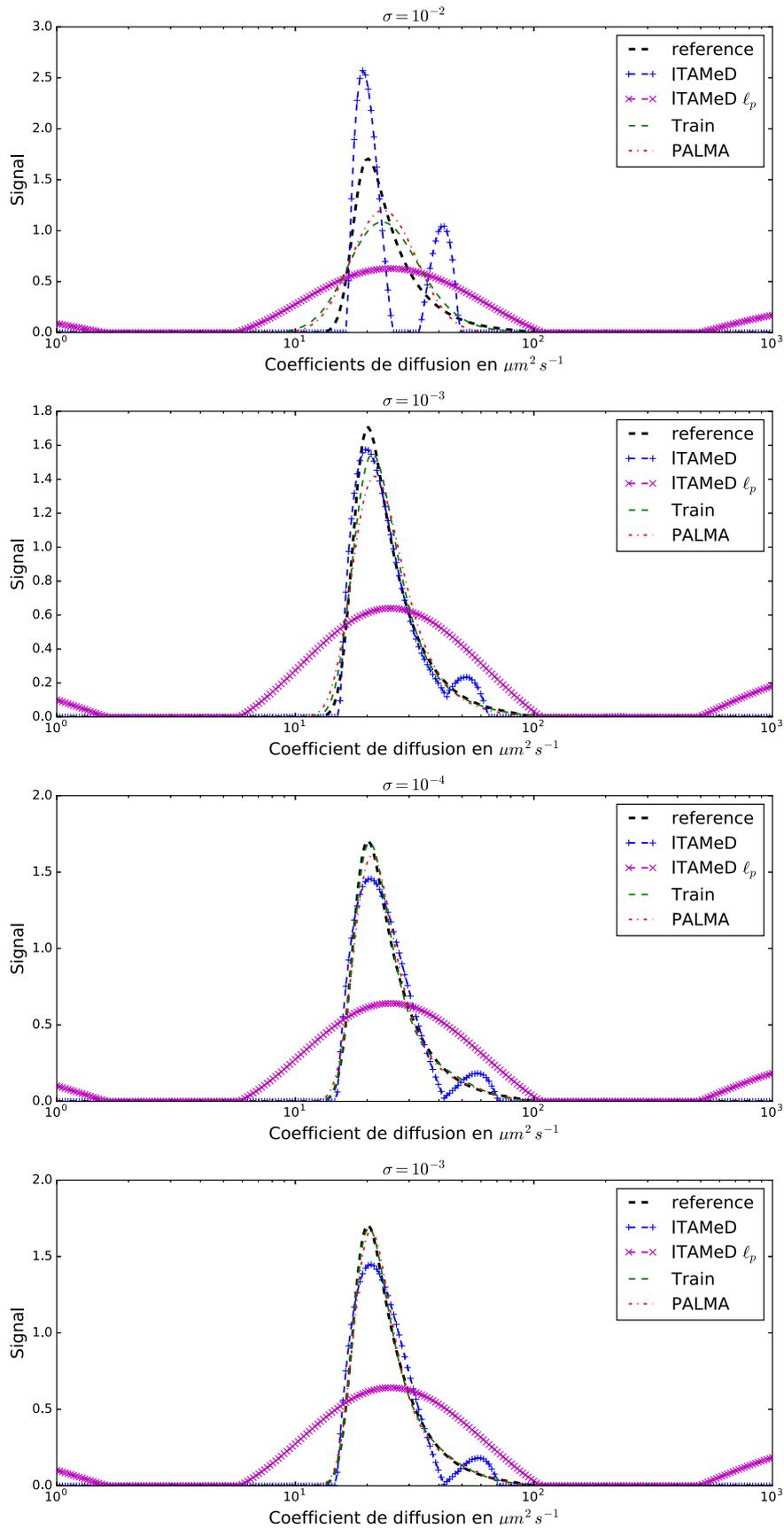


FIGURE 4.7 – Résultats de reconstruction du signal C avec les différents algorithmes proposés pour les différents niveaux de bruit

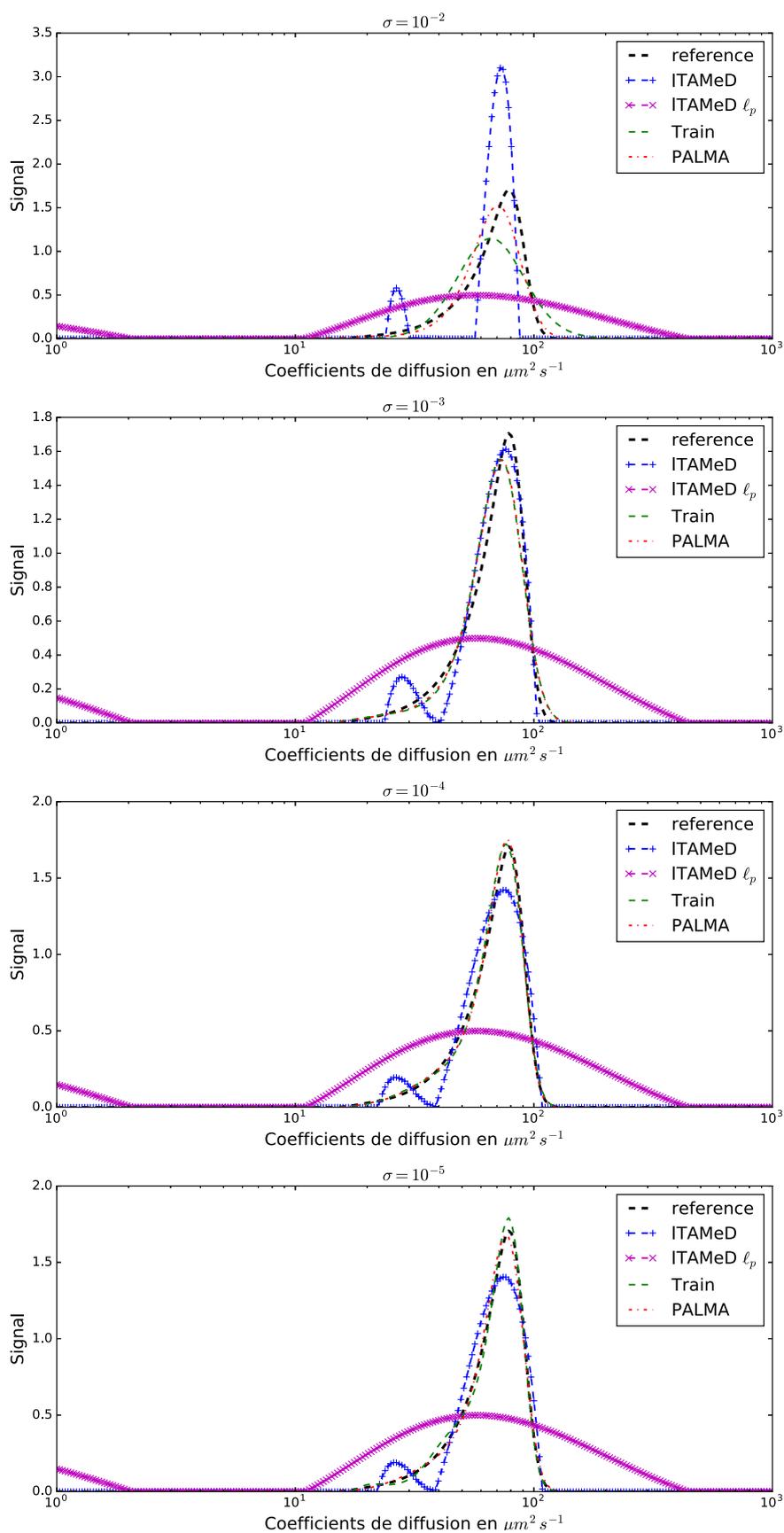


FIGURE 4.8 – Résultats de reconstruction du signal D avec les différents algorithmes proposés pour les différents niveaux de bruit

4.4 Application de PALMA sur des données réelles

Notre premier test sur des données réelles s'intéresse à l'analyse des données DOSY ^{19}F de la molécule 1H,2H,2H,2H-Perfluorooctanol FTOH^a de concentration 1mM mesurées dans le solvant DMSO- d_6 ^b. Notons que le FTOH est une molécule fluorée très utilisée dans les matériaux de protection de surface car elle possède la propriété d'hydrophobicité. Cependant, cette molécule ne se dégrade pas dans la nature et son accumulation sera systématiquement nocive, ce qui fait l'intérêt de plusieurs travaux de recherches depuis quelques années [Marchione 2009, Li 2018].

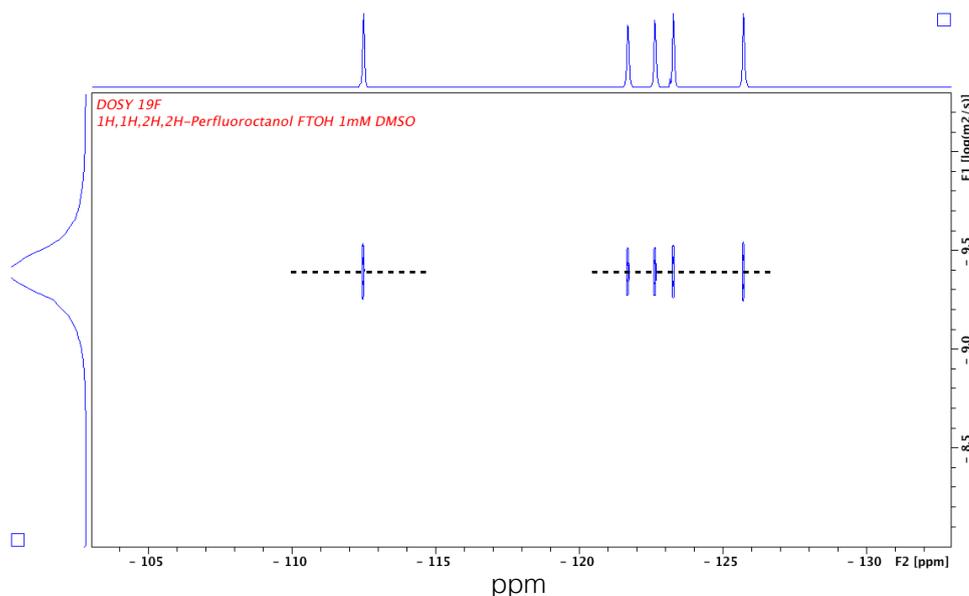


FIGURE 4.9 – Spectre 2D DOSY ^{19}F de la FTOH dans le DMSO. L'expérience a duré 15 min 22 s à une température $T=289\text{K}$

Dans ce genre d'analyse DOSY, la présence d'une molécule particulière ou même d'une famille de molécules est caractérisée par l'alignement au même niveau de diffusion avec de différentes positions de déplacements chimiques.

La Figure 4.9 indique la présence d'une famille de molécules de coefficient de diffusion $269\mu\text{m}^2\text{s}^{-1}$. Cela sert à détecter la présence des molécules fluorées dans des mélanges complexes. Par exemple, l'analyse PALMA des données DOSY d'un extrait du sol d'une rivière (Figure 4.10) montre l'existence des molécules de coefficients de diffusion $295\mu\text{m}^2\text{s}^{-1}$. Cela provient de la présence de Fluor organique dans le sol de cette rivière. Cette interprétation très intéressante implique la pollution du sol qui influe pernicieusement sur l'eau de la rivière.

En tant que second test, nous proposons d'analyser un extrait éthanolique d'algue avec notre algorithme PALMA. En fait, l'analyse des mélanges complexes a été largement utilisée pour l'étude de produits naturels, c'est pourquoi nous avons pensé à ce choix de données.

La Figure 4.11 présente les résultats de l'expérience DOSY où les composants les plus abondants sont visibles à ce niveau expérimental. On voit que sur la même ligne horizontale, les chaînes d'acides gras sont bien présentes à $220\mu\text{m}^2\text{s}^{-1}$, le glycérol aussi à $380\mu\text{m}^2\text{s}^{-1}$ et le méthanol

a. FTOH : Fluorotolomer alcohols, une molécule fluorée avec un groupe alcool.

b. DMSO : Diméthylsulfoxyde, c'est un solvant organique de formule brute $\text{C}_2\text{H}_6\text{OS}$

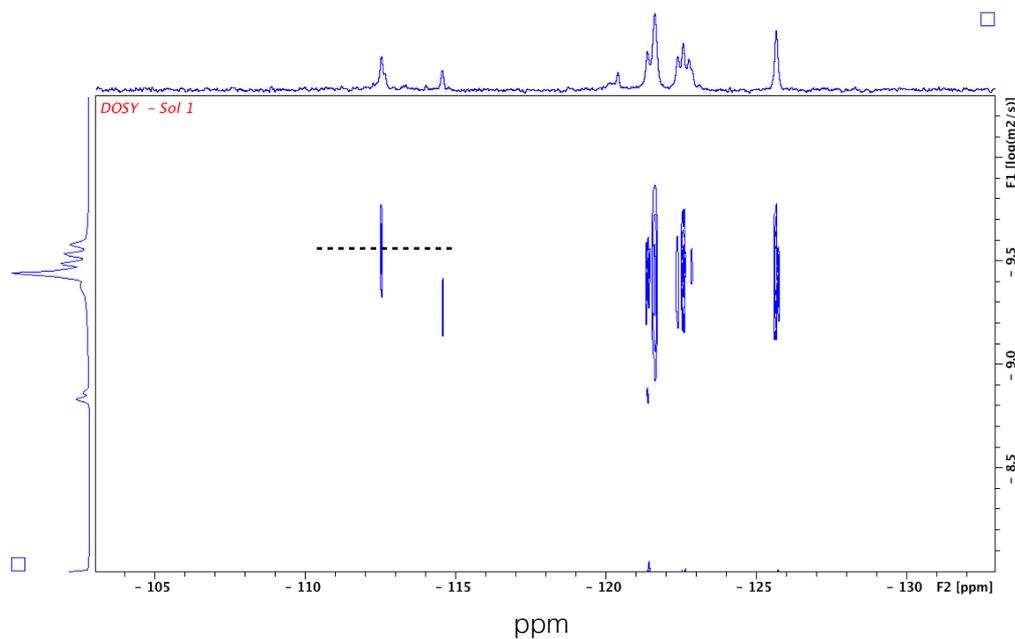


FIGURE 4.10 – Spectre 2D DOSY ^{19}F de de l'échantillon du sol dans le DMSO. L'expérience a durée 15 min à une température $T=289\text{K}$

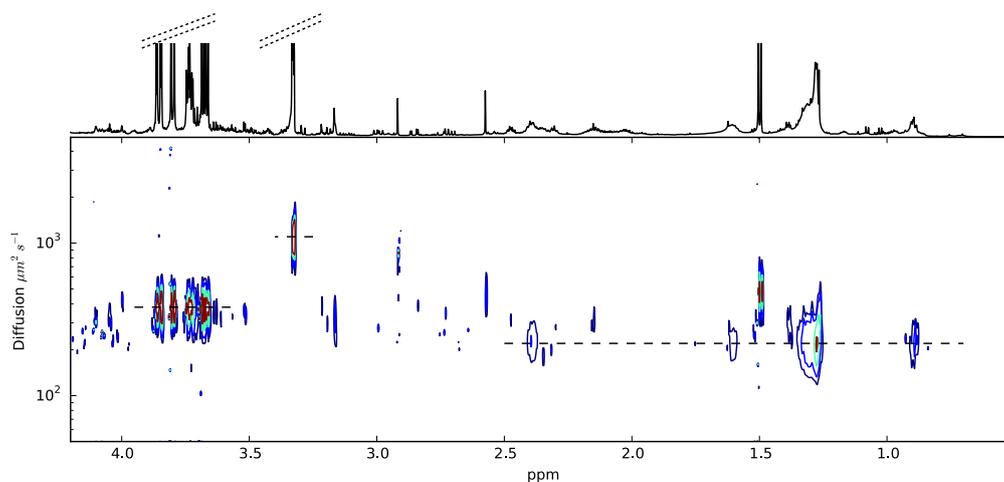


FIGURE 4.11 – Expérience DOSY sur un extrait éthanologique d'algue. Seules les composantes principales sont présentées. La ligne horizontale indique les chaînes d'acides aminés à $220 \mu\text{m}^2\text{s}^{-1}$, le glycérol à $380 \mu\text{m}^2\text{s}^{-1}$ et le méthanol à $1100 \mu\text{m}^2\text{s}^{-1}$

à $1100 \mu\text{m}^2\text{s}^{-1}$. L'expérience DOSY est très utile dans ces analyses car elle fournit une étude dynamique très élevée telle que les lignes les plus intenses ne dominent pas les moins intenses. Par conséquent, toutes les composantes peuvent apparaître avec leurs intensités exactes et l'analyse de la matière sera précisément effectuée.

Avec les mêmes données et les mêmes conditions, nous nous proposons d'ajouter de la chloroquine à une faible concentration (1% de l'extrait sec d'algue). La chloroquine est une molécule antipaludéenne (très utilisée pour limiter la maladie du paludisme). L'objectif de cette procédure est de

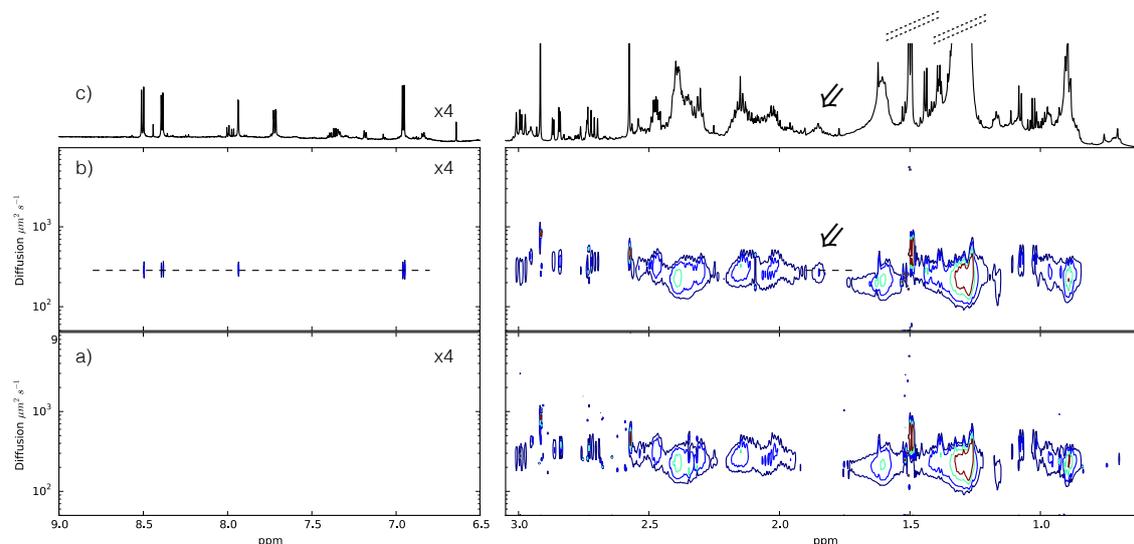


FIGURE 4.12 – Comparaison entre deux expériences DOSY sur un extrait éthanolique d’algue : (a) résultat de l’expérience initiale (Figure 4.11), (b) même expérience avec 0.16 mg mL^{-1} de chloroquine, (c) le spectre 1D de l’extrait d’algue avec la chloroquine

vérifier si l’algorithme PALMA est capable de détecter cette molécule présente en faible quantité. Les signaux aromatiques de la chloroquine, situés dans une région vide comme il est montré dans la Figure 4.12 illustrant le résultat de PALMA, sont facilement détectés, contrairement aux signaux de la chaîne aliphatique situés dans une région encombrée. Les groupes de méthyles à la position 1.3 ppm^c et 1.45 ppm sont complètement enterrés sous les signaux d’acides gras et les coefficients de diffusion possèdent presque la même valeur. De plus, la présence de la chloroquine est détectée à 1.9 ppm . Même si l’intensité de ces signaux est 100 fois plus faible par rapport aux grandes intensités des chaînes d’acides gras, la séparation des signaux est bien établie. Par conséquent, l’algorithme PALMA assure efficacement l’analyse des données réelles de DOSY et il détecte toutes les composantes, y compris celles de faibles intensités.

4.5 Serveur Web

Après avoir validé notre approche avec des signaux synthétiques et réels, nous proposons de mettre en œuvre l’usage de l’algorithme PALMA. Par conséquent, nous développons un serveur web <http://palma.labo.igbmc.fr/> qui permet facilement d’analyser toutes mesures DOSY. L’interaction entre l’utilisateur et le serveur est fondée sur un téléchargement de données DOSY avec l’insertion d’une adresse mail valide de la part de l’utilisateur (Figure 4.13). D’autre part, le serveur PALMA est censé vérifier le format des données et envoyer l’analyse de données à l’utilisateur si les formats sont adéquats. Sinon, l’utilisateur reçoit une notification d’erreur. L’historique de l’analyse (l’heure de la réception des données, les données DOSY, les résultats d’analyse, l’adresse mail de l’utilisateur, ses commentaires) sera automatiquement sauvegardé par le serveur. À ce jour, le serveur PALMA a été utilisé par 232 utilisateurs et une valorisation du serveur web a été effectuée début 2018 par la SATT CONECTUS^d de la région Alsace.

c. ppm : particule par million, c’est une unité de déplacement chimique en RMN, équivalente à la fraction à 10^{-6} (un millionième).

d. SATT : Société d’Accélération du Transfert de Technologies

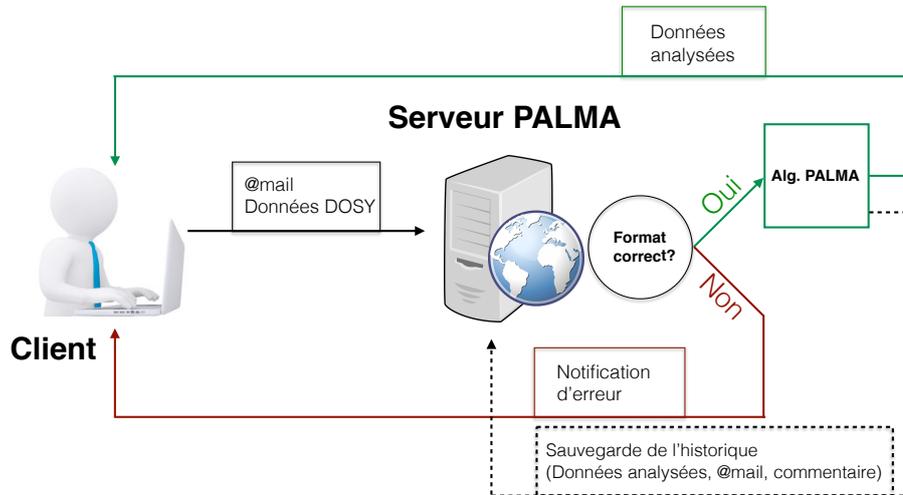


FIGURE 4.13 – Représentation schématique des interactions avec le serveur PALMA

L'interface graphique du serveur est formée de cinq fenêtres :

- **Home** : Une page d'accueil de notre serveur web présente les grands titres de notre travail et les collaborations effectuées pour le créer. Tous les intervenants dans ce projet d'un point de vue développement algorithmique, programmation, et création du serveur web, sont mentionnés sur cette page.
- **Data analysis** : Sur cette page web, on peut faire la demande d'analyse de données DOSY. Il suffit de soumettre les données en format compressé comme il est mentionné. Tout autre format sera rejeté et l'utilisateur sera informé immédiatement. De plus, un choix de la valeur α sera demandé, ce choix est très limité. La valeur $\alpha = 1$ correspond théoriquement à l'entropie maximale. Sinon il est préférable de choisir la valeur $\alpha = 0.01$ ou $\alpha = 0.05$. L'utilisateur est mené à préciser une adresse e-mail valide sur laquelle le serveur va envoyer les résultats d'analyse dans les plus brefs délais.
- **Methods** : Sur cette page, nous expliquons brièvement la théorie de notre approche et quelques indications sur la nature des fichiers à soumettre et à recevoir.
- **How to** : La page "How to" est consacrée à expliquer le mode d'utilisation du serveur PALMA. Toutes les informations nécessaires pour analyser les données et recevoir les résultats sont détaillées.
- **Contact us** : Nous créons cet espace web pour se mettre à la disponibilité de l'utilisateur pour toutes les remarques ou questions éventuelles.

Les détails de ces interfaces sont présentés dans l'Annexe D. Nous mettons aussi notre code PALMA en langage python à la disponibilité de tous les utilisateurs sur GitHub <https://github.com/delsuc/PALMA>.

4.6 Conclusion

Dans ce chapitre, nous nous sommes concentrés sur la régularisation hybride faisant une combinaison entre l'entropie de Shannon et la parcimonie ℓ_1 . Nous avons proposé une stratégie algorithmique nommée PALMA et nous avons discuté le choix de tous les paramètres de cet algorithme

pour pouvoir reconstruire efficacement un signal DOSY quelle que soit sa nature : monodisperse ou polydisperse. Selon les différentes comparaisons de PALMA avec les algorithmes existants dédiés à la DOSY, les résultats montrent que notre algorithme PALMA est supérieur en termes de convergence, robustesse et précision. De plus, il bénéficie de garanties de convergence solides. Cette contribution a fait l'objet d'un article de journal [Cherni 2017].

Chapitre 5

Nouvelle approche par dictionnaire pour résoudre le problème inverse en SM

Sommaire

5.1	Introduction	62
5.2	Revue de la littérature	62
5.2.1	Principe de la SM	62
5.2.2	Le spectre de masse	64
5.2.3	Problème des massifs isotopiques	64
5.3	Méthodes de traitement de données en SM	66
5.4	Nouvelle approche par dictionnaire	67
5.4.1	Modélisation du problème	67
5.4.2	Approche par dictionnaire	68
5.4.3	Stratégie d'optimisation	68
5.4.4	Approximation par blocs	69
5.5	Résultats et validation	71
5.5.1	Tests de simulations	71
5.5.2	Tests sur des données réelles	76
5.6	Conclusion	77

5.1 Introduction

La minimisation d'un critère pénalisé conduit généralement à une solution efficace pour un problème inverse mal posé. Nous avons vu dans les chapitres précédents l'utilité et l'efficacité de cette stratégie pour restaurer les signaux DOSY. Lorsque le problème inverse est non linéaire et de grande dimension, une complexité de calcul importante accompagnée d'un coût mémoire élevé s'ajoutent aux contraintes du problème initial. Nous nous intéresserons dans ce chapitre à résoudre le problème inverse qui se pose en spectrométrie de masse. Nous présenterons dans la Section 5.2 un état de l'art sur cette application et en Section 5.3 les méthodes proposées pour analyser les spectres de masses. Puis, une nouvelle modélisation du problème sera proposée dans la Section 5.4, ainsi qu'une nouvelle approche par dictionnaire. Nous finirons ce chapitre avec la Section 5.5 où nous présenterons une application de notre approche par dictionnaire ainsi qu'une version approchée bénéficiant d'un coût mémoire limité pour restaurer des spectres de masse synthétiques et réels.

5.2 Revue de la littérature

5.2.1 Principe de la SM

L'histoire de la spectrométrie de masse a débuté avec le britannique Joseph John Thomson en 1897 quand il a découvert la notion d'électron. Au début du XX^{ième} siècle, Thomson a construit le premier spectromètre de masse pour déterminer les masses et les charges des entités chimiques dans un mélange donné et il a reçu pour cela le prix Nobel de Physique en 1906. Cette application assure la détection, l'identification et la caractérisation des structures chimiques des molécules. Elle est caractérisée par une sensibilité importante et une haute sélectivité. Les récents développements de cette application ont élargi son champ d'application tel que l'anti-dopage [Hemmersbach 2008], la qualité de contrôle pharmaceutique et bio-pharmaceutique [Goto 2003, Paudel 2015], le diagnostic et le profilage des tumeurs [Schwartz 2004], le métabolomique [Dettmer 2007] et l'analyse protéomique [Aebersold 2003], etc.

Son principe réside à ioniser un composé chimique et séparer ses ions selon leur rapport masse sur charge (m/z). Puis, un détecteur compte les ions les plus abondants et amplifie leurs réponses comme il est schématisé dans la Figure 5.1. Un spectre de masse est obtenu à la fin du processus qui présente l'abondance des ions en fonction de leur rapport masse sur charge.

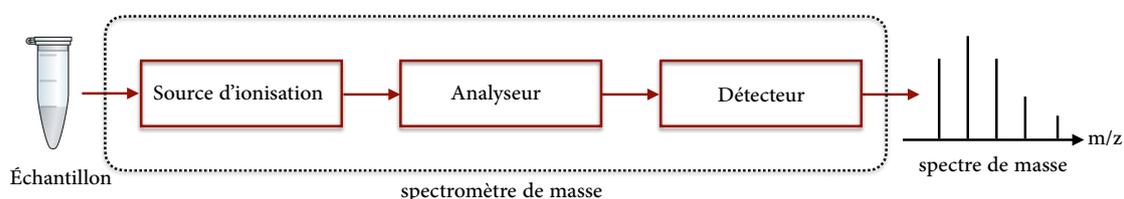


FIGURE 5.1 – Représentation schématique d'un spectromètre de masse

La spectrométrie de masse se caractérise par une grande diversité des spectromètres disponibles sur le marché ou développés au sein des laboratoires de recherche. Cette diversité se retrouve au niveau de la source d'ionisation et l'analyseur.

5.2.1.1 Source d'ionisation

La première étape de mesure en spectrométrie de masse est l'ionisation de l'échantillon pour passer de son état initial à un état gazeux. Différentes sources d'ionisation peuvent être utilisées en spectrométrie de masse soit en mode négatif pour étudier les ions négatifs, soit en mode positif pour étudier les ions positifs.

- Ionisation chimique [Munson 1966] : Cette méthode utilise un gaz réactif ionisé par un faisceau d'électrons et qui donne des ions qui interagissent à leur tour avec les entités à analyser.
- Source d'ionisation laser assistée par matrice (MALDI) [Karas 1988] : L'ionisation MALDI (Matrix-Assisted Laser Desorption Ionisation) utilise un faisceau laser pulsé. Cette technique s'adresse souvent aux molécules polaires, et peu volatiles comme les peptides et les protéines.
- Source electrospray (ESI) [Whitehouse 1989] : L'ionisation par électronébulisation s'effectue à pression atmosphérique et consiste à nébuliser une solution sous une tension élevée, et de transférer dans le vide les ions formés dans la solution. Elle est connue par sa technique douce qui contrôle l'énergie interne du système et qui assure l'étude de la stabilité des données en phase gazeuse.

La liste complète des sources d'ionisation est très large, chacune des sources repose sur un principe physique différent.

5.2.1.2 Analyseur

Après avoir ionisé toutes les entités chimiques de l'échantillon, un analyseur prend en charge la séparation des ions selon leur rapport masse sur charge. De nombreux analyseurs sont disponibles. Toutefois, comme pour les sources d'ionisation, le choix de l'analyseur se base sur le type de l'analyse à réaliser.

- TOF [Chernushevich 2001] : L'analyseur à temps de vol (Time Of Flight) est souvent associé à une source MALDI. Il consiste à accélérer les ions par un champ électrique. Les ions de même charge subissent la même quantité de mouvement. Ils ont une vitesse inversement proportionnelle à la racine carrée de leur masse. Pour déterminer la masse de chaque ion, il suffit de calculer le temps que chaque ion nécessite pour atteindre un détecteur soumis à une distance connue. On appelle ce temps "temps de vol".
- ORBITRAP [Hu 2005] : Cet analyseur est constitué d'une électrode externe en forme de cylindre et d'une électrode coaxiale interne qui piège les ions dans un mouvement orbital. Le signal émis par les ions piégés est détecté et converti en un spectre de masse par transformée de Fourier, dont la fréquence est inversement proportionnelle à la racine carrée de leur masse.
- FTICR [Amster 1996] : Le spectromètre à transformée de Fourier est caractérisé par la plus haute résolution et précision de masse. Cette technique est basée sur l'excitation des ions dans une cellule résonance cyclotronique des ions (ICR) sous l'action d'un champ magnétique. Le signal mesuré en temps sera converti en spectre de fréquence inversement proportionnelle à la masse via la transformée de Fourier. Le spectromètre FTICR est très utilisé dans le domaine pharmaceutique et biologique du fait de sa grande exactitude et ses hautes performances. Dans cette thèse, nous utiliserons dans nos tests des données mesurées avec ce type de spectromètre.

5.2.1.3 Détecteur

Le rôle d'un détecteur est de compter les ions et transformer le courant ionique en un signal mesurable même s'il est de faible intensité. On utilise souvent le détecteur destructif qui assure la détection des ions et amplifie le courant ionique émis en un signal mesurable. Il est dit destructif car il décharge les ions après les avoir détectés. Notons que l'analyseur Orbitrap et FT-ICR sont des détecteurs non-destructifs.

5.2.2 Le spectre de masse

Le spectre de masse est le résultat fourni par un spectromètre. Il s'agit généralement d'un graphe qui présente une série de pics à différentes positions. L'intensité de chaque pic présente l'abondance de l'ion ayant le rapport masse sur charge m/z sur l'axe des abscisses. L'unité de masse est le Dalton en l'honneur de J. Dalton telle que 1 Dalton correspond à $1/12$ de la masse d'un atome de Carbone. Le pic le plus intense correspond à l'ion le plus abondant appelé "pic de base". Les autres pics se présentent en % de l'intensité du pic de base. L'information la plus importante recherchée dans un spectre de masse est la masse mono-isotopique qui correspond à la masse exacte du pic constitué des isotopes majoritaires de chaque élément présent. Nous expliquerons cette notion dans la suite de ce manuscrit.

En spectrométrie de masse, pour quantifier la distinction entre deux pics, la résolution est définie par $R = m/\Delta m$ où Δ présente la différence de masse entre deux pics voisins ayant des intensités proches et m la masse du premier pic (Figure 5.2). Aussi, on peut estimer la résolution avec la finesse du pic donnée par $m/\Delta m$ où Δm est la largeur du pic à mi-hauteur et m la masse du pic. Plus la résolution est élevée, plus le pic est fin et plus il sera possible de distinguer des pics proches.

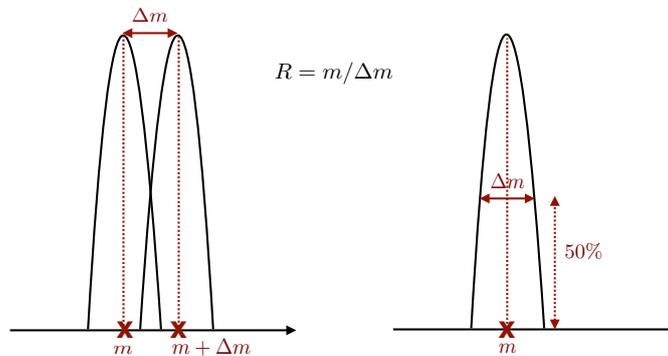


FIGURE 5.2 – Résolution du pic

5.2.3 Problème des massifs isotopiques

5.2.3.1 Notion d'isotope

Rappelons que l'atome est l'unité élémentaire de la matière quel que soit son état (gazeux, solide ou liquide). Chaque atome est présenté dans le tableau périodique par un symbole \mathbf{X} et possède un nombre atomique \mathbf{Z} qui présente le nombre de protons, et un nombre de masse \mathbf{A} qui définit la somme du nombre de protons avec le nombre de neutrons. Dans la nature, chaque atome se

présente toujours avec un nombre atomique unique, mais avec différents nombres de neutrons, l'équivalent de différents nombres de masse. On appelle ces formes du même élément chimique, les isotopes.

Les isotopes d'un même élément ont des propriétés chimiques identiques mais des propriétés physiques différentes (stabilité, radioactivité). Citons l'exemple de l'atome de Carbone ayant le symbole C et un nombre atomique $Z=6$. Le Carbone présente deux isotopes stables : le ^{12}C avec 6 neutrons, le ^{13}C avec 7 neutrons, et un isotope radioactif : le ^{14}C avec 8 neutrons. Chacun des isotopes se présente avec une abondance spécifique dans la nature.

La masse mono-isotopique est la masse du premier pic du profil isotopique fourni par le spectromètre (Figure 5.3). Cette masse ne prend en compte que les masses des isotopes les plus abondants. Une identification avec une résolution suffisante de la structure chimique d'un échantillon peut être facilement déduite à partir de cette masse. C'est l'information la plus importante recherchée dans un spectre de masse.

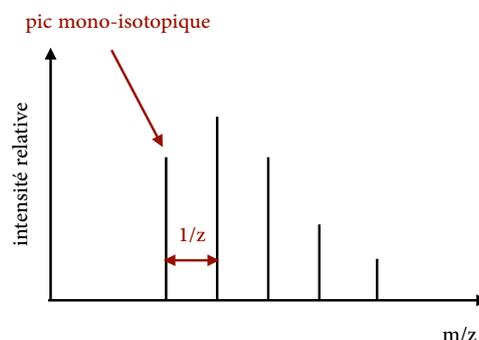


FIGURE 5.3 – Exemple du profil isotopique

La différence entre deux pics successifs est égale à $1.003/z$ qui s'approxime en $1/z$. Par conséquent, la charge z peut être déterminée en comptant le nombre des pics sur une unité du spectre.

5.2.3.2 Distribution isotopique

La spectrométrie de masse est tellement précise qu'elle permet de mesurer non seulement la masse du pic de base, mais plutôt les masses de différents isotopes des atomes présents dans la molécule cible, incluant idéalement la masse mono-isotopique. Par conséquent, différents pics seront présents dans le spectre de masse.

Prenons l'exemple de la série des alcanes de formule $\text{C}_n\text{H}_{2n+2}$. En traçant le profil isotopique de cette molécule pour différentes valeurs de n (Figure 5.4), on voit bien que plus la molécule est grande, plus le nombre de pics augmente et le pic mono-isotopique sera décalé. Bien que la masse mono-isotopique permette de déterminer la structure chimique de l'échantillon, sa mesure pour les molécules de grande taille est donc difficile car l'intensité du pic mono-isotopique devient très faible. D'autre part dans le cas d'un mélange complexe, la probabilité d'avoir différents états de charge conduit à une superposition des profils isotopiques, ce qui rend le problème encore plus compliqué. En outre, un niveau de bruit élevé peut cacher complètement les pics de faibles intensités et conduit à une perte d'information.

Une partie de ce travail de thèse consiste à développer une méthode qui assure la détection précise des pics mono-isotopiques en analyse protéomique. Seules les molécules de protéines seront prises

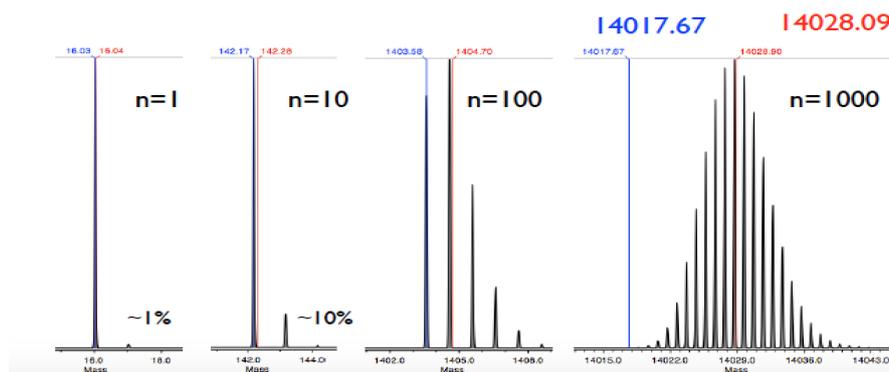


FIGURE 5.4 – Profil isotopique des alcanes pour différentes valeurs de n . (Valeur en bleu : masse mono-isotopique, Valeur en rouge : masse moyenne du massif)

en compte dans notre étude. Rappelons que les protéines sont de grosses molécules fondamentales de la matière vivante. Elles sont formées par de longues chaînes d'acides aminés et elles assurent une multitude de fonctions biologiques comme la structure des cellules, la régulation des gènes, le rôle de catalyseur du processus biologique, etc. La formule générale de la protéine est $C_{N_C}H_{N_H}O_{N_O}N_{N_N}S_{N_S}$ telle que N_C, N_H, N_O, N_N , et N_S définissent respectivement le nombre de Carbone C, Hydrogène H, Oxygène O, Azote N et Soufre S. Nous présentons dans le Tableau 5.1 les isotopes stables de chaque atome présent dans la formule de la protéine ainsi que leur masse et leur abondance relative.

Atome	Masse (en Dalton)	Abondance relative
^{12}C	12 (<i>par définition</i>)	0.9893
^{13}C	13.0033548378	0.0107
^1H	1.00782503207	0.999885
^2H	2.0141017778	0.000115
^{16}O	15.99491461956	0.99757
^{17}O	16.99913170	0.00038
^{18}O	17.9991610	0.00205
^{14}N	14.0030740048	0.99636
^{15}N	15.0001088982	0.00364
^{32}S	31.97207100	0.9499
^{33}S	32.97145876	0.0075
^{34}S	33.96786690	0.0425

Tableau 5.1 – Isotopes et abondances relatives des atomes de la protéine [Audi 1993, Rosman 1999]

Pour les protéines ne contenant que les éléments C, H, O, N , et S , dont l'isotope majoritaire est celui de plus faible masse, le pic mono-isotopique est le pic de plus faible masse. Cependant, il n'est pas toujours présent surtout pour des protéines de taille importante où la statistique conduit à une intensité négligeable pour ce pic monoisotopique.

5.3 Méthodes de traitement de données en SM

En 1991, Skilling a proposé une méthode bayésienne basée sur le maximum d'entropie pour déconvoluer des spectres de masses bruités. La déconvolution en spectrométrie de masse consiste à améliorer la résolution des pics du spectre mono-isotopique restauré. L'application de cette approche permet d'assurer la séparation de plusieurs pics chevauchés et les masses mono-isotopiques obtenues étaient très proches des masses mono-isotopiques théoriques. La résolution du problème

de spectrométrie de masse par cette méthode, si elle mène à des résultats efficaces, présente par contre une complexité de calcul très importante en grande dimension [Skilling 1984].

À ce jour, la technique la plus utilisée pour analyser un spectre de masse est le "peak picking" qu'on trouve dans tous les logiciels d'analyse biophysique (TOPSpin, NMRNotebook, OpenMS, etc). Le principe de cette technique consiste à seuiller le spectre de masse : seuls les pics d'intensités élevés sont traités et les pics de faibles intensités sont ignorés. La valeur du seuil peut être égale à une valeur liée au niveau de bruit estimé, ou choisie manuellement. Néanmoins, en présence d'un niveau de bruit élevé, et avec des pics mono-isotopiques de faibles intensités, cette opération conduit systématiquement à une perte d'information.

Fondé sur ce principe de seuillage, un algorithme NITPICK (Non-greedy Iterative Template-based peak PICKer) [Renard 2008] a été proposé en 2008 pour résoudre le problème sous la forme d'une régularisation avec la norme ℓ_1 [Tibshirani 1996]. Avec des signaux simulés et réels, Renard *et al.* ont montré l'efficacité de leur algorithme. Dans le même contexte, Slawski [Slawski 2010] a proposé une autre méthode qui repose sur une étape du pré-traitement avec du peak picking suivi d'une pénalisation avec une contrainte de positivité [Slawski 2010]. Bien que les résultats montrés dans [Renard 2008, Slawski 2012] assurent l'analyse précise d'un spectre de masse, le recours à la technique du peak picking pose toujours une contrainte d'instabilité et une sensibilité au niveau de bruit. En outre, l'applicabilité de ces deux algorithmes reste limitée à des dimensions réduites des spectres de masses.

5.4 Nouvelle approche par dictionnaire

5.4.1 Modélisation du problème

Nous considérons qu'un spectre de masse est décrit par une série des $(m_p^{\text{iso}}, z_p, a_p)_{1 \leq p \leq P}$ où $m_p \in]0, +\infty[$ définit la masse mono-isotopique, $z_p \in \mathbb{N}^*$ désigne l'état de charge et a_p présente l'abondance de la protéine p dans un échantillon formé de P protéines. Nous modélisons ce spectre noté y par la somme pondérée des différentes distributions isotopiques :

$$y = \sum_{p=1}^P a_p D(m_p^{\text{iso}}, z_p) + b \quad (5.1)$$

où b modélise le bruit d'acquisition. L'objectif principal est alors de déterminer les valeurs $(m_p^{\text{iso}}, z_p, a_p)_{1 \leq p \leq P}$ à partir du spectre y .

Étant donné M valeurs de masse et P protéines, nous considérons que le spectre de masse peut être donné par :

$$\mathbf{y} = \sum_{p=1}^P a_p \mathbf{d}(m_p^{\text{iso}}, z_p) + \mathbf{b} \quad (5.2)$$

où $\mathbf{y} \in \mathbb{R}^M$, $\mathbf{d}(m_p^{\text{iso}}, z_p) \in [0, +\infty[^M$ et $\mathbf{b} \in \mathbb{R}^M$. La résolution de ce problème présente une difficulté d'une part dans la complexité de la fonction \mathbf{d} et sa non linéarité et d'autre part dans la grande dimension de y .

Le modèle "averagine" proposé par Senko *et al.* [Senko 1994] détermine la formule de la protéine avec des arrondis à l'entier des nombres d'atomes moyens pour toute masse donnée m . Il est défini

par l'opérateur \mathcal{A} tel que :

$$\begin{aligned} \mathcal{A} : \mathbb{R} &\rightarrow \mathbb{R}^5 \\ m &\rightarrow (N_C, N_H, N_O, N_N, N_S) \end{aligned} \quad (5.3)$$

Par conséquent, pour toute masse m la distribution isotopique D peut être systématiquement déduite, ainsi que sa forme échantillonnée \mathbf{d} . Nous adoptons ce modèle pour résoudre le problème de (5.2) sous la condition que l'axe de masse et les états de charge des P protéines soient donnés.

5.4.2 Approche par dictionnaire

Soit un spectre de masse $\mathbf{y} \in \mathbb{R}^M$ défini sur un axe de masse $[m_{\min}, m_{\max}]$ ayant Z états de charge donnés dans $[z_{\min}, z_{\max}]$. Nous définissons le dictionnaire $\mathbf{D} \in \mathbb{R}^{M \times T}$ avec $T := M \times Z$:

$$\mathbf{D} = [\mathbf{D}_1 | \dots | \mathbf{D}_\ell | \dots | \mathbf{D}_Z] \quad (5.4)$$

où pour tout état de charge $\ell \in \{1, \dots, Z\}$, la sous-matrice $\mathbf{D}_\ell \in \mathbb{R}^{M \times M}$ définit le dictionnaire associé à la charge ℓ comme il est présenté dans la Figure 5.5. Ainsi, la $i^{\text{ième}}$ colonne de la matrice \mathbf{D} est la version échantillonnée \mathbf{d} de la distribution isotopique donnée par le modèle averagine pour la $i^{\text{ième}}$ masse sur la grille de taille T définie par :

$$(\forall i \in \{1, \dots, T\}) \quad m_i = m_{\min} + (j - 1)m_{\max} \quad (5.5)$$

$$z_i = z_{\min} + (\ell - 1)z_{\max} \quad (5.6)$$

avec $i = \ell M + j$, $j \in \{1, \dots, M\}$ et $\ell \in \{1, \dots, Z\}$.

Selon cette modélisation, le problème SM peut être reformulé par :

$$\mathbf{y} = \mathbf{D}\bar{\mathbf{x}} + \underline{\mathbf{b}} \quad (5.7)$$

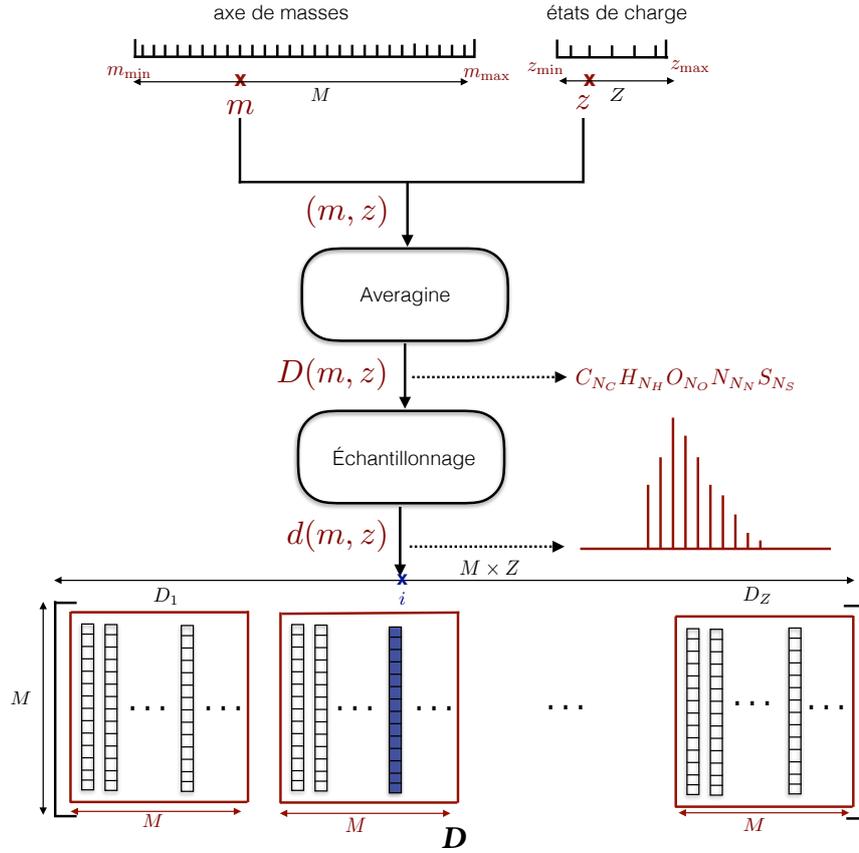
où $\bar{\mathbf{x}} \in \mathbb{R}^T$ représente un signal parcimonieux de degré de parcimonie P , tels que les P coefficients positifs non nuls nous mènent à déterminer la masse mono-isotopique et l'état de charge des différentes protéines. Nous désignons par $\underline{\mathbf{b}}$ la somme du bruit d'acquisition et des erreurs de modélisation.

5.4.3 Stratégie d'optimisation

Pour résoudre le problème (5.7) (vu le mauvais conditionnement du dictionnaire \mathbf{D} et la présence du bruit) nous nous proposons d'utiliser l'approche de minimisation sous contrainte :

$$\underset{\mathbf{x} \in \mathbb{R}_+^T}{\text{minimiser}} \quad \Psi(\mathbf{x}) \quad \text{tel que} \quad \|\mathbf{D}\mathbf{x} - \mathbf{y}\| \leq \eta \quad (5.8)$$

où Ψ est une fonction de pénalisation proposée pour renforcer la parcimonie du signal cible $\bar{\mathbf{x}}$. Nous choisissons dans ce cadre d'utiliser la fonction de pénalité ℓ_1 (2.8). Le paramètre η dépend du niveau du bruit, et une estimation de sa valeur sera envisagée à partir du spectre de masse mesuré. À cet égard, une solution de (5.8) peut être obtenue en utilisant l'algorithme Primal-Dual présenté dans le Chapitre 2 selon la forme :

FIGURE 5.5 – Représentation schématique de la construction du dictionnaire \mathbf{D}

où $\text{prox}_{\kappa(\Psi + \iota_{[0, +\infty[})}$ présente l'opérateur proximal de $\kappa\Psi$ sous contrainte de positivité pour $\kappa > 0$ et $\text{proj}_{\|\cdot - \mathbf{y}\| \leq \eta}$ est définie en (3.46). Le choix précis de cet algorithme est justifié par le fait que le Primal-Dual est un algorithme efficace pour les problèmes d'optimisation convexe. De plus, aucune inversion du dictionnaire \mathbf{D} n'est exigée par cet algorithme, ce qui réduit les coûts de calculs et l'encombrement mémoire dans notre cas.

5.4.4 Approximation par blocs

Bien que l'approche par dictionnaire paraisse relativement efficace pour résoudre (5.8), elle s'avère coûteuse à mettre en œuvre numériquement avec les grandes dimensions rencontrées en spectrométrie de masse. Un spectre de masse contient généralement un nombre de points $> 500\text{k}$. Par conséquent, une grande mémoire de stockage est indispensable pour stocker \mathbf{D} .

Pour pallier cette contrainte, nous proposons une approximation notée $\overline{\mathbf{D}}$ du dictionnaire exact \mathbf{D} . Cette approximation se fonde sur deux hypothèses :

- On suppose que la distribution isotopique est localement stable : pour des valeurs de masses proches, le profil isotopique diffère par une légère variation d'intensité.
- Le profil isotopique est présenté par un signal parcimonieux à valeurs positives.

En partant de ces hypothèses, l'axe de masses peut être décomposé en fenêtres de largeur $L \leq M$ sur lesquelles le caractère de stabilité est vérifié. Nous pouvons considérer le problème en spectrométrie de masse comme étant un problème de déconvolution où le spectre mesuré est un produit de

Algorithme 8 Primal-Dual pour résoudre (5.8)

$$\begin{array}{l}
 \mathbf{u}^{(0)} \in \mathbb{R}^M, \mathbf{x}^{(0)} \in \mathbb{R}^T \\
 0 < \nu\kappa\|\mathbf{D}\|^2 < 1, \kappa \in \mathbb{R}_+^* \\
 \rho \in (0, 2) \\
 \text{Pour } k = 0, 1, \dots \\
 \left[\begin{array}{l}
 \tilde{\mathbf{x}}^{(k)} = \text{prox}_{\kappa(\Psi + \iota_{[0, +\infty[})}(\mathbf{x}^{(k-1)} - \kappa\mathbf{D}^\top(\mathbf{u}^{(k-1)})) \\
 \mathbf{v}^{(k)} = \mathbf{u}^{(k-1)} + \nu\mathbf{D}(2\tilde{\mathbf{x}}^{(k)} - \mathbf{x}^{(k-1)}) \\
 \tilde{\mathbf{u}}^{(k)} = \mathbf{v}^{(k)} - \nu\text{proj}_{\|\cdot\|_{-\mathbf{y}} \leq \tau}(\mathbf{v}^{(k)}/\nu) \\
 \mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + \rho(\tilde{\mathbf{x}}^{(k)} - \mathbf{x}^{(k-1)}) \\
 \mathbf{u}^{(k)} = \mathbf{u}^{(k-1)} + \rho(\tilde{\mathbf{u}}^{(k)} - \mathbf{u}^{(k-1)})
 \end{array} \right.
 \end{array}$$

convolution entre le pic mono-isotopique de masse m et la distribution isotopique qui correspond à la masse moyenne de la fenêtre de largeur $L \leq M$ contenant la masse m .

Soit $\bar{\mathbf{d}}_{j,\ell}$ la distribution isotopique correspondant à la masse moyenne de la $j^{\text{ième}}$ fenêtre de charge ℓ . L'approximation proposée de chaque sous-matrice \mathbf{D}_ℓ a la forme :

$$\bar{\mathbf{D}}_\ell = \text{BDiag} \left([\text{Circ}(\bar{\mathbf{d}}_{j,\ell})]_{1 \leq j \leq M/L} \right) \quad (5.9)$$

où $\text{BDiag}(\cdot)$ définit une matrice bloc diagonale de M/L matrices circulantes (Circ) ayant $\bar{\mathbf{d}}_{j,\ell}$ en première ligne pour tout $j \in \{1, \dots, M/L\}$ (Figure 5.6).

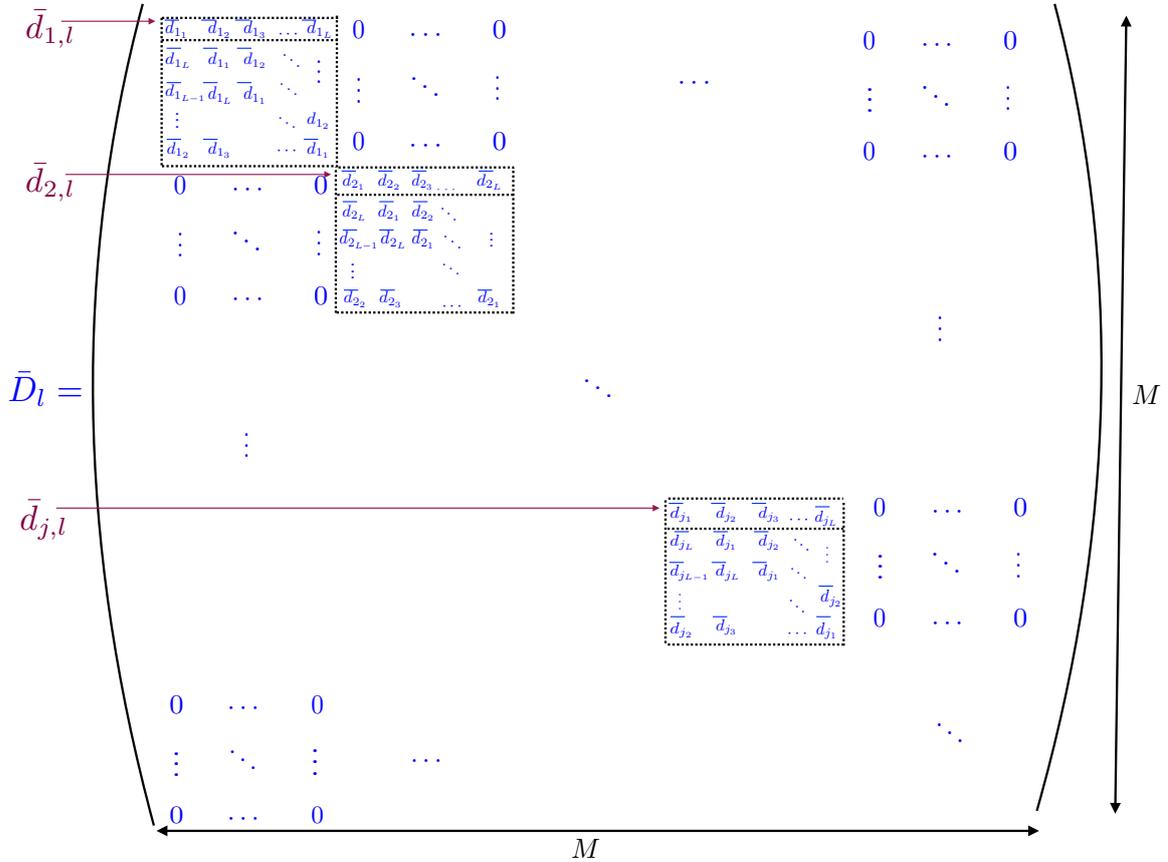


FIGURE 5.6 – Représentation schématique de l'approximation par blocs $\bar{\mathbf{D}}_\ell$ de M/L matrices circulantes

Par conséquent, l'approximation du dictionnaire \mathbf{D} est donnée par :

$$\overline{\mathbf{D}} = [\overline{\mathbf{D}}_1 | \dots | \overline{\mathbf{D}}_\ell | \dots | \overline{\mathbf{D}}_Z] \quad (5.10)$$

Le produit matriciel de $\overline{\mathbf{D}}_\ell$ et $\overline{\mathbf{D}}_\ell^T$ avec un vecteur peut être simplement calculé par la transformée de Fourier. Par conséquent, l'algorithme Primal-Dual reste utile et facile à utiliser en remplaçant la matrice \mathbf{D} par $\overline{\mathbf{D}}$.

5.5 Résultats et validation

Afin de valider notre approche par dictionnaire et son approximation par blocs, nous proposons d'effectuer différents tests avec des signaux simulés de différents états de charge. En outre, pour illustrer la pertinence de notre nouvelle approche, nous proposons d'analyser des données expérimentales réelles.

5.5.1 Tests de simulations

5.5.1.1 Signaux simulés

Nous considérons les signaux suivants :

- un signal A mono-chargé ($Z = 1$) formé de 10 protéines, et de taille $M = 3000$ (Figure 5.7).
- un signal B multi-chargé ($Z = 3, z \in (1, 2, 3)$) formé de 50 protéines et possédant une taille $M = 4000$ (Figure 5.8).
- un signal C multi-chargé ($Z = 4, z \in (1, 2, 3, 4)$) formé de 100 protéines et possédant une taille $M = 5000$ Figure (5.9).

Les masses varient de $m_{\min} = 1000$ Dalton jusqu'à $m_{\max} = 1100$ Dalton. Un bruit gaussien d'écart type $\sigma > 0$ est ajouté aux données simulées.

Nous avons effectué nos tests sous Python, sur une machine Macintosh Mac Pro Intel Xeon qui contient 8 processeurs de capacité 12 GB et le système d'exploitation MacOSX.

Un code couleur est utilisé pour distinguer les états de charge.

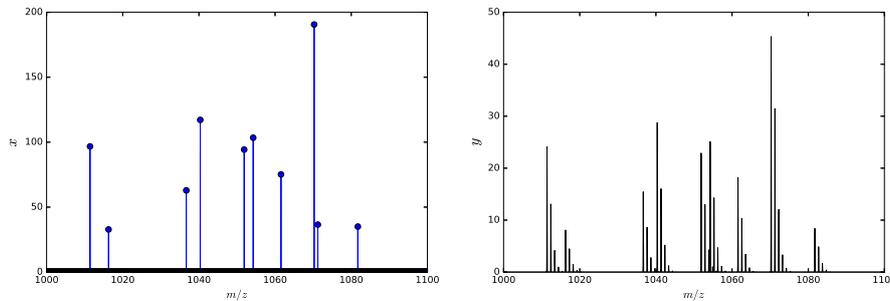


FIGURE 5.7 – Signal synthétique A mono-chargé ($M = 3000, P = 10, Z = 1$) avec $\sigma = 10^{-2}$ (à gauche : le signal cible, à droite : le spectre mesuré)

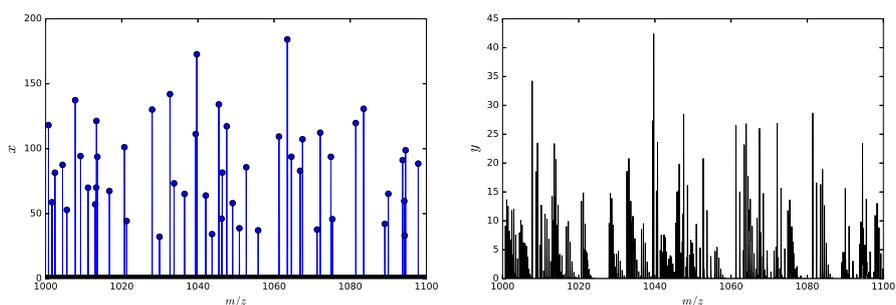


FIGURE 5.8 – Signal synthétique B multi-chargé ($M = 4000, P = 50, Z = 3$) avec $\sigma = 10^{-2}$ (à gauche : le signal cible, à droite : le spectre mesuré)

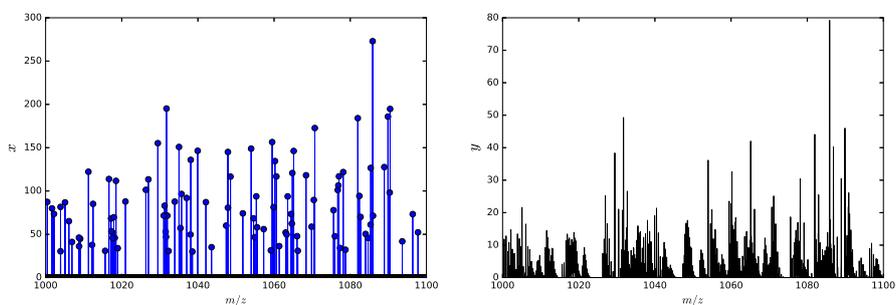


FIGURE 5.9 – Signal synthétique C multi-chargé ($M = 5000, P = 100, Z = 4$) avec $\sigma = 10^{-2}$ (à gauche : le signal cible, à droite : le spectre mesuré)

5.5.1.2 Test de l'approche avec dictionnaire exact

Dans cette section, nous nous proposons de valider l'approche par dictionnaire exact **D**. Les Figures 5.10, 5.11 et 5.12 présentent les résultats de reconstruction des signaux A, B et C avec l'Algorithme 8 pour 1000 itérations.

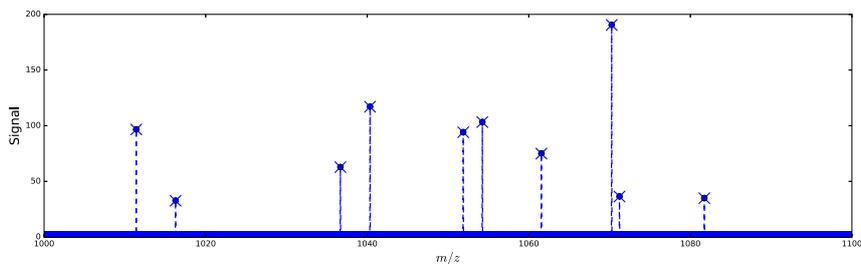


FIGURE 5.10 – Résultat de reconstruction du signal synthétique A mono-chargé ($Z = 1$) avec **D** (le signal original est présenté par les " — • " et le signal reconstruit avec les " × ")

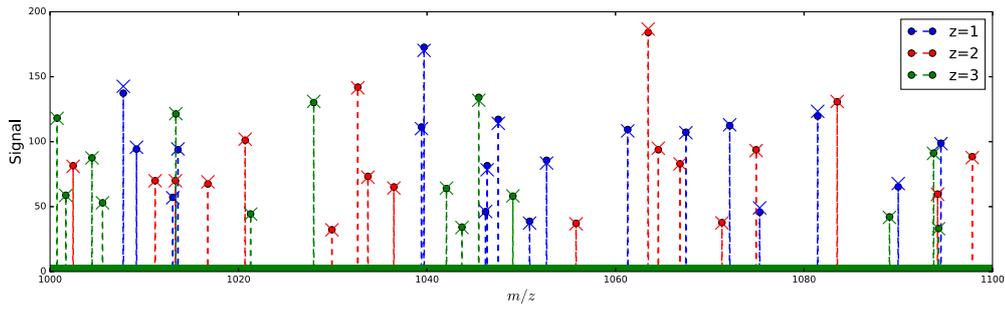


FIGURE 5.11 – Résultat de reconstruction du signal synthétique B multi-chargé ($Z = 3$) avec \mathbf{D} (le signal original est présenté par les " — • " et le signal reconstruit avec les " × ")

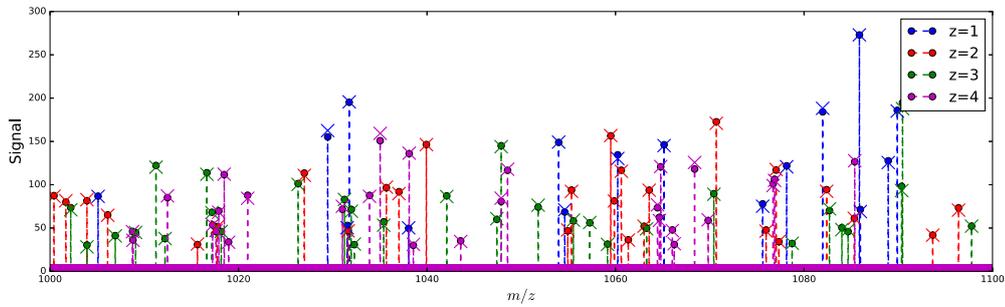


FIGURE 5.12 – Résultat de reconstruction du signal synthétique C multi-chargé ($Z = 4$) avec \mathbf{D} (le signal original est présenté par les " — • " et le signal reconstruit avec les " × ")

Les résultats montrent que l'approche par dictionnaire exact assure une reconstruction efficace des profils mono-isotopiques d'un spectre de masse avec un $\text{RSB} = 33.3$ dB pour le signal A, un $\text{RSB} = 35.3$ dB pour le signal B et un $\text{RSB} = 32.5$ dB pour le signal C. Nous allons par la suite comparer cette approche avec l'approximation par blocs $\overline{\mathbf{D}}$.

5.5.1.3 Validation de l'approximation par blocs

Cette section est consacrée à la validation de l'approximation par blocs $\overline{\mathbf{D}}$ et la comparaison de ses résultats avec l'utilisation du dictionnaire exact \mathbf{D} .

Nous présentons dans la Figure 5.15 le résultat de reconstruction des signaux A, B et C avec l'approche par dictionnaire exact \mathbf{D} et son approximation par blocs $\overline{\mathbf{D}}$ pour une largeur de fenêtres $L = 10$.

La qualité de reconstruction des signaux A, B et C avec l'approximation par blocs $\overline{\mathbf{D}}$ paraît équivalente à celle obtenue avec l'approche par dictionnaire exacte \mathbf{D} .

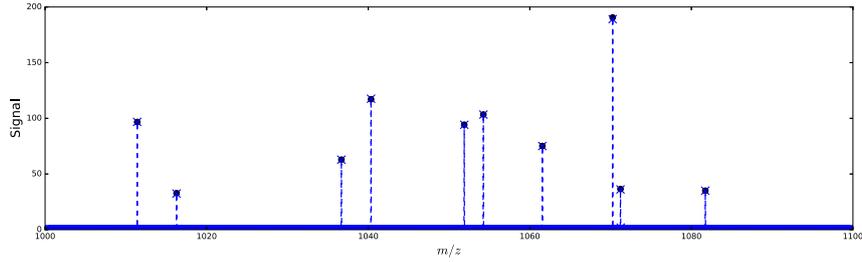


FIGURE 5.13 – Résultats de reconstruction du signal A mono-chargé : Le signal original est en "•", le signal reconstruit avec l'approche par dictionnaire \mathbf{D} en ("×") et le signal reconstruit avec l'approximation par blocs $\overline{\mathbf{D}}$ ($L = 10$) en ("--").

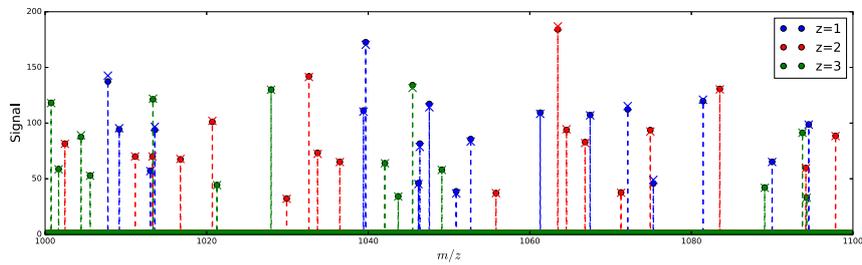


FIGURE 5.14 – Résultats de reconstruction du signal B multi-chargé ($z \in (1, 2, 3)$) : Le signal original est en "•", le signal reconstruit avec l'approche par dictionnaire \mathbf{D} en ("×") et le signal reconstruit avec l'approximation par blocs $\overline{\mathbf{D}}$ ($L = 10$) en ("--").

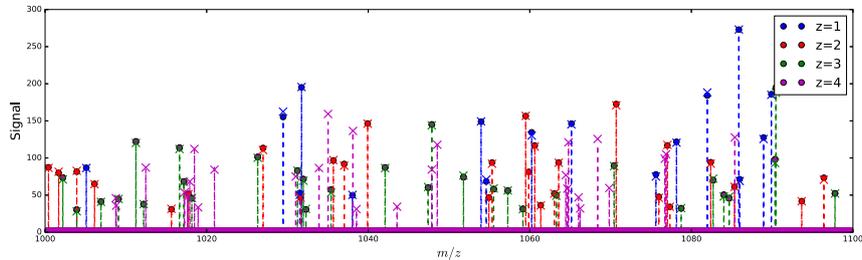


FIGURE 5.15 – Résultats de reconstruction du signal C multi-chargé ($z \in (1, 2, 3, 4)$) : Le signal original est en "•", le signal reconstruit avec l'approche par dictionnaire \mathbf{D} en ("×") et le signal reconstruit avec l'approximation par blocs $\overline{\mathbf{D}}$ ($L = 10$) en ("--").

Le Tableau 5.2 présente le RSB et le temps de calcul obtenus avec \mathbf{D} et $\overline{\mathbf{D}}$ ainsi que la mémoire exigée par le stockage du dictionnaire \mathbf{D} . Pour les différents niveaux de bruit, le signal reconstruit avec les deux méthodes possède le degré de parcimonie \hat{P} exact avec des RSB satisfaisants. Nous pouvons conclure que les deux approches : par dictionnaire exact et son approximation par blocs sont robustes par rapport au niveau de bruit.

Par ailleurs, la construction de \mathbf{D} exige une grande mémoire et un temps de calcul plus important que $\overline{\mathbf{D}}$. Par exemple, la reconstruction du signal B avec $\overline{\mathbf{D}}$ pour $\sigma = 10^{-1}$ et $L = 10$ produit une qualité de reconstruction équivalente à celle donnée par le dictionnaire exact \mathbf{D} et est 4 est fois plus rapide.

		Approche par dictionnaire				Approximation par blocs												
		σ	RSB	Temps	\hat{P}	Mémoire	$L = 2$			$L = 10$			$L = 50$			$L = 100$		
							RSB	Temps	\hat{P}	RSB	Temps	\hat{P}	RSB	Temps	\hat{P}	RSB	Temps	\hat{P}
Signal A $P=10$	10^{-1}	12.36	10.40	10	72	12.36	15.59	10	12.38	3.44	10	12.43	0.98	10	12.41	0.60	10	
	10^{-2}	33.36	10.35	10		33.82	15.53	10	33.37	3.56	10	34.49	0.96	10	32.82	0.63	10	
	10^{-3}	59.09	12.66	10		57.53	15.50	10	41.66	3.47	10	41.06	1.00	10	34.22	0.63	10	
Signal B $P=50$	10^{-1}	14.67	54.94	50	384	14.70	56.04	50	14.73	11.85	50	14.66	2.87	50	14.73	11.55	50	
	10^{-2}	34.67	54.78	50		35.00	57.70	50	35.15	13.06	50	32.44	3.32	50	35.15	11.89	50	
	10^{-3}	35.32	55.87	50		38.24	54.79	50	39.19	11.66	50	34.29	3.23	50	39.19	11.58	50	
Signal C $P=100$	10^{-1}	13.91	220.84	100	800	13.91	134.09	100	13.903	22.70	100	13.84	5.61	100	13.90	39.99	100	
	10^{-2}	33.94	118.03	100		33.84	112.55	100	33.55	26.81	100	31.34	4.52	100	33.55	19.02	100	
	10^{-3}	32.52	104.70	100		36.91	85.54	100	40.00	18.52	100	33.91	4.71	100	40.00	18.78	100	

Tableau 5.2 – RSB (en dB), Temps de calcul (en s), mémoire de stockage de \mathbf{D} (en Mb) et nombre de pics \hat{P} détectés pour différentes valeurs de σ . L'approximation par blocs $\bar{\mathbf{D}}$ est testée pour différentes largeurs de fenêtres L (en Dalton)

De plus, nous remarquons que plus L est grand, plus la reconstruction avec $\bar{\mathbf{D}}$ est rapide mais avec une légère détérioration en qualité de reconstruction ce qui était attendu. La Figure 5.16 montre que la qualité de reconstruction avec l'approximation par blocs est assez stable par rapport à la valeur de L .

En conclusion, même si la qualité de restauration obtenue avec l'approximation par blocs présente une légère dégradation en comparaison avec l'approche par dictionnaire exact, elle paraît très utile et satisfaisante pour éviter les contraintes de mémoires et les calculs numériques coûteux.

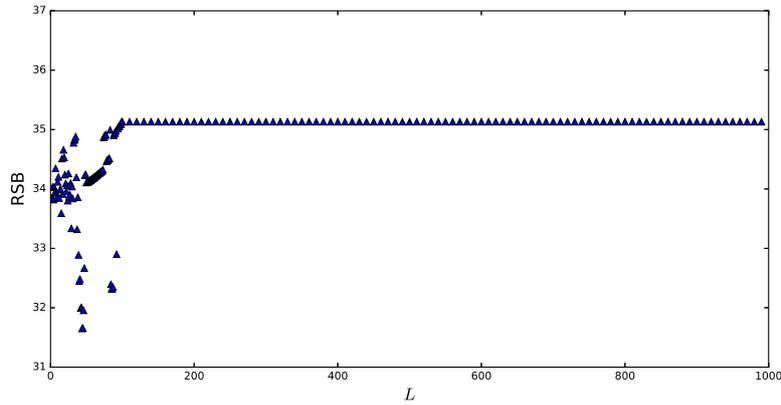


FIGURE 5.16 – Résultats de reconstruction du signal A avec l'approximation par blocs pour différentes valeurs de L ($\sigma = 10^{-2}$)

Remarque : En comparant les résultats avec les signaux originaux, nous avons remarqué une légère sous-estimation d'intensités (Figure 5.17). Cela provient fort probablement de la fonction de pénalité ℓ_1 utilisée dans ce cas. Nous y reviendrons dans le Chapitre 6.

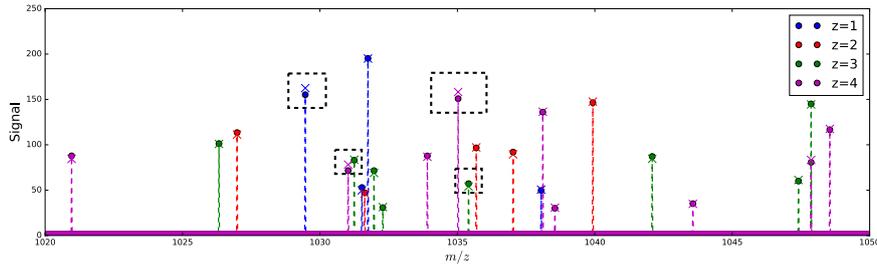


FIGURE 5.17 – Zoom sur la reconstruction du signal C avec l’approche dictionnaire exacte \mathbf{D} (présenté avec les ” \times ”) et l’approximation par blocs $\overline{\mathbf{D}}$ pour $L = 10$ (présenté avec les ” $-$ ”). Le signal originale est présenté avec ” \bullet ”

5.5.2 Tests sur des données réelles

Nous considérons un spectre de masse réel, mesuré avec Bruker Solarix 15T, en utilisant un spectromètre de masse caractérisé par une source d’ionisation ESI et un analyseur FT-ICR. L’échantillon cible est constitué de $3\mu\text{M}$ du peptide de séquence EVEALEKKVAALESKVQALEKKVEALEHG-NH₂ et de formule brute $\text{C}_{140}\text{H}_{240}\text{N}_{38}\text{O}_{45}$. Cet échantillon est mesuré dans des conditions natives dans 50 mM de NH_4OAc . Le spectre mesuré présente une taille $M = 18.130.981$ avec des états de charges $z \in (1, 2, 3, 4, 5)$ et un axe de masse $[m_{\min}, m_{\max}]$ où $m_{\min} = 153.57$ Dalton, $m_{\max} = 4999.96$ Dalton. Le peptide d’intérêt est de charge $z = 5$.

Supposons qu’on applique l’approche par dictionnaire exact pour restaurer le profil mono-isotopique de ce signal, le dictionnaire \mathbf{D} possèdera les dimensions $(M, M \times Z)$ avec $M \times Z = 90.654.905$, l’équivalent de quelques Térabytes. Il est clair que ce sont des dimensions prohibitives. Ceci prouve l’importance de notre approximation par blocs.

La Figure 5.18 présente notre résultat d’analyse de ce spectre de masse avec l’approximation par blocs pour $L = 10$ sur la zone $[1890, 1920]$ Dalton. L’objectif final est de déterminer les rapports m/z de tous les pics obtenus. La Figure 5.18 présente la mesure et le profil isotopique correspondant. La distribution isotopique possède un état de charge $z = 5$. Nous avons un pic majeur à une masse $m = 9526.439$ Dalton. Cette valeur correspond bien à la masse mono-isotopique théorique du peptide qui égale à $m = 9526.337$ Dalton. Aussi, nous avons un deuxième pic à une distance de $+21.959$ Dalton par rapport au premier pic majeur. Cette information confirme la présence du Sodium car la masse mono-isotopique du Sodium est égale $+21.982$ Dalton.

À ce jour, nos analyses des spectres de masses mono-chargés et multi-chargés produisent des résultats très proches des résultats théoriques attendus et l’identification chimique des échantillons est efficacement déterminée. Ceci confirme la grande fiabilité de notre approche.

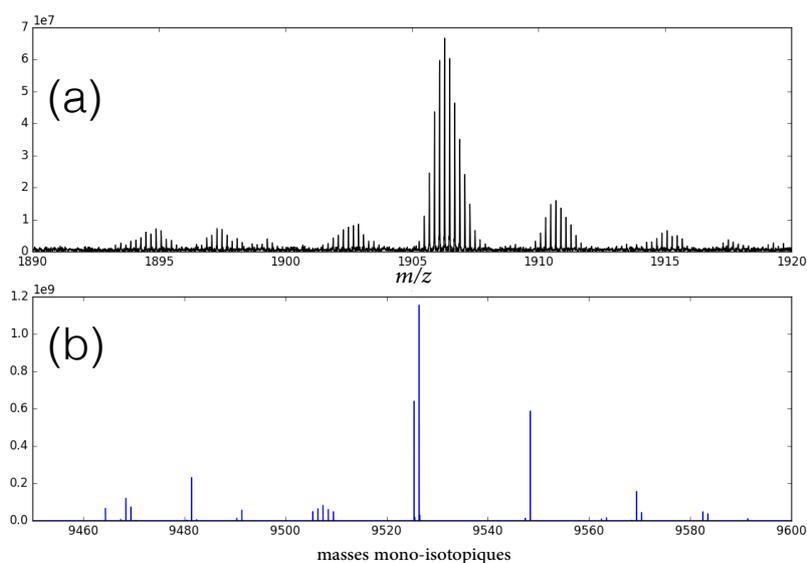


FIGURE 5.18 – Zoom sur le résultat d’analyse du peptide EVEALEKKVAALESKVQALEKKVEALEHG-NH₂. a- spectre de masse, b- signal reconstruit avec l’approximation par blocs $\overline{\mathbf{D}}$ pour $L = 10$ et $z = 5$

5.6 Conclusion

Ce chapitre a été consacré à la résolution du problème inverse en spectrométrie de masse. Pour répondre à cet objectif, nous avons proposé une nouvelle approche par dictionnaire. Cette méthode se fonde sur le modèle averagine dédié à l’analyse protéomique. Avec une stratégie de minimisation d’un critère pénalisé, nous avons résolu le problème en utilisant l’algorithme Primal-Dual. Nous avons montré que la norme ℓ_1 assure une restauration efficace des profils mono-isotopiques. L’approche par dictionnaire exact présente cependant une grande complexité de calcul et nécessite une mémoire de stockage importante. Pour résoudre ce problème, nous avons proposé une approximation par blocs du dictionnaire qui mène à une qualité de reconstruction très satisfaisante avec un temps de calcul réduit et une occupation de mémoire limitée. Les résultats que nous avons montrés dans la Section 5.5 confirment l’efficacité de notre contribution et ils ont fait l’objet d’un article de conférence [Cherni 2018]. Suite à la légère sous-estimation de l’intensité dans le signal reconstruit que nous avons remarquée avec les signaux synthétiques (Section 5.5), nous proposerons dans le chapitre suivant une nouvelle fonction de pénalité capable de mieux estimer l’intensité des signaux parcimonieux.

Chapitre 6

Nouvelle pénalité pour la résolution des signaux parcimonieux en SM

Sommaire

6.1	Introduction	80
6.2	Présentation de la pénalisation ℓ_p/ℓ_q	80
6.2.1	Version exacte de ℓ_p/ℓ_q	81
6.2.2	Approximation lisse de ℓ_p/ℓ_q	81
6.2.3	Propriétés de la fonction Ψ	82
6.3	Algorithme de minimisation	87
6.3.1	Forward-Backward à métrique variable localement ajustée	87
6.3.2	Validité de l'algorithme	87
6.3.3	Calcul de l'opérateur proximal inexact	89
6.4	Analyse de convergence	89
6.5	Résultats expérimentaux	90
6.5.1	Données synthétiques	90
6.5.2	Test de reconstruction	90
6.6	Conclusion	95

6.1 Introduction

Comme nous l'avons vu dans le chapitre 5, les signaux en spectrométrie de masse sont des signaux parcimonieux. Nous avons proposé une approche par dictionnaire pour restaurer le signal cible. Par ailleurs, pour éviter les coûts de calcul élevés et le besoin d'une grande mémoire de stockage, nous avons développé une approximation par blocs. Les résultats confirment l'utilité et l'efficacité de notre méthode. Cependant, dans les cas de simulations, nous avons remarqué une légère différence d'intensité entre le signal estimé et le signal original. Cela est fort probablement dû au choix de la fonction de régularité ℓ_1 . Une stratégie classique consiste à utiliser une approche d'élimination du biais, menant à résoudre de nouveau le problème des moindres carrés, sous une contrainte de support sur la position des pics. Une idée alternative, pour surpasser ce problème, consiste à modifier la fonction de régularisation, pour se retrouver plus près de la pénalité ℓ_0 et éviter les problèmes d'échelle [Mitrinovic 2012]. En particulier, citons le travail [Repetti 2015] proposant une pénalisation basée sur le ratio de normes ℓ_1/ℓ_2 , fournissant des résultats très encourageants sur un problème de déconvolution aveugle de signaux sismiques. Nous proposons par la suite d'étendre cette pénalité avec la pseudo-norme ℓ_p/ℓ_q .

Le plan de ce chapitre est le suivant : Dans la Section 6.2 nous présenterons la fonction de pénalité ℓ_p/ℓ_q et en proposerons une approximation lisse. Nous listerons ses propriétés mathématiques qui seront utiles pour la résolution de notre problème d'optimisation. Ensuite, dans la Section 6.3 nous proposerons un nouvel algorithme résultant de la résolution du problème de minimisation sous contrainte. La Section 6.4 présentera une analyse de convergence de notre algorithme. Finalement, nous présenterons dans la Section 6.5 des résultats de simulation qui montrent l'utilité et la fiabilité de la pénalité ℓ_p/ℓ_q pour restaurer des signaux parcimonieux en spectrométrie de masse.

6.2 Présentation de la pénalisation ℓ_p/ℓ_q

Pour restaurer des signaux parcimonieux, nous rappelons que notre problème d'optimisation (5.8) donné par :

$$\underset{\mathbf{x} \in \mathbb{R}_+^T}{\text{minimiser}} \Psi(\mathbf{x}) \quad \text{tel que} \quad \|\mathbf{D}\mathbf{x} - \mathbf{y}\| \leq \eta \quad (6.1)$$

revient généralement à utiliser une pénalité Ψ qui renforce le degré de parcimonie du signal \mathbf{x} mesuré avec un opérateur \mathbf{D} . Comme nous l'avons précisé dans les chapitres précédents, la pseudo-norme ℓ_0 représente un choix classique dans ce cas. Cependant, son utilisation reste limitée car cette pseudo-norme est non dérivable et non convexe, ce qui rend le problème d'optimisation de type NP^a [Natarajan 1995]. Une généralisation de cette pénalité définie pour tout $\mathbf{x} \in \mathbb{R}^N$ par :

$$(\forall p \in]0, +\infty[) \quad \ell_p(\mathbf{x}) = \left(\sum_{n=1}^N |x_n|^p \right)^{1/p} \quad (6.2)$$

a été utilisée dans différentes applications pour différentes valeurs de p comme le cas de la norme ℓ_1 , obtenue pour $p = 1$, que nous avons adoptée dans le Chapitre 5 [Cherni 2018]. Notons qu'une normalisation de cette fonction est équivalente à la fonction de moyennes généralisées (minimum si $p \rightarrow -\infty$, maximum si $p \rightarrow +\infty$, moyenne géométrique si $p = 0$, moyenne quadratique si $p = 2$, etc) [Bullen 2013].

a. NP : Nondeterministic Polynomial time, cela signifie que le problème ne peut pas être résolu en temps polynomial.

On trouve aussi le rapport de pénalités utilisé comme étant une approximation de ℓ_0 tel que ℓ_1/ℓ_∞ pour la reconstruction efficace des signaux parcimonieux [Demagnet 2014], ℓ_∞/ℓ_0 pour le partitionnement des données [Chang 2014] et aussi le rapport de pénalités ℓ_1/ℓ_2 [Repetti 2015] proposé très récemment pour assurer la déconvolution aveugle des signaux sismiques. Les résultats obtenus avec ces rapports montrent leur efficacité à éliminer le biais des signaux parcimonieux.

Nous proposons donc de résoudre le problème (6.1) avec une fonction de pénalité Ψ basée sur le rapport de pénalités ℓ_p/ℓ_q .

6.2.1 Version exacte de ℓ_p/ℓ_q

Pour tout $p \in]0, 2]$ et $q \in [2, +\infty[$, nous définissons :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \ell_p(\mathbf{x})/\ell_q(\mathbf{x}) = \left(\sum_{n=1}^N \left(\frac{|x_n|^q}{\sum_{n'=1}^N |x_{n'}|^q} \right)^{p/q} \right)^{1/p} \quad (6.3)$$

La pénalité ℓ_p/ℓ_q ne vérifie pas les propriétés de convexité et de différentiabilité ce qui rend son utilisation difficile. Nous proposerons dans la suite une approximation lisse de cette fonction généralisant l'approche proposée dans [Repetti 2015] et nous établirons des propriétés mathématiques de cette approximation qui seront utiles pour la résolution du problème d'optimisation.

6.2.2 Approximation lisse de ℓ_p/ℓ_q

Une approximation lisse des deux fonctions ℓ_p et ℓ_q pour tout $p \in]0, 2]$ et $q \in [2, +\infty[$, peut être définie pour $(\alpha, \eta) \in]0, +\infty[^2$ par :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \ell_{p,\alpha}(\mathbf{x}) = \left(\sum_{n=1}^N \left((x_n^2 + \alpha^2)^{p/2} - \alpha^p \right) \right)^{1/p} \quad (6.4)$$

et

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \ell_{q,\eta}(\mathbf{x}) = \left(\eta^q + \sum_{n=1}^N |x_n|^q \right)^{1/q} \quad (6.5)$$

On retrouve la forme (6.3) pour $\alpha = \eta = 0$. Afin de pallier la difficulté liée à la présence d'un ratio dans la fonction de pénalité, nous proposons de remplacer la fonction ℓ_p/ℓ_q par l'approximation suivante :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \Psi(\mathbf{x}) = \log \left(\frac{(\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p)^{1/p}}{\ell_{q,\eta}(\mathbf{x})} \right) \quad (6.6)$$

où $\beta \in]0, +\infty[$. Nous proposons d'utiliser la fonction logarithmique pour rendre la pénalité ℓ_p/ℓ_q plus facile à gérer d'une part, et pour renforcer la parcimonie du signal cible à travers la concavité de cette fonction d'autre part. De plus, nous avons ajouté le paramètre β pour que la fonction logarithmique soit définie en 0. Nous proposons d'appeler notre approximation de la pénalité ℓ_p/ℓ_q SPOQ (Smoothed p -Over- q norm). Lorsque $p = 1$ et $q = 2$, la fonction (6.6) est équivalente à la pénalisation SOOT (Smoothed One-Over-Two) proposée dans [Repetti 2015].

Nous notons pour tout $(\alpha, \eta) \in]0, +\infty[^2$:

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \Psi_1(\mathbf{x}) = \frac{1}{p} \log(\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p) \quad (6.7)$$

et

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \Psi_2(\mathbf{x}) = \log(\ell_{q,\eta}(\mathbf{x})) \quad (6.8)$$

Le problème (6.1) peut être reformulé par :

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} \quad \Psi(\mathbf{x}) + \Phi(\mathbf{x}) = \underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} \quad \Omega(\mathbf{x}) \quad (6.9)$$

tel que :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \Omega(\mathbf{x}) = \Psi(\mathbf{x}) + \Phi(\mathbf{x}) \quad (6.10)$$

avec :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \Psi(\mathbf{x}) = \Psi_1(\mathbf{x}) - \Psi_2(\mathbf{x}) \quad (6.11)$$

et :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \Phi(\mathbf{x}) = \iota_{[0,+\infty[^N}(\mathbf{x}) + \iota_{\|\mathbf{D} \cdot \mathbf{y}\| \leq \eta}(\mathbf{x}) \quad (6.12)$$

6.2.3 Propriétés de la fonction Ψ

Nous présentons ici une proposition importante, listant deux propriétés mathématiques utiles de la fonction Ψ , à savoir son caractère Lipschitz différentiable, et sa majoration par une fonction quadratique :

Proposition 6.1. La fonction Ψ est une fonction Lipschitz-différentiable sur \mathbb{R}^N et pour tout $\rho \in [0, +\infty[$, elle admet une majorante tangente quadratique en tout point $\mathbf{x} = (x_n)_{1 \leq n \leq N} \in \bar{\mathcal{B}}_{q,\rho} = \{(y_n)_{1 \leq n \leq N} \in \mathbb{R}^N \mid \sum_{n=1}^N |y_n|^q \geq \rho^q\}$:

$$(\forall \mathbf{x}' \in \bar{\mathcal{B}}_{q,\rho}) \quad \Psi(\mathbf{x}') \leq \Psi(\mathbf{x}) + (\mathbf{x}' - \mathbf{x})^\top \nabla \Psi(\mathbf{x}) + \frac{1}{2} (\mathbf{x}' - \mathbf{x})^\top A_{q,\rho}(\mathbf{x}) (\mathbf{x}' - \mathbf{x}) \quad (6.13)$$

où

$$\nabla \Psi(\mathbf{x}) = \left(\frac{x_n(x_n^2 + \alpha^2)^{\frac{q}{2}-1}}{\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p} - \frac{\text{sign}(x_n)|x_n|^{q-1}}{\ell_{q,\eta}^q(\mathbf{x})} \right)_{1 \leq n \leq N} \quad (6.14)$$

$$A_{q,\rho}(\mathbf{x}) = \chi_{q,\rho} \mathbf{I}_N + \frac{1}{\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p} \mathcal{D} \left(((x_n^2 + \alpha^2)^{p/2-1})_{1 \leq n \leq N} \right) \quad (6.15)$$

et

$$\chi_{q,\rho} = \frac{q-1}{(\eta^q + \rho^q)^{2/q}} \quad (6.16)$$

Démonstration. Pour tout $\mathbf{u} = (u_n)_{1 \leq n \leq N} \in \mathbb{R}^N$, posons $\mathcal{D}(\mathbf{u}) = \text{Diag} \left((u_n)_{1 \leq n \leq N} \right)$. Soient $p \in]0, 2]$ et $q \in [2, +\infty[$.

Pour $\ell_{q,\eta}^q$, on a :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \nabla \ell_{q,\eta}^q(\mathbf{x}) = q [\text{sign}(x_1)|x_1|^{q-1}, \dots, \text{sign}(x_N)|x_N|^{q-1}]^\top \quad (6.17)$$

$$\nabla^2 \ell_{q,\eta}^q(\mathbf{x}) = q(q-1) \mathcal{D} (|x_1|^{q-2}, \dots, |x_N|^{q-2}) \quad (6.18)$$

et pour $\ell_{p,\alpha}^p$, on a :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \nabla \ell_{p,\alpha}^p(\mathbf{x}) = pY \quad (6.19)$$

$$\nabla^2 \ell_{p,\alpha}^p(\mathbf{x}) = p \mathcal{D}(Z) \quad (6.20)$$

où $Y = (x_n(x_n^2 + \alpha^2)^{\frac{p}{2}-1})_{1 \leq n \leq N}$ et $Z = (((p-1)x_n^2 + \alpha^2)(x_n^2 + \alpha^2)^{\frac{p}{2}-2})_{1 \leq n \leq N}$.

Nous avons par la suite, pour tout $\mathbf{x} \in \mathbb{R}^N$:

$$\nabla \Psi_1(\mathbf{x}) = \frac{1}{p} \frac{\nabla \ell_{p,\alpha}^p(\mathbf{x})}{\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p} \quad (6.21)$$

$$\nabla \Psi_2(\mathbf{x}) = \frac{1}{q} \frac{\nabla \ell_{q,\eta}^q(\mathbf{x})}{\ell_{q,\eta}^q(\mathbf{x})} \quad (6.22)$$

$$\nabla^2 \Psi_1(\mathbf{x}) = \frac{1}{p} \left(\frac{\nabla^2 \ell_{p,\alpha}^p(\mathbf{x})}{\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p} - \frac{\nabla \ell_{p,\alpha}^p(\mathbf{x}) (\nabla \ell_{p,\alpha}^p(\mathbf{x}))^\top}{(\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p)^2} \right) \quad (6.23)$$

$$\nabla^2 \Psi_2(\mathbf{x}) = \frac{1}{q} \left(\frac{\nabla^2 \ell_{q,\eta}^q(\mathbf{x})}{\ell_{q,\eta}^q(\mathbf{x})} - \frac{\nabla \ell_{q,\eta}^q(\mathbf{x}) (\nabla \ell_{q,\eta}^q(\mathbf{x}))^\top}{\ell_{q,\eta}^{2q}(\mathbf{x})} \right) \quad (6.24)$$

Afin d'établir le caractère gradient Lipschitz de Ψ , nous étudierons les propriétés de $\nabla^2 \Psi$. Nous allons tout d'abord montrer que la norme du Hessien $\nabla^2 \Psi_1$ est bornée. Pour cela, nous étudions le développement asymptotique de $|||\nabla^2 \Psi_1(\mathbf{x})|||$ quand $\|\mathbf{x}\| \rightarrow +\infty$ où $|||\cdot|||$ désigne la norme spectrale. Les équations (6.19), (6.20), et (6.23) donnent :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad |||\nabla^2 \Psi_1(\mathbf{x})||| \leq \frac{|||\mathcal{D}(Z)|||}{\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p} + \frac{p\|Y\|^2}{(\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p)^2} \quad (6.25)$$

Posons :

$$\nu = \sum_{n=1}^N (x_n^2 + \alpha^2)^{p/2} \quad (6.26)$$

Pour $\|\mathbf{x}\| \rightarrow +\infty$, $\nu \rightarrow +\infty$. De plus, quand $\nu \rightarrow +\infty$:

$$\begin{aligned} \frac{|||\mathcal{D}(Z)|||}{\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p} &= \frac{1}{\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p} \sup_{1 \leq n \leq N} \left(((p-1)x_n^2 + \alpha^2)(x_n^2 + \alpha^2)^{\frac{p}{2}-2} \right) \\ &= \frac{1}{\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p} \sup_{1 \leq n \leq N} \left((p-1) \frac{x_n^2}{x_n^2 + \alpha^2} (x_n^2 + \alpha^2)^{\frac{p}{2}-1} + \alpha^2 (x_n^2 + \alpha^2)^{\frac{p}{2}-2} \right) \\ &\leq \frac{1}{\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p} \sup_{1 \leq n \leq N} \left((p-1)(x_n^2 + \alpha^2)^{\frac{p}{2}-1} + \alpha^2 (x_n^2 + \alpha^2)^{\frac{p}{2}-2} \right) \\ &\leq p \frac{\alpha^{p-2}}{\beta^p} \end{aligned} \quad (6.27)$$

$$\begin{aligned} \frac{\|Y\|^2}{(\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p)^2} &= \frac{1}{(\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p)^2} \sum_{n=1}^N x_n^2 (x_n^2 + \alpha^2)^{p-2} \\ &\leq \frac{1}{(\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p)^2} \sum_{n=1}^N \frac{x_n^2}{x_n^2 + \alpha^2} (x_n^2 + \alpha^2)^{p-1} \\ &\leq \frac{1}{\alpha^2 (\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p)^2} \sum_{n=1}^N (x_n^2 + \alpha^2)^p \\ &\leq \frac{1}{\alpha^2 (\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p)^2} \left(\sum_{n=1}^N (x_n^2 + \alpha^2)^{p/2} \right)^2 \\ &= \frac{\nu^2}{\alpha^2 (\nu - N\alpha^p + \beta^p)^2} \\ &\sim \frac{1}{\alpha^2} \end{aligned} \quad (6.28)$$

Cela montre que $\nabla^2\Psi_1$ est asymptotiquement bornée et puisque $\nabla^2\Psi_1$ est une fonction continue, nous déduisons que la fonction $\nabla^2\Psi_1$ est bornée sur \mathbb{R}^N .

Considérons maintenant la matrice Hessienne de Ψ_2 . Soit $\epsilon \in]0, +\infty[$, $\mathbf{x} \in \mathbb{R}^N$ et

$$\Lambda_\epsilon = \frac{\nabla^2 \ell_{q,\eta}^q(\mathbf{x}) + q(q-1)\epsilon \mathbf{I}_N}{\ell_{q,\eta}^q(\mathbf{x})} \quad (6.29)$$

ainsi,

$$\nabla_\epsilon^2 \Psi_2(\mathbf{x}) = \frac{1}{q} \left(\Lambda_\epsilon - \frac{\nabla \ell_{q,\eta}^q(\mathbf{x}) (\nabla \ell_{q,\eta}^q(\mathbf{x}))^\top}{\ell_{q,\eta}^{2q}(\mathbf{x})} \right) \quad (6.30)$$

Avec la propriété de continuité, on a :

$$\lim_{\epsilon \rightarrow 0} \|\nabla_\epsilon^2 \Psi_2(\mathbf{x})\| = \|\nabla^2 \Psi_2(\mathbf{x})\| \quad (6.31)$$

D'autre part, puisque Λ_ϵ est une matrice définie positive pour tout $p \in]0, 2]$ et $q \in [2, +\infty[$, on a :

$$\nabla_\epsilon^2 \Psi_2(\mathbf{x}) = \frac{1}{q} \Lambda_\epsilon^{1/2} (\mathbf{I}_N - v_\epsilon v_\epsilon^\top) \Lambda_\epsilon^{1/2} \quad (6.32)$$

où

$$v_\epsilon = \Lambda_\epsilon^{-1/2} \frac{\nabla \ell_{q,\eta}^q(\mathbf{x})}{\ell_{q,\eta}^q(\mathbf{x})} = \sqrt{\frac{q}{q-1}} \frac{1}{\ell_{q,\eta}^{q/2}(\mathbf{x})} \left[\frac{\text{sign}(x_1)|x_1|^{q-1}}{\sqrt{|x_1|^{q-2} + \epsilon}}, \dots, \frac{\text{sign}(x_N)|x_N|^{q-1}}{\sqrt{|x_N|^{q-2} + \epsilon}} \right]^\top \quad (6.33)$$

Donc :

$$\|\nabla_\epsilon^2 \Psi_2(\mathbf{x})\| = \left\| \frac{1}{q} \Lambda_\epsilon^{1/2} (\mathbf{I}_N - v_\epsilon v_\epsilon^\top) \Lambda_\epsilon^{1/2} \right\| \quad (6.34)$$

$$\leq \frac{1}{q} \|\Lambda_\epsilon\| \|\mathbf{I}_N - v_\epsilon v_\epsilon^\top\| \quad (6.35)$$

d'après (6.33) on a :

$$\Lambda_\epsilon = \left(\sqrt{\frac{q}{q-1}} \frac{1}{\ell_{q,\eta}^{q/2}(\mathbf{x})} \left[\frac{\text{sign}(x_1)|x_1|^{q-1}}{\sqrt{|x_1|^{q-2} + \epsilon}}, \dots, \frac{\text{sign}(x_N)|x_N|^{q-1}}{\sqrt{|x_N|^{q-2} + \epsilon}} \right]^\top \right)^{-2} \frac{(\nabla \ell_{q,\eta}^q(\mathbf{x}))^2}{(\ell_{q,\eta}^q(\mathbf{x}))^2} \quad (6.36)$$

$$= \frac{q-1}{q} \frac{\ell_{q,\eta}^q(\mathbf{x})}{\ell_{q,\eta}^{2q}(\mathbf{x})} \left[\frac{|x_1|^{q-2} + \epsilon}{|x_1|^{2q-2}}, \dots, \frac{|x_N|^{q-2} + \epsilon}{|x_N|^{2q-2}} \right]^\top \frac{(\nabla \ell_{q,\eta}^q(\mathbf{x}))^2}{(\ell_{q,\eta}^q(\mathbf{x}))^2} \quad (6.37)$$

$$= \frac{q-1}{q} \frac{(\nabla \ell_{q,\eta}^q(\mathbf{x}))^2}{\ell_{q,\eta}^{2q}(\mathbf{x})} \left[\frac{|x_1|^{q-2} + \epsilon}{|x_1|^{2q-2}}, \dots, \frac{|x_N|^{q-2} + \epsilon}{|x_N|^{2q-2}} \right]^\top \quad (6.38)$$

$$= \frac{q-1}{q} \frac{q^2 \left[|x_1|^{2q-2}, \dots, |x_N|^{2q-2} \right]^\top}{\ell_{q,\eta}^{2q}(\mathbf{x})} \left[\frac{|x_1|^{q-2} + \epsilon}{|x_1|^{2q-2}}, \dots, \frac{|x_N|^{q-2} + \epsilon}{|x_N|^{2q-2}} \right]^\top \text{ d'après (6.17)} \quad (6.39)$$

$$= q(q-1) \frac{\left[|x_1|^{2q-2}, \dots, |x_N|^{2q-2} \right]^\top}{\ell_{q,\eta}^{2q}(\mathbf{x})} \left[\frac{|x_1|^{q-2} + \epsilon}{|x_1|^{2q-2}}, \dots, \frac{|x_N|^{q-2} + \epsilon}{|x_N|^{2q-2}} \right]^\top \quad (6.40)$$

donc :

$$\|\Lambda_\epsilon\| = q(q-1) \frac{1}{\ell_{q,\eta}^q(\mathbf{x})} \sup_{1 \leq n \leq N} (|x_n|^{q-2} + \epsilon)_{1 \leq n \leq N} \quad (6.41)$$

Par conséquent, nous obtenons :

$$\|\|\nabla_\epsilon^2 \Psi_2(\mathbf{x})\|\| \leq \frac{1}{q} \|\|\Lambda_\epsilon\|\| \|\|I_N - v_\epsilon v_\epsilon^\top\|\| = \frac{q-1}{\ell_{q,\eta}^q(\mathbf{x})} \sup_{1 \leq n \leq N} (|x_n|^{q-2} + \epsilon) \max\{1, \|v_\epsilon\|^2 - 1\} \quad (6.42)$$

Pour $\epsilon \rightarrow 0$, (6.31) donne :

$$\|\|\nabla^2 \Psi_2(\mathbf{x})\|\| \leq \frac{q-1}{\ell_{q,\eta}^q(\mathbf{x})} \sup_{1 \leq n \leq N} |x_n|^{q-2} \max\{1, \|v\|^2 - 1\} \quad (6.43)$$

où

$$\|v\|^2 = \lim_{\epsilon \rightarrow 0} \|v_\epsilon\|^2 = \frac{q}{q-1} \frac{\sum_{n=1}^N |x_n|^q}{\ell_{q,\eta}^q(\mathbf{x})} \quad (6.44)$$

$$\Rightarrow \|v\|^2 - 1 = \frac{\sum_{n=1}^N |x_n|^q - (q-1)\eta^q}{(q-1)(\eta^q + \sum_{n=1}^N |x_n|^q)} = \frac{1}{q-1} \left(1 - \frac{q\eta^q}{\ell_{q,\eta}^q(\mathbf{x})}\right) < 1 \quad (6.45)$$

car $\left(1 - \frac{q\eta^q}{\ell_{q,\eta}^q(\mathbf{x})}\right) < 1$ et $\frac{1}{q-1} < 1$ pour tout $q \in [2, +\infty[$.

Nous déduisons par la suite que :

$$\begin{aligned} \|\|\nabla^2 \Psi_2(\mathbf{x})\|\| &\leq \frac{q-1}{\ell_{q,\eta}^q(\mathbf{x})} \sup_{1 \leq n \leq N} |x_n|^{q-2} \\ &\leq \frac{q-1}{(\eta^q + \sum_{n=1}^N |x_n|^q)^{2/q}} \left(\frac{\sup_{1 \leq n \leq N} |x_n|^q}{\eta^q + \sum_{n=1}^N |x_n|^q} \right)^{\frac{q-2}{q}} \\ &\leq \frac{q-1}{(\eta^q + \sum_{n=1}^N |x_n|^q)^{2/q}} \leq \frac{q-1}{\eta^2}. \end{aligned} \quad (6.46)$$

Puisque $\nabla^2 \Psi_1$ et $\nabla^2 \Psi_2$ sont bornées, $\nabla^2 \Psi = \nabla^2 \Psi_1 - \nabla^2 \Psi_2$ est par la suite bornée, et Ψ est une fonction Lipschitz-différentiable.

Conclusion : Nous avons montré que Ψ est une fonction Lipschitz-différentiable. L'expression de son gradient est donnée par les équations (6.17), (6.19), (6.21), et (6.22).

Passons maintenant à la preuve de la propriété de majoration. Nous utilisons le fait que $\xi \mapsto (\xi + \alpha^2)^{p/2}$ est une fonction concave^b sur \mathbb{R}^+ . En appliquant la propriété de majoration d'une fonction concave par sa tangente, on obtient par composition avec la fonction $x \mapsto x^2$ l'inégalité suivante pour tout $\mathbf{x}' = (x'_n)_{1 \leq n \leq N} \in \mathbb{R}^N$, $\mathbf{x} = (x_n)_{1 \leq n \leq N} \in \mathbb{R}^N$ et $n \in \{1, \dots, N\}$:

$$(x_n'^2 + \alpha^2)^{p/2} \leq (x_n^2 + \alpha^2)^{p/2} + p x_n (x_n^2 + \alpha^2)^{p/2-1} (x_n' - x_n) + \frac{p}{2} (x_n^2 + \alpha^2)^{p/2-1} (x_n' - x_n)^2 \quad (6.47)$$

Par conséquent :

$$\ell_{p,\alpha}^p(\mathbf{x}') \leq \ell_{p,\alpha}^p(\mathbf{x}) + (\mathbf{x}' - \mathbf{x})^\top \nabla \ell_{p,\alpha}^p(\mathbf{x}) + \frac{p}{2} (\mathbf{x}' - \mathbf{x})^\top A_1(\mathbf{x}) (\mathbf{x}' - \mathbf{x}) \quad (6.48)$$

où $A_1(\mathbf{x}) = \mathcal{D} \left(((x_n^2 + \alpha^2)^{p/2-1})_{1 \leq n \leq N} \right)$. Avec l'inégalité de Napier :

$$(\forall (u, v) \in]0, +\infty[^2) \quad \log u \leq \log v + \frac{u-v}{v} \quad (6.49)$$

b. Une fonction f est concave si $-f$ est une fonction convexe

nous obtenons :

$$\Psi_1(\mathbf{x}') \leq \Psi_1(\mathbf{x}) + (\mathbf{x}' - \mathbf{x})^\top \nabla \Psi_1(\mathbf{x}) + \frac{1}{2(\ell_{p,\alpha}^p(\mathbf{x}) + \beta^p)} (\mathbf{x}' - \mathbf{x})^\top A_1(\mathbf{x})(\mathbf{x}' - \mathbf{x}) \quad (6.50)$$

Appliquons enfin le lemme de descente [Peypouquet 2015, Bolte 2017] sur la fonction $-\Psi_2$. En utilisant (6.46), nous avons :

$$(\forall (\mathbf{x}, \mathbf{x}') \in \overline{B}_{q,\rho}^2) \quad -\Psi_2(\mathbf{x}') \leq -\Psi_2(\mathbf{x}) - (\mathbf{x}' - \mathbf{x})^\top \nabla \Psi_2(\mathbf{x}) + \frac{\chi_{q,\rho}}{2} \|\mathbf{x}' - \mathbf{x}\|^2. \quad (6.51)$$

La propriété de majoration est déduite, en utilisant (6.50) et (6.51). \square

Nous notons que notre proposition 6.1 étend les résultats établis dans [Repetti 2015] pour $p = 1$ et $q = 2$ (ℓ_1/ℓ_2).

Remarque : Nous remarquons que $\nabla \ell_{p,\alpha}^p(0) = \nabla \ell_{q,\eta}^q(0) = 0$, le point 0 est donc un point critique de la fonction Ψ . En outre :

$$\nabla^2 \Psi_1(0) = \frac{\alpha^{p-2}}{\beta^p} \mathbf{I}_N \quad (6.52)$$

$$(6.53)$$

De plus si $q > 2$ on a :

$$\nabla^2 \Psi_2(0) = \frac{1}{q} \left(\frac{\nabla^2 \ell_{q,\eta}^q(0)}{\ell_{q,\eta}^q(0)} - \frac{\nabla \ell_{q,\eta}^q(0) (\nabla \ell_{q,\eta}^q(0))^\top}{\ell_{q,\eta}^{2q}(0)} \right) \quad (6.54)$$

$$= \frac{1}{q} \left(\frac{q(q-1)\mathcal{D}(0^{q-2})}{\eta^q} - q \frac{0}{\eta^{2q}} \right) \quad (6.55)$$

$$= 0 \text{ car } 0^{q-2} = 0 \text{ pour } q - 2 > 0 \quad (6.56)$$

et si $q = 2$:

$$\nabla^2 \Psi_2(0) = \frac{1}{2} \left(\frac{\nabla^2 \ell_{2,\eta}^2(0)}{\ell_{2,\eta}^2(0)} - \frac{\nabla \ell_{2,\eta}^2(0) (\nabla \ell_{2,\eta}^2(0))^\top}{\ell_{2,\eta}^4(0)} \right) \quad (6.57)$$

$$= \frac{1}{2} \left(\frac{2(2-1)\mathcal{D}(0^0)}{\eta^2} - 2 \frac{0}{\eta^{2q}} \right) \quad (6.58)$$

$$= \frac{1}{\eta^2} \mathbf{I}_N \text{ car } 0^0 = 1 \text{ par convention} \quad (6.59)$$

Par conséquent :

$$\nabla^2 \Psi_2(0) = \begin{cases} \frac{1}{\eta^2} \mathbf{I}_N & \text{si } q = 2 \\ 0 & \text{sinon} \end{cases} \quad (6.60)$$

Nous déduisons que $\nabla^2 \Psi(0)$ est une matrice définie positive si ($q = 2$ et $\eta^2 \alpha^{p-2} > \beta^p$) ou si $q > 2$. Dans le cas où ces conditions sont vérifiées, le point 0 est un minimum local de Ψ , ce qui est attendu par une pénalité choisie pour renforcer la parcimonie.

6.3 Algorithme de minimisation

6.3.1 Forward-Backward à métrique variable localement ajustée

Comme nous l'avons vu dans la Proposition 6.1, la fonction Ψ peut être majorée par une fonction quadratique, dans un domaine limité, paramétré par une valeur ρ , dépendant du point où est calculée la majorante. Nous proposons de tirer partie de cette propriété intéressante de majoration locale, pour construire une extension de la stratégie Forward-Backward à métrique variable [Chouzenoux 2014] que nous avons introduite dans le Chapitre 2.

L'algorithme sera donné par l'Algorithme 9. A chaque itération, on va appliquer une étape de l'algorithme de Forward-Backward à métrique variable. L'opérateur proximal est calculé de façon non nécessairement exacte, à condition de vérifier Alg :9(A) et Alg :9(B). La métrique utilisée est déduite de la proposition 6.1. Elle dépend donc du paramètre ρ dont la valeur est déterminée par une stratégie de "backtracking" afin de garantir la validité de la propriété de majoration au point obtenu après l'étape proximale approchée.

Algorithme 9 Forward-Backward à métrique variable localement ajustée

Soient $\beta \in]0, +\infty[$ et $\mathbf{x}_0 \in \text{dom}(\Phi)$
 Pour $k = 0, 1, \dots$:

Pour $i = 1, \dots, B$:	
Si $i = 1$:	$\rho_{k,1} = \sum_{n=1}^N x_{n,k} ^q$
Sinon si $i = B$:	$\rho_{k,B} = 0$
Sinon :	$\rho_{k,i} = \sum_{n=1}^N y_{n,k,i-1} ^q$
Fin si	
$\mathcal{B}_{k,i} = \overline{\mathcal{B}}_{q,\rho_{k,i}}$	
$\mathbf{A}_{k,i} = \mathbf{A}_{q,\rho_{k,i}}(\mathbf{x}_k)$	
Soit $\gamma_k \in]0, 1[$	
Trouver $\mathbf{y}_{k,i} \in \mathbb{R}^N$ et $\mathbf{r}_{k,i} \in \partial\Phi(\mathbf{y}_{k,i})$ tels que :	
$\Phi(\mathbf{y}_{k,i}) + (\mathbf{y}_{k,i} - \mathbf{x}_k)^\top \nabla\Psi(\mathbf{x}_k) + \gamma_k \ \mathbf{y}_{k,i} - \mathbf{x}_k\ _{\mathbf{A}_{k,i}}^2 \leq \Phi(\mathbf{x}_k)$	Alg :9(A)
$\ \nabla\Psi(\mathbf{x}_k) + \mathbf{r}_{k,i}\ \leq \beta \ \mathbf{y}_{k,i} - \mathbf{x}_k\ _{\mathbf{A}_{k,i}}$	Alg :9(B)
Si $\mathbf{y}_{k,i} \in \mathcal{B}_{k,i}$:	
$\mathbf{y}_k = \mathbf{y}_{k,i}$	
Fin si.	
$\mathbf{x}_{k+1} = \mathbf{y}_k$	

La séquence $(\mathbf{x}_k)_{k \in \mathbb{N}}$ converge vers un point critique du problème (6.9) sous certaines hypothèses que nous précisons dans la section suivante.

6.3.2 Validité de l'algorithme

Nous montrons ici que l'Algorithme 9 est bien défini, au sens où l'on est garanti de sortir de la boucle interne en un nombre fini d'itérations. Nous listons tout d'abord des propriétés utiles. Les preuves sont omises pour raison de place, et peuvent être déduites facilement de la Propriété 6.1.

Lemme 1.

(i) $\Phi : \mathbb{R}^N \rightarrow]+\infty, +\infty]$ est une fonction propre semi-continue inférieurement, bornée et continue sur $\text{dom}(\Phi)$.

(ii) $\Psi : \mathbb{R}^N \rightarrow \mathbb{R}$ est Lipschitz différentiable sur $\text{dom}(\Phi)$.

(iii) $\Omega = \Psi + \Phi$ est une fonction coercive.

Lemme 2.

(i) Par construction, on a pour tout $k \in \mathbb{N}$ et pour tout $i \in \{1, \dots, B\} : \mathbf{A}_{k,i} \in \mathcal{S}_N^c$, $B \geq 1$, $\mathcal{B}_k^1 \subset \mathcal{B}_k^2 \dots \subset \mathcal{B}_k^B$ et $\mathbf{x}_k \in \mathcal{B}_{k,1}$, $\mathcal{B}_{k,B} = \mathbb{R}^N$ et on a :

$$(\forall \mathbf{x} \in \mathcal{B}_{k,i}) \quad \Psi(\mathbf{x}) \leq \Psi(\mathbf{x}_k) + (\mathbf{x} - \mathbf{x}_k)^\top \nabla \Psi(\mathbf{x}_k) + \frac{1}{2} \|\mathbf{x} - \mathbf{x}_k\|_{\mathbf{A}_{k,i}}^2 \quad (6.61)$$

(ii) Il existe $(\underline{\nu}, \bar{\nu}) \in]0, +\infty[^2$, tel que pour tout $k \in \mathbb{N}$ et tout $i \in \{1, \dots, B\}$, on a :

$$\underline{\nu} \mathbf{I}_N \leq \mathbf{A}_{k,i} \leq \bar{\nu} \mathbf{I}_N$$

Hypothèse 6.1.

Il existe $(\underline{\gamma}, \bar{\gamma}) \in]0, +\infty[^2$ tel que pour tout $k \in \mathbb{N}$, $\underline{\gamma} \leq \gamma_k \leq 1 - \bar{\gamma}$.

Le lemme ci-dessous permet de confirmer que l'algorithme proposé est bien défini.

Lemme 3.

Selon le Lemme 1, le Lemme 2 et l'Hypothèse 6.1, pour tout $i \in \{1, \dots, B\}$, il existe $\mathbf{y}_{k,i} \in \mathbb{R}^N$ et $\mathbf{r}_{k,i} \in \partial \Psi(\mathbf{y}_{k,i})$ tel que les conditions Alg :9(A) et Alg :9(B) sont vérifiées.

Démonstration. Soit $k \in \mathbb{N}$ et $i \in \{1, \dots, B\}$. Posons :

$$\mathbf{y}_{k,i} = \text{prox}_{\gamma_k^{-1} \mathbf{A}_{k,i}, \Psi}(\mathbf{x}_k - \gamma_k (\mathbf{A}_{k,i})^{-1} \nabla \Psi(\mathbf{x}_k)). \quad (6.62)$$

En utilisant la définition de l'opérateur proximal, on a :

$$\Psi(\mathbf{y}_{k,i}) + (\mathbf{y}_{k,i} - \mathbf{x}_k)^\top \nabla \Psi(\mathbf{x}_k) + \gamma_k^{-1} \|\mathbf{y}_{k,i} - \mathbf{x}_k\|_{\mathbf{A}_{k,i}}^2 \leq \Psi(\mathbf{x}_k) \quad (6.63)$$

Par conséquent, on déduit la condition Alg :9(A). De plus, il existe $\mathbf{r}_{k,i} \in \partial \Psi(\mathbf{y}_{k,i})$ tel que :

$$\mathbf{r}_{k,i} = -\nabla \Psi(\mathbf{x}_k) + \gamma_k^{-1} \mathbf{A}_{k,i} (\mathbf{x}_k - \mathbf{y}_{k,i}). \quad (6.64)$$

Sous le Lemme 2(ii) et l'Hypothèse 6.1, on obtient :

$$\|\mathbf{r}_{k,i} + \nabla \Psi(\mathbf{x}_k)\| = \gamma_k^{-1} \|\mathbf{A}_{k,i} (\mathbf{x}_k - \mathbf{y}_{k,i})\| \leq \underline{\gamma}^{-1} \sqrt{\bar{\nu}} \|\mathbf{x}_k - \mathbf{y}_{k,i}\|_{\mathbf{A}_{k,i}} \quad (6.65)$$

d'où la condition Alg :9(B) avec $\beta = \underline{\gamma}^{-1} \sqrt{\bar{\nu}}$.

□

Remarque 6.1.

Sous l'Hypothèse 6.1, il existe $i \in \{1, \dots, B\}$ tel que $\mathbf{y}_{k,i} \in \mathcal{B}_{k,i}$.

c. On désigne par \mathcal{S}_N l'ensemble des matrices symétriques définies positives dans $\mathbb{R}^{N \times N}$

En effet, il suffit de prendre $i = B$ (car $\mathcal{B}_{k,B} = \mathbb{R}^N$) et d'appliquer le lemme ci dessus. Nous en déduisons que l'Algorithme 9 est bien défini.

6.3.3 Calcul de l'opérateur proximal inexact

Pour calculer l'opérateur proximal de la fonction Φ donnée en (6.12) et obtenir les $(\mathbf{y}_{k,i})_{k \in \mathbb{N}, 1 \leq i \leq B}$, nous devons résoudre un problème de la forme :

$$\tilde{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{x} - \tilde{\mathbf{x}}\|_{\mathbf{B}}^2 + \Phi(\mathbf{x}) \quad (6.66)$$

$$= \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{x} - \tilde{\mathbf{x}}\|_{\mathbf{B}}^2 + \iota_{[0, +\infty[^N}(\mathbf{x}) + \iota_{\|\mathbf{D} \cdot - \mathbf{y}\| \leq \eta}(\mathbf{x}) \quad (6.67)$$

avec la métrique $\mathbf{B} = \gamma^{-1} \mathbf{A}$ et \mathbf{A} étant une matrice diagonale à entrées strictement positives, de la forme (6.15). Nous proposons d'utiliser l'algorithme Dual Forward-Backward à métrique variable [Abboud 2017] pour la résolution de (6.67).

Algorithme 10 Calcul de l'opérateur proximal avec le Dual FB

$$\left[\begin{array}{l} \mathbf{y}_0 \in \mathbb{R}^M \\ \mathbf{z}_0 = -\mathbf{B}^{-1} \mathbf{D}^T \mathbf{y}_0 \\ \sigma = \|\mathbf{D} \mathbf{B}^{-1} \mathbf{D}^T\| \\ \theta \in]0, 2[\\ \text{Pour } j = 0, 1, \dots \\ \left[\begin{array}{l} \mathbf{x}_j = \operatorname{prox}_{\mathbf{B}, \iota_{[0, +\infty[^N}}(\tilde{\mathbf{x}} + \mathbf{z}_j) \\ \tilde{\mathbf{y}}_j = \mathbf{y}_j + \theta_j \sigma^{-1} \mathbf{D} \mathbf{x}_j \\ \mathbf{y}_{j+1} = \tilde{\mathbf{y}}_j - \theta_j \sigma^{-1} \mathcal{P}_{\|\cdot - \mathbf{y}\| \leq \eta}(\theta_j \sigma \tilde{\mathbf{y}}_j) \\ \mathbf{z}_{j+1} = \mathbf{z}_j - \mathbf{B}^{-1} \mathbf{D}^T (\mathbf{y}_{j+1} - \mathbf{y}_j) \end{array} \right. \end{array} \right.$$

La métrique \mathbf{B} est une matrice diagonale à entrées strictement positives, par conséquent :

$$\operatorname{prox}_{\mathbf{B}, \iota_{[0, +\infty[^N}}(\mathbf{x}) = \mathcal{P}_{[0, +\infty[^N}(\mathbf{x}) \quad (6.68)$$

où l'opérateur $\mathcal{P}_{[0, +\infty[^N}$ désigne la projection sur l'ensemble des réels positifs. La série $(\mathbf{z}_j)_{j \in \mathbb{N}}$ converge vers la solution de (6.67). Notons que nous avons aussi considéré d'autres méthodes pour calculer l'opérateur proximal tel que l'algorithme PPXA+ (Annexe E.1) et la version accélérée de Chambolle-Pock (Annexe E.2). Néanmoins, celles-ci se sont avérées moins performantes en terme de complexité de calcul sur nos exemples pratiques dans le cadre de la SM.

6.4 Analyse de convergence

Cette section est consacrée à la présentation des propriétés de convergence de l'algorithme Forward-Backward à métrique variable localement ajustée (Algorithme 9).

Lemme 4.

Sous le Lemme 1, le Lemme 2, et l'Hypothèse 6.1, il existe $\mu \in]0, +\infty[$ tel que pour tout $k \in \mathbb{N}$:

$$\Omega(\mathbf{x}_{k+1}) \leq \Omega(\mathbf{x}_k) - \frac{\mu}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \quad (6.69)$$

Démonstration. On a :

$$(\forall k \in \mathbb{N}) \quad \Omega(\mathbf{x}_{k+1}) = \Psi(\mathbf{x}_{k+1}) + \Phi(\mathbf{x}_{k+1}) \quad (6.70)$$

D'après la condition Alg :9(A) et l'Hypothèse 6.1 :

$$\Phi(\mathbf{x}_{k+1}) + (\mathbf{x}_{k+1} - \mathbf{x}_k)^\top \nabla \Psi(\mathbf{x}_k) + \alpha \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_{\mathbf{A}^{k, i_k}}^2 \leq \Phi(\mathbf{x}_k) \quad (6.71)$$

où $\alpha = (1 - \bar{\gamma})$. Par construction, on a $\mathbf{x}_{k+1} \in \mathcal{B}_{k, i_k}$, d'après le Lemme 2(i) :

$$\Psi(\mathbf{x}_{k+1}) \leq \Psi(\mathbf{x}_k) + (\mathbf{x}_{k+1} - \mathbf{x}_k)^\top \nabla \Psi(\mathbf{x}_k) + \frac{1}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_{\mathbf{A}^{k, i_k}}^2 \quad (6.72)$$

Donc,

$$\begin{aligned} \Omega(\mathbf{x}_{k+1}) &\leq \Psi(\mathbf{x}_k) + \Phi(\mathbf{x}_k) + \frac{1}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_{\mathbf{A}^{k, i_k}}^2 - \alpha \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_{\mathbf{A}^{k, i_k}}^2 \\ &\leq \Omega(\mathbf{x}_k) - \left(\alpha - \frac{1}{2}\right) \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_{\mathbf{A}^{k, i_k}}^2 \end{aligned}$$

Par conséquent, (6.69) est déduite en appliquant le Lemme 2(ii) et en posant $\mu = \underline{\nu}(2\alpha - 1)$ pour $\alpha \in]1/2, +\infty[$. \square

Théorème 6.1.

Sous l'Hypothèse 6.1, la suite $(\mathbf{x}_k)_{k \in \mathbb{N}}$ générée par l'Algorithme 9 converge vers un point critique $\hat{\mathbf{x}}$ de la fonction Ω .

Démonstration. La preuve de convergence de $(\mathbf{x}_k)_{k \in \mathbb{N}}$ se déduit de [Repetti 2015]. \square

6.5 Résultats expérimentaux

6.5.1 Données synthétiques

Nous proposons d'appliquer notre Algorithme 9 pour reconstruire des spectres synthétiques de SM (Figure 6.1). Le modèle d'observation est similaire à celui du chapitre précédent. Nous considérons :

- un spectre de masse A de dimension $N = 2000$ mono-charge $z = 1$ avec $P = 10$
- un spectre de masse B de dimension $N = 1000$, avec $P = 100$ pics de différents états de charge $z \in (1, 2, 3)$

Nous considérons un axe de masse $[m_{\min}, m_{\max}]$ avec $m_{\min} = 1000$ Dalton et $m_{\max} = 1050$ Dalton. Nous ajoutons un bruit gaussien d'écart type $\sigma = 10^{-2}$.

6.5.2 Test de reconstruction

Nous reconstruisons les signaux en utilisant l'Algorithme 8 basé sur la pénalité ℓ_1 avec 1000 itérations et l'Algorithme 9 basé sur notre approche SPOQ. Pour que la reconstruction des signaux soit moins sensible aux minima locaux, nous initialisons l'Algorithme 9 par 10 itérations de l'Algorithme 8. De plus, l'Algorithme 10 est utilisé pour calculer l'opérateur proximal approché, avec

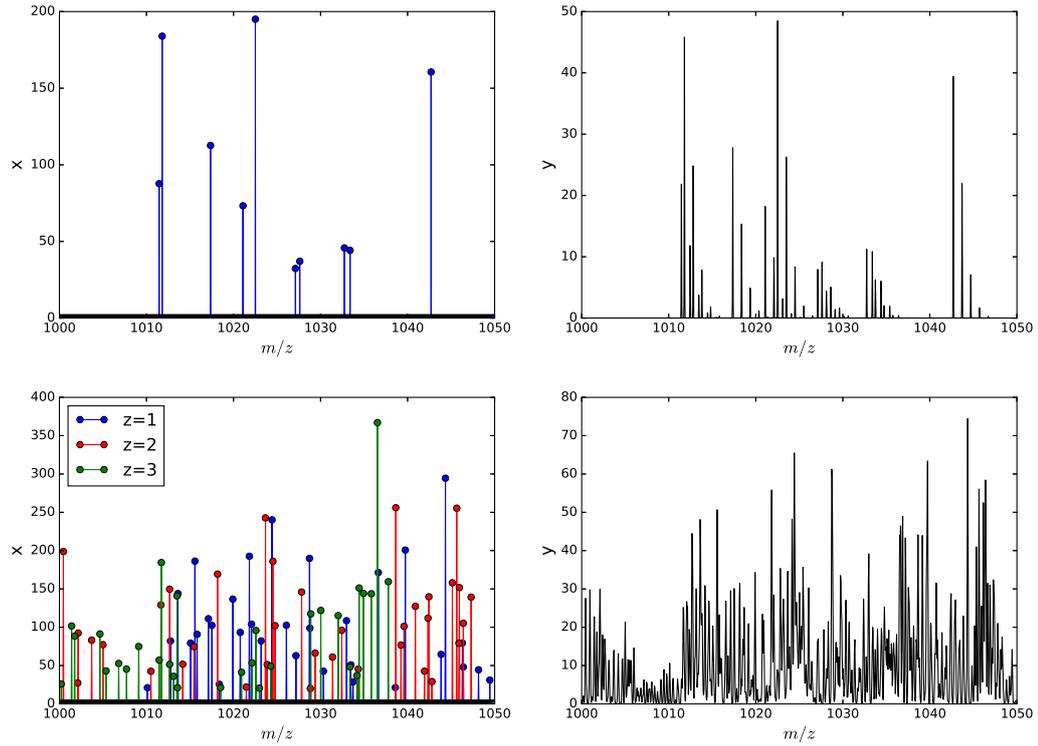


FIGURE 6.1 – Signaux synthétiques : signal A mono-chargé de dimension $N = 1000$ et $P = 10$ (en haut). Signal B de $N = 1000$ ayant les états de charge $z \in (1, 2, 3)$ (en bas)

un maximum de 1000 itérations. Les paramètres α , β , et η sont optimisés de manière manuelle pour maximiser le RSB.

La pénalité ℓ_p/ℓ_q est testée pour différentes valeurs de $p \in]0, 2]$ et $q \in [2, +\infty[$. La tableau 6.1 présente les qualités de reconstruction obtenues où nous proposons de calculer le RSB classique donné par (2.10) et qu'on note RSB_{ℓ_2} ainsi que le RSB calculé avec la norme ℓ_1 donné par :

$$\text{RSB}_{\ell_1}(\bar{\mathbf{x}}, \hat{\mathbf{x}}) = -20 \log_{10} \left(\frac{\sum_{n=1}^N |\bar{x}_n - \hat{x}_n|}{\sum_{n=1}^N |\bar{x}_n|} \right) \quad (6.73)$$

où $\hat{\mathbf{x}} = (\hat{x}_n)_{1 \leq n \leq N}$ désigne l'estimation du signal cible $\bar{\mathbf{x}} = (\bar{x}_n)_{1 \leq n \leq N}$.

Selon les résultats du tableau 6.1, dans le cas du signal A mono-chargé, la pénalité ℓ_1 donne un RSB équivalent à celui donné par la pénalité $\ell_{1/4}/\ell_2$ et $\ell_{1/4}/\ell_3$. On trouve que le meilleur RSB est obtenu avec la pénalité ℓ_1/ℓ_2 en calculant le RSB_{ℓ_2} et $\ell_{1/4}/\ell_2$ avec le RSB_{ℓ_1} .

Pour le signal B multi-chargé, toutes les pénalités pour les différentes valeurs de p et q que nous avons proposées, présentent des RSB plus élevés que le RSB obtenu avec la pénalité ℓ_1 . Le meilleur RSB est donné par la pénalité $\ell_{1/4}/\ell_2$ soit en utilisant le RSB classique (2.10) soit celui donné en (6.73).

		$q \backslash p$	2	3	4	5	10	ℓ_1
		Signal A						
Signal A	RSB $_{\ell_2}$	0.25	34.779377	34.778418	37.621941	37.629219	36.828765	34.77
		0.50	34.782178	38.011798	37.300695	37.015847	36.677098	
		0.75	37.609228	36.936144	36.738387	36.637137	41.707722	
		1.00	43.050516	42.590367	41.762935	41.195306	40.514449	
		1.25	40.683816	40.403777	40.391314	40.387593	40.379915	
		1.50	40.404176	40.365398	40.360246	40.358867	40.355675	
		2.00	40.422442	40.333073	40.348125	40.349422	40.351232	
	RSB $_{\ell_1}$	0.25	33.403519	33.407192	28.997798	28.249926	27.319098	33.40
		0.50	33.394760	28.463171	27.986844	27.607383	27.092849	
		0.75	28.312057	27.463137	27.175018	26.954358	19.023019	
		1.00	19.199416	19.104378	18.930786	18.784865	18.561345	
		1.25	18.534763	18.502106	18.509457	18.514503	18.523087	
		1.50	18.494037	18.517687	18.521399	18.523308	18.526200	
		2.00	18.531357	18.520168	18.523187	18.524528	18.526659	
Signal B	RSB $_{\ell_2}$	0.25	47.299629	46.492635	41.654001	37.859961	36.018852	18.54
		0.50	44.379107	42.523241	40.113274	37.817467	34.090610	
		0.75	43.869759	42.778173	39.858742	38.227077	36.742793	
		1.00	44.119362	42.551960	37.488475	38.794064	36.801415	
		1.25	44.052773	38.528645	40.301975	38.827252	33.681357	
		1.50	43.983570	42.765253	40.450645	36.395626	33.484941	
		2.00	40.567852	42.094655	39.988724	36.228957	33.097467	
	RSB $_{\ell_1}$	0.25	48.453824	48.262395	46.203242	43.606076	41.503628	26.96
		0.50	46.721416	45.773877	44.500191	43.069180	39.548848	
		0.75	46.178205	45.605351	44.148571	43.094743	41.808120	
		1.00	46.285428	45.454566	42.475882	43.315208	41.886469	
		1.25	46.250520	43.225235	44.204488	43.343778	39.232458	
		1.50	46.217752	45.566790	44.287818	41.542458	39.062958	
		2.00	44.212347	45.228568	44.050747	41.348742	38.768255	

Tableau 6.1 – RSB de la reconstruction des signaux synthétiques A et B pour différentes valeurs de $p \in]0, 2[$ et $q \in [2, +\infty[$

Nous présentons dans la Figure 6.2 le résultat de reconstruction des signaux synthétiques A et B pour trois choix de pénalité : ℓ_1/ℓ_2 , $\ell_{1/4}/\ell_2$ et $\ell_{3/4}/\ell_2$.

Les résultats de la Figure 6.2 montrent que les différentes pénalités assurent une restauration satisfaisante de la position des pics, mais avec une sous-estimation d'intensités dans certains cas.

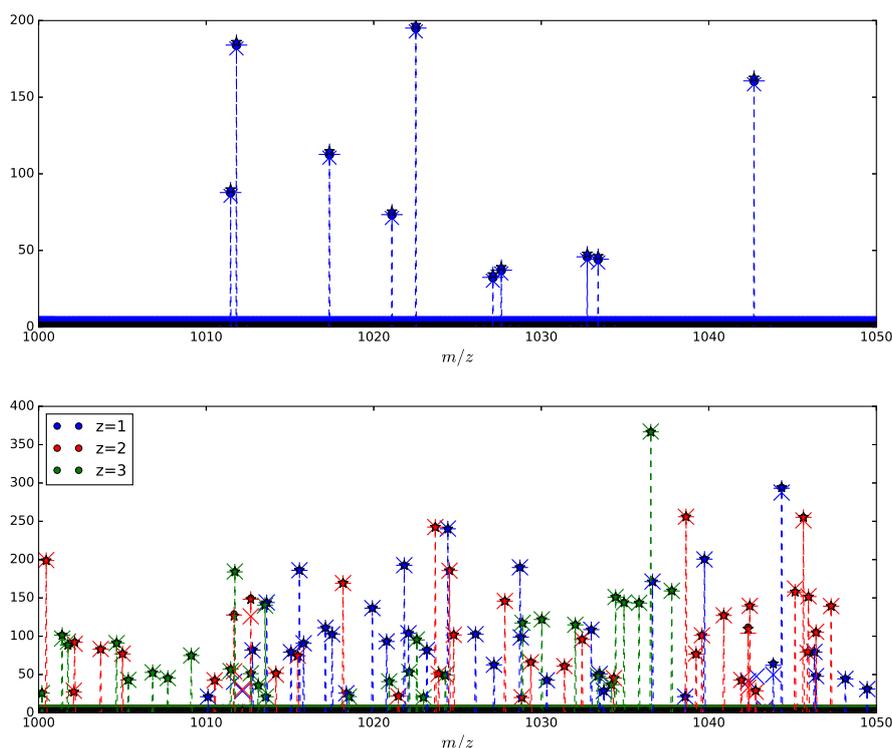


FIGURE 6.2 – Résultat de reconstruction des signaux A (en haut) et B (en bas) (Figure 6.1) avec la pénalité : ℓ_1 en " \times ", ℓ_1/ℓ_2 en " +", $\ell_{1/4}/\ell_2$ en " \star " et $\ell_{3/4}/\ell_3$ en " -". Le signal original est en " \bullet "

Un zoom sur deux zones de l'axe de masse : $[1025, 1030]$ et $[1030, 1035]$ pour le signal A et $[1010, 1014]$ et $[1041, 1043]$ pour le signal B, est présenté dans la Figure 6.3. Nous constatons que la fonction de pénalité ℓ_1 présente une sous-estimation remarquable pour quelques pics.

Nous montrons aussi plus précisément dans la Figure 6.4 l'erreur d'estimation pour chaque pic des deux signaux A et B, calculée comme étant la différence entre l'intensité du pic original et l'intensité du pic reconstruit. Ces résultats prouvent que la pénalité classique ℓ_1 assure une reconstruction satisfaisante des signaux parcimonieux. Cependant, dans notre application SM où l'analyse et la séparation des pics mono-isotopiques est compliquée dans le cas de multi-charges, l'approche SPOQ paraît plus efficace pour certaines valeurs de p et q , en particulier pour $p = 1/4$ et $q = 2$ ou $q = 3$. Notons que cette amélioration de qualité de reconstruction est accompagnée d'une augmentation du temps de calcul. Par exemple, la pénalité ℓ_1/ℓ_2 nécessite 5 minutes pour restaurer le signal A. Cependant avec la pénalité ℓ_1 , la restauration du signal B est obtenue après 4.6 secondes.

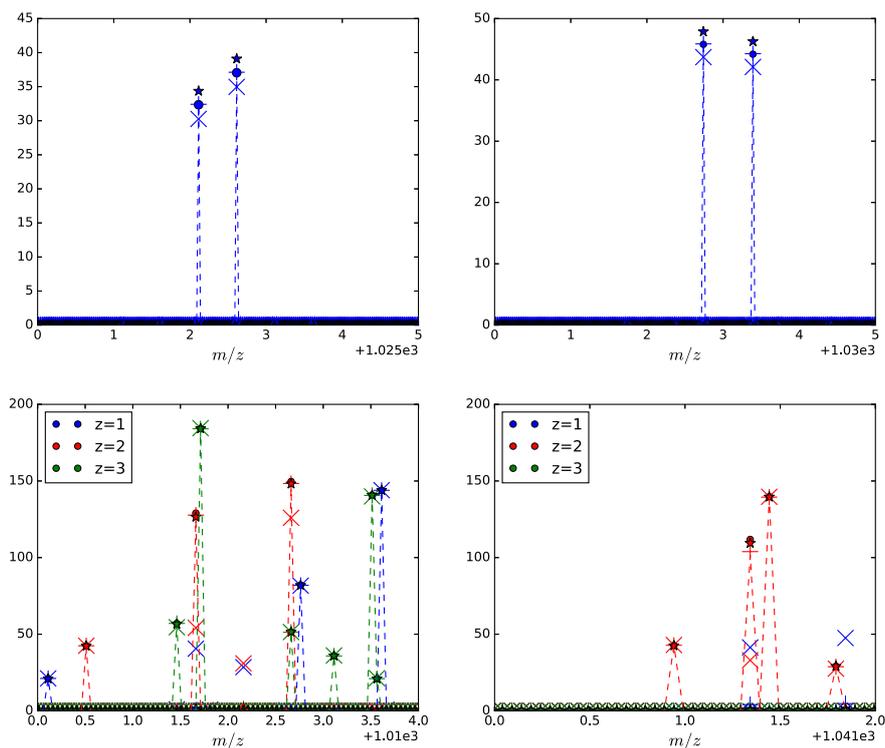


FIGURE 6.3 – Zoom sur le résultat de reconstruction des signaux A (en haut) et B (en bas) (Figure 6.2) avec la pénalité : ℓ_1 en " × ", ℓ_1/ℓ_2 en " + ", $\ell_{1/4}/\ell_2$ en " ★ " et $\ell_{3/4}/\ell_3$ en " - - ". Le signal original est en " ● "

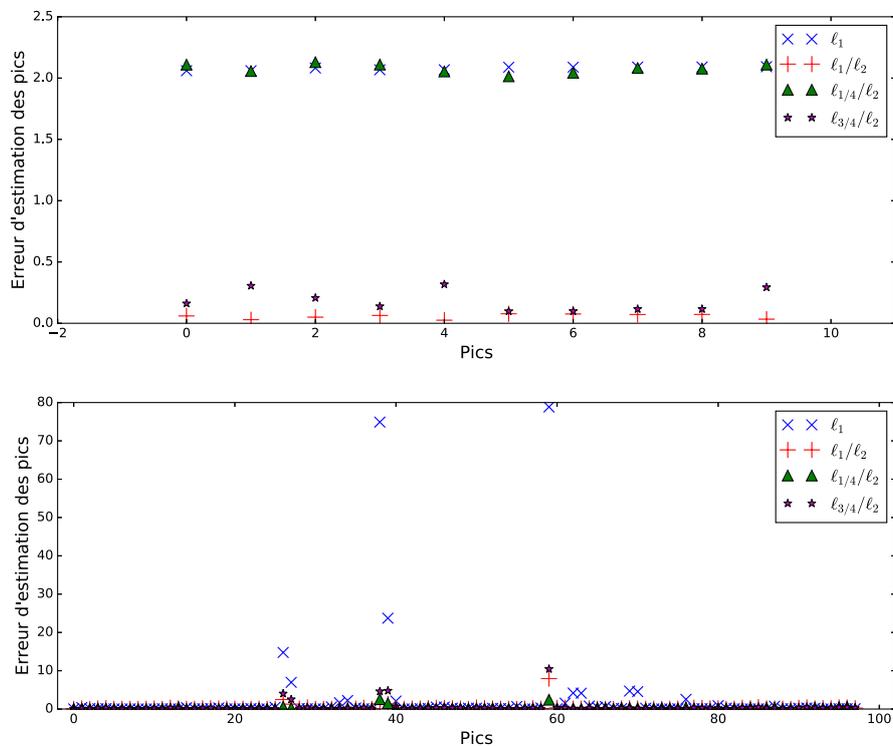


FIGURE 6.4 – Erreur d'estimation d'intensités des différentes reconstructions des signaux A mono-chargé (en haut) et B multi-chargé (en bas), obtenus avec les pénalités : ℓ_1 , ℓ_1/ℓ_2 , $\ell_{1/4}/\ell_2$ et $\ell_{3/4}/\ell_2$

6.6 Conclusion

Dans ce chapitre, nous avons introduit une nouvelle approche SPOQ fondée sur une approximation lisse de la pénalité ℓ_p/ℓ_q pour $p \in]0, 2]$ et $q \in [2, +\infty[$, comme étant une généralisation de la pénalité ℓ_1/ℓ_2 [Repetti 2015]. Nous avons étudié cette fonction et nous avons proposé un nouvel algorithme à métrique variable localement ajustée, basé sur la forme générique du FB (Algorithme 2) et la stratégie MM [Chouzenoux 2014]. Dans ce cadre, une étude de convergence a été réalisée. Notre application de cette pénalité pour la restauration des signaux en spectrométrie de masse a montré des résultats efficaces pour quelques valeurs de p et de q , meilleurs que ceux rapportés par la régularisation ℓ_1 .

Par conséquent, notre contribution à la résolution du problème inverse en spectrométrie de masse porte sur deux algorithmes efficaces : Algorithme 8 et Algorithme 9. Ces deux algorithmes se distinguent par le choix de la pénalité et la stratégie de résolution utilisée. Nous avons vu dans le Chapitre 5 que l'Algorithme 8 assure une reconstruction efficace des signaux parcimonieux en SM mais avec une sous-estimation d'intensité dans certains cas. Dans ce chapitre, nous avons proposé un nouvel Algorithme 9 qui résout ce problème avec une estimation précise de positions et d'intensités des signaux parcimonieux en SM. Néanmoins, cela est au prix d'un accroissement important du coût de calcul d'au minimum un facteur 20 par rapport au temps de calcul requis par l'Algorithme 8.

Méthode de débruitage des données spectrales de grande dimension

Sommaire

7.1	Introduction	97
7.2	L'échantillonnage	97
7.2.1	Théorème de Shannon-Nyquist	97
7.2.2	Échantillonnage non uniforme	98
7.3	Méthodes de débruitage en biophysique	99
7.3.1	Papoulis-Gerchberg	99
7.3.2	Cadzow	100
7.3.3	urQRd	100
7.4	Nouvelle méthode PGSane	101
7.4.1	Modélisation du problème	101
7.4.2	Nouvelle proposition	103
7.5	Application sur des signaux synthétiques	103
7.6	Conclusion	109

7.1 Introduction

La restauration de l'information nécessite toujours la compréhension de la procédure de mesure. Selon le principe d'Heisenberg, toute mesure est accompagnée d'une incertitude [Heisenberg 1966]. L'échantillonnage régulier, sous la condition de Shannon consiste à minimiser l'incertitude de mesure pour garantir une perte minimale d'information du signal [Shannon 1948]. À l'inverse, un échantillonnage non uniforme parcimonieux peut conduire à une perte d'information et une augmentation d'incertitude qu'on appelle "bruit d'échantillonnage".

Dans les modèles de mesure de grande dimension, l'échantillonnage de Shannon est très coûteux et le recours à un échantillonnage non uniforme parcimonieux est indispensable. Cependant, le bruit d'échantillonnage devient très important dans ce cas.

Nous avons vu dans le Chapitre 5 que les spectres de masses sont de grande dimension et l'un des problèmes majeurs en SM est que le bruit élevé peut cacher des informations importantes. L'analyse des données sera donc bien meilleure si les mesures sont préalablement débruitées. De ce concept, nous avons eu l'idée de développer une approche pour débruiter les mesures de grande dimension en RMN et en SM. Notre but est de proposer des méthodes pour restaurer un spectre de Fourier, correspondant à une mesure RMN ou SM à partir des échantillons non uniforme du signal temporel, bénéficiant d'un cadre théorique solide. Nous commencerons par présenter l'échantillonnage classique et l'intérêt de l'échantillonnage non uniforme parcimonieux dans la Section 7.2. Dans la Section 7.3, nous présenterons quelques méthodes de débruitage de spectres de Fourier utilisées en biophysique. Nous proposerons ensuite une modélisation du problème et une nouvelle méthode de débruitage basée sur une approche proximale dans la Section 7.4. Finalement, la Section 7.5 sera consacrée à présenter une application de notre approche et une comparaison avec les autres méthodes existantes sur des spectres synthétiques.

7.2 L'échantillonnage

7.2.1 Théorème de Shannon-Nyquist

Le principe d'échantillonnage permet de présenter une information en temps continu avec un signal en temps discret $((t_n)_{n \in \mathbb{Z}})$. Cela revient généralement à multiplier le signal temporel $x(t)$ par une somme d'impulsions de Dirac $u(t)$ définie par :

$$u(t) = \sum_{n \in \mathbb{Z}} \delta(t - t_n) \quad (7.1)$$

Par la suite, le signal échantillonné est donné par :

$$x_e(t) = \sum_{n \in \mathbb{Z}} x(t_n) \delta(t - t_n) \quad (7.2)$$

Dans le domaine fréquentiel, la transformée de Fourier de la fonction de Dirac (7.1) est donnée par :

$$U(f) = \sum_{n \in \mathbb{Z}} \exp(-j2\pi f t_n) \quad (7.3)$$

Par conséquent, l'opération d'échantillonnage est un produit de convolution entre le signal dans le

domaine fréquentiel $X(f)$ et la fonction $U(f)$:

$$X_e(f) = X(f) * U(f) = \sum_{n \in \mathbb{Z}} x(t_n) \exp(-j2\pi f t_n) \quad (7.4)$$

Selon une grille d'échantillonnage régulière, pour convertir un signal analogique en un signal numérique sans perte d'informations, le théorème de Shannon-Nyquist consiste à échantillonner le signal cible avec une fréquence d'échantillonnage f_e de valeur supérieure ou égale à deux fois la plus haute fréquence de son spectre f_{\max} [Shannon 1948, Shannon 1949] :

$$f_e \geq 2f_{\max} \quad (7.5)$$

Cette fréquence d'échantillonnage f_e conditionne la dimension du signal mesuré. La technique d'échantillonnage la plus utilisée est l'échantillonnage uniforme. Cependant, cette technique pose de nombreux problèmes, notamment une forte consommation d'énergie si le nombre de mesures à effectuer N est grand. Nous avons vu par exemple en spectrométrie de masse que l'axe des m/z présente un grand nombre d'échantillons N sur une longue durée d'acquisition allant de quelques minutes jusqu'à quelques heures [Qi 2014]. L'échantillonnage non uniforme représente une approche fondamentale pour contourner ce problème.

7.2.2 Échantillonnage non uniforme

Un échantillonnage non uniforme NUS (Non Uniform Sampling) vise à dépasser la limite du théorème de Shannon-Nyquist (7.5) pour s'adapter aux limites et à la dimension du modèle d'acquisition. Le principe de cette méthode consiste à échantillonner le signal selon une grille d'échantillonnage régulière (qui suit le théorème de Shannon-Nyquist) mais avec des trous (Figure 7.1). Plus on a d'échantillons manquants, plus les écarts sont importants et le théorème de Shannon-Nyquist n'est plus valable même si la grille est initialement échantillonnée régulièrement.

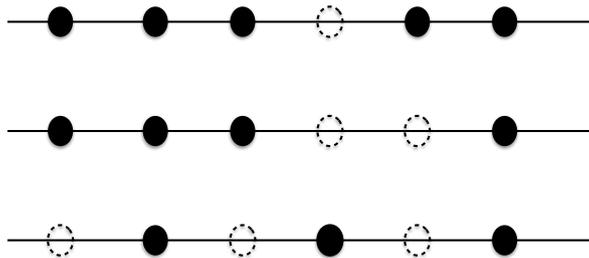


FIGURE 7.1 – Séquence incomplète d'échantillons relevés à une grille régulière

Le degré de parcimonie de ce sous-échantillonnage est donné par le rapport du nombre des échantillons mesurés et le nombre d'échantillons de Shannon-Nyquist. Plusieurs travaux [Candès 2006, Donoho 2006] ont montré la possibilité de restaurer un signal sous-échantillonné sous quelques conditions sur son spectre de Fourier. Cette stratégie a été utilisée dans différentes applications telles que la RMN [Hoch 2014, Kazimierczuk 2015] et la SM [Bray 2017] qui possèdent de grandes dimensions. Néanmoins, mesurer seulement M échantillons choisis aléatoirement à partir des N

échantillons réguliers de la grille temporelle fait apparaître un bruit d'échantillonnage non uniforme sur le spectre de Fourier associé. Le spectre sera dans ce cas bruité et plus on réduit le nombre d'échantillons M , plus ce bruit devient important. À ce niveau, l'analyse efficace des données nécessite une approche de débruitage accompagnée d'une méthode de restauration des mesures manquantes. Nous présenterons dans la section suivante quelques approches proposées en biophysique pour résoudre ce problème.

7.3 Méthodes de débruitage en biophysique

Le NUS représente un outil efficace pour assurer l'acquisition de données de grande dimension. En biophysique, comme en SM et en RMN, il est souvent utilisé pour éviter les temps d'acquisition coûteux. Néanmoins, son utilisation nécessite des approches avancées pour assurer une analyse efficace. Nous présentons dans cette section une approche qui s'intéresse à la restauration des échantillons manquants dans le domaine temporel et quelques méthodes dédiées à la minimisation du bruit d'échantillonnage non uniforme, en particulier la méthode de "Cadzow" et "urQRd".

7.3.1 Papoulis-Gerchberg

Avec une grille régulière, le spectre de Fourier d'un signal sous-échantillonné sur l'axe temporel obtenu en remplaçant les échantillons manquants par des zéros, produit un bruit d'échantillonnage important. Une procédure itérative proposée par Papoulis-Gerchberg consiste à restaurer les échantillons manquants selon les étapes suivantes, basées sur l'hypothèse que le spectre de Fourier du signal cible est parcimonieux :

1. Selon une grille régulière, mettre à zéros tous les points manquants.
2. Passer du signal sous-échantillonné ($\mathbf{x} \in \mathbb{R}^N$) à un spectre de Fourier :

$$X \leftarrow \mathcal{F}(\mathbf{x})$$

3. Appliquer un seuillage dur (Hard Thresholding) avec un seuil σ donné pour mettre à zéros toutes les amplitudes inférieures à ce seuil :

$$\tilde{X} \leftarrow \text{prox}_{\sigma \ell_0}(X)$$

4. Revenir à l'espace de mesure avec une transformée de Fourier inverse :

$$\hat{\mathbf{x}} \leftarrow \mathcal{F}^{-1}(\tilde{X})$$

5. Re-injecter les échantillons aléatoires initiaux

Après quelques itérations, le vecteur $\hat{\mathbf{x}}$ tend vers signal complet, sans artefacts sous l'hypothèse de bien choisir le seuil à utiliser dans l'étape de seuillage dur. L'approche est itérative de sorte que l'inverse de la transformée de Fourier du spectre soit le plus proche possible du signal sous-échantillonné en ignorant les points manquants.

Une application efficace de cette stratégie de Papoulis-Gerchberg (notée PG) a été effectuée en imagerie [Papoulis 1973], en optique [Gerchberg 1974] et en RMN [Kuethe 1999]. Néanmoins, sa qualité de restauration reste limitée à un échantillonnage non uniforme avec un nombre de mesures suffisant et un degré de parcimonie du spectre important, et requiert une estimation efficace du niveau de bruit.

7.3.2 Cadzow

La décomposition en valeurs singulières SVD (Singular Value Decomposition) [Muller 2004] est un outil efficace pour retirer le maximum d'artéfacts d'un signal donné. C'est une technique mathématique qui décompose une matrice donnée \mathbf{A} en trois matrices :

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V} \quad (7.6)$$

où Σ est une matrice diagonale qui contient les valeurs singulières de la matrice \mathbf{A} telle que $\Sigma = \text{Diag}(\sigma_1, \sigma_2, \dots, \sigma_N)$ avec $\sigma_1 > \sigma_2 > \dots > \sigma_N$. Sur la diagonale de Σ , les valeurs singulières sont classées par ordre décroissant : les plus grandes correspondent aux intensités importantes du signal, et les valeurs les plus faibles correspondent au bruit. Une troncature de r colonnes de la matrice Σ , l'équivalent de la conservation des r échantillons les plus importants du signal d'intérêt, conduit systématiquement à une réduction du bruit de la matrice \mathbf{A} donnée par la multiplication des matrices tronquées de rang réduit r . Cette approche est connue sous le nom de la méthode de Cadzow [Cadzow 1988], utilisée dans différentes applications comme la compression d'image [Aharon 2006], l'analyse en composante principale [Wall 2003], etc. Pour débruiter un signal composé d'une somme de peu de sinusoides (i.e., de spectre de Fourier parcimonieux), on peut utiliser l'approche de Cadzow sur sa matrice de Hankel notée $\mathbf{H} \in \mathbb{R}^{M \times T}$ construite à partir du signal sous-échantillonné $\mathbf{x} = (x_n)_{1 \leq n \leq N}$ tel que pour $T = N - M + 1$, $i = 1, \dots, M$ et $j = 1, \dots, T$:

$\mathbf{H}_{i,j} \leftarrow x_{i+j-1}$, c'est à dire :

$$\mathbf{H} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & \cdot & \cdot & \cdot & x_M \\ x_2 & x_3 & x_4 & \cdot & \cdot & \cdot & \cdot & x_{M+1} \\ x_3 & x_4 & \cdot & \cdot & \cdot & \cdot & \cdot & x_{M+2} \\ x_4 & \cdot \\ \cdot & \cdot \\ \cdot & \cdot & x_M & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & x_M & x_{M+1} & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_M & x_{M+1} & x_{M+2} & \cdot & \cdot & \cdot & \cdot & x_T \end{pmatrix} \quad (7.7)$$

Néanmoins, l'application de cette approche reste limitée à des signaux de dimension réduite. En effet, le calcul de la SVD des spectres en très grande dimension est prohibitive ce qui rend cet algorithme inapplicable en spectrométrie de masse ou en résonance magnétique nucléaire.

7.3.3 urQRd

Une nouvelle stratégie de débruitage fondée sur la décomposition QR a été proposée en 2014 [Chiron 2014]. Cette stratégie consiste à effectuer les étapes suivantes pour un signal $\mathbf{x} \in \mathbb{R}^N$:

1. Construire la matrice de forme Hankel ($\mathbf{H} \in \mathbb{R}^{M \times T}$) (7.7) à partir du signal temporel.
2. Projeter cette matrice sur une base aléatoire de dimension réduite avec une matrice aléatoire $\Theta \in \mathbb{R}^{T \times K}$:

$$\mathbf{Y} \leftarrow \mathbf{H}\Theta$$

3. Décomposer la matrice obtenue sous la forme QR [Trefethen 1997] :

$$(\mathbf{Q}, \mathbf{R}) \leftarrow QR(\mathbf{Y})$$

4. Projeter la matrice de \mathbf{H} sur la base \mathbf{Q} de rang réduit $r \leq M$:

$$\tilde{\mathbf{H}} \leftarrow \mathbf{Q}\mathbf{Q}^*\mathbf{H}$$

5. Restaurer le signal $\hat{\mathbf{x}} = (\hat{x}_n)_{n \in \mathbb{N}}$ à partir de la matrice obtenue pour $n = 1, \dots, N$:

$$\hat{x}_n \leftarrow \langle \tilde{\mathbf{H}}_{i,j} \rangle_{i+j=n+1}$$

Itérativement, le signal $\hat{\mathbf{x}}$ tend vers un signal débruité sous la condition de bien choisir le rang de la matrice de Hankel. Cette procédure est appelée rQRd (random QR decomposition). Ce nom vient du fait que l'algorithme se fonde sur la décomposition QR avec la projection sur des matrices aléatoires. Une nouvelle version accélérée du rQRd a été proposée par Chiron *et al.* [Chiron 2014] sous le nom de urQRd (uncloied random QR decomposition) qui possède un calcul rapide des matrices de Hankel avec les nouvelles outils de calculs "Fast Hankel" [Korobeynikov 2009, Golub 2012]. Sane (Support selection Algorithm for Noise Elimination) est une version plus élaborée de l'urQRd qui possède le même principe mais avec une projection sur la base \mathbf{Q} au lieu d'une matrice aléatoire (étape 2) [Chiron 2016]. L'intérêt de ce changement vient du fait qu'avec la matrice \mathbf{Q} , la solution reste proche des données originales dès les premières itérations.

L'application de Sane a prouvé son efficacité, sa robustesse et sa rapidité pour débruiter des mesures FT-ICR avec différents types de bruit (bruit additif, bruit de scintillation, bruit d'échantillonnage, etc.) [Chiron 2014]. De plus, l'utilisation de la projection aléatoire et la décomposition en QR contrairement aux différentes stratégies qui utilisent la SVD [Cadzow 1988, Martinsson 2006], fournit une rapidité remarquable lors de l'analyse de données de grande dimension.

Bien que cet algorithme a servit récemment à débruiter plusieurs spectres dans plusieurs applications SM [Floris 2016, Cheng 2016], l'utilisation d'un signal temporel échantillonné sur une grille régulière suffisamment fine est indispensable.

7.4 Nouvelle méthode PGSane

7.4.1 Modélisation du problème

Le principe de l'algorithme Papoulis-Gerchberg consiste à restaurer les mesures manquantes en partant du spectre vers le signal temporel. Par ailleurs, les algorithmes urQRd et sa version améliorée Sane reposent sur des projections alternées, faisant intervenir des espaces non convexes avec la pénalité ℓ_0 et le rang faible des matrices de Hankel. Bien que ces algorithmes soient efficaces à la restauration de données de grande dimension, leurs stratégies algorithmiques ne bénéficient pas de garanties de convergence. Dans cette section, nous proposons de redéfinir des algorithmes traitant des problèmes similaires, dans le cadre des approches proximales. Une nouvelle approche sera ensuite proposée qui combinera la restauration les échantillons manquants (comme le fait l'approche de Papoulis-Gerchberg) et la réduction du bruit d'échantillonnage non uniforme (comme le fait l'algorithme Sane). Notre approche sera fondée sur la recherche d'un vecteur à l'intersection :

- l'espace des mesures sous-échantillonnées
- l'espace des signaux harmoniques pour utiliser les matrices de Hankel et introduire le principe de l'algorithme Sane
- l'espace de Fourier (domaine fréquentiel du signal mesuré) pour utiliser le principe de Papoulis-Gerchberg

Nous expliquerons cette procédure dans la suite de cette section.

7.4.1.1 Contrainte de parcimonie dans le domaine fréquentiel

Commençons par la problématique traitée par l'algorithme Papoulis-Gerchberg, nous modélisons le problème de débruitage comme suit :

$$\underset{\mathbf{x} \in \mathbb{S}}{\text{minimiser}} \quad \Psi(\mathcal{F}\mathbf{x}) \quad (7.8)$$

où $\mathbf{x} = (x_n)_{1 \leq n \leq N}$ présente le signal complet à restaurer de taille N , \mathcal{F} présente la transformée de Fourier et Ψ la fonction de pénalité. Nous désignons par \mathbb{S} l'espace d'échantillonnage non uniforme qui présente les mesures brutes dans le domaine temporel tel que $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^N \mid x_n = y_n\}$ si n appartient à la liste des échantillons mesurés et y_n la mesure d'indice temporel n .

Le problème (7.8) est équivalent à :

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimiser}} \quad \iota_{\mathbb{S}}(\mathbf{x}) + \Psi(\mathcal{F}\mathbf{x}) \quad (7.9)$$

où la fonction de pénalité Ψ sera définie dans ce cas par la norme convexe ℓ_1 , la plus proche de la pseudo-norme ℓ_0 utilisée dans l'algorithme de Papoulis-Gerchberg. L'Algorithme 11 résout le problème (7.9) en s'appuyant sur la base générique de l'algorithme Douglas-Rachford, donné dans ce cas par :

Algorithme 11 Douglas-Rachford sous contrainte de parcimonie dans le domaine fréquentiel

$$\left[\begin{array}{l} \mathbf{x}_0 \in \mathbb{R}^N \\ \gamma \in]0, +\infty[, \lambda \in]0, 2[\\ \text{Pour } k = 0, 1, \dots \\ \quad \mathbf{y}_k = \mathcal{F}^{-1}(\text{prox}_{\gamma\Psi}(\mathcal{F}\mathbf{x}_k)) \\ \quad \mathbf{z}_k = \mathcal{P}_{\mathbb{S}}(2\mathbf{y}_k - \mathbf{x}_k) \\ \quad \mathbf{x}_{k+1} = \mathbf{x}_k + \lambda(\mathbf{z}_k - \mathbf{y}_k) \end{array} \right.$$

où l'opérateur $\mathcal{P}_{\mathbb{S}}$ désigne la projection sur l'ensemble des mesures sous-échantillonnées. La suite $(\mathbf{x}_k)_{k \in \mathbb{N}}$ de cet algorithme converge vers la solution du problème (7.9).

7.4.1.2 Contrainte du rang réduit de la matrice de Hankel

L'idée principale de l'urQRd consiste à restaurer le signal sous une contrainte d'attache aux données mesurées et une contrainte d'un rang réduit de la matrice de Hankel. Nous traduisons ce concept par le modèle suivant :

$$\text{Trouver } \mathbf{x} \text{ tel que } \mathbf{x} \in \mathbb{S} \cap \mathbb{H}_r \quad (7.10)$$

où $\mathbb{H}_r = \{\mathbf{x} \in \mathbb{R}^N \mid \text{rang}(\mathbf{H}(\mathbf{x})) \leq r\}$, $\mathbf{H}(\mathbf{x}) \in \mathbb{R}^{M \times T}$ désigne la matrice de Hankel d'ordre $M \leq N/2$ associée à \mathbf{x} telle que :

$$(\forall i \in \{1, \dots, M\})(\forall j \in \{1, \dots, T\}) \quad \mathbf{H}_{ij}(\mathbf{x}) = x_{i+j-1} \quad (7.11)$$

avec la dimension $T = N - M + 1$ et le rang $r \in [1, M]$ [Chiron 2014]. Le recours à l'algorithme Douglas-Rachford (Algorithme 3) avec ces conditions donne lieu à l'algorithme suivant :

Algorithme 12 Douglas-Rachford sous la contrainte du rang réduit

$$\left[\begin{array}{l}
\mathbf{x}_0 \in \mathbb{R}^N \\
\gamma \in]0, +\infty[, \lambda \in]0, 2[\\
r \in [1, M] \\
\text{Pour } k = 0, 1, \dots \\
\left[\begin{array}{l}
\mathbf{y}_k = \mathcal{P}_{\mathbb{H}_r}(\mathbf{x}_k) \\
\mathbf{z}_k = \mathcal{P}_{\mathbb{S}}(2\mathbf{y}_k - \mathbf{x}_k) \\
\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda(\mathbf{z}_k - \mathbf{y}_k)
\end{array} \right.
\end{array} \right.$$

avec $\mathcal{P}_{\mathbb{S}}$ présente toujours la projection sur l'ensemble des mesures sous-échantillonnées et $\mathcal{P}_{\mathbb{H}_r}(\mathbf{x})$ désigne la projection estimée avec la technique de [Chiron 2014] :

1. Construire la matrice de Hankel \mathbf{H} à partir de \mathbf{x}
2. Construire $\tilde{\mathbf{H}}$, la projection de \mathbf{H} du rang réduit r en utilisant la décomposition QR.
3. Calculer $\bar{\mathbf{x}} = (\bar{x}_n)_{n \in \mathbb{N}}$ tel que $\bar{x}_n = \left\langle \tilde{H}_{ij} \right\rangle_{i+j=n+1}$ pour tout $n \in \{1, \dots, N\}$.

La suite $(\mathbf{x}_k)_{k \in \mathbb{N}}$ de l'Algorithme 12 tend vers la solution du problème (7.10)

7.4.2 Nouvelle proposition

La fusion du principe de l'algorithme PG (7.9) et du principe de l'algorithme Sane (7.10), nommée PGSane et menant à restaurer les mesures du signal temporel et débruiter le spectre de Fourier, peut être décrite par le problème suivant :

$$\underset{\mathbf{x} \in \mathbb{S} \cap \mathbb{H}_r}{\text{minimiser}} \Psi(\mathcal{F}\mathbf{x}) \quad (7.12)$$

Autrement dit, il s'agit d'une minimisation de la fonction de pénalité Ψ du spectre de Fourier $\mathcal{F}\mathbf{x}$ où le signal recherché \mathbf{x} doit vérifier les deux contraintes $\mathbf{x} \in \mathbb{S}$ et $\mathbf{x} \in \mathbb{H}_r$. Cela revient à alterner les projections entre trois espaces : l'espace temporel dans lequel l'échantillonnage irrégulier est effectué, l'espace des signaux harmoniques dans lequel nous pouvons construire les matrices de Hankel à rang réduit, et l'espace fréquentiel dans lequel se présente le spectre de Fourier du signal mesuré.

À ce niveau, nous pouvons utiliser l'algorithme PPXA [Combettes 2011]. Une solution du problème (7.12) sera obtenue par la limite de la suite $(\mathbf{v}_k)_{k \in \mathbb{N}}$ générée par l'Algorithme 13.

Notons que la projection sur l'espace \mathbb{H}_r des matrices de Hankel définies par (7.7) n'est pas convexe. Par conséquent, aucune garantie théorique de convergence du PPXA n'est disponible dans l'état actuel de la littérature. Toutefois, nos résultats numériques que nous présenterons dans la suite montre que la méthode est stable.

7.5 Application sur des signaux synthétiques

Nous créons un signal synthétique A qui correspond à un FID^a de 128.000 échantillons ayant un spectre composé de 10 fréquences de différentes intensités allant de 1 à 10. La Figure 7.2 présente ce signal complet et sa mesure sous-échantillonnée. La Figure 7.2-b) présente un zoom sur le signal dans le domaine temporel. Nous présentons en bleu les échantillons irréguliers que nous mesurons à partir du signal continu (en noir).

a. FID : Free Induction Decay, c'est le signal d'induction libre en résonance magnétique

Algorithme 13 PPXA sous contraintes de parcimonie dans le domaine fréquentiel et du rang réduit de la matrice de Hankel

$$\begin{array}{l}
 \gamma \in]0, +\infty[, \lambda \in]0, 2[\\
 (\mathbf{x}_{0,i})_{1 \leq i \leq 3} \in \mathbb{R}^N, \\
 \sum_{i=1}^3 w_i = 1, \quad w_i \in]0, 1[, \\
 \mathbf{v}_0 = w_1 \mathbf{x}_{0,1} + w_2 \mathbf{x}_{0,2} + w_3 \mathbf{x}_{0,3} \\
 \text{Pour } k = 0, 1, \dots \\
 \left[\begin{array}{l}
 \mathbf{y}_{k,1} = \mathcal{P}_{\mathbb{H}_r}(\mathbf{x}_{k,1}) \\
 \mathbf{y}_{k,2} = \mathcal{P}_{\mathbb{S}}(\mathbf{x}_{k,2}) \\
 \mathbf{y}_{k,3} = \mathcal{F}^{-1}(\text{prox}_{\gamma\Psi/w_3}(\mathcal{F}\mathbf{x}_{k,3})) \\
 \mathbf{z}_k = w_1 \mathbf{y}_{0,1} + w_2 \mathbf{y}_{0,2} + w_3 \mathbf{y}_{0,3} \\
 \mathbf{x}_{k+1,1} = \mathbf{x}_{k,1} + \lambda(2\mathbf{z}_k - \mathbf{v}_k - \mathbf{y}_{k,1}) \\
 \mathbf{x}_{k+1,2} = \mathbf{x}_{k,2} + \lambda(2\mathbf{z}_k - \mathbf{v}_k - \mathbf{y}_{k,2}) \\
 \mathbf{x}_{k+1,3} = \mathbf{x}_{k,3} + \lambda(2\mathbf{z}_k - \mathbf{v}_k - \mathbf{y}_{k,3}) \\
 \mathbf{v}_{k+1} = \mathbf{v}_k + \lambda(\mathbf{z}_k - \mathbf{v}_k)
 \end{array} \right.
 \end{array}$$

Nous proposons d'appliquer les deux algorithmes Papoulis-Gerchberg et Sane, ainsi que les algorithmes que nous avons proposés pour résoudre les mêmes problèmes : Algorithme 11, Algorithme 12, pour débruiter le signal synthétisé.

Selon les résultats de la Figure 7.3 et la Figure 7.4, bien que le niveau de bruit est réduit dans le spectre de Fourier du signal restauré par PG (Figure 7.3-a) en comparaison avec le bruit d'échantillonnage (Figure 7.2-c), le résidu de reconstruction reste encore non négligeable comme nous le voyons dans la Figure 7.3-b.

Nous présentons dans la Figure 7.4 le résultat de restauration du signal A avec l'Algorithme 11 en utilisant la fonction de pénalité ℓ_1 . Le résultat de restauration obtenu avec l'Algorithme 11, apparaît équivalent à celui obtenu avec l'algorithme Papoulis-Gerchberg. Néanmoins, seul l'Algorithme 11 avec la pénalité de ℓ_1 présente une garantie de convergence. Rappelons que notre but est de développer un nouvel algorithme de débruitage fondé sur une base algorithmique convergente.

Selon les résultats de la Figure 7.6 et la Figure 7.5, l'Algorithme 12 assure un débruitage du signal A avec 100 itérations et un rang réduit $r = 10$ plus efficace que celui de l'algorithme Sane (Figure 7.5).

Puisque l'Algorithme 12 et l'Algorithme 11 donnent des résultats satisfaisants, nous proposons d'appliquer la combinaison des 2 formalismes, donnant lieu à l'Algorithme 13. Le résultat de débruitage présenté dans la Figure 7.7 montre que le principe d'alterner les contraintes produit une stratégie de débruitage efficace.

Le même algorithme est appliqué pour restaurer un signal B qui correspond à un FID de 128.000 échantillons d'un spectre composé de 10 fréquences d'intensités variant de 1 à 10 avec un bruit gaussien additif de niveau 5%, l'équivalent d'un RSB initial = 6.37 dB (Figure 7.8).

Le débruitage du signal B avec l'Algorithme 13 donne un RSB = 8.11 dB avec une détection des pics de faibles intensités (Figure 7.9-b).

Ces résultats sont encourageants pour mieux développer l'idée d'alterner les contraintes (7.12) et créer une nouvelle approche de débruitage avec des garanties de convergence. Bien que la forme générique du PPXA est proposée pour des fonctions de pénalités convexes, ce qui n'est pas vérifié dans notre cas avec la contrainte du rang réduit utilisée dans l'algorithme Sane et la norme ℓ_0 dans l'algorithme Papoulis-Gerchberg, les résultats montrent la stabilité de cette approche.

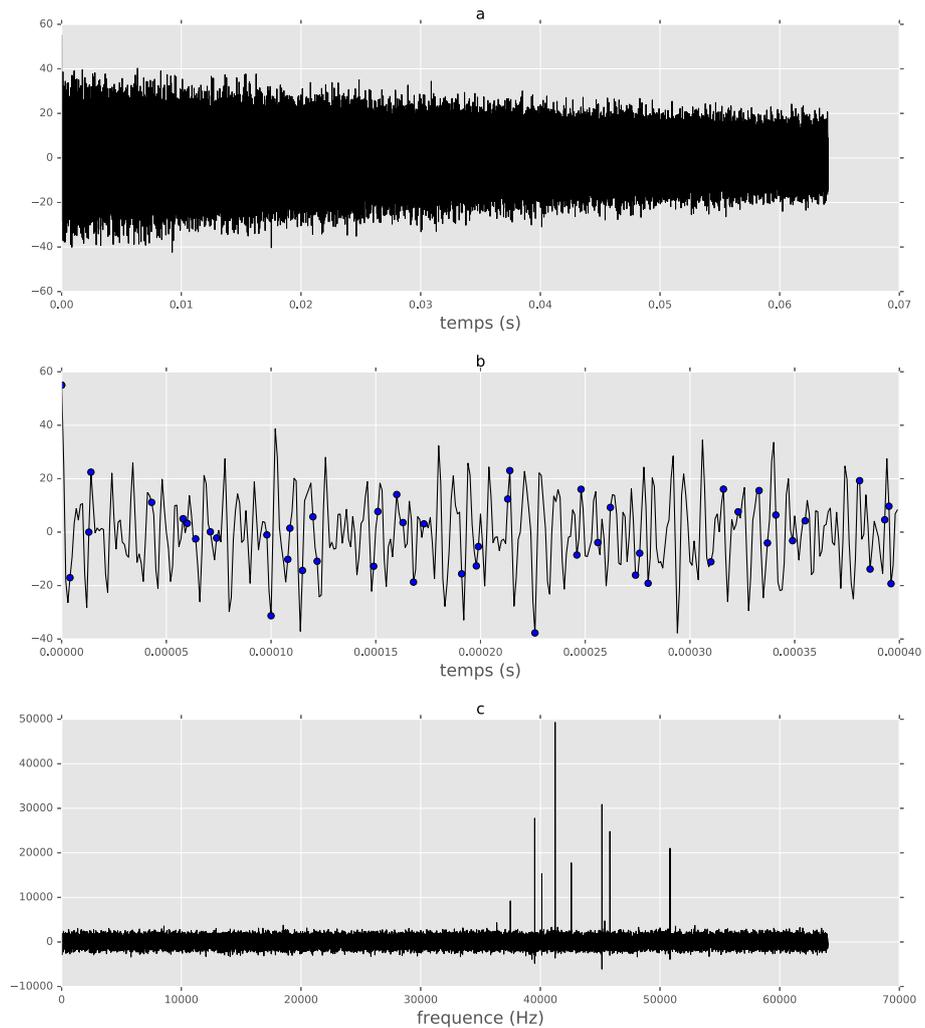


FIGURE 7.2 – a) FID à 128.000 échantillons d'un spectre de 10 fréquences de différentes intensités allant de 1 à 10, b) zoom sur une partie du FID sous-échantillonné avec un degré de parcimonie égale à $1/8$, c) FT du FID qui montre le bruit d'échantillonnage

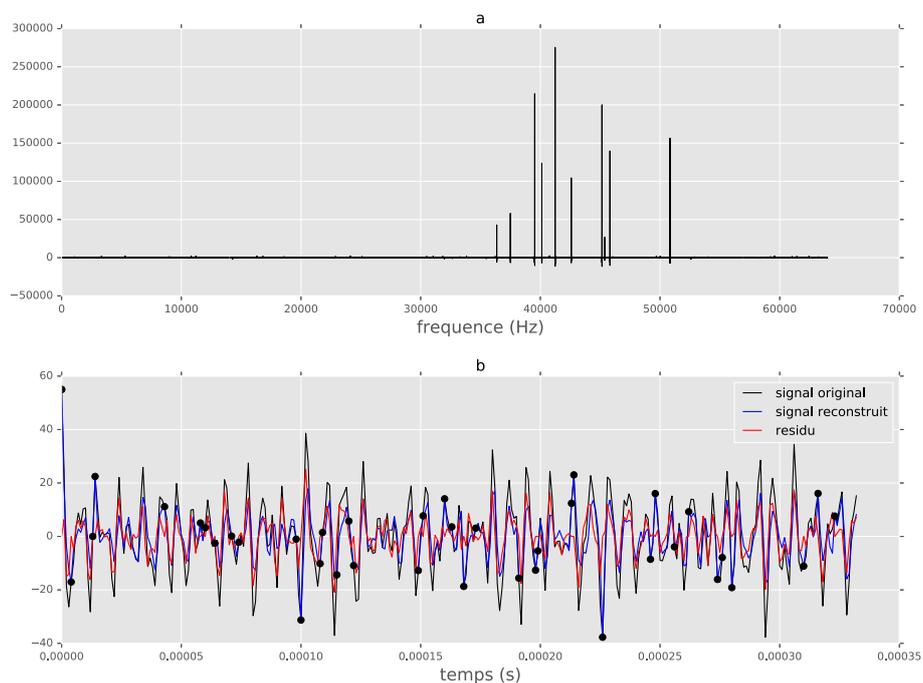


FIGURE 7.3 – Résultats de restauration du signal sous-échantillonné A donné dans la Figure 7.2 avec l’algorithme Papoulis-Gerchberg : a) FT du FID reconstruit avec 100 itérations, b) zoom sur le signal restauré dans le domaine temporel.

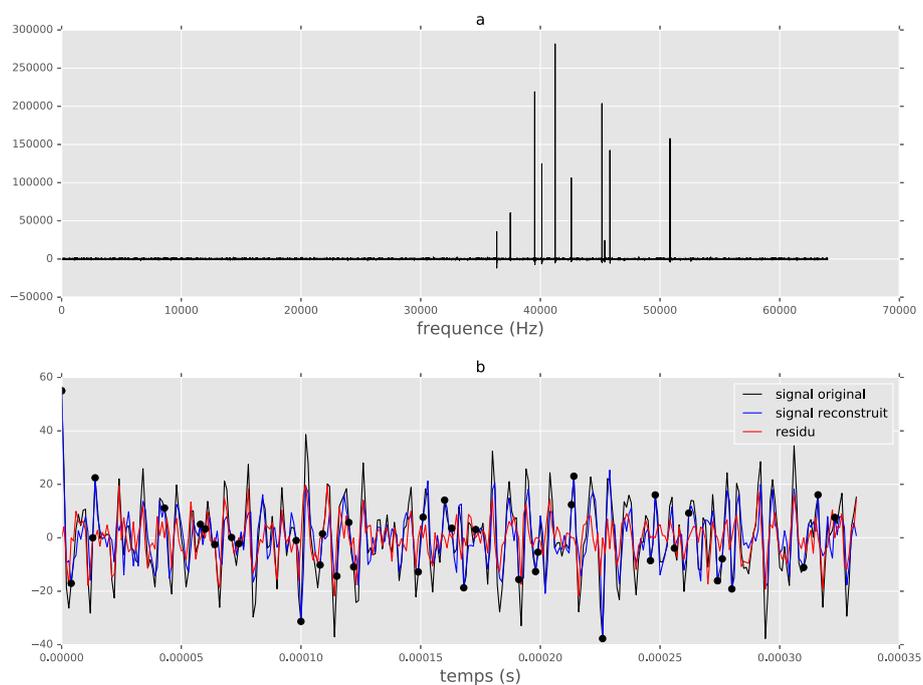


FIGURE 7.4 – Résultats de restauration du signal sous-échantillonné A donné dans la Figure 7.2 avec l’Algorithme 11 : a) FT du FID reconstruit avec 100 itérations, b) zoom sur le signal restauré dans le domaine temporel.

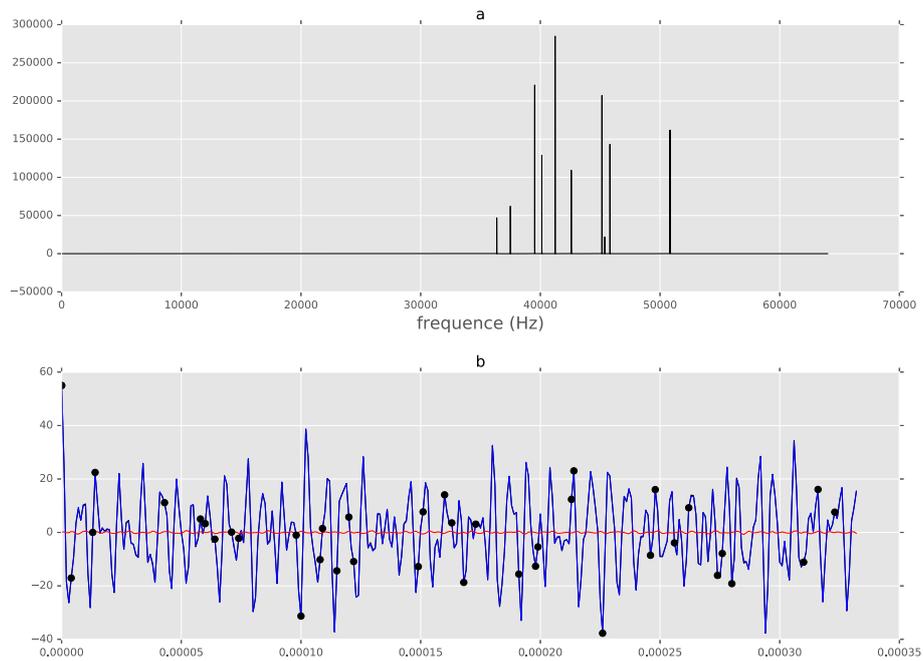


FIGURE 7.5 – Résultats de restauration du signal sous-échantillonné A donné dans la Figure 7.2 avec l’Algorithme Sane : a) FT du FID reconstruit (100 itérations et rang=10), b) zoom sur le signal restauré dans le domaine temporel

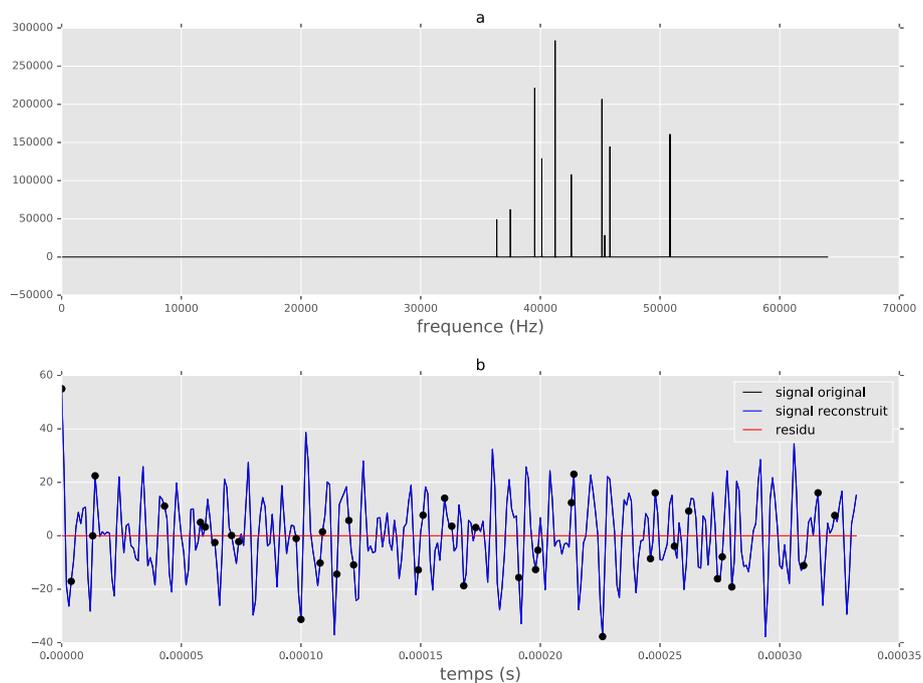


FIGURE 7.6 – Résultats de restauration du signal sous-échantillonné A donné dans la Figure 7.2 avec l’Algorithme 12 : a) FT du FID reconstruit (100 itérations et rang=10), b) zoom sur le signal restauré dans le domaine temporel

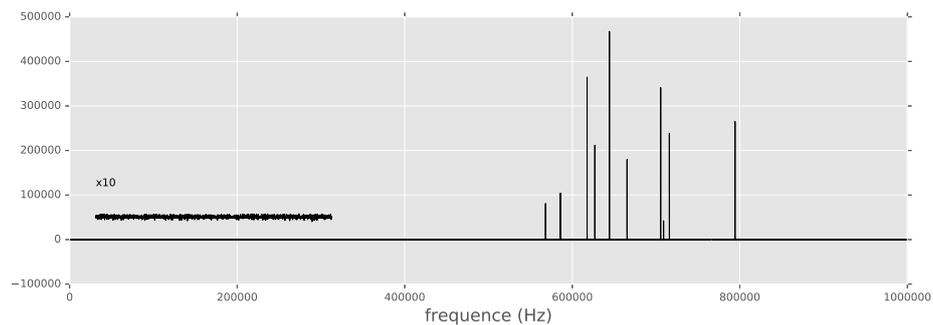


FIGURE 7.7 – Résultats de restauration du signal sous-échantillonné A donné dans la Figure 7.2 avec l’Algorithme 13, FT du FID reconstruit (100 itérations et rang=10)

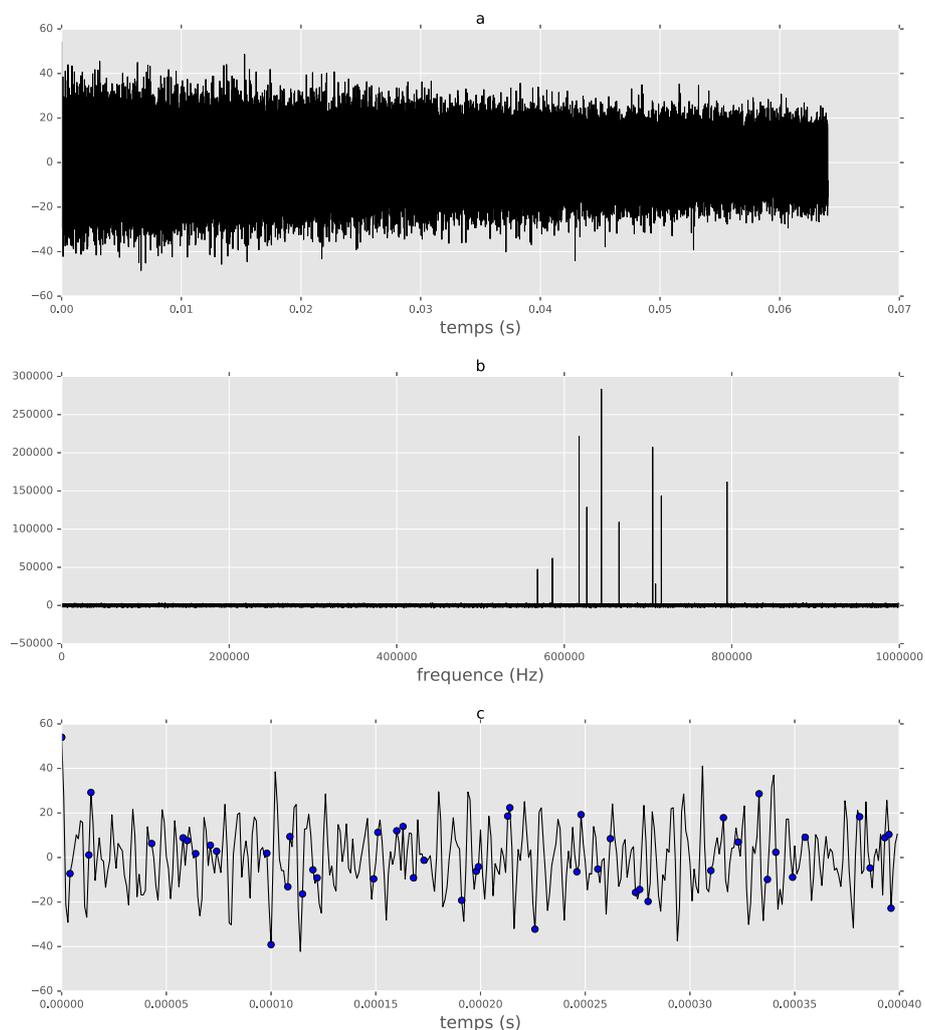


FIGURE 7.8 – a) FID à 128.000 échantillons d’un spectre de 10 fréquences de différentes intensités allant de 1 à 10 avec un bruit gaussien additif de niveau 5%, b) FT du FID qui montre le bruit d’échantillonnage, c) zoom sur une partie du FID sous-échantillonné avec un degré de parcimonie égale à $1/8$

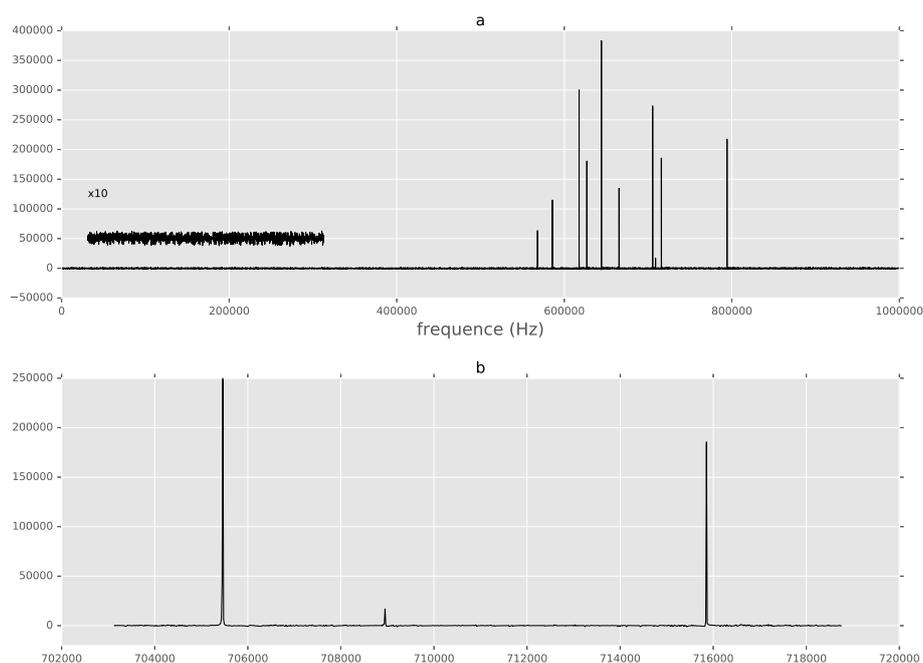


FIGURE 7.9 – Résultats de restauration du signal sous-échantillonné B donné dans la Figure 7.8 avec l’Algorithme 13 a) FT du FID reconstruit (100 itérations et rang=10), b) zoom sur un pic de faible intensité

7.6 Conclusion

Dans ce chapitre, nous avons abordé la notion du débruitage des données biophysiques de grande dimension dans le cas où leur spectre de Fourier est parcimonieux. Nous avons présenté les algorithmes (PG et Sane) les plus connus et les plus efficaces pour débruiter et restaurer ces signaux dans le cas où ils sont sous-échantillonnés non uniformément en Spectrométrie de Masse ou en Résonance Magnétique Nucléaire. Chacun des deux algorithmes présente des contraintes soit au niveau de la dimension des données, soit au niveau du coût de calcul demandé. Une combinaison des deux stratégies nous paraît alors comme une approche naturelle à tester. Cette idée a fait l’objet de ce chapitre où nous avons montré qu’une qualité de restauration très satisfaisante peut être garantie avec l’algorithme 13 basé sur le schéma proximal PPXA sur des données synthétiques. Une perspective à court terme de ce travail est d’analyser les propriétés de convergence théoriques de l’Algorithme 13. La combinaison de cette approche avec les algorithmes d’analyse de données SM des chapitres 5 et 6 devra également être investiguée.

Conclusion & perspectives

Conclusion

Ce travail de thèse s'intéresse à la résolution des problèmes inverses en biophysique, et en particulier aux expériences de RMN de type DOSY et à la SM. En effet, ces deux modalités donnent lieu à des problèmes inverses ayant le caractère mal posé, ce qui nécessite le recours à des méthodes d'analyse avancées pour assurer la restauration efficace de l'information cible.

Dans un premier temps, nous avons commencé par étudier le problème inverse en DOSY. Pour le résoudre, nous avons proposé une méthode de régularisation hybride qui combine la pénalité entropique et le degré de parcimonie du signal cible. Différentes combinaisons ont été proposées entre l'entropie de Shannon et l'entropie de Burg avec les pénalités de parcimonie comme la norme ℓ_1 , la pseudo-norme ℓ_0 , la somme de Cauchy et la pénalité log-sum. Après avoir calculé les opérateurs proximaux qui correspondent à nos fonctions de régularisation hybride [Cherni 2016], nous avons montré que la pénalisation hybride entre l'entropie de Shannon et la norme ℓ_1 assure une meilleure qualité de reconstruction des signaux DOSY monodisperses et polydisperses. Par conséquent, un nouvel algorithme répondant à cette pénalisation a été proposé sous le nom de PALMA. La convergence, la rapidité et la fiabilité de PALMA ont été prouvées par différents tests sur des données synthétiques et réelles [Cherni 2017]. De plus, nous avons créé un serveur web <http://palma.labo.igbmc.fr/> grâce auquel tous les utilisateurs peuvent analyser leurs données DOSY expérimentales facilement.

Dans un deuxième temps, nous avons abordé le problème inverse en SM. Cette application présente un modèle d'observation non linéaire compliqué, et des données de très grande dimension. Nous avons traité ce problème dans le cadre de l'analyse protéomique. Nous avons proposé une nouvelle approche par dictionnaire fondée sur le modèle averagine, ce qui nous permet de nous ramener à l'inversion d'un modèle linéaire sous contraintes de parcimonie. Notre démarche a montré que cette approche fournit des résultats de bonne qualité mais elle présente une complexité de calcul et nécessite une large mémoire de stockage due à la grande dimension des données en spectrométrie de masse. Par conséquent, nous avons développé une approximation par blocs du dictionnaire averagine menant à une qualité de reconstruction très satisfaisante avec un temps de calcul réduit et une occupation de mémoire limitée [Cherni 2018]. Une campagne

d'analyse des résultats synthétiques et expérimentaux a été menée et nous avons constaté qu'avec la pénalité ℓ_1 , une sous-estimation de l'intensité des pics mono-isotopiques peut se produire. Pour éliminer ce biais, nous avons proposé une nouvelle approche SPOQ fondée sur une version lisse du ratio de pénalités ℓ_p/ℓ_q pour $p \in]0, 2]$ et $q \in [2, +\infty[$. Nous avons étudié les propriétés de cette pénalité et un nouvel algorithme Forward-Backward à métrique variable localement ajustée a été développé. Cet algorithme est très efficace pour la restauration des signaux parcimonieux en spectrométrie de masse sans avoir de problèmes d'échelle mais il requiert un coût de calcul important en comparaison avec le temps de calcul requis par la pénalité ℓ_1 .

La dernière partie de ce manuscrit traite la restauration de signaux issus d'échantillons bruités de type NUS des données RMN ou SM de grande dimension. Nous avons proposé plusieurs algorithmes proximaux pour traiter ce problème de débruitage et interpolation, de façon séparée ou simultanée. Des tests numériques ont été présentés, pour valider nos approches sur des données synthétiques.

Perspectives

Optimisation non convexe

Pour résoudre le problème inverse mal posé en DOSY dans le Chapitre 3, nous avons proposé une régularisation hybride qui combine l'entropie et la parcimonie. Différentes fonctions ont été proposées [Cherni 2016]. Bien que notre choix se soit focalisé sur le cas convexe avec la régularisation hybride entre l'entropie de Shannon et la norme ℓ_1 , de part sa grande efficacité sur notre application RMN DOSY, nos opérateurs proximaux calculés peuvent être utiles pour d'autres applications. Citons en particulier le cas de l'application de régression logistique parcimonieuse, qui a fait l'objet d'une publication de conférence [Chierchia 2017]. Une perspective serait de proposer une nouvelle approche de minimisation pour pouvoir traiter les fonctions de pénalité non convexes que nous n'avons pas testées en pratique dans cette thèse telles que la pseudo-norme de parcimonie ℓ_0 .

Déconvolution aveugle

Pour la résolution du problème inverse en RMN de type DOSY, nous avons proposé une approche de régularisation hybride accompagnée d'un nouvel algorithme PALMA [Cherni 2017]. L'opérateur de mesure (opérateur de Laplace) est connu dans ce cas. Néanmoins, dans d'autres expériences RMN, l'opérateur de mesure est inconnu et l'approche de déconvolution aveugle peut être envisagée. Une perspective de notre travail sera donc de proposer une nouvelle version de l'algorithme PALMA qui assure la déconvolution aveugle des données RMN.

Modèle de convolution estimé à partir des mesures

Dans le Chapitre 5, nous avons posé le problème de spectrométrie de masse comme un problème de déconvolution où le spectre de masse mesuré est un produit de convolution entre le pic mono-isotopique de masse m et de charge z , et la distribution obtenue pour cette masse avec le modèle averaginé. Selon un axe de masse donné, nous avons proposé de calculer le profil isotopique qui correspond à chaque masse. Néanmoins, l'utilisation d'une distribution isotopique à partir du spectre mesuré peut être envisagée. L'utilisation d'une distribution estimée à partir de mesures paraît plus réaliste car elle présente le même paramétrage que le spectre de masse mesuré (résolution, niveau

de bruit, etc). Par conséquent, une perspective à court terme serait d'utiliser notre approche par dictionnaire et son approximation par blocs avec une distribution isotopique estimée à partir des mesures.

Méthode de débruitage et de restauration convexe

L'approche de débruitage que nous avons proposée dans le Chapitre 7 se fonde sur un ensemble de projections alternées. Notre idée consiste à développer un nouvel algorithme avec des garanties de convergence qui assure la restauration des données manquantes lors d'un échantillonnage de type NUS avec une minimisation du bruit d'échantillonnage non uniforme. Selon les résultats que nous avons obtenus, la combinaison des deux modalités PG et Sane fondée sur la forme générique de PPXA, représente un outil efficace pour l'analyse des données de grande dimension en RMN. Afin d'ajouter des garanties de convergence à notre approche, nous proposons d'utiliser une pénalité convexe ℓ_1 au lieu de la pénalité ℓ_0 avec une relaxation du rang réduit des matrices de Hankel avec la norme nucléaire ^a [Lewis 1996].

Apprentissage par réseau de neurones en biophysique

La Résonance Magnétique Nucléaire (RMN) et la Spectrométrie de Masse (SM) sont deux méthodes d'analyses biophysiques très puissantes utilisées intensivement dans les laboratoires de biologie afin d'étudier les protéines en solution sous leur forme native, de mesurer leurs interactions avec différents partenaires moléculaires et d'obtenir des informations structurales sur les différentes espèces présentes dans la solution. Ces deux techniques ont également en commun de produire une très grande quantité de données de haute résolution, ce qui dépasse la capacité du scientifique à analyser directement et nécessite l'utilisation de méthodes d'extraction automatique. Bien que nous ayons proposé deux solutions : PALMA pour le problème DOSY et l'approche par dictionnaire pour la déconvolution des massifs isotopiques, les nouveaux outils d'apprentissage par réseau de neurones (Deep Learning en anglais) [Dong 2014, Wang 2015] tels que la CNN ou ConvNet semblent être une technique efficace pour déconvoluer les données de grande dimension [Goodfellow 2016]. Une perspective de notre travail serait donc de développer des approches d'apprentissage automatique dédiées à l'analyse d'ensembles de données spectroscopiques de RMN et de SM en utilisant la modélisation des processus de mesure que nous avons proposée.

a. <http://proximity-operator.net>



Annexe A

Calcul des opérateurs proximaux

Rappelons que la fonction Ψ est séparable :

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \Psi(\mathbf{x}) = \sum_{n=1}^N \psi(x_n)$$

Ainsi, l'opérateur proximal peut être donné par :

$$\text{prox}_{\Psi}(\mathbf{x}) = (\text{prox}_{\psi}(x_n))_{1 \leq n \leq N}$$

Par définition, l'opérateur proximal d'une fonction ψ propre et semi-continue inférieurement est donné par l'opérateur multi-valué :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{\psi}(x) = \underset{u \in \mathbb{R}}{\text{Argmin}} \left(\frac{1}{2} \|u - x\|^2 + \psi(u) \right) \quad (\text{A.1})$$

Nous notons dans la suite pour tout $x \in \mathbb{R}$:

$$(\forall u \in \mathbb{R}) \quad \Phi(u, x) := \frac{1}{2} \|u - x\|^2 + \psi(u) \quad (\text{A.2})$$

Nous avons proposé dans le Chapitre 3 différentes régularisations hybrides faisant la combinaison entre la fonction d'entropie et une mesure de parcimonie. Les régularisations sont toutes propres et semi-continues inférieurement sur \mathbb{R} . Certaines régularisations sont convexes telles que l'entropie de Shannon combinée à la norme ℓ_1 et l'entropie de Burg combinée à la norme ℓ_1 . Dans ce cas, l'opérateur proximal défini en (A.1) est mono-valué. Autrement dit, pour tout $x \in \mathbb{R}$, $\text{prox}_{\psi}(x)$ prend une valeur unique, égale à l'unique minimiseur de $u \mapsto \Phi(u, x)$.

Cependant, plusieurs régularisations sont non convexes. Pour chaque $x \in \mathbb{R}$, l'opérateur proximal défini par (A.1) fournit un ensemble de valeurs qui n'est pas forcément un singleton. Notons que cet ensemble est nécessairement non vide, car les régularisations considérées sont toutes bornées inférieurement par une fonction affine [Attouch 2011].

Afin de chercher le minimiseur de chaque proposition, nous considérons deux cas de figure :

- ψ est convexe : La définition de l'opérateur proximal implique l'équivalence suivante :

$$p = \text{prox}_{\psi}(x) \Leftrightarrow 0 \in \partial \Phi(p, x) \quad (\text{A.3})$$

— ψ n'est pas convexe : l'équivalence (A.3) n'est plus valide. On a seulement :

$$p = \text{prox}_\psi(x) \implies 0 \in \partial\Phi(p, x) \quad (\text{A.4})$$

En pratique, nous chercherons tous les 0 du sous-gradient de $u \mapsto \Phi(u, x)$, notés \hat{p} , et nous devrons chercher les valeurs minimisant effectivement $\Phi(\hat{p})$.

A.1 Entropie de Shannon

Notons, pour tout $x \in \mathbb{R}$:

$$\psi(x) = \begin{cases} \alpha x \log(x) & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.5})$$

Cette fonction est convexe sur \mathbb{R} . Pour tout $\alpha > 0$, on a :

$$(\forall u \in \mathbb{R}) \quad \Phi(u, x) = \begin{cases} \frac{1}{2}\|u - x\|^2 + \alpha u \log(u) & \text{si } u > 0 \\ \frac{1}{2}x^2 & \text{si } u = 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.6})$$

La fonction (A.6) est égale à $+\infty$ pour les entrées négatives, il suffit de chercher le minimiseur de $\Phi(\cdot, x)$ pour les valeurs $u \geq 0$. La pente infinie de la dérivée de Φ en 0^+ permet d'éliminer la solution $u = 0$. Nous nous limitons donc à $u \in]0, +\infty[$:

$$(\forall u > 0) \quad \nabla\Phi(u, x) = 0 \implies u - x + \alpha \log(u) + \alpha = 0 \quad (\text{A.7})$$

$$\implies \log(u) = -\frac{1}{\alpha}(u - x + \alpha) \quad (\text{A.8})$$

$$\implies u = \exp\left(\frac{-u + x - \alpha}{\alpha}\right) \quad (\text{A.9})$$

$$\implies u = \alpha \mathcal{W}\left(\frac{1}{\alpha} \exp\left(\frac{x}{\alpha} - 1\right)\right) \quad (\text{A.10})$$

Il est facile de vérifier que l'expression en (A.10) est strictement positive. Par conséquent l'opérateur proximal de l'entropie de Shannon est donné par :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_\psi(x) = \alpha \mathcal{W}\left(\frac{1}{\alpha} \exp\left(\frac{x}{\alpha} - 1\right)\right) \quad (\text{A.11})$$

A.2 Entropie de Shannon + ℓ_0

Notons, pour tout $x \in \mathbb{R}$ et $(\alpha, \beta) \in]0, +\infty[^2$:

$$\psi(x) = \begin{cases} \alpha x \log x + \beta & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.12})$$

Remarquons que ψ est convexe sur $]0, +\infty[$. Néanmoins, elle est non convexe sur \mathbb{R} , et discontinue en 0 (à droite et à gauche). De plus, on a pour tout $x \in \mathbb{R}$:

$$(\forall u \in \mathbb{R}) \quad \Phi(u, x) = \begin{cases} \frac{1}{2}\|u - x\|^2 + \alpha u \log(u) + \beta & \text{si } u > 0 \\ \frac{1}{2}x^2 & \text{si } u = 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.13})$$

Du fait de la discontinuité en 0 de (A.13), nous allons chercher le minimiseur de $u \mapsto \Phi(u, x)$ pour $u > 0$. La valeur obtenue, notée p , devra ensuite être comparée à celle de $\Phi(0, x)$, pour pouvoir déterminer la ou les valeurs de l'opérateur proximal souhaité.

$$(\forall u > 0) \quad \nabla \Phi(u, x) = 0 \Rightarrow u - x + \alpha \log(u) + \alpha = 0 \quad (\text{A.14})$$

$$\Rightarrow \log(u) = \frac{1}{\alpha}(-u + x - \alpha) \quad (\text{A.15})$$

$$\Rightarrow u = \exp\left(\frac{-u + x - \alpha}{\alpha}\right) \quad (\text{A.16})$$

$$\Rightarrow u = \alpha \mathcal{W}\left(\frac{1}{\alpha} \exp\left(\frac{x}{\alpha} - 1\right)\right) \quad (\text{A.17})$$

Posons :

$$p = \alpha \mathcal{W}\left(\frac{1}{\alpha} \exp\left(\frac{x}{\alpha} - 1\right)\right) \quad (\text{A.18})$$

On a :

$$\Phi(p, x) = \frac{1}{2}\|p - x\|^2 + \alpha p \log(p) + \beta \quad (\text{A.19})$$

$$= \frac{1}{2}p^2 - px + \frac{1}{2}x^2 + \beta - p^2 + px - \alpha p \quad \text{en utilisant (A.14)} \quad (\text{A.20})$$

$$= \frac{1}{2}x^2 - \frac{1}{2}p^2 + \beta - \alpha p \quad (\text{A.21})$$

$$(\text{A.22})$$

et $\Phi(0, x) = \frac{1}{2}x^2$.

$\Phi(p) - \Phi(0) = -\frac{1}{2}p^2 + \beta - \alpha p$. On remarque que $\Phi(p) > \Phi(0)$ si $\beta > \bar{\beta}$ avec $\bar{\beta} = \frac{1}{2}p^2 + \alpha p$.

Par conséquent l'opérateur proximal de l'entropie de Shannon combinée avec ℓ_0 est :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_\psi(x) = \begin{cases} p & \text{si } \beta < \bar{\beta} \\ \{0, p\} & \text{si } \beta = \bar{\beta} \\ 0 & \text{sinon} \end{cases} \quad (\text{A.23})$$

tel que $\bar{\beta} = \frac{1}{2}p^2 + \alpha p$ et $p = \alpha \mathcal{W}\left(\frac{1}{\alpha} \exp\left(\frac{x}{\alpha} - 1\right)\right)$

A.3 Entropie de Shannon + ℓ_1

Notons, pour tout $x \in \mathbb{R}$, et $(\alpha, \beta) \in]0, +\infty[^2$:

$$(\forall x \in \mathbb{R}) \quad \psi(x) = \begin{cases} \alpha x \log x + \beta x & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.24})$$

Cette fonction est convexe sur \mathbb{R} et pour tout $x \in \mathbb{R}$ on a :

$$(\forall u \in \mathbb{R}) \quad \Phi(u, x) = \begin{cases} \frac{1}{2}\|u - x\|^2 + \alpha u \log(u) + \beta u & \text{si } u > 0 \\ \frac{1}{2}x^2 & \text{si } u = 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.25})$$

Comme dans le cas de la Section A.1, nous pouvons nous limiter à l'étude de $u \mapsto \Phi(u, x)$ sur $]0, +\infty[$:

$$(\forall u > 0) \quad \nabla \Phi(u, x) = 0 \Rightarrow u - x + \alpha \log(u) + \alpha + \beta = 0 \quad (\text{A.26})$$

$$\Rightarrow \log(u) = -\frac{1}{\alpha}(u - x + \alpha + \beta) \quad (\text{A.27})$$

$$\Rightarrow u = \exp\left(\frac{-u + x - \alpha - \beta}{\alpha}\right) \quad (\text{A.28})$$

$$\Rightarrow u = \alpha \mathcal{W}\left(\frac{1}{\alpha} \exp\left(\frac{x - \beta}{\alpha} - 1\right)\right) \quad (\text{A.29})$$

D'où la formule de l'opérateur proximal de l'entropie de Shannon combinée avec la norme ℓ_1 sous l'hypothèse $(\alpha, \beta) \in]0, +\infty[^2$:

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{\psi}(x) = \alpha \mathcal{W}\left(\frac{1}{\alpha} \exp\left(\frac{x - \beta}{\alpha} - 1\right)\right) \quad (\text{A.30})$$

A.4 Entropie de Shannon + $\mathbf{f}_{\log\text{-sum}}$

Notons, pour tout $x \in \mathbb{R}$ et $(\alpha, \beta) \in]0, +\infty[^2$:

$$\psi(x) = \begin{cases} \alpha x \log x + \beta \log(\delta + x) & \text{si } x > 0 \\ \beta \log(\delta) & \text{si } x = 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.31})$$

Cette fonction est différentiable sur $]0, +\infty[$. Elle est continue en 0 (à droite), et non convexe sur \mathbb{R} . Dans ce cas, on a pour tout $x \in \mathbb{R}$:

$$(\forall u \in \mathbb{R}) \quad \Phi(u, x) = \begin{cases} \frac{1}{2}\|u - x\|^2 + \alpha u \log(u) + \beta \log(\delta + u) & \text{si } u > 0 \\ \frac{1}{2}x^2 + \beta \log(\delta) & \text{si } u = 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.32})$$

La valeur infinie de Φ sur l'orthant négatif, la pente infinie de la dérivée de $\Phi(\cdot, x)$ en 0, et sa continuité en 0^+ , nous assurent que son ou ses minimiseurs ne peuvent être que strictement positifs. De plus :

$$(\forall u > 0) \quad \nabla \Phi(u, x) = 0 \Rightarrow u - x + \alpha \log(u) + \alpha + \frac{\beta}{\delta + u} = 0 \quad (\text{A.33})$$

$$\Rightarrow u^2 + (\delta - x + \alpha)u + \alpha(\delta + u) \log(u) + \beta - \delta x + \delta \alpha = 0 \quad (\text{A.34})$$

$$\Rightarrow u \text{ est une solution de } \varphi(p) = 0 \text{ où :} \quad (\text{A.35})$$

$$\varphi(p) = p^2 + (\delta - x + \alpha)p + \alpha(\delta + u) \log(p) + \beta + \delta(\alpha - x) \quad (\text{A.36})$$

Par conséquent :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{\psi}(x) = \underset{p \in]0, +\infty[\text{ tel que } \varphi(p)=0}{\text{Argmin}} \left(\frac{1}{2}(x - p)^2 + \psi(p) \right) \quad (\text{A.37})$$

Remarquons qu'il est nécessaire dans ce cas de chercher le ou les zéros de la fonction $\varphi(p)$ (A.36) et de chercher manuellement le ou lesquels minimisent $\Phi(\cdot, x)$.

A.5 Entropie de Shannon + f_{Cauchy}

Notons, pour tout $x \in \mathbb{R}$ et $(\alpha, \beta, \delta) \in]0, +\infty[^3$:

$$\psi(x) = \begin{cases} \alpha x \log x + \beta \log(\delta + x^2) & \text{si } x > 0 \\ \beta \log(\delta) & \text{si } x = 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.38})$$

Comme dans la Section A.4, la fonction est non convexe sur \mathbb{R} . Elle est différentiable sur $]0, +\infty[$, et continue (à droite) en 0. De plus, pour tout $x \in \mathbb{R}$:

$$(\forall u \in \mathbb{R}) \quad \Phi(u, x) = \begin{cases} \frac{1}{2}\|u - x\|^2 + \alpha u \log(u) + \beta \log(\delta + u^2) & \text{si } u > 0 \\ \frac{1}{2}x^2 + \beta \log(\delta) & \text{si } u = 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.39})$$

Suivant les mêmes arguments que la Section A.4, nous étudions uniquement $\Phi(\cdot, x)$ sur $]0, +\infty[$:

$$(\forall u > 0) \quad \nabla \Phi(u, x) = 0 \Rightarrow u - x + \alpha \log(u) + \alpha + \frac{2u\beta}{\delta + u^2} = 0 \quad (\text{A.40})$$

$$\Rightarrow u^3 + (\alpha - x)u^2 + (2\beta + \delta)u + \alpha(\delta + u^2) \log(u) - \delta x + \delta \alpha = 0 \quad (\text{A.41})$$

$$\Rightarrow u \text{ est une solution de } \varphi(p) = 0 \text{ où :} \quad (\text{A.42})$$

$$\varphi(p) = p^3 + (\alpha - x)p^2 + (2\beta + \delta)p + \alpha(\delta + p^2) \log(p) + \delta(\alpha - x) \quad (\text{A.43})$$

Par conséquent :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_{\psi}(x) = \underset{p \in]0, +\infty[\text{ tel que } \varphi(p)=0}{\text{Argmin}} \left(\frac{1}{2}(x - p)^2 + \psi(p) \right) \quad (\text{A.44})$$

Il est nécessaire de chercher le ou les zéros de la fonction $\varphi(p)$ (A.43) et de chercher manuellement le ou lesquels minimisent $\Phi(\cdot, x)$.

A.6 Entropie de Burg

Pour tout $x \in \mathbb{R}$ et $\alpha > 0$ on a :

$$\psi(x) = \begin{cases} -\alpha \log(x) & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.45})$$

Cette fonction est convexe sur \mathbb{R} . De plus, pour tout $x \in \mathbb{R}$:

$$(\forall u \in \mathbb{R}) \quad \Phi(u, x) = \begin{cases} \frac{1}{2}\|u - x\|^2 - \alpha \log u & \text{si } u > 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.46})$$

Il est clair que le minimiseur de Φ est nécessairement strictement positif. Pour tout $x \in \mathbb{R}$, on a :

$$(\forall u > 0) \quad \nabla \Phi(u, x) = 0 \Rightarrow u - x - \frac{\alpha}{u} = 0 \quad (\text{A.47})$$

$$\Rightarrow u^2 - xu - \alpha = 0 \quad (\text{A.48})$$

$$\Rightarrow u = \pm \frac{x + \sqrt{x^2 + 4\alpha}}{2} \quad (\text{A.49})$$

Du fait de la contrainte de positivité, on déduit :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_\psi(x) = \frac{x + \sqrt{x^2 + 4\alpha}}{2} \quad (\text{A.50})$$

A.7 Entropie de Burg + ℓ_0

Pour tout $x \in \mathbb{R}$ et $(\alpha, \beta) \in]0, +\infty[^2$ on note :

$$\psi(x) = \begin{cases} -\alpha \log x + \beta & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.51})$$

Cette fonction est convexe sur $]0, +\infty[$. De plus, pour tout $x \in \mathbb{R}$:

$$(\forall u \in \mathbb{R}) \quad \Phi(u, x) = \begin{cases} \frac{1}{2}\|u - x\|^2 - \alpha \log u + \beta & \text{si } u > 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.52})$$

A nouveau, il est clair que le minimiseur de Φ est nécessairement strictement positif. Pour tout $x \in \mathbb{R}$, on a :

$$(\forall u > 0) \quad \nabla \Phi(u, x) = 0 \Rightarrow u - x - \frac{\alpha}{u} = 0 \quad (\text{A.53})$$

$$\Rightarrow u^2 - xu - \alpha = 0 \quad (\text{A.54})$$

$$\Rightarrow u = \pm \frac{x + \sqrt{x^2 + 4\alpha}}{2} \quad (\text{A.55})$$

En utilisant la contrainte de positivité, on déduit :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_\psi(x) = \frac{x + \sqrt{x^2 + 4\alpha}}{2} \quad (\text{A.56})$$

A.8 Entropie de Burg + ℓ_1

Pour tout $x \in \mathbb{R}$, et $(\alpha, \beta) \in]0, +\infty[^2$, notons :

$$\psi(x) = \begin{cases} -\alpha \log x + \beta x & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.57})$$

Cette fonction est convexe sur \mathbb{R} . De plus, la fonction Φ dans ce cas est donnée pour tout $x \in \mathbb{R}$ par :

$$(\forall u \in \mathbb{R}) \quad \Phi(u, x) = \begin{cases} \frac{1}{2}\|u - x\|^2 - \alpha \log u + \beta u & \text{si } u > 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.58})$$

Le minimiseur est nécessairement strictement positif. Pour tout $x \in \mathbb{R}$:

$$(\forall u > 0) \quad \nabla \Phi(u, x) = 0 \Rightarrow u - x - \frac{\alpha}{u} + \beta = 0 \quad (\text{A.59})$$

$$\Rightarrow u^2 - (x - \beta)u - \alpha = 0 \quad (\text{A.60})$$

$$\Rightarrow u = \pm \frac{x - \beta + \sqrt{(x - \beta)^2 + 4\alpha}}{2} \quad (\text{A.61})$$

A nouveau, sous la contrainte de positivité, nous déduisons :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_\psi(x) = \frac{x - \beta + \sqrt{(x - \beta)^2 + 4\alpha}}{2} \quad (\text{A.62})$$

A.9 Entropie de Burg + $\mathbf{f}_{\log\text{-sum}}$

Notons, pour tout $x \in \mathbb{R}$, $(\alpha, \beta, \delta) \in]0, +\infty[^3$:

$$\psi(x) = \begin{cases} -\alpha \log x + \beta \log(\delta + x) & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.63})$$

Cette fonction est non convexe sur \mathbb{R} . De plus pour tout $x \in \mathbb{R}$:

$$(\forall u \in \mathbb{R}) \quad \Phi(u, x) = \begin{cases} \frac{1}{2}\|u - x\|^2 - \alpha \log u + \beta \log(u + \delta) & \text{si } u > 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.64})$$

Cette fonction peut avoir un ou plusieurs minimiseurs. Néanmoins, on est assuré qu'ils appartiennent à $]0, +\infty[$. De plus, pour tout $x \in \mathbb{R}$:

$$(\forall u > 0) \quad \nabla \Phi(u, x) = 0 \Rightarrow u - x - \frac{\alpha}{u} + \frac{\beta}{u + \delta} = 0 \quad (\text{A.65})$$

$$\Rightarrow u^2 \delta + u^3 - x \delta u - x u^2 + \beta u - \alpha(\delta + u) = 0 \quad (\text{A.66})$$

$$\Rightarrow u^3 + (\delta - x)u^2 + (\beta - \alpha - \delta x)u - \alpha \delta = 0 \quad (\text{A.67})$$

$$\Rightarrow u \text{ est une solution de } \varphi(p) = 0 \text{ où :} \quad (\text{A.68})$$

$$\varphi(p) = p^3 + (\delta - x)p^2 + (\beta - \alpha - \delta x)p - \alpha \delta \quad (\text{A.69})$$

Par conséquent :

$$\text{prox}_\psi(x) = \underset{p \in]0, +\infty[\text{ tel que } \varphi(p) = 0}{\text{Argmin}} \left(\frac{1}{2}(x - p)^2 + \psi(p) \right) \quad (\text{A.70})$$

Il est nécessaire de chercher le ou les zéros de la fonction $\varphi(p)$ (A.69) et de chercher manuellement le ou lesquels minimisent $\Phi(\cdot, x)$.

A.10 Entropie de Burg + f_{Cauchy}

La dernière proposition fait la combinaison entre l'entropie de Burg et la fonction log-sum, ce qui donne pour tout $x \in \mathbb{R}$ et $(\alpha, \beta, \delta) \in]0, +\infty[^3$:

$$\psi(x) = \begin{cases} -\alpha \log x + \beta \log(\delta + x^2) & \text{si } x > 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.71})$$

Cette fonction est non convexe sur \mathbb{R} . De plus, pour tout $x \in \mathbb{R}$, on a :

$$(\forall u \in \mathbb{R}) \quad \Phi(u, x) = \begin{cases} \frac{1}{2}\|u - x\|^2 - \alpha \log u + \beta \log(u^2 + \delta) & \text{si } u > 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{A.72})$$

Pour tout $x \in \mathbb{R}$, il suffit de chercher le ou les minimiseurs de $\Phi(\cdot, x)$ sur $]0, +\infty[$:

$$(\forall u > 0) \quad \nabla \Phi(u, x) = 0 \Rightarrow u - x - \frac{\alpha}{u} + \frac{2\beta u}{\delta + u^2} = 0 \quad (\text{A.73})$$

$$\Rightarrow \delta u^2 + u^4 - x \delta u - x u^3 + 2\beta u^2 - \alpha \delta - \alpha u^2 = 0 \quad (\text{A.74})$$

$$\Rightarrow u^4 - x u^3 + (\delta + 2\beta - \alpha)u^2 - x \delta u - \alpha \delta = 0 \quad (\text{A.75})$$

$$\Rightarrow u \text{ est une solution de } \varphi(p) = 0 \text{ où :} \quad (\text{A.76})$$

$$\varphi(p) = p^4 - x p^3 + (\delta + 2\beta - \alpha)p^2 - x \delta p - \alpha \delta \quad (\text{A.77})$$

Par conséquent :

$$\text{prox}_\psi(x) = \underset{p \in]0, +\infty[\text{ tel que } \varphi(p) = 0}{\text{Argmin}} \left(\frac{1}{2}(x - p)^2 + \psi(p) \right) \quad (\text{A.78})$$

Il est nécessaire dans ce cas de chercher le ou les zéros de la fonction $\varphi(p)$ (A.77) et de chercher manuellement le ou lesquels minimisent $\Phi(\cdot, x)$.

Fonction W de Lambert

B.1 Définition

La fonction \mathcal{W} de Lambert que nous avons définie en (3.23), proposée par Jean-Henri Lambert [Gray 1978], peut être généralisée par :

$$z = wa^w \iff w = \frac{\mathcal{W}(z \log(a))}{\log(a)} \quad (\text{B.1})$$

Cette fonction est multi-valuée et elle prend deux branches pour les valeurs réelles $x \geq -\frac{1}{e}$: une branche principale notée \mathcal{W}_0 définie pour $x \geq -\frac{1}{e}$ et $\mathcal{W}(x) > -1$ et une branche secondaire notée \mathcal{W}_{-1} définie pour $x < 0$ et $\mathcal{W}(x) < -1$. Nous avons présenté cette fonction dans la Figure 3.6.

B.2 Utilisation pratique

Commençons par un exemple :

$$2^t = 5t \quad (\text{B.2})$$

$$\implies 1 = \frac{5t}{2^t} \quad (\text{B.3})$$

$$\implies 1 = 5t \exp(-t \log(2)) \quad (\text{B.4})$$

$$\implies \frac{1}{5} = t \exp(-t \log(2)) \quad (\text{B.5})$$

$$\implies \frac{-\log(2)}{5} = (-t \log(2)) \exp(-t \log(2)) \quad (\text{B.6})$$

$$\implies \mathcal{W}\left(\frac{-\log(2)}{5}\right) = -t \log(2) \quad (\text{B.7})$$

$$\implies t = -\frac{\mathcal{W}\left(\frac{-\log(2)}{5}\right)}{\log(2)} \quad (\text{B.8})$$

$$(\text{B.9})$$

En général, l'équation :

$$p^x = ax + b \quad (\text{B.10})$$

pour $p > 0$ et $p \neq 1$ et $a \neq 0$ peut être reformulée avec la variable :

$$-t = x + \frac{b}{a} \quad (\text{B.11})$$

par :

$$tp^t = -\frac{1}{a}p^{\frac{-b}{a}} = R \quad (\text{B.12})$$

ce qui donne :

$$t = \frac{\mathcal{W}(R \log(p))}{\log(p)} \quad (\text{B.13})$$

La solution finale sera donnée par :

$$x = \frac{\mathcal{W}\left(-\frac{\log(p)}{a}p^{\frac{-b}{a}}\right)}{\log(p)} - \frac{b}{a} \quad (\text{B.14})$$

Annexe C

Calcul de l'opérateur proximal de PALMA

C.1 Opérateur proximal : $\alpha \text{Ent} + (1 - \alpha)\ell_1$

Nous utilisons la même notation de l'Annexe A pour calculer la fonction Φ (A.2). L'entropie de Shannon combinée avec la norme ℓ_1 avec un compromis $\alpha + \beta = 1$ correspond à $(\forall x \in \mathbb{R})$:

$$\psi(x) = \begin{cases} \alpha \frac{x}{a} \log\left(\frac{x}{a}\right) + (1 - \alpha)x & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{C.1})$$

Cette fonction est convexe sur \mathbb{R} . De plus, pour tout $x \in \mathbb{R}$ on a :

$$(\forall u \in \mathbb{R}) \quad \Phi(u, x) = \begin{cases} \frac{1}{2}\|u - x\|^2 + \alpha \frac{u}{a} \log\left(\frac{u}{a}\right) + \beta u & \text{si } u > 0 \\ \frac{1}{2}x^2 & \text{si } u = 0 \\ +\infty & \text{sinon} \end{cases} \quad (\text{C.2})$$

• Si $\alpha \in]0, 1]$, la solution est nécessairement dans $]0, +\infty[$. On a pour tout $x \in \mathbb{R}$:

$$(\forall u > 0) \quad \nabla \Phi(u, x) = 0 \implies u - x_n + \frac{\alpha}{a} \log(u) + \frac{\alpha}{a} - \frac{\alpha}{a} \log(a) + 1 - \alpha = 0 \quad (\text{C.3})$$

$$\implies \log(u) = \frac{a}{\alpha} \left(-u + x - \frac{\alpha}{a} + \frac{\alpha}{a} \log(a) - 1 + \alpha \right) \quad (\text{C.4})$$

$$\implies u = \exp \left(-\frac{a}{\alpha} u + \frac{ax - a(1 - \alpha)}{\alpha} + \log(a) - 1 \right) \quad (\text{C.5})$$

En utilisant (B.2), la solution sera donnée par :

$$u = \frac{\alpha}{a} \mathcal{W} \left(\frac{a}{\alpha} \exp \left(\frac{ax - a(1 - \alpha)}{\alpha} + \log(a) - 1 \right) \right) \quad (\text{C.6})$$

- Si $\alpha = 0$, pour tout $x \in \mathbb{R}$:

$$x - u \in \partial|u| \implies x - u \in \begin{cases} \text{sign}(u) & \text{si } u \neq 0 \\ [-1, 1] & \text{sinon} \end{cases}$$

$$\implies u = \begin{cases} x - 1 & \text{si } x > 1 \\ 0 & \text{si } x \in [-1, 1] \\ x + 1 & \text{sinon} \end{cases}$$

L'expression de l'opérateur proximal sera donc :

$$(\forall x \in \mathbb{R}) \quad \text{prox}_\psi(x) = \begin{cases} \frac{\alpha}{a} \mathcal{W} \left[\frac{a}{\alpha} \exp\left(\frac{ax - a(1-\alpha)}{\alpha} + \log(a) - 1\right) \right] & \text{si } \alpha \in]0, 1[\\ \text{sign}(x) \max(|x| - 1, 0) & \text{si } \alpha = 0 \end{cases} \quad (\text{C.7})$$

C.2 Mesures

Nous présentons dans la Figure C.1 un zoom sur une partie des mesures des signaux A, B, C et D présentés dans le Chapitre 4 pour montrer la présence du bruit :

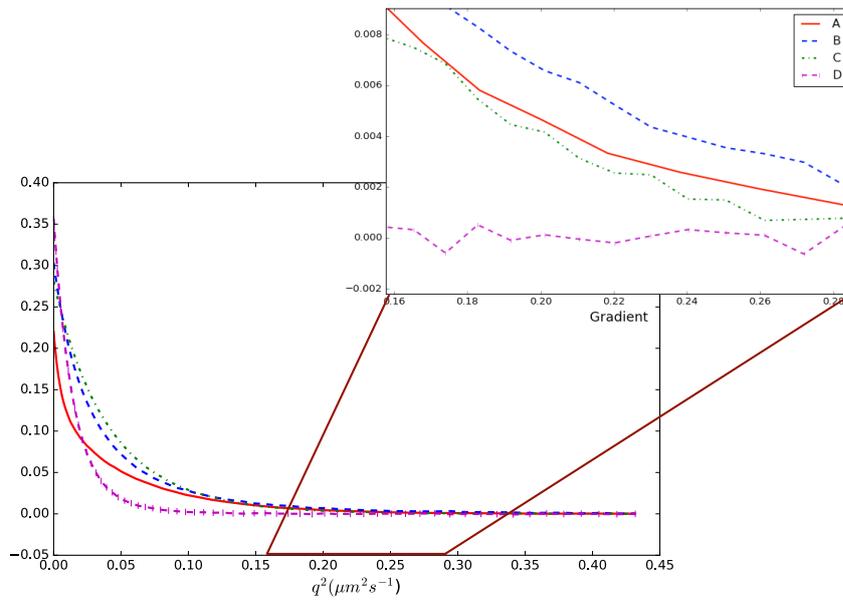


FIGURE C.1 – Mise en évidence de la présence du bruit

Le traçage en forme logarithmique montre la présence du bruit :

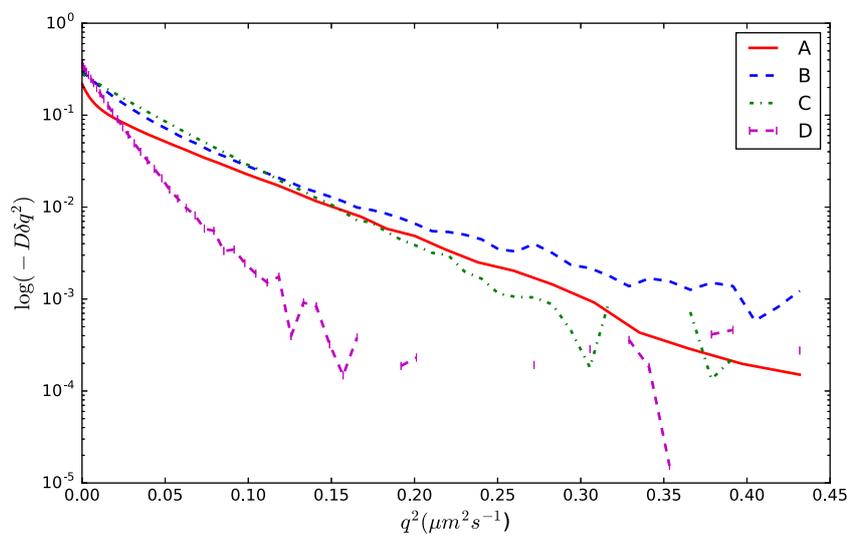


FIGURE C.2 – Présentation en échelle logarithmique des mesures (C.1)

Annexe D

Serveur Web PALMA

Nous présentons dans cette partie du manuscrit l'interface de communication entre le serveur PALMA et l'utilisateur :

- "Home" :

Theory

A polydisperse sample has to be analysed with a distribution $X(D)$ of diffusion coefficients:

$$I(q) = \int_{D_{\min}}^{D_{\max}} X(D) \exp(-D\Delta q^2) dD$$

Determining the distribution $X(D)$ from $I(q)$ requires to solve the Laplace inversion of the q^2 dependence of $I(q)$. Given that the experiment was performed over a series of M values of q and measured as a series of intensities y_m for one given chemical shift value. The problem stated by the equation above can be discretised:

$$y_m = \sum_{n=1}^N x_n \exp(-D_n \Delta q_m^2)$$

with D_n sampling the D_{\min} to D_{\max} axis in an exponential manner.

As this expression is linear in x_n , it can be rewritten as follows:

$$Y = \mathbf{H}X$$

where $Y = \{y_m, 1 \leq m \leq M\}$ is the experimental series, $X = \{x_n, 1 \leq n \leq N\}$ is the reconstructed distribution, and \mathbf{H} is a $M \times N$ matrix with values $H_{m,n} = \exp(-D_n \Delta q_m^2)$.

A general approach to solving this linear equation is to generate a solution \hat{X} which solves the following constrained optimisation problem:

$$\text{minimize}_{X \in \mathbb{R}^N} \Psi(X) \quad \text{subject to} \quad \|\mathbf{H}X - Y\| \leq \eta$$

where η is an estimate of the expected quality of the fit, based on an estimate of the experimental noise.

In order to favour both smooth and sparse shapes in the estimated signal, we use a regularisation defined as follows:

$$\Psi(X) = \lambda \text{ent}(X, a) + (1 - \lambda) \ell_1(X)$$

where $\text{ent}(X, a)$ is given by

$$\text{ent}(X, a) = \begin{cases} \sum_{n=1}^N \frac{x_n}{a} \log\left(\frac{x_n}{a}\right) & \text{if } x_n > 0 \\ 0 & \text{if } x_n = 0 \\ +\infty & \text{elsewhere,} \end{cases}$$

corresponds to the MaxEnt regularisation with a flat prior $a > 0$.

$\ell_1(X)$ is the ℓ_1 norm of the vector X :

FIGURE D.1 – Page d'accueil

• "Data analysis" :

Submit form - DOSY analysis x

palma.labo.igbmc.fr/form.html

PALMA for DOSY Analysis

igbmc
Institut de génétique et de
biologie moléculaire et cellulaire

Home Data analysis Methods How-To Contact us

This program will only work for experiment acquired on a Bruker instrument running Topspin (tested for 2.0 to 3.2)

- Acquire your DOSY as usual, using a sequence from the standard library (check the [How-To](#) for hints on acquisition).
- set the processing parameters for the F2 (classic) dimension to your taste. (*favor SNR to resolution, and avoid having too many points in F2*)
- Process it in the F2 dimension with the `xF2` command
- create a zipfile containing the whole experiment at the expno level, that is, if your experiment is `data_dir/user_name/nmr/project_name/expno/pdata/procno` zip the whole expno folder (*remarks: -there should be only one procno in the experiment; - only the `2rx`, `diff1ist` and `parameter` files are used, the `ser` and `2ii` files are not required and you may remove them before zipping if you wish*)

Remarks

- Some Diffusion coefficients typical values:
 - solvents are usually within 2000 (water) to 5000 $\mu\text{m}^2/\text{s}$
 - small organic molecules are usually within 200 to 1000 $\mu\text{m}^2/\text{s}$
 - proteins are usually within 50 (globular proteins) to 200 (peptides) $\mu\text{m}^2/\text{s}$
 - polymers can present wide diffusion distributions...
- The program only handles ^1H data-set for the moment (*we're working on it*)
- No need to use a lot of point in the F2 horizontal dimension. 4k-8k points are usually enough, and make nicer spectra than a larger size (*and faster too!*).
- The program should give reliable diffusion coefficients to the condition that your spectrometer was correctly calibrated (סטנדרט) and if you are using a pulse sequence from the standard library.

Your data

Email address

Your results will be sent to this address, usually in less than an hour.

Optional comment

Data file input (.zip)

Choisir un fichier Aucun fichier choisi

Algorithm parameters

Choose Diffusion limits and robustness

Default values lambda=0.05 Robust conditions - good for mixtures

FIGURE D.2 – Demande d'analyse des données DOSY

• "Methods" :

PALMA for DOSY Analysis

Home Data analysis Methods How-To Contact us

Theory

A polydisperse sample has to be analysed with a distribution $X(D)$ of diffusion coefficients:

$$I(q) = \int_{D_{\min}}^{D_{\max}} X(D) \exp(-D\Delta q^2) dD$$

Determining the distribution $X(D)$ from $I(q)$ requires to solve the Laplace inversion of the q^2 dependence of $I(q)$. Given that the experiment was performed over a series of M values of q and measured as a series of intensities y_m for one given chemical shift value. The problem stated by the equation above can be discretised:

$$y_m = \sum_{n=1}^N x_n \exp(-D_n \Delta q_m^2)$$

with D_n sampling the D_{\min} to D_{\max} axis in an exponential manner.

As this expression is linear in x_n , it can be rewritten as follows:

$$Y = \mathbf{H}X$$

where $Y = \{y_m, 1 \leq m \leq M\}$ is the experimental series, $X = \{x_n, 1 \leq n \leq N\}$ is the reconstructed distribution, and \mathbf{H} is a $M \times N$ matrix with values $\mathbf{H}_{m,n} = \exp(-D_n \Delta q_m^2)$.

A general approach to solving this linear equation is to generate a solution \hat{X} which solves the following constrained optimisation problem:

$$\text{minimize}_{X \in \mathbb{R}^N} \Psi(X) \quad \text{subject to} \quad \|\mathbf{H}X - Y\| \leq \eta$$

where η is an estimate of the expected quality of the fit, based on an estimate of the experimental noise.

In order to favour both smooth and sparse shapes in the estimated signal, we use a regularisation defined as follows:

$$\Psi(X) = \lambda \text{ent}(X, a) + (1 - \lambda) \ell_1(X)$$

where $\text{ent}(X, a)$ is given by

$$\text{ent}(X, a) = \begin{cases} \sum_{n=1}^N \frac{x_n}{a} \log\left(\frac{x_n}{a}\right) & \text{if } x_n > 0 \\ 0 & \text{if } x_n = 0 \\ +\infty & \text{elsewhere,} \end{cases}$$

corresponds to the MaxEnt regularisation with a flat prior $a > 0$.

$\ell_1(X)$ is the ℓ_1 norm of the vector X :

FIGURE D.3 – Présentation illustrée de la théorie derrière PALMA

• "How to" :

PALMA for DOSY Analysis

igbmc
Institut de génétique et de
biologie moléculaire et cellulaire

Home Data analysis Methods How-To Contact us

How to measure the best possible DOSY experiment...

You know that the DOSY experiment is not an easy one, it may fail in a number of manner. We provide here a short check-list covering the most typical problems that you may face

This involves three stages

- 1/ ACQUIRE
- 2/ PROCESS
- 3/ ANALYSE
- IN CASE OF PROBLEMS

1/ Acquire the best possible DOSY experiment...

First Convection is your enemy ! you're always better in D₂O and DMSO than in CDCl₃ (water is more viscous than chloroform !), you prefer 3mm tubes than 5mm, you prefer room temperature than anything else.

Also check the temperature control !

- eventually increase the gaz-flow to be sure that the temperature is homogeneous in the tube
- display the $\Sigma T E$ panel, and go to monitoring. If you observe oscillation there, you're off. Re-tune the PDI of the temperature control in your current conditions.

Then insure that all the parameters have been set correctly, typically the following ones should be checked before *each* experiments:

- the P90 should have been carefully calibrated.
- use a reasonable gradient list, typically not less than 20-24 points, and not more 64 points as it is probably useless. Ranging from 1% to 99% is Ok on the modern probes and the modern PFG systems.
- use a correct pulse sequence, `stexxx` and `1edxxx` are usually good choices. Use `dstebpxxx` if you suspect convection.
- calibrate the Δ (diffusion delay) and δ (PFG duration) parameters, so that, when comparing the first and the last PFG increment, the slowest decaying signal is left in the final experiment to a few % of its initial intensity. I usually do this by running a 2 or 3 points DOSY.
- Should should have a descent SNR. PALMA is able to extract information down to about a Signal-to-Noise Ratio of x20; however in the current version, this is limited to x32. So signals lower than that will not be analyzed.
- the gain `RG`, should not be set too high as non-linearity in the acquisition chain may perturb the measure. The rule of thumb I advise is the following : in the block where the DOSY acquisition is going to take place, run the `RG` command, then divide the obtained `RG` value by 2, just to be sure.
- use a large enough gradient recovery delay (usually `D16`). Increase it if you observe phase distortion on the right hand of the sharp peaks.

FIGURE D.4 – Explications du mode d'utilisation du serveur

• "Contact us" :

PALMA for DOSY Analysis

igbmc
Institut de génétique et de
biologie moléculaire et cellulaire

Home Data analysis Methods How-To Contact us

Contact information

For any information please contact Marc-André Delsuc at the following address:

- Email : madelsuc@unistra.fr
- Address : 1 rue Laurent Fries BP 10142 67404 Illkirch CEDEX, France

For a bug report, please join the **JobId** in which the problem was encountered (*given in the web page when submitting the process*). If a jobid was not generated, or if you lost it, provide us with the date and time at which the problem was encountered.

Made with [Bootstrap 3](#) and [Flask](#) microframework.

FIGURE D.5 – Page de contacts

Annexe **E**

Calcul de l'opérateur proximal

Rappelons que :

$$\tilde{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{x} - \tilde{\mathbf{x}}\|_{\mathbf{B}}^2 + \iota_{[0, +\infty[^N}(\mathbf{x}) + \iota_{\|\cdot - y\| \leq \eta}(\mathbf{D}\mathbf{x}) \quad (\text{E.1})$$

E.1 Calcul de l'opérateur proximal avec l'algorithme PPXA+

Algorithme 14 Calcul de l'opérateur proximal avec PPXA+

$$\left[\begin{array}{l} \mathbf{x}_{1,0} \in \mathbb{R}^N \\ \mathbf{x}_{2,0} \in \mathbb{R}^M \\ \mathbf{z}_0 = (\mathbf{I} + \mathbf{D}^\top \mathbf{D})^{-1} (\mathbf{x}_{1,0} + \mathbf{D}^\top \mathbf{x}_{2,0}) \\ \theta \geq 0 \\ \text{For } j = 0, 1, \dots \\ \left[\begin{array}{l} \mathbf{y}_{1,j} = \operatorname{prox}_{\iota_{[0, +\infty[^N + \theta/2\|\cdot - \tilde{\mathbf{x}}\|_{\mathbf{B}}^2}}(\mathbf{x}_{1,j}) \\ \mathbf{y}_{2,j} = \mathcal{P}_{\|\cdot - y\| \leq \eta}(\mathbf{x}_{2,j}) \\ \mathbf{v}_i = (\mathbf{I} + \mathbf{D}^\top \mathbf{D})^{-1} (\mathbf{y}_{1,j} + \mathbf{D}^\top \mathbf{y}_{2,j}) \\ \mathbf{x}_{1,j+1} = \mathbf{x}_{1,j} + 2\mathbf{v}_i - \mathbf{z}_i - \mathbf{y}_{1,j} \\ \mathbf{x}_{2,j+1} = \mathbf{x}_{2,j} + \mathbf{D}(2\mathbf{v}_i - \mathbf{z}_i) - \mathbf{y}_{2,j} \\ \mathbf{z}_{i+1} = \mathbf{v}_i \end{array} \right. \end{array} \right.$$

Comme la métrique \mathbf{B} est diagonale à entrées strictement positives, on a :

$$\operatorname{prox}_{\iota_{[0, +\infty[^N + \theta/2\|\cdot - \tilde{\mathbf{x}}\|_{\mathbf{B}}^2}}(\mathbf{x}) = (p_n)_{1 \leq n \leq N}, \quad (\text{E.2})$$

avec pour tout $n \in \{1, \dots, N\}$:

$$p_n = \mathcal{P}_{[0, +\infty[} \left(\frac{x_n + \theta b_n \tilde{x}_n}{1 + \theta b_n} \right) \quad (\text{E.3})$$

où on note b_n le n -ième terme de la diagonale de \mathbf{B} .

L'opérateur proximal est obtenu comme limite de la suite $(\mathbf{z}_i)_{i \in \mathbb{N}}$.

E.2 Calcul de l'opérateur proximal avec la version accélérée de Chambolle-Pock

Algorithme 15 Calcul de l'opérateur proximal avec la version accélérée de Chambolle-Pock

$$\begin{array}{l}
 \mathbf{x}_0 \in \mathbb{R}^N, \bar{\mathbf{x}}_0 = \mathbf{x}_0 \\
 (\tau_0, \sigma_0) \text{ tel que } \tau_0 \sigma_0 \|\mathbf{D}\|^2 \leq 1 \\
 \gamma = \min(b_n)_{1 \leq n \leq N} \\
 \text{Pour } j = 0, 1, \dots \\
 \left[\begin{array}{l}
 \tilde{\mathbf{y}}_{j+1} = \mathcal{P}_{\|\cdot - \mathbf{y}\| \leq \eta}(\sigma_j^{-1} \mathbf{y}_j + \mathbf{D} \bar{\mathbf{x}}_j) \\
 \mathbf{y}_{j+1} = \mathbf{y}_j + \sigma_j (\mathbf{D} \bar{\mathbf{x}}_j - \tilde{\mathbf{y}}_{j+1}) \\
 \mathbf{x}_{j+1} = \text{prox}_{\iota_{[0, +\infty[}^N + \tau_j/2 \|\cdot - \bar{\mathbf{x}}\|_{\mathbb{B}}^2}(\mathbf{x}_j) - \tau_j \mathbf{D}^\top \mathbf{y}_{j+1}) \\
 \theta_j = 1/\sqrt{1 + 2\gamma\tau_j} \\
 \tau_{j+1} = \theta_j \tau_j \\
 \sigma_{(j+1)} = \sigma_j / \theta_j \\
 \bar{\mathbf{x}}_{j+1} = \mathbf{x}_{j+1} + \theta_j (\mathbf{x}_{j+1} - \mathbf{x}_j)
 \end{array} \right.
 \end{array}$$

L'opérateur proximal est obtenu comme limite de la suite $(\mathbf{x}_j)_{j \in \mathbb{N}}$.

Annexe **F**

Publications

PROXIMITY OPERATORS FOR A CLASS OF HYBRID SPARSITY + ENTROPY PRIORS. APPLICATION TO DOSY NMR SIGNAL RECONSTRUCTION.

Cherni Afef^{1,2}, Chouzenoux Émilie¹, Delsuc Marc-André²

¹ Université Paris-Est, LIGM (UMR 8049), CNRS, ENPC, ESIEE Paris, UPEM, Marne-la-Vallée, France.

² Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U596, CNRS UMR 7104
Université de Strasbourg, 67404 Illkirch-Graffenstaden, France

ABSTRACT

Inverse problems arising from Laplace transform inversion are ill-posed, and require suitable regularization strategies. Although the maximum entropy regularization approach usually appears as an adequate strategy due to its ability to recover regular positive valued signals, it was observed to lead to poor reconstruction results when the sought signal contains narrow peaks. In that case, a sparsity promoting penalty such as the ℓ_1 norm, combined with a positivity constraint, is more suitable. In order to derive a flexible resolution method, hybrid approaches combining both entropy and sparsity regularization strategies should be envisaged. However, the choice of an efficient optimization algorithm remains a challenging task. Among available optimization techniques, proximal methods have shown their efficiency in solving large scale possibly nonsmooth problems. This paper provides an extensive list of new proximity operators for the sum of entropy and sparsity penalties. The applicability of these results is illustrated by means of experiments, in the context of DOSY NMR signal reconstruction.

Index Terms— Inverse problems, Laplace inversion, entropy regularization, sparsity prior, proximity operator, spectroscopy, nuclear magnetic resonance.

1. INTRODUCTION

The fast resolution of large size ill-posed linear inverse problems presents a big challenge in the context of biophysical data processing [1]. A difficult inverse problem, arising for instance in the context of diffusion ordered nuclear magnetic resonance spectroscopy (DOSY NMR) [2], consists in the estimation of an original unknown positive-valued signal $\bar{x} = (x_n)_{1 \leq n \leq N} \in [0, +\infty[^N$, from degraded measurements $y = (y_m)_{1 \leq m \leq M} \in \mathbb{R}^M$ related to \bar{x} through a linear model

$$y = K\bar{x} + w, \quad (1)$$

This work was supported by the CNRS MASTODONS project under grant 2016TABASCO and by the Agence Nationale pour la Recherche (ANR, France) under grant 2010FT-ICR2D and grant Défi de tous savoirs 2014, ONE-SHOT-FT-ICR-MS-2D).

where $K = (K_{m,n})_{1 \leq m \leq M, 1 \leq n \leq N} \in \mathbb{R}^{M \times N}$ is the observation matrix corresponding to a discretized version of a Laplace transform, and $w = (w_m)_{1 \leq m \leq M} \in \mathbb{R}^M$ represents some additive acquisition noise. An efficient resolution strategy is to employ a penalized approach that defines an estimate $\hat{x} \in \mathbb{R}^N$ of \bar{x} as a solution to the following constrained minimization problem [3]:

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \Psi(x) \quad \text{subject to} \quad \|Kx - y\| \leq \tau, \quad (2)$$

where $\tau > 0$ is a parameter depending on the noise characteristics, and Ψ is a regularization function that allows to add prior information on the sought signal and to impose the positivity of its entries. The so-called maximum entropy reconstruction approach [4], corresponding to the choice of the Shannon entropy function [5, 6] for the penalization term Ψ , has been at the core of several papers dealing with regularized inverse Laplace transform [7, 8, 9, 10, 11, 12]. A more recent approach consists in adopting for Ψ a criterion enforcing both sparsity and positivity, with the aim to improve the resolution of narrow peaks possibly present in the sought signal [13, 14, 15, 16], but this strategy may be at the price of losing the smoothness of the solution. Hybrid regularization approaches combining both entropy and sparsity terms in Ψ should thus be envisaged so as to derive a flexible resolution method. However, the choice of an efficient optimization algorithm to solve the resulting constrained problem (2) (or its Lagrangian formulation) remains a challenging task.

On the one hand, in the case of entropy regularization, the optimization techniques proposed in the literature are usually *ad hoc* and difficult to extend to other types of priors [17, 18, 19]. On the other hand, in the case of sparsity regularization, the minimization step is usually handled by proximal optimization methods [20]. These methods, grounded on the use of the proximity operator [21] are highly flexible, and particularly efficient for handling functionals for which the proximity operators have a closed form expression. In this paper, we provide a comprehensive list of formula for the proximity operators of new hybrid regularization functions combining a non necessarily convex sparsity enhancing term and an entropy penalty. This list constitutes a very helpful tool for

deriving fast resolution methods for the regularized inversion of (1). The applicability of our method is illustrated by means of an example in the context of DOSY NMR.

The paper is organized as follows: In section 2, we introduce our notation and recall the definition of the proximity operator. Section 3 presents our main contribution, that is the expression of the proximity operators of several combination of entropy and sparsity promoting penalties. Then, Section 4 presents an application of our results to the resolution of an inverse problem arising in DOSY NMR signal reconstruction. Finally, some conclusions are drawn in Section 5.

2. PROBLEM STATEMENT

2.1. Hybrid regularization

Let us define the following family of hybrid regularization functions of the form:

$$(\forall x \in \mathbb{R}^N) \quad \Psi(x) = \alpha\Psi_1(x) + \beta\Psi_2(x), \quad (3)$$

where $\Psi_1 : \mathbb{R}^N \rightarrow]-\infty, +\infty]$ and $\Psi_2 : \mathbb{R}^N \rightarrow \mathbb{R}$ are lower semi continuous (lsc) and proper functions and (α, β) are some positive weights. In this paper, we focus on the case when Ψ_1 is either the Shannon or the Burg entropy [22] and Ψ_2 is a sparsity promoting prior. An efficient strategy to promote the sparsity of the sought signal is to choose Ψ_2 as the ℓ_0 penalty that counts the number of nonzeros in x . However, the later function is neither differentiable nor continuous at 0, so that its convex approximation, the ℓ_1 norm, that sums the absolute value of the signal entries is often used instead [23]. Continuous, but non convex, approximations have also been proposed for the ℓ_0 penalty, namely the log-sum penalty [24] and the Cauchy penalty [25]. Note that all the aforementioned regularization terms are separable, so that (3) can be rewritten as

$$(\forall x \in \mathbb{R}^N) \quad \Psi(x) = \sum_{n=1}^N \psi(x_n), \quad (4)$$

with $\psi = \alpha\psi_1 + \beta\psi_2$, $\psi_1 : \mathbb{R} \rightarrow]-\infty, +\infty]$ and $\psi_2 : \mathbb{R} \rightarrow \mathbb{R}$. Figures 1 and 2 illustrate the considered functions, for the entropy and sparse penalty terms ψ_1 and ψ_2 , respectively.

2.2. Proximity Operator

When Ψ is convex on \mathbb{R}^N , its proximity operator at $x \in \mathbb{R}^N$ is defined as the unique minimizer of $\Psi + \frac{1}{2}\|\cdot - x\|^2$ [26]. This operator has been generalized for lsc proper functions that are not necessarily convex in [27, Sec.XV-4], as the multi-valued operator:

$$\text{prox}_\Psi : x \mapsto \underset{y \in \mathbb{R}^N}{\text{Argmin}} \left(\frac{1}{2}\|y - x\|^2 + \Psi(y) \right). \quad (5)$$

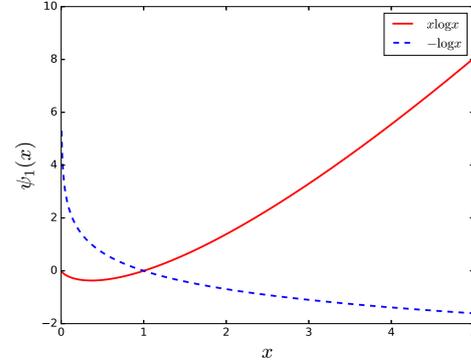


Fig. 1. Examples of entropy based penalties, in the scalar case: Shannon (continuous red line) and Burg (dashed blue line) entropy priors. Note that $\psi_1 = +\infty$ for negative values of x .

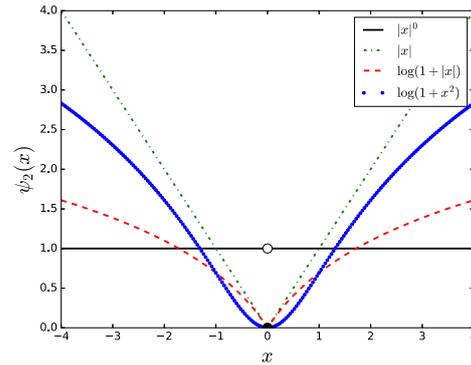


Fig. 2. Examples of sparsity promoting penalties, in the scalar case: ℓ_0 (continuous black line), ℓ_1 (dashed-dotted green line), log-sum (dashed red line) and Cauchy (dotted blue line) priors.

Since Ψ in (4) takes a separable form, its proximity operator (5) is given by ([21]):

$$\text{prox}_\Psi(x) = (p_n(x_n))_{1 \leq n \leq N}, \quad (6)$$

where, for every $n \in \{1, \dots, N\}$,

$$p_n(x_n) = \text{prox}_{\alpha\psi_1 + \beta\psi_2}(x_n). \quad (7)$$

In the sequel, we will focus our attention on the resolution of the scalar optimization problem (7), when $\alpha \in]0, +\infty[$ and $\beta \in [0, +\infty[$.

3. PROXIMITY OPERATORS FOR HYBRID SPARSE + ENTROPY PRIORS

3.1. Shannon entropy

The Shannon entropy regularization is defined, for all $x \in \mathbb{R}$, as

$$\psi_1(x) = \begin{cases} x \log x & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ +\infty & \text{elsewhere.} \end{cases} \quad (8)$$

According to [20], the proximity operator of $\alpha\psi_1$ at $x \in \mathbb{R}^N$, with $\alpha \in]0, +\infty[$, reads:

$$\text{prox}_{\alpha\psi_1}(x) = \alpha W \left(\frac{1}{\alpha} \exp \left(\frac{x}{\alpha} - 1 \right) \right), \quad (9)$$

where W states for the W Lambert function [28]. The remainder of this section shows how to generalize this expression in the case when $\beta \in]0, +\infty[$.

3.1.1. Shannon entropy + ℓ_1

Set $(\alpha, \beta) \in]0, +\infty[^2$. The hybrid Shannon entropy + ℓ_1 penalty is defined, for every $x \in \mathbb{R}$, as

$$\psi(x) = \begin{cases} \alpha x \log x + \beta x & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ +\infty & \text{elsewhere.} \end{cases} \quad (10)$$

Its proximity operator reads, for every $x \in \mathbb{R}$:

$$\text{prox}_{\psi}(x) = \alpha W \left(\frac{1}{\alpha} \exp \left(\frac{x - \beta}{\alpha} - 1 \right) \right). \quad (11)$$

3.1.2. Shannon entropy + ℓ_0

Set $(\alpha, \beta) \in]0, +\infty[^2$. The hybrid Shannon entropy + ℓ_0 penalty is defined, for every $x \in \mathbb{R}$, as

$$\psi(x) = \begin{cases} \alpha x \log x + \beta & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ +\infty & \text{elsewhere.} \end{cases} \quad (12)$$

Its proximity operator reads, for every $x \in \mathbb{R}$:

$$\text{prox}_{\psi}(x) = \begin{cases} p & \text{if } \beta < \bar{\beta} \\ \{0, p\} & \text{if } \beta = \bar{\beta} \\ 0 & \text{elsewhere,} \end{cases} \quad (13)$$

where $\bar{\beta} = \frac{1}{2}p^2 + \alpha p \in]0, +\infty[$ and

$$p = \alpha W \left(\frac{1}{\alpha} \exp \left(\frac{x}{\alpha} - 1 \right) \right). \quad (14)$$

Note that, due to the discontinuity of the ℓ_0 penalty in 0, prox_{ψ} is multi-valued in the limit case when $\beta = \bar{\beta}$.

3.1.3. Shannon + log-sum

Set $(\alpha, \beta, \delta) \in]0, +\infty[^3$. The hybrid Shannon entropy + log-sum penalty is defined, for every $x \in \mathbb{R}$, as

$$\psi(x) = \begin{cases} \alpha x \log x + \beta \log(\delta + x) & \text{if } x > 0 \\ \beta \log(\delta) & \text{if } x = 0 \\ +\infty & \text{elsewhere.} \end{cases} \quad (15)$$

For every $x \in \mathbb{R}$, its proximity operator is given by

$$\text{prox}_{\psi}(x) = \underset{p \in]0, +\infty[\text{ s.t. } \varphi(p)=0}{\text{Argmin}} \left(\frac{1}{2}(x-p)^2 + \psi(p) \right),$$

with $\varphi(p) = p^2 + (\delta - x + \alpha)p + \alpha(\delta + p) \log(p) + \delta(\alpha - x) + \beta$.

3.1.4. Shannon + Cauchy

Set $(\alpha, \beta, \delta) \in]0, +\infty[^3$. The hybrid Shannon entropy + Cauchy penalty is defined, for every $x \in \mathbb{R}$, as

$$\psi(x) = \begin{cases} \alpha x \log x + \beta \log(\delta + x^2) & \text{if } x > 0 \\ \beta \log(\delta) & \text{if } x = 0 \\ +\infty & \text{elsewhere.} \end{cases} \quad (16)$$

For every $x \in \mathbb{R}$, its proximity operator is given by

$$\text{prox}_{\psi}(x) = \underset{p \in]0, +\infty[\text{ s.t. } \varphi(p)=0}{\text{Argmin}} \left(\frac{1}{2}(x-p)^2 + \psi(p) \right),$$

with

$\varphi(p) = p^3 + (\alpha - x)p^2 + (\delta + 2\beta)p + \alpha(\delta + p^2) \log(p) + \delta(\alpha - x)$.

3.2. Case of Burg entropy

The Burg entropy regularization is defined, for all $x \in \mathbb{R}$, as

$$\psi_1(x) = \begin{cases} -\log x & \text{if } x > 0 \\ +\infty & \text{elsewhere.} \end{cases} \quad (17)$$

According to [29], the proximity operator of the Burg entropy function reads, for every $x \in \mathbb{R}^N$, for every $\alpha \in]0, +\infty[$:

$$\text{prox}_{\alpha\psi_1}(x) = \frac{x + \sqrt{x^2 + 4\alpha}}{2}. \quad (18)$$

Let us now present the extension of this result to the case when $\beta \in]0, +\infty[$.

3.2.1. Burg entropy + ℓ_1

Set $(\alpha, \beta) \in]0, +\infty[^2$. The hybrid Burg entropy + ℓ_1 penalty is defined, for every $x \in \mathbb{R}$, as

$$\psi(x) = \begin{cases} -\alpha \log x + \beta x & \text{if } x > 0 \\ +\infty & \text{elsewhere.} \end{cases} \quad (19)$$

Its proximity operator reads, for every $x \in \mathbb{R}$ ([29]):

$$\text{prox}_{\psi}(x) = \frac{x - \beta + \sqrt{(\beta - x)^2 + 4\alpha}}{2}. \quad (20)$$

3.2.2. Burg + ℓ_0

Set $(\alpha, \beta) \in]0, +\infty[^2$. The hybrid Burg entropy + ℓ_0 penalty is defined, for every $x \in \mathbb{R}$, as

$$\psi(x) = \begin{cases} -\alpha \log x + \beta & \text{if } x > 0 \\ +\infty & \text{elsewhere.} \end{cases} \quad (21)$$

Its proximity operator reads, for every $x \in \mathbb{R}$:

$$\text{prox}_\psi(x) = \frac{x + \sqrt{x^2 + 4\alpha}}{2}. \quad (22)$$

3.2.3. Burg + log-sum

Set $(\alpha, \beta, \delta) \in]0, +\infty[^3$. The hybrid Burg entropy +log-sum penalty is defined, for every $x \in \mathbb{R}$, as

$$\psi(x) = \begin{cases} -\alpha \log x + \beta \log(\delta + x) & \text{if } x > 0 \\ +\infty & \text{elsewhere.} \end{cases} \quad (23)$$

For every $x \in \mathbb{R}$, its proximity operator is given by

$$\text{prox}_\psi(x) = \underset{p \in]0, +\infty[\text{ s.t. } \varphi(p)=0}{\text{Argmin}} \left(\frac{1}{2}(x-p)^2 + \psi(p) \right),$$

with $\varphi(p) = p^3 + (\delta - x)p^2 + p(\beta - \delta x - \alpha) - \delta\alpha$.

3.2.4. Burg + Cauchy

Set $(\alpha, \beta, \delta) \in]0, +\infty[^3$. The hybrid Burg entropy + Cauchy penalty is defined, for every $x \in \mathbb{R}$, as

$$\psi(x) = \begin{cases} -\alpha \log x + \beta \log(\delta + x^2) & \text{if } x > 0 \\ +\infty & \text{elsewhere.} \end{cases} \quad (24)$$

For every $x \in \mathbb{R}$, its proximity operator is given by

$$\text{prox}_\psi(x) = \underset{p \in]0, +\infty[\text{ s.t. } \varphi(p)=0}{\text{Argmin}} \left(\frac{1}{2}(x-p)^2 + \psi(p) \right),$$

with $\varphi(p) = p^4 - xp^3 + (\delta + 2\beta - \alpha)p^2 - \delta xp - \delta\alpha$.

4. APPLICATION TO NMR SIGNAL RECONSTRUCTION

4.1. Problem formulation

During the DOSY (Diffusion Order Spectroscopy) experiment, proposed by [30] to analyze the properties of complex chemical mixtures, a series of measurements are acquired for different pulsed field gradient strengths, the data are then processed with the aim to separate different species according to their diffusion coefficient. The DOSY NMR data $y \in \mathbb{R}^M$ gathers the results of $M \geq 1$ experiments corresponding to different acquisition settings characterized by the set of parameters $t = (t_m)_{1 \leq m \leq M}$. The problem is then to estimate,

from these measurements, the values of the true diffusion distribution $\bar{\chi}(T)$ at given positions $T = (T_n)_{1 \leq n \leq N}$. The relation between y and the sought signal $\bar{x} = (\bar{\chi}(T_n))_{1 \leq n \leq N} \in \mathbb{R}^N$ can be written under the form (1) where $K \in \mathbb{R}^{M \times N}$ is given, for every $m \in \{1, \dots, M\}$ and every $n \in \{1, \dots, N\}$ by

$$K_{m,n} = \exp(-T_n t_m), \quad (25)$$

where $w \in \mathbb{R}^M$ is a perturbation noise. In practice, the noise is assumed to be zero-mean Gaussian, i.i.d, with known standard deviation $\sigma > 0$ so that we propose to find an estimate $\hat{x} \in \mathbb{R}^N$ of \bar{x} by solving (2) with $\tau = \eta\sqrt{M}\sigma$, where $\eta > 0$ is a weight closed to 1 [31].

4.2. Experimental results

We now present the experimental results obtained when, for every $n \in \{1, \dots, N\}$, for every $m \in \{1, \dots, M\}$,

$$\begin{cases} t_m = t_{\min} + \frac{m-1}{M} t_{\max}, \\ T_n = T_{\min} \exp\left(-\frac{(n-1)}{N} \log\left(\frac{T_{\min}}{T_{\max}}\right)\right), \end{cases} \quad (26)$$

with $M = 50$, $N = 200$, $T_{\min} = 1$, $T_{\max} = 10^3$, $t_{\min} = 0$, $t_{\max} = 1.5$. We consider two synthetic signals A and B. The synthetic signal A models a monodisperse distribution with symmetric log-normal shape located at $\bar{T} = 16$ while signal B corresponds to a polydisperse distribution that is the sum of two log-normal patterns located at $\bar{T}_1 = 4$ and $\bar{T}_2 = 32$. Table 1 presents the resulting signal to noise ratio (SNR) defined as $10 \log_{10} (\|\bar{x}\|^2 / \|\hat{x} - \bar{x}\|^2)$ which was obtained when solving Problem (2) for both datasets, for different values of σ , and several choices for the regularization term Ψ , namely (8), (10), (17) and (19). Here, we choose to focus only on convex priors so that the optimization problem can be efficiently solved with the PPXA+ algorithm [32]. Parameter η is set equals to 1.2, since it was observed to lead to the best reconstruction results. Moreover, when hybrid penalties are considered, we fix $\beta = 1 - \alpha$, and we optimize $\alpha \in]0, 1]$ manually so as to obtain a minimal reconstruction error.

		σ	Shannon prior	Shannon + ℓ_1 prior	Burg prior	Burg + ℓ_1 prior
Dataset A	10^{-2}		12.45	13.16	12.92	12.92
	10^{-3}		18.16	20.86	12.11	13.44
	10^{-4}		20.87	25.95	12.03	15.53
Dataset B	10^{-3}		11.14	18.52	6.20	9.54
	10^{-4}		18.11	20.23	7.41	10.95
	10^{-5}		19.05	26.30	7.54	10.98

Table 1. SNR in dB of the restored signals A and B for various choices of the penalization function Ψ .

One can observe that the addition of the ℓ_1 norm in the penalty functions improves significantly the quality of restoration results, especially in the case of the polydisperse signal B. The best results are obtained with the combination of Shannon entropy and ℓ_1 penalties. This is also confirmed by visual inspection, as it can be noticed in Figure 3 displaying an example of reconstruction result we obtained in the case of dataset B, with $\sigma = 10^{-5}$.

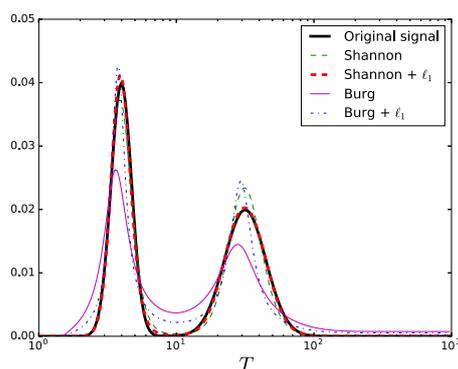


Fig. 3. Reconstruction results for dataset B and $\sigma = 10^{-5}$.

5. CONCLUSION

This work expands the extensive list of proximity operators available in the literature [20, 29, 33] by considering the case of separable functions combining entropy and sparsity promoting terms. Through numerical experiments, we show that these functions may serve as efficient hybrid penalties for solving ill-posed inverse problems in the context of DOSY NMR spectroscopy. Since only convex penalties were tested in our experiments, the PPXA+ algorithm was retained. It should however be emphasized that the novel proximity operators we derive could be applied in a variety of proximal algorithms, in the convex [34] or the non-convex case [35, 36]. The latter case should be of particular interest in the context of blind signal restoration problems such as those encountered in [37, 38] where the proposed hybrid penalties could be beneficial.

6. REFERENCES

- [1] H. E. Pence and A. J. Williams, "Big data and chemical education," *Journal of Chemical Education*, vol. 93, no. 3, pp. 504–508, Mar. 2016.
- [2] C. S. Johnson Jr, "Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 34, pp. 203–256, 1999.
- [3] J. Idier, *Bayesian Approach to Inverse Problems*, ISTE Ltd and John Wiley & Sons Inc, Apr. 2008.
- [4] P. P. B. Eggermont, "Maximum entropy regularization for Fredholm integral equations of the first kind," *SIAM Journal on Mathematical Analysis*, vol. 24, no. 6, pp. 1557–1576, 1993.
- [5] R. M. Gray, *Entropy and Information Theory*, Springer US, 2nd edition, 2011.
- [6] M. Basseville and J. F. Cardoso, "On entropies, divergences, and mean values," in *IEEE International Symposium on Information Theory (ISIT 1995)*, Whistler, BC, 17–22 Sep. 1995, p. 330.
- [7] E. D. Laue, J. Skilling, J. Staunton, S. Sibisi, and R. G. Brereton, "Maximum entropy method in nuclear magnetic resonance spectroscopy," *Journal of Magnetic Resonance*, vol. 62, no. 3, pp. 437–452, 1985.
- [8] A. K. Livesey, P. Licinio, and M. Delaye, "Maximum entropy analysis of quasielastic light scattering from colloidal dispersions," *The Journal of chemical physics*, 1986.
- [9] A.K. Livesey and J.C. Brochon, "Analysing the distribution of decay constants in pulse fluorimetry using the maximum entropy method," *Biophysical Journal*, vol. 52, pp. 693–706, 1987.
- [10] M.-A. Delsuc and T. E. Malliavin, "Maximum entropy processing of DOSY NMR spectra," *Analytical Chemistry*, vol. 70, pp. 2146–2148, 1998.
- [11] F. Mariette, J.-P. Gullieument, C. Tellier, and P. Marchal, "Continuous relaxation time distribution decomposition by MEM," in *Signal Treatment and Signal Analysis in NMR*, vol. 18 of *Data Handling in Science and Technology*, pp. 218–234, 1996.
- [12] G. J. Daniell and P. J. Hore, "Maximum entropy and bayesian methods: Cambridge, england, 1988," in *The Inverse Problem for Nuclear Magnetic Resonance*, J. Skilling, Ed., pp. 297–302. Springer Netherlands, Dordrecht, 1989.
- [13] M. Urbanczyk, D. Bernin, W. Kozminski, and K. Kazimierzuk, "Iterative thresholding algorithm for multiexponential decay applied to pgse nmr data," *Analytical chemistry*, vol. 85, no. 3, pp. 1828–1833, 2013.
- [14] X. Kaipin and Z. Shanmin, "Trust-region algorithm for the inversion of molecular diffusion nmr data," *Analytical chemistry*, vol. 86, no. 1, pp. 592–599, 2013.
- [15] B. Worley, "Convex accelerated maximum entropy reconstruction," *Journal of Magnetic Resonance*, vol. 265, pp. 90–98, Apr. 2016.
- [16] P. Berman, O. Levi, Y. Parmet, M. Saunders, and Z. Wiesman, "Laplace inversion of low-resolution NMR relaxometry data using sparse representation methods," *Concepts in Magnetic Resonance. Part A*, vol. 42, no. 3, pp. 72–88, 2013.
- [17] I. F. Gorodnitsky, "An extension of an interior-point method for entropy minimization," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1999)*, vol. 3, pp. 1697–1700, 15–19 Mar. 1999.
- [18] E. Chouzenoux, S. Moussaoui, J. Idier, and F. Mariette, "Efficient maximum entropy reconstruction of nuclear magnetic resonance t1-t2 spectra," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6040–6051, 2010.

- [19] M. Willis, B. Jeffes, and D. Long, "Maximum entropy image restoration revisited," *Proceedings of the IEEE International Conference on Image Processing (ICIP 2000)*, pp. 89–92, 10–13 Sept. 2000.
- [20] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, S. R. Burachik, L. P. Combettes, V. Elser, R. D. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer, New York, NY, 2011.
- [21] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, CMS Books in Mathematics. Springer, New York, NY, 2011.
- [22] R. Nityananda and R. Narayan, "Maximum entropy image reconstruction—a practical non-information-theoretic approach," *Journal of Astrophysics and Astronomy*, vol. 3, no. 4, Dec. 1982.
- [23] M.B. Candès, E.J. Wakin, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.
- [24] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 317–329, 2010.
- [25] A. Antoniadis, D. Leporini, and J.-C. Pesquet, "Wavelet thresholding for some classes of non-Gaussian noise," *Statistica Neerlandica*, vol. 56, pp. 434–453, 2002.
- [26] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société mathématique de France*, vol. 93, pp. 273–299, 1965.
- [27] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms II: Advanced theory and bundle methods*, vol. 306 of *Grundlehren der mathematischen Wissenschaften*, Springer-Verlag, New York, 1993.
- [28] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Advances in Computational mathematics*, vol. 5, no. 1, pp. 329–359, 1996.
- [29] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Problems*, vol. 23, no. 4, pp. 1495, 2007.
- [30] K. F. Morris and Johnson Jr C. S., "Diffusion-ordered 2D NMR spectroscopy," *Journal of the American Chemical Society*, vol. 114, pp. 3139–3141, 1992.
- [31] H.J. Trussell, "Convergence criteria for iterative restoration methods," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 31, no. 1, pp. 129–136, 1983.
- [32] J.-C. Pesquet and N. Pustelnik, "A parallel inertial proximal optimization method," *Pacific Journal of Optimization*, vol. 8, no. 2, pp. 273–305, 2012.
- [33] M. El Gheche, J. C. Pesquet, and J. Farah, "A proximal approach for optimization problems involving kullback divergences," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*, Vancouver, BC, Canada, 26–31 May 2013, pp. 5984–5988.
- [34] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 31–54, Oct. 2015.
- [35] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function," *Journal of Optimization Theory and Applications*, vol. 162, no. 1, pp. 107–132, Jul. 2014.
- [36] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "A block coordinate variable metric forward-backward algorithm," *Journal of Global Optimization*, pp. 1–29, 2016.
- [37] A. Repetti, M.-Q. Pham, L. Duval, E. Chouzenoux, and J.-C. Pesquet, "Euclid in a taxicab: Sparse blind deconvolution with smoothed ℓ_1/ℓ_2 regularization," *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 539–543, 2015.
- [38] I. Toumi, S. Caldarelli, and B. Torrèsani, "A review of blind source separation in nmr spectroscopy," *Progress in nuclear magnetic resonance spectroscopy*, vol. 81, pp. 37–64, 2014.



Cite this: DOI: 10.1039/c6an01902a

PALMA, an improved algorithm for DOSY signal processing†

Afef Cherni,^{a,b} Emilie Chouzenoux^{b,c} and Marc-André Delsuc^{*a}

NMR is a tool of choice for the measurement of diffusion coefficients of species in solution. The DOSY experiment, a 2D implementation of this measurement, has been proven to be particularly useful for the study of complex mixtures, molecular interactions, polymers, etc. However, DOSY data analysis requires to resort to the inverse Laplace transform, in particular for polydisperse samples. This is a known difficult numerical task for which we present here a novel approach. A new algorithm based on a splitting scheme and on the use of proximity operators is introduced. Used in conjunction with a Maximum Entropy and ℓ_1 hybrid regularisation, this algorithm converges rapidly and produces results robust against experimental noise. This method has been called PALMA. It is able to reproduce faithfully monodisperse as well as polydisperse systems, and numerous simulated and experimental examples are presented. It has been implemented on the server <http://palma.labo.igbmc.fr> where users can have their datasets processed automatically.

Received 25th August 2016,
Accepted 3rd January 2017

DOI: 10.1039/c6an01902a

www.rsc.org/analyst

1. Introduction

Diffusion coefficients can be efficiently measured by NMR using magnetic field gradients. The most classical approach consists of applying a symmetric pair of pulsed field gradients (PFG) of varying intensity, separated by a diffusion delay Δ . Random displacements of the molecule during Δ because of Brownian motion result in modulation of the signal intensity I following the Stejskal–Tanner equation:^{1,2}

$$I(q) = I_0 \exp(-D\Delta q^2) \quad (1)$$

where D is the diffusion coefficient of the molecular species and $q = \gamma\delta g$, the measure of the phase dispersion created by the PFG. Here, γ is the gyromagnetic ratio of the studied spin, and δ and g are the duration and intensity of the PFG, respectively. A least squares fit of the experimental values to an exponential decay provides an estimate of the value of D . The DOSY experiment, introduced by Johnson,³ is a representation of this measure as a 2D spectrum, with chemical shifts presented horizontally and diffusion coefficients vertically. DOSY has

been used intensively to analyse molecular interactions, to sort the components in complex mixtures, or to evaluate molecular size distributions.^{4–6}

A monodisperse sample presents a well-defined diffusion coefficient, and a simple exponential adjustment of eqn (1) allows the determination of D . When several compounds share the same chemical shift, resulting in overlapping lines in the NMR spectrum, the result of a mono-exponential fit becomes incorrect. A simple column-wise least squares fit to two or more exponentials presents instabilities in noisy datasets, which make this approach difficult to use in complex cases.

Several methods have emerged in the literature for the analysis of complex mixtures of monodisperse species, where the difficulty mostly arises from the presence of many overlapping species in the spectrum. Approaches based on a global analysis of the whole experimental matrix have been proposed based on a clever decomposition of the 2D spectrum matrix into a multivariate model, which allows to extract the spectra of each species along with their respective diffusion profiles.^{7,8} Some developments on this approach have been based on a harmonic analysis of the decay.^{9–14} The exponential hypothesis can even be relaxed using methods related to blind source deconvolution.^{15–18}

All these approaches model the sample as a mixture of species with a characteristic decay pattern. Some samples such as polydisperse polymers, dendrimers, nanoparticles, gels and aggregated species, because of the variation in the size, length or aggregate state of the different molecules in the sample, present a distribution of diffusion coefficients rather than a single coefficient. Moreover, the presence of common decay

^aInstitut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U596, CNRS UMR 7104, Université de Strasbourg, 67404 Illkirch-Graffenstaden, France. E-mail: madelsuc@unistra.fr

^bUniversité Paris-Est, LIGM (UMR 8049), CNRS, ENPC, ESIEE Paris, UPEM, Marne-la-Vallée, France

^cCenter for Visual Computing, CentraleSupélec and GALEN group, INRIA, Paris-Saclay, Châtenay-Malabry, France

† Electronic supplementary information (ESI) available. See DOI: 10.1039/C6AN01902A

patterns at different chemical shifts cannot really be assumed anymore, as in these complex systems a subtle coupling between the chemical shift and the size usually broadens the spectral line, with each spectral channel sampling a slightly different subset of the species, so each spectral channel has to be processed independently.

For these strongly polydisperse samples a precise determination of the diffusion distribution is of great analytical importance, it is however a difficult task. Because of this difficulty, polydispersity is commonly measured by a polydispersity index (PDI) defined as the ratio of the mass average molar mass M_w to its number average molar mass M_n : $PDI = M_w/M_n$. This quantity characterises the breadth of the distribution independently of the details of its shape, a PDI of 1.0 indicates monodispersity. The PDI is commonly measured by size exclusion chromatography, (electron) microscopy, light or X-ray scattering, or even NMR-DOSY.¹⁹

A polydisperse sample has to be analysed with a distribution $X(D)$ of diffusion coefficients and eqn (1) becomes:

$$I(q) = \int_{D_{\min}}^{D_{\max}} X(D) \exp(-D\Delta q^2) dD \quad (2)$$

Determining the distribution $X(D)$ from $I(q)$ requires to solve the Laplace inversion of the q^2 dependency of $I(q)$.

The shape of the distribution X can be modelled by a Gaussian function or by any other symmetric or asymmetric analytical shape, and the parameters for this shape are fitted to the experimental data.²⁰ This straightforward approach is very sensitive to the choice of the shape, and will fail if it is not well adapted to the data, or if the distribution contains several isolated wide shapes and it should be used with care.

In this work, we present a general approach that solves the Laplace inversion problem presented in eqn (2). A new algorithmic approach based on a splitting scheme and on the use of proximity operators is introduced. Used in conjunction with Maximum Entropy and ℓ_1 regularisations, the algorithm is stable against experimental noise, reproduces faithfully monodisperse as well as polydisperse situations, and converges rapidly.

2. Theory

2.1 Problem description

We assume that the diffusion experiment was performed over a series of M values of q (by varying δ , g or both) and measured as a series of intensities y_m for a given chemical shift value. The problem stated by eqn (2) can be rewritten in a discrete form to be solved numerically:

$$y_m = \sum_{n=1}^N x_n \exp(-D_n \Delta q_m^2)$$

with D_n ranging from D_{\min} to D_{\max} . As this expression is linear in x_n , it can be rewritten as follows:

$$Y = HX \quad (3)$$

where $Y = \{y_m, 1 \leq m \leq M\}$ is the experimental series, $X = \{x_n, 1 \leq n \leq N\}$ is a sampling of the distribution, and H is an $M \times N$ matrix with entries $H_{m,n} = \exp(-D_n \Delta q_m^2)$. In this work, we call X the Laplace spectrum of Y . Determining X from the knowledge of H and Y is an ill-posed problem as the experimental points are inevitably tainted with noise, and H is usually a non-square matrix, with $N > M$. A simple inversion does not provide a valid solution, and one has to resort to alternative approaches.

2.2 Lagrangian formulation

A general approach for solving eqn (3) is to generate a solution X that solves the following regularised minimisation problem:

$$\underset{X \in \mathbb{R}^N}{\text{minimise}} \|HX - Y\|^2 + \mu \Psi(X). \quad (4)$$

The first term evaluates the distance between data and the reconstruction, while the second term is the regulariser, the Lagrangian coefficient $\mu > 0$ acting as a weight between the two. The regularisation function Ψ is aimed at selecting among all possible distributions, the most natural one, given the experimental evidence, using some *a priori* information on the problem. It is usually built as a measure of the cost of the reconstruction (in terms of energy, information, number of signals, *etc.* see below) and tends to favour an empty spectrum. Depending on the expression chosen for Ψ , the problem can be solved by different approaches. The CONTIN method²¹ solves this problem for $\Psi(X) = \|GX\|^2$, where G is a matrix that contains prior assumptions about the data. Classical choices are $G = Id$ which selects the solution with the least energy, or the first or second derivatives thus removing fluctuations not required for a faithful reconstruction. CONTIN has shown great success since its introduction more than 30 years ago, however it suffers from slow convergence and over-smoothed solutions. Choosing the opposite of the entropy as the regularisation function ($\Psi(X) = \sum (x_n/a) \log(x_n/a)$) allows to produce the distribution with the least information in the sense of Shannon.²² This Maximum Entropy (MaxEnt) approach has been shown to be of great efficiency and robustness for solving the DOSY problem²³ and has been widely used. However, because of the strong curvature of the entropy function, the classical implementations^{23,24} of this approach are known to present slow convergence rates. Kazimierzczuk *et al.*²⁵ proposed recently to use $\Psi(X) = \sum |x_n| = \|X\|_1$, where Ψ is the ℓ_1 norm of X . Their approach relies on previous reports that have shown that this is equivalent to selecting the spectrum with the fewest non-null values. The principal advantage of their approach is to rely on recent major advances in the field of convex minimisation and compressed sensing. The algorithm ITAMeD they developed is based on the soft thresholding approach and allows a rapid convergence toward the solution. Enforcing a minimum number of non-null values in X is a good approach when the sample is a mixture of monodisperse compounds, and the Laplace spectrum a set of sharp lines, it is not appropriate however for the analysis of polydisperse samples that may present very large distributions. Urbańczyk *et al.*²⁶

recently extended this work to minimising the ℓ_p norm ($\|X\|_p = (\sum |x_n|^p)^{\frac{1}{p}}$) with $1 \leq p \leq 2$. As expected, the authors show that adapting the p parameter to this kind of data allows reconstruction of spectra of various widths. However, the choice of p is somewhat *ad hoc* and has to be adapted to each situation. In addition, the authors rely here on the IRLS²⁷ algorithm that is slower than ITAMeD. The TRAI method²⁸ that has been proposed recently is not explicitly based on a regularisation approach, but rather on an early stopping strategy, in conjunction with a trust region algorithm. This method is claimed to be efficient in polydisperse distributions. The early stopping approach creates a bias in the final solution which can be assimilated to an implicit regularisation albeit with no analytical definition, and this could be seen as a deficiency.²⁹

2.3 Constrained formulation

The Lagrangian coefficient μ involved in eqn (4) may be difficult to adjust in practice. However, one has often some precise knowledge about the level of noise corrupting the data. We propose to adopt a more practical formulation by solving the following constrained optimisation problem:

$$\underset{X \in \mathbb{R}^N}{\text{minimise}} \Psi(X) \quad \text{subject to} \quad \|HX - Y\| \leq \eta \quad (5)$$

where $\eta > 0$ is related to the expected quality of the fit, based on an estimate of the experimental noise. This has the advantage to shift the burden of determining the adequate value of a Lagrangian coefficient to the much simpler task of estimating a noise level.

2.4 Hybrid regularisation

In order to favour both smooth (polydisperse) and sparse (monodisperse) shapes in the estimated signal, we propose a novel regularisation defined as follows:

$$\Psi(X) = \lambda \text{ent}(X, a) + (1 - \lambda) \ell_1(X) \quad (6)$$

where $\text{ent}(X, a)$ is given by

$$\text{ent}(X, a) = \begin{cases} \sum_{n=1}^N \frac{x_n}{a} \log\left(\frac{x_n}{a}\right) & \text{if } x_n > 0 \\ 0 & \text{if } x_n = 0 \\ +\infty & \text{elsewhere,} \end{cases}$$

and is the opposite of the Shannon entropy with a flat prior $a > 0$, typically chosen here from an estimate value of $\sum x_n$. $\ell_1(X)$ is the ℓ_1 norm of the vector X defined as:³⁰

$$\ell_1(X) = \sum_{n=1}^N |x_n|$$

and $\lambda \in [0, 1]$ allows to control the balance between the sparsity prior and the entropy prior.

2.5 Proximity operator

The general problem expressed in eqn (5) can be elegantly solved by the convex optimisation algorithm PPXA+, based on

the use of proximity operators.³¹ This allows a generic solver for any choice of convex function Ψ while ensuring a very rapid convergence toward the optimal solution. Note that the proximal approach is also the basis of the ITAMeD method. The FISTA algorithm on which ITAMeD is based³² is derived from the soft-thresholding operator, the proximity operator of the ℓ_1 norm operator. To our knowledge, it has never been applied to the MaxEnt penalty nor to a hybrid approach as proposed here.

It can be shown³³ that the proximity operator of the functional Ψ used in eqn (6) can be expressed as follows (see ESI sections 1 and 2† for details):

$$\text{prop}_{\Psi}(X) = (p(x_n))_{1 \leq n \leq N} \quad (7)$$

where

$$p(x_n) = \begin{cases} \frac{\lambda}{a} \mathcal{W} \left[\frac{a}{\lambda} \exp \left(\frac{ax_n - a(1-\lambda)}{\lambda} + \log(a) - 1 \right) \right], & \text{if } \lambda \in]0, 1], \\ \text{sign}(x_n) \max(|x_n| - (1-\lambda), 0), & \text{if } \lambda = 0. \end{cases} \quad (8)$$

In the above expression, \mathcal{W} stands for the Lambert function, defined as the inverse function of $f: z \rightarrow ze^z$ for all $z \in \mathbb{C}$,³⁴ *i.e.*:

$$z = we^w \Leftrightarrow w = \mathcal{W}(z).$$

In the current context, only a restriction of \mathcal{W} to \mathbb{R}^+ is required. It should be noted that for pure Maximum Entropy ($\lambda = 1$) we recover the result from Combettes and Pesquet:³⁵

$$p_{\text{ent}}(x) = \frac{1}{a} \mathcal{W}(a \exp(x) + \log(a) - 1).$$

Similarly, pure ℓ_1 regularisation ($\lambda = 0$) brings the soft thresholding operator:

$$p_{\ell_1}(x) = \text{sign}(x) \max(|x| - 1, 0).$$

2.6 Algorithm

With the expression of the proximity operator given in eqn (7) and (8), the convex optimisation problem (5) can be easily solved using a proximal splitting algorithm. At each iteration, such a method alternates between the proximity operator Ψ and the proximity operator associated with the constraint $\|HX - Y\| \leq \eta$ (*i.e.* the projection operator onto this constrained set).

In order to ensure good convergence properties of our algorithm, we adopt the PPXA+ approach from Pustelnik *et al.*,³⁶ generalizing the PPXA method from Combettes and Pesquet.³⁷ These algorithms both rely on the Douglas–Rachford scheme,³⁸ which consists of replacing the involved proximity operators by their reflections (see ESI sections 3, 4 and 5† for details).

This leads us to the so-called PALMA algorithm, standing for “Proximal Algorithm for L_1 combined with MAXent prior”. This algorithm is fully detailed in the ESI.†

3. Materials & methods

3.1 Simulations

Several simulated datasets, chosen to represent various analytical situations, were used for the evaluation of the algorithm. Set A consists of three monodisperse components with diffusion coefficients $16 \mu\text{m}^2 \text{s}^{-1}$, $63 \mu\text{m}^2 \text{s}^{-1}$ and $230 \mu\text{m}^2 \text{s}^{-1}$, with the respective intensities of 1.0, 0.33 and 0.66. This dataset is equivalent to the simulation used in Urbańczyk *et al.*²⁵ Set B is a wide distribution, simulated as a log-normal distribution centred at $35 \mu\text{m}^2 \text{s}^{-1}$, which presents a PDI estimated to be 6.26. Sets C1 and C2 are asymmetric distributions built from 15 log-normal components, ranging from 18 to $85 \mu\text{m}^2 \text{s}^{-1}$, with intensities ranging from 0.1 to 10, and they have PDIs estimated respectively to be 1.79 and 1.32. In all simulations, 64 gradient values were simulated, and a Gaussian noise equal to 0.001%, 0.01%, 0.1%, or 1% of the initial point was added. The gradient values were chosen with a harmonic progression for set A, and with linear increments for sets B, C1, and C2. All Laplace spectra were reconstructed on 256 logarithmically sampled points. Other simulations with varying distributions and noise levels are also presented in the ESI (see section 6†).

3.2 NMR measurement

A set of PEO standards were purchased from American Polymer Standards Corporation (Mentor, OH, USA), and 3 samples were prepared. Sample (a) is a standard PEO with $M_w = 2343.3 \text{ g mol}^{-1}$ and PDI = 1.07; sample (b) is a standard PEO with $M_w = 4051.2 \text{ g mol}^{-1}$ and PDI = 1.28; sample (c) is a mixture prepared from standard PEOs ranging from 350 to 5250 g mol^{-1} for a theoretical M_w of $3238.5 \text{ g mol}^{-1}$ and a theoretical PDI of 2.01. Each sample was prepared and measured as described by Viéville *et al.*¹⁹

The crude plant extract was obtained from the brown algae *Sargassum muticum* as described in Vonthron-Sénécheau.³⁹ Two equivalent samples were prepared by dissolving each time 18.3 mg in 0.75 mL MeOD plus 0.4 mL D₂O, and chloroquine was added to one sample at a concentration of 0.16 mg mL^{-1} (1% w/w of plant dry extract). The NMR experiments were run on an Avance III Bruker spectrometer, operating at 700 MHz, and equipped with a TXI cryo-probe. DOSY spectra were acquired with the convection compensated experiment using bipolar pulses⁴⁰ (dstebpgp3spr pulse program). 50 gradient increments from 0.5 G cm^{-1} to 52.5 G cm^{-1} were used, with a cosine roll-off PFG shape. Each elementary PFG had a duration of $\delta = 1.1 \text{ ms}$ and the diffusion delay Δ was set to 150 ms. For each gradient intensity, 64 scans were acquired with a relaxation of 1.5 s, for a total experimental time of 1 hour 50 minutes. Each 1D spectrum was apodised with an unshifted sine-bell, zero-filled once and Fourier transformed. A spline baseline correction was applied, as well as a correction of small shifts caused by possible instabilities of the temperature control.⁴¹

3.3 Processing

The PALMA algorithm was implemented with the programming language python version 2.7 using the numpy/scipy

libraries. The algorithm was then packaged in a plugin to the SPIKE program developed by our group.⁴² All programs are available from the authors. All computations were performed on a Macintosh Mac Pro dual Xeon with a total of 8 cores, equipped with 32 GB of memory and running MacOSX 10.7.

The DOSY experiments presented in Fig. 2 and 3 were processed column-wise with $\lambda = 0.01$. The noise level along the gradient axis was estimated by the difference of the actual data from the result of a polynomial smoothing, this value being used to estimate η in eqn (5) (see ESI section 5.2† for more details). The diffusion axis was logarithmically sampled from 50 to $10\,000 \mu\text{m}^2 \text{s}^{-1}$ over 256 points.

Experiments on PEO were processed with a maximum of 20 000 iterations, for a total time of about 30 seconds for 110 DOSY profiles computed. Experiments on plant extract were performed taking all signals with an estimated SNR above 20 (26 dB) in the first 1D spectrum of the experimental matrix, with a maximum of 200 000 iterations, for a total time of about 1 hour for 1200 to 1400 DOSY profiles.

The algorithms ITAMeD, ℓ_p tailored-ITAMeD, and TRAIIn utilised in the ESI† were used as downloaded from their respective websites, using MATLAB program version R_2013b on MacOs.

3.4 Server

A Web server is available at <http://palma.labo.igbmc.fr>, where users may submit datasets for automatic processing. The python code of the PALMA algorithm is available on the same server as well as at <https://github.com/delsuc/PALMA>.

4. Results

4.1 Tests on simulated data

The PALMA algorithm described above was first tested on a series of simulated datasets. Fig. 1a presents the results obtained on the simulated experiment A consisting of the superposition of three monodisperse species, separated by less than a factor of 4 in diffusion coefficients, equivalent to the test used in Urbańczyk *et al.*²⁵ When analysed with PALMA using a null λ value, indicating a pure ℓ_1 regularisation, a Laplace spectrum consisting of 3 sharp peaks is produced as expected. Using the pure MaxEnt mode ($\lambda = 1$) on the same dataset, a broader spectrum is reconstructed. PALMA allows the weight between the two approaches to be freely varied. When doing so a narrowing of the MaxEnt distribution for the decreasing value of λ , characteristic of a bias toward monodisperse distributions, can be observed, with a sudden transition to sharp lines for $\lambda = 0$. As shown in Fig. 1b the same procedure was applied on a broad Gaussian line simulating a polydisperse polymer with a PDI of about 6, corresponding to experiment B. Again it can be observed that a null λ gives rise to sparse spectra with sharp lines this time only sampling the broad line in an inadequate manner. However, in contrast to the previous case, all the non-null values of λ lead to a nearly perfect reconstruction of the line-shape, with a correct determination of its width. In this example, the minimal

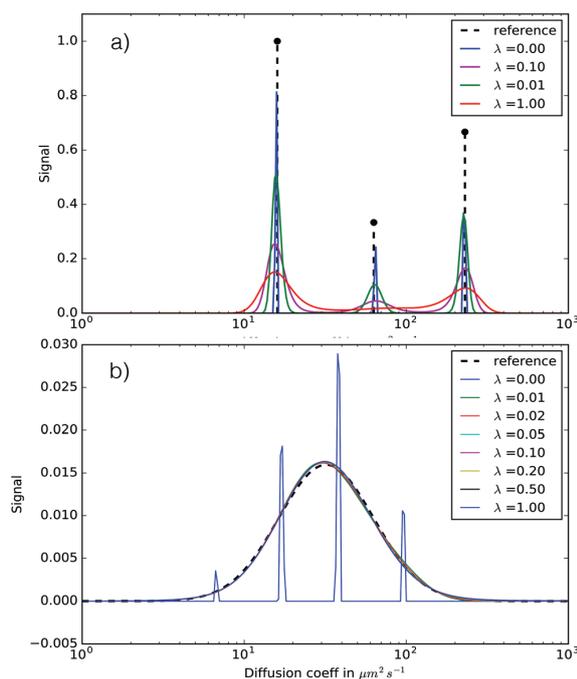


Fig. 1 PALMA reconstruction for various values of λ of the simulated experiments, with an added 0.1% Gaussian noise. (a) Experiment A with 3 monodisperse species, indicated by the black vertical lines, and the $\lambda = 0$ spectrum was divided by 3 for clarity; (b) experiment B with a large polydisperse Gaussian profile $\lambda = 0$ spectrum was divided by 8 for clarity.

reconstruction error was obtained for $\lambda = 0.05$. The tests performed on asymmetric distributions (see ESI Fig. S5 and S6 and Table S1†) show the same tendency, with stable results for all non-null values of λ . It should be noted that, since the algorithm maintains the analysis within the noise distance of the data, *i.e.* $\|HX - Y\| \lesssim \eta$ (see eqn (5)) all the reconstructed Laplace spectra fit equally well the data. They differ only in how they match the regularisation term, a term which holds and expresses the *a priori* information we have on the dataset.

The PALMA algorithm was tested against ITAMeD,²⁵ ℓ_p tailored-ITAMeD²⁶ and TRAIN²⁸ algorithms, using the same simulated data as above. Table 1 presents the synthetic results,

Table 1 Quality of reconstruction of signal B with different algorithms for various noise levels. Quality is computed as $\frac{\|X_{sim}\|}{\|X_{sim} - X_{calc}\|}$ expressed in dB. For each noise level, the highest quality results are given in *italics*

Algorithm	Noise level			
	1%	0.1%	0.01%	0.001%
ITAMeD	3.37	18.65	29.04	29.40
ITAMeD with ℓ_p	6.06	25.26	36.69	37.08
TRAIN	<i>24.75</i>	<i>28.63</i>	<i>26.53</i>	<i>19.47</i>
PALMA $\lambda = 0.01$	20.54	28.57	41.69	53.25
PALMA $\lambda = 0.05$	24.01	<i>32.51</i>	<i>48.28</i>	51.37

and the extensive results are presented in the ESI (see Fig. S7 to S14 and Tables S2 and S3†).

In our hands, PALMA and TRAIN present the best results in terms of faithfulness and robustness, in particular for polydisperse datasets, with TRAIN showing better results on the C2 dataset, while PALMA behaving better for data sampled with a small number of data points.

4.2 Application to polydisperse polymers

It is acknowledged that experimental data are quite different from the simulated data, with a mixture of sharp and large diffusion distribution, tainted with instrumental artifacts and non-stationary noise. With respect to the behaviour of the method on polydisperse systems, the program was first applied on DOSY experiments measured from poly-ethylene-oxide (PEO) polymers in water with calibrated chain lengths and polydispersity. Fig. 2 presents the results obtained for three PEO samples with polydispersity ranging from 1.07 to 2.0 measured in a standard manner, and processed with PALMA. The polydispersity of the different samples can be clearly seen in the profile widths. Sample (a) is a standard PEO polymer, with a rather low polydispersity. Sample (c) was prepared from a set of rather monodisperse polymers, in order to cover regularly a wide range of chain lengths. Sample (b) on the other hand is a PEO polymer given to have a standard polydispersity, however the details of the composition are not known.

4.3 Application to plant extract

To test the robustness of the approach, it was applied to crude ethanolic plant extract obtained from brown algae.

Fig. 3 shows the aliphatic region of the DOSY experiment performed on this algae extract. Only the more abundant species are visible at this plot level. This kind of analysis on complex mixtures has been extensively used to analyse natural

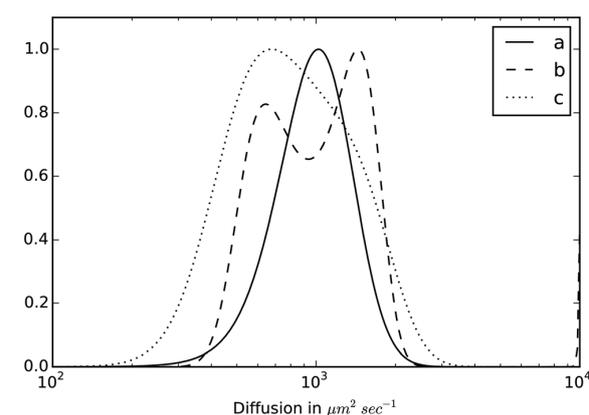


Fig. 2 DOSY profile of the main NMR signal of different standard PEO samples, (a) a reference PEO with a PDI of 1.07, (b) a reference PEO with a PDI of 1.28, and (c) a mixture of reference PEOs with a global PDI of 2.01.

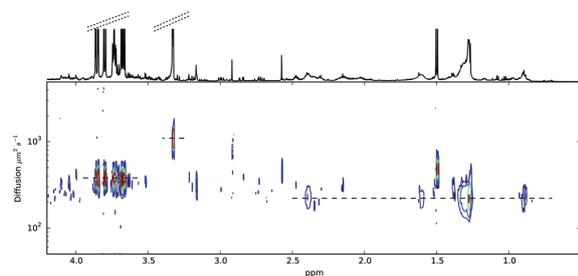


Fig. 3 DOSY experiment on a brown algae methanol/water extract showing only the major constituents. Dashed horizontal label lines are indicated for fatty acid chains ($220 \mu\text{m}^2 \text{s}^{-1}$), glycerol and short polyol ($380 \mu\text{m}^2 \text{s}^{-1}$), and methanol ($1100 \mu\text{m}^2 \text{s}^{-1}$).

products, plant extract,⁴³ or even adulteration of herbal and dietary supplements.⁴⁴ Here the presence of a particular molecule or a family of molecules is characterised by the alignment at the same diffusion coefficient of the characteristic lines located at their corresponding chemical shift positions. For instance, in Fig. 3, the polyol and fatty acid signals are outlined. The fatty acids are certainly partly aggregated in this sample, as indicated by the width of the line, both along the spectral and Laplace axes. The DOSY is useful in this context because it provides a high dynamic analysis, where the most intense lines do not “hide” less intense ones.

In Fig. 4, the chloroquine molecule was added at a low concentration (1% w/w) to the same sample. While the aromatic signals of chloroquine, located in a rather empty region of the spectrum, are easily detected, the aliphatic chain signals fall in the crowded region where they are difficult to observe. The methyl groups that fall at positions 1.3 ppm and 1.45 ppm are completely buried under the fatty acid signals and the diffusion coefficients are not different enough. In contrast, the signals from the methylene moieties are observed around

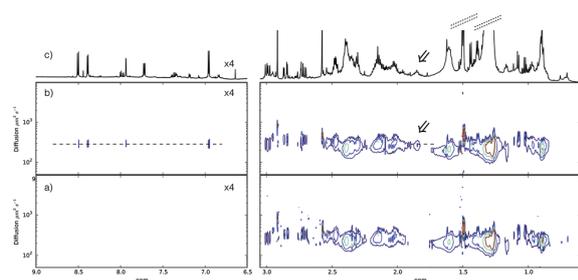


Fig. 4 Comparison of two DOSY experiments on brown algae extracts. (a) The aromatic (empty) and aliphatic regions of the same experiment as shown in Fig. 3, but plotted at a level four times lower; (b) the same brown algae extract with 0.16 mg mL^{-1} of chloroquine added, showing the aromatic signals the methyl signal outlined with an arrow in the spectrum, and the dashed horizontal label is at $285 \mu\text{m}^2 \text{s}^{-1}$; (c) the 1D spectrum of the brown algae extract with added chloroquine. In all three spectra, the aromatic panel is plotted four times lower than the aliphatic panel.

1.9 ppm in a relatively free spectral region. Despite being about $\times 100$ smaller than the larger signals (methanol, polyol or fatty acid chains), the signal is well separated in the DOSY spectrum, and is aligned with the aromatic signals.

One can also observe the reproducibility of the PALMA-processed DOSY spectra, as the main features of the spectra are nearly identical for both samples.

5. Discussion

The hybrid regularisation proposed in eqn (6) implements two well-known approaches, namely the ℓ_1 regularisation, which tends to minimise the number of non-null signals required to explain the data, and the MaxEnt regularisation which is usually presented as a way to maximise the posterior probability of the analysis while preserving the positivity of the retrieved spectra.⁴⁵ While both regularisations are well established and based on clear principles, the MaxEnt is known to be somewhat more difficult to implement. The hybrid regularisation proposed here allows to obtain very robust results even in the case of complex signals such as the one presented in the experimental examples.

The constrained problem is solved using a new convex optimisation algorithm, based on the use of proximity operators and a split version of the Douglas–Rachford procedure. The use of the proximity operators allows implementation of a simple incremental step, requiring no inner line-search minimisation step, where the main burden is three applications of the linear operator \mathbf{H} or of its generalised inverse $\mathbf{B} = (\mathbf{I} + \mathbf{H}^t\mathbf{H})^{-1}$. This algorithm allows a rapid convergence even with the hybrid regularisation used here. In the simulation presented here, an approximate solution is obtained very rapidly (in less than a second). The results presented in this work were obtained with longer convergences, using typically 10 000 to 100 000 iterations, however thanks to the rapidity of the iterative step, this corresponds typically to a few seconds on a laptop.

Because of the constrained approach used here, there is no need to determine a Lagrangian parameter as in most other techniques (sometimes called a smoothing parameter). Nevertheless, the approach requires some parameters, of which the prior a and the noise threshold η can readily be estimated from the experimental dataset, using respectively the first point of the decay and an estimate of the noise level from a polynomial smoothing of the experimental dataset (see the ESI section for details†). The weight λ between the MaxEnt and the ℓ_1 regularisations embodies a prior assumption on the presence of sparse components in the Laplace spectrum. In the simulations of a sparse theoretical spectrum (Fig. 1a) $\lambda = 0$ corresponding to a pure ℓ_1 regularisation provides the best reconstruction as expected. Simulations performed on several wide distributions (Fig. 1b and ESI Fig. S4, S6, S15, and S16†) show that the method recovers faithfully the position of the signal and the theoretical profile, for most non-null values of the λ parameter, even in the presence of noise. The results are

more contrasted for the sparse spectrum in Fig. 1a, where the pure MaxEnt analysis ($\lambda = 1.0$) presents large features located at the position of the three components, and a narrowing of the MaxEnt distribution for decreasing value of λ can be observed. The results of pure MaxEnt analysis produced by the PALMA algorithm were checked to be equivalent to the results obtained with the original algorithm²³ based on a fixed point approach, however the convergence is much faster, and processing times are about 10 times shorter for the same results. It should be recalled that the MaxEnt analysis produces a statistical analysis of the data, where the final spectrum is the density distribution which maximises the posterior probability of finding a signal.^{45,46} The width of the actual signals can thus be considered as an uncertainty on the position of the monodisperse components, uncertainty that is present in the pure ℓ_1 case as errors in the position of the lines, but not directly manifest. These results suffer however from a lack of resolution power, and the possibility to bias toward a sparser result is certainly a plus. In a general approach, the optimal value for λ should be chosen from assumptions on the data based on explicit previous knowledge, however on a practical point, this is not feasible. Even on a sample known to be composed solely of monodisperse species, the choice of a null λ is problematic. On the one hand, some polydisperse impurities might be present with the risk of overlooking them as we observe in Fig. 2 and 3; on the other hand, most instrument imperfections such as temperature drift, gradient non-linearity, phase distortions, convection, *etc.* will distort the pure exponential decay and create some apparent polydispersity. Confronted with the same difficulty Urbańczyk *et al.*²⁶ chose to vary the p parameter of their tailored-ITAMeD algorithm, somewhat similar to λ and chose the larger value which allows a minimal residual. The same approach could easily be used here, however considering the fact that polydisperse samples are correctly analysed for most non-null values of λ , with the better results obtained for small values, we suggest using values between 0.01 and 0.05 as monodisperse data are well described with these values.

The quality of these results is in sharp contrast to equivalent analyses presented in the literature. The $\lambda = 1$ mode reproduces the classical MaxEnt regularisation²³ although in less processing time. In the $\lambda = 0$ mode, compared to the ℓ_1 based ITAMeD approach,²⁵ the final resolution resulting from PALMA reconstruction is much higher, as observed in our simulations, in agreement with what has been published. Intermediate values of λ , creating a bias of the MaxEnt solution toward more sparse data, produce more resolved spectra, which usually better match the patterns expected in solution NMR.

To summarize, we recommend using a value of λ in the range 0.01 to 0.05 for safer results, with the possibility to adapt this value in particular cases (for instance extreme polydispersity or spectral superposition). We do not recommend using the pure ℓ_1 mode ($\lambda = 0$), even in the case of monodisperse samples, because the instrumental fluctuations mentioned above certainly disturb this pure behaviour, and

because of the difficulty to estimate the uncertainty of the result from a sparse spectrum alone.

6. Conclusion

The DOSY experiment holds a special position among the different experiments available to NMR spectroscopists. Whilst it provides invaluable information on the size and the interaction of the molecules in solution, it presents an important challenge for its acquisition and its analysis. The acquired data are usually corrupted by many artefacts produced by the very high sensitivity of this experiment to instrument imperfections such as temperature drift, non-linearity of the gradients and of the detection electronics, phase distortions, convection, *etc.* In parallel, while the parameter dependency expressed in the basic evolution equations is extremely simple, it is well established that a simple fit of the data to this equation usually fails to provide a faithful analysis of the data in the general case, and one has to resort to the inverse Laplace transform for the analysis step, a problem well known to be of extreme noise sensitivity. For these reasons, the stability and robustness of the acquisition and processing schemes are of great importance for the quality of DOSY spectroscopy, and many acquisition schemes and many processing procedures have already been proposed in the literature for this purpose. In this work, we have introduced a general method to solve the inverse problem as found in the analysis of DOSY experiments that we believe provides an unequalled level of quality and robustness in the processing step.

The method is based on a constrained regularisation of the least squares problem, and we showed that a hybrid regularisation, combining the maximum sparsity and Maximum Entropy, provides optimal results in the general case. This approach is controlled by a single parameter λ weighting between these two criteria, and the results are not very sensitive to the exact value of this parameter as long as extreme values are not chosen. This method, which we called PALMA, is faster and more robust than previous MaxEnt implementation and provides better results. It requires only weak assumptions from the user, and can be run in a fully automatic manner. It has been implemented on a server freely available at <http://palma.labo.igbmc.fr>, where users may submit their datasets for automatic processing. The code of the algorithm is available at the same address.

Acknowledgements

This work was supported by the CNRS MASTODONS project under grant 2016TABASCO, by the Agence Nationale pour la Recherche, grant ANR2014 ONE_SHOT_2D_FT_ICR. The authors thank Jean-Christophe Pesquet for the initial idea and for discussion all over this project. We thank Catherine Vonthron-Sénécheau, Mélanie Bourgeot and Laure Marguerite for the plant extract sample. We thank Samuel Nicaise for the first version of the WEB server interface and Julien Seiler for helping in the deployment of the WEB server.

References

- 1 E. O. Stejskal and J. E. Tanner, *J. Chem. Phys.*, 1965, **42**, 288–292.
- 2 D. Sinnaeve, *Concepts Magn. Reson., Part A*, 2012, **40**, 39–65.
- 3 C. Johnson, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1999, **34**, 203–256.
- 4 G. A. Morris, Reference Deconvolution, *eMagRes*, 2007.
- 5 W. S. Price, *NMR Studies of Translational Motion: Principles and Applications*, Cambridge University Press, Cambridge, UK, 2009.
- 6 P. T. Callaghan, *Translational dynamics and magnetic resonance: principles of pulsed gradient spin echo NMR*, Oxford University Press, Oxford, UK, 2011.
- 7 P. Stilbs, K. Paulsen and P. Griffiths, *J. Phys. Chem.*, 1996, **100**, 8180–8189.
- 8 L. Van Gorkom and T. Hancewicz, *J. Magn. Reson.*, 1998, **130**, 125–130.
- 9 W. Windig and B. Antalek, *Chemom. Intell. Lab. Syst.*, 1997, **37**, 241–254.
- 10 G. Armstrong, N. Loening, J. Curtis, A. Shaka and V. Mandelshtam, *J. Magn. Reson.*, 2003, **163**, 139–148.
- 11 R. Huo, R. Wehrens and L. Buydens, *J. Magn. Reson.*, 2004, **169**, 257–269.
- 12 M. Nilsson and G. Morris, *Anal. Chem.*, 2008, **80**, 3777–3782.
- 13 P. Stilbs, *J. Magn. Reson.*, 2010, **207**, 332–336.
- 14 B. R. Martini, V. A. Mandelshtam, G. A. Morris, A. A. Colbourne and M. Nilsson, *J. Magn. Reson.*, 2013, **234**, 125–134.
- 15 D. Nuzillard, S. Bourgand and J.-M. Nuzillard, *J. Magn. Reson.*, 1998, **133**, 358–363.
- 16 W. Naanaa and J.-M. Nuzillard, *Signal Process.*, 2005, **85**, 1711–1722.
- 17 A. Colbourne, G. Morris and M. Nilsson, *J. Am. Chem. Soc.*, 2011, **133**, 7640–7643.
- 18 I. Toumi, B. Torrèsani and S. Caldarelli, *Anal. Chem.*, 2013, **85**, 11344–11351.
- 19 J. Viéville, M. Tanty and M.-A. Delsuc, *J. Magn. Reson.*, 2011, **212**, 169–173.
- 20 N. H. Williamson, M. Nydén and M. Röding, *J. Magn. Reson.*, 2016, **267**, 54–62.
- 21 S. W. Provencher, *Comput. Phys.*, 1982, **27**, 213–227.
- 22 R. Nityananda and R. Narayan, *Astron. Astrophys.*, 1982, **3**, 419–450.
- 23 M.-A. Delsuc and T. E. Malliavin, *Anal. Chem.*, 1998, **70**, 2146–2148.
- 24 J. Skilling and R. Bryan, *Mon. Not. R. Astron. Soc.*, 1984, **211**, 111.
- 25 M. Urbańczyk, D. Bernin, W. Kozminski and K. Kazimierzuk, *Anal. Chem.*, 2013, **85**, 1828–1833.
- 26 M. Urbańczyk, D. Bernin, A. Czuroń and K. Kazimierzuk, *Analyst*, 2016, **141**, 1745–1752.
- 27 K. Kazimierzuk and V. Y. Orekhov, *Angew. Chem., Int. Ed.*, 2011, **50**, 5556–5559.
- 28 K. Xu and S. Zhang, *Anal. Chem.*, 2013, **86**, 592–599.
- 29 *Bayesian Approach to Inverse Problems*, ed. J. Idier, ISTE, 2008.
- 30 M. Elad, *Sparse and Redundant Representations*, Springer, New York, NY, 2010.
- 31 H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, New York, NY, 2011.
- 32 A. Beck and M. Teboulle, *SIAM J. Imaging. Sci.*, 2009, **2**, 183–202.
- 33 A. Cherni, E. Chouzenoux and M.-A. Delsuc, *Proc. 8th Int. Symp. Signal, Image, Video and Commun*, 2016, pp. x–x+6.
- 34 R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey and D. E. Knuth, *Adv. Comput. Math.*, 1996, **5**, 329–359.
- 35 P. L. Combettes and J.-C. Pesquet, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer, New York, NY, 2011, pp. 185–212.
- 36 N. Pustelnik, C. Chau and J.-C. Pesquet, *IEEE Trans. Image Process.*, 2011, **20**, 2450–2462.
- 37 P. L. Combettes and J.-C. Pesquet, *Inverse Probl.*, 2008, **24**, 564–574.
- 38 P. L. Combettes and J.-C. Pesquet, *IEEE J. Sel. Top. Signal Process.*, 2007, **1**, 564–574.
- 39 C. Vonthron-Sénécheau, M. Kaiser, I. Devambe, A. Vastel, I. Mussio and A.-M. Rusig, *Mar. Drugs*, 2011, **9**, 922–933.
- 40 A. Jerschow and N. Müller, *J. Magn. Reson.*, 1997, **125**, 372–375.
- 41 O. Assemat, M.-A. Coutouly, R. Hajjar and M.-A. Delsuc, *C. R. Chim.*, 2010, **13**, 412–415.
- 42 L. Chiron, M.-A. Coutouly, J.-P. Starck, C. Rolando and M.-A. Delsuc, 2016, arXiv, 1608.06777, 1–13.
- 43 R. Novoa-Carballal, E. Fernandez-Megia, C. Jimenez and R. Riguera, *Nat. Prod. Rep.*, 2011, **28**, 78–98.
- 44 S. Balayssac, S. Trefi, V. Gilard, M. Malet-Martino, R. Martino and M.-A. Delsuc, *J. Pharm. Biomed. Anal.*, 2009, **50**, 602–612.
- 45 E. Chouzenoux, S. Moussaoui, J. Idier and F. Mariette, *IEEE Trans. Signal Process.*, 2010, **58**, 6040–6051.
- 46 E. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.

FAST DICTIONARY-BASED APPROACH FOR MASS SPECTROMETRY DATA ANALYSIS

Cherni Afef^{1,3}, Chouzenoux Émilie^{1,2}, Delsuc Marc-André³

¹ Université Paris-Est Marne-la-Vallée, LIGM, UMR CNRS 8049, Champs-sur-Marne, France.

² Centre pour la Vision Numérique, CentraleSupélec, INRIA Saclay, Gif-sur-Yvette, France.

³ Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U596, UMR CNRS 7104, Université de Strasbourg, Illkirch-Graffenstaden, France.

ABSTRACT

Mass spectrometry (MS) is a fundamental technology of analytical chemistry for measuring the structure of molecules, with many application fields such as clinical biomarker analysis or pharmacokinetics. In the context of proteomic analysis with MS, the superposition of the isotopic patterns of different proteins, in various charge-states produces MS spectra difficult to decipher. The complexity of the pattern models and the large size of the data again increase the difficulty of the analysis step. In this paper, we propose to formulate the problem of proteins characterization as the estimation of a positive-valued sparse signal thanks to a dictionary-based approach relying on the protein average concept. A proximal primal-dual splitting convex optimization method is considered to solve the resulting variational problem. Moreover, the large size of the dictionary matrix is circumvented by proposing a suitable block circulant approximation of it, allowing to limit the computational burden of the method. Numerical experiments on synthetic and real MS datasets illustrate the good performance of our approach.

Index Terms— Mass spectrometry, Proteomic Analysis, Sparsity, Dictionary-based strategy, Primal-Dual algorithm.

1. INTRODUCTION

Mass Spectrometry (MS) is a powerful tool used for robust, accurate, and sensitive detection and quantification of molecules of interest. Thanks to its sensibility and selectivity, MS is widely used in proteomics such anti-doping, metabolomics, medicine or structural biology [1, 2]. In particular, it has applications in clinical research [3], personalized medicine [2], diagnosis process and tumours profiling [4] and pharmaceutical quality control [5]. In a MS experiment, the raw signal arising from the molecule ionization in an ion beam is measured as a function of time via Fourier Transform-based measures such as Ion Cyclotron Resonance (FT-ICR) and Orbitrap. A spectral analysis step is then performed, possibly involving a series of operations/algorithms [6, 7] to improve the quality of data, transforming the time-domain data into the frequency domain. The frequency spectrum is then converted to the so-called MS spectrum through a calibration function. This spectrum presents a set of positive-valued peaks distributed according to the charge state and the isotopic distribution of the studied molecule, generating several typical patterns in the signal. The goal is then to determine from

this observed pattern distribution the most probable chemical composition of the sample, through the determination of the monoisotopic mass, charge state and abundance of each present molecule. Unfortunately, the superposition of the isotopic patterns in different charge-states can produce MS spectra difficult to decipher, and the complexity of the problem again increases with the number and the size of molecules. Proteins, the family of molecules that will be targeted in this work, present a particular challenge since they are rather large molecules with wide isotopic patterns, ionized by ElectroSpray Ionization (E.S.I) leading to a mixture of several charge states. Additionally, the usual high resolution and thus very large size of the measurements in this context (usually > 500k data point) make their analysis cumbersome. Peak-peaking and pattern recognition approaches are the most common methods currently used for MS spectrum analysis in the context of proteins samples [8, 9, 10]. Although these methods may be quite fast, they suffer from slow performance, instability and sensitivity to high noise level since they all require a preprocessing step to threshold the data. In particular, their performance can be highly degraded when a strong peak overlap masks the position of other peaks, which usually happens when several distinct proteins and/or several charge states are in presence [9].

In this paper, we propose a new dictionary-based approach to solve efficiently and automatically this problem. We start by introducing in Section 2 the chemical problem statement, the measurement model and our dictionary-based strategy. In Section 3, we formulate the selection of the dictionary elements as the resolution of a convex non-smooth optimization problem, and describe a primal-dual proximal algorithm to solve it efficiently. In Section 4, we discuss the practical implementation of our method, and propose a block-circulant approximation of the dictionary matrix to reduce the computational cost of the processing. Finally, an illustration of the good performance of our method for recovering synthetic and real data in the context of Mass Spectrometry is presented in Section 5.

2. PROBLEM STATEMENT

2.1. Some reminders on chemistry

An atom is the basic unit of matter and the smallest defining structure of elements, defined in the periodic table with a symbol and a nucleon number. An atom can be present under different forms with different numbers of neutron, called isotopes. Each stable isotope is present in the nature with a specific abundance. Each unique molecule presents thus a specific mass in Daltons, depending on the sum of the masses of each of its constituting isotopes and on its charge state z .

This work was supported by the CNRS MASTODONS project under grant 2016TABASCO and by the Agence Nationale pour la Recherche (ANR, France) under grant 2010FT-ICR2D and grant Défi de tous savoirs 2014, ONE-SHOT-FT-ICR-MS-2D).

When a large number of samples with various charge states is considered, for instance when measuring a MS spectrum, a distribution of peaks, named multi-charged isotopic pattern, is observed on the mass over charge (i.e. (m/z)) axis, following the composition of the elementary distributions of all atoms [11, 12] and their charge distribution [13]. For a fixed charge state $z > 0$, the smallest (m/z) position of the peaks is associated with the most abundant isotope mass and allows to determine the monoisotopic mass at charge z . This quantity is independent on the relative isotopic abundances, and helps in a non-ambiguous determination of the molecule. However, for large molecules, the probability of having a single charge state and no isotopes is extremely low so that the peak intensity at the monoisotopic masses can be vanishingly small and their direct detection impossible.

The purpose of this work is to provide an automatic tool to characterize monoisotopic mass and charge state quantities from the measured MS data, in the context of proteomic analysis, i.e. when the chemical sample to be studied is made of several proteins. We recall that a protein is a large molecule having the generic formula $C_{N_C}H_{N_H}O_{N_O}N_{N_N}S_{N_S}$ where $(N_C, N_H, N_O, N_N, N_S)$ are respectively the number of Carbon C, Hydrogen H, Oxygen O, Nitrogen N and Sulfur S. The most probable isotopes of the latter atoms, along with their mass at the neutral state and their associated abundances, are given in Table 1. Note that for $z > 0$, as it is the case in MS data, the mass values are all shifted proportionally to z and to the mass of the hydrogen adduct ions [13].

Atom	Mass (in Dalton)	Relative Abundance
^{12}C	12 (<i>by definition</i>)	0.9893
^{13}C	13.0033548378	0.0107
^1H	1.00782503207	0.999885
^2H	2.0141017778	0.000115
^{16}O	15.99491461956	0.99757
^{17}O	16.99913170	0.00038
^{18}O	17.9991610	0.00205
^{14}N	14.0030740048	0.99636
^{15}N	15.0001088982	0.00364
^{32}S	31.97207100	0.9499
^{33}S	32.97145876	0.0075
^{34}S	33.96786690	0.0425

Table 1: Isotopic mass and natural abundance of atoms found in proteins [14, 15].

2.2. Observation model

Let us consider a given chemical sample, composed of P different proteins with monoisotopic mass $m_p^{\text{iso}} \in (0, +\infty)$, charge state $z_p \in \mathbb{N}^*$ and abundance $a_p \in (0, +\infty)$, for $p \in \{1, \dots, P\}$. The acquired MS spectrum y can be modeled as the weighted sum of each individual isotopic pattern $y = \sum_{p=1}^P a_p D(m_p^{\text{iso}}, z_p) + n$ where n models the acquisition noise and possible errors arising from the spectral analysis preprocessing step. The measurements are taken on a discrete grid of (m/z) values with size M , so that the observation model finally reads:

$$\mathbf{y} = \sum_{p=1}^P a_p \mathbf{d}(m_p^{\text{iso}}, z_p) + \mathbf{n} \quad (1)$$

with $\mathbf{y} \in \mathbb{R}^M$, $\mathbf{d}(m_p^{\text{iso}}, z_p) \in [0, +\infty]^M$ and $\mathbf{n} \in \mathbb{R}^M$. The aim of the MS spectrum analysis is then to reconstruct the set of coefficients $(a_p, m_p^{\text{iso}}, z_p)_{1 \leq p \leq P}$ from \mathbf{y} . The lack of knowledge of the

complicated and nonlinear function $\mathbf{d}(m, z)$ and the large value of M present the main locks of this problem.

2.3. Proposed dictionary-based strategy

The mass distribution function $D(m, z)$, at a given (m, z) is actually easy to evaluate from the molecular formula using a recursive program [11]. We thus propose to adopt a dictionary-based approach for solving the estimation problem, under the assumption that we know approximately the range of mass and charge for the P proteins present in the sample. To do so, let us define a search grid, with size $T := MZ$ which defines M possible values of isotopic masses and Z possible values for the charges. From this grid, we build the dictionary $\mathbf{D} \in \mathbb{R}^{M \times T}$ so that each column $i \in \{1, \dots, T\}$ of \mathbf{D} is $\mathbf{d}(m_i, z_i)$ where (m_i, z_i) is the couple charge-mass in the i -th position of the grid. Then, the problem is reformulated as:

$$\mathbf{y} = \mathbf{D}\bar{\mathbf{x}} + \mathbf{n}' \quad (2)$$

where $\bar{\mathbf{x}}$ is a sparse vector with positive entries, for which the P non-zeros coefficients allow to determine the isotopic mass and charge state of each protein, along with their abundance. Moreover, $\mathbf{n}' = \mathbf{n} + \mathbf{e}$ models the acquisition noise and possible errors arising from the spectral analysis and discretization steps ($\mathbf{e} \rightarrow 0$ with high accuracy).

3. OPTIMIZATION STRATEGY

3.1. Variational formulation

Because of the presence of noise in \mathbf{y} and the ill-conditioning of the dictionary matrix \mathbf{D} , direct inversion is not suitable to find an estimate of $\bar{\mathbf{x}}$. We propose instead to employ a penalized approach that defines the estimate $\hat{\mathbf{x}} \in \mathbb{R}^T$ as a solution of the constrained minimization problem:

$$\underset{\mathbf{x} \in \mathbb{R}^T}{\text{minimize}} \Phi(\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{D}\mathbf{x} - \mathbf{y}\| \leq \eta \quad (3)$$

where $\Phi: \mathbb{R}^T \mapsto (-\infty, +\infty]$ is a proper, lower semicontinuous, convex regularization function used to enforce positivity and sparsity on the solution, and $\eta > 0$ is a parameter that depends on the noise characteristics. When P is unknown, as it is usually the case in practical MS experiments, a simple choice for Φ is to consider, for every $\mathbf{x} \in \mathbb{R}^T$, $\Phi(\mathbf{x}) = \sum_{i=1}^T \max(0, x_i)$. More sophisticated penalties can also be used, involving for instance block-sparsity regularizers [16], or entropy-like priors [17].

3.2. Primal-dual optimization strategy

To resolve Problem (3), we propose to use the proximal Primal-Dual Splitting algorithm from [18] which is an efficient algorithm for convex optimization. In particular, this approach allows to treat efficiently the non-necessarily differentiable function Φ , and does not require any inversion step on the linear operator \mathbf{D} [19].

Hereabove, $\text{prox}_{\tau\Phi}(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^T$, $\tau > 0$, states for the proximity operator of function $\tau\Phi$ at \mathbf{x} [20] which is defined as the unique minimizer of $\tau\Phi + 1/2\|\cdot - \mathbf{x}\|^2$ [21]. Moreover, the projection operator $\text{proj}_{\|\cdot - \mathbf{y}\| \leq \eta}$ is defined, for every $(\mathbf{y}, \mathbf{v}) \in (\mathbb{R}^N)^2$, as:

$$\text{proj}_{\|\cdot - \mathbf{y}\| \leq \eta}(\mathbf{v}) = \mathbf{v} + (\mathbf{v} - \mathbf{y}) \min\left(\frac{\eta}{\|\mathbf{v} - \mathbf{y}\|}, 1\right) - \mathbf{y}. \quad (4)$$

The convergence of the iterates $(\mathbf{x}^{(k)})_{k \in \mathbb{N}}$ to a solution of Problem (3) is ensured, according to [18, 22].

Algorithm 1 Primal-Dual Splitting Algorithm

Initialization
 $\mathbf{u}^{(0)} \in \mathbb{R}^M, \mathbf{x}^{(0)} \in \mathbb{R}^T$
 $0 < \sigma < \|\mathbf{D}\|^2/\tau, \rho \in (0, 2), \tau > 0$
Minimization
For $k = 0, 1, \dots$

$$\begin{cases} \tilde{\mathbf{x}}^{(k)} = \text{prox}_{\tau\Phi}(\mathbf{x}^{(k-1)} - \tau\mathbf{D}^\top(\mathbf{u}^{(k-1)})) \\ \mathbf{v}^{(k)} = \mathbf{u}^{(k-1)} + \sigma\mathbf{D}(2\tilde{\mathbf{x}}^{(k)} - \mathbf{x}^{(k-1)}) \\ \tilde{\mathbf{u}}^{(k)} = \mathbf{v}^{(k)} - \sigma\text{proj}_{\|\cdot\| \leq \eta}(\mathbf{v}^{(k)}/\sigma) \\ \mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + \rho(\tilde{\mathbf{x}}^{(k)} - \mathbf{x}^{(k-1)}) \\ \mathbf{u}^{(k)} = \mathbf{u}^{(k-1)} + \rho(\tilde{\mathbf{u}}^{(k)} - \mathbf{u}^{(k-1)}) \end{cases}$$

4. PRACTICAL IMPLEMENTATION

We now discuss the practical implementation of our approach in the applicative context of MS spectrum analysis. In particular, the question of the computation and storage of matrix \mathbf{D} is raised, and an approximated strategy is proposed for dealing with very large dimension.

4.1. Dictionary construction

As already mentioned, for any given (m, z) , it is possible to estimate precisely the isotopic distribution $D(m, z)$ of an average protein having a monoisotopic mass m and charge state z using the so-called average model [23]. Given a range of masses $[m_{\min}, m_{\max}]$ and charges $[z_{\min}, z_{\max}]$, we define a regular grid:

$$(\forall i \in \{1, \dots, T\}) \quad m_i = m_{\min} + (j-1)m_{\max}, \quad (5)$$

$$z_i = z_{\min} + (\ell-1)z_{\max}, \quad (6)$$

with the convention $i = \ell M + j$, $j \in \{1, \dots, M\}$ and $\ell \in \{1, \dots, Z\}$. Then, the i -th column of \mathbf{D} is taken as $\mathbf{d}(m_i, z_i)$ which corresponds to a sampled version of $D(m, z)$ on the mass grid with size M . Here, we propose to perform this sampling in the Fourier domain, and we normalize the result, so as to preserve the sum of squared amplitudes from $D(m, z)$ to $\mathbf{d}(m, z)$.

4.2. Circulant approximation

Mass spectrometry aims at providing a very high mass accuracy, so that the value of M can be very large. Even for small Z , the large number of columns of \mathbf{D} presents a computational challenge as large memory resources may be needed to store this matrix. In order to avoid such memory issues, we propose an approximation $\bar{\mathbf{D}}$ of \mathbf{D} whose structure will allow the use of Fourier transform operations for computing the products of \mathbf{D} and \mathbf{D}^\top with vectors. Our approximation relies on the important facts that (i) the isotopic patterns for similar mass values mainly differs by a simple shift of peaks positions, (ii) these patterns are sparse with non-zero elements located in a limited range of indexes near the monoisotopic mass value. Therefore, an alternative to the storage of isotopic patterns for each mass/charge couple that was proposed in Section 4.1, is to decompose the mass axis into windows onto which the isotopic pattern is assumed to be constant up to a circular shift. Let us introduce the notation $\mathbf{D} = [\mathbf{D}_1 | \dots | \mathbf{D}_\ell | \dots | \mathbf{D}_Z]$ where, for every $\ell \in \{1, \dots, Z\}$, $\mathbf{D}_\ell \in \mathbb{R}^{M \times M}$ maps for the dictionary associated to charge $z_{\min} + (\ell-1)z_{\max}$. Let $L \leq M$ the chosen window width and $\bar{\mathbf{d}}_{s,\ell}$ the average isotopic pattern for a mass within the range $[(s-1)L + 1, sL]$, and a fixed charge state $z_{\min} + (\ell-1)z_{\max}$.

We propose to approximate each \mathbf{D}_ℓ by the following block diagonal (BDiag) matrix made of M/L blocks assumed to be circulant (Circ) matrices with first line $\bar{\mathbf{d}}_{s,\ell}$, $s \in \{1, \dots, M/L\}$:

$$\bar{\mathbf{D}}_\ell = \text{BDiag} \left([\text{Circ}(\bar{\mathbf{d}}_{s,\ell})]_{1 \leq s \leq M/L} \right). \quad (7)$$

As a consequence, for every charge value, the products $\bar{\mathbf{D}}_\ell$ and $\bar{\mathbf{D}}_\ell^\top$ with vectors can be easily computed using Fourier operations. Under this approximation, Algorithm 1 can still be used to estimate the mass and charge positions, where \mathbf{D} has now been replaced by $\bar{\mathbf{D}} = [\bar{\mathbf{D}}_1 | \dots | \bar{\mathbf{D}}_\ell | \dots | \bar{\mathbf{D}}_Z]$, and the norm of $\bar{\mathbf{D}}$ is computed using power iteration.

5. EXPERIMENTAL RESULTS

5.1. Synthetic data

In order to evaluate the performance of our method, we simulate two synthetic sparse signals A and B. Signal A models a mono-charge MS spectrum with $M = 3000$, $Z = 1$, $z_{\min} = z_{\max} = 1$ and $P = 10$ proteins. Signal B represents a multi-charge MS spectrum with $M = 5000$, containing $Z = 3$ charge values with $z_{\min} = 1$, $z_{\max} = 3$, and $P = 50$ proteins. In both cases, $m_{\min} = 1000$ Daltons, and $m_{\max} = 1100$ Daltons. Noisy data are then created using the linear model (2), where the noise is assumed to be zero-mean Gaussian, i.i.d, with known standard deviation $\sigma > 0$. We solve (3) with the penalty function Φ described in Section 3.1, and $\eta = \theta\sigma\sqrt{M}$ where $\theta > 0$ is a weight closed to 1.

Tab. 2 presents the results obtained for several noise levels. The quality of the results is evaluated in terms of signal to noise ratio (SNR) in dB defined as $10 \log_{10} (\|\bar{\mathbf{x}}\|^2 / \|\bar{\mathbf{x}} - \hat{\mathbf{x}}\|^2)$, and in terms of the number of recovered peaks \hat{P} . As one can observe, the exact dictionary approach and its block-circulant approximation both allow to recover the required number of peaks with good SNR values even for high noise level. The reconstruction quality can also be confirmed by a visual inspection of Fig. 1 which displays the input MS spectrum, the exact signal and the result of our method (using \mathbf{D} or $\bar{\mathbf{D}}$) in the case of the multi-charged dataset B, and $\sigma = 10^{-2}$. Note that the signals are represented here along the abscissa axis (m/z), which is a standard representation in the context of MS. We also provide the computation time when running our method on a Macintosh Mac Pro Intel Xeon with a total of 8 cores, equipped with 12 GB of memory and running MacOSX 10.7.5. Algorithm 1 is used with $\rho = 1.99$, $\tau = \|\mathbf{D}\|^{-1}$, $\sigma = 0.9\tau$, and it is stopped whenever the relative error between two consecutive iterates is lower than 10^{-8} . In our tests, a maximum of 1000 iterations was sufficient to reach such precision. As expected, when using the block-circulant approximation of the dictionary matrix, the computation time depends on the value of L , the fastest reconstruction being obtained with a large L . Moreover, even if the reconstruction quality is slightly deteriorated, the SNR values remain stable as soon as L is sufficiently large as it can be observed on Fig. 2. According to our practical observations, the value of L mainly impacts on the peak intensity values, and has little effect on the estimation quality of peak positions, which is of main interest in the MS application. We also provide in Tab. 2 the memory required for the storage of the dictionary elements when using the exact matrix \mathbf{D} . As expected, the memory requirement in that case may be quite high, especially in the multi-charged case (dataset B). In contrast, the block-circulant approximation we proposed allows to avoid any dictionary storage, since the products with $\bar{\mathbf{D}}$ and its adjoint are performed using Fourier operators where the first lines of each circulant term are computed on the fly.

Dataset	σ	Exact dictionary approach			Block-circulant approximation													
		SNR	Time	\hat{P}	$L = 2$			$L = 10$			$L = 50$			$L = 100$				
					SNR	Time	\hat{P}	SNR	Time	\hat{P}	SNR	Time	\hat{P}	SNR	Time	\hat{P}		
Dataset A	P=10	1	15.07	10.18	10	72	15.05	15.16	10	11.31	3.31	10	13.88	0.95	10	11.37	0.61	10
		0.1	35.09	10.40	10		34.98	15.22	10	31.95	3.32	10	32.60	0.27	10	32.28	0.24	10
		0.01	38.38	10.46	10		34.33	15.14	10	34.44	3.35	10	34.92	0.94	10	36.91	0.61	10
Dataset B	P=50	1	16.18	303.33	50	600	15.57	127.85	50	13.99	27.20	50	14.01	6.98	50	15.59	3.52	50
		0.1	35.73	206.84	50		35.43	44.48	50	33.95	8.19	50	33.53	3.09	50	28.26	2.06	50
		0.01	39.56	377.80	50		38.38	290.56	50	34.40	58.92	50	35.44	11.66	50	28.99	16.67	50

Table 2: SNR (in dB), computation time (in s), memory storage (in Mb) for matrix \mathbf{D} and number \hat{P} of detected peaks for the restored signals A and B for various values of noise level. Block-circulant approximation $\bar{\mathbf{D}}$ is tested for four L values.

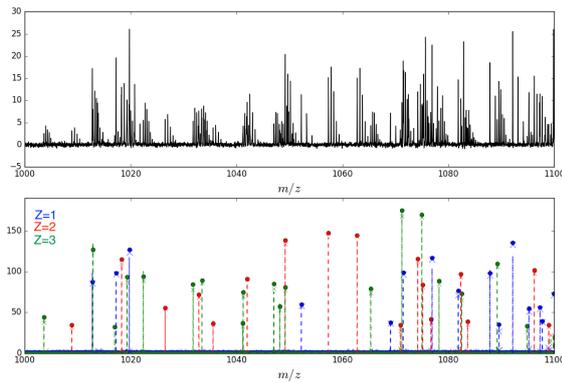


Fig. 1: Reconstruction results of the signal B with $\sigma = 10^{-2}$: (top) input data \mathbf{y} , (bottom) exact spectrum (dots), restored spectrum with exact dictionary (dashed line), and with its block-circulant approximation for $L = 10$ (asterisks).

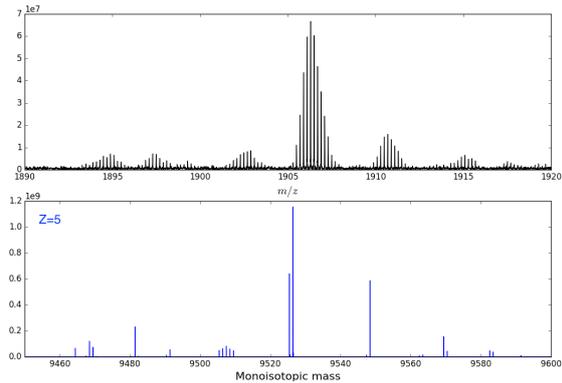


Fig. 3: Analysis of the real FT-ICR-MS spectrum of a peptide in trimer form: (top) zoom on the acquired data; (bottom) recovered spectrum at $z = 5$ using block-circulant approximated dictionary with $L = 10$.

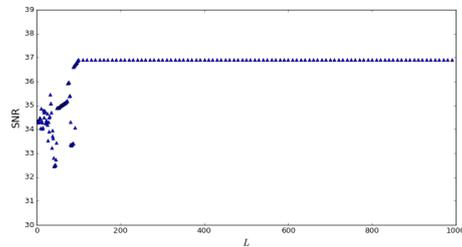


Fig. 2: Reconstruction results of the signal A with $\sigma = 10^{-2}$: SNR of the restored spectrum using $\bar{\mathbf{D}}$ for various values of L .

5.2. Real data

We also perform numerical tests on a real MS dataset measured on a Bruker Solarix 15 T, FT-ICR instrument with an ESI source. The considered sample was constituted of $3 \mu\text{M}$ of the peptide EVEALEKKVAALSKVQALEKKVEALEHG-NH2 ($\text{C}_{140}\text{H}_{240}\text{N}_{38}\text{O}_{45}$) in its trimer form within 50 mM of NH_4OAc , acquired in native conditions. The input data is of size $M = 8130981$ with $m_{\min} = 153.57$ Daltons, $m_{\max} = 4999.96$ Daltons, and $Z = 5$ with $z_{\min} = 1$ and $z_{\max} = 5$. Parameter η is set in a similar manner than for the synthetic data, the noise level σ being estimated from an empty frame of the measured signal.

Fig. 3(top) displays a zoom on the input signal \mathbf{y} for m/z within [1890, 1920] Daltons. We also show on Fig. 3(bottom) the recovered

signal along the monoisotopic mass axis, for the same mass range, and $z = 5$ charge state. It is worth mentioning that, due to the very large size of the dataset, such results were only made possible by the block-circulant approximation we proposed, since the storage of a dictionary would have required GigaBytes of memory. Here, the best visual results were obtained using $L = 10$, with a processing time of 108 minutes. A major peak can be distinguished at $m = 9526.439$ Daltons which fits well with the theoretical monoisotopic mass of the studied peptide equals to $m = 9526.337$ Daltons. A second peak, shifted by -1 Dalton, is also observed, due to unavoidable grid ambiguity. Finally, we observe a third important peak distant with +21.959 Daltons of the main peak, which allows to identify the sodium adduct (with theoretical relative position of +21.982 Daltons), thus validating the faithfulness of our approach.

6. CONCLUSION

This work presents a new dictionary-based approach based on the average function to solve the isotopic pattern analysis problem arising in Mass Spectrometry of proteins. We propose a sparsity-aware variational strategy to determine the dictionary elements, associated with a primal-dual splitting minimization strategy. To counteract any computation burden, we propose a suitable block-circulant approximation of the dictionary. Our experimental results illustrate the efficiency of our method to solve the MS problem. Future work will address the extension of the approach to the processing of multi-dimensional MS spectra [24, 25].

7. REFERENCES

- [1] L.-M. Heaney, D.J. Jones, and T. Suzuki, "Mass spectrometry in medicine: a technology for the future?," *Future Science OA*, vol. 3, no. 3, June 2017.
- [2] I.-E. Sodal, "The medical mass spectrometer," *Biomedical Instrumentation & Technology*, vol. 23, no. 6, pp. 469–476, 1989.
- [3] O. Gaillot, N. Blondiaux, C. Loiez, F. Wallet, N. Lemaitre, S. Herwegh, and R. Courcol, "Cost-effectiveness of switch to matrix-assisted laser desorption ionization-time of flight mass spectrometry for routine bacterial identification," *Journal of Clinical Microbiology*, vol. 49, no. 12, pp. 4412–4412, 2011.
- [4] S. A. Schwartz, R. J. Weil, M. D. Johnson, S. A. Toms, and R. M. Caprioli, "Protein profiling in brain tumors using mass spectrometry," *Clinical Cancer Research*, vol. 10, no. 3, pp. 981–987, 2004.
- [5] F. J. Belas and I. A. Blair, "Mass spectrometry in pharmaceutical analysis," *Journal of liposome research*, vol. 11, no. 4, pp. 309–342, 2001.
- [6] A.-G. Ferrige, M.-J. Seddon, B.-N. Green, S.-A. Jarvis, J. Skilling, and J. Staunton, "Disentangling electrospray spectra with maximum entropy," *Rapid Communications in Mass Spectrometry*, vol. 6, no. 11, pp. 707–711, 1992.
- [7] A. Mohammad-Djafari, J.-F. Giovannelli, G. Demoment, and J. Idier, "Regularization, maximum entropy and probabilistic methods in mass spectrometry data processing problems," *International Journal of Mass Spectrometry*, vol. 215, no. 1, pp. 175 – 193, 2002, Detectors and the Measurement of Mass Spectra.
- [8] J. Samuelsson, D. Dalevi, F. Levander, and T. Rognvaldsson, "Modular, scriptable and automated analysis tools for high-throughput peptide mass fingerprinting," *Bioinformatics*, vol. 20, no. 18, pp. 3628–3635, 2004.
- [9] B. Y. Renard, M. Kirchner, H. Steen, J. A. J. Steen, and F. A. Hamprecht, "Nitpick: peak identification for mass spectrometry data," *BMC Bioinformatics*, vol. 9, no. 1, pp. 355, 2008.
- [10] D. P. A. Kilgour, S. Hughes, S. L. Kilgour, C. L. Mackay, M. Palmblad, B. Q. Tran, Y. A. Goo, R. K. Ernst, D. J. Clarke, and D. R. Goodlett, "Autopicker - a robust and reliable peak detection algorithm for mass spectrometry," *Journal of The American Society for Mass Spectrometry*, vol. 28, no. 2, pp. 253–262, Feb. 2017.
- [11] J.-A. Yergey, "A general approach to calculating isotopic distributions for mass spectrometry," *International Journal of Mass Spectrometry and Ion Physics*, vol. 52, no. 2-3, pp. 337–349, 1983.
- [12] P. Kaur and P.-B. O'Connor, "Use of statistical methods for estimation of total number of charges in a mass spectrometry experiment," *Analytical Chemistry*, vol. 76, no. 10, pp. 2756–2762, 2004.
- [13] M. Mann, C. K. Meng, and J. B. Fenn, "Interpreting mass spectra of multiply charged ions," *Analytical Chemistry*, vol. 61, no. 15, pp. 1702–1708, 1989.
- [14] G. Audi and A.-H. Wapstra, "The 1993 atomic mass evaluation:(i) atomic mass table," *Nuclear Physics A*, vol. 565, no. 1, pp. 1–65, 1993.
- [15] K.-J.-R. Rosman and P.-D.-P. Taylor, "Table of isotopic masses and natural abundances," *Pure and Applied Chemistry*, vol. 71, pp. 1593–1607, 1999.
- [16] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Structured sparsity through convex optimization," *Statistical Science*, vol. 27, no. 4, pp. 450–468, 2012.
- [17] A. Cherni, E. Chouzenoux, and Delsuc M.-A., "Proximity for a class of hybrid sparsity + entropy prior. Application to DOSY NMR signal reconstruction," in *Proceedings of the 8th International Symposium on Signal, Image, Video and Communications (ISIVC 2016)*, Tunis, Tunisia, 21-23 Nov. 2016, pp. x–x+6.
- [18] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.
- [19] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 31–54, Oct. 2015.
- [20] H.H. Bauschke and P.L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, CMS Books in Mathematics. Springer, New York, NY, 2011.
- [21] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société mathématique de France*, vol. 93, pp. 273–299, 1965.
- [22] L. Condat, "A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460–479, Aug. 2013.
- [23] M.-W. Senko, J.-P. Speir, and F.-W. McLafferty, "Collisional activation of large multiply charged ions using Fourier transform mass spectrometry," *Analytical Chemistry*, vol. 66, no. 18, pp. 2801–2808, 1994.
- [24] M.-A. Agthoven, Delsuc M.-A., G. Bodenhausen, and C. Rolando, "Towards analytically useful two-dimensional fourier transform ion cyclotron resonance mass spectrometry," *Analytical and Bioanalytical Chemistry*, vol. 405, pp. 51–61, Oct. 2013.
- [25] F. Floris, M.-V. Agthoven, L. Chiron, A.-J. Soulby, C. A. Wootton, Y.-P.-Y. Lam, M.-P. Barrow, M.-A. Delsuc, and P.-B. O'Connor, "2D FT-ICR MS of Calmodulin: A Top-Down and Bottom-Up Approach.," *Journal of the American Society for Mass Spectrometry*, vol. 27, no. 9, pp. 1531–1538, Sep. 2016.

Approche de Douglas-Rachford aléatoire par blocs appliquée à la régression logistique parcimonieuse

Giovanni CHERCHIA¹, Afef CHERNI^{1,2}, Emilie CHOUZENOUX^{1,3}, Jean-Christophe PESQUET³

¹Université Paris Est, LIGM UMR 8049, CNRS, ENPC, ESIEE Paris, UPEM, Noisy-le-Grand, France.

²IGBMC, CNRS UMR 7104, Inserm U 964, Illkirch-Graffenstaden, France.

³Centre pour la Vision Numérique, CentraleSupélec, Université Paris-Saclay, Châtenay-Malabry, France.

Résumé – Dans cet article, nous proposons un nouvel algorithme d’optimisation stochastique pour la régression logistique parcimonieuse, basé sur une version aléatoire par blocs de l’algorithme de Douglas-Rachford. Notre algorithme parcourt la base d’apprentissage en sélectionnant aléatoirement un bloc de données à chaque itération, puis réalise l’étape de mise à jour en utilisant l’opérateur proximal de la fonction de coût logistique, dont nous donnons une expression explicite. Les tests réalisés sur plusieurs jeux de données démontrent l’efficacité de notre algorithme par rapport aux méthodes de type gradient stochastique.

Abstract – We propose a stochastic optimization algorithm for logistic regression based on a randomized version of Douglas-Rachford splitting method. Our algorithm sweeps the training set by randomly selecting a mini-batch of data at each iteration, and it performs the update step by leveraging the proximity operator of the logistic loss, for which a closed-form expression is derived. Experiments carried out on standard datasets compare the efficiency of our algorithm to stochastic gradient-like methods.

1 Introduction

Les méthodes d’apprentissage parcimonieux sont devenues très populaires dans le contexte de l’apprentissage supervisé, de par leur capacité à éliminer les caractéristiques non pertinentes pendant la phase d’entraînement. Ces approches visent à apprendre une combinaison linéaire pondérée de fonctions de base liant les données d’entraînement, en forçant autant de poids que possible à être nuls. Ceci revient à résoudre un problème d’optimisation qui met en jeu une fonction de coût et une régularisation parcimonieuse. Différents classifieurs sont obtenus en faisant varier la fonction de coût, les plus populaires étant la fonction *hinge* et la fonction logistique [20].

Dans le cadre de l’apprentissage supervisé, l’utilisation de régularisations parcimonieuses remonte aux travaux de Bradley et Mangasarian [3], qui ont montré qu’une sélection efficace des fonctions de bases était obtenue grâce à un seuillage à zéro des faibles poids, ce qui revient à l’utilisation d’une régularisation ℓ_1 . D’autres formes de régularisations ont été également étudiées [10, 14].

De nombreux algorithmes d’apprentissage existent pour traiter le cas de la régularisation quadratique, exploitant la forme simple du dual du problème au sens de Lagrange [2]. Cependant, une telle approche perd son avantage en présence d’une régularisation parcimonieuse, car le problème dual devient dans ce cas aussi difficile à traiter que le primal. Les classifieurs linéaires parcimonieux sont

donc habituellement entraînés en résolvant directement le problème d’optimisation primal associé. Parmi les algorithmes possibles, nous pouvons mentionner les méthodes de majoration-minimisation [12] et les approches proximales [1, 5, 10].

En pratique, l’usage de règles de mise à jour stochastiques permet de réduire significativement le temps de calcul des méthodes d’apprentissage supervisé, lorsqu’une régularisation quadratique est employée [2]. De nombreux travaux récents ont visé à généraliser ces méthodes d’optimisation stochastiques pour pouvoir traiter des régularisations favorisant la parcimonie [18]. Citons, en particulier, les stratégies de minimisation par blocs [22], les approches explicites-implicites stochastiques [8, 21], les algorithmes primaux-duaux stochastiques [9, 19], les techniques de descente duale stochastiques [24], et les méthodes de majoration-minimisation stochastiques [6, 13].

Parmi toutes ces approches, les algorithmes primaux-duaux stochastiques sont en fait les seules techniques permettant de s’affranchir de l’hypothèse de différentiabilité de la fonction de coût. Ils permettent en effet de traiter toute fonction de coût convexe par son opérateur proximal, en supposant que celui-ci puisse être calculé efficacement. C’est le cas de la fonction *hinge*, mais pas de la régression logistique, pour laquelle il n’existe pas d’expression explicite dans la littérature.

Dans cet article, nous proposons un algorithme proximal stochastique pour la régression logistique parcimonieuse,

basé sur un schéma de Douglas-Rachford. Notre approche traite la fonction de coût logistique par son opérateur proximal, ce qui contraste avec les méthodes existantes utilisant plutôt son gradient. De ce fait, l'algorithme proposé n'est pas contraint par la valeur de la constante de Lipschitz du gradient de la fonction de coût, ce qui, par conséquent, peut conduire à une convergence plus rapide. D'un point de vue théorique, notre contribution principale est de montrer que l'opérateur proximal de la fonction logistique peut en fait s'exprimer sous une forme quasi explicite, et ainsi être calculé efficacement. Des comparaisons avec plusieurs approches d'optimisation stochastique de l'état de l'art sont réalisées sur des bases de données de référence.

L'article s'organise comme suit. Dans la section 2, nous explicitons la formulation du problème de régression logistique parcimonieuse. Dans la section 3, nous décrivons l'algorithme proposé, et dérivons l'expression de l'opérateur proximal de la fonction logistique. Dans la section 4, nous évaluons les performances de notre méthode, et la comparons avec trois algorithmes de classification parcimonieuse de la littérature [4, 21, 24]. Enfin, la section 5 conclut ce travail et dresse quelques perspectives.

Notations: $\Gamma_0(\mathcal{H})$ représente l'ensemble des fonctions propres, convexes, semi-continues inférieurement d'un espace de Hilbert \mathcal{H} vers $\mathbb{R} \cup \{+\infty\}$. Pour tout $\nu \in \mathcal{H}$, le sous-différentiel de $\psi \in \Gamma_0(\mathcal{H})$ en ν est noté $\partial\psi(\nu) = \{\xi \in \mathcal{H} \mid (\forall \zeta \in \mathcal{H}) \langle \zeta - \nu \mid \xi \rangle + \psi(\nu) \leq \psi(\zeta)\}$.

2 Régression logistique

Un classifieur linéaire binaire vise à associer, à chaque donnée en entrée $x \in \mathbb{R}^N$, une valeur binaire $y \in \{-1, +1\}$ via un modèle linéaire de la forme:

$$d_w(x) = \text{signe}(x^\top w). \quad (1)$$

En apprentissage supervisé, le vecteur de poids $w \in \mathbb{R}^N$ défini ci-dessus est estimé à partir d'un ensemble de paires entrée/sortie:

$$\mathcal{S} = \{(x_\ell, y_\ell) \in \mathbb{R}^N \times \{-1, +1\} \mid \ell \in \{1, \dots, L\}\}, \quad (2)$$

qu'on appelle la *base d'entraînement*.

Une approche très populaire en apprentissage pour effectuer la tâche de prédiction se base sur la notion de risque structurel. L'objectif est de réaliser un compromis entre la fidélité à la base d'entraînement et la réduction de la complexité du modèle, en résolvant le problème d'optimisation suivant:

$$\underset{w \in \mathbb{R}^N}{\text{minimiser}} f(w) + \sum_{\ell=1}^L h(y_\ell x_\ell^\top w), \quad (3)$$

où $f \in \Gamma_0(\mathbb{R}^N)$ est une fonction de régularisation, tandis que $h \in \Gamma_0(\mathbb{R})$ est la fonction de coût.

De nombreux choix ont été proposés dans la littérature pour la fonction de régularisation. Dans le contexte

de l'apprentissage parcimonieux, un choix courant est d'utiliser la norme ℓ_1 , $f = \lambda \|\cdot\|_1$, avec $\lambda > 0$. Pour la fonction de coût, là encore, plusieurs choix sont possibles, tels que, dans le cas de la classification, la fonction *hinge*, la fonction *hinge* carrée, ou le coût logistique, et, dans le cas de la régression, le coût quadratique ou le coût de Huber [20]. Dans ce travail, nous nous intéressons plus particulièrement au coût logistique, défini par :

$$(\forall v \in \mathbb{R}) \quad h(v) = \log(1 + \exp(-v)). \quad (4)$$

3 Méthode d'optimisation

La résolution du problème (3) requiert un algorithme d'optimisation capable de traiter un terme de régularisation non différentiable favorisant la parcimonie. Dans ce contexte, les méthodes proximales apparaissent comme une solution de choix, de par leur capacité à gérer efficacement des problèmes convexes non lisses [7, 17]. Ces méthodes reposent sur un outil fondamental, appelé *l'opérateur proximal* [16], défini pour une fonction $\psi \in \Gamma_0(\mathcal{H})$ comme

$$(\forall \zeta \in \mathcal{H}) \quad \text{prox}_\psi(\zeta) = \underset{\xi \in \mathcal{H}}{\text{argmin}} \frac{1}{2} \|\xi - \zeta\|^2 + \psi(\xi). \quad (5)$$

L'opérateur proximal peut être interprété comme une étape de sous-gradient implicite, sur la fonction ψ . En effet, $\nu = \text{prox}_\psi(\zeta)$ est défini de façon unique par l'inclusion $\zeta - \nu \in \partial\psi(\nu)$. De nombreux algorithmes proximaux existent dans la littérature, permettant de résoudre différentes classes de problèmes. Dans ce travail, nous proposons d'utiliser une version aléatoire par blocs de l'algorithme proximal de Douglas-Rachford, introduite dans [8]. Un aspect remarquable de cet algorithme est qu'il permet une mise à jour par blocs des variables du problème. Ici, nous nous intéressons à une mise à jour partielle de variables auxiliaires qui nous permet de travailler à chaque itération sur un sous-ensemble des données de la base d'entraînement, plutôt que sur la base complète. Une telle stratégie se rapproche en ce sens des méthodes de gradient incrémental / stochastique reconnues pour leur efficacité dans les problèmes d'apprentissage supervisé.

Nous obtenons, dans le cadre de la résolution du Problème (3), l'algorithme 1. A chaque itération i de la méthode, un sous-ensemble \mathbb{L}_i de $\{1, \dots, L\}$ est sélectionné, de façon aléatoire, et seules les composantes $(v_\ell^{[i]})_{\ell \in \mathbb{L}_i}$ de la variable auxiliaire v sont mises à jour, à partir des données $(x_\ell, y_\ell)_{\ell \in \mathbb{L}_i}$ extraites de la base. Notons que le calcul de la matrice Q définie dans l'algorithme 1 est généralement peu coûteux dans les applications en apprentissage, la dimension N étant souvent petite devant L . Les approches proximales présentent l'avantage de posséder des propriétés de convergence solides, même dans un contexte stochastique. Dans notre cas, nous pouvons déduire du résultat de [8] que, si la probabilité d'activation de chaque variable est non nulle et indépendante de i , sous des hypothèses

de qualification du problème, la suite $(w^{[i]})_{i \in \mathbb{N}}$ générée par l'algorithme 1 converge presque sûrement vers une solution du Problème (3).

Algorithm 1 Algorithme de Douglas-Rachford aléatoire par blocs

INITIALISATION

$$\left[\begin{array}{l} \text{Calculer } Q = \left(\text{Id} + \sum_{\ell=1}^L x_{\ell} x_{\ell}^{\top} \right)^{-1} \\ \text{Choisir } t^{[0]} \in \mathbb{R}^N \text{ et } (v_1^{[0]}, \dots, v_L^{[0]}) \in \mathbb{R}^L \\ \text{Fixer } u^{[0]} = \sum_{\ell=1}^L y_{\ell} x_{\ell} v_{\ell}^{[0]}. \\ \text{Choisir } \gamma \in]0, +\infty[\text{ et } \mu \in]0, 2[\end{array} \right.$$

POUR $i = 0, 1, \dots$

$$\left[\begin{array}{l} \text{Sélectionner } \mathbb{L}_i \subset \{1, \dots, L\} \\ w^{[i]} = Q(t^{[i]} + u^{[i]}) \\ t^{[i+1]} = t^{[i]} + \mu \left(\text{prox}_{\gamma_i f}(2w^{[i]} - t^{[i]}) - w^{[i]} \right) \\ (\forall \ell \in \mathbb{L}_i) \quad v_{\ell}^{[i+1]} = v_{\ell}^{[i]} \\ \quad \quad \quad + \mu \left(\text{prox}_{\gamma_i h}(2y_{\ell} x_{\ell}^{\top} w^{[i]} - v_{\ell}^{[i]}) - y_{\ell} x_{\ell}^{\top} w^{[i]} \right) \\ (\forall \ell \notin \mathbb{L}_i) \quad v_{\ell}^{[i+1]} = v_{\ell}^{[i]} \\ u^{[i+1]} = u^{[i]} + \sum_{\ell \in \mathbb{L}_i} (v_{\ell}^{[i+1]} - v_{\ell}^{[i]}) y_{\ell} x_{\ell}. \end{array} \right.$$

A chaque itération de l'algorithme 1, il est nécessaire d'évaluer les opérateurs proximaux de f et h . De ce fait, un calcul rapide de ces opérateurs est nécessaire pour garantir de bonnes performances numériques de l'algorithme. Un grand nombre de fonctions présentent une forme explicite pour leur opérateur proximal [7, 17]. Par exemple, si f est la norme ℓ_1 , son opérateur proximal se réduit à un simple seuillage doux. En ce qui concerne la fonction de coût logistique, quelques stratégies de calcul numérique ont été proposées [17, 23], mais à notre connaissance, une forme explicite de son opérateur proximal est inconnue. La contribution principale de cet article, présentée dans la proposition ci-dessous, permet de combler cette lacune.

Proposition 3.1. *Soit $\gamma \in]0, +\infty[$. L'opérateur proximal de la fonction de coût logistique (4) est*

$$(\forall v \in \mathbb{R}) \quad \text{prox}_{\gamma h}(v) = v + W_{\exp(-v)} \left(\gamma \exp(-v) \right), \quad (6)$$

où W . est la fonction de *W-Lambert généralisée* [11, 15].

La fonction de *W-Lambert généralisée* (aussi appelée fonction de *r-Lambert*) mentionnée ci-dessus permet de résoudre les équations transcendantes de la forme $p \exp(p) + rp = q$, où $p = W_r(q)$. Elle peut être évaluée de façon

rapide et très précise, par la méthode de type Newton développée par Mező *et al.* [15], disponible en ligne¹.

4 Résultats expérimentaux

Nous évaluons maintenant les performances de notre approche sur les bases de données MNIST² ($N = 717$ et $L = 60000$) et W8A³ ($N = 300$ et $L = 49749$). Nous comparons notre algorithme, avec les méthodes suivantes : l'algorithme stochastique explicite-implicite (SFB), l'approche duale régularisée moyennée (RDA) [24], et l'algorithme primal-dual par blocs (BCPD) [4]. Pour chacun de ces algorithmes, nous traçons l'évolution du critère défini en (3) en fonction du temps de calcul.

Les paramètres des différentes méthodes sont fixés manuellement afin d'optimiser leurs performances : le *taux d'apprentissage* de SFB et RDA est $\gamma = 10^{-4}$, tandis que $\gamma = 1$ et $\mu = 1.8$ pour l'algorithme 1. Des blocs de données de taille 1000 sont sélectionnés aléatoirement à chaque itération, suivant une loi uniforme, et une régularisation ℓ_1 pondérée par $\lambda = 1$ est employée. Tous les tests ont été menés sous Matlab 2015a, avec un CPU Intel i7 cadencé à 3.40 GHz et doté de 12 Go de RAM.

La Figure 1 illustre les performances des différentes méthodes, pour les deux jeux de données testés. Les résultats indiquent que notre méthode permet d'atteindre une convergence rapide vers une valeur faible de la fonction objectif. Ceci s'explique en partie par le rôle de préconditionneur joué par la matrice Q intégrée dans l'algorithme 1. Un point notable est que notre méthode laisse un choix libre pour les paramètres γ et μ . Cela contraste avec les méthodes SFB et RDA, où le paramètre γ doit être attentivement réglé à la main, sans quoi les méthodes deviennent très lentes, ou bien à l'inverse, divergent.

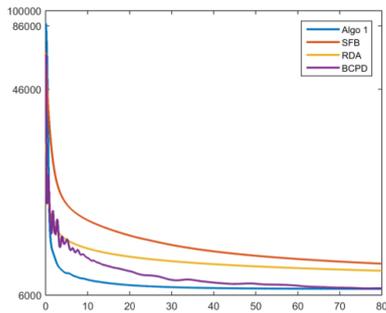
5 Conclusion

Dans cet article, nous avons proposé un algorithme de Douglas-Rachford aléatoire par blocs, pour résoudre un problème de régression logistique parcimonieuse. Notre approche se base sur l'opérateur proximal de la fonction de coût logistique, pour lequel nous avons obtenu une expression explicite facile à évaluer. Ces contributions nous ont permis de mettre en place une méthode de résolution rapide, et dont les paramètres de pas peuvent être choisis de manière flexible, sans risquer de faire diverger l'algorithme. Ceci constitue un avantage par rapport aux techniques d'optimisation stochastique usuelles en apprentissage. Nos résultats expérimentaux confirment les bonnes performances de notre algorithme.

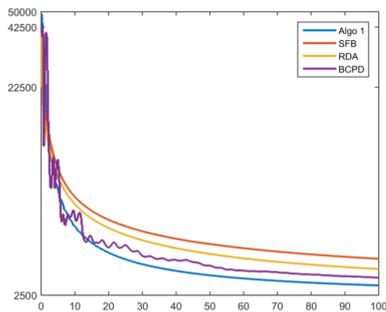
¹<https://sites.google.com/site/istvanmezo81/others>

²yann.lecun.com/exdb/mnist

³www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html



(a) Dataset W8A : $N = 300$ et $L = 49749$.



(b) Dataset MNIST : $N = 717$ et $L = 60000$.

Figure 1: Critère (3) en fonction du temps.

Remerciements

Ce travail a été financé par le défi CNRS MASTODONS 2016TABASCO.

Références

- [1] M. Barlaud, W. Belhajali, P. L. Combettes, and L. Fillatre. Classification and regression using a constrained convex splitting method. *IEEE Trans. Signal Process.*, 2017.
- [2] M. Blondel, A. Fujino, and N. Ueda. Large-scale multiclass support vector machine training via euclidean projection onto the simplex. In *Proc. of ICPR*, pages 1289–1294, Stockholm, Sweden, 24-28 August 2014.
- [3] P. S. Bradley and O. L. Mangasarian. Feature selection via concave minimization and support vector machines. In *Proc. of ICML*, pages 82–90, Madison, USA, 1998.
- [4] G. Chierchia, N. Pustelnik, and J.-C. Pesquet. Random primal-dual proximal iterations for sparse multiclass SVM. In *Proc. of MLSP*, Salerno, Italy, September 2016.
- [5] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu. A proximal approach for sparse multiclass SVM. *Preprint arXiv:1501.03669*, February 2015.
- [6] E. Chouzenoux and J.-C. Pesquet. A stochastic majorize-minimize subspace algorithm for online penalized least squares estimation. *IEEE Trans. Signal Process.*, 2017.
- [7] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, New York, 2011.
- [8] P. L. Combettes and J.-C. Pesquet. Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping. *SIAM J. Opt.*, 25(2):1221–1248, July 2015.
- [9] P.L. Combettes and J.-C. Pesquet. Stochastic approximations and perturbations in forward-backward splitting for monotone operators. *Pure and Applied Functional Analysis*, 1(1):13–37, January 2016.
- [10] L. Laporte, R. Flamary, S. Canu, S. Déjean, and J. Mothe. Non-convex regularizations for feature selection in ranking with sparse SVM. *IEEE Trans. Neural Netw. Learn. Syst.*, 25(6):1118 – 1130, June 2014.
- [11] A. Maignan and T. Scott. Fleshing out the generalized lambert W function. *ACM Communications in Computer Algebra*, 50(2):45–60, June 2016.
- [12] J. Mairal. Optimization with first-order surrogate functions. In *Proc. of ICML*, pages 783–791, 2013.
- [13] J. Mairal. Stochastic majorization-minimization algorithms for large-scale optimization. In *Proc. of NIPS*, pages 2283–2291, 2013.
- [14] L. Meier, S. Van De Geer, and P. Bühlmann. The group Lasso for logistic regression. *J. R. Stat. Soc. B*, 70(1):53–71, 2008.
- [15] I. Mező and Á. Baricz. On the generalization of the lambert W function. *Trans. Amer. Math. Soc.*, 2017.
- [16] J. J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.
- [17] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):123–231, 2014.
- [18] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. Hero, and S. McLaughlin. A survey of stochastic simulation and optimization methods in signal processing. *IEEE J. Sel. Topics Signal Process.*, 10(2):224–241, March 2016.
- [19] J.-C. Pesquet and A. Repetti. A class of randomized primal-dual algorithms for distributed optimization. *J. Nonlinear Convex Anal.*, 16(12):2453–2490, 2015.
- [20] L. Rosasco, E. De Vito, A. Caponnetto, M. Piana, and A. Verri. Are loss functions all the same? *Neural Comput.*, 16(5):1063–1076, May 2004.
- [21] L. Rosasco, S. Villa, and B. C. Vũ. Stochastic forward-backward splitting for monotone inclusions. *J. Optim. Theory Appl.*, 169(2):388–406, May 2016.
- [22] S. Shalev-Shwartz and A. Tewari. Stochastic methods for l_1 -regularized loss minimization. *J. Mach. Learn. Res.*, 12:1865–1892, June 2011.
- [23] P.-W. Wang, M. Wytock, and J. Z. Kolter. Epigraph projections for fast general convex programming. In *Proc. of ICML*, 2016.
- [24] L. Xiao. Dual averaging methods for regularized stochastic learning and online optimization. *J. Mach. Learn. Res.*, 11:2543–2596, October 2010.

Bibliographie

- [Abboud 2017] F. Abboud, E. Chouzenoux, J.-C. Pesquet, J.-H. Chenot and L. Laborelli. *Dual block-coordinate forward-backward algorithm with application to deconvolution and deinterlacing of video sequences*. Journal of Mathematical Imaging and Vision, vol. 59, no. 3, pages 415–431, 2017.
- [Aebersold 2003] R. Aebersold and M. Mann. *Mass spectrometry-based proteomics*. Nature, vol. 422, no. 6928, pages 198–207, 2003.
- [Aharon 2006] M. Aharon, M. Elad and A. Bruckstein. *rm k -SVD : An algorithm for designing overcomplete dictionaries for sparse representation*. IEEE Transactions on signal processing, vol. 54, no. 11, pages 4311–4322, 2006.
- [Amster 1996] I.-J. Amster. *Fourier transform mass spectrometry*. Journal of mass spectrometry, vol. 31, no. 12, pages 1325–1337, 1996.
- [Antoniadis 2002] A. Antoniadis, D. Leporini and J.-C. Pesquet. *Wavelet thresholding for some classes of non Gaussian noise*. Statistica neerlandica, vol. 56, no. 4, pages 434–453, 2002.
- [Attouch 2009] H. Attouch and J. Bolte. *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*. Mathematical Programming, vol. 116, no. 1-2, pages 5–16, 2009.
- [Attouch 2011] H. Attouch, J. Bolte and B.-F. Svaiter. *Convergence of descent methods for semi-algebraic and tame problems : proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*. Mathematical Programming, vol. 137, no. 1-2, pages 91–129, 2011.
- [Audi 1993] G. Audi and A.-H. Wapstra. *The 1993 atomic mass evaluation :(I) Atomic mass table*. Nuclear Physics A, vol. 565, no. 1, pages 1–65, 1993.
- [Bauschke 2011] H.-H. Bauschke, R.-S. Burachik, P.-L. Combettes, V. Elser, D.-R. Luke and H. Wolkowicz. *Fixed-point algorithms for inverse problems in science and engineering*, volume 49. Springer Science & Business Media, 2011.
- [Bauschke 2017] H.-H. Bauschke and P.-L. Combettes. *Convex analysis and monotone operator theory in hilbert spaces*, volume 2011. Springer, 2017.
- [Bertsekas 1999] D.-P. Bertsekas. *Nonlinear programming*. Athena scientific Belmont, 1999.
- [Bilen 2017] C. Bilen, A. Ozerov and P. Perez. *Solving Time Domain Audio Inverse Problems using Nonnegative Tensor Factorization*. 2017.
- [Boccacci 2002] P. Boccacci and M. Bertero. *Image-restoration methods : basics and algorithms*. Confocal and Two-Photon Microscopy : Foundations, Applications, and Advances, New York : Wiley-Liss, pages 253–269, 2002.
- [Bolte 2017] J. Bolte, T.-P. Nguyen, J. Peypouquet and B. Suter. *From error bounds to the complexity of first-order descent methods for convex functions*. Mathematical Programming, vol. 165, no. 2, pages 471–507, 2017.
- [Bray 2017] F. Bray, J. Bouclon, L. Chiron, M. Witt, M.-A. Delsuc and C. Rolando. *Nonuniform Sampling Acquisition of Two-Dimensional Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Increased Mass Resolution of Tandem Mass Spectrometry Precursor Ions*. Analytical chemistry, vol. 89, no. 17, pages 8589–8593, 2017.

- [Briceno 2017] A. Briceno, M. Luis, G. Chierchia, E. Chouzenoux and J.-C. Pesquet. *A Random Block-Coordinate Douglas-Rachford Splitting Method with Low Computational Complexity for Binary Logistic Regression*. arXiv preprint arXiv :1712.09131, 2017.
- [Bruckstein 2009] A.-M. Bruckstein, D.-L. Donoho and M. Elad. *From sparse solutions of systems of equations to sparse modeling of signals and images*. SIAM review, vol. 51, no. 1, pages 34–81, 2009.
- [Buck 1991] B. Buck and V-A. Macaulay. *Maximum entropy in action : a collection of expository essays*. Oxford University Press, USA, 1991.
- [Bullen 2013] P.-S Bullen. *Handbook of means and their inequalities*, volume 560. Springer Science and Business Media, 2013.
- [Cadzow 1988] J.-A. Cadzow. *Signal enhancement-a composite property mapping algorithm*. IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 36, no. 1, pages 49–62, 1988.
- [Candès 2006] E.-J. Candès, J. Romberg and T. Tao. *Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information*. IEEE Transactions on information theory, vol. 52, no. 2, pages 489–509, 2006.
- [Chambolle 2011] A. Chambolle and T. Pock. *A first-order primal-dual algorithm for convex problems with applications to imaging*. Journal of mathematical imaging and vision, vol. 40, no. 1, pages 120–145, 2011.
- [Chang 2014] X. Chang, Y. Wang, R. Li and Z. Xu. *Sparse K-Means with ℓ_∞/ℓ_0 Penalty for High-Dimensional Data Clustering*. arXiv preprint arXiv :1403.7890, 2014.
- [Chardon 2010] G. Chardon, A. Peillot, L Daudet and F. Ollivier. *Le Compressed sensing pour l’holographie acoustique de champ proche-I : Aspects algorithmiques et simulations*. In 10ème Congrès Français d’Acoustique, 2010.
- [Chaux 2007] C. Chaux, P.-L. Combettes, J.-C. Pesquet and V.-R. Wajs. *A variational formulation for frame-based inverse problems*. Inverse Problems, vol. 23, no. 4, page 1495, 2007.
- [Chen 2001] S.-S. Chen, D.-L. Donoho and M.-A. Saunders. *Atomic decomposition by basis pursuit*. SIAM review, vol. 43, no. 1, pages 129–159, 2001.
- [Cheng 2016] J. Cheng and M.-D. Sacchi. *Fast dual-domain reduced-rank algorithm for 3D deblending via randomized QR decomposition*. Geophysics, vol. 81, no. 1, pages V89–V101, 2016.
- [Cherni 2016] A. Cherni, E. Chouzenoux and M.-A. Delsuc. *Proximity operators for a class of hybrid sparsity+ entropy priors. application to DOSY NMR signal reconstruction*. International Symposium on Signal, Image, Video and Communications (ISIVC 2016), pages 816–820, 2016.
- [Cherni 2017] A. Cherni, E. Chouzenoux and M.-A. Delsuc. *PALMA, an improved algorithm for DOSY signal processing*. Analyst, vol. 142, no. 5, pages 772–779, 2017.
- [Cherni 2018] A. Cherni, E. Chouzenoux and M.-A. Delsuc. *Fast dictionary-based approach for mass spectrometry data analysis*. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2018), pages 1–5, 2018.
- [Chernushevich 2001] I.-V. Chernushevich, A.-V. Loboda and B.-A. Thomson. *An introduction to quadrupole-time-of-flight mass spectrometry*. Journal of Mass Spectrometry, vol. 36, no. 8, pages 849–865, 2001.

- [Chiang 2005] Y.-W. Chiang, P.-P Borbat and J.-H. Freed. *The determination of pair distance distributions by pulsed ESR using Tikhonov regularization*. Journal of Magnetic Resonance, vol. 172, no. 2, pages 279–295, 2005.
- [Chierchia 2017] G. Chierchia, A. Cherni, E. Chouzenoux and J.-C. Pesquet. *Approche de Douglas-Rachford aléatoire par blocs appliquée à la régression logistique parcimonieuse*. In GRETSI 2017, pages 1–4, 2017.
- [Chiron 2014] L. Chiron, M.-A. van Agthoven, B. Kieffer, C. Rolando and M.-A. Delsuc. *Efficient denoising algorithms for large experimental datasets and their applications in Fourier transform ion cyclotron resonance mass spectrometry*. Proceedings of the National Academy of Sciences, vol. 111, no. 4, pages 1385–1390, 2014.
- [Chiron 2016] L. Chiron, M.-A. Coutouly, J.-P. Starck, C. Rolando and M.-A Delsuc. *SPIKE a Processing Software dedicated to Fourier Spectroscopies*. arXiv preprint arXiv :1608.06777, 2016.
- [Chouzenoux 2010] E. Chouzenoux, S. Moussaoui, J. Idier and F. Mariette. *Efficient maximum entropy reconstruction of nuclear magnetic resonance T1-T2 spectra*. IEEE Transactions on Signal Processing, vol. 58, no. 12, pages 6040–6051, 2010.
- [Chouzenoux 2014] E. Chouzenoux, J.-C. Pesquet and A. Repetti. *Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function*. Journal of Optimization Theory and Applications, vol. 162, no. 1, pages 107–132, 2014.
- [Combettes 2005] P.-L. Combettes and V.-R. Wajs. *Signal recovery by proximal forward-backward splitting*. Multiscale Modeling & Simulation, vol. 4, no. 4, pages 1168–1200, 2005.
- [Combettes 2007] P.-L Combettes and J.-C. Pesquet. *A Douglas–Rachford splitting approach to nonsmooth convex variational signal recovery*. IEEE Journal of Selected Topics in Signal Processing, vol. 1, no. 4, pages 564–574, 2007.
- [Combettes 2011] P.-L. Combettes and J.-C. Pesquet. *Proximal splitting methods in signal processing*. In Fixed-point algorithms for inverse problems in science and engineering, pages 185–212. Springer, 2011.
- [Corless 1996] R.-M. Corless, G.-H. Gonnet, D.-E. Hare, D.-J. Jeffrey and D.-E. Knuth. *On the Lambert W function*. Advances in Computational mathematics, vol. 5, no. 1, pages 329–359, 1996.
- [Courtès 2015] C. Courtès, G. Dusson, R. Hatchi, R. Molina and A. Thomas. *Compressed sensing pour l'imagerie radar*. PhD thesis, 2015.
- [Davenport 2011] M.-A. Davenport, M.-F. Duarte, Y.-C. Eldar and G. Kutyniok. *Introduction to compressed sensing*. preprint, vol. 93, no. 1, page 2, 2011.
- [Delsuc 1989] M.-A. Delsuc. *A new maximum entropy processing algorithm, with applications to nuclear magnetic resonance experiments*. In Maximum entropy and bayesian methods, pages 285–290. Springer, 1989.
- [Delsuc 1998] M.-A. Delsuc and T.-E. Malliavin. *Maximum entropy processing of DOSY NMR spectra*. Analytical Chemistry, vol. 70, no. 10, pages 2146–2148, 1998.
- [Demagnet 2014] L. Demagnet and P. Hand. *Scaling law for recovering the sparsest element in a subspace*. Information and Inference : A Journal of the IMA, vol. 3, no. 4, pages 295–309, 2014.

- [Dettmer 2007] K. Dettmer, P.-A. Aronov and B.-D. Hammock. *Mass spectrometry-based metabolomics*. Mass spectrometry reviews, vol. 26, no. 1, pages 51–78, 2007.
- [Djafari 1988] M. Djafari and A.-G. Demoment. *Maximum entropy image reconstruction in X-ray and diffraction tomography*. IEEE transactions on Medical Imaging, vol. 7, no. 4, pages 345–354, 1988.
- [Djafari 1994] M. Djafari. *Maximum d'entropie et problèmes inverses en imagerie*. Signal, vol. 2, pages 87–116, 1994.
- [Dong 2014] C. Dong, C.-C. Loy, K. He and X. Tang. *Learning a deep convolutional network for image super-resolution*. In European conference on computer vision, pages 184–199. Springer, 2014.
- [Donoho 2003] D.-L. Donoho and M. Elad. *Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization*. Proceedings of the National Academy of Sciences, vol. 100, no. 5, pages 2197–2202, 2003.
- [Donoho 2006] D.-L. Donoho. *Compressed sensing*. IEEE Transactions on information theory, vol. 52, no. 4, pages 1289–1306, 2006.
- [Douglas 1956] J. Douglas and H.-H. Rachford. *On the numerical solution of heat conduction problems in two and three space variables*. Transactions of the American mathematical Society, vol. 82, no. 2, pages 421–439, 1956.
- [Einstein 1905] A. Einstein. *On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat*. Annalen der physik, vol. 17, pages 549–560, 1905.
- [Elad 2010] M. Elad. *Sparse and redundant representations from theory to applications in signal and image processing*. Springer New York Dordrecht Heidelberg London, 2010.
- [Floris 2016] F. Floris, M. van Agthoven, L. Chiron, A.-J. Soulby, C.-A. Wootton, Y.-P. Lam, M.-P. Barrow, M.-A. Delsuc and P.-B. O'Connor. *2D FT-ICR MS of calmodulin : a top-down and bottom-up approach*. Journal of The American Society for Mass Spectrometry, vol. 27, no. 9, pages 1531–1538, 2016.
- [Foucart 2013] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*, volume 1. Birkhäuser Basel, 2013.
- [Gagneux 1995] G. Gagneux and T.-M. Madaune. *Analyse mathématique de modèles non linéaires de l'ingénierie pétrolière*, volume 22. Springer Science & Business Media, 1995.
- [Galatsanos 1992] N.-P. Galatsanos and A.-K. Katsaggelos. *Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation*. IEEE Transactions on image processing, vol. 1, no. 3, pages 322–336, 1992.
- [Gavrilyuk 2017] A.-L. Gavrilyuk, D.-A. Osinkin and D.-I. Bronin. *The use of Tikhonov regularization method for calculating the distribution function of relaxation times in impedance spectroscopy*. Russian Journal of Electrochemistry, vol. 53, no. 6, pages 575–588, 2017.
- [Gerchberg 1974] R.-W. Gerchberg. *Super-resolution through error energy reduction*. Optica Acta : International Journal of Optics, vol. 21, no. 9, pages 709–720, 1974.
- [Gilard 2008] V. Gilard, S. Trefi, S. Balayssac, M.-A. Delsuc, T. Gostan, M.-M. Malet, R. Martino, Y. Prigent and F. Taulelle. *DOSY NMR for drug analysis*. In NMR spectroscopy in pharmaceutical analysis, pages 269–289. Elsevier, 2008.
- [Golub 2012] G.-H. Golub and C.-F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.

- [Gomez 2008] M.-V. Gomez, J. Guerra, A.-H. Velders and R.-M. Crooks. *NMR characterization of fourth-generation PAMAM dendrimers in the presence and absence of palladium dendrimer-encapsulated nanoparticles*. Journal of the American Chemical Society, vol. 131, no. 1, pages 341–350, 2008.
- [Goodfellow 2016] I. Goodfellow, Y. Bengio, A. Courville and Y. Bengio. Deep learning, volume 1. MIT press Cambridge, 2016.
- [Gostan 2004] T. Gostan, D. Tramesel, E. Brun, Y. Prigent and M.-A. Delsuc. *L'expérience DOSY, une puissante méthode RMN pour l'analyse de mélanges complexes et la détection de traces*. Spectra analyse, vol. 33, pages 26–32, 2004.
- [Goto 2003] J. Goto and M. Nariyasu. *Biomedical and biological mass spectrometry*. Analytical Sciences, vol. 19, no. 1, pages 3–14, 2003.
- [Gray 1978] J.-J. Gray. *Johann Heinrich Lambert, mathematician and scientist, 1728–1777*. Historia mathematica, vol. 5, no. 1, pages 13–41, 1978.
- [Hadamard 1902] J. Hadamard. *Sur les problèmes aux dérivées partielles et leur signification physique*. Bulletin of Princeton University, vol. 13, pages 1–20, 1902.
- [He 2001] Y. He and Y.-C. Sun. *Neural network-based l1-norm optimisation approach for fault diagnosis of nonlinear circuits with tolerance*. IEE Proceedings-Circuits, Devices and Systems, vol. 148, no. 4, pages 223–228, 2001.
- [Heisenberg 1966] W.-K Heisenberg. *The uncertainty principle*. The World of the Atom, Volume 1. Edited with commentaries by Henry A. Boorse and Lloyd Motz, with a foreword by II Rabi. Published by Basic Books, New York, 1966, p. 1094, page 1094, 1966.
- [Hemmersbach 2008] P. Hemmersbach. *History of mass spectrometry at the Olympic Games*. Journal of mass spectrometry, vol. 43, no. 7, pages 839–853, 2008.
- [Hiriart-Urruty 1993] J.-B. Hiriart-Urruty and C. Lemaréchal. Convex analysis and minimization algorithms ii : Advanced theory and bundle methods, volume 306 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, New York, 1993.
- [Hoch 2014] J.-C Hoch, M.-W Maciejewski, M. Mobli, A.-D Schuyler and A.-S Stern. *Nonuniform sampling and maximum entropy reconstruction in multidimensional NMR*. Accounts of chemical research, vol. 47, no. 2, pages 708–717, 2014.
- [Hu 2005] Q. Hu, R.-J. Noll, H. Li, A. Makarov, M. Hardman and R. Graham Cooks. *The Orbitrap : a new mass spectrometer*. Journal of mass spectrometry, vol. 40, no. 4, pages 430–443, 2005.
- [Johnson 1999] C.-S. Johnson. *Diffusion ordered nuclear magnetic resonance spectroscopy : principles and applications*. Prog. Nucl. Magn. Reson. Spectrosc., vol. 34, no. 3, pages 203–256, 1999.
- [Karas 1988] M. Karas and F. Hillenkamp. *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons*. Analytical chemistry, vol. 60, no. 20, pages 2299–2301, 1988.
- [Kazimierczuk 2012] K. Kazimierczuk and V.-Y. Orekhov. *A comparison of convex and non-convex compressed sensing applied to multidimensional NMR*. J. Magn. Reson., vol. 223, pages 1–10, 2012.
- [Kazimierczuk 2015] K. Kazimierczuk and V. Orekhov. *Non-uniform sampling : post-Fourier era of NMR data collection and processing*. Magnetic Resonance in Chemistry, vol. 53, no. 11, pages 921–926, 2015.

- [Korobeynikov 2009] A. Korobeynikov. *Computation-and space-efficient implementation of SSA*. arXiv preprint arXiv :0911.4498, 2009.
- [Kuethe 1999] D.-O. Kuethe, A. Caprihan, I.-J. Lowe, D.-P. Madio and H.-M. Gach. *Transforming NMR data despite missing points*. Journal of Magnetic Resonance, vol. 139, no. 1, pages 18–25, 1999.
- [Landau 1970] L.-D. Landau and E.-M. Lifchitz. *Physique théorique. Tome II, Théorie des champs*. 1970.
- [Le Gall 2016] Y. Le Gall, S.-E. Dosso, F.-X. Socheleau and J. Bonnel. *Bayesian source localization with uncertain Green's function in an uncertain shallow water ocean*. The Journal of the Acoustical Society of America, vol. 139, no. 3, pages 993–1004, 2016.
- [Lewis 1996] A.-S. Lewis. *Derivatives of Spectral Functions*. Mathematics of Operations Research, vol. 21, no. 3, pages 576–588, 1996.
- [Li 2018] F. Li, Q. Su, Z. Zhou, X. Liao, J. Zou, B. Yuan and W. Sun. *Anaerobic biodegradation of 8 : 2 fluorotelomer alcohol in anaerobic activated sludge : Metabolic products and pathways*. Chemosphere, vol. 200, pages 124–132, 2018.
- [Marchione 2009] A.-A. Marchione and R.-C. Buck. *Complete multinuclear magnetic resonance characterization of a set of polyfluorinated acids and alcohols*. Magnetic Resonance in Chemistry, vol. 47, no. 2, pages 194–198, 2009.
- [Martinssona 2006] P.-G. Martinssona, V. Rokhlinb and M. Tygertb. *A randomized algorithm for the approximation of matrices*. Rapport technique, Technical Report 1361, Department of Computer Science, Yale University, New Haven, CT, 2006.
- [Mary 2007] D. Mary and O.-J. Michel. *Quelques Applications du Compressed Sensing en Astronomie*. In 21 Colloque GRETSI, Troyes, FRA, 11-14 septembre 2007. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 2007.
- [Mazet 2013] V. Mazet, C. Soussen and E.-H. Djermoune. *Décomposition de spectres en motifs paramétriques par approximation parcimonieuse*. In Proc. GRETSI, pages 1–4, 2013.
- [Mitrinovic 2012] D.-S. Mitrinovic, J. Pecaric and A.-M. Fink. *Inequalities involving functions and their integrals and derivatives*, volume 53. Springer Science & Business Media, 2012.
- [Moreau 1965] J.-J. Moreau. *Proximité et dualité dans un espace hilbertien*. Bulletin de la Société mathématique de France, vol. 93, pages 273–299, 1965.
- [Muller 2004] N. Muller, L. Magaia and B.-M. Herbst. *Singular value decomposition, eigenfaces, and 3D reconstructions*. SIAM review, vol. 46, no. 3, pages 518–545, 2004.
- [Munson 1966] M.-S. Munson and F.-H. Field. *Chemical ionization mass spectrometry. I. General introduction*. Journal of the American Chemical Society, vol. 88, no. 12, pages 2621–2630, 1966.
- [Natarajan 1995] B.-K. Natarajan. *Sparse approximate solutions to linear systems*. SIAM journal on computing, vol. 24, no. 2, pages 227–234, 1995.
- [Nicoleau 2000] F. Nicoleau. *An inverse scattering problem with the Aharonov–Bohm effect*. Journal of Mathematical Physics, vol. 41, no. 8, pages 5223–5237, 2000.
- [Nityananda 1982] R. Nityananda and R. Narayan. *Maximum entropy image reconstruction—a practical non-information-theoretic approach*. Journal of Astrophysics and Astronomy, vol. 3, no. 4, pages 419–450, 1982.

- [Papoulis 1973] A. Papoulis. *A new method in image restoration*. JSTAC, Paper VI-3, 1973.
- [Paudel 2015] A. Paudel, V. Kumar and S. Singh. *Mass spectrometry in pharmaceutical analysis*. Advanced drug delivery reviews, vol. 89, pages 3–20, 2015.
- [Pesquet 2012] J.-C. Pesquet and N. Pustelnik. *A parallel inertial proximal optimization method*. Pacific Journal of Optimization, vol. 8, no. 2, pages 273–305, 2012.
- [Peypouquet 2015] J. Peypouquet. *Convex optimization in normed spaces : theory, methods and examples*. Springer, 2015.
- [Pichon 1998] C. Pichon and E. Thiébaud. *Non-parametric reconstruction of distribution functions from observed galactic discs*. Monthly Notices of the Royal Astronomical Society, vol. 301, no. 2, pages 419–434, 1998.
- [Powell 1975] M. James D. Powell. *Convergence properties of a class of minimization algorithms*. In Nonlinear programming 2, pages 1–27. Elsevier, 1975.
- [Provencher 1979] S.-W. Provencher. *Inverse problems in polymer characterization : direct analysis of polydispersity with photon correlation spectroscopy*. Macromolecular Chemistry and Physics, vol. 180, no. 1, pages 201–209, 1979.
- [Provencher 1982a] S.-W. Provencher. *A constrained regularization method for inverting data represented by linear algebraic or integral equations*. Computer Physics Communications, vol. 27, no. 3, pages 213–227, 1982.
- [Provencher 1982b] S.-W. Provencher. *CONTIN : a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations*. Computer Physics Communications, vol. 27, no. 3, pages 229–242, 1982.
- [Pustelnik 2011] N. Pustelnik, C. Chaux and J.-C. Pesquet. *Parallel proximal algorithm for image restoration using hybrid regularization*. IEEE transactions on Image Processing, vol. 20, no. 9, pages 2450–2462, 2011.
- [Qi 2014] Y. Qi and P.-B. O’Connor. *Data processing in Fourier transform ion cyclotron resonance mass spectrometry*. Mass spectrometry reviews, vol. 33, no. 5, pages 333–352, 2014.
- [Reimer 2010] P. Reimer, P.-M. Parizel, J.-F.-M. Meaney and F.-A. Stichnoth. *Clinical mr imaging*. Springer, 2010.
- [Renard 2008] B.-Y. Renard, M. Kirchner, H. Steen, J.-A. Steen and F.-A. Hamprecht. *NITPICK : peak identification for mass spectrometry data*. BMC bioinformatics, vol. 9, no. 1, page 355, 2008.
- [Repetti 2015] A. Repetti, M.-Q. Pham, L. Duval, E. Chouzenoux and J.-C. Pesquet. *Euclid in a Taxicab : Sparse Blind Deconvolution with Smoothed l_1/l_2 Regularization*. IEEE Signal Processing Letters, vol. 22, no. 5, pages 539–543, 2015.
- [Rivet 1997] J.-P. Rivet, C. Barbachoux and F. Debbasch. *Le mouvement brownien : une trajectoire hors du commun*. Le Bulletin de l’ADION, 1997.
- [Rockafellar 1976] R.-T. Rockafellar. *Monotone operators and the proximal point algorithm*. SIAM journal on control and optimization, vol. 14, no. 5, pages 877–898, 1976.
- [Rosman 1999] K.-J.-R. Rosman and P.-D.-P. Taylor. *Table of isotopic masses and natural abundances*. Pure and Applied Chemistry, vol. 71, pages 1593–1607, 1999.
- [Schwartz 2004] S.-A. Schwartz, R.-J. Weil, M.-D. Johnson, S.-A. Toms and R.-M. Caprioli. *Protein profiling in brain tumors using mass spectrometry*. Clinical Cancer Research, vol. 10, no. 3, pages 981–987, 2004.

- [Senko 1994] M.-W. Senko, J.-P. Speir and F.-W. McLafferty. *Collisional activation of large multiply charged ions using Fourier transform mass spectrometry*. Analytical Chemistry, vol. 66, no. 18, pages 2801–2808, 1994.
- [Shannon 1948] C.-E. Shannon. *A mathematical theory of communication*. vol. 27, pages 623–656, 1948.
- [Shannon 1949] C.-E. Shannon. *Communication in the presence of noise*. Proceedings of the IRE, vol. 37, no. 1, pages 10–21, 1949.
- [Shell 2008] M. Shell. *The relative entropy is fundamental to multiscale and inverse thermodynamic problems*. The Journal of chemical physics, vol. 129, no. 14, page 144108, 2008.
- [Sinnaeve 2012] D. Sinnaeve. *The Stejskal-Tanner equation generalized for any gradient shape-an overview of most pulse sequences measuring free diffusion*. Concepts in Magnetic Resonance Part A, vol. 40A, no. 2, pages 39–65, 2012.
- [Skilling 1984] J. Skilling and R.-K. Bryan. *Maximum entropy image reconstruction : general algorithm*. Monthly notices of the royal astronomical society, vol. 211, no. 1, pages 111–124, 1984.
- [Slawski 2010] M. Slawski, M. Hein and E. Campus. *Sparse recovery for protein mass spectrometry data*. In Practical Applications of Sparse Modeling., pages 79–98. MIT Press, 2010.
- [Slawski 2012] M. Slawski, R. Hussong, A. Tholey, T. Jakoby, B. Gregorius, A. Hildebrandt and M. Hein. *Isotope pattern deconvolution for peptide mass spectrometry by non-negative least squares/least absolute deviation template matching*. BMC bioinformatics, vol. 13, no. 1, page 291, 2012.
- [Soussen 2015] C. Soussen, J. Idier, J. Duan and D. Brie. *Homotopy based algorithms for l_0 -regularized least-squares*. hand, vol. 2, 2015.
- [Stejskal 1965] E.-O. Stejskal and J.-E. Tanner. *Spin diffusion measurements : spin echoes in the presence of a time-dependent field gradient*. J. Chem. Phys., vol. 42, no. 1, pages 288–292, 1965.
- [Tarantola 2005] A. Tarantola. *Inverse problem theory and methods for model parameter estimation*, volume 89. siam, 2005.
- [Tibshirani 1996] R. Tibshirani. *Regression shrinkage and selection via the lasso*. Journal of the Royal Statistical Society. Series B (Methodological), pages 267–288, 1996.
- [Tikhonov 1976] A.-N. Tikhonov, V.-A. Arsenin and V. Kotliar. *Méthodes de résolution de problèmes mal posés*. 1976.
- [Tikhonov 1986] A.-N. Tikhonov and V.-Y. Arsenin. *Solutions of Ill-Posed Problems (Winston, Washington, DC, 1977)*. Google Scholar, page 100, 1986.
- [Trefethen 1997] L.-N. Trefethen and D. Bau. *Numerical linear algebra*, volume 50. Siam, 1997.
- [Trussell 1983] H.-J. Trussell. *Convergence Criteria for Iterative Restoration Methods*. IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 31, no. 1, pages 129–136, 1983.
- [Urbańczyk 2013] M. Urbańczyk, D. Bernin, W. Kozminski and K. Kazimierczuk. *Iterative thresholding algorithm for multiexponential decay applied to PGSE NMR data*. Anal. Chem., vol. 85, no. 3, pages 1828–1833, 2013.
- [Van der Pas 1971] P.-W. Van der Pas. *The discovery of the Brownian motion*. Scientiarum Historia : Tijdschrift voor de Geschiedenis van de Wetenschappen en de Geneeskunde, vol. 13, no. 1, pages 27–35, 1971.

- [Viéville 2011] J. Viéville, M. Tanty and M.-A. Delsuc. *Polydispersity index of polymers revealed by DOSY NMR*. Journal of magnetic resonance, vol. 212, no. 1, pages 169–173, 2011.
- [Wall 2003] M.-E. Wall, A. Rechtsteiner and L.-M. Rocha. *Singular value decomposition and principal component analysis*. In A practical approach to microarray data analysis, pages 91–109. Springer, 2003.
- [Wang 2015] Z. Wang, D. Liu, J. Yang, W. Han and T. Huang. *Deep networks for image super-resolution with sparse prior*. In Proceedings of the IEEE International Conference on Computer Vision, pages 370–378, 2015.
- [Whitehouse 1989] C.-M. Whitehouse, R.-N. Dreyer, M. Yamashita and J.-B. Fenn. *Electrospray ionization for mass-spectrometry of large biomolecules*. Science, vol. 246, no. 4926, pages 64–71, 1989.
- [Wilkins 1999] D.-K. Wilkins, S.-B. Grimshaw, V. Receveur, Christopher M. Dobson, J.-A. Jones and L.-J. Smith. *Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques*. Biochemistry, vol. 38, no. 50, pages 16424–16431, 1999.
- [Williamson 2016] N.-H Williamson, M. Nydén and M. Röding. *The lognormal and gamma distribution models for estimating molecular weight distributions of polymers using PGSE NMR*. J Magn Reson, vol. 267, pages 54–62, 2016.
- [Wipf 2010] D. Wipf and S. Nagarajan. *Iterative Reweighted Methods for Finding Sparse Solutions*. IEEE Journal of Selected Topics in Signal Processing, vol. 4, no. 2, pages 317–329, 2010.
- [Xu 2013] K. Xu and S. Zhang. *Trust-Region Algorithm for the Inversion of Molecular Diffusion NMR Data*. Anal. Chem., vol. 86, no. 1, pages 592–599, 2013.
- [Yarlagadda 1985] R. Yarlagadda, J. Bednar and T. Watt. *Fast algorithms for lp deconvolution*. IEEE transactions on acoustics, speech, and signal processing, vol. 33, no. 1, pages 174–182, 1985.
- [Yu 2017] Y. Yu, X. Peng J. and Han and A. Cui. *A Primal Douglas-Rachford Splitting Method for the Constrained Minimization Problem in Compressive Sensing*. Circuits, Systems, and Signal Processing, vol. 36, no. 10, pages 4022–4049, 2017.

Méthodes modernes d'analyse de données en biophysique analytique

Résolution des problèmes inverses en RMN DOSY et SM

Résumé

Cette thèse s'intéresse à la création de nouvelles approches algorithmiques pour la résolution du problème inverse en biophysiques. Dans un premier temps, on vise l'application RMN de type DOSY: une nouvelle approche de régularisation hybride a été proposée avec un nouvel algorithme PALMA (<http://palma.labo.igbmc.fr/>). Cet algorithme permet d'analyser des données réelles DOSY avec une précision importante quelque soit leur type. Dans un deuxième temps, notre intérêt s'est tourné vers l'application de spectrométrie de masse. Nous avons proposé une nouvelle approche par dictionnaire dédiée à l'analyse protéomique en utilisant le modèle averagine et une stratégie de minimisation sous contraintes d'une pénalité de parcimonie. Afin d'améliorer la précision de l'information obtenue, nous avons proposé une nouvelle méthode SPOQ, basée sur une nouvelle fonction de pénalisation, résolue par un nouvel algorithme Forward-Backward à métrique variable localement ajustée. Tous nos algorithmes bénéficient de garanties théoriques de convergence, et ont été validés expérimentalement sur des spectres synthétisés et des données réelles.

Mots clés : Problème inverse, DOSY, RMN, Régularisation hybride, Entropie de Shannon, Parcimonie, Approche proximale, PALMA, SM, Déconvolution, Analyse protéomique, Massifs Isotopiques, Approche par dictionnaire, SPOQ.

Abstract

This thesis aims at proposing new approaches to solve the inverse problem in biophysics. Firstly, we study the DOSY NMR experiment: a new hybrid regularization approach has been proposed with a novel PALMA algorithm (<http://palma.labo.igbmc.fr/>). This algorithm ensures the efficient analysis of real DOSY data with a high precision for all different type. In a second time, we study the mass spectrometry application. We have proposed a new dictionary based approach dedicated to proteomic analysis using the averagine model and the constrained minimization approach associated with a sparsity inducing penalty. In order to improve the accuracy of the information, we proposed a new SPOQ method based on a new penalization, solved with a new Forward-Backward algorithm with a variable metric locally adjusted. All our algorithms benefit from sounded convergence guarantees, and have been validated experimentally on synthetics and real data.

Key words: Inverse problem, DOSY, NMR, Hybrid regularization, Shannon Entropy, Sparsity, Proximal approach, PALMA, MS, Deconvolution, Proteomic analysis, Isotopic distribution, Dictionary-based approach, SPOQ.