UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Contributions à l'estimation fréquentielle multidimensionnelle et à la sélection de variables en spectroscopie

Mémoire de recherche

présenté publiquement le 1er décembre 2017 pour l'obtention d'une

Habilitation à Diriger des Recherches de l'Université de Lorraine

(mention automatique, traitement du signal et génie informatique)

par

El-Hadi DJERMOUNE

Composition du jury

Présidente :	Laure BLANC-FERAUD	Directrice de Recherche CNRS à Sophia-Antipolis
Rapporteurs :	Pascal LARZABAL	Professeur à l'Université Paris-Saclay
	Corinne MAILHES	Professeur à l'INP Toulouse
	Hervé CARFANTAN	MdC HDR à l'Université Paul Sabatier
Examinateurs :	André FERRARI	Professeur à l'Université de Sophia-Antipolis
	Cédric CARTERET	Professeur à l'Université de Lorraine
	David BRIE	Professeur à l'Université de Lorraine

Centre de Recherche en Automatique de Nancy
CRAN UMR 7039, Université de Lorraine, CNRS
Faculté des Sciences et Technologies, BP 70239, 54506 Vandœuvre Cedex
Tél. : 03 72 74 53 39 • Courriel : el-hadi.djermoune@univ-lorraine.fr

À Magalie, Célian, Chloé et Clément.

Table des matières

Table des figures

Liste des tableaux

Organisation du manuscrit

Ce document présente une synthèse de mes activités de recherche et d'enseignement réalisées depuis ma nomination en qualité de maître de conférences en septembre 2004. Mes activités d'enseignement ont été effectuées à la Faculté des Sciences et Technologies de l'Université de Lorraine (Université Henri Poincaré avant 2013). Mon travail de recherche a été réalisé au Centre de Recherche en Automatique de Nancy (CRAN, Université de Lorraine, CNRS) au sein du groupe thématique IRIS (*Identification, Restauration, Images et Signaux*) jusqu'en 2012, puis dans le département SBS (*Santé, Biologie, Signal*) qui regroupe des chercheurs en traitement du signal et de l'image, en biologie et en santé depuis le 1er janvier 2013. Les travaux de recherche que j'ai menés s'inscrivent dans les domaines des problèmes inverses en traitement du signal et des images (estimation, déconvolution), de l'approximation parcimonieuse (développement d'algorithmes, applications à l'analyse spectrale et à la sélection de variables en spectroscopie), et de l'analyse d'images de réseaux vasculaires tumoraux.

Ce manuscrit est organisé en trois parties.

La première partie présente mon *curriculum vitae* ainsi qu'une synthèse générale de mes activités d'enseignement, de recherche, d'encadrement et de gestion de contrats. Elle se termine par un bilan global de ma production scientifique.

La deuxième partie est dédiée à la présentation de mes principales contributions méthodologiques et appliquées dans les domaines de l'**estimation fréquentielle multidimensionnelle**, de l'**approximation parcimonieuse** et du **traitement de signaux/images spectroscopiques**. Cette partie est composée de trois chapitres. Chaque chapitre constitue une revue synthétique d'une contribution. Les détails concernant certains éléments techniques et les méthodes développées sont fournis dans une sélection d'articles annexée à ce document.

La troisième partie dresse quelques perspectives scientifiques à court et à moyen terme. J'esquisserai également les grandes lignes de mon projet de recherche à plus long terme dans des applications liées à la biologie et à la caractérisation de la progression tumorale en particulier.

Première partie

Présentation générale

Chapitre 1

Curriculum vitae détaillé

1.1 Identification

Nom :	DJERMOUNE	**Prénom :**	El-Hadi
Naissance :	12/01/1974	**Lieu :**	Amizour (Algérie)
Statut familial :	pacsé, 3 enfants		
Affectation :	Université de Lorraine (UL)	**Date :**	1er septembre 2004
Section CNU :	61	**Corps :**	Maître de conférences
Grade :	6ème échelon depuis le 15 juillet 2016		

1.2 Fonctions principales

2004-17 Maître de conférences à la Faculté des Sciences et Technologies de l'UL.
Recherche au département SBS du CRAN UMR 7039 CNRS-UL.
Enseignement au Département d'Automatique.

2003-04 Attaché Temporaire d'Enseignement et de Recherche (ATER) à l'IUT R&T Nancy-Brabois. Recherche dans le groupe IRIS du CRAN.

2002-03 ATER à la Faculté des Sciences de l'Université Henri Poincaré. Recherche au CRAN.

1998-02 Étudiant de 3e cycle titulaire d'une bourse Franco-Algérienne.

1.3 Titres universitaires

2003 Doctorat en Automatique, Traitement du signal et Génie informatique
Thèse soutenue en juillet 2003. Titre : *Estimation des paramètres de sinusoïdes amorties en sous-bandes adaptative*. Encadrants : A. Richard, M. Tomczak.
Université Henri Poincaré, Centre de Recherche en Automatique de Nancy

1999 DEA Automatique et Traitement Numérique du Signal, Major d'option « Diagnostic »
Université Henri Poincaré, Centre de Recherche en Automatique de Nancy

1998 Ingénieur d'État en Électronique, mention Très Bien, Major de promotion
Université Abderrahmane Mira de Béjaïa, Algérie

1992 Baccalauréat Mathématiques, mention Bien
Lycée d'Amizour, Algérie

1.4 Activités d'enseignement

1.4.1 Enseignements dispensés

J'ai intégré le Département d'Automatique de la Faculté des Sciences et Technologies en septembre 2004. J'enseigne principalement en Licence Sciences pour l'Ingénieur (SPI) et en Master Ingénierie de Systèmes Complexe (ISC) dans des matières liées au traitement du signal, aux communications numériques et aux réseaux de communication. Depuis septembre 2011, j'assure également des enseignement en traitement du signal à TelecomNancy. Le tableau 1.1 donne un aperçu général de ma charge annuelle d'enseignement.

Niveau	Enseignement	Heures		
		CM	TD	TP
L3 SPI	Communication numérique	12	14	
L3 SPI	Réseaux industriels	6	4	9
LP TTAM	Communication industrielle	6	6	8
1re année TelecomNancy	Signal, information, communication	10	18	
Master 1 ISC	Méthodes et algorithmes en traitement du signal	12	12	16
Master 1 ISC	Informatique temps-réel	6	4	
Master 2 ISC	Traitement d'images biologiques multi-échelle	6		

TABLE 1.1 : Résumé de ma charge annuelle d'enseignements à la Faculté des Sciences et à TelecomNancy

Concernant les contenus, mes interventions portent sur les notions suivantes :

- *Communication numérique* (L3) : systèmes de communication, introduction au traitement du signal (transformées de Fourier, échantillonnage, processus aléatoires, densité spectrale de puissance), étude des modulations numériques en bande de base et sur fréquence porteuse, études des performances en présence de bruit (canal de Nyquist).
- *Réseaux industriels* (L3) : description des couches physique et liaison de données dans les réseaux locaux industriels (RLI), études de quelques réseaux (AS-i, Profibus et Ethernet industriel).
- *Communication industrielle* (Licence professionnelle) : modèle OSI, description des couches fonctionnelles, architecture des réseaux, détection d'erreur (parité, Checksum, CRC), étude de cas.
- *Signal, information, communication* (1re année TelecomNancy) : classification des signaux, énergie, puissance, produit de convolution, série et transformées de Fourier, auto et intercorrélation, mise en œuvre sous Matlab.
- *Méthodes et algorithme en traitement du signal* (M1) : Processus aléatoires, stationnarité et ergodicité, densité spectrale de puissance (DSP), estimateurs non paramétriques et paramétriques de la DSP, estimateurs fréquentiels MUSIC et ESPRIT, mise en œuvre sous Matlab.
- *Informatique temps-réel* (M1) : Ordonnancement de tâches périodiques et apériodiques, dépendantes ou indépendantes, système RTAI.

- *Traitement d'images biologiques multi-échelle* (M2) : filtrage d'images, détection de contours, restauration, morphologie mathématique, décomposition en ondelettes.

1.4.2 Responsabilités pédagogiques

- Responsable de la licence CASI de l'IUP GEII en 2004–2005.
- Responsable d'année (M1) du Master ISC depuis septembre 2010 : entretiens pédagogiques et animation des commissions M1.
- Évaluation de dossiers d'entrée en M1 et M2 ISC.
- Contribution à l'élaboration de la maquette d'accréditation du Master ISC 2018-2022 (co-responsable avec J.-P. Georges du parcours *Réseaux, Signaux, éco-TIC*).
- Responsable des unités d'enseignement (UE) *Communication numérique* et *Réseaux industriels* (L3), *Communication industrielle* (LP), *Méthodes et algorithme en traitement du signal* (M1 ISC) et *Traitement d'images biologiques multi-échelle* (M2 ISC).
- Mise en place des CM et TD de l'UE *Communication numérique*.
- Mise en place des CM, TD et TP des UE *Réseaux industriels* et *Communication industrielle*.
- Mise en place des CM, TD et TP de l'UE *Méthodes et algorithmes en traitement du signal* avec C. Soussen (qui est responsable de la partie *analyse en composantes principale* de l'UE).
- Mise en place de l'UE *Traitement d'images biologiques multi-échelle* avec C. Soussen et D. Brie.

1.5 Activités de recherche

Cette partie présente mes travaux de recherche effectués :

- entre 1999 et 2003 en thèse au CRAN sous la direction de Alain Richard ;
- depuis 2004 en tant que Maître de Conférences au sein du département SBS du CRAN.

1.5.1 Travaux de thèse

J'ai soutenu ma thèse de doctorat intitulée : *Estimation des paramètres de sinusoïdes amorties par décomposition en sous-bandes adaptative. Application à la spectroscopie RMN*, le 09 juillet 2003 devant le jury suivant :

Président :	Régis Lengellé, Professeur à l'Université de Technologie de Troyes
Rapporteurs :	Pascal Larzabal, Professeur à l'Université Paris-Sud
	Eric Moreau, Professeur à l'Université de Toulon et du Var
Examinateurs :	Pierre Mutzenhardt, Professeur à l'Université Henri Poincaré, Nancy 1
	Alain Richard, Professeur à l'Université Henri Poincaré, Nancy 1
	Marc Tomczak, Maître de conférences à l'Univ. Henri Poincaré, Nancy 1

Les recherches que j'ai menées portent sur l'analyse spectrale et l'estimation fréquentielle. Je me suis plus particulièrement intéressé aux signaux de spectroscopie de résonance magnétique nucléaire (RMN) qui présentent plusieurs difficultés en traitement du signal notamment à cause de leur complexité (nombre élevé de composantes, problèmes de résolution dynamique et fréquentielle) et la présence de facteurs de relaxation exponentiels. Mes principales contributions sont décrites ci-dessous.

Estimation paramétrique en sous-bandes Pour réduire la complexité des signaux, nous avons proposé une méthode basée sur la décomposition en sous-bandes avant l'estimation. Nous avons proposé un banc de filtres ayant des caractéristiques intéressantes pour l'estimation fréquentielle. Il permet d'éviter l'atténuation de modes et le repliement spectral. L'application de cette technique à des signaux de spectroscopie RMN a permis de mettre en évidence plusieurs avantages de l'estimation en sous-bandes. Parmi ceux-ci, on peut citer l'amélioration du taux de détection et la réduction du taux de fausses alarmes. En terme de complexité numérique, nous avons montré que l'estimation en sous-bandes réduit sensiblement le temps de calcul grâce à la réduction de l'ordre du modèle et du nombre d'échantillons du signal [A1].

Décomposition en sous-bandes adaptative Lorsque l'on veut réaliser une décomposition en sous-bandes, un des problèmes qui se pose est celui du choix du facteur de décimation car celui-ci conditionne les résultats de l'estimation. Nous avons proposé une procédure qui utilise une règle d'arrêt basée sur une mesure de platitude spectrale des résidus d'estimation. Nous avons montré au travers de simulations numériques que cette méthode permet d'atteindre de meilleures performances, plus particulièrement sur le plan de la détection et du nombre de bandes finales, comparativement à la décomposition uniforme et aux techniques adaptatives fondées sur d'autres règles d'arrêt. De plus, cette technique assure parallèlement une variance d'estimation raisonnable, induite par la maximisation du taux de détection. Ces résultats sont confirmés par une application à des signaux de spectroscopie RMN de complexité très élevée (constitués d'une centaine de composantes) [A2, O1].

1.5.2 Travaux de recherche après la thèse

1.5.2.1 Estimation fréquentielle 1-D et 2-D

Ces travaux s'inscrivent dans la continuité de mon travail de thèse. Mes contributions portent sur deux volets :

- l'estimation fréquentielle par décomposition en sous-bandes 2-D ;
- l'analyse théorique des performances de méthodes d'estimation en sous-espace.

Concernant la décomposition en sous-bandes 2-D, l'approche que j'ai proposée est une extension de la méthode développée pendant ma thèse. Je me suis focalisé sur les approches de décomposition adaptatives. Une méthode de décomposition/estimation, utilisant un critère d'arrêt basé sur la platitude spectrale, a été proposée [Ci8, Ci7] et validée par des applications aux signaux spectroscopie RMN 13C [A6].

Les méthodes d'estimation qui se basent sur la séparation des sous-espaces signal et bruit sont souvent utilisées pour estimer les paramètres d'un modèle sinusoïdal. Dans le cas de sinusoïdes pures,

les performances de ces méthodes sont connues et largement étudiées dans la littérature. Mon objectif était d'analyser les performances de ces méthodes en présence de *signaux amortis*. Les méthodes étudiées sont : Kumaresan-Tufts (KT), *matrix pencil* (MP) et *direct data approximation* (DDA), en m'appuyant sur la théorie des perturbations. Cette étude a abouti aux expressions analytiques de la variance des estimées [A3, Ci9]. Ce résultat nous a permis de démontrer l'équivalence des méthodes MP et DDA et leur supériorité comparativement à KT. Par la suite, nous avons démontré la convergence linéaire de la variance de toutes ces méthodes vers la borne de Cramér-Rao.

1.5.2.2 Modélisation du soudage MIG/MAG et segmentation de vidéos par contour actif

Ce travail a été réalisé dans le cadre de la thèse CIFRE de Jean-Pierre Planckaert[1]. Le problème était de proposer un modèle physique du procédé de soudage MIG/MAG[2] et le valider par comparaison à des données expérimentales. À cause de la complexité des phénomènes physiques sous-jacents (plasma), des simplifications ont été effectuées pour obtenir un modèle représentant les tendances majeures et suffisamment simple pour construire un simulateur numériquement stable. Le modèle proposé est un modèle hybride comportant deux états continus : un état d'arc pendant lequel une goutte se forme et un état de court-circuit correspondant au transfert de métal dans le bain. Afin de valider le simulateur basé sur ce modèle, des séquences vidéo haute vitesse (10000 images/s) d'opérations de soudage ont été réalisées. Un algorithme de contour actif a été développé pour extraire des informations sur l'évolution de la géométrie et la dynamique des gouttes de métal fondu [A4].

1.5.2.3 Estimation modale multidimensionnelle

À partir de l'année 2009, je me suis intéressé au problème de l'estimation modale multidimensionnelle dans le cadre de la thèse de Souleymen Sahnoun[3] (R-D). L'objectif était de développer des algorithmes rapides pour l'estimation des paramètres de ces signaux en présence de problèmes de résolution et de complexité numérique. Nos contributions portent sur les trois points décrits ci-dessous.

Algorithme mutirésolution exploitant la parcimonie Les méthodes d'approximation parcimonieuses sont des techniques en plein essor dans la communauté de traitement du signal. Ces méthodes peuvent être utilisées pour décomposer un signal en termes d'éléments d'un dictionnaire redondant (signaux élémentaires, de forme connue, appelés *atomes*). L'idée est de trouver la meilleure combinaison linéaire de ces atomes permettant de minimiser l'erreur de reconstruction en impliquant le moins d'atomes possible. Dans le cas de l'estimation modale, le dictionnaire est obtenu par la discrétisation de fonctions exponentielles complexes amorties et chaque point de la grille de discrétisation permet d'obtenir un atome. Afin d'atteindre une bonne résolution spectrale, il est nécessaire de choisir une grille très fine, ce qui conduit à la manipulation d'un dictionnaire de grande taille avec tous les problèmes de calcul sous-jacents. La méthode que nous avons proposée consiste à combiner une approximation parcimonieuse et une approche multigrille. Il s'agit de réaliser l'approximation

[1] Doctorant (2004-2007).
[2] *Metal Inert Gas/Metal Active Gas.*
[3] Doctorant (2009-2012).

sur plusieurs niveaux de résolution. Dans le premier niveau, le dictionnaire correspond à une grille grossière ; cette dernière est ensuite affinée de façon adaptative en fonction des atomes activés par l'approximation parcimonieuse, en insérant des atomes au voisinage de ceux activés. Cette procédure est réitérée jusqu'à ce qu'un critère d'arrêt soit satisfait : niveau de résolution atteint ou erreur de reconstruction atteinte. Cette contribution a fait l'objet de communications en congrès nationaux et internationaux (par exemple [Ci11, Ci10, Cn1]) et de deux papiers de revues internationales [A11, A5].

Estimation tensorielle en sous-espace L'approche multirésolution précédente est performante numériquement mais n'offre pas de garanties de convergence à cause de la méthode d'approximation gloutonne (S-OMP) utilisée à chaque niveau[4]. En revanche, si les modes R-D sont spatialement bien séparés, la convergence peut être facilement établie. Afin de garantir la convergence dans tous les cas, et en particulier lorsque le nombre de modes $F > 1$, nous avons exploité une représentation alternative du modèle qui permet de décomposer le tenseur de données en F tenseurs chacun contenant *un seul mode*.

Bornes de Cramér-Rao des paramètres du modèle exponentiel complexe R-D Une des approches classiques permettant d'évaluer les performances d'un estimateur est de comparer la variance des estimées à la borne de Cramér-Rao. Dans le cas du modèle exponentiel complexe, ces bornes sont déjà établies pour les signaux 1-D et 2-D. En revanche, pour les signaux R-D, seul le modèle exponentiel pur (non amorti) a été considéré dans la littérature. Nous avons pu établir les bornes de Cramér-Rao dans le cas général d'un modèle exponentiel complexe R-D dans lequel les amortissements ne sont pas nécessairement nuls [A11].

1.5.2.4 Approches parcimonieuses pour la sélection de variables et la classification

Ce problème est issu du projet FUI TRISPIRABOIS qui sera décrit dans le paragraphe 1.7. L'objectif est de détecter la présence de polluants dans les déchets de bois par spectroscopie proche infrarouge. Ce problème est formulé comme un problème de classification binaire dans lequel on cherche à minimiser le volume de bois pollué qui entre dans la fabrication de nouveaux panneaux. Comme les spectres sont composés de plus de 1600 bandes, une étape de sélection de variables est nécessaire pour, d'une part, réduire les temps de calcul en sélectionnant seulement les bandes susceptibles de révéler la présence de polluants et, d'autre part, d'éviter le phénomène de surapprentissage qui nuit à la généralisation du modèle. Ce travail de recherche a été réalisé dans le cadre de la thèse de Leïla Belmerhnia[5]. Les approches proposées pour la sélection de variables sont basée sur des représentations parcimonieuses simultanées impliquant tous les spectres d'une base d'apprentissage [Ci14, Ci17, Cn8, Ci20]. Les méthodes proposées sont détaillées dans le chapitre 4.

1.5.2.5 Restauration d'images hyperspectrales

Une autre facette du projet TRISPIRABOIS concerne la restauration d'images hyperspectrales acquises par un spectro-imageur rapide (par exemple, un imageur *pushbroom*). Le problème concerne

[4] La convergence peut bien sûr être établie dans le cas où chaque dimension du tenseur de données contient un mode unique.

[5] Doctorante (2013-2016).

la déconvolution et le débruitage du cube de données résultant. Plusieurs méthodes de déconvolution d'images hyperspectrales sont proposées dans la litérature, mais leur mise en œuvre nécessite le réglage des paramètres de régularisation qui conditionnent directement leurs performances. Le premier objectif du travail de thèse de Yingying Song[6] que je co-encadre avec David Brie a été de proposer des critères permettant de sélectionner automatiquement ces paramètres. Le problème est d'abord formulé comme un problème d'optimisation multi-objectifs impliquant des termes (quadratiques) de régularisation spectrale et spatiale ainsi qu'une contrainte de positivité. L'étude des propriétés de la surface de réponse, qui est convexe dans les cas contraint et non-contraint, nous a permis de proposer deux critères pour estimer les paramètres de régularisation : (1) le *minimum distance criterion* (MDC) ; (2) le *maximum curvature criterion* (MCC). La supériorité de ces critères par rapport à la courbe (ou l'hypersurface) en L a été démontrée au moyen de simulation de résultats expérimentaux [A8, Ci21].

1.5.2.6 Identification de modèles dynamiques linéaires à effets mixtes en biologie

L'identification de systèmes apparaît de plus en plus dans la modélisation de systèmes biologiques. Dans ce contexte, la variabilité inter-essais est importante et il est souvent nécessaire de répéter plusieurs fois une même expérience afin de quantifier cette variabilité. L'objectif est alors de prendre en compte l'ensemble des répétitions dans l'estimation des paramètres pour assurer qu'ils soient groupés dans l'espace des solutions, permettant ainsi une meilleure prise en compte de l'information à travers les répétitions. Les approches par population sont connues en statistique sous le nom de modèles à effets mixtes. L'objectif de la thèse de Levy Batista[7], co-encadrée avec Thierry Bastogne, est d'intégrer les modèles à effets mixtes dans les structures de modèles dit « boîtes-noires » en identification de systèmes et d'adapter les méthodes d'estimation associées. Dans cette thèse, une méthode basée sur une structure ARX[8] a été proposée ; les paramètres sont estimés en utilisant l'algorithme EM (Espérance-Maximisation) [Ci16, Ci18, Cs2, Ci22]. Cette approche est ensuite étendue à l'ensemble des modèles boîtes-noires en utilisant la version stochastique de l'algorithme EM, l'algorithme SAEM.

1.6 Activités d'encadrement

1.6.1 Thèses de doctorat

[D.7] **Pauline Guyot** (2016–2019) : *Modélisation et simulation de l'électrocardiogramme d'un patient numérique.*

- Financement Hôpital Virtuel de Lorraine (HVL) et start-up CYBERnano. Thèse débutée en octobre 2016.
- Directeur : T. Bastogne (40%), co-directeurs : **E.-H. Djermoune** (30%) et B. Chenuel (CHU Nancy).
- Publication : [Ci23].

[6] Doctorante (2015-2018).
[7] Doctorant (2014-2017).
[8] *Auto-Regressive model with eXternal inputs.*

[D.6] **Thi-Thanh Nguyen** (2016–2019) : *Algorithmes d'inversion haute résolution régularisée par la parcimonie.*

- Financement ANR BECOSE. Thèse débutée en octobre 2016.
- Directeur : C. Soussen (50%), co-directeur : J. Idier (IRCCyN, 30%).
- Co-encadrant : **E.-H. Djermoune** (20%).
- Publication : [Cn9].

[D.5] **Yingying Song** (2015–2018) : *Amélioration de la résolution d'un spectro-imageur IR industriel par déconvolution/séparation conjointe. Application à la caractérisation de déchets bois.*

- Financement FUI (projet Trispirabois) et Région Lorraine. Thèse débutée en octobre 2015.
- Directeur : D. Brie (50%), co-directeur : **E.-H. Djermoune** (50%).
- Publications : [A8, Ci24, Ci21, Cn10].

[D.4] **Levy Batista** (2014–2017) : *Identification de populations de systèmes dynamiques représentés par des modèles à effets mixtes.*

- Contrat Cifre (start-up CYBERnano). Thèse débutée en octobre 2014 et soutenue le 6 décembre 2017.
- Directeur : T. Bastogne (60%), co-directeur : **E.-H. Djermoune** (40%).
- Publications : [Ci22, Ci18, Cs2, Cn7, Cn5].

[D.3] **Leila Belmerhnia** (2013–2016) : *Approches parcimonieuses pour la sélection de variables et la classification. Application à la spectroscopie IR de déchets de bois.*

- Financement FUI (projet Trispirabois). Thèse débutée en octobre 2013 et soutenue le 2 mai 2017.
- Directeur : D. Brie (50%), co-directeur : **E.-H. Djermoune** (50%).
- Situation depuis mars 2017 : ingénieur au laboratoire PROTEE (Université de Toulon).
- Publications : [Ci20, Ci17, Ci14, Cn8].

[D.2] **Souleymen Sahnoun** (2009–2012) : *Développement de méthodes d'estimation modale de signaux multidimensionnels. Application à la spectroscopie RMN multidimensionnelle.*

- Bourse ministérielle (contrat doctoral). Thèse débutée en octobre 2009 et soutenue le 27 novembre 2012.
- Directeur : D. Brie (50%), co-directeur : **E.-H. Djermoune** (50%).
- ATER en 2012-2013 à l'Université de Lorraine.
- Qualifié aux fonctions de Maître de conférences en 2013.
- Post-doctorant au Gipsa-lab (Grenoble) en 2013-2016 sous la direction de P. Comon.
- Situation depuis septembre 2016 : ingénieur R&D en traitement du signal dans la téléphonie mobile (start-up Situ8ed, www.situ8ed.com), Grenoble.

- Publications : [A5, Ci13, Ci11, Ci10, Cn3, Cn1].

[D.1] **Jean-Pierre Planckaert** (2004–2007) : *Modélisation du soudage MIG/MAG en mode short-arc.*

- Convention Cifre Air Liquide. Thèse débutée en novembre 2004 et soutenue le 1er juillet 2008.
- Directeur : D. Brie (50%), co-directeur : **E.-H. Djermoune** (50%).
- Situation depuis novembre 2007 : ingénieur R&D et expert international à Air Liquide à Jouy-en-Josas.
- Publications : [A4, Ci6, Ci5].

1.6.2 Stages de master 2

[S.7] Thomas Aiguier. *Mise en œuvre d'un système de classification de déchets bois par spectroscopie NIR.* Master Ingénierie de Systèmes Complexes, CRAN, Université de Lorraine, 2017.

[S.6] Levy Batista. *Identification et analyse spectrale de réponses d'impédance.* Master Ingénierie de Systèmes Complexes, CRAN, Université de Lorraine, 2014.

[S.5] El Waled Nemane. *Apport de la cohérence spectrale en classification par machines à vecteurs de support.* Master Ingénierie Mathématique et Outils Informatiques, CRAN, Université de Lorraine, 2014.

[S.4] Leila Belmerhnia. *Approches parcimonieuses pour la séparation d'images hyperspectrales.* Master Ingénierie de Systèmes Complexes, CRAN, Université de Lorraine, 2013.

[S.3] Oualid Challaf. *Filtrage robuste d'images de spectroscopie d'émission optique couplée à l'ablation laser.* Master Ingénierie de Systèmes Complexes, CRAN, Université de Lorraine, 2010. Publication : [Ci12].

[S.2] Gordana Kasalica. *Méthodes à haute résolution pour l'analyse des signaux de spectroscopie RMN multidimensionnelle.* Master Ingénierie de Systèmes Complexes, CRAN, Université de Lorraine, 2007. Publications : [Ci7, Cs1].

[S.1] Jean-Marc Schram. *Détection et classification de défauts de soudage MIG.* Stage d'ingénieur CNAM, CRAN, Université de Lorraine, 2006.

1.6.3 Projet de master 1

[P.3] Jérémy Hraman. *Détection de défauts de soudage en fabrication additive.* Projet 2A INP Grenoble-PHELMA, CRAN, Université de Lorraine, 2017.

[P.2] Mohammed Elaynousse. *Classification temps-fréquence multivoie pour la surveillance acoustique de générateurs de vapeurs.* Master 1 Ingénierie Mathématique, CRAN, Université de Lorraine, 2014.

[P.1] Jérémy Tan Luong Ann. *Développement d'un logiciel interactif pour l'analyse spectrale de signaux RMN.* Projet 2A TelecomNancy, CRAN, Université de Lorraine, 2006.

1.7 Contrats de recherche

1.7.1 Au delà de l'échantillonnage compressé : Algorithmes d'approximation parcimonieuse pour les problèmes inverses mal conditionnés (BECOSE)

- Financement : ANR
- Laboratoires partenaires : CRAN, IRCCyN, INRIA Centre Rennes, ONERA, IRSTEA
- Durée : 4 ans (2016–2019)
- Porteur : C. Soussen (CRAN)
- Autres chercheurs : T. Nguyen, **E.-H. Djermoune** (CRAN), J. Idier, S. Bourguignon (IRCCyN), C. Herzet, R. Gribonval (INRIA), F. Champagnat, B. Leclaire, C. Illoul (ONERA), D. Heitz (IRSTEA)
- Budget : 480 k€

Résumé : Le projet concerne les algorithmes d'approximation parcimonieuse dédiés aux problèmes inverses mal conditionnés. Si beaucoup d'algorithmes rapides sont adaptés aux problèmes bien conditionnés, les problèmes plus difficiles nécessitent d'utiliser des algorithmes plus complexes basés sur la minimisation d'une fonction coût non convexe. L'analyse théorique de reconstruction exacte est un verrou car les conditions habituelles sont trop pessimistes. Nous proposons d'analyser certains algorithmes efficaces (utilisant la norme ℓ_0) afin de mieux refléter le comportement réel des algorithmes dans le contexte des problèmes mal conditionnés. Les algorithmes proposés et analysés seront appliqués à la PIV[9] tomographique, modalité récente de mesure de vitesse 3D d'écoulements par imagerie de particules. L'enjeu de ces traitements numériques est majeur dans différents secteurs d'application de la mécanique des fluides comme l'industrie automobile, l'aéronautique et l'environnement.

1.7.2 Décomposition spectroscopique en imagerie multispectrale (DSIM)

- Financement : ANR JCJC
- Laboratoires partenaires : ICube, CRAN, CRAL
- Durée : 4 ans (2015–2018)
- Porteur : V. Mazet (ICube)
- Responsables CRAN : C. Soussen, **E.-H. Djermoune**
- Budget : 180 k€

[9] *Particle Image Velocimetry.*

Résumé : La compréhension de l'Univers (et en particulier son histoire et son évolution) demande d'analyser la lumière des galaxies. De nos jours, les télescopes fournissent des images multispectrales qui sont des images 3D dont la troisième dimension correspond à la longueur d'onde. Chaque pixel d'une image multispectrale est un spectre électromagnétique présentant des raies d'émission. La connaissance de leurs paramètres (longueur d'onde, intensité, ...) fournit des informations sur les propriétés physiques des galaxies observées. Le but du projet DSIM est de développer des algorithmes permettant d'estimer les paramètres et le nombre de raies. La décomposition spectroscopique est considérée comme un problème inverse et les raies sont modélisées par une fonction paramétrique dont les paramètres sont inconnus. L'évolution spatiale est modélisée à l'aide de régularisations et d'*a priori* adéquats. Le problème est traité, d'une part, dans un cadre bayésien avec des algorithmes de type RJMCMC et Hamiltonian MCMC, et d'autre part, par des méthodes d'approximation parcimonieuse afin de réduire le coût de calcul.

1.7.3 Tri par différentes spectroscopies, dont l'infra-rouge, des déchets de bois (TRISPIRABOIS)

- Programme : FUI EcoIndustries 2012
- Financement : BPI France, régions Lorraine et PACA, Conseil Général des Vosges
- Consortium : Egger Rambervillers, Critt Bois, CRAN, LCPME, Pellenc ST
- Durée : 4 ans (2013–2016)
- Porteur : Egger Rambervillers
- Responsables CRAN : D. Brie
- Autres chercheurs : **E.-H. Djermoune**, C. Carteret (LCPME)
- Budget CRAN : 240 k€

Résumé : Le projet TRISPIRABOIS a pour objectif général la gestion des déchets bois et la mise en place de filières de revalorisation pertinentes. Il est le fruit d'une collaboration entre plusieurs partenaires et regroupe des acteurs dans les domaines de valorisation de déchets (Egger, Pellenc ST), de promotion du matériau bois (Crittbois), de chercheurs (CRAN, LCPME) et des pouvoirs publics. Dans l'industrie de la collecte et du tri sur plateforme, certaines techniques de tri et de séparation existent pour le bois. La spectroscopie proche infra-rouge (SPIR) est aujourd'hui capable de séparer les bois des non-bois, ainsi que les bois non-traités (bois de classe A), les bois faiblement adjuvantés : mélaminés, agglomérés, peints et vernis, contreplaqués (bois de classe B) et les bois fortement adjuvantés (bois de classe C), avec une efficacité encore perfectible. La spectroscopie est utilisée également dans la reconnaissance des matériaux (polymères entre eux, papier, bois). Avec les techniques actuelles de spectroscopie IR, chaque spectre est composé d'un millier de longueurs d'ondes. Afin de réduire les coûts de calcul (le temps d'acquisition et de traitement doit être de quelques dizaines de micro-secondes par spectre) et répondre à la contrainte « temps-réel » de l'application, il est nécessaire de sélectionner dans l'ensemble des bandes spectrales celles qui sont susceptibles de révéler la présence de polluants. La première partie du travail de recherche a pour

but de sélectionner un sous-ensemble de variables explicatives (longueurs d'onde) partagées par tous les spectres déjà étiquettés. Ce problème de sélection de variables est abordé dans le cadre de l'approximation parcimonieuse simultanée dans laquelle les coefficients sont également contraints d'être constants par morceaux. Cette contrainte permet de prendre en compte le fait que les spectres soient rangés dans une matrice en fonction de leurs classes d'appartenance. La deuxième partie vise à la mise en place d'un classifieur exploitant les variables sélectionnées afin de séparer les bois recyclables des bois indésirables.

1.7.4 Surveillance de la réaction sodium-eau dans les générateurs de vapeur : Détection et caractérisation des signatures acoustiques de fuites faibles (DARSE)

- Financement : NEEDS (Nucléaire, Energie, Environnement, Déchets et Société) CNRS
- Laboratoires : STMR UMR 6279, CRAN, CEA Cadarache
- Durée : 2 ans (2013–2014)
- Porteur : P. Beauseroy (STMR)
- Responsable CRAN : **E.-H. Djermoune**
- Budget CRAN : 18 k€

Résumé : Dans le cadre des études sur des réacteurs nucléaires innovants et de l'amélioration de leur sûreté de fonctionnement, le développement de solutions utilisant le sodium liquide est en cours (prototype ASTRID). Les conditions spécifiques associées à ce caloporteur impliquent de garantir l'opérabilité d'un système de conversion d'énergie utilisant de l'eau (dans lequel se trouve des échangeurs thermiques : les générateurs de vapeur). Ainsi, la détection rapide de petites fuites d'eau dans le sodium constitue une voie d'étude importante qui conduit à réaliser des injections d'eau volontaires dans le sodium et à qualifier les différents systèmes de détection possibles. Parmi ceux-ci, la mesure acoustique constitue un axe privilégié en raison de sa sensibilité et de son faible temps de réponse : le principe est d'identifier une signature acoustique spécifique. Toutefois, le bruit de fond de l'installation industrielle est susceptible de cacher le signal utile et les bruits générés par d'autres sources peuvent conduire à des fausses alarmes. C'est pourquoi, le traitement des signaux acoustiques représente une partie incontournable de ces études. Le système de détection développé par le CRAN est basé sur un classifieur SVM utilisant comme variables une estimation dans la densité spectrale de puissance (spectrogramme) des signaux acoustiques.

1.7.5 Décomposition parcimonieuse de signaux spectroscopiques (SpectroDec)

- Financement : PEPS CNRS
- Laboratoires partenaires : ICube, CRAN, Observatoire Astronomique de Strasbourg, Laboratoire Francis Perrin (Saclay)
- Durée : 2 ans (2012–2013)

- Porteur : V. Mazet (ICube)
- Chercheurs CRAN : **E.-H. Djermoune**, C. Soussen
- Budget : 16 k€

Résumé : Ce projet avait pour but de développer des méthodes originales de décomposition d'un signal spectroscopique (estimation de la position des raies, de leurs largeurs et de leurs amplitudes). Le premier objectif concerne la décomposition conjointe, à l'aide de méthodes statistiques, d'une séquence de spectres dont les raies évoluent et sont en nombre inconnu. Le second objectif concerne la conception de méthodes d'approximation parcimonieuse adaptées à la décomposition spectroscopique dans le but d'accélérer les méthodes actuelles. Le défi est d'utiliser un dictionnaire très grand composé de fonctions gaussiennes fortement corrélées.

1.8 Activités de valorisation et de transfert

1.8.1 Contrats industriels

1.8.1.1 Détection de défauts de soudage à l'arc en fabrication additive

- Contrat de collaboration CRAN–Air Liquide
- Durée : 6 mois (2017)
- Responsable : **E.-H. Djermoune**
- Autre chercheur : D. Brie (CRAN)
- Budget : 25 k€

Résumé : Ce contrat de recherche correspond à une demande exprimée par la société Air Liquide pour la surveillance des soudures réalisées par des robots en fabrication additive. Les systèmes de fabrications additives utilisent de fines coupes horizontales tirées de modèles CAO, de scans 3D ou de scanners médicaux des objets à fabriquer et la pièce réelle est créée en superposant ces couches de matériaux. Une stratégie est d'utiliser les procédés de soudages conventionnels comme le MIG-MAG, le TIG, le plasma et le laser afin de générer les couches successives par la fusion d'un fil. Cette famille de procédés de fabrication additive est appelée WAAM (*Wire Arc Additive Manufacturing*). Une étape d'usinage pour le parachèvement est indispensable pour assurer la précision dimensionnelle de la pièce. Une étape de traitement thermique est souvent appliquée également. Lors de l'opération de soudage, il est classique de mesurer certains paramètres caractéristiques susceptibles de traduire un dysfonctionnement comme, par exemple, l'intensité de courant, la tension, la pression du gaz de protection et la vitesse d'amenée du fil. Le but est de développer une méthode permettant de détecter des défauts de soudage à partir des mesures de courant, de tension, de température et de profilométrie laser. L'idée est de rechercher des attributs dans les signaux mesurés qui vont permettre de diagnostiquer et éventuellement de classifier les défauts détectés en fonction de la cause de leur apparition, afin d'aider l'opérateur à remédier à un possible dysfonctionnement.

1.8.1.2 Détection de phénomènes tuyère dans les hauts fourneaux

- Contrat de collaboration CRAN–Paul Wurth
- Durée : 1 an (mars 2016 – février 2017)
- Responsables : **E.-H. Djermoune**, J.-C. Ponsart
- Autre chercheur : D. Theilliol (CRAN)
- Budget : 25 k€

Résumé : Durant les 3 dernières années, Paul Wurth a développé une nouvelle technologie de surveillance pour les tuyères de hauts fourneaux. Un système intégré utilisant des caméras numériques permet une visualisation continue de toutes les tuyères à partir d'un ordinateur connecté au réseau. Des méthodes de vision industrielle sont appliquées pour la détection des phénomènes de tuyères. Un lien direct vers le système d'automatisation du haut fourneau permet de mettre en œuvre des mesures de sécurité supplémentaires (par exemple, arrêt immédiat de l'injection) permettant l'amélioration de la sécurité du personnel et des équipements. Au cours du fonctionnement normal, la surveillance à distance de l'injection de la tuyère améliore la stabilité du procédé permettant des taux plus élevés d'injections. Au travers des images rafraichies de manière périodique, l'objectif de cette étude consiste à réaliser une identification des caractéristiques d'une tuyère : localisation de la tuyère sur une image (centre et rayon) ainsi que de la lance d'injection de forme rectangulaire (recherche des coordonnées des quatre sommets). Il est important de préciser que la qualité des images peut être altérée pour différentes raisons : encrassement (poussière) de l'objectif de la caméra ; condensation (vapeur d'eau) ; problèmes d'alignement de la tuyère ; taux d'injection élevé conduisant à des difficultés de localisation de la lance.

1.8.1.3 Surveillance des réacteurs SFR : Analyse des données acoustiques des générateurs de vapeur

- Contrat de collaboration CRAN–CEA, dans le cadre du GIS-3SGS
- Durée : 1 an (2009)
- Responsable : **E.-H. Djermoune**
- Autres chercheurs : S. Henrot, D. Brie (CRAN)
- Budget : 15 k€

Résumé : A l'arrêt du réacteur à neutrons rapides britannique *Prototype Fast Reactor* (PFR) en 1994, des essais de fin de vie furent organisés. En particulier, les essais *Leak Detection System* (LDS) concernent les techniques de détection de fuites d'eau dans le sodium des générateurs de vapeur. Le but de l'étude est le développement d'une méthode de détection acoustique sur une série d'enregistrements issus des essais LDS5 qui consistent à injecter de l'argon, de l'hydrogène et de l'eau dans le sodium de l'un des évaporateurs de PFR. L'idée est de fournir une approche complémentaire aux méthodes chimiques déjà existantes. La méthode proposée est basée sur le spectrogramme des signaux acoustiques en combinaison avec un classifieur SVM.

1.8.1.4 Modélisation du soudage MIG/MAG en mode short arc

- Contrat de collaboration CRAN–Air Liquide
- Durée : 3 ans (2004–2007)
- Responsables : D. Brie, **E.-H. Djermoune**
- Autre chercheurs : J.-P. Planckaert (doctorant), F. Briand, F. Richard (Air Liquide)

Résumé : Cette collaboration a été réalisée dans le cadre d'une convention CIFRE. L'objectif est de modéliser le processus de soudage MIG/MAG pour un transfert de type *short-arc*. Il s'agit donc de réaliser la modélisation de l'ensemble du système (fil, gaz, arc, tôle, régulation) en fonction d'un certain nombre de grandeurs physiques parmi lesquelles : la tension de consigne du générateur en mode « arc » et le profil d'évolution de l'intensité du générateur en mode « court-circuit » ; la distance tube contact-pièce ; la vitesse d'avance du fil ; le type de gaz utilisé ; la nuance de l'acier à souder, etc. Le travail effectué au cours de la thèse comporte des contributions qui concernent la modélisation physique du procédé de soudage ainsi que la segmentation de vidéos rapides. L'approche développée pour modéliser le processus de soudage est une approche analytique avec le but de comprendre la physique du procédé. Ce dernier est représenté sous forme de système hybride avec deux états continus (arc et court-circuit) ; le passage de l'un à l'autre étant contrôlé par une variable de commutation. Par ailleurs, une méthode de suivi de contour sur des vidéos rapides en utilisant l'approche de contour actif dynamique a été développée. Cette approche a d'abord été appliquée en mode arc, puis en mode court-circuit en intégrant une détection de la surface du bain de soudure. Grâce aux mesures effectuées sur les vidéos segmentées, le comportement des modèles physiques a été validé et des pistes pour améliorer la cohérence entre simulation et expérience ont été énoncées.

1.8.1.5 Détection de défauts de soudure à l'arc

- Contrat de collaboration CRAN–Air Liquide
- Durée : 1 an (2004–2005)
- Chercheurs CRAN : D. Brie, **E.-H. Djermoune**
- Budget : 25 k€

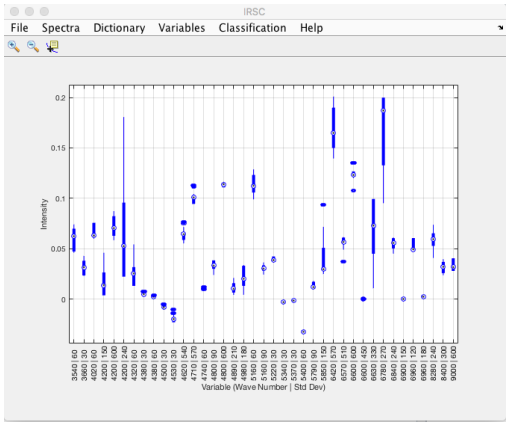
Résumé : L'objectif de cette étude est de détecter des défauts de soudage à l'arc à partir des enregistrements de tension, d'intensité, de vitesse fil et de pression de gaz. Le but est de pouvoir installer un système électronique directement sur des chaînes de production afin de faire le tri automatiquement et en temps réel entre pièces conformes et pièces défectueuses. Il est impossible d'envisager toutes les conditions de soudage que l'on peut rencontrer dans l'industrie. Dans cette étude, nous avons considéré un soudage à clin faible épaisseur (2 mm) avec 3 types de défauts qui se produisent lorsque les pièces à souder et la torche de soudage sont mal positionnées. La méthode développée est basée sur la tension et le courant moyens. Ces deux grandeurs alimentent ensuite une version robuste du détecteur de saut de moyenne de Page-Hinkley.

1.8.2 Développement de logiciels

Mes travaux de recherche académiques et industriels ont souvent donné lieu à des logiciels Matlab qui se présentent sous la forme de GUI[10] pour faciliter leur utilisation. Je donne ici une liste non exhaustive de ces boîtes à outils.

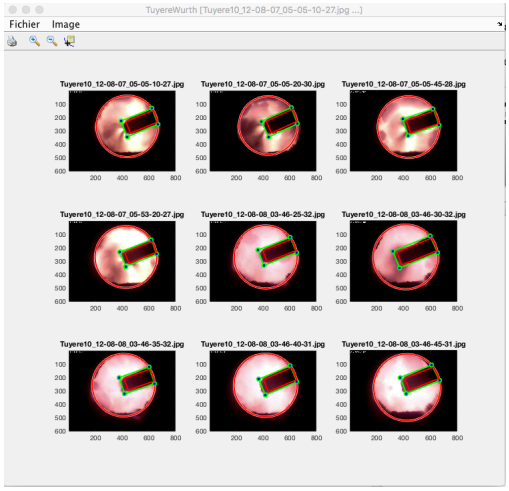
1.8.2.1 IRSC : Infrared Spectra Classifier (2016)

- Logiciel d'analyse, de sélection de variables et de classification en spectroscopie infrarouge
- Projet : TRISPIRABOIS
- Développeur : E.-H. Djermoune
- Contributeurs : L. Belmerhnia, D. Brie (CRAN)
- Publications associées : [Ci20]



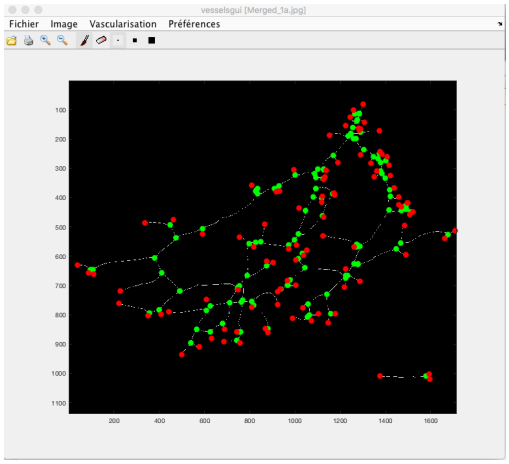
1.8.2.2 TuyereWurth (2016)

- Logiciel dédié à la détection de tuyère et de lance de hauts fourneaux
- Projet : contrat CRAN–Paul Wurth
- Développeur : E.-H. Djermoune
- Publications associées : [R5]



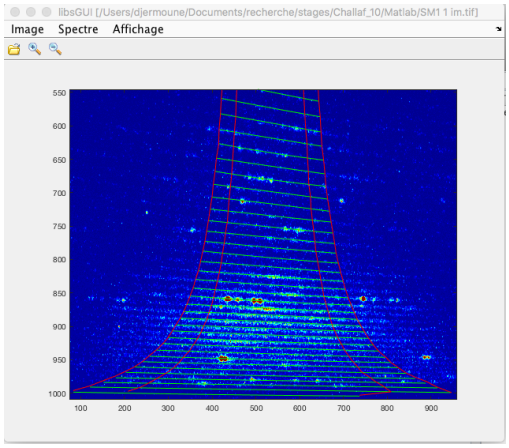
[10] *Graphical User Interface.*

1.8.2.3 vesselsGUI et angioGUI (2014–2015)



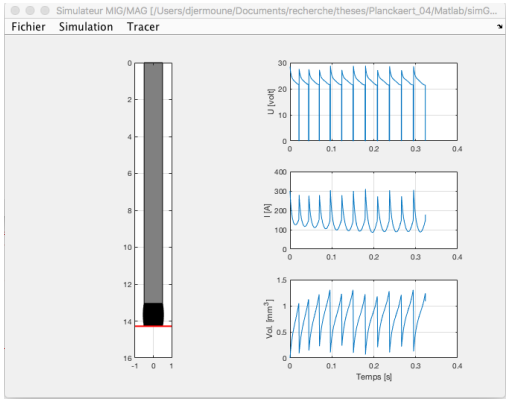
- Logiciels d'analyse et de quantification de réseaux vasculaires (angiogénèse, glandes mammaires)
- Projet : collaboration avec des biologistes du CRAN
- Développeur : E.-H. Djermoune
- Contributeurs : J.-B. Tylcz, K. El Alami-El Ismaili, N. Thomas, B. Faivre, H. Dumond
- Publications associées : [A9, A10, Ci19, Cn6, Cn4]

1.8.2.4 libsGUI (2010)



- Logiciel dédié au filtrage d'images de spectroscopie d'émission optique couplée à l'ablation laser (LIBS)
- Projet : collaboration CRAN–GeoRessources
- Développeur : E.-H. Djermoune
- Contributeurs : O. Challaf (stagiaire), D. Brie (CRAN), P. Robert (GeoRessources)
- Publications associées : [Ci12]

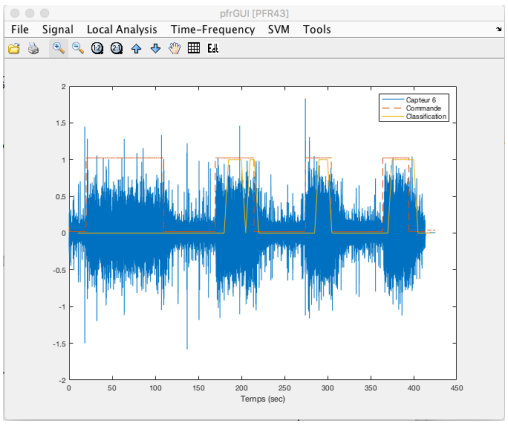
1.8.2.5 simMIGMAG (2008)



- Simulateur de soudage MIG/MAG
- Projet : contrat CRAN–Air Liquide
- Développeur : E.-H. Djermoune
- Contributeurs : J.-P. Planckaert, D. Brie (CRAN)
- Publications associées : [A4]

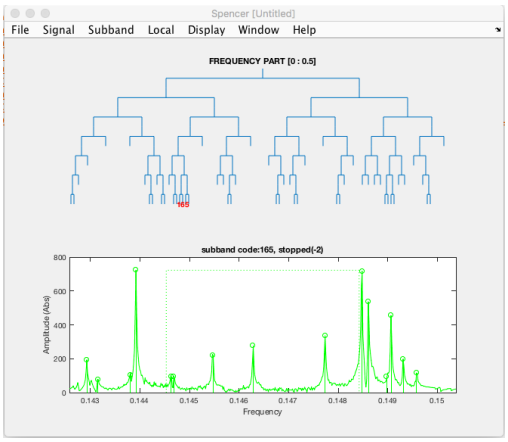
1.8.2.6 pfrGUI (2009)

- Logiciel d'analyse et de détection de défauts (signaux acoustiques des générateurs de vapeur de PFR)
- Projet : contrat CRAN–CEA
- Développeur : E.-H. Djermoune
- Contributeurs : S. Henrot, D. Brie (CRAN)
- Publications associées : [R4]



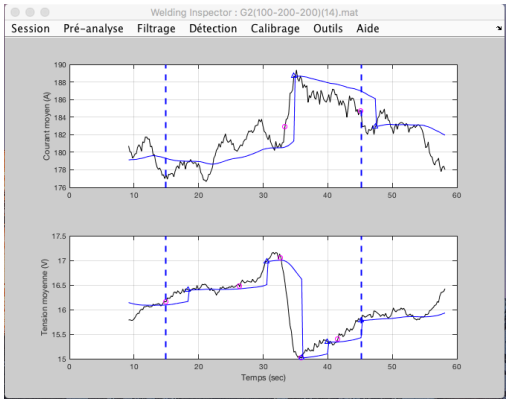
1.8.2.7 SPENCER (2005–2006)

- Logiciel d'estimation des paramètres de signaux de RMN 1–D
- Projet : aucun
- Développeur : E.-H. Djermoune
- Contributeurs : J. Tan Luong Ann (Telecom-Nancy), M. Tomczak (CRAN)
- Publications associées : [O1]



1.8.2.8 WeldingInspector (2003–2004)

- Logiciel d'analyse et de détection de défauts de soudage à l'arc
- Projet : contrat CRAN–Air Liquide
- Développeur : E.-H. Djermoune
- Contributeurs : D. Brie (CRAN)
- Publications associées : [R3, R2, R1]



1.9 Implication dans la vie collective

1.9.1 Collaborations

- Jérôme Idier (LS2N, Nantes) et Charles Soussen (CRAN) dans le cadre de la thèse de Thi-Thanh Nguyen.
- Vincent Mazet (ICube, Strasbourg) depuis 2012 : approches parcimonieuses pour la décomposition de signaux spectroscopiques.
- Pascal Robert (laboratoire GeoRessources, Nancy) : spectroscopie d'émission optique couplée à l'ablation laser.
- Cédric Carteret (laboratoire LCPME, Nancy) : analyse de signaux de spectroscopie proche infrarouge.
- Béatrice Faivre, Noémie Thomas et Hélène Dumond (Laboratoire de biologie Sigreto, intégré au CRAN en 2012) depuis 2010 : quantification de réseaux vasculaires tumoraux.

1.9.2 Animation scientifique

- Animateur de l'équipe-projet SiMul (Signaux Multidimensionnels) du département SBS depuis août 2017.
- Participation au jury de thèse de K. El Alaoui Lasmaili (Université de Lorraine, 04/04/2017, Examinateur).
- Responsable et co-responsable de contrats de recherche.
- Révision d'articles de revues : *Applied Mathematical Modelling*, *IEEE Transactions on Signal Processing*, *IEEE Signal Processing Letters*, *IEEE Transactions on Image Processing*, *EURASIP Journal on Advances in Signal Processing*, *Signal Processing*, *Signal, Image and Video Processing*, *Mechanical Systems and Signal Processing*, *ISA Transactions*.
- Révision d'articles pour les conférences : *European Signal Processing Conference* (EUSIPCO), *International Conference on Image Processing* (ICIP), *International Conference on Industrial Technology* (ICIT), *Colloques Gretsi*.

1.9.3 Rayonnement scientifique

- Présentation des travaux de thèse de Leila Belmerhnia à un Forum de la Fédération Charles Hermite (janvier 2016) et à la journée « Algorithmes gloutons pour l'optimisation sous contrainte de parcimonie » du GdR ISIS (juin 2016).
- Présentation des travaux de thèse de Yingying Song à la journée « Imagerie hyperspectrale : quelles données ? quels traitements ? quelles applications ? » du GdR ISIS (avril 2016).
- Titulaire de la prime d'encadrement doctorale et de recherche 2017-2021 (avis CNU : A).

1.9.4 Responsabilités

- Responsable d'année en master Ingénierie de Systèmes Complexe, Université de Lorraine, depuis 2010.
- Rapporteur au comité de sélection pour le poste PRAG n°1186, Université de Lorraine, 2012.
- Membre de la commission IST (information scientifique et technique) du CRAN depuis 2013.
- Responsable bibliothèque CRAN, site Faculté des Sciences, 2005–2010.
- Responsable licence IUP GEII, Université Henri Poincaré, 2004–2005.

1.10 Production scientifique

1.10.1 Mémoire de thèse

[Th] E.-H. Djermoune. *Estimation des paramètres de sinusoïdes amorties par décomposition en sous-bandes adaptative. Application à la spectroscopie RMN.* PhD thesis, Université Henri Poincaré Nancy 1, juillet 2003.

1.10.2 Chapitres d'ouvrages

[O3] S. Henrot, **E.-H. Djermoune**, and D. Brie. Supervision and safety of complex systems. In Yves Vandenboomgaerde Nada Matta and Jean Arlat, editors, *SVM time-frequency classification for the detection of injection states*, pages 231–245. ISTE Ltd and John Wiley Sons Inc, 2012.

[O2] S. Henrot, **E.-H. Djermoune**, and D. Brie. Supervision, surveillance et sûreté de fonctionnement des grands systèmes. In Yves Vandenboomgaerde et Jean Arlat Nada Matta, editor, *Classification temps-fréquence par SVM pour la détection d'états d'injection*, pages 273–288. Hermès, 2012.

[O1] M. Tomczak, **E.-H. Djermoune**, and P. Mutzenhardt. Progress in magnetic resonance research. In Bernard C. Castelman, editor, *High-resolution MR spectroscopy via adaptive subband decomposition*, pages 241–289. Novascience Publishers, 2007.

1.10.3 Articles de revues internationales à comité de lecture

[A11] S. Sahnoun, **E.-H. Djermoune**, D. Brie, and P. Comon. A simultaneous sparse approximation method for multidimensional harmonic retrieval. *Signal Processing*, 131 :36–48, 2017.

[A10] K. El Alaoui-Lasmaili, **E.-H. Djermoune**, J.-B. Tylcz, D. Meng, F. Plénat, N. Thomas, and B. Faivre. A new algorithm for a better characterization and timing of the anti-VEGF vascular effect named "normalization". *Angiogenesis*, 20(1) :149–162, 2017.

[A9] C. Thiebaut, C. Chamard-Jovenin, A. Chesnel, C. Morel, **E.-H. Djermoune**, T. Boukhobza, and H. Dumond. Mammary epithelial cell phenotype disruption in vitro and in vivo through ERalpha36 overexpression. *Plos One*, 2017.

[A8] Y. Song, D. Brie, **E.-H. Djermoune**, and S. Henrot. Regularization parameter estimation for non-negative hyperspectral image deconvolution. *IEEE Transactions on Image Processing*, 25(11) :5316–5330, 2016.

[A7] J.-B. Tylcz, K. El Alaoui-Lasmaili, **E.-H. Djermoune**, N. Thomas, B. Faivre, and T. Bastogne. Data-driven modeling and characterization of anti-angiogenic molecule effects on tumoral vascular density. *Biomedical Signal Processing and Control*, 20 :52–60, July 2015.

[A6] **E.-H. Djermoune**, M. Tomczak, and D. Brie. NMR data analysis : A time-domain parametric approach using adaptive sub band decomposition. *Oil & Gas Science and Technology - Rev. IFP Energies nouvelles* Special issue on Advances in Signal Processing and Image Analysis for Physico-Chemical, Analytical Chemistry and Chemical Sensing, 69 :229–244, 2014.

[A5] S. Sahnoun, **E.-H. Djermoune**, C. Soussen, and D. Brie. Sparse multidimensional modal analysis using a multigrid dictionary refinement. *EURASIP Journal on Advances in Signal Processing*, 60, 2012.

[A4] J.-P. Planckaert, **E.-H. Djermoune**, D. Brie, F. Briand, and F. Richard. Modeling of MIG/MAG welding with experimental validation using an active contour algorithm applied on high speed movies. *Applied Mathematical Modelling*, 34(4) :1004–1020, 2010.

[A3] **E.-H. Djermoune** and M. Tomczak. Perturbation analysis of subspace-based methods in estimating a damped complex exponential. *IEEE Transactions on Signal Processing*, 57(11) :4558–4563, 2009.

[A2] **E.-H. Djermoune**, M. Tomczak, and P. Mutzenhardt. A new adaptive subband decomposition approach for automatic analysis of NMR data. *Journal of Magnetic Resonance*, 169 :73–84, 2004.

[A1] M. Tomczak and **E.-H. Djermoune**. A subband ARMA modeling approach to high resolution NMR spectroscopy. *Journal of Magnetic Resonance*, 158(1) :86–98, 2002.

1.10.4 Conférences internationales avec actes et comité de lecture

[Ci24] Y. Song, **E.-H. Djermoune**, J. Chen, C. Richard, and D. Brie. Online deconvolution for pushbroom hyperspectral imaging systems. In *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2017*, Curaçao, Dutch Antilles, 2017.

[Ci23] P. Guyot, L. Batista, **E.-H. Djermoune**, J.-M. Moureaux, L. Doerr, M. Beckler, and T. Bastogne. Comparison of compression solutions for impedance and field potential signals of cardiomyocytes. In *Proc. of the 44th Annual Conference on Computing in Cardiology, CinC'17*, Rennes, France, 2017.

[Ci22] L. Batista, **E.-H. Djermoune**, and T. Bastogne. Identification of dynamical systems population described by a mixed effect ARX model structure. In *20th IFAC World Congress*, Toulouse, France, 2017.

[Ci21] Y. Song, D. Brie, **E.-H. Djermoune**, and S. Henrot. Minimum distance criterion for non-negative image deconvolution. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2016*, Shanghai, China, 2016.

[Ci20] L. Belmerhnia, **E.-H. Djermoune**, C. Carteret, and D. Brie. A regularized sparse approximation method for hyperspectral image classification. In *IEEE Statistical Signal Processing Conference, SSP 2016*, Palma de Mallorca, Spain, 2016.

[Ci19] C. Chamard-Jovenin, A. Chesnel, E. Bresso, C. Morel, C. Thiébaut, M. Smail-Tabbone, **E.-H. Djermoune**, M.-D. Devignes, T. Boukhobza, and H. Dumond. Transgenerational effects of ERalpha36 over-expression on mammary gland development and molecular phenotype : clinical perspective for breast cancer risk and therapy. In *21st World Congress on Advances in Oncology and 19th International Symposium on Molecular Medicine*, Athens, Greece, 2016.

[Ci18] L. Batista, T. Bastogne, and **E.-H. Djermoune**. Identification of dynamical biological systems based on mixed-effect models. In *Symposium on Applied Computing, ACM-SAC*, Pisa, Italy, 2016.

[Ci17] L. Belmerhnia, **E.-H. Djermoune**, C. Carteret, and D. Brie. Simultaneous regularized sparse approximation for wood wastes NIR spectra features selection. In *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2015*, Cancun, Mexico, 2015.

[Ci16] L. Batista, T. Bastogne, and **E.-H. Djermoune**. Identification of biological dynamic systems described by random effects models. In *IEEE Engineering in Medicine and Biology Conference, EMBC 2015*, Milan, Italy, 2015.

[Ci15] C. Chamard, A. Chesnel, **E.-H. Djermoune**, C. Morel, T. Boukhobza, and H. Dumond. Long chain alkylphenol mixture promotes breast cancer initiation and progression through an ERα36-mediated mechanism. In *2ème colloque scientifique de la Fondation Rovaltain*, Valence, France, 2015.

[Ci14] L. Belmerhnia, **E.-H. Djermoune**, and D. Brie. Greedy methods for simultaneous sparse approximation. In *European Signal Processing Conference, EUSIPCO 2014*, Lisbon, Portugal, 2014.

[Ci13] S. Sahnoun, **E.-H. Djermoune**, and D. Brie. Sparse modal estimation of 2-D NMR signals. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2013*, Vancouver, Canada, 2013.

[Ci12] P. Robert, O. Challaf, **E.-H. Djermoune**, D. Brie, and C. Fabre. Spectra processing for minor elements detection using single pulse LIBS at microscopic scale. In *Euro-Mediterranean Symposium on LIBS, EMSLIBS 2013*, Bari, Italie, 2013.

[Ci11] S. Sahnoun, **E.-H. Djermoune**, and D. Brie. Sparse multigrid modal estimation : Initial grid selection. In *European Signal Processing Conference, EUSIPCO 2012*, Bucharest, Roumania, 2012.

[Ci10] S. Sahnoun, **E.-H. Djermoune**, and D. Brie. Sparse multiresolution modal estimation. In *IEEE Statistical Signal Processing Workshop, SSP 2011*, Nice, France, 2011.

[Ci9] **E.-H. Djermoune**, M. Thomassin, and M. Tomczak. First-order analysis of the mode and amplitude estimates of a damped sinusoid using Matrix Pencil. In *European Signal Processing Conference, EUSIPCO 2009*, Glasgow, Scotland, 2009.

[Ci8] **E.-H. Djermoune**, D. Brie, and M. Tomczak. A subband algorithm for estimating the parameters of two-dimensional exponential signals. In *European Signal Processing Conference, EUSIPCO 2009*, Glasgow, Scotland, 2009.

[Ci7] **E.-H. Djermoune**, G. Kasalica, and D. Brie. Estimation of the parameters of two-dimensional NMR spectroscopy signals using and adapted subband decomposition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2008*, Las Vegas, USA, April 2008.

[Ci6] J.-P. Planckaert, **E.-H. Djermoune**, D. Brie, F. Briand, and F. Richard. Metal transfer characterization with an active contour algorithm in MIG/MAG welding movies. In *IEEE International Conference on Automation Science and Engineering, CASE 2007*, pages 933–938, Scottsdale, USA, 2007.

[Ci5] J.-P. Planckaert, **E.-H. Djermoune**, D. Brie, F. Briand, and F. Richard. Droplet features extraction with a dynamic active contour for MIG/MAG welding modelling. In *International Conference on Systems Engineering, ICSE 2006*, pages 365–370, Coventry, England, 2006.

[Ci4] **E.-H. Djermoune** and M. Tomczak. Statistical analysis of the Kumaresan-Tufts and Matrix Pencil methods in estimating a damped sinusoid. In *European Signal Processing Conference, EUSIPCO 2004*, Vienna, Austria, 2004.

[Ci3] **E.-H. Djermoune** and M. Tomczak. An adapted filterbank for frequency estimation. In *European Signal Processing Conference, EUSIPCO 2004*, Vienna, Austria, 2004.

[Ci2] **E.-H. Djermoune** and M. Tomczak. An adaptive subband decomposition method for high resolution nuclear magnetic resonance spectroscopy. In *International Symposium on Physics in Signal and Image Processing, PSIP 2003*, Grenoble, France, 2003.

[Ci1] **E.-H. Djermoune** and M. Tomczak. SNR enhancement of damped exponential signals in noise. In *European Signal Processing Conference, EUSIPCO 2002*, Toulouse, France, 2002.

1.10.5 Conférences nationales avec actes et comité de lecture

[Cn10] Y. Song, **E.-H. Djermoune**, J. Chen, C. Richard, and D. Brie. Déconvolution en ligne d'images hyperspectrales pour les systèmes d'imagerie pushbroom. In *26ème Colloque GRETSI Traitement du Signal & des Images*, Juan-les-Pins, France, 2017.

[Cn9] T. Nguyen, C. Soussen, J. Idier, and **E.-H. Djermoune**. An optimized version of non-negative OMP. In *26ème Colloque GRETSI Traitement du Signal & des Images*, Juan-les-Pins, France, 2017.

[Cn8] L. Belmerhnia, **E.-H. Djermoune**, C. Carteret, and D. Brie. Approximation parcimonieuse simultanée constante par morceaux. In *25ème Colloque GRETSI Traitement du Signal & des Images*, Lyon, France, 2015.

[Cn7] L. Batista, T. Bastogne, and **E.-H. Djermoune**. Identification de systèmes biologiques dynamiques à effets aléatoires. In *25ème Colloque GRETSI Traitement du Signal & des Images*, Lyon, France, 2015.

[Cn6] K. El Alaoui Lasmaili, J.-B. Tyclz, **E.-H. Djermoune**, N. Thomas, and B. Faivre. Optimization of the use of anti-angiogenics in glioblastoma using mathematical modeling. In *6ème Congrès de la Société Française d'Angiogenèse*, Paris, France, 2015.

[Cn5] L. Batista, T. Bastogne, and **E.-H. Djermoune**. Classification des réponses de cellules cancéreuses fondée sur l'analyse de signaux d'impédancemétrie cellulaire. In *Congrès Chimiométrie XVI*, Geneva, Switzerland, 2015.

[Cn4] K. El Alaoui Lasmaili, J.-B. Tylcz, **E.-H. Djermoune**, N. Thomas, and B. Faivre. Characterization of the anti-angiogenic and anti-vascular effects of bevacizumab using mathematical tools : an effective way to determine the normalization window ? In *5ème Congrès de la Société Française d'Angiogenèse*, Chamonix, France, April 2014.

[Cn3] S. Sahnoun, **E.-H. Djermoune**, and D. Brie. Estimation modale de signaux 2-D par approximation parcimonieuse simultanée. In *24ème Colloque GRETSI Traitement du Signal & des Images*, Brest, France, 2013.

[Cn2] V. Mazet, C. Soussen, and **E.-H. Djermoune**. Décomposition de spectres en motifs paramétriques par approximation parcimonieuse. In *24ème Colloque GRETSI Traitement du Signal & des Images*, Brest, France, 2013.

[Cn1] S. Sahnoun, **E.-H. Djermoune**, C. Soussen, and D. Brie. Analyse modale bidimensionnelle par approximation parcimonieuse et multirésolution. In *23ème Colloque GRETSI Traitement du Signal & des Images*, Bordeaux, France, 2011.

1.10.6 Conférences internationales sans comité de lecture

[Cs2] L. Batista, T. Bastogne, and **E.-H. Djermoune**. Parameters identification of a population of systems based on EM algorithm. In *The Population Approach Group Meeting, PAGE*, Lisbon, Portugal, 2016.

[Cs1] **E.-H. Djermoune**, G. Kasalica, and D. Brie. Two-dimensional NMR signal analysis with an adapted subband decomposition. In *22nd IAR Annual Meeting*, Grenoble, France, 2007.

1.10.7 Rapports de contrats industriels

[R5] **E.-H. Djermoune**, D. Theilliol, and J.-C. Ponsart. Détection de phénomènes tuyère dans les hauts fourneaux. Rapport de fin de contrat, CRAN-Paul Wurth, 31 pages, février 2017.

[R4] S. Henrot, **E.-H. Djermoune**, and D. Brie. Surveillance des réacteurs SFR : Analyse des données acoustiques des GV de PFR. Rapport de fin de contrat, CRAN-CEA, 59 pages, décembre 2009.

[R3] **E.-H. Djermoune**, D. Brie, F. Briand, and F. Richard. Calibrage de l'algorithme de détection : Utilisation des temps arc et court-circuit. Rapport de contrat, CRAN-Air Liquide, 26 pages, octobre 2004.

[R2] **E.-H. Djermoune**, D. Brie, F. Briand, and F. Richard. Détection et classification de défauts de soudure à l'arc : Constitution d'une boîte à outils de pré-analyse. Rapport de contrat CRAN-Air Liquide, 20 pages, février 2004.

[R1] **E.-H. Djermoune**, D. Brie, F. Briand, and F. Richard. Détection et classification de défauts de soudure à l'arc : Approche expérimentale. Rapport de contrat CRAN-Air Liquide, 32 pages, décembre 2003.

Deuxième partie

Synthèse des activités de recherche

Avant-propos

Dans la suite du manuscrit, je vais détailler la partie de mes travaux de recherche consacrés à l'estimation modale mono et multidimensionnelle ainsi qu'à la sélection de variables pour la classification de signaux de spectroscopie proche infrarouge. Ce choix est dicté principalement par le temps que j'ai consacré à chacune de ces activités. En terme de structure, j'ai opté pour un découpage qui reflète mes contributions théoriques, algorithmiques et appliquées sur trois sujets. Enfin, j'ai choisi de respecter l'ordre chronologique des travaux réalisés.

La synthèse de mes contributions est organisée en trois chapitres :

- Chapitre 2. Analyse statistique d'algorithmes d'estimation modale 1-D. Les résultats présentés sont issues de mes travaux d'après-thèse.
- Chapitre 3. Estimation modale multidimensionnelle. Les développements ont été réalisés durant et après la thèse de Souleymen Sahnoun.
- Chapitre 4. Analyse des signaux de spectroscopie infrarouge : sélection de variables et classification. La méthode de sélection de variable proposée est issue de la thèse de Leïla Belmerhnia.

Notations

x, λ, N	:	scalaires
$\boldsymbol{x}, \boldsymbol{\phi}$	:	vecteurs
$\boldsymbol{X}, \boldsymbol{\Phi}$	:	matrices
$\boldsymbol{x}_i, \boldsymbol{x}^j$	:	colonne i et ligne j d'un matrice $\boldsymbol{X}$
$\boldsymbol{\mathcal{X}}$	:	tenseur
$\|\|\boldsymbol{x}\|\|_p$	:	norme ℓ_p, $p > 0$; $\|\|\boldsymbol{x}\|\|_p = (\sum_i \|x_i\|^p)^{1/p}$
$\|\|\boldsymbol{x}\|\|_0$	:	pseudo-norme ℓ_0 ; $\|\|\boldsymbol{x}\|\|_0 = \sum_{i,x_i\neq 0} \|x_i\|^0$
$\|\|\boldsymbol{X}\|\|_F$	:	norme de Frobenius ; $\|\|\boldsymbol{X}\|\|_F = (\sum_{i,j} \|x_{ij}\|^2)^{1/2}$
$\|\|\boldsymbol{X}\|\|_{p,q}$	:	norme mixte ℓ_p/ℓ_q ; $\|\|\boldsymbol{X}\|\|_{p,q} = (\sum_i \|\|\boldsymbol{x}^i\|\|_q^p)^{1/p}$ où $\boldsymbol{x}^i$ est la i^{e} ligne de $\boldsymbol{X}$
$\boldsymbol{X}^\mathsf{T}$	:	transposée de $\boldsymbol{X}$
$\boldsymbol{X}^\mathsf{H}$	:	transposée et conjugué de $\boldsymbol{X}$
$\boldsymbol{X}^\dagger$	:	pseudo-inverse de Moore-Penrose de $\boldsymbol{X}$
$\boldsymbol{A} \otimes \boldsymbol{B}$	:	produit de Kronecker de matrices
$\boldsymbol{A} \odot \boldsymbol{B}$	:	produit de Khatri-Rao de matrices (ayant un même nombre de colonnes)

$\boldsymbol{A} * \boldsymbol{B}$	:	produit de Hadamard (terme à terme) de matrices de mêmes dimensions
$\boldsymbol{a} \circ \boldsymbol{b}$	:	produit extérieur de vecteurs ; $(\boldsymbol{a} \circ \boldsymbol{b})_{ij} = a_i b_j$
$\boldsymbol{\mathcal{A}} \underset{r}{\bullet} \boldsymbol{B}$	:	produit selon le mode r d'un tenseur et d'une matrice ; si $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{M_1 \times \ldots \times M_R}$ et $\boldsymbol{B} \in K \times M_r$, alors $(\boldsymbol{\mathcal{A}} \underset{r}{\bullet} \boldsymbol{B})_{m_1 \ldots m_{r-1} k m_{r+1} \ldots m_R} = \sum_{m_r=1}^{M_r} x_{m_1 \ldots m_R} b_{k m_r}$
$\boldsymbol{\mathcal{A}} \underset{r}{\sqcup} \boldsymbol{\mathcal{B}}$	:	concaténation de tenseurs selon le mode r ; si $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{M_1 \times \ldots M_{r-1} \times K_1 \times M_{r+1} \times \ldots \times M_R}$ et $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{M_1 \times \ldots M_{r-1} \times K_2 \times M_{r+1} \times \ldots \times M_R}$, alors $\boldsymbol{\mathcal{A}} \underset{r}{\sqcup} \boldsymbol{\mathcal{B}} \in \mathbb{R}^{M_1 \times \ldots M_{r-1} \times (K_1 + K_2) \times M_{r+1} \times \ldots \times M_R}$
$\boldsymbol{A}_{(r)}$	:	matricisation du tenseur $\boldsymbol{\mathcal{A}}$ selon le mode r ; si $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{M_1 \times \ldots \times M_R}$, alors $\boldsymbol{A}_{(r)} \in \mathbb{R}^{M_r \times (M_1 \ldots \times M_{r-1} M_{r+1} \ldots M_R)}$
$f \diamond g$	:	opérateur de composition de fonctions ; $f \diamond g = f(g)$

Chapitre 2

Analyse statistique d'algorithmes d'estimation modale 1-D

L'estimation modale est le problème qui consiste à déterminer les paramètres d'un signal composé de F sinusoïdes complexes amorties dont le modèle s'écrit :

$$\tilde{y}(m) = y(m) + e(m) \triangleq \sum_{f=1}^{F} c_f a_f^{m-1} + e(m), \quad m = 1, \ldots, M \tag{2.1}$$

où $a_f = \exp(\alpha_f + j\omega_f)$ représente un « mode » du signal avec une amplitude $c_f = \lambda_f \exp(j\phi_f)$, où $\lambda_f \in \mathbb{R}^+$ et $\phi_f \in \mathbb{R}$. Les paramètres $\alpha_f \in \mathbb{R}^-$ et $\omega_f \in \mathbb{R}$ sont le facteur d'amortissement et la pulsation, respectivement. Le bruit complexe $e(m)$ est supposé blanc, gaussien, de moyenne nulle et de variance σ^2. Le problème consiste à estimer les quadruplets $\{\alpha_f, \omega_f, \lambda_f, \phi_f\}_{f=1}^{F}$ à partir des données bruitées $\{\tilde{y}(m)\}_{m=1}^{M}$.

Ce problème est très ancien en traitement du signal et a donné lieu à plusieurs algorithmes d'estimation. Les méthodes haute-résolution[1] utilisent la prédiction linéaire [Kumaresan82, Tufts82], les sous-espaces signal ou bruit [Schmidt79, Roy86, Stoica89a], les faisceaux de matrices [Hua90], l'espace d'état [Kung83] ou encore le principe du maximum de vraisemblance [Bresler86, Yau93]. L'analyse statistique des algorithmes associés a donné lieu à plusieurs publications, plus particulièrement dans le cas de signaux harmoniques (i.e. lorsque $\alpha_f = 0, \forall f$) [Hua88, Rao88, Stoica89a, Stoica89b, Rao89, Swindlehurst92, Rao92]. Très peu de travaux se sont intéressés au cas de signaux amortis [Porat87, Okhovat89, Kot93, Steedly94]. Ma contribution a été d'étudier les propriétés statistiques de trois algorithmes dans le cas de signaux exponentiellement amortis. Mon objectif a été d'obtenir des expressions sur la variance des paramètres qui soient facilement exploitables afin, par exemple, de choisir les hyperparamètres permettant d'atteindre la variance minimale.

2.1 Algorithmes étudiés

Algorithme *Backward Linear Prediction* Pour les signaux sinusoïdaux amortis, Kumaresan et Tufts [Kumaresan82] ont proposé un algorithme d'estimation basé sur la prédiction linéaire arrière

[1] Ce qui exlcut les méthodes à base de transformée de Fourier comme le périodogramme ou le corrélogramme.

(*backward linear prediction* : BLP). Cet algorithme peut être résumé par les étapes suivantes :

1. En utilisant les données disponibles, former le système d'équations $\tilde{\boldsymbol{Y}}_1\tilde{\boldsymbol{b}} \approx -\tilde{\boldsymbol{y}}_0$:

$$\begin{bmatrix} \tilde{y}(2) & \tilde{y}(3) & \cdots & \tilde{y}(L+1) \\ \tilde{y}(3) & y(4) & \cdots & \tilde{y}(L+2) \\ \vdots & \vdots & & \vdots \\ \tilde{y}(M-L+1) & \tilde{y}(M-L+2) & \cdots & \tilde{y}(M) \end{bmatrix} \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \vdots \\ \tilde{b}_L \end{bmatrix} \approx - \begin{bmatrix} \tilde{y}(1) \\ \tilde{y}(2) \\ \vdots \\ \tilde{y}(M-L) \end{bmatrix}. \tag{2.2}$$

   Le vecteur $\tilde{\boldsymbol{b}}$ contient les coefficients de prédiction arrière et $L \geq F$ est appelé « ordre de prédiction ».

2. Obtenir la décomposition en valeurs singulières (SVD) de $\tilde{\boldsymbol{Y}}_1$ et mettre à 0 les $L-F$ valeurs singulières les plus faibles. La matrice $\bar{\boldsymbol{Y}}_1$ obtenue est la meilleure approximation de rang F, au sens de la norme de Frobenius, de la matrice $\tilde{\boldsymbol{Y}}_1$.

3. Obtenir une estimation $\tilde{\boldsymbol{b}}$ en utilisant la pseudoinverse de rang réduit de $\tilde{\boldsymbol{Y}}_1$:

$$\tilde{\boldsymbol{b}} = -\bar{\boldsymbol{Y}}_1^{\dagger}\tilde{\boldsymbol{y}}_0 \tag{2.3}$$

   où † désigne la pseudoinverse.

4. Calculer les racines $\{\tilde{z}_i\}_{i=1}^{L}$ du polynôme $\tilde{B}(z) = 1 + \sum_{i=1}^{L} \tilde{b}_i z^{-i} = \prod_{i=1}^{L}(1 - \tilde{z}_i z^{-1})$ et sélectionner celles qui sont à l'extérieur du cercle unité dans le plan complexe. Elles correspondent à l'inverse des modes du signal (i.e. $\tilde{a}_f = 1/\tilde{z}_f$ pour $f = 1, \ldots, F$).

Algorithme *Matrix Pencil* La méthode *matrix pencil* (MP) [Hua90] consiste également en quatre étapes :

1. Former deux matrices Hankel $\tilde{\boldsymbol{Y}}_0$ et $\tilde{\boldsymbol{Y}}_1$. La matrice $\tilde{\boldsymbol{Y}}_1$ est la même que précédemment et $\tilde{\boldsymbol{Y}}_0$ en est une version décalée :

$$\tilde{\boldsymbol{Y}}_0 = \begin{bmatrix} \tilde{y}(1) & \tilde{y}(2) & \cdots & \tilde{y}(L) \\ \tilde{y}(2) & y(3) & \cdots & \tilde{y}(L+1) \\ \vdots & \vdots & & \vdots \\ \tilde{y}(M-L) & \tilde{y}(M-L+1) & \cdots & \tilde{y}(M-1) \end{bmatrix} \tag{2.4}$$

2. Calculer l'approximation de rang F ($\bar{\boldsymbol{Y}}_1$) de la matrice $\tilde{\boldsymbol{Y}}_1$ en utilisant la SVD.

3. Obtenir une estimation de la matrice $\tilde{\boldsymbol{Z}}$:

$$\tilde{\boldsymbol{Z}} = \bar{\boldsymbol{Y}}_1^{\dagger}\tilde{\boldsymbol{Y}}_0 \tag{2.5}$$

4. Les modes du signal sont les inverses des F valeurs propres de $\tilde{\boldsymbol{Z}}$ situées à l'extérieur du cercle unité.

Algorithme *Direct Data Approximation* Dans la méthode *direct data approximation* (DDA) [Kung83], le signal $y(m)$ est vu comme la réponse d'un système linéaire discret dont la matrice de transition $\boldsymbol{A}$ a des valeurs propres égales aux modes recherchés. Le problème revient donc à estimer cette matrice :

1. Former la matrice $\tilde{\boldsymbol{Y}}$ de la façon suivante :

$$\tilde{\boldsymbol{Y}} = \begin{bmatrix} \tilde{y}(1) & \tilde{y}(2) & \cdots & \tilde{y}(L+1) \\ \tilde{y}(2) & y(3) & \cdots & \tilde{y}(L+2) \\ \vdots & \vdots & & \vdots \\ \tilde{y}(M-L) & \tilde{y}(M-L+1) & \cdots & \tilde{y}(M) \end{bmatrix} \tag{2.6}$$

2. Obtenir la SVD de $\tilde{\boldsymbol{Y}}$:

$$\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{V}}^{\mathsf{H}} \triangleq \tilde{\boldsymbol{U}}_F\tilde{\boldsymbol{S}}_F\tilde{\boldsymbol{V}}_F^{\mathsf{H}} + \tilde{\boldsymbol{U}}_b\tilde{\boldsymbol{S}}_b\tilde{\boldsymbol{V}}_b^{\mathsf{H}} \tag{2.7}$$

   où $\tilde{\boldsymbol{S}} \in \mathbb{R}^{(M-L)\times(L+1)}$ est une matrice diagonale contenant les valeurs singulières rangées dans un ordre décroissant. Les matrices $\tilde{\boldsymbol{U}} \in \mathbb{C}^{(M-L)\times(M-L)}$ et $\tilde{\boldsymbol{V}} \in \mathbb{C}^{(L+1)\times(L+1)}$ sont des matrices unitaires contenant les vecteurs singuliers à gauche et à droite pour $\tilde{\boldsymbol{Y}}$, respectivement. $\tilde{\boldsymbol{U}}_F$ (resp. $\tilde{\boldsymbol{V}}_F$) est la matrice formée par les F premières colonnes de $\tilde{\boldsymbol{U}}$ (resp. $\tilde{\boldsymbol{V}}$) et $\tilde{\boldsymbol{S}}_F \in \mathbb{R}^{F\times F}$ contient les F plus grandes valeurs singulières. Les matrices $\tilde{\boldsymbol{U}}_b$, $\tilde{\boldsymbol{V}}_b$ et $\tilde{\boldsymbol{S}}_b$ contiennent le reste des éléments dans $\tilde{\boldsymbol{U}}$, $\tilde{\boldsymbol{S}}$ et $\tilde{\boldsymbol{V}}$, respectivement.

3. Obtenir une estimation de la matrice d'observabilité $\tilde{\boldsymbol{O}}$:

$$\tilde{\boldsymbol{O}} = \tilde{\boldsymbol{U}}_F\tilde{\boldsymbol{S}}_F^{\frac{1}{2}}. \tag{2.8}$$

   Une estimation de la matrice $\tilde{\boldsymbol{A}}$ est alors donnée par :

$$\tilde{\boldsymbol{A}} = \tilde{\boldsymbol{O}}_1^{\dagger}\tilde{\boldsymbol{O}}_2 \tag{2.9}$$

   où $\tilde{\boldsymbol{O}}_1$ et $\tilde{\boldsymbol{O}}_2$ sont obtenues à partir de $\tilde{\boldsymbol{O}}$ en éliminant la dernière et la première ligne, respectivement.

4. Les modes du signal sont les F valeurs propres de $\tilde{\boldsymbol{A}}$.

2.2 Perturbation au premier ordre du mode du signal

Afin d'obtenir des variances d'estimation ayant des expressions compactes, je me suis intéressé au cas où le signal $y(m)$ ne contient qu'une seule composante ($F = 1$). Pour alléger les notations, les paramètres ne seront pas indexés par f.

Tous les algorithmes étudiés partagent l'étape de « filtrage des données » qui revient à calculer la pseudoinverse de rang réduit des matrices de données ($\tilde{\boldsymbol{Y}}_1$ ou $\tilde{\boldsymbol{Y}}$). La démarche utilisée pour déterminer la variance d'estimation commence donc par l'analyse de perturbation au premier ordre de la valeur singulière principale de ces matrices et du vecteur singulier associé en exploitant les résultats de Wilkinson [Wilkinson65]. L'erreur est ensuite propagée jusqu'à l'estimation du mode

a. Lorsque le niveau de bruit est faible[2], les expressions de la perturbation au 1er ordre de $\tilde{a}$ sont [Djermoune09b] :

$$\text{BLP :} \qquad \tilde{a} = a + \frac{a^2}{\sigma_1 \beta a^L} \boldsymbol{u}_1^{\mathsf{H}} (\boldsymbol{E}_1 \boldsymbol{v}_1 - \sqrt{r\zeta_L}\boldsymbol{e}_0) \tag{2.10}$$

$$\text{MP :} \qquad \tilde{a} = a + \frac{a}{\sigma_1} \boldsymbol{u}_1^{\mathsf{H}} (\boldsymbol{E}_1 - a\boldsymbol{E}_0) \boldsymbol{v}_1 \tag{2.11}$$

$$\text{DDA :} \qquad \tilde{a} = a + \frac{1}{\text{trace}(\boldsymbol{O}_1^{\mathsf{H}} \boldsymbol{O}_1)} \boldsymbol{O}_1^{\mathsf{H}} (\tilde{\boldsymbol{O}}_1 - a\tilde{\boldsymbol{O}}_2) \tag{2.12}$$

où $r = |a|^2 = \exp(2\alpha)$, $\zeta_L = \sum_{m=0}^{L-1} r^m$, $\beta = (\zeta_L - Lr^L)/a^{L-1}(1-r^L)$. σ_1 est la valeur singulière principale de $\boldsymbol{Y}_1$ associée aux vecteurs à gauche et à droite $\boldsymbol{u}_1$ et $\boldsymbol{v}_1$, respectivement. $\boldsymbol{E}_0$, $\boldsymbol{E}_1$ et $\boldsymbol{e}_0$ sont formés d'échantillons du bruit $e(m)$ de la même façon que $\boldsymbol{Y}_0$, $\boldsymbol{Y}_1$ et $\boldsymbol{y}_0$, respectivement.

Calcul de variance

Nous avons montré que les algorithmes étudiés donnent des estimations non biaisées du mode a. La variance d'estimation par BLP de la pulsation et de l'amortissement est donnée par :

$$\text{var}\{\tilde{\omega}\}^{BLP} = \text{var}\{\tilde{\alpha}\}^{BLP} = \frac{\sigma^2(1-r)^3}{2\lambda^2 r} \cdot \begin{cases} \dfrac{(1+r^{M-L})[1-(2L+1)(1-r)r^L - r^{2L+1}]}{(1-r^{M-L})^2[1-r^L - L(1-r)r^L]^2} & \text{si } L \leq M/2, \\ \dfrac{(1-r^{M-L})(1+r^L)(1+r^{L+1}) - 2(M-L)(1-r)(1+r^{M-L})r^L}{(1-r^{M-L})^2[1-r^L - L(1-r)r^L]^2} & \text{si } L \geq M/2. \end{cases} \tag{2.13}$$

Nous avons également montré que les algorithmes MP et DDA sont équivalents. Leur variance d'estimation s'exprime par :

$$\text{var}\{\tilde{\omega}\}^{MP} = \text{var}\{\tilde{\alpha}\}^{MP} = \frac{\sigma^2(1-r)^3}{2\lambda^2 r} \cdot \begin{cases} \dfrac{1+r^{M-L}}{(1-r^{M-L})^2(1-r^L)} & \text{si } L \leq M/2, \\ \dfrac{1+r^L}{(1-r^{M-L})(1-r^L)^2} & \text{si } L \geq M/2. \end{cases} \tag{2.14}$$

Comparaion des algorithmes

Les résultats dans (2.13) et (2.14) permettent d'établir la supériorité des algorithmes MP et DDA en comparaison avec BLP.

Proposition 2.1 ([Djermoune09b]). *Pour tout $r \in]0,1]$ et $1 \leq L \leq M-1$, nous avons :*

$$\text{var}\{\tilde{\omega}\}^{BLP} \geq \text{var}\{\tilde{\omega}\}^{MP} = \text{var}\{\tilde{\omega}\}^{DDA}. \tag{2.15}$$

□

[2]Cette hypothèse permet de garantir que l'ordre des valeurs singulières dans $\tilde{\boldsymbol{S}}$ est le même avec et sans bruit. Dans notre cas, cela concerne la valeur singulière principale.

Convergence vers la borne de Cramér-Rao

La borne inférieure de Cramér-Rao (CRLB) sur la pulsation et le facteur d'amortissement dans le cas mono-composante est donnée par [Yao95] :

$$CRLB(\alpha) = CRLB(\omega) = \frac{\sigma^2(1-r)^3}{2\lambda^2 r} \cdot \frac{1-r^M}{(1-r^M)^2 - M^2 r^{M-1}(1-r)^2}. \tag{2.16}$$

La variance des estimateurs BLP, MP et DDA converge (quand $M \to \infty$) linéairement vers la borne de Cramér-Rao.

Proposition 2.2 ([Djermoune09b]). *Soit ε_M l'écart de l'une des variances dans (2.13) ou (2.14) de la borne de Cramér-Rao : $\varepsilon_M = \mathrm{var}\{\tilde{\omega}\} - CRLB(\omega)$. Alors, $\forall r \in]0,1[$ et $L = \mu M$, $\mu \in]0,1[$, ε_M converge linéairement vers 0 avec le taux suivant :*

$$\lim_{M\to\infty} \frac{\varepsilon_{M+1}}{\varepsilon_M} = r^{\min(\mu, 1-\mu)} \tag{2.17}$$

□

Hyperparamètre optimal

Dans toutes les méthodes haute résolution, l'hyperparamètre L joue un rôle crucial sur les performances des estimées. Il est donc important de le fixer de manière optimale. L'un des critères de choix est celui qui permet d'atteindre la variance minimale. Comme la variance théorique correspondant aux algorithmes MP et DDA dans l'équation (2.14) est une fonction rationnelle en r^L, il est possible de calculer analytiquement la valeur de L qui minimise la variance. Le résultat obtenu est donné par :

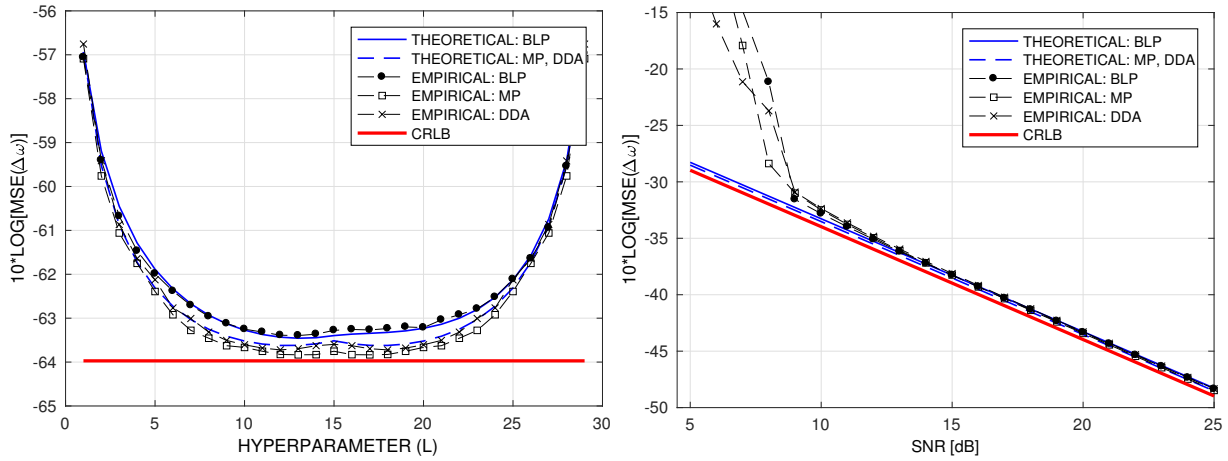
$$L_{opt} = \frac{M}{2} \pm \frac{1}{\ln r} \ln\left(\tan\frac{\pi - \tan^{-1} r^{-M/2}}{3}\right). \tag{2.18}$$

Cette valeur dépend du nombre d'échantillons M mais aussi du facteur d'amortissement α. Dans le cas où $\alpha = 0$ (sinusoïde pure), on retrouve le résultat déjà connu dans la littérature : $L_{opt} \in \{M/3, 2M/3\}$. Si $\alpha < 0$ et M suffisamment grand pour que $r^M \ll 1$, alors la valeur optimale tend vers $L_{opt} = M/2$. Ceci indique que L_{opt} dépend de α mais que sa valeur est dans l'intervalle $[M/3, 2M/3]$.

Exemple de simulation

Afin de montrer expérimentalement la validité des résultats théoriques obtenus, on considère un signal mono-composante de $M = 30$ échantillons et de paramètre $\alpha = -0.1$. L'amplitude c est fixée de telle sorte que $10\log(|c|^2/\sigma^2) = 40$ dB. La figure 2.1(a) montre l'erreur quadratique moyenne (EQM) sur l'estimation de la pulsation ω. On observe que les résultats théoriques sont très proches des EQM empiriques. Par ailleurs, il apparaît nettement qu'il existe une valeur de L qui permet d'obtenir la variance minimale. Pour MP et DDA, les valeurs obtenues par l'expression (2.18) sont $\{12, 18\}$. Ces valeurs sont assez proches des minima expérimentaux de l'EQM.

Les algorithmes étudiés ici, comme d'ailleurs tous les algorithmes à base de SVD, ont des performances qui décrochent à partir d'un certain seuil de RSB. La figure 2.1(b) montre que pour notre



(a) Variances en fonction de L

(b) Variances en fonction du RSB

FIGURE 2.1 : Variances empiriques et théoriques de la pulsation avec les méthodes BLP, MP et DDA

exemple, ce seuil est autour de 8 dB. On constate également que ce seuil correspond aussi au seuil de validité des résultats théoriques.

2.3 Perturbation au premier ordre de l'amplitude complexe

Dans ce paragraphe, on s'intéresse à l'analyse de l'erreur sur l'amplitude c, toujours dans le cas mono-composante. Une fois le mode a estimé, l'amplitude peut être calculée en résolvant le système linéaire suivant :

$$\tilde{\boldsymbol{a}}_K \cdot \tilde{c} \approx \tilde{\boldsymbol{y}}_K \tag{2.19}$$

où $\tilde{\boldsymbol{a}}_K = [1, \tilde{a}, \ldots, \tilde{a}^{K-1}]^{\mathsf{T}}$, $\tilde{\boldsymbol{y}}_K = [\tilde{y}(1), \ldots, \tilde{x}(K)]^{\mathsf{T}}$ et $K \leq M$ est un hyperparamètre qui impose le nombre d'équations mises en œuvre pour estimer l'amplitude[3]. La solution au sens des moindres carrés de cette équation s'écrit :

$$\tilde{c} = (\tilde{\boldsymbol{a}}_K^{\mathsf{H}} \tilde{\boldsymbol{a}}_K)^{-1} \tilde{\boldsymbol{a}}_K^{\mathsf{H}} \tilde{\boldsymbol{y}}_K. \tag{2.20}$$

La prise en compte du bruit dans le vecteur $\tilde{\boldsymbol{y}}_K$ et des erreurs d'estimation dans $\tilde{\boldsymbol{a}}_K$ permet d'obtenir la perturbation à l'ordre 1 sur l'amplitude, qui est donnée par :

$$\Delta c = \frac{1}{\zeta_K} \boldsymbol{a}_K^{\mathsf{H}} (\boldsymbol{e}_K - a \Delta \boldsymbol{a}_K) \tag{2.21}$$

où $\Delta a = \tilde{a} - a$, $\Delta c = \tilde{c} - c$, $\zeta_K = \sum_{m=0}^{K-1} r^m$, $\boldsymbol{e}_K = [e(1), \cdots, e(K)]^{\mathsf{T}}$ et $\Delta \boldsymbol{a}_K = \Delta a [0, 1, 2a, \cdots, (K-1)a^{K-2}]^T$. À partir de cette expression, il est clair que pour toute estimation non-biaisée $\tilde{a}$, l'estimation de $\tilde{c}$ au sens des moindres carrés l'est également. De plus, on peut facilement montrer que $E\{(\Delta c)^2\} = 0$. Dans le cas où le mode est obtenu par les algorithmes MP ou DDA, la variance de

[3] Il s'agit également ici de trouver la valeur optimale de cet hyperparamètre

l'amplitude complexe $E\{|\Delta c|^2\}$ est donnée par l'expression suivante [Djermoune09a] :

$$E\{|\Delta c|^2\} = \frac{1-r}{1-r^K}\sigma^2 + \frac{\lambda^2}{r}\left(\frac{r}{1-r} - \frac{Kr^K}{1-r^K}\right)^2 E\{|\Delta a|^2\} + \frac{2\sigma^2}{r\zeta_L\zeta_{M-L}}\left(\frac{r}{1-r} - \frac{Kr^K}{1-r^K}\right)\frac{(1-r)r^K m_K}{1-r^K}. \tag{2.22}$$

où $m_K = \min(K, M-K, L, M-L)$. Les variances de l'amplitude et de la phase peuvent être retrouvées en utilisant la relation : $\text{var}\{\lambda\} = \tilde{\lambda}^2\text{var}\{\tilde{\phi}\} = \frac{1}{2}E\{|\Delta c|^2\}$.

Le cas d'une sinusoïde pure

Afin de comparer l'expression dans (2.22) avec celles de la littérature, considérons le cas où $\alpha = 0$ qui donne :

$$E\{|\Delta c|^2\} = \frac{\sigma^2}{K} + \frac{1}{4}\lambda^2(K-1)^2 E\{|\Delta a|^2\} + \frac{\sigma^2(K-1)\min(K, M-K, L, M-L)}{L(M-L)K}. \tag{2.23}$$

Les deux premiers termes de l'équation correspondent bien a ceux déjà rapportés dans la littérature [Ducasse95, Ying99]. Le dernier terme, lié à l'intercorrélation entre $\boldsymbol{e}_K$ et Δa, est négligé dans les références précédentes. En réalité, ce terme ne peut pas toujours être négligé : cela dépend des valeurs de L et K. Par exemple, si l'on choisit $L = L_{opt} = M/3$, alors la valeur de K permettant d'obtenir la variance minimale sur l'amplitude est $K_{opt} \approx 0.86M$. Comme nous le verrons dans une simulation, ce résultat est plus précis que celui présenté dans [Ying99], c'est-à-dire $K_{opt} = 0.66M$. Pour terminer, les deux minima de $E\{|\Delta c|^2\}$ pour $L = M/3$ sont atteints pour $K \in \{0.53M, 0.86M\}$.

Le cas d'une sinusoïde amortie

Dans le cas d'une sinusoïde amortie, la valeur optimale de K doit être calculée à partir de l'expression de la variance dans l'équation (2.22). Ceci n'est pas facile à cause de sa dépendance non linéaire par rapport à K. Pour des raisons pratiques, nous en donnons ici une approximation sous les hypothèses suivantes :

i) $L \in [M/3, 2M/3]$ et $K \geqslant L$,

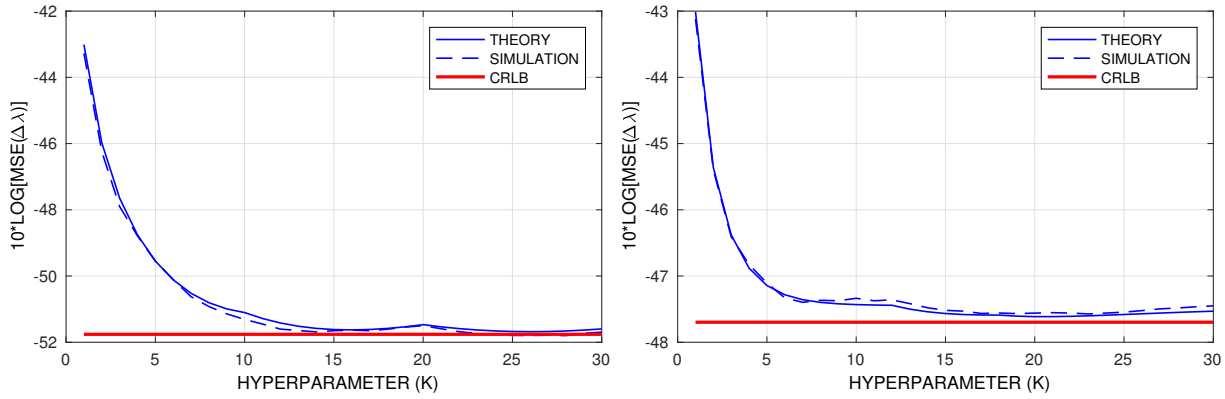
ii) $r^L \ll 1$.

L'hypothèse $L \in [M/3, 2M/3]$ n'est pas très restrictive car on sait que la valeur optimale de L est située justement dans cet intervalle. La deuxième hypothèse implique que le nombre d'échantillons est suffisamment élevé pour que la sinusoïde tende vers 0 ($r^L \ll 1 \Rightarrow r^M \ll 1$). Le résultat obtenu est :

$$K_{opt} \approx \begin{cases} \min(L, M-L) + \dfrac{0.5 + r^2}{1 - r^2}, & \text{si } m_K \in \{L, M-L\} \\ 0.5M + \dfrac{1 + r^2}{4(1-r^2)}, & \text{si } m_K = M - K. \end{cases} \tag{2.24}$$

Exemple de simulation

Soit un signal sinusoïdal complexe d'amplitude $c = 1$ composé de $M = 30$ échantillons. On considère deux cas pour le facteur d'amortissement : $\alpha = 0$ et $\alpha = -0.1$. La variance du bruit additif est choisie de telle sorte que $10\log(|c|^2/\sigma^2) = 40$ dB. Le mode est estimé avec la méthode MP. En fixant L à sa valeur optimale ($L_{opt} = 10$ pour $\alpha = 0$ et $L_{opt} = 12$ pour $\alpha = -0.1$), les variances théoriques et expérimentales sont représentées sur la figure 2.2. On constate que les courbes théoriques sont très proches des courbes expérimentales. Dans le cas $\alpha = 0$, la variance minimale est atteinte pour $K \approx 16$ et $K \approx 26$, ce qui correspond bien aux valeurs prédites précédemment (i.e. $K_{opt} \in \{0.53M, 0.86M\}$). Dans le cas amorti, le minimum est atteint pour $K \approx 20$ assez proche de la valeur donnée par l'équation (2.24). La position du second minimum local ($K \approx 10$) ne satisfait pas l'hypothèse (i). De manière général, les résultats théoriques et expérimentaux montrent que l'estimation de l'amplitude dans le cas amorti nécessite moins d'équations que dans le cas harmonique.



(a) $\alpha = 0$ (b) $\alpha = -0.1$

FIGURE 2.2 : Variances empiriques et théoriques de l'amplitude λ en fonction de K

La figure 2.3 montre les variances théorique et simulée en fonction du RSB pour $\alpha = -0.1$. Les paramètres L et K sont fixés à leurs valeurs optimales. L'expression théorique est valide pour cet exemple à partir de 8 dB. Bien sûr, ce résultat ne peut pas être généralisé car cela dépend aussi de la valeur de α.

2.4 Conclusion

Nous nous sommes intéressés dans ce chapitre à l'analyse de perturbation au premier ordre de méthodes d'estimations modales. Les expressions analytiques obtenues permettent de calculer les hyperparamètres qui conduisent à la minimisation de la variance d'estimation des paramètres d'intérêt. Par rapport aux études déjà présentées dans la littérature, ma contribution théorique se situe dans la prise en compte du facteur d'amortissement. Nous avons montré que pour l'estimation du mode (fréquence et facteur amortissement), le paramètre optimal (L) est situé dans l'intervalle $[M/3, 2M/3]$ où M est le nombre d'échantillons. Dans le cas de l'amplitude complexe, le paramètre K doit être choisi dans l'intervalle $[0.53M, 0.86M]$. En pratique, l'extension de ces résultats au cas d'un signal multi-composantes est possible si les modes sont suffisamment séparés et que les facteurs

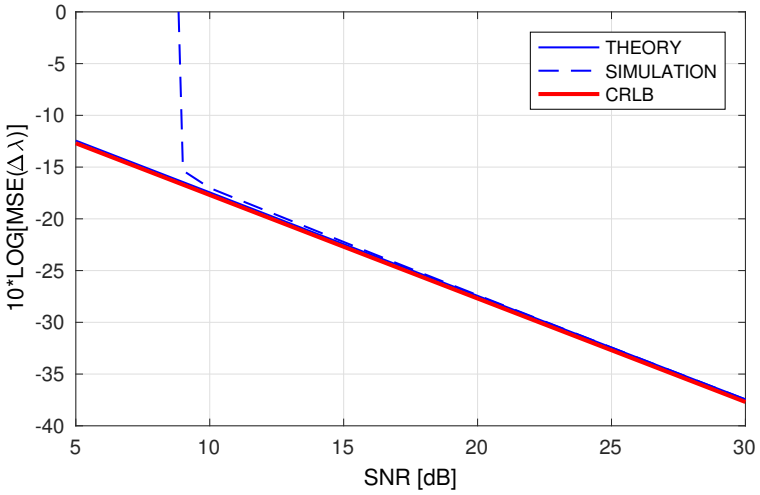


FIGURE 2.3 : Variances empiriques et théoriques de l'amplitude λ en fonction du RSB

d'amortissement sont de même ordre de grandeur. La généralisation de ces contributions au cas d'un signal multi-composantes et multidimensionnel a été récemment proposée dans [Sahnoun17b] pour la méthode ESPRIT.

Chapitre 3

Estimation modale multidimensionnelle

Un signal modal multidimensionel (R-D) est modélisé par la superposition de F exponentielles complexes :

$$\tilde{y}(m_1,\ldots,m_R) = y(m_1,\ldots,m_R) + e(m_1,\ldots,m_R) \triangleq \sum_{f=1}^{F} c_f \prod_{r=1}^{R} a_{f,r}^{m_r-1} + e(m_1,\ldots,m_R) \qquad (3.1)$$

où $m_r = 1,\ldots,M_r$ pour $r = 1,\ldots,R$. M_r désigne la taille du support temporel dans la r^{e} dimension, $a_{f,r} = \exp(\alpha_{f,r} + j\omega_{f,r}) \in \mathbb{C}$ est le f^{e} mode dans la r^{e} dimension, $\{\alpha_{f,r}\}_{f=1,r=1}^{F,R}$, $\alpha_{f,r} \in \mathbb{R}^-$, sont les facteurs d'amortissement, $\{\omega_{f,r} = 2\pi\nu_{f,r}\}_{f=1,r=1}^{F,R}$ sont les pulsations, et $c_f = \lambda_f \exp(j\phi_f)$ est l'amplitude complexe du f^{e} mode. $\lambda_f = |c_f|$ et $\phi_f = \arg(c)$ sont le module et la phase de c_f. Le bruit complexe $e(m_1, m_2, \ldots, m_R)$ est supposé gaussien, de moyenne nulle, de variance σ^2 et mutuellement indépendant dans toutes les dimensions. L'objectif est d'estimer les $(R+1)$-uplets $\{a_{f,1},\ldots,a_{f,R},c_f\}_{f=1}^{F}$ en utilisant les $M = \prod_{r=1}^{R} M_r$ échantillons de $\tilde{y}$. Un exemple d'un signal modal 2-D est représenté en figure 3.1.

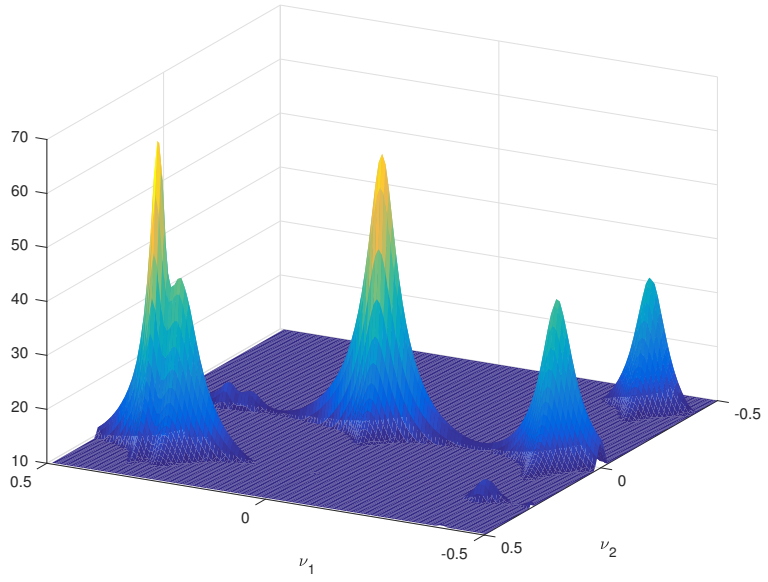


FIGURE 3.1 : Spectre d'un signal composé de $F = 5$ modes bidimensionnels

Comme pour le cas 1-D discuté au chapitre 2, ce problème est également ancien en traitement du signal. Il a reçu un regain d'intérêt ces dernières années grâce à l'émergence de nouvelles applications dont la spectroscopie de résonance magnétique nucléaire multidimensionnelle [Li98], les canaux de communication sans fil [Gershman05], l'imagerie radar MIMO [Nion10], etc. Plusieurs méthodes paramétriques sont proposées pour résoudre ce problème pour $R \geq 2$. On retrouve les méthodes basées sur la prédiction linéaire étendue au problème 2-D comme TLS-Prony 2-D [Sacchini93], les méthodes de sous-espaces telles que *matrix enhancement and matrix pencil* (MEMP) [Hua92] et ESPRIT 2-D [Rouquette01]. On peut également citer les méthodes R-D MUSIC [Li98], *multidimensional folding* (MDF) [Mokios04], *improved multidimensional folding* (IMDF) [Liu06, Liu07], Tensor-ESPRIT [Haardt08], *principal-singular-vector utilization for modal analysis* (PUMA) [Sun12, So10] ainsi que celles proposées dans [Huang12, Lin13].

Les méthodes d'estimation modale multidimensionnelles sont confrontées à deux problèmes principaux : (1) l'alignement de modes dans certaines dimensions (i.e. $\exists r \in \{1, \ldots, R\}$ et $f \neq f'$ tels que $a_{f,r} = a_{f',r}$) ; (2) la reformation (ou le couplage) des composantes R-D. L'alignement de modes engendre une déficience de rang dans les algorithmes de sous-espaces basés sur le sous-espace signal [Haardt08]. Une solution à ce problème consiste à utiliser une procédure de filtrage spatial (*spatial smoothing*) [Shan85, Hua92, Liu06, Haardt08]. Le couplage des modes (*pairing* en anglais) est une étape nécessaire dans tous les algorithmes qui réalisent des estimations séparées sur chacune des dimensions. Ce couplage peut être réalisé soit par la minimisation d'un critère après l'estimation des modes [Sacchini93, Hua92, Pesavento04] ou par diagonalisation conjointe de R matrices [Haardt08].

Les méthodes prédédentes ont des performances différentes mais il est généralement admis qu'elles donnent des résultats précis (proches des bornes de Cramér-Rao) quand le bruit est faible et que les composantes du signal ne sont pas trop proches. Malheureusement, la plupart de ces approches nécessite la construction de grandes matrices qu'il faut décomposer en valeurs propres ou en valeurs singulières. Plus récemment, des approches basées sur le concept de parcimonie ont été proposées dans le cadre de l'estimation fréquentielle [Goodwin99, Malioutov05, Stoica11, Sward14, Tang13, Sahnoun17a]. Le dictionnaire est composé d'un ensemble de fonctions exponentielles complexes potentiellement présentes dans le signal analysé, ce qui permet d'inclure très facilement une information *a priori* sur la position de certains modes.

Ce chapitre est consacré à la présentation des travaux développés dans le cadre de la thèse de Souleymen Sahnoun[1] et poursuivis jusqu'en 2016. Nos contributions portent principalement sur les point suivants :

- estimation fréquentielle basée sur la parcimonie et la mise à jour adaptative (multigrille) du dictionnaire ;
- développement d'un algorithme d'estimation modale multidimensionnelle ;
- calcul des bornes de Cramér-Rao des paramètres du modèle R-D.

[1] Thèse de doctorat, 2009-2012.

3.1 Estimation modale et approximation parcimonieuse

3.1.1 Principe

Le problème inverse d'estimation modale multidimensionnelle peut être formulé comme un problème d'approximation parcimonieuse. Considérons d'abord le cas d'un signal 1-D sans bruit[2] :

$$y(m) \triangleq \sum_{f=1}^{F} c_f a_f^{m-1} \quad m = 1, \ldots, M, \tag{3.2}$$

qui s'écrit sous forme matricielle :

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{c}, \tag{3.3}$$

où $\boldsymbol{y} = [y(1), \ldots, y(M)]^{\mathsf{T}}, \boldsymbol{c} = [c_1, \ldots, c_F]^{\mathsf{T}}$, et $\boldsymbol{A} \in \mathbb{C}^{M \times F}$ est une matrice de Vandermonde :

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_F \\ \vdots & \vdots & & \vdots \\ a_1^{M-1} & a_2^{M-1} & \cdots & a_F^{M-1} \end{bmatrix}. \tag{3.4}$$

On définit une grille uniforme bidimensionnelle avec un pas δ_α et δ_ω couvrant les intervalles $[\alpha_{\min}, 0]$ et $[0, 2\pi[$, respectivement, où $\alpha_{\min}$ est une borne inférieure sur $\{\alpha_f\}_{f=1}^F$. On pose $K_\alpha = \lfloor -\alpha_{\min}/\delta_\alpha \rfloor$, $K_\omega = \lfloor 2\pi/\delta_\omega \rfloor$, $N = K_\alpha K_\omega \gg F$, et $\boldsymbol{\phi}(\beta, \mu) = [1, e^{\beta + j\mu}, \ldots, e^{(\beta+j\mu)(M-1)}]^{\mathsf{T}}$. On définit le dictionnaire :

$$\boldsymbol{\Phi} = [\boldsymbol{\phi}(0,0), \cdots, \boldsymbol{\phi}(0, (K_\omega - 1)\delta_\omega), \cdots, \boldsymbol{\phi}((K_\alpha - 1)\delta_\alpha, (K_\omega - 1)\delta_\omega)] \in \mathbb{C}^{M \times N} \tag{3.5}$$

Le modèle du signal modal peut alors s'écrire approximativement :

$$\boldsymbol{y} \approx \boldsymbol{\Phi}\boldsymbol{x}$$

où $\boldsymbol{x} \in \mathbb{C}^N$ est un vecteur de coefficients qui peut être obtenu à partir des données en résolvant le problème suivant :

$$\hat{\boldsymbol{x}} = \arg\min_x ||\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}||_2^2 \quad \text{tel que} \quad ||\boldsymbol{x}||_0 \leq F. \tag{3.6}$$

où $||\boldsymbol{x}||_0$ est la pseudo-norme zéro d'un vecteur (cardinalité). Dans le cas idéal (pas de bruit, le dictionnaire contient les modes du signal), la solution à ce problème donnera les modes et les amplitudes complexes qui correspondront aux éléments non nuls dans $\hat{\boldsymbol{x}}$.

Remarque 3.1. *Il est important de noter ici que les méthodes d'approximation parcimonieuse basées sur des dictionnaires continus (dits* gridless *ou* off the grid*) est une tendance de fond dans littérature [Tang13, Candès14, Chen14, Xu14, Yang16b, Yang16a]. Quoique les algorithmes proposés ont de plus fortes garanties théoriques (dans le cas harmonique), elles nécessitent néanmoins beaucoup plus d'effort de calcul par rapport aux méthodes basées sur un dictionnaire discret.*

[2] Pour simplifier les notations, on omet temporairement la dépendance des paramètres de la dimension r.

Cette idée peut être généralisée au cas d'un signal R-D. On définit d'abord le vecteur $\boldsymbol{y}$:

$$\boldsymbol{y} = \begin{bmatrix} y(1,\ldots,1,1) \\ \vdots \\ y(1,\ldots,1,M_R) \\ y(1,1,\ldots,2,1) \\ \vdots \\ y(1,\ldots,2,M_R) \\ \vdots \\ y(M_1,\ldots,M_{R-1},M_R) \end{bmatrix} \tag{3.7}$$

et les matrices de Vandermonde dans chaque dimension $r = 1,\ldots,R$:

$$\boldsymbol{A}_r = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ a_{1,r} & a_{2,r} & \cdots & a_{F,r} \\ \vdots & \vdots & & \vdots \\ a_{1,r}^{M_r-1} & a_{2,r}^{M_r-1} & \cdots & a_{F,r}^{M_r-1} \end{bmatrix}. \tag{3.8}$$

On peut alors établir que :

$$\boldsymbol{y} = (\boldsymbol{A}_1 \odot \boldsymbol{A}_2 \odot \cdots \odot \boldsymbol{A}_R)\boldsymbol{c} \triangleq \boldsymbol{A}\boldsymbol{c}, \tag{3.9}$$

où $\odot$ désigne le produit de Khatri-Rao. Ainsi, le problème d'estimation modal R-D peut être formulé comme un problème d'approximation parcimonieuse dans laquelle le dictionnaire peut être écrit sous la forme [Sahnoun12b] :

$$\boldsymbol{\Phi} = \boldsymbol{\Phi}_1 \otimes \boldsymbol{\Phi}_2 \otimes \cdots \otimes \boldsymbol{\Phi}_R \tag{3.10}$$

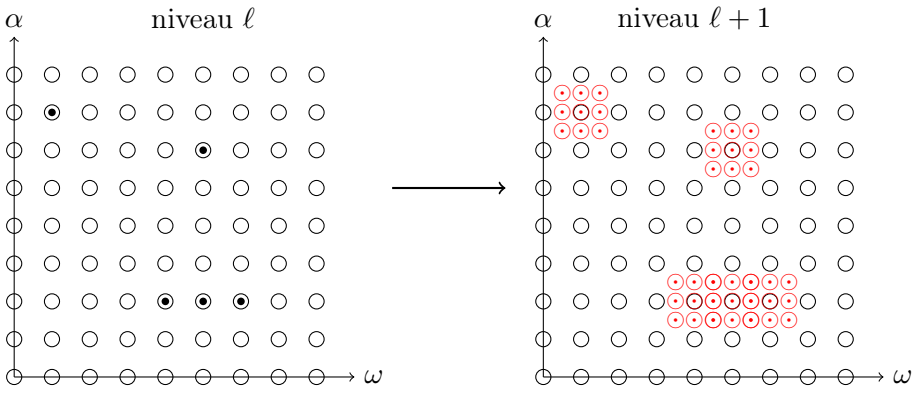
où $\otimes$ est le produit de Kronecker et $\boldsymbol{\Phi}_r$ le dictionnaire construit dans la dimension r.

3.1.2 Approche multigrille

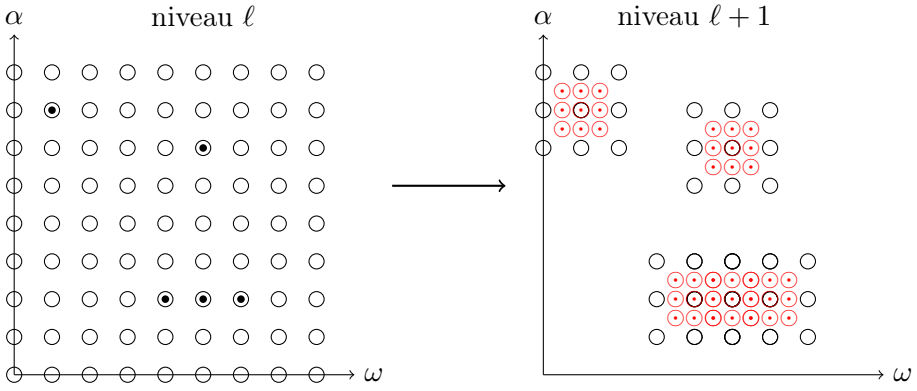
Nous avons vu que le problème de l'estimation modale peut être formulée comme un problème d'approximation parcimonieuse dans lequel nous faisons implicitement l'hypothèse que le dictionnaire inclut les vrais modes présents dans le signal. Une première approche pour garantir cela (au moins approximativement) est de définir $\boldsymbol{\Phi}$ sur une grille très fine ($\delta_\omega, \delta_\alpha \ll 1$) qui mène à un dictionnaire de grande dimension. La principale limitation d'une telle approche réside dans l'augmentation considérable du coût de calcul. Nous avons alors proposé une approche alternative qui consiste en un raffinement adaptatif du dictionnaire en partant d'un dictionnaire initialement grossier.

Soit ℓ le niveau de la grille courante ($\ell = 0,\ldots,L-1$) où $\ell = 0$ correspond à la grille la plus grossière (par exemple, celle de Fourier pour les fréquences). Au niveau ℓ, on commence par restaurer le signal $\boldsymbol{y}$ en utilisant le dictionnaire $\boldsymbol{\Phi}^{(\ell)}$. Le dictionnaire est ensuite mis à jour en insérant des atomes au voisinage de ceux activés et, éventuellement, en retirant les atomes qui ne sont pas activés (cf. figure 3.2). On obtient ainsi un nouveau dictionnaire $\boldsymbol{\Phi}^{(\ell+1)}$. Ce processus est répété jusqu'à ce que le niveau de résolution souhaité soit atteint.

La procédure qui vient d'être décrite consiste à échantillonner conjointement les fréquences et



(a) Politique d'ajout



(b) Poltique d'ajout/retrait

FIGURE 3.2 : Approche multigrille pour un dictionnaire 1-D. Les paramètres α et ω sont raffinés *conjointement*

les facteurs d'amortissement sur un grille 2-D. Il est également possible de scinder ce raffinement en deux étapes en utilisant une grille linéaire dans chaque étape [Sward14, Sahnoun17a] :

1. affiner les fréquences en supposant que les facteurs d'amortissement sont nuls ;
2. affiner ensuite les amortissements en utilisant les fréquences obtenues dans l'étape précédente.

Le schéma de principe de cette stratégie est représenté en figure 3.3. Celle-ci a l'avantage de réduire les temps de calcul et permet également d'établir des conditions de convergence pour un signal mono-composante.

3.1.3 Résultat de convergence

La stratégie multigrille utilise une règle heuristique qui consiste à ajouter dans le dictionnaire des atomes au voisinage de ceux activés par la méthode parcimonieuse choisie. De manière générale, il n'y a aucune garantie pour que l'approche converge vers les « vraies » composantes du signal. En fait, cela dépend d'une part de l'algorithme choisi pour résoudre le critère (3.6), mais aussi du niveau de bruit et de la distance entre les composantes du signal. En revanche, dans le cas d'un

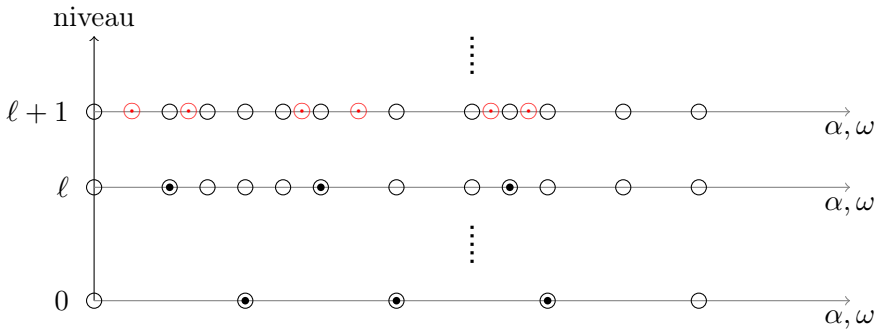


FIGURE 3.3 : Approche multigrille pour un dictionnaire 1-D. Les paramètres α et ω sont raffinés *séparément*

signal mono-composante, il est possible d'établir une condition de convergence. Nous verrons dans le paragraphe 3.2 comment exploiter cette condition pour un signal R-D multi-composantes.

Prenons la cas d'un signal sans bruit et mono-composante : $F = 1$. Sans perte de généralité, on suppose également que $R = 1$. Le modèle du signal est donné par l'équation (3.2). Soit le dictionnaire $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N]$ dont les colonnes sont normalisées ($||\boldsymbol{\phi}_n||_2 = 1, n = 1, \ldots, N$). Chaque colonne du dictionnaire est générée par le mode hypothétique $\phi_n = \exp(\beta_n + j\mu_n), \beta_n \in [\alpha_{\min}, 0]$ et $\mu_n \in [0, 2\pi[$. L'approximation parcimonieuse de $\boldsymbol{y}$ relativement à $\boldsymbol{\Phi}$ est la solution du critère :

$$\min_{\boldsymbol{x}} J(\boldsymbol{x}) = ||\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}||_2^2 \quad \text{tel que} \quad ||\boldsymbol{x}||_0 = 1. \tag{3.11}$$

La solution optimale est donnée par :

$$x_n^* = \boldsymbol{\phi}_n^{\mathsf{H}}\boldsymbol{y}, \quad \boldsymbol{x}^*_{\{1,\ldots,N\}\setminus n} = \boldsymbol{0}, \quad J(\boldsymbol{x}^*) = ||\mathbf{y}||_2^2 - \boldsymbol{y}^{\mathsf{H}}\boldsymbol{\phi}_n\boldsymbol{\phi}_n^{\mathsf{H}}\mathbf{y} \tag{3.12}$$

où n désigne la colonne activée dans $\boldsymbol{\Phi}$. Le minimum de $J(\boldsymbol{x}^*)$ est atteint pour un atome $\boldsymbol{\phi}_n$ qui maximise $J'(\boldsymbol{\phi}_n) = \boldsymbol{y}^{\mathsf{H}}\boldsymbol{\phi}_n\boldsymbol{\phi}_n^{\mathsf{H}}\boldsymbol{y} = |\boldsymbol{\phi}_n^{\mathsf{H}}\boldsymbol{y}|^2$, $n = 1, \ldots, N$.

Estimation de la fréquence : dictionnaire harmonique

Pour estimer la fréquence, on utilise un dictionnaire harmonique (i.e. $\beta_n = 0, \forall n$). On cherche alors à maximiser :

$$J'(\mu_n) = \frac{|c|^2}{M}\left|\frac{1 - e^{\alpha M + j(\omega - \mu_n)M}}{1 - e^{\alpha + j(\omega - \mu_n)}}\right|^2. \tag{3.13}$$

Théorème 3.1 ([Sahnoun17a])**.** *Soit $y(m)$ un signal modal mono-composante de longueur M et $\boldsymbol{\Phi}^{(0)} = [\boldsymbol{\phi}_1^{(0)}, \ldots, \boldsymbol{\phi}_{N_0}^{(0)}]$ le dictionnaire harmonique initial ($\ell = 0$) dans lequel les colonnes sont rangées dans l'ordre croissant des pulsations $\mu_n^{(0)}$ telles que $\mu_1^{(0)} = 0$ et $\mu_{N_0}^{(0)} = 2\pi(1 - 1/M)$. La procédure de raffinement multigrille converge $\forall\alpha$ (i.e. $\exists n \in \{1, \ldots, N_\ell\}$ tel que $\lim_{\ell\to\infty} \mu_n^{(\ell)} = \omega$) si :*

$$\max_{n\in\{1,\ldots,N_0-1\}} |\mu_{n+1}^{(0)} - \mu_n^{(0)}| < 2\zeta_M \tag{3.14}$$

où $\zeta_M > 1/2M$ est une constante. □

Corollaire 3.1. *La procédure de raffinement converge si la grille initiale est la grille de Fourier.* □

Estimation de l'amortissement : dictionnaire modal

Supposons que l'approximation parcimonieuse multigrille précédente a convergé vers $\hat{\mu} = \omega$. Pour estimer α, on construit un dictionnaire modal en utilisant une grille d'amortissements β_n et la pulsation ω : $\phi_n = \exp(\beta_n + j\omega)$. On obtient alors :

$$J'(\beta_n) = \frac{|c|^2(1-e^{2\beta_n})}{1-e^{2\beta_n M}} \left(\frac{1-e^{(\alpha+\beta_n)M}}{1-e^{(\alpha+\beta_n)}} \right)^2 . \tag{3.15}$$

Théorème 3.2 ([Sahnoun17a])**.** *Soit $y(m)$ un signal modal mono-composante de longueur M et $\boldsymbol{\Phi}^{(0)} = [\boldsymbol{\phi}_1^{(0)}, \ldots, \boldsymbol{\phi}_{N_0}^{(0)}]$ le dictionnaire modal initial formé des modes $\exp(\beta_n^{(0)} + j\omega)$, où ω est la pulsation de $\boldsymbol{y}$. Les colonnes de $\boldsymbol{\Phi}^{(0)}$ sont rangées dans l'ordre croissant des $\beta_n^{(0)}$: $\beta_1^{(0)} = \alpha_{\min}$ et $\beta_{N_0}^{(0)} = 0$. La procédure de raffinement multigrille converge (i.e. $\exists n \in \{1, \ldots, N_\ell\}$ tel que $\lim_{\ell\to\infty} \beta_n^{(\ell)} = \alpha$) si $\alpha \in [\alpha_{\min}, 0]$.* □

Corollaire 3.2. *Dans la seconde étape de la stratégie multigrille, le dictionnaire initial peut être formé en utilisant uniquement deux amortissements : $\beta_1^{(0)} = \alpha_{\min}$ et $\beta_2^{(0)} = 0$.* □

En conclusion, dans le cas d'un signal mono-composante, la stratégie qui consiste à raffiner le dictionnaire en deux étapes converge sous des conditions non restrictives. Il s'agit maintenant d'exploiter ce résultat pour l'estimation modale dans le cas multi-composantes.

3.2 Algorithme d'estimation modale R-D basé sur la parcimonie

3.2.1 Approximation parcimonieuse simultanée

En notation tensorielle, le modèle (3.1) s'écrit :

$$\widetilde{\boldsymbol{\mathcal{Y}}} = \boldsymbol{\mathcal{Y}} + \boldsymbol{\mathcal{E}} \tag{3.16}$$

où $\{\widetilde{\boldsymbol{\mathcal{Y}}}, \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{E}}\} \in \mathbb{C}^{M_1 \times M_2 \times \cdots \times M_R}$. Le tenseur $\boldsymbol{\mathcal{Y}}$ peut également s'écrire sou la forme :

$$\boldsymbol{\mathcal{Y}} = \sum_{f=1}^{F} c_f \; \boldsymbol{a}_{f,1} \circ \boldsymbol{a}_{f,2} \circ \cdots \circ \boldsymbol{a}_{f,R} \tag{3.17}$$

où $\boldsymbol{a}_{f,r} = [1, a_{f,r}, \ldots, a_{f,r}^{M_r-1}]^{\mathsf{T}}$, $r = 1, \ldots, R$ et $\circ$ désigne le produit extérieur. Cette équation est appelée décomposition canonique polyadique ou décomposition Candecomp/Parafac du tenseur $\boldsymbol{\mathcal{Y}}$ [Comon14, Kolda09]. La matrice $\boldsymbol{Y}_{(r)} \in \mathbb{C}^{M_r \times M_r'}$ ($M_r' = M/M_r$) obtenue par le dépliage de $\boldsymbol{\mathcal{Y}}$ selon la r^{e} dimension s'écrit :

$$\boldsymbol{Y}_{(r)} = \boldsymbol{A}_r \underbrace{\boldsymbol{\Delta}_{\boldsymbol{c}}(\boldsymbol{A}_R \odot \cdots \odot \boldsymbol{A}_{r+1} \odot \boldsymbol{A}_{r-1} \odot \cdots \odot \boldsymbol{A}_1)^{\mathsf{T}}}_{\boldsymbol{H}_r} \tag{3.18}$$

avec $\boldsymbol{\Delta}_{\boldsymbol{c}} = \text{diag}(\boldsymbol{c})$. Finalement, on obtient :

$$\widetilde{\boldsymbol{Y}}_{(r)} = \boldsymbol{A}_r \boldsymbol{H}_r + \boldsymbol{E}_{(r)} \tag{3.19}$$

où $\boldsymbol{H}_r \in \mathbb{C}^{F \times M'_r}$. Ainsi, chaque colonne de la matrice $\boldsymbol{Y}_{(r)}$ est une combinaison linéaire des modes de la r^{e} dimension. En effet, cette matrice a la forme suivante :

$$\begin{aligned} \boldsymbol{Y}_{(r)} &\triangleq [\boldsymbol{y}_{(r),1}, \ldots, \boldsymbol{y}_{(r),M'_r}] \\ &= \left[\sum_{f=1}^{F} h_r(f,1)\boldsymbol{a}_{f,r}, \ldots, \sum_{f=1}^{F} h_r(f,M'_r)\boldsymbol{a}_{f,r} \right] \end{aligned} \tag{3.20}$$

où $h_r(f, m'_r)$ est l'élément (f, m'_r) de la matrice $\boldsymbol{H}_r$ pour $f = 1, \ldots, F$ et $m'_r = 1, \ldots, M'_r$. Par conséquent, si $M_r > F$ alors il est possible d'identifier les coordonnées des modes $\{a_{f,r}\}_{f=1}^{F_r}$ ($F_r < F$) dans la dimension r à partir de n'importe quelle colonne de $\boldsymbol{Y}_{(r)}$. Dans notre cas, il faut remplacer la matrice $\boldsymbol{Y}_{(r)}$ par sa version perturbée $\widetilde{\boldsymbol{Y}}_{(r)}$. Étant donné un dictionnaire $\boldsymbol{\Phi}_r \in \mathbb{C}^{M_r \times N}$, l'approximation parcimonieuse d'un vecteur $\tilde{\boldsymbol{y}}_{(r),m'_r}$ de cette matrice correspond au problème d'optimisation suivant :

$$\boldsymbol{x}_{m'_r} = \arg\min_{\boldsymbol{x}} ||\boldsymbol{x}||_0 \quad \text{s.c.} \quad ||\tilde{\boldsymbol{y}}_{(r),m'_r} - \boldsymbol{\Phi}_r \boldsymbol{x}||_2^2 \leq \epsilon. \tag{3.21}$$

Comme chaque vecteur $\tilde{\boldsymbol{y}}_{(r),m'_r}$, $m'_r = 1, \ldots, M'_r$ peut être vu comme un signal 1-D composé des mêmes modes $\{a_{f,r}\}_{f=1}^{F_r}$, la position des éléments non nuls est la même dans tous les vecteurs $\boldsymbol{x}_{m'_r}$. Finalement, si l'on veut exploiter toutes les données dans la matrice $\widetilde{\boldsymbol{Y}}_{(r)}$, il faut résoudre un problème d'approximation parcimonieuse simultanée :

$$\boldsymbol{X}_r = \arg\min_{\boldsymbol{X}} ||\boldsymbol{X}||_{0,2} \quad \text{s.c.} \quad \|\widetilde{\boldsymbol{Y}}_{(r)} - \boldsymbol{\Phi}_r \boldsymbol{X}\|_F^2 \leq \epsilon \tag{3.22}$$

où $||\boldsymbol{X}||_{0,2}$ est la norme mixte ℓ_0/ℓ_2 de $\boldsymbol{X}$ (nombre de lignes ayant une norme ℓ_2 non nulle). Pour des signaux de grandes dimensions, ce problème peut être résolu par des algorithmes gloutons comme S-OMP [Tropp06].

Une méthode simple pour obtenir les R-uplets $\{(a_{f,1}, \ldots, a_{f,R})\}_{f=1}^{F}$ consiste à estimer les modes $a_{f,r}$ dans les R dimensions puis à les associer pour former les modes multidimensionnels. Pour que cette méthode donne des résultats satisfaisants, il est nécessaire de vérifier deux conditions :

1. le dictionnaire $\boldsymbol{\Phi}_r$ contient les modes recherchés ;

2. l'algorithme d'approximation parcimonieuse possède des garanties de reconstruction parfaite (si les données ne sont pas bruitées).

En général, même si la première condition est vérifiée, aucun algorithme glouton ne peut garantir la deuxième condition pour un signal multi-composantes. L'approche qui va être présentée dans le prochain paragraphe revient à décomposer le tenseur de données d'ordre R en F tenseurs d'ordre $R-1$, chacun contenant une seule composante. Cette décomposition permettra alors de garantir la convergence de la stratégie multigrille et de satisfaire la condition de reconstruction exacte.

3.2.2 Séparation des F modes

On suppose ici qu'il existe une dimension r où tous les modes sont distincts (non alignées) et que dans cette dimension l'on a $M_r > F$. On suppose, pour simplifier, que cette dimension est $r = 1$. D'après l'équation (3.17), le tenseur $\boldsymbol{\mathcal{Y}}$ peut s'écrire :

$$\begin{aligned}\boldsymbol{\mathcal{Y}} &= \boldsymbol{\mathcal{I}} \underset{1}{\bullet} \boldsymbol{A}_1 \underset{2}{\bullet} \cdots \underset{R}{\bullet} \boldsymbol{A}_R \underset{R+1}{\bullet} \boldsymbol{c}^{\mathsf{T}} \\ &= \underbrace{\left(\boldsymbol{\mathcal{I}} \underset{2}{\bullet} \boldsymbol{A}_2 \underset{3}{\bullet} \cdots \underset{R}{\bullet} \boldsymbol{A}_R \underset{R+1}{\bullet} \boldsymbol{c}^{\mathsf{T}}\right)}_{\boldsymbol{\mathcal{B}}} \underset{1}{\bullet} \boldsymbol{A}_1 \\ &= \boldsymbol{\mathcal{B}} \underset{1}{\bullet} \boldsymbol{A}_1 \end{aligned} \tag{3.23}$$

où $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{F \times \cdots \times F}$ est un tenseur identité d'ordre $R+1$ et $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{F \times M_2 \times \cdots \times M_R}$. On peut montrer que la matrice $\boldsymbol{B}_{(1)}$ correspondant au dépliage du tenseur $\boldsymbol{\mathcal{B}}$ selon la 1re dimension a la forme suivante :

$$\boldsymbol{B}_{(1)} = (\boldsymbol{c}^{\mathsf{T}} \odot \boldsymbol{A}_R \odot \cdots \odot \boldsymbol{A}_2)^{\mathsf{T}} = \begin{bmatrix} c_1(\boldsymbol{a}_{1,R} \odot \cdots \odot \boldsymbol{a}_{1,2})^{\mathsf{T}} \\ c_2(\boldsymbol{a}_{2,R} \odot \cdots \odot \boldsymbol{a}_{2,2})^{\mathsf{T}} \\ \vdots \\ c_F(\boldsymbol{a}_{F,R} \odot \cdots \odot \boldsymbol{a}_{F,2})^{\mathsf{T}} \end{bmatrix}. \tag{3.24}$$

On constate que chaque ligne $f = 1, \ldots, F$ de cette matrice correspond à un signal 1-D qui contient exactement une composante $(R-1)$-D. Par ailleurs, chacune de ces lignes peut être obtenue en dépliant les tenseurs $\boldsymbol{\mathcal{B}}_f \in \mathbb{C}^{1 \times M_2 \times \cdots \times M_R}$ selon la première dimension, où :

$$\boldsymbol{\mathcal{B}}_f = c_f \ \boldsymbol{a}_{f,2} \circ \boldsymbol{a}_{f,3} \circ \cdots \circ \boldsymbol{a}_{f,R}. \tag{3.25}$$

Le tenseur $\boldsymbol{\mathcal{B}}$ peut alors être vu comme le résultat de la concaténation des sous-tenseurs $\boldsymbol{\mathcal{B}}_f$ selon la 1re dimension :

$$\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{B}}_1 \underset{1}{\sqcup} \boldsymbol{\mathcal{B}}_2 \underset{1}{\sqcup} \cdots \underset{1}{\sqcup} \boldsymbol{\mathcal{B}}_F. \tag{3.26}$$

C'est l'ensemble de ces propriétés qui permet de transformer un problème d'estimation modal multi-composantes en F problèmes d'estimation mono-composante. Pour ce faire, il faut d'abord estimer les coordonnées des modes dans la première dimension, c'est-à-dire $\boldsymbol{A}_1$.

Estimation des modes de la 1re dimension

Pour obtenir les modes de la 1re dimension, on utilise la matrice $\widetilde{\boldsymbol{Y}}_{(1)}$ dont la SVD est donnée par :

$$\widetilde{\boldsymbol{Y}}_{(1)} = \boldsymbol{U}_F \boldsymbol{S}_F \boldsymbol{V}_F^{\mathsf{H}} + \boldsymbol{U}_b \boldsymbol{S}_b \boldsymbol{V}_b^{\mathsf{H}}. \tag{3.27}$$

où $\boldsymbol{U} = [\boldsymbol{U}_F, \boldsymbol{U}_b] \in \mathbb{C}^{M_1 \times L}$ et $\boldsymbol{V} = [\boldsymbol{V}_F, \boldsymbol{U}_b] \in \mathbb{C}^{M_1' \times L}$ sont des matrices orthonormales contenant les vecteurs singuliers à gauche et à droite de $\widetilde{\boldsymbol{Y}}_{(1)}$, avec $L = \min\{M_1, M_1'\}$. Les matrices $\boldsymbol{S}_F$ et $\boldsymbol{S}_b$ sont des matrices diagonales contenant les valeurs singulières associées aux sous-espaces signal et bruit, respectivement. On suppose que ces valeurs singulières sont rangées dans un ordre décroissant. Soit

$\widehat{\boldsymbol{Y}}_{(1)}$ la matrice obtenue à partir des valeurs et vecteurs singuliers principaux :

$$\widehat{\boldsymbol{Y}}_{(1)} = \boldsymbol{U}_F \boldsymbol{S}_F \boldsymbol{V}_F^{\mathsf{H}}. \tag{3.28}$$

On peut montrer en utilisant l'équation (3.19) que $\boldsymbol{A}_1$ et $\boldsymbol{U}_F$ engendrent le même sous-espace signal si le niveau de bruit est faible. Par conséquent, en notant $\boldsymbol{T}$ la matrice des vecteurs propres de $\underline{\boldsymbol{U}}_F^{\dagger}\overline{\boldsymbol{U}}_F$, alors $\boldsymbol{A}_1$ peut être estimée par :

$$\widehat{\boldsymbol{A}}_1 = \boldsymbol{U}_F \boldsymbol{T}. \tag{3.29}$$

où $\underline{\boldsymbol{U}}_F$ et $\overline{\boldsymbol{U}}_F$ sont obtenues à partir de $\boldsymbol{U}_F$ en éliminant la première et la dernière ligne, respectivement. On note ici que la matrice $\underline{\boldsymbol{U}}_F$ est de rang colonne plein (égal à F) car les modes dans la première dimension sont supposés distincts et que $M_1 > F$.

Estimation des modes des autres dimensions

En utilisant l'équation (3.23), on peut estimer le tenseur $\boldsymbol{\mathcal{B}}$:

$$\widehat{\boldsymbol{\mathcal{B}}} = \widetilde{\boldsymbol{\mathcal{Y}}} \underset{1}{\bullet} \widehat{\boldsymbol{A}}_1^{\dagger}, \tag{3.30}$$

ce qui permet d'obtenir également les sous-tenseurs $\widehat{\boldsymbol{\mathcal{B}}}_f$. Les modes des autres dimensions, c'est-à-dire $\{\boldsymbol{a}_{f,r}\}_{f=1,r=2}^{F,R}$, peuvent ensuite être estimés par une méthode parcimonieuse avec raffinement de la grille en utilisant les tenseurs :

$$\bar{\boldsymbol{\mathcal{Y}}}_f = \widehat{\boldsymbol{\mathcal{B}}}_f \underset{1}{\bullet} \hat{\boldsymbol{a}}_{f,1}. \tag{3.31}$$

3.2.3 Complexité de l'algorithme

Le tableau 3.1 donne la complexité numérique de la méthode proposée (séparation des modes et estimation des paramètres par S-OMP multigrille) en comparaison à trois autres algorithmes pour l'estimation modale multidimensionnelle. On rappelle les notations utilisées :

$M =$ nombre d'échantillons dans le tenseur de données ;

$F =$ nombre de modes ;

$R =$ nombre de dimensions ;

$N =$ nombre d'atomes dans le dictionnaire (même nombre pour les fréquences et les amortissements) ;

$L =$ nombre de mises à jour du dictionnaire ;

$k_t =$ constante liée à la méthode de calcul de la SVD ;

$P =$ paramètre utilisateur (typiquement $P \approx M/3$).

En termes de multiplications, seule la méthode proposée a une complexité linéaire vis-à-vis du nombre d'échantillons M, ce qui lui permet d'être très compétitive pour l'analyse de grands signaux.

Méthode	Nombre de multiplications
Proposée	$(N \cdot L \cdot (2 \cdot F \cdot (R-1)+1)+2 \cdot F) \cdot M$
Tensor-ESPRIT [Haardt08]	$k_t \cdot F \cdot (R+1) \cdot P \cdot M$
PUMA [So10]	M^3
TPUMA [Sun12]	$k_t \cdot M \cdot (R+F-1) + \sum_{r=1}^{R} K \cdot (F+1) \cdot M_r^3$

TABLE 3.1 : Complexité de l'algorithme en comparaison avec Tensor-ESPRIT, PUMA et TPUMA

3.3 Bornes de Cramér-Rao des paramètres du signal modal amorti

Dans cette partie, je décris la démarche utilisée pour déterminer les bornes de Cramér-Rao (CRLB) des paramètres du modèle exponentiel amorti multidimensionnel. Soit

$$\boldsymbol{\theta} = [\omega_{1,1} \dots \omega_{1,R} \quad \omega_{2,1} \dots \omega_{F,R} \quad \alpha_{1,1} \dots \alpha_{1,R} \quad \alpha_{2,1} \dots \alpha_{F,R} \quad \lambda_1 \dots \lambda_F \quad \phi_1 \dots \phi_F]^T \tag{3.32}$$

le vecteur des paramètres. Étant donné le modèle du bruit, la densité de probabilité conjointe du vecteur $\tilde{\boldsymbol{y}}$ ayant la forme donnée dans (3.7) est :

$$p(\tilde{\boldsymbol{y}};\boldsymbol{\theta}) = \frac{1}{(\sigma^2\pi)^M} \exp\left\{-\frac{1}{\sigma^2}(\tilde{\boldsymbol{y}} - \boldsymbol{y}(\boldsymbol{\theta}))^{\mathsf{H}}(\tilde{\boldsymbol{y}} - \boldsymbol{y}(\boldsymbol{\theta}))\right\} \tag{3.33}$$

Le m^{e} échantillon de $\boldsymbol{y}(\boldsymbol{\theta})$ peut s'écrire :

$$\boldsymbol{y}(\boldsymbol{\theta})_m = \sum_{f=1}^{F} c_f \prod_{r=1}^{R} a_{f,r}^{t_{m,r}}, \tag{3.34}$$

pour $m = 1, \dots, M$, avec

$$t_{m,r} = \left\lfloor \frac{m-1}{\prod_{i=r+1}^{R} M_\ell} \right\rfloor \mod M_r, \tag{3.35}$$

et $\lfloor \cdot \rfloor$ désigne l'arrondi à l'entier inférieur.

3.3.1 Calcul des CRLB

En utilisant l'équation (3.33), l'élément (k,l) de la matrice d'information de Fisher est donné par [Yao95, Kay93] :

$$[\mathbf{I}(\boldsymbol{\theta})]_{kl} = \frac{2}{\sigma^2}\text{Re}\left\{\left[\frac{\partial \boldsymbol{y}(\boldsymbol{\theta})}{\partial \theta_k}\right]^{\mathsf{H}} \frac{\partial \boldsymbol{y}(\boldsymbol{\theta})}{\partial \theta_l}\right\}. \tag{3.36}$$

où $\text{Re}\{\cdot\}$ désigne la partie réelle d'un complexe. Nous devons maintenant exprimer $\partial \boldsymbol{y}(\boldsymbol{\theta})_m / \partial \theta_k$ pour $m = 1, \dots, M$ et $k = 1, \dots, 2RF + 2F$.

- Pour $k = 1, \dots, RF$:

$$\frac{\partial \boldsymbol{y}(\boldsymbol{\theta})_m}{\partial \theta_k} = j t_{m,r_k} c_{f_k} \prod_{r=1}^{R} a_{f_k,r}^{t_{m,r}} \tag{3.37}$$

avec $r_k = [(k-1) \mod R] + 1$, $f_k = \lfloor (k-1)/R \rfloor + 1$ et $n \mod R$ désigne le reste de la division de n par R.

- Pour $k = RF + 1, \ldots, 2RF$:

$$\frac{\partial \boldsymbol{y}(\boldsymbol{\theta})_m}{\partial \theta_k} = t_{m,r_k} c_{f_k} \prod_{r=1}^{R} a_{f_k,r}^{t_{m,r}} \tag{3.38}$$

avec $r_k = [(k - RF - 1) \mod R] + 1$ et $f_k = \lfloor (k - RF - 1)/R \rfloor + 1$.

- Pour $k = 2RF + 1, \ldots, 2RF + F$:

$$\frac{\partial \boldsymbol{y}(\boldsymbol{\theta})_m}{\partial \theta_k} = e^{j\phi_{f_k}} \prod_{r=1}^{R} a_{f_k,r}^{t_{m,r}} \tag{3.39}$$

où $f_k = k - 2RF$.

- Pour $k = 2RF + F + 1, \ldots, 2RF + 2F$:

$$\frac{\partial \boldsymbol{y}(\boldsymbol{\theta})_m}{\partial \theta_k} = j c_{f_k} \prod_{r=1}^{R} a_{f,r}^{t_{m,r}} \tag{3.40}$$

où $f_k = k - 2RF - F$.

Ainsi, la matrice $\partial \boldsymbol{y}(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \in \mathbb{C}^{M \times (2RF+2F)}$ peut être mise sous la forme :

$$\frac{\partial \boldsymbol{y}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \underbrace{[j\boldsymbol{Z}'\boldsymbol{\Phi} \quad \boldsymbol{Z}'\boldsymbol{\Phi} \quad \boldsymbol{Z}\boldsymbol{\phi} \quad j\boldsymbol{Z}\boldsymbol{\phi}]}_{\boldsymbol{V}} \cdot \underbrace{\text{blkdiag}(\boldsymbol{\Lambda}, \boldsymbol{\Lambda}, \boldsymbol{I}_F, \boldsymbol{\lambda})}_{\boldsymbol{S}} \tag{3.41}$$

où

$$\begin{aligned}
&\boldsymbol{Z}' = [\boldsymbol{Z}'_1, \ldots, \boldsymbol{Z}'_F] \in \mathbb{C}^{M \times RF}, \text{avec } \boldsymbol{Z}'_f(m,l) = t_{m,l} \prod_{r=1}^{R} a_{f,r}^{t_{m,r}} \\
&\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_F] \in \mathbb{C}^{M \times F}, \text{avec } \boldsymbol{z}_f(m) = \prod_{r=1}^{R} a_{f,r}^{t_{m,r}} \\
&\boldsymbol{\lambda} = \text{diag}([\lambda_1, \ldots, \lambda_F]) \in \mathbb{R}^{F \times F}, \\
&\boldsymbol{\phi} = \text{diag}([e^{j\phi_1}, \ldots, e^{j\phi_F}]) \in \mathbb{C}^{F \times F} \\
&\boldsymbol{\Lambda} = \boldsymbol{\lambda} \otimes \boldsymbol{I}_R \in \mathbb{R}^{RF \times RF}, \\
&\boldsymbol{\Phi} = \boldsymbol{\phi} \otimes \boldsymbol{I}_R \in \mathbb{C}^{RF \times RF}.
\end{aligned}$$

Finalement, l'inverse de la matrice d'information de Fisher est donnée par :

$$\boldsymbol{I}^{-1}(\boldsymbol{\theta}) = \frac{\sigma^2}{2} \boldsymbol{S}^{-1} \left[\text{Re}\{\boldsymbol{V}^{\mathsf{H}} \boldsymbol{V}\} \right]^{-1} \boldsymbol{S}^{-1} = \frac{\sigma^2}{2} \boldsymbol{S}^{-1} \boldsymbol{W} \boldsymbol{S}^{-1}. \tag{3.42}$$

La borne de Cramér-Rao de θ_k correspond à l'élément $[\boldsymbol{I}^{-1}(\boldsymbol{\theta})]_{kk}$:

$$\begin{aligned}
\mathrm{CRLB}(\omega_{f,r}) &= \frac{2\sigma^2}{\lambda_f^2}\boldsymbol{W}_{R(f-1)+r,R(f-1)+r} \\
\mathrm{CRLB}(\alpha_{f,r}) &= \frac{2\sigma^2}{\lambda_f^2}\boldsymbol{W}_{RF+R(f-1)+r,RF+R(f-1)+r} \\
\mathrm{CRLB}(\lambda_f) &= 2\sigma^2\boldsymbol{W}_{2RF+f,2RF+f} \\
\mathrm{CRLB}(\phi_f) &= \frac{2\sigma^2}{\lambda_f^2}\boldsymbol{W}_{2RF+F+f,2RF+F+f}
\end{aligned}$$

pour $f = 1, \ldots, F$ et $r = 1, \ldots, R$.

Théorème 3.3. *Pour un processus exponentiel R-D, les bornes de Cramér-Rao vérifient :*

$$\mathrm{CRLB}(\omega_{f,r}) = \mathrm{CRLB}(\alpha_{f,r}) \tag{3.43}$$
$$\mathrm{CRLB}(\lambda_f) = \lambda^2\mathrm{CRLB}(\phi_f) \tag{3.44}$$

pour $f = 1, \ldots, F$ et $r = 1, \ldots, R$.

3.3.2 Cas d'une seule composante

Dans ce paragraphe, les CRLB sont simplifiées pour le cas particulier d'un signal mono-composante ($F = 1$). Cela permet d'obtenir des expressions qui se prêtent plus facilement à l'analyse et à l'interprétation. Dans le but de simplifier les notations, la référence au mode unique $f = 1$ sera omise.

Nous devons d'abord exprimer les produits $\mathbf{Z'}^{\mathsf{H}}\mathbf{Z'}$, $\mathbf{Z}^{\mathsf{H}}\mathbf{Z}$ et $\mathbf{Z'}^{\mathsf{H}}\mathbf{Z}$. Supposons que $|a_r| = \exp(\alpha_r) < 1, \forall r$, alors :

$$[\mathbf{Z'}^{\mathsf{H}}\mathbf{Z'}]_{nk} = \prod_{\substack{r=1\\r\neq n,k}}^{R}\left(\frac{1-|a_r|^{2M_r}}{1-|a_r|^2}\right)\cdot\begin{cases}\left(\sum_{m=0}^{M_n-1} m|a_n|^{2m}\right)\left(\sum_{m=0}^{M_k-1} m|a_k|^{2m}\right), & \text{si } n\neq k\\ \left(\sum_{m=0}^{M_n-1} m^2|a_n|^{2m}\right), & \text{si } n=k\end{cases}$$

$$\mathbf{Z}^{\mathsf{H}}\mathbf{Z} = \prod_{r=1}^{R}\left(\frac{1-|a_r|^{2M_r}}{1-|a_r|^2}\right)$$

$$[\mathbf{Z'}^{\mathsf{H}}\mathbf{Z}]_n = \prod_{\substack{r=1\\r\neq n}}^{R}\left(\frac{1-|a_r|^{2M_r}}{1-|a_r|^2}\right)\cdot\sum_{m=0}^{M_n-1} m|a_n|^{2m}.$$

On pose :

$$
\begin{aligned}
M^{(\alpha)} &= \prod_{r=1}^{R}(1-|a_r|^{2M_r})/(1-|a_r|^2) \\
q_i(n) &= \sum_{m=0}^{M_n-1} m^i |a_n|^{2m} / \sum_{m=0}^{M_n-1} |a_n|^{2m}, \quad i \in \{1,2\} \\
[\mathbf{P}]_{nk} &= M^{(\alpha)} \times \begin{cases} q_1(n)q_1(k), & \text{si } n \neq k \\ q_2(n), & \text{si } n = k \end{cases} \\
\mathbf{G} &= M^{(\alpha)} \\
[\mathbf{Q}]_n &= M^{(\alpha)} q_1(n),
\end{aligned}
$$

alors

$$
\mathrm{Re}\{\mathbf{V}^{\mathsf{H}}\mathbf{V}\} = \begin{bmatrix} \mathbf{P} & 0 & 0 & \mathbf{Q} \\ 0 & \mathbf{P} & \mathbf{Q} & 0 \\ 0 & \mathbf{Q}^T & \mathbf{G} & 0 \\ \mathbf{Q}^T & 0 & 0 & \mathbf{G} \end{bmatrix}. \tag{3.45}
$$

L'inversion de cette matrice donne les expressions suivantes :

$$
\mathrm{CRLB}(\omega_r) = \mathrm{CRLB}(\alpha_r) = \frac{\sigma^2}{2\lambda^2 M^{(\alpha)}} \cdot \frac{(1-|a_r|^2)^2(1-|a_r|^{2M_r})^2}{-M_r^2|a_r|^{2M_r}(1-|a_r|^2)^2 + |a_r|^2(1-|a_r|^{2M_r})^2} \tag{3.46}
$$

$$
\frac{\mathrm{CRLB}(\lambda)}{\lambda^2} = \mathrm{CRLB}(\phi) = \frac{\sigma^2}{2\lambda^2 M^{(\alpha)}} \left(1 + \sum_{r=1}^{R} \frac{q_1^2(r)}{q_2(r) - q_1^2(r)}\right). \tag{3.47}
$$

Finalement, dans le cas harmonique ($\alpha_r = 0, \forall r$), nous avons $M^{(\alpha)} = \prod_{r=1}^{R} M_r \triangleq M$ et en prenant la limite des CRLB quand $\alpha_r \to 0$, on obtient :

$$
\lim_{\alpha_r \to 0} \mathrm{CRLB}(\omega_r) = \frac{6\sigma^2}{\lambda^2 M(M_r^2 - 1)} \tag{3.48}
$$

$$
\lim_{\alpha_r \to 0} \frac{\mathrm{CRLB}(\lambda)}{\lambda^2} = \frac{\sigma^2}{2\lambda^2 M} \left(1 + 3\sum_{r=1}^{R} \frac{M_r - 1}{M_r + 1}\right). \tag{3.49}
$$

Dans le cas non amorti, l'expression (3.48) correspond bien au résultat déjà rapporté dans [Sun12].

3.4 Conclusion

Le problème traité dans ce chapitre concerne l'estimation modale (signal harmonique ou amorti) multidimensionnelle (R-D) formulé comme un problème d'approximation parcimonieuse simultanée. Cette idée permet de remplacer un problème d'estimation joint en R sous-problèmes résolus séparément. La première contribution porte sur la mise à jour multigrille du dictionnaire pour réduire le coût de calcul. Nous avons donné les conditions de convergence de cette procédure dans le cas mono-composante. Pour un signal multi-composantes, le tenseur de données est décomposé de façon à estimer séparément chaque mode multidimensionnel, ce qui constitue notre deuxième contribution

méthodologique. Par ailleurs, les modes obtenus sont automatiquement couplés ce qui permet d'éviter une étape supplémentaire de reformation de modes. L'algorithme ainsi obtenu est très compétitif en terme de performance et, surtout, de complexité numérique. Enfin, nous avons également établi les bornes de Cramér-Rao des paramètres du signal R-D.

Chapitre 4

Analyse des signaux de spectroscopie infrarouge : sélection de variables et classification

Ce chapitre est consacré aux travaux que j'ai réalisé récemment sur le problème de sélection de variables pour la classification en spectroscopie proche infrarouge. Ce travail a été développé dans le cadre du FUI Trispirabois résumé dans le chapitre 1.

La spectroscopie proche infrarouge (en anglais *NIR : Near-Infrared*) est une technique d'analyse qui permet d'obtenir des informations sur la composition et les interactions dans un échantillon [Stuart05, Siesler02]. Comme le spectre obtenu constitue une sorte de signature caractérisant le matériau analysé, cette technique est utilisée dans plusieurs applications comme l'identification, la caractérisation ou le contrôle non destructif [Adams95, Jonsson05]. Ces dernières années, la spectroscopie NIR a été renforcée par le développement de systèmes rapides d'imagerie hyperspectrale ayant ouvert la voie à ce que l'on appelle « imagerie chimique ». Il existe essentiellement quatre techniques d'acquisition d'images hyperspectrales [Li13, Willett14] : *whiskbroom* (scan ponctuel), *pushbroom* (scan linéaire), filtre accordable (scan par longueur d'onde) et *snapshot* (acquisition en une seule passe de l'image hyperspectrale).

Dans le projet FUI, nous nous sommes focalisés sur les imageurs pushbroom qui sont utilisés dans de nombreuses applications telles que le contrôle de qualité dans l'agroalimentaire [Foley98, Elmasry12, Lorente12, Huang14], le géo-référencement [Cariou08] et le tri de matériaux [PellencST17, Tatzer05, Yoon11, Serranti12]. La finalité du projet est le tri de matériaux bois formulé comme un problème de classification supervisée, en conditions industrielles. Pour faire face au problème de dimensionalité, une étape clef est la sélection de variables. L'idée est de trouver un sous-ensemble de variables explicatives qui décrit au mieux les principales caractéristiques des réponses associées aux différentes classes. Dans une approche similaire à celle de Turlach *et al.* [Turlach05], nous avons proposé des algorithmes de sélection basés sur la parcimonie simultanée. Les développements présentés dans ce chapitre sont issus des travaux de thèse de Leila Belmerhnia[1].

[1] Thèse de doctorat, 2013-2016.

4.1 Formulation du problème

On considère une matrice de données $\boldsymbol{Y} \in \mathbb{R}^{M \times K}$ composée de K spectres (les réponses) chacun de taille M. On veut décomposer cette matrice de la façon suivante :

$$\boldsymbol{Y} \approx \boldsymbol{\Phi}\boldsymbol{X}, \tag{4.1}$$

où $\boldsymbol{X} \in \mathbb{R}^{N \times K}$ est une matrice *parcimonieuse* de coefficients, ce qui signifie que seulement une petite partie de ses lignes est non nulle. Les colonnes du dictionnaire $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N] \in \mathbb{R}^{M \times N}$ contiennent les variables explicatives. Il est construit de façon à concentrer l'énergie des signaux dans $\boldsymbol{Y}$ sur un ensemble réduit d'atomes. Le choix de ce dictionnaire est donc crucial et dépend grandement de l'application. Comme en spectroscopie NIR les pics observés sont typiquement larges, nous supposons ici que les colonnes $\boldsymbol{\phi}_n$ sont des fonctions gaussiennes dont les centres et les largeurs couvrent tout l'intervalle des longueurs d'onde NIR.

L'approximation parcimonieuse simultanée [Cotter05, Chen06] consiste à trouver une solution $\boldsymbol{X}$ ayant un nombre limité de lignes actives. Le problème peut donc être formulé de la façon suivante :

$$\underset{\boldsymbol{X}}{\text{minimiser}} \quad \tfrac{1}{2}||\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{X}||_F^2, \tag{4.2a}$$

$$\text{sous contrainte} \quad ||\boldsymbol{X}||_{0,2} \leq s, \tag{4.2b}$$

où $||\boldsymbol{X}||_{0,2}$ désigne la pseudo-nome mixte ℓ_0/ℓ_2 de la matrice $\boldsymbol{X}$ (nombre de lignes ayant une norme ℓ_2 non nulle) et $s \ll N$ est la cardinalité du support de $\boldsymbol{X}$:

$$\text{supp}(\boldsymbol{X}) = \{1 \leq n \leq N \,|\, \boldsymbol{x}^n \neq \boldsymbol{0}\}, \tag{4.3}$$

où $\boldsymbol{x}^n$ est le vecteur formé par les éléments de la n^{e} ligne de $\boldsymbol{X}$. L'objectif de la reconstruction simultanée est de sélectionner un *sous-ensemble commun de variables actives* pour tous les signaux dans $\boldsymbol{Y}$. Ce sous-ensemble est indexé par le support de la solution $\boldsymbol{X}$.

L'approximation parcimonieuse simultanée et régularisée a pour but de reconstruire une matrice de coefficients dont les lignes sont constantes par morceaux. Le problème associé peut être formulé comme suit :

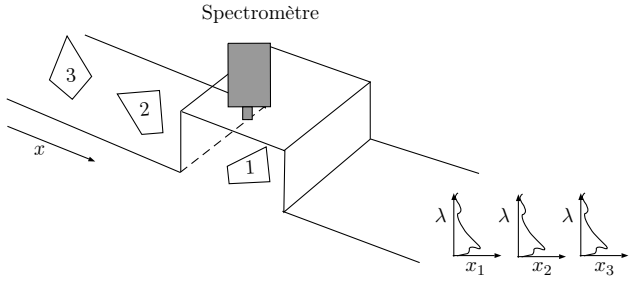
$$\underset{\boldsymbol{X}}{\text{minimiser}} \quad \tfrac{1}{2}||\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{X}||_F^2 + \lambda_2||\boldsymbol{D}\boldsymbol{X}^\mathsf{T}||_{1,1}, \tag{4.4a}$$

$$\text{sous contrainte} \quad ||\boldsymbol{X}||_{0,2} \leq s, \tag{4.4b}$$

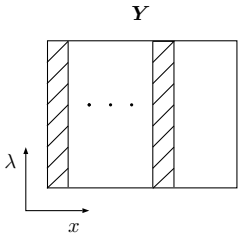
où $\boldsymbol{D} \in \mathbb{R}^{(K-1)\times K}$ est une matrice de différences finies d'ordre 1 :

$$\boldsymbol{D} = \begin{bmatrix} -1 & 1 & & \boldsymbol{0} \\ & \ddots & \ddots & \\ \boldsymbol{0} & & -1 & 1 \end{bmatrix}. \tag{4.5}$$

Le critère (4.4) inclut une pénalité de type variation totale qui impose une parcimonie sur la différence entre les colonnes de $\boldsymbol{X}$: $||\boldsymbol{D}\boldsymbol{X}^\mathsf{T}||_{1,1} = \sum_{i=1}^{K-1} ||\boldsymbol{x}_{i+1} - \boldsymbol{x}_i||_1$. C'est ce qui permet d'obtenir une solution $\boldsymbol{X}$ dont les lignes sont constantes par morceaux. Il est important de noter ici que cette pénalité n'a de

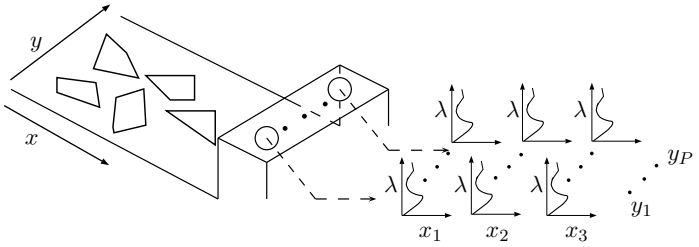


(a) Spectromètre et tapis roulant

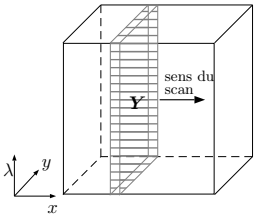


(b) Matrice de données

FIGURE 4.1 : Construction de la matrice de données composée de spectres acquis par un spectromètre placé dans une cabine. Le tapis roulant entraîne les pièces à scanner. (x) dimension spatiale ; (λ) dimension spectrale



(a) Imageur hyperspectral et tapis roulant



(b) Cube de données

FIGURE 4.2 : Images hyperspectrales acquises par un spectro-imageur NIR industriel. La matrice $\boldsymbol{Y}$ est une tranche du cube de données. (x, y) dimensions spatiales ; (λ) dimension spectrale

sens que si les signaux dans $\boldsymbol{Y}$ sont ordonnés. C'est le cas par exemple quand ces signaux sont rangés selon leurs étiquettes ou leurs classes d'appartenance dans la phase d'apprentissage (fig. 4.1). En classification d'images hyperspectrales, les signaux sont naturellement ordonnés selon leur position spatiale (fig. 4.2). Par la suite, la pénalité ℓ_1 sera désignée par le terme de « fusion ». Comme la contrainte dans (4.4) est formulée par une peudo-norme mixte ℓ_0/ℓ_2, le problème sous-jacent est NP-difficile. À moins d'utiliser une quelconque heuristique, il est impossible de garantir une solution efficace au problème. Dans le paragraphe suivant, nous proposons des relaxations convexes à la norme ℓ_0/ℓ_1, ce qui permettra de trouver des solutions en utilisant des algorithmes plus simples.

4.2 Approximation parcimonieuse simultanée et régularisée : relaxations convexes

4.2.1 Fused Sparse Lasso (FSL)

La première relaxation du problème (4.4) consiste à utiliser le critère suivant :

$$J_1(\boldsymbol{X}; \lambda_1, \lambda_2) = \tfrac{1}{2}||\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{X}||_F^2 + \lambda_1||\boldsymbol{X}||_{1,1} + \lambda_2||\boldsymbol{D}\boldsymbol{X}^\mathsf{T}||_{1,1} \tag{4.6}$$

où $||\boldsymbol{X}||_{1,1} = \sum_{i=1}^{K} ||\boldsymbol{x}_i||_1 = \sum_{n=1}^{N} ||\boldsymbol{x}^n||_1$. Les hyperparamètres $\lambda_1, \lambda_2 \geq 0$ contrôlent le compromis entre la fidélité aux données, la parcimonie globale $||\boldsymbol{X}||_{1,1}$ et la pénalité de fusion $||\boldsymbol{D}\boldsymbol{X}^\mathsf{T}||_{1,1}$.

Ce critère constitue une extension au cas MMV (*multiple measurement vector*) du problème *sparse fused Lasso* étudié dans [Tibshirani05] où une solution basée sur l'algorithme des contraintes actives à deux phases [Gill98] développé pour les problèmes de programmation quadratique avec des contraintes linéaires de parcimonie. Dans [Tibshirani11], le problème est étendu au fused Lasso généralisé (GFL) dans lequel la matrice $\boldsymbol{D}$ est quelconque mais sans le terme de parcimonie. Friedman *et al.* [Friedman07] ont proposé une solution au problème fused Lasso intégrant les termes de parcimonie et de fusion dans le cas particulier où $\boldsymbol{\Phi}$ est une matrice identité. L'approche est ensuite étendue dans [Xin14] pour un dictionnaire quelconque et un terme de fusion qui n'agit pas nécessairement sur des variables consécutives. Nous proposons une solution à ce problème dans le cas particulier où le terme de fusion agit uniquement sur les lignes de $\boldsymbol{X}$: cette spécificité permet d'obtenir un algorithme très efficace et rapide qui permet son utilisation pour résoudre des problèmes de grande dimension.

Remarque 4.1. *Le terme de parcimonie dans le critère (4.6) ne correspond pas vraiment à une relaxation de* $||\boldsymbol{X}||_{0,2}$. *En fait, la norme mixte* $||\boldsymbol{X}||_{1,2}$ *est plus appropriée (nous reviendrons sur ce point dans le paragraphe suivant). Cependant, la combinaison des termes* $||\boldsymbol{X}||_{1,1}$ *et* $||\boldsymbol{D}\boldsymbol{X}^{\mathsf{T}}||_{1,1}$ *va permettre d'aboutir, si* λ_2 *est grand, à une approximation parcimonieuse simultanée : la simultanéité étant imposée par le terme de fusion. Par contre, si* λ_2 *est faible, le nombre de lignes actives n'est plus contrôlé directement par* λ_1.

Soit vec$(\cdot)$ la transformation qui convertit une matrice en vecteur en superposant l'une sur l'autre les colonnes de la matrice. On pose :

$$\boldsymbol{x} = \text{vec}(\boldsymbol{X}^{\mathsf{T}}), \tag{4.7}$$

$$\boldsymbol{y} = \text{vec}(\boldsymbol{Y}^{\mathsf{T}}). \tag{4.8}$$

On peut facilement montrer que le critère (4.6) peut s'écrire sous la forme vectorielle suivante :

$$J_1(\boldsymbol{x}; \lambda_1, \lambda_2) = \tfrac{1}{2}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2 + \lambda_1||\boldsymbol{x}||_1 + \lambda_2||\boldsymbol{F}\boldsymbol{x}||_1 \tag{4.9}$$

avec $\boldsymbol{A} = \boldsymbol{\Phi} \otimes \boldsymbol{I}_K \in \mathbb{R}^{NK \times MK}$, $\boldsymbol{F} = \boldsymbol{I}_N \otimes \boldsymbol{D} \in \mathbb{R}^{N(K-1) \times NK}$ et $\boldsymbol{I}_N$ désigne la matrice identité de taille $N \times N$. Pour minimiser (4.9), nous utilisons l'algorithme FISTA (*fast iterative shrinkage-thresholding algorithm*) [Beck09]. En notant $f(x)$ le terme de fidélité aux données et $\Omega(x)$ le terme de régularisation (non différentiable) :

$$f(\boldsymbol{x}) = \tfrac{1}{2}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2, \tag{4.10}$$

$$\Omega(\boldsymbol{x}) = \lambda_1||\boldsymbol{x}||_1 + \lambda_2||\boldsymbol{F}\boldsymbol{x}||_1, \tag{4.11}$$

alors la mise à jour à l'itération $k+1$ de $\boldsymbol{x}$ s'exprime par :

$$\boldsymbol{x}_{(k+1)} = \arg\min_{\boldsymbol{x} \in \mathbb{R}^{NK}} \left\{ \Omega(\boldsymbol{x}) + \tfrac{L}{2}||\boldsymbol{x} - \boldsymbol{v}_{(k)}||_2^2 \right\} \tag{4.12}$$

où

$$\boldsymbol{v}_{(k)} = \boldsymbol{x}_{(k)} - \tfrac{1}{L}\nabla f(\boldsymbol{x}_{(k)}) \tag{4.13}$$

et $\nabla f(\boldsymbol{x}) = \boldsymbol{A}^{\mathsf{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})$ est le gradient de $f(\boldsymbol{x})$. L est la constante de Lipschitz de $\nabla f(\boldsymbol{x})$. On note

Algorithme 1 : Fused Sparse Lasso (FSL)

Entrées : $\boldsymbol{Y} \in \mathbb{C}^{M\times K}$, $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$, λ_1, λ_2, *maxiter*

1 Initialisation : $\boldsymbol{X}_{(0)} = \boldsymbol{0}$, $\boldsymbol{Z}_{(1)} = \boldsymbol{0}$, $t_{(1)} = 1$, $L \leq ||\boldsymbol{\Phi}||_2^2$;
2 **pour** $k \leftarrow 1$ **à** *maxiter* **faire**
3 $\quad \boldsymbol{V}_{(k)} \leftarrow \boldsymbol{Z}_{(k)} - \frac{1}{L}\boldsymbol{\Phi}^\mathsf{T}(\boldsymbol{\Phi}\boldsymbol{Z}_{(k)} - \boldsymbol{Y})$;
4 $\quad$ **pour** $n \leftarrow 1$ **à** N **faire**
5 $\quad\quad \mathbf{x}^n_{(k)} \leftarrow \arg\min_{\boldsymbol{x}} \frac{1}{2}||\boldsymbol{x} - \boldsymbol{v}^i_{(k)}||_2^2 + \frac{\lambda_1}{L}||\boldsymbol{x}||_1 + \frac{\lambda_2}{L}||\boldsymbol{D}\boldsymbol{x}||_1$; `// résolution par flsa`
6 $\quad$ **fin**
7 $\quad t_{(k+1)} \leftarrow \frac{1+\sqrt{1+4t^2_{(k)}}}{2}$; `// prochaine itération FISTA`
8 $\quad \boldsymbol{Z}_{(k+1)} \leftarrow \boldsymbol{X}_{(k)} + \frac{t_{(k)}-1}{t_{(k+1)}}(\boldsymbol{X}_{(k)} - \boldsymbol{X}_{(k-1)})$;
9 **fin**

Sortie : $\boldsymbol{X} \in \mathbb{R}^{N\times K}$

que la mise à jour de $\boldsymbol{v}_{(k)}$ selon (4.13) nécessite le calcul et la sauvegarde de $\boldsymbol{A}^\mathsf{T}\boldsymbol{A}$ et $\boldsymbol{A}^\mathsf{T}\boldsymbol{y}$. Pour réduire les coûts de calcul et l'empreinte mémoire, on peut mettre à jour la matrice $\boldsymbol{V}_{(k)}$:

$$\boldsymbol{V}_{(k)} = \boldsymbol{X}_{(k)} - \tfrac{1}{L}\boldsymbol{\Phi}^\mathsf{T}(\boldsymbol{\Phi}\boldsymbol{X}_{(k)} - \boldsymbol{Y}), \tag{4.14}$$

où $\boldsymbol{V}_{(k)}$ est telle que $\boldsymbol{v}_{(k)} = \mathrm{vec}(\boldsymbol{V}^\mathsf{T}_{(k)})$. Par conséquent, seulement des matrices de tailles réduites doivent être calculées : $\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{\Phi}$ et $\boldsymbol{\Phi}^\mathsf{T}\boldsymbol{Y}$. Le problème (4.12) est similaire au *fused Lasso signal approximator* (FLSA) [Hoefling10]. De plus, grâce à la structure bloc diagonale de la matrice $\boldsymbol{F}$, on peut montrer que ce problème peut être résolu indépendamment pour chaque ligne $\boldsymbol{x}^n$:

$$\boldsymbol{x}^n_{(k+1)} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^K} \tfrac{1}{2}||\boldsymbol{x} - \boldsymbol{v}^n_{(k)}||_2^2 + \tfrac{\lambda_1}{L}||\boldsymbol{x}||_1 + \tfrac{\lambda_2}{L}||\boldsymbol{D}\boldsymbol{x}||_1. \tag{4.15}$$

Théorème 4.1 ([Friedman07, Liu10])**.** *Soit*

$$\boldsymbol{x}(\lambda_1, \lambda_2) = \arg\min_{\boldsymbol{x}} \tfrac{1}{2}||\boldsymbol{x} - \boldsymbol{v}||_2^2 + \lambda_1||\boldsymbol{x}||_1 + \lambda_2||\boldsymbol{D}\boldsymbol{x}||_1. \tag{4.16}$$

Alors, pour tout λ_1, $\lambda_2 \geq 0$ *:*

$$\boldsymbol{x}(\lambda_1, \lambda_2) = \mathrm{sign}(\boldsymbol{x}(0, \lambda_2)) * \max(|\boldsymbol{x}(0, \lambda_2)| - \lambda_1, 0). \tag{4.17}$$

où $*$ *est le produit de Hadamard (terme à terme).* □

Ce résultat est utilisé dans [Friedman07, Hoefling10] pour développer des algorithmes de recherche du chemin de régularisation selon λ_2 avec une valeur fixée de λ_1 (typiquement $\lambda_1 = 0$). Dans le pseudo-code présenté dans Algorithme 1, le problème (4.15) est résolu en utilisant la routine `flsa` implémentée dans le paquetage SLEP[2]. Les lignes 7 et 8 sont associées à la procédure d'accélération de FISTA. La ligne 8 correspond à une combinaison linéaire très particulière des deux points précédents $\{\boldsymbol{X}_{(k)}, \boldsymbol{X}_{(k-1)}\}$.

[2] http://yelab.net/software/SLEP/

4.2.2 Fused Sparse Group Lasso (FSGL)

Comme mentionné précédemment, le problème fused sparse Lasso n'est pas une bonne relaxation du problème (4.4). En effet, si λ_2 est faible, il est possible d'obtenir une matrice parcimonieuse $\boldsymbol{X}$ dont les coefficient actifs ne sont pas nécessairement regroupés dans un nombre limité de lignes. Une façon d'imposer la simultanéité est de remplacer la pseudo-norme ℓ_0/ℓ_2 par la norme mixte ℓ_1/ℓ_2 définie par : $||\boldsymbol{X}||_{1,2} = \sum_{n=1}^{N} ||\boldsymbol{x}^n||_2$. Cette dernière constitue une instance particulière de la pénalité Lasso par groupe. Le critère correspondant est donné par :

$$J_2(\boldsymbol{x}; \lambda_1, \lambda_2, \lambda_3) = \tfrac{1}{2}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2 + \lambda_1||\boldsymbol{x}||_1 + \lambda_2||\boldsymbol{F}\boldsymbol{x}||_1 + \lambda_3 \sum_{n=1}^{N} ||\boldsymbol{x}^n||_2. \tag{4.18}$$

À noter que la pénalité $||\boldsymbol{x}||_1$ est maintenue pour permettre le contrôle de la parcimonie globale de la solution. L'opérateur proximal associé aux trois derniers termes non lisses de (4.18) s'écrit :

$$\begin{aligned} \text{prox}_{FSGL}(\boldsymbol{v}) = \arg\min_{\boldsymbol{x}} & \tfrac{1}{2}||\boldsymbol{x} - \boldsymbol{v}||_2^2 \\ & + \tfrac{\lambda_1}{L}||\boldsymbol{x}||_1 + \tfrac{\lambda_2}{L}||\boldsymbol{F}\boldsymbol{x}||_1 + \tfrac{\lambda_3}{L} \sum_{n=1}^{N} ||\boldsymbol{x}^n||_2. \end{aligned} \tag{4.19}$$

De la même façon que précédemment, comme les lignes de $\boldsymbol{X}$ sont découplées dans l'équation (4.19), il suffit de résoudre le problème d'optimisation pour chaque ligne $n = 1, \ldots, N$:

$$\begin{aligned} \text{prox}_{FSGL}(\boldsymbol{v}^n) = \arg\min_{\boldsymbol{x}^n} & \tfrac{1}{2}||\boldsymbol{x}^n - \boldsymbol{v}^n||_2^2 \\ & + \tfrac{\lambda_1}{L}||\boldsymbol{x}^n||_1 + \tfrac{\lambda_2}{L}||\boldsymbol{D}\boldsymbol{x}^n||_1 + \tfrac{\lambda_3}{L}||\boldsymbol{x}^n||_2. \end{aligned} \tag{4.20}$$

L'opérateur proximal dans (4.20) possède une propriété de décomposition qui permet son calcul en deux étapes.

Théorème 4.2 ([Zhou12]). *Soient*

$$\text{prox}_{FSL}(\boldsymbol{v}) = \arg\min_{\boldsymbol{x}} \tfrac{1}{2}||\boldsymbol{x} - \boldsymbol{v}||_2^2 + \lambda_1||\boldsymbol{x}||_1 + \lambda_2||\boldsymbol{D}\boldsymbol{x}||_1, \tag{4.21}$$

$$\text{prox}_{GL}(\boldsymbol{v}) = \arg\min_{\boldsymbol{x}} \tfrac{1}{2}||\boldsymbol{x} - \boldsymbol{v}||_2^2 + \lambda_3||\boldsymbol{x}||_2. \tag{4.22}$$

Alors, pour tout $\lambda_1, \lambda_2, \lambda_3 \geq 0$ *:*

$$\text{prox}_{FSGL}(\boldsymbol{v}) = (prox_{GL} \diamond \text{prox}_{FSL})(\boldsymbol{v}). \tag{4.23}$$

où $\diamond$ *est l'opérateur de composition.* □

Finalement, le pseudo-code de l'algorithme FSGL est donné dans Algorithme 2. Le problème d'optimisation groupe-Lasso dans la ligne 6 est résolu par la routine `altra` également disponible dans SLEP.

Algorithme 2 : Fused Sparse Group Lasso (FSGL)

Entrées : $\boldsymbol{Y} \in \mathbb{C}^{M\times K}$, $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$, λ_1, λ_2, λ_3, *maxiter*

Initialisation : $\boldsymbol{X}_{(0)} = \boldsymbol{0}$, $\boldsymbol{Z}_{(1)} = \boldsymbol{0}$, $t_{(1)} = 1$, $L \leq ||\boldsymbol{\Phi}||_2^2$;
pour $k \leftarrow 1$ **à** *maxiter* **faire**
  $\boldsymbol{V}_{(k)} \leftarrow \boldsymbol{Z}_{(k)} - \frac{1}{L}\boldsymbol{\Phi}^\mathsf{T}(\boldsymbol{\Phi}\boldsymbol{Z}_{(k)} - \boldsymbol{Y})$;
  pour $n \leftarrow 1$ **à** N **faire**
    $\mathbf{w}_{(k)}^n \leftarrow \arg\min_{\boldsymbol{x}} \frac{1}{2}||\boldsymbol{x} - \boldsymbol{v}_{(k)}^n||_2^2 + \frac{\lambda_1}{L}||\boldsymbol{x}||_1 + \frac{\lambda_2}{L}||\boldsymbol{D}\boldsymbol{x}||_1$; `// résolution par flsa`
    $\boldsymbol{x}_{(k)}^n \leftarrow \arg\min_{\boldsymbol{x}} \frac{1}{2}||\boldsymbol{x} - \mathbf{w}_{(k)}^n||_2^2 + \frac{\lambda_3}{L}||\boldsymbol{x}||_2$; `// résolution par altra`
  fin
  $t_{(k+1)} \leftarrow \frac{1+\sqrt{1+4t_{(k)}^2}}{2}$; `// prochaine itération FISTA`
  $\boldsymbol{Z}_{(k+1)} \leftarrow \boldsymbol{X}_{(k)} + \frac{t_{(k)}-1}{t_{(k+1)}}(\boldsymbol{X}_{(k)} - \boldsymbol{X}_{(k-1)})$;
fin

Sortie : $\boldsymbol{X} \in \mathbb{R}^{N\times K}$

4.2.3 Fused Sparse Group Lasso non négatif

En spectroscopie, tout comme beaucoup d'autres applications, il est judicieux d'imposer une contrainte de non-négativité à la solution. Cela permet de garantir que chacun des spectres dans la matrice $\boldsymbol{Y}$ ne peut être représenté que par la somme de « spectres élémentaires » (donc positifs). Une telle décomposition offre également l'avantage d'avoir une meilleure interprétation physique. La version non négative du critère FSGL s'écrit :

$$\underset{\boldsymbol{x}}{\text{minimiser}} \quad J_2(\boldsymbol{x}; \lambda_1, \lambda_2, \lambda_3), \tag{4.24a}$$

$$\text{sous contrainte} \quad \boldsymbol{x} \geq 0. \tag{4.24b}$$

Plusieurs méthodes peuvent être utilisées pour optimiser cet objectif parmi lesquelles on peut citer la méthode de la pénalité quadratique [Nocedal06] et l'algorithme ADMM [Boyd11, Parikh13]. Je présente ici une solution basée sur la pénalité quadratique. Pour ce faire, on remplace les contraintes d'inégalité par des contraintes d'égalité en définissant une variable auxiliaire vectorielle $\boldsymbol{u} \in \mathbb{R}^{NK}$, ce qui conduit au problème :

$$\underset{\boldsymbol{x}}{\text{minimiser}} \quad J_2(\boldsymbol{x}; \lambda_1, \lambda_2, \lambda_3), \tag{4.25a}$$

$$\text{sous contrainte} \quad \boldsymbol{x} - \boldsymbol{u} = 0, \boldsymbol{u} \geq 0. \tag{4.25b}$$

La pénalité quadratique consiste à remplacer les contraintes par des termes de pénalité dans la fonction objectif qui devient :

$$\begin{aligned} J_3(\boldsymbol{x}, \boldsymbol{u}; \lambda_1, \lambda_2, \lambda_3) &= \tfrac{1}{2}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2 + \tfrac{\xi}{2}||\boldsymbol{x} - \boldsymbol{u}||_2^2 \\ &+ \lambda_1||\boldsymbol{x}||_1 + \lambda_2||\boldsymbol{F}\boldsymbol{x}||_1 + \lambda_3 \sum_{n=1}^{N} ||\boldsymbol{x}^n||_2, \quad \boldsymbol{u} \geq 0 \end{aligned} \tag{4.26}$$

Algorithme 3 : Fused Sparse Group Lasso non négatif (NN-FSGL)

Entrées : $\boldsymbol{Y} \in \mathbb{C}^{M\times K}$, $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$, λ_1, λ_2, λ_3, β, *maxiter*, *nniter*

Initialisation : $\boldsymbol{X}_{(0)} = \boldsymbol{0}$, $\boldsymbol{Z}_{(1)} = \boldsymbol{0}$, $t_{(1)} = 1$, $\boldsymbol{U}_{(1)} = 0$, $\xi_{(1)} = 1$, $L \leq ||\boldsymbol{\Phi}||_2^2$;
pour $\ell \leftarrow 1$ **à** *nniter* **faire**
  $L_p \leftarrow L + \xi_{(\ell)}$;
  pour $k \leftarrow 1$ **à** *maxiter* **faire**
    $\boldsymbol{V}_{(k)} \leftarrow \boldsymbol{Z}_{(k)} - \frac{1}{L_p}\boldsymbol{\Phi}^\mathsf{T}(\boldsymbol{\Phi}\boldsymbol{Z}_{(k)} - \boldsymbol{Y}) - \frac{\xi_{(\ell)}}{L_p}(\boldsymbol{Z} - \mathbf{U}_{(\ell)})$;
    pour $n \leftarrow 1$ **à** N **faire**
      $\mathbf{w}_{(k)}^n \leftarrow \arg\min_{\boldsymbol{x}} \frac{1}{2}||\boldsymbol{x} - \boldsymbol{v}_{(k)}^n||_2^2 + \frac{\lambda_1}{L_p}||\boldsymbol{x}||_1 + \frac{\lambda_2}{L_p}||\boldsymbol{D}\boldsymbol{x}||_1$; `// résolution par flsa`
      $\boldsymbol{x}_{(k)}^n \leftarrow \arg\min_{\boldsymbol{x}} \frac{1}{2}||\boldsymbol{x} - \mathbf{w}_{(k)}^n||_2^2 + \frac{\lambda_3}{L_p}||\boldsymbol{x}||_2$; `// résolution par altra`
    fin
    $t_{(k+1)} \leftarrow \frac{1+\sqrt{1+4t_{(k)}^2}}{2}$; `// prochaine itération FISTA`
    $\boldsymbol{Z}_{(k+1)} \leftarrow \boldsymbol{X}_{(k)} + \frac{t_{(k)}-1}{t_{(k+1)}}(\boldsymbol{X}_{(k)} - \boldsymbol{X}_{(k-1)})$;
  fin
  $\boldsymbol{u}_{(\ell+1)} \leftarrow \max(0, \boldsymbol{x}_{(\text{maxiter})})$; `// seuillage dur`
  $\xi_{(\ell+1)} \leftarrow \beta\xi_{(\ell)}$;
fin

Sortie : $\boldsymbol{X} \in \mathbb{R}^{N\times K}$

où ξ est le paramètre qui pénalise la violation des contraintes. Ainsi, quand $\xi \to \infty$, les éléments du vecteur $\boldsymbol{x}$ tendent vers ceux du vecteur $\boldsymbol{u}$ et la contrainte d'inégalité $\boldsymbol{x} \geq 0$ est satisfaite asymptotiquement. En combinant les termes quadratiques dans le critère auxiliaire (4.26), on obtient :

$$J_3(\boldsymbol{x}, \boldsymbol{u}; \lambda_1, \lambda_2, \lambda_3) = \tfrac{1}{2}||\boldsymbol{z}_{\boldsymbol{u}} - \boldsymbol{B}\boldsymbol{x}||_2^2 + \Lambda(\boldsymbol{x}), \quad \boldsymbol{u} \geq 0 \tag{4.27}$$

où $\boldsymbol{B} = [\boldsymbol{A}^\mathsf{T}, \sqrt{\xi}\boldsymbol{I}]^\mathsf{T}$, $\boldsymbol{z}_{\boldsymbol{u}} = [\boldsymbol{y}^\mathsf{T}, \sqrt{\xi}\boldsymbol{u}^\mathsf{T}]^\mathsf{T}$ et $\boldsymbol{I}$ est une matrice identité de même taille que $\boldsymbol{A}$. Le terme $\Lambda(\boldsymbol{x})$ est donné par :

$$\Lambda(\boldsymbol{x}) = \lambda_1||\boldsymbol{x}||_1 + \lambda_2||\boldsymbol{F}\boldsymbol{x}||_1 + \lambda_3 \sum_{i=1}^{N} ||\boldsymbol{x}^i||_2. \tag{4.28}$$

La minimisation de J_3 par rapport à $(\boldsymbol{x}, \boldsymbol{u})$ s'effectue en alternant la minimisation sans contrainte par rapport à $\boldsymbol{x}$, minimisation contrainte par rapport à $\boldsymbol{u}$ et augmentation du paramètre ξ. La minimisation de J_3 par rapport à $\boldsymbol{x}$ aboutit à une itération similaire à celle de FSGL :

$$\begin{cases} \boldsymbol{v}_{(k)} = \boldsymbol{x}_{(k)} - \frac{1}{L_p}\nabla g(\boldsymbol{x}_{(k)}) \\ \boldsymbol{x}_{(k+1)} = \arg\min_{\boldsymbol{x}} \frac{1}{2}||\boldsymbol{x} - \boldsymbol{v}_{(k)}||_2^2 + \frac{1}{L_p}\Lambda(\boldsymbol{x}), \end{cases} \tag{4.29}$$

où $g(\boldsymbol{x}) = \frac{1}{2}||\boldsymbol{z}_{\boldsymbol{u}} - \boldsymbol{B}\boldsymbol{x}||_2^2$, $\nabla g(\boldsymbol{x}) = \boldsymbol{B}^\mathsf{T}(\boldsymbol{B}\boldsymbol{x} - \boldsymbol{z}_{\boldsymbol{u}})$ et $L_p = L + \xi$ est la constante de Lipschitz de $\nabla g(\boldsymbol{x})$. Encore une fois, l'optimisation peut être découplée pour chaque ligne $\boldsymbol{x}^n$. On définit la

matrice $\boldsymbol{U}$ telle que $\boldsymbol{u} = \text{vec}(\boldsymbol{U}^\mathsf{T})$. En remplaçant $\boldsymbol{B}$ et $\boldsymbol{z_u}$ par leurs expressions, on obtient :

$$\begin{cases} \boldsymbol{V}_{(k)} & = \boldsymbol{X}_{(k)} - \frac{1}{L_p}\boldsymbol{\Phi}^\mathsf{T}(\boldsymbol{\Phi}\boldsymbol{X}_{(k)} - \boldsymbol{Y}) - \frac{\xi_{(\ell)}}{L_p}(\boldsymbol{X}_{(k)} - \mathbf{U}_{(\ell)}), \\ \boldsymbol{x}^n_{(k+1)} & = \arg\min_{\boldsymbol{x}\in\mathbb{R}^K} \frac{1}{2}||\boldsymbol{x} - \boldsymbol{v}^n_{(k)}||_2^2 + \frac{\lambda_1}{L_p}||\boldsymbol{x}||_1 + \frac{\lambda_2}{L_p}||\boldsymbol{D}\boldsymbol{x}||_1 + \frac{\lambda_3}{L_p}||\boldsymbol{x}||_2, \\ & \text{pour } n = 1, \ldots, N. \end{cases} \tag{4.30}$$

Une boucle extérieure (ℓ) est ajoutée afin de mettre à jour la variable $\boldsymbol{u}$. La minimisation sous contrainte de J_3 par rapport à $\boldsymbol{u}$ est un seuillage dur :

$$\boldsymbol{u}_{(\ell+1)} = \max(0, \boldsymbol{x}^*), \tag{4.31}$$

où $\boldsymbol{x}^*$ est la valeur de $\boldsymbol{x}_{(k)}$ à la dernière itération k. Le poids de la pénalité quadratique ξ doit être augmenté à chaque itération ℓ pour garantir la convergence vers le minimum contraint de J_3. Le schéma le plus simple consiste à augmenter progressivement sa valeur en utilisant la règle linéaire par exemple : $\xi_{(\ell+1)} = \beta\xi_{(\ell)}$ avec $\beta > 1$ et $\xi_1 = 1$. Les étapes de cet algorithme sont résumées dans Algorithme 3.

4.3 Application au tri de déchets bois

Les méthodes présentées précédemment ont été implantées en langage Matlab. Afin de simplifier leur utilisation, une interface graphique a été développée (cf. §1.8.2.1). L'application visée est la sélection de variables en spectroscopie NIR afin d'optimiser la classification de déchets de bois en conditions industrielles[3]. Nous espérons que le fait d'exploiter conjointement toutes les réponses spectrales pour sélectionner un ensemble de longueurs d'onde caractéristiques nous permettra d'éviter les problèmes de sur-apprentissage. Les déchets doivent être triés dans deux grandes catégories : les bois recyclables et les bois indésirables. Chaque catégorie est composée d'un certain nombre de groupes donnés dans le tableau 4.1. La base utilisée ici est composée de 290 échantillons de déchets bois étiquetés par des experts dans 11 groupes distincts.

4.3.1 Acquisition des données

Les spectres NIR sont mesurés par un spectromètre Nicolet 8700 FTIR purgé par de l'azote liquide. Les reflectances sont mesurées dans l'intervalle 3562–10000 cm^{-1} (correspondant à 2.8–1 μm) avec une résolution de 16 cm^{-1} et un pas spectral de 4 cm^{-1}. Chaque spectre, composé de 1647 longueurs d'onde, est ensuite traité afin de supprimer la ligne de base [Mazet05] et de normaliser les intensités. Quelques uns des spectres des deux catégories sont représentés dans la figure 4.3.

Les spectres pré-traités sont rangés dans les colonnes de la matrice de données $\boldsymbol{Y} \in \mathbb{R}^{1647\times 294}$. Les spectres correspondants aux bois recyclables (classe 1) sont rangés dans les premières colonnes et ceux des bois indésirables dans les dernières colonnes.

[3]Le système de tri intégrant le spectromètre et l'algorithme de classification sera installé en tête de ligne de recyclage. Le temps d'acquisition et de traitement doit être de quelques dizaines de micro-secondes par spectre. La sélection de variables est une étape indispensable pour réduire les coûts de calcul.

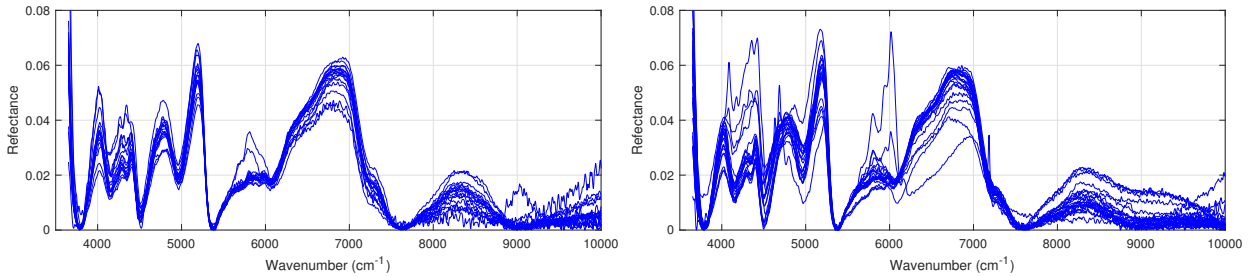
(a) Classe 1 : bois à recycler

Groupe	Nature	Nombre
1.1	bois brut	32
1.2	bois massif paint	36
1.3	bois massif avec finition	35
1.4	contreplaqué brut	16
1.5	contreplaqué avec finition	16
1.6	panneau de particules brut	28
1.7	panneau de particules paint	6

(b) Classe 2 : bois à rejeter

Groupe	Nature	Nombre
2.1	bois traité par sels métalliques	35
2.2	panneau MDF/HDF brut	28
2.3	panneau MDF/HDF peint	50
2.4	panneau de fibre dure	8

TABLE 4.1 : Composition des deux catégories de déchets de bois



(a) Classe 1 : bois à recycler

(b) Classe 2 : bois à rejeter

FIGURE 4.3 : Quelques spectres NIR des deux classes de déchets bois

4.3.2 Construction du dictionnaire

Les colonnes du dictionnaire $\boldsymbol{\Phi}$ sont des fonctions gaussiennes normalisées avec des centres $c_i \in [3600, 10000]$ cm^{-1} et des largeurs $\sigma_j \in [30, 600]$ cm^{-1}, de façon à couvrir uniformément l'intervalle des nombres d'onde observés. La discrétisation des largeurs a donné lieu à 20 valeurs différentes de σ_j ($\{\sigma_j = 30j\}_{j=1}^{20}$). Pour chaque valeur de σ_j, les centres sont choisis de telle sorte que deux gaussiennes consécutives de même largeur soient distantes de σ_j. Ainsi, on obtient $N = 778$ atomes dans le dictionnaire ; chacun étant normalisé ($||\boldsymbol{\phi}_n|| = 1, n = 1, \ldots, N$).

4.3.3 Sélection de variables

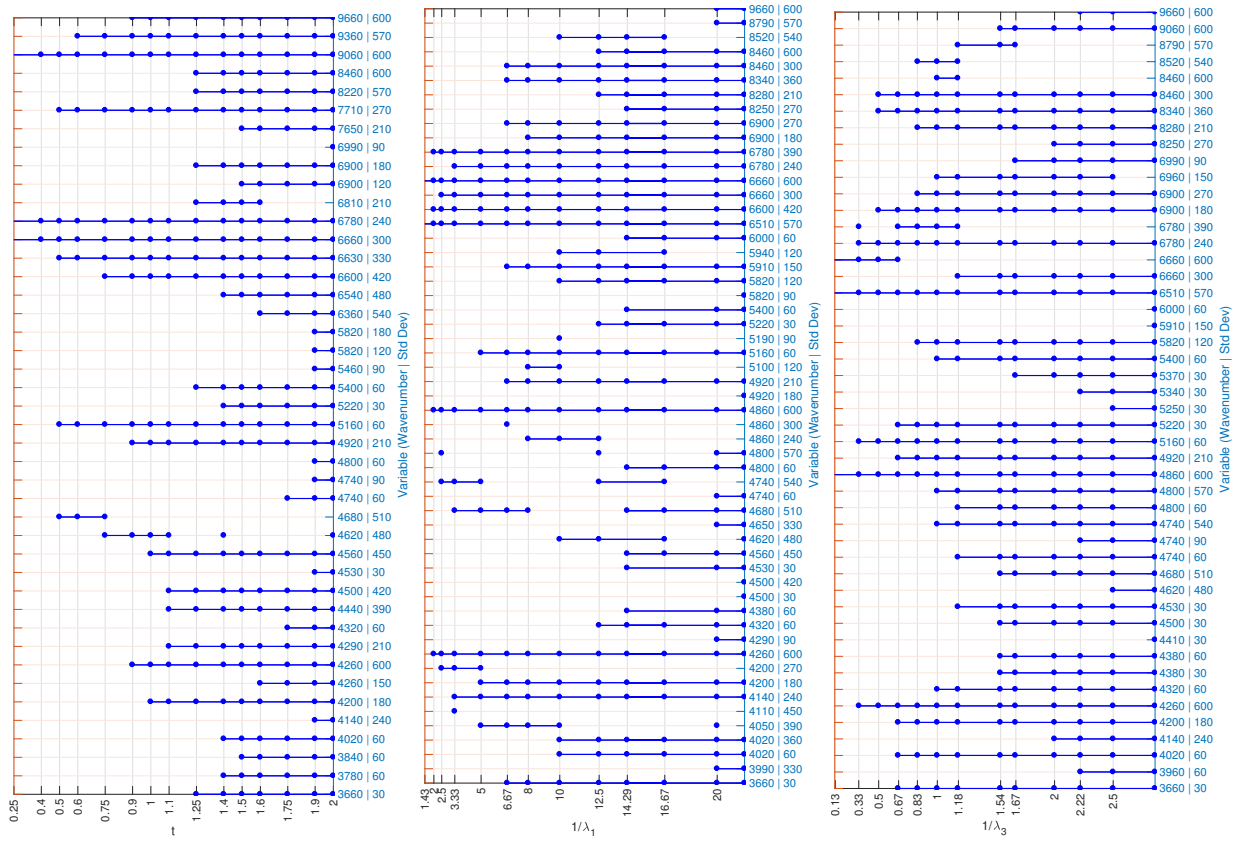
Les algorithmes FSL, FSGL et NN-FSGL sont comparés à la méthode SVS [Turlach05] pour la sélection simultanée de variables. Cette dernière est une extension de la méthodologie Lasso au problème dans lequel plusieurs réponses corrélées sont disponibles. Le problème d'optimisation correspondant s'écrit :

$$\underset{\boldsymbol{X}}{\text{minimiser}} \ \tfrac{1}{2}||\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{X}||_F^2, \tag{4.32a}$$

$$\text{s. c.} \quad \sum_{n=1}^{N} ||\boldsymbol{x}^n||_\infty \leq t, \tag{4.32b}$$

où t est un paramètre qui contrôle la parcimonie de la solution. Ce problème est résolu par une méthode de point intérieur implémentée en langage C[4].

[4]Merci à Berwin Turlach pour les codes sources de la méthode SVS et le Makefile qui a simplifié la compilation.



(a) SVS [Turlach05] (b) FSL ($\lambda_2 = 0.5$) (c) FSGL ($\lambda_1 = 0, \lambda_2 = 0.5$)

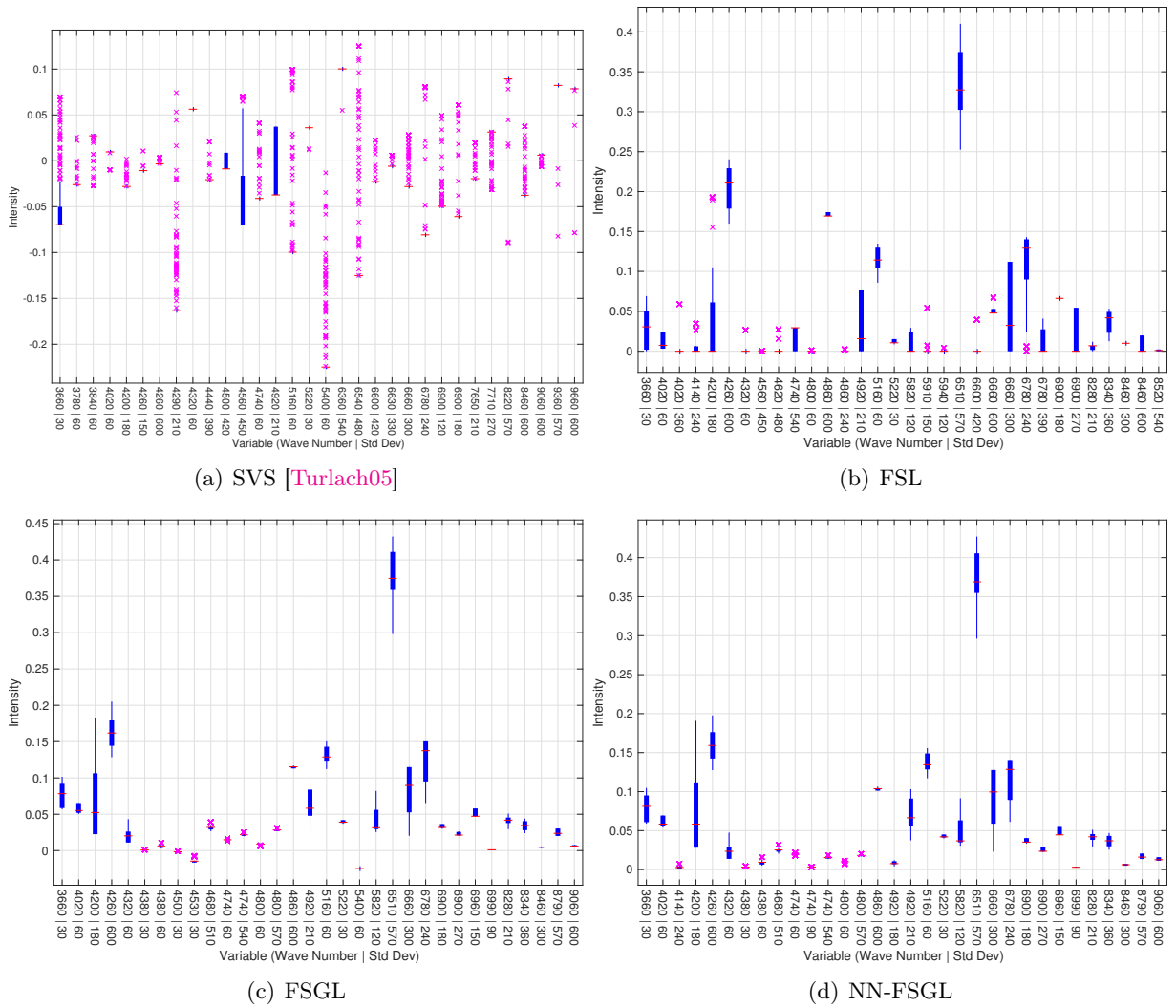
FIGURE 4.4 : Variables sélectionnées en utilisant l'ensemble de la base de données

La figure 4.4 présente les variables sélectionnées avec les méthodes SVS, FSL et FSGL pour différentes valeurs des paramètres t, $1/\lambda_1$ et $1/\lambda_3$, respectivement. Les lignes horizontales relient les valeurs des paramètres pour lesquels les coefficients restent non nuls. Lorsque les paramètres précédents sont faibles, les variables sélectionnées correspondent principalement à la zone spectrale 6600–6700 cm^{-1} où des pics larges et intenses peuvent être observés (cf. figure 4.3). En augmentant la valeur des paramètres, d'autres variables sont choisies. Pour une quarantaine d'atomes, SVS partage 28 variables communes avec FSL et FSGL. Ces dernières méthodes partagent 38 variables communes.

À titre d'exemple, pour sélectionner 32 variables en utilisant l'ensemble des spectres de la base de données, on peut fixer les paramètres de la façon suivante :

$$\begin{array}{lcl} \text{SVS} & : & t = 1.75 \\ \text{FSL} & : & \lambda_1 = 0.075, \lambda_2 = 0.5 \\ \text{FSGL} & : & \lambda_1 = 0, \lambda_2 = 0.5, \lambda_3 = 0.625 \\ \text{NN-FSGL} & : & \lambda_1 = 0, \lambda_2 = 0.5, \lambda_3 = 0.6, \beta = 1.1 \end{array}$$

avec $maxiter = 1000$. NN-FSGL est initialisé avec la solution obtenue avec FSGL puis en fixant $maxiter = 100$ et $nniter = 100$. La figure 4.5 présente le digramme de dispersion des coefficients non nuls. À noter que la méthode SVS nécessite une normalisation différente des données et des éléments du dictionnaire : moyenne nulle et variance unité. En outre, aucune contrainte n'est imposée



(a) SVS [Turlach05]

(b) FSL

(c) FSGL

(d) NN-FSGL

FIGURE 4.5 : Diagrammes de dispersion des 32 variables sélectionnées en utilisant tous les spectres de la base de données

aux coefficients associés à la même variable, ce qui explique la différence de dispersion avec les autres méthodes. Globalement, les intensités relatives renvoyées par les algorithmes proposés ont une dispersion plus faible grâce à la pénalité de fusion. Sur les 32 variables, l'algorithme FSGL partage 14 variables communes avec SVS, 19 avec FSL et 29 avec NN-FSGL.

4.3.4 Classification des déchets bois

Les 32 atomes sélectionnés précédemment sont utilisés pour estimer, par moindres carrés, les coefficients restreints au support des variables actives. Ces coefficients sont ensuite passés dans un classifieur SVM à noyau quadratique. Les résultats obtenus sont comparés à :

- SVM : c'est l'algorithme SVM classique (la classification est réalisée sur les spectres entiers, sans sélection de variables) ;
- G-SVM : cet algorithme est proposé dans [Flamary14]. Il consiste à résoudre un critère SVM

Sélection de variables	Classifieur	Paramètres			$\lvert\Gamma\rvert$	Précision		
		λ_1	λ_2	λ_3		globale	classe 1	classe 2
–	SVM	–	–	–	–	82.2%	80.4%	84.8%
G-SVM [Flamary14]		$C = 150$	–	–	32	76.9%	81.4%	71.3%
SVS [Turlach05]	SVM	$t = 1.75$	–	–	32	84.1%	83.1%	85.6%
FSL	SVM	0.075	0.5	–	32	85.9%	85.6%	86.3%
FSGL	SVM	0	0.5	0.625	32	**87.8%**	**85.9%**	**90.3%**
		0.04	0.2	0.295	32	86.5%	86.4%	86.7%
NN-FSGL	SVM	0	0.5	0.6	32	86.9%	85.5%	88.8%

TABLE 4.2 : Précision de la classification des déchets bois

augmenté où une contrainte de parcimonie (norme ℓ_1) est imposée aux vecteurs de support. Le nombre de coefficients non nuls est contrôlé par le paramètre C[5]. Ici, $C = 150$, ce qui permet une prise de décision sur 32 coefficients.

Les résultats de classification en validation croisée sont résumés dans le tableau 4.2. En terme de précision globale, les méthodes FSL et FSGL sont nettement meilleures que celles développées dans [Turlach05, Flamary14]. En outre, la sélection de variables contribue à la généralisation de la règle apprise comme en témoigne les résultats obtenus avec le classifieur SVM appliqué directement aux spectres originaux avec 1647 variables. Le meilleur résultat est de 88% de bonne classification globale avec FSGL combiné au classifieur SVM.

4.3.5 Influence des paramètres

Le but de ce paragraphe est d'étudier l'influence des paramètres de régularisation sur l'ensemble des variables sélectionnées et, par conséquent, sur le résultat de classification. Pour $\lambda_2 = 0.2$ et $\lambda_3 = 0$, le taux d'erreur de classification et la cardinalité du support sont représentés dans la figure 4.6(a) pour plusieurs valeurs de λ_1. Le taux d'erreur décroit progressivement lorsque λ_1 augmente entre 10^{-2} et 0.18 pour atteindre un minimum de 14% pour une trentaine de variables sélectionnées. Les performances se dégradent rapidement quand $\lambda_1 > 0.3$, ce qui correspond à moins de 20 variables retenues.

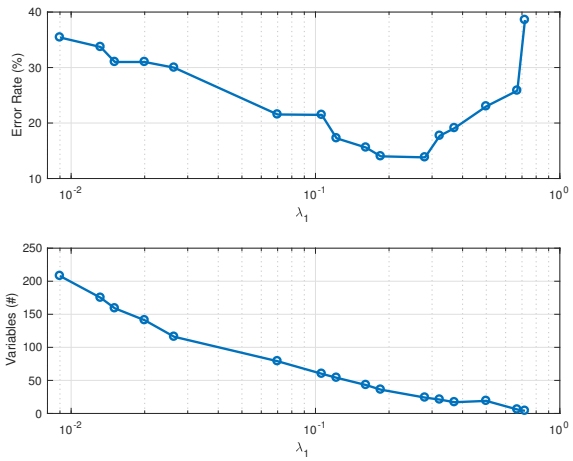
Pour $\lambda_2 = 0.5$ et $\lambda_1 = 0$, les résultats sont représentés sur la figure 4.6(b) pour différentes valeurs de λ_3. Le taux minimum d'erreur (inférieur à 15%) est ici obtenu pour $\lambda_3 \in [0.1, 1.5]$. En particulier, $\lambda_3 = 0.5$ donne le meilleur taux avec 12% d'erreur.

L'influence du paramètre de fusion λ_2 est représentée sur la figure 4.6(c). Ici, $\lambda_1 = 0$ et λ_3 est choisi de façon à obtenir 40 variables. L'erreur de classification est inférieure à 15% dans l'intervalle $\lambda_2 \in [0.1, 0.6]$. L'erreur minimale (11%) est atteinte pour $\lambda_2 = 0.55$.

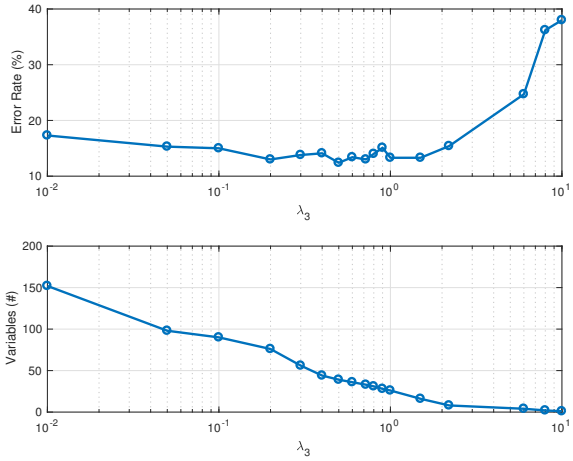
4.4 Conclusion

J'ai présenté dans ce chapitre une approche pour la sélection de variables appliquée à la spectroscopie proche infrarouge. Les spectres sont d'abord modélisés comme une combinaison linéaire de

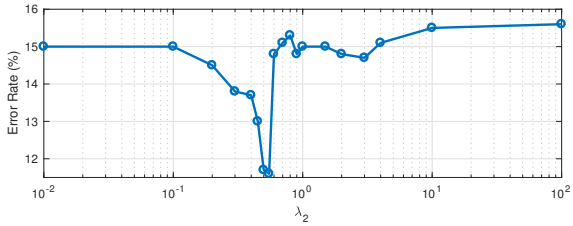
[5] http://remi.flamary.com/soft/soft-gsvm.html



(a) Paramètre de parcimonie ($\lambda_2 = 0.2, \lambda_3 = 0$)



(b) Paramètre de groupe ($\lambda_1 = 0, \lambda_2 = 0.5$)



(c) Paramètre de fusion ($\lambda_1 = 0$, λ_3 est fixé tel que $|\Gamma| = 40$)

FIGURE 4.6 : Taux d'erreur de classification en fonction des paramètres de régularisation

quelques variables explicatives tirées d'une collection de formes élémentaires (dictionnaire). L'objectif est alors de sélectionner un sous-ensemble de variables partagées par tous les signaux d'apprentissage en utilisant une décomposition parcimonieuse simultanée.

La première contribution de ce travail est la contrainte de constance par morceaux imposée aux lignes de la matrice des coefficients de la décomposition $\boldsymbol{X}$. Celle-ci a pour effet de réduire la variance intraclasse. La deuxième contribution est la proposition d'un schéma de résolution numériquement efficace. En effet, à chaque itération des algorithmes proposés, la matrice $\boldsymbol{X}$ est calculée ligne par ligne de façon indépendante. Le développement de versions de calcul parallèle de ces algorithmes s'en trouve donc grandement simplifiée. L'application au problème de tri de déchets de bois a montré l'intérêt de la méthode de sélection en classification bi-classe.

La méthodologie proposée ici consiste à rechercher les variables qui contribuent le plus à la synthèse des signaux. Par conséquent, les variables sélectionnées ne seront pas forcément celles qui permettent la meilleure discrimination des classes. Une approche alternative serait d'intégrer dans le critère un terme favorisant la séparabilité. Cette idée est pour l'instant reléguée au chapitre des perspectives (§5.1.2).

Troisième partie

Perspectives

Chapitre 5

Projet scientifique

Je présente dans ce chapitre mes principaux projets de recherche qui s'inspirent des développements déjà réalisés et des résultats obtenus. Pour chaque thème abordé, je m'efforcerai de décliner mes perspectives à court et à moyen termes. À plus long terme, je pense m'impliquer davantage dans les applications liées à la biologie en collaboration avec des collègues automaticiens (Magalie Thomassin, Thierry Bastogne, Alain Richard) et biologistes (Muriel Barberi-Heyob, Béatrice Faivre, Noémie Thomas, Hélène Dumond).

5.1 Développement d'algorithmes d'approximation parcimonieuse

5.1.1 Algorithmes $\ell_2 - \ell_0$ avec contrainte de positivité

La restauration de signaux sous contrainte de positivité est un problème que l'on rencontre naturellement dans plusieurs applications comme la spectroscopie, la microscopie optique, la factorisation en matrices non-négatives, etc. Dans le cadre de l'approximation parcimonieuse, cette contrainte est relativement facile à prendre en compte lorsque la parcimonie est induite par la norme ℓ_1 :

$$\min_{\boldsymbol{x}} \left\{ \tfrac{1}{2} ||\boldsymbol{y} - \boldsymbol{\Phi x}||_2^2 + \lambda_1 ||\boldsymbol{x}||_1 \right\} \quad \text{s.c.} \quad \boldsymbol{x} \geq \boldsymbol{0}. \tag{5.1}$$

Plusieurs méthodes sont proposées dans la littérature pour résoudre ce type de problème en utilisant, par exemple, des algorithmes de seuillage itératif ou l'algorithme ADMM. Dans un critère $\ell_2 - \ell_0$ ayant l'une des formes :

$$\min_{\boldsymbol{x} \geq \boldsymbol{0}} \left\{ \tfrac{1}{2} ||\boldsymbol{y} - \boldsymbol{\Phi x}||_2^2 \right\} \quad \text{s.c.} \quad ||\boldsymbol{x}||_0 \leq s \tag{5.2}$$

$$\min_{\boldsymbol{x} \geq \boldsymbol{0}} \left\{ \tfrac{1}{2} ||\boldsymbol{y} - \boldsymbol{\Phi x}||_2^2 + \lambda ||\boldsymbol{x}||_0 \right\}, \tag{5.3}$$

la prise en compte de la positivité dans les algorithmes gloutons – classiquement utilisés pour résoudre ce problème – est plus délicate et coûteuse numériquement. Par exemple, un algorithme OMP avec contrainte de non-négativité a été proposé dans [Bruckstein08]. L'implémentation standard de cet algorithme (NNOMP) consiste à remplacer la projection orthogonale dans chaque itération de OMP par la résolution d'un problème de moindres carrés non-négatif. Une implémentation plus rapide mais moins précise de cet algorithme a été proposée dans [Yaghoobi15]. Dans le cadre de la thèse

de Thi-Thanh NGUYEN, nous avons proposé une version « exacte » et plus efficace dans laquelle les sous-problèmes de moindres carrés sont résolus, récursivement, par l'algorithme des contraintes actives de Nocedal et Wright [Nocedal06, ch. 16] qui autorise un démarrage à chaud.

À court terme, je poursuivrai ces travaux pour étendre le résultat obtenu avec OMP à d'autres algorithmes gloutons unidirectionnels tels que MP [Mallat93] et OLS [Blumensath07] et bidirectionnels comme Bayesian-OMP [Herzet10] et SBR [Soussen11]. Ces derniers conduisent généralement à une meilleure sélection des atomes [Soussen15].

À moyen terme, il serait intéressant de généraliser les algorithmes développés aux problèmes de parcimonie structurée avec contrainte de positivité. Un instance particulière est la parcimonie simultanée qui a des applications en analyse spectrale ou en sélection de variables.

5.1.2 Apprentissage de dictionnaire

La performance des algorithmes d'approximation parcimonieuse en termes d'erreur d'approximation et de parcimonie de la solution dépend non seulement du signal analysé mais également du dictionnaire utilisé. De manière générale, ce dernier peut être choisi de deux façons différentes [Rubinstein10a] : (1) construction des atomes sur la base du modèle mathématique ayant engendré les données ou (2) apprentissage du dictionnaire en utilisant des données d'entraînement. Depuis le travail de Olshausen et Field [Olshausen97], plusieurs algorithmes d'apprentissage ont été proposés. Si les données d'entraînement sont rangées dans le colonne de la matrice $\boldsymbol{Y} \in \mathbb{R}^{M\times K}$, le problème d'optimisation associé peut être exprimé par l'une des deux formes génériques suivantes :

$$(\boldsymbol{\Phi}^*, \boldsymbol{X}^*) = \arg\min_{\boldsymbol{\Phi},\boldsymbol{X}} \left\{ \tfrac{1}{2}||\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{X}||_F^2 \right\} \quad \text{s.c.} \quad ||\boldsymbol{x}_k||_0 \leq s, \forall k, \tag{5.4}$$

$$(\boldsymbol{\Phi}^*, \boldsymbol{X}^*) = \arg\min_{\boldsymbol{\Phi},\boldsymbol{X}} \left\{ \tfrac{1}{2}||\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{X}||_F^2 + \lambda||\boldsymbol{X}||_{1,1} \right\}. \tag{5.5}$$

L'optimisation par rapport à $\boldsymbol{\Phi}$ est un problème délicat pour lequel il n'y a pas de solution analytique. La stratégie généralement utilisée consiste à alterner entre l'approximation parcimonieuse et la mise à jour du dictionnaire. C'est un schéma qui aboutit à une bonne approximation de $\boldsymbol{Y}$ avec un dictionnaire générique qui n'a pas de forme particulière.

À moyen terme, il serait intéressant d'étudier des schémas d'apprentissage dans lesquels chaque atome du dictionnaire est décrit par un modèle paramétrique (par exemple, une fonction gaussienne pour les signaux de spectroscopie intrarouge, une ondelette ou tout autre fonction de forme connue). Un première possibilité serait d'utiliser une technique similaire à celle proposée dans [Rubinstein10b] dans laquelle la dictionnaire prend la forme $\boldsymbol{\Phi} = \boldsymbol{D}\boldsymbol{A}$ où $\boldsymbol{D}$ est un dictionnaire de base fixé qui regroupe les fonctions imposées et $\boldsymbol{A}$ une matrice parcimonieuse. Le problème à résoudre peut alors s'écrire :

$$(\boldsymbol{A}^*, \boldsymbol{X}^*) = \arg\min_{\boldsymbol{A},\boldsymbol{X}} \left\{ \tfrac{1}{2}||\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{A}\boldsymbol{X}||_F^2 \right\} \quad \text{s.c.} \quad \begin{aligned} &||\boldsymbol{x}_k||_0 \leq s, \forall k, \\ &||\boldsymbol{a}_n||_0 \leq t, \forall n. \end{aligned} \tag{5.6}$$

Il également possible d'ajouter au critère un terme qui favorise l'apprentissage d'un dictionnaire contenant des atomes mutuellement incohérents. Cette propriété est souvent recherchée lorsque les

algorithmes d'approximation parcimonieuses sont utilisés pour révéler une structure ou une agrégation dans les données [Barchiesi13]. Les applications visées vont de la sélection de variables à la factorisation en matrices non-négatives.

Les méthodes d'apprentissage du dictionnaire capturent le plus souvent les composantes les plus significatives des signaux d'apprentissage pour en obtenir une bonne reconstruction. Dans le contexte de la classification, il n'est pas garanti que le dictionnaire soit optimal pour la discrimination des classes [Mairal08].

À long terme, il serait donc intéressant d'ajouter au critère un terme qui favorise cette séparabilité :

$$(\boldsymbol{\Phi}^*, \boldsymbol{X}^*) = \arg\min_{\boldsymbol{\Phi},\boldsymbol{X}} \left\{ \tfrac{1}{2}||\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{X}||_F^2 + \lambda||\boldsymbol{X}||_{1,1} + \mu C(\boldsymbol{X}, \boldsymbol{\Phi}, \theta) \right\}. \tag{5.7}$$

où $C(\boldsymbol{X}, \boldsymbol{\Phi}, \theta)$ est une fonction discriminante qui dépend des paramètres θ du modèle de classification. La fonction $C(\boldsymbol{X}, \boldsymbol{\Phi}, \theta)$ interviendra à la fois dans l'estimation de $\boldsymbol{X}$ et la mise à jour du dictionnaire de sorte que le dictionnaire appris sera fortement dépendant de la méthode de classification choisie.

5.1.3 Décomposition conjointe de signaux en motifs élémentaires

La décomposition d'un signal en motifs paramétriques constitue une généralisation du problème de déconvolution impulsionnelle. Par exemple, en spectroscopie de résonance magnétique nucléaire ou de photo-électrons, les signaux peuvent être vus comme une superposition de motifs élémentaires ayant une forme lorentzienne ou gaussienne. Le problème consiste alors à déterminer l'amplitude, la position et le paramètre de forme de chacun des motifs. Dans la thèse de Souleymen Sahnoun [Sahnoun12a], la solution proposée revient à construire un dictionnaire en discrétisant les positions et les largeurs des motifs.

Dans le cadre du projet PEPS SpectroDec (2012-2013) puis de l'ANR DSIM (2015-2018), portés par Vincent Mazet du laboratoire ICube, le problème concerne la décomposition conjointe sur le même dictionnaire d'une séquence de spectres acquis à des temps différents. La décomposition conjointe est codée par la notion de voisinage : deux spectres voisins temporellement doivent avoir une décomposition similaire (les amplitudes, les positions et les formes évoluent lentement comme le montre la figure 5.1). Lorsque les motifs dans deux spectres voisins se chevauchent fortement, la première solution envisagée consiste à minimiser la fonction de coût :

$$\sum_{s=1}^{S} ||\boldsymbol{y}_s - \boldsymbol{\Phi}\boldsymbol{x}_s||^2 + \mu \sum_{s=1}^{S-1} ||\boldsymbol{\Phi}\boldsymbol{x}_{s+1} - \boldsymbol{\Phi}\boldsymbol{x}_1||^2 + \lambda||\boldsymbol{x}_s||_0 \tag{5.8}$$

où $\boldsymbol{y}_s$ est un spectre ($s = 1, \ldots, S$). Dans le cas contraire, nous avons proposé une approche plus complexe qui revient à remplacer le deuxième terme de l'équation précédente par la distance (discrète) de Hausdorff entre les supports des solutions $\boldsymbol{x}_s$ et $\boldsymbol{x}_{s+1}$ [Mazet13].

À court terme, il serait intéressant de comparer cette méthode avec les approches basées sur la parcimonie sociale [Kowalski13] qui reposent sur une norme mixte permettant de favoriser la sélection d'atomes connexes entre deux ou plusieurs spectres consécutifs.

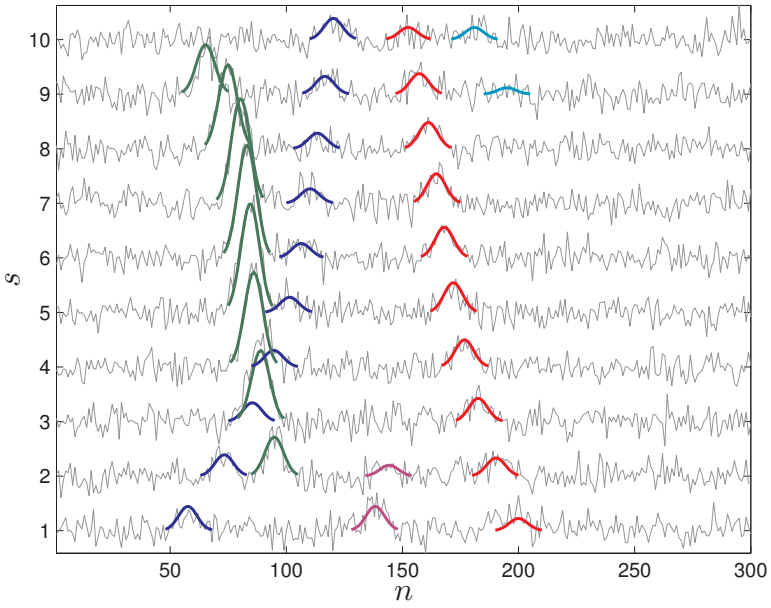


FIGURE 5.1 : Exemple de séquence simulée de 10 spectres de photo-électrons de 300 échantillons et 5 trajectoires. Chaque spectre est composé de motifs gaussiens de paramètres inconnus

5.2 Traitement d'images hyperspectrales

Equipés de capteurs optiques avancées, les spectro-imageurs modernes permettent d'acquérir des images hyperspectrales à hautes résolutions spectrale et spatiale. Ces images $\boldsymbol{\mathcal{I}}(\lambda, x, y)$ se présentent comme des cubes de données représentant l'intensité de la lumière transmise ou reflétée dans une scène à une longueur d'onde λ et à la position (x, y). Comme les éléments chimiques ont des signatures spectrales différentes, les données hyperspectrales hautement résolues permettent d'obtenir des informations plus précises sur les matériaux observés que ne peut fournir une simple image à trois couleurs. Les domaines d'application étant variés (télédétection, astronomie, microscopie de fluorescence, etc.), l'analyse d'images hyperspectrales a suscité un immense intérêt de la communauté du traitement du signal et des images. En particulier, de nombreux auteurs se sont intéressés à la classification [Camps-Valls14], à la déconvolution [Henrot13, Song16] et au démélange d'images [Bioucas-Dias12, Dobigeon14].

La dimensionnalité des images hyperspectrales empêche les algorithmes déjà développés à produire rapidement des résultats précis [Willett14]. Ce constat nous a conduit dans le cadre de la thèse de Yingying SONG à poser le problème de restauration d'images hyperspectrales comme un problème de déconvolution adaptative. En effet, sur certains spectro-imageurs industriels rapides (*pushbroom*), les objets à scanner défilent sous l'imageur. Les images hyperspectrales sont acquises tranche par tranche, chaque tranche $\boldsymbol{\mathcal{I}}(\lambda, x, y_i)$ est mesurée à une valeur y_i fixée et assimilable au temps. Nous développons actuellement des algorithmes inspirés de LMS avec des contraintes de régularité (spectrale et/ou spatiale) et de positivité.

À court terme, je compte poursuivre ces travaux dans lesquels plusieurs problèmes sont encore posés, en particulier, le réglage du pas du gradient pour éviter les instabilités et le calcul de la précision asymptotique des images restaurées.

5.3 Modèles et outils pour la caractérisation de la progression tumorale *in vivo*

Depuis les années 1960, différentes approches ont été adoptées pour construire des modèles mathématiques permettant de décrire la croissance tumorale avec et sans traitement. Ces modèles peuvent être classés selon leur échelle (microscopique ou macroscopique). À l'échelle cellulaire, les modèles basés sur les automates cellulaires [Gerlee09] ou les modèles d'agents [Anderson09] sont bien adaptés à l'étude des interactions cellulaires et leurs implications sur une population. Cependant, ces modèles discrets ne peuvent en général être implémentés que pour décrire un volume maximal de l'ordre du mm^3 représentant quelques millions de cellules [Benzekry14]. Les modèles continus quant à eux permettent de décrire le comportement moyen, au niveau tissulaire, d'un amas de cellules cancéreuses en utilisant des équations aux dérivées partielles : système à réaction-diffusion [Gatenby96] ou mécanique des milieux continus [Bresch10]. Ces modèles permettent de prendre en compte les interactions avec l'environnement comme l'apport en oxygène/nutriments et l'élasticité des membranes [Lagaert11]. Enfin, des modèles macroscopiques, basés sur des équations différentielles ordinaires, liant le volume tumoral à celui des cellules endothéliales du réseau vasculaire ont également été développés [Hahnfeldt99, d'Onofrio09].

Les modèles de croissance tumorale ont permis une meilleure compréhension de certains phénomènes biologiques et, dans certains cas, l'effet d'un traitement direct ou indirect de la tumeur (chimiothérapie, rayonnements ionisants, traitement anti-angiogénique) sur cette croissance. Les modèles peuvent également servir à la simulation, à la prédiction ou encore à l'amélioration des protocoles thérapeutiques. La principale difficulté à la mise en oeuvre d'un modèle dans une situation réelle est liée au choix de la structure de modèle (en fonction de la tumeur étudiée) et à l'**estimation des paramètres et coefficients** intervenant dans le modèle.

Depuis 2010, je travaille en collaboration avec des biologistes de l'ancien laboratoire Sigreto (Signalisation, Génomique et Recherche Translationnelle en Oncologie, EA 4421), intégré au CRAN depuis janvier 2013 dans le département SBS (Santé, Biologie, Signal). Le but de ce département est de favoriser l'interaction de chercheurs en biologie, ingénierie pour la santé et traitement du signal/image pour le développement de méthodes permettant l'analyse et la compréhension du cancer. Jusqu'à présent, ma collaboration avec Béatrice Faivre, Noémie Thomas et Hélène Dumond porte sur le développement d'un logiciel de traitement d'images de microscopie intravitale pour la quantification d'un réseau vasculaire (figure 5.2). L'objectif est de caractériser le processus d'angiogénèse dans le cas du glioblastome [Alaoui-Lasmaili17] et de cellules tumorales mammaires [Thiebaut17]. Les grandeurs déterminées à partir des images (volume tumoral, densité vasculaire, surface vasculaire, etc.) ont également servi à l'identification des coefficients d'un modèle « boîte noire » pour décrire l'effet d'un traitement anti-angiogénique sur la croissance tumorale [Tylcz15].

À moyen et long termes, il serait intéressant de poursuivre ces travaux en intégrant d'autres modalités non invasives d'acquisition de données *in vivo* comme l'IRM et la spectroscopie par résonance magnétique nucléaire (RMN). Les séquences d'IRM hautement résolues spatialement et temporellement permettent d'obtenir des données extrêmement précises avec différents niveaux d'informations : morphologiques, fonctionnelles et moléculaires. La spectroscopie RMN est un outil de quantification des grands métabolites marqueurs du métabolisme tumoral [Toussaint16]. Les méthodes de traitement du signal de l'image seront alors indispensables pour, d'une part, **extraire les**

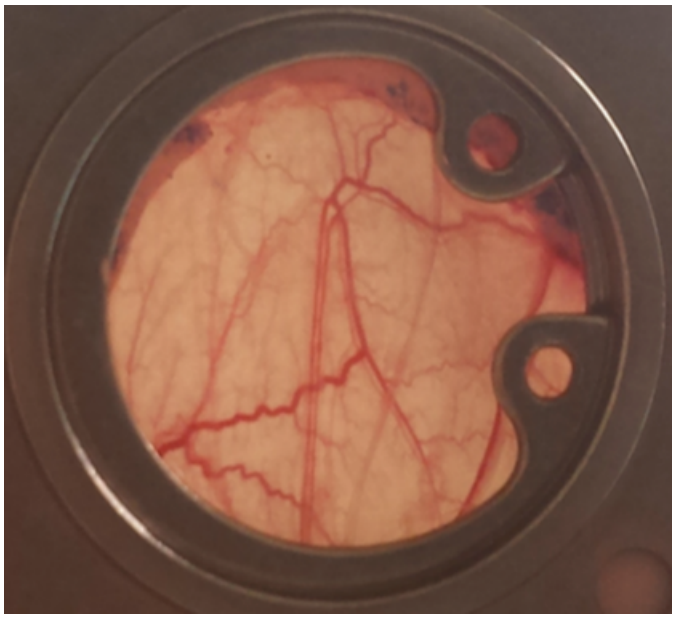


FIGURE 5.2 : Chambre dorsale implantée sur une souris immunodéficiente *nude*. Elle permet la xénogreffe de tumeurs sous forme de cellules et l'observation, par microscope, de l'évolution de la tumeur et du réseau vasculaire

informations d'intérêt à partir des données et, d'autre part, **résoudre des problèmes inverses** pour estimer les paramètres de modèles de croissance tumorale. Ces modèles pourront ensuite être exploités dans des contextes pré-clinique ou clinique. Pour prendre en compte la variabilité des réponses aux traitements selon les individus, je pense en particulier m'appuyer sur les **approches de population**[1] (modèles à effets mixtes) pour estimer la distributions des paramètres du modèle sur plusieurs réponses.

[1]La thèse de Levy BATISTA qui se termine fin 2017 porte justement sur le développement de méthodes d'identification de populations de systèmes dynamiques.

Bibliographie

[Adams95] Adams M.J. *Chemometrics in analytical spectroscopy.* Royal Society of Chemistry, Cambridge, UK, 1995.

[Alaoui-Lasmaili17] Alaoui-Lasmaili K.E., Djermoune E.H., Tylcz J.B., Meng D., Plénat F., Thomas N., and Faivre B. A new algorithm for a better characterization and timing of the anti-VEGF vascular effect named "normalization". *Angiogenesis*, available online http://link.springer.com/article/10.1007/s10456-016-9536-3, 2017.

[Anderson09] Anderson A., Rejniak K., Gerlee P., and Quaranta V. Microenvironment driven invasion : A multiscale multimodel investigation. *Journal of Mathematical Biology*, 58(4) :579–624, 2009.

[Barchiesi13] Barchiesi D. and Plumbley M.D. Learning incoherent dictionaries for sparse approximation using iterative projections and rotations. *IEEE Transactions on Signal Processing*, 61(8) :2055–2065, 2013.

[Beck09] Beck A. and Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1) :183–202, 2009.

[Benzekry14] Benzekry S., Lamont C., Beheshti A., Tracz A., Ebos J., Hlatky L., and Hahnfeldt P. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Computational Biology*, 10(8) :e1003800, 2014.

[Bioucas-Dias12] Bioucas-Dias J.M., Plaza A., Dobigeon N., Parente M., Du Q., Gader P., and Chanussot J. Hyperspectral unmixing overview : Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2) :354–379, 2012.

[Blumensath07] Blumensath T. and Davies M.E. On the difference between orthogonal matching pursuit and orthogonal least squares. Technical report, http://www.personal.soton.ac.uk/tb1m08/papers/BDOMPvsOLS07.pdf, 2007.

[Boyd11] Boyd S., Parikh N., Chu E., Peleato B., and Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1) :1–122, 2011.

[Bresch10] Bresch D., Colin T., Grenier E., Ribba B., and Saut O. Computational modeling of solid tumor growth : The avascular stage. *SIAM Journal on Scientific Computing*, 32(4) :2321–2344, 2010.

[Bresler86] Bresler Y. and Makovski A. Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5) :1081–1089, 1986.

[Bruckstein08] Bruckstein A.M., Elad M., and Zibulevsky M. On the uniqueness of nonnegative sparse solutions to underdetermined systems of equation. *IEEE Transactions on Information Theory*, 54(11) :4813–4820, 2008.

[Camps-Valls14] Camps-Valls G., Tuia D., Bruzzone L., and Benediktsson J.A. Advances in hyperspectral image classification : Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine*, 31(1) :45–54, 2014.

[Candès14] Candès E. and Fernandez-Grenada C. Compressed sensing off the grid. *Communications on Pure and Applied Mathematics*, 67(6) :906–956, 2014.

[Cariou08] Cariou C. and Chehdi K. Automatic georeferencing of airborne pushbroom scanner images with missing ancillary data using mutual information. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5) :1290–1300, 2008.

[Chen06] Chen J. and Huo X. Theoretical results on sparse representations of multiple-measurement vectors. *IEEE Transactions on Signal Processing*, 54(12) :4634–4643, 2006.

[Chen14] Chen Y. and Chi Y. Compressed sensing off the grid. *IEEE Transactions on Information Theory*, 60(10) :6576–6601, 2014.

[Comon14] Comon P. Tensors : A brief introduction. *IEEE Signal Processing Magazine*, 31(3), 2014. Special issue on BSS. hal-00923279.

[Cotter05] Cotter S.F., Rao B.D., Engan K., and Kreutz-Delgado K. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Transactions on Signal Processing*, 53(7) :2477–2488, 2005.

[Djermoune09a] Djermoune E.H., Thomassin M., and Tomczak M. First-order analysis of the mode and amplitude estimates of a damped sinusoid using Matrix Pencil. In *European Signal Processing Conference, EUSIPCO 2009*. Glasgow, Scotland, 2009.

[Djermoune09b] Djermoune E.H. and Tomczak M. Perturbation analysis of subspace-based methods in estimating a damped complex exponential. *IEEE Transactions on Signal Processing*, 57(11) :4558–4563, 2009.

[Dobigeon14] Dobigeon N., Tourneret J.Y., Richard C., Bermudez J.C.M., McLaughlin S., and Hero A.O. Nonlinear unmixing of hyperspectral images : Models and algorithms. *IEEE Signal Processing Magazine*, 31(1) :82–94, 2014.

[d'Onofrio09] d'Onofrio A. and Gandolfi A. A family of models of angiogenesis and anti-angiogenesis anti-cancer therapy. *Mathematical Medecine and Biology*, 26(1) :63–95, 2009.

[Ducasse95] Ducasse A., Mailhes C., and Castanie F. Amplitude and phase estimator study in Prony method for noisy exponential data. In *Proc. IEEE ICASSP*, pages 1796–1799. Detroit, MI, May 9-14 1995.

[Elmasry12] Elmasry G., Kamruzzaman M., Sun D.W., and Allen P. Principles and applications of hyperspectral imaging in quality evaluation of agro-food products : a review. *Critical Reviews in Food Science and Nutrition*, 52(11) :999–1023, 2012.

[Flamary14] Flamary R., Jrad N., Phlypo R., Congedo M., and Rakotomamonjy A. Mixed-norm regularization for brain decoding. *Computational and Mathematical Methods in Medicine*, 2014 :ID 317056, 2014.

[Foley98] Foley W.J., McIlwee A., Lawler I., Aragones L., Woolnough A.P., and Berding N. Ecological applications of near infrared reflectance spectroscopy–a tool for rapid, cost-effective prediction of the composition of plant and animal tissues and aspects of animal performance. *Oecologia*, 116(3) :293–305, 1998.

[Friedman07] Friedman J., Hastie T., Höfling H., and Tibshirani R. Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2) :302–332, 2007.

[Gatenby96] Gatenby R. and Gawlinski E. A reaction-diffusion model of cancer invasion. *Cancer Research*, 56(24) :5745–5753, 1996.

[Gerlee09] Gerlee P. and Anderson A. Evolution of cell motility in an individual-based model of tumor growth. *Journal of Theoretical Biology*, 259(1) :67–83, 2009.

[Gershman05] Gershman A.B. and Sidiropoulos N.D. *Space-time processing for MIMO communications*. Wiley Online Library, 2005.

[Gill98] Gill P.E., Murray W., and Saunders M.A. *User's guide for SNOPT 5.3 : A Fortran package for large-scale nonlinear programming*. Department of Mathematics, University of California, San Diego, USA, 1998.

[Goodwin99] Goodwin M. and Vetterli M. Matching pursuit and atomic signal models based on recursive filter banks. *IEEE Transactions on Signal Processing*, 47(7) :1890–1902, 1999.

[Haardt08] Haardt M., Roemer F., and Del Galdo G. Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems. *IEEE Transactions on Signal Processing*, 56(7) :3198–3213, 2008.

[Hahnfeldt99] Hahnfeldt P., Panigrahy D., Folkman J., and Hlatky L. Tumor development under angiogenic signaling : A dynamical theory of tumor growth, treatment

response, and postvascular dormancy. *Cancer Research*, 59(19) :4770–4775, 1999.

[Henrot13] Henrot S., Soussen C., and Brie D. Fast positive deconvolution of hyperspectral images. *IEEE Transactions on Image Processing*, 22(2) :828–833, 2013.

[Herzet10] Herzet C. and Drémaud A. Bayesian pursuit algorithms. In *Proc. European Signal Processing Conference (EUSIPCO)*. Aalborg, Denmark, 2010.

[Hoefling10] Hoefling H. A path algorithm for the fused lasso signal approximator. *Journal of Computational and Graphical Statistics*, 19(4) :984–1006, 2010.

[Hua88] Hua Y. and Sarkar T.K. Perturbation analysis of TK method for harmonic retrieval problem. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(2) :228–240, 1988.

[Hua90] Hua Y. and Sarkar T.K. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(5) :814–824, 1990.

[Hua92] Hua Y. Estimating two-dimensional frequencies by matrix enhancement and matrix pencil. *IEEE Transactions on Signal Processing*, 40(9) :2267–2280, 1992.

[Huang12] Huang L., Wu Y., So H., Zhang Y., and Huang L. Multidimensional sinusoidal frequency estimation using subspace and projection separation approaches. *IEEE Transactions on Signal Processing*, 60(10) :5536–5543, 2012.

[Huang14] Huang H., Liu L., and Ngadi M.O. Recent developments in hyperspectral imaging for assessment of food quality and safety. *Sensors*, 14(4) :7248–7276, 2014.

[Jonsson05] Jonsson P., Bruce S.J., Moritz T., Trygg J., Sjöström M., Plumb R., Granger J., Maibaum E., Nicholson J.K., Holmes E., and Antti H. Extraction, interpretation and validation of information for comparing samples in metabolic LC/MS data sets. *Analyst*, 130(5) :701–707, 2005.

[Kay93] Kay S. *Fundamentals of statistical signal processing. Estimation theory*. Prentice Hall International Editions, Englewood Cliffs, New Jersey, 1993.

[Kolda09] Kolda T. and Bader B. Tensor decompositions and applications. *SIAM Review*, 51(3) :455–500, 2009.

[Kot93] Kot A.C., Tufts D.W., and Vaccaro R.J. Analysis of linear prediction by matrix approximation. *IEEE Transactions on Signal Processing*, 41(11) :3174–3177, 1993.

[Kowalski13] Kowalski M., Siedenburg K., and Dörfler M. Social sparsity ! neighborhood systems enrich structured shrinkage operators. *IEEE Transactions on Signal Processing*, 61(10) :2498–2511, 2013.

[Kumaresan82] Kumaresan R. and Tufts D.W. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30(6) :833–840, 1982.

[Kung83] Kung R., Arun K.S., and Rao D.V.B. State-space and singular value decomposition-based approximation methods for the harmonic retrieval problem. *Journal of the Optical Society of America*, 73(12) :1799–1811, 1983.

[Lagaert11] Lagaert J.B. *Modélisation de la croissance tumorale.* Ph.D. thesis, Université de Bordeaux 1, 2011.

[Li98] Li Y., Razavilar J., and Liu K. A high-resolution technique for multidimensional NMR spectroscopy. *IEEE Transactions on Biomedical Engineering*, 45(1) :78–86, 1998.

[Li13] Li Q., He X., Wang Y., Liu H., Xu D., and Guo F. Review of spectral imaging technology in biomedical engineering : achievements and challenges. *Journal of Biomedical Optics*, 18(10) :100901–100901, 2013.

[Lin13] Lin C. and Fang W. Efficient multidimensional harmonic retrieval : A hierarchical signal separation framework. *IEEE Signal Processing Letters*, 20(5) :427–430, 2013.

[Liu06] Liu J. and Liu X. An eigenvector-based approach for multidimensional frequency estimation with improved identifiability. *IEEE Transactions on Signal Processing*, 54(12) :4543–4556, 2006.

[Liu07] Liu J., Liu X., and Ma X. Multidimensional frequency estimation with finite snapshots in the presence of identical frequencies. *IEEE Transactions on Signal Processing*, 55 :5179–5194, 2007.

[Liu10] Liu J., Yuan L., and Ye J. An efficient algorithm for a class of fused lasso problems. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 323–332. ACM, 2010.

[Lorente12] Lorente D., Aleixos N., Gómez-Sanchis J., Cubero S., García-Navarrete O.L., and Blasco J. Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. *Food and Bioprocess Technology*, 5(4) :1121–1142, 2012.

[Mairal08] Mairal J., Bach F., Ponce J., Sapiro G., and Zisserman A. Supervised dictionary learning. Research report, INRIA, https://hal.inria.fr/inria-00322431, 2008.

[Malioutov05] Malioutov D., Cetin M., and Willsky A. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Transactions on Signal Processing*, 53(8) :3010–3022, 2005.

[Mallat93] Mallat S. and Zhang Z. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12) :3397–3415, 1993.

[Mazet05] Mazet V., Carteret C., Brie D., Idier J., and Humbert B. Background removal from spectra by designing and minimising a non-quadratic cost function. *Chemometrics and Intelligent Laboratory Systems*, 76(2) :121–133, 2005.

[Mazet13] Mazet V., Soussen C., and Djermoune E.H. Décomposition de spectres en motifs paramétriques par approximation parcimonieuse. In *24ème Colloque GRETSI Traitement du Signal & des Images*. Brest, France, 2013.

[Mokios04] Mokios K., Sidiropoulos N., Pesavento M., and Mecklenbrauker C. On 3-D harmonic retrieval for wireless channel sounding. In *Proc. IEEE ICASSP*, pages ii89–ii92. Montreal, Canada, May 2004.

[Nion10] Nion D. and Sidiropoulos S. Tensor algebra and multidimensional retrieval in signal processing for MIMO radar. *IEEE Transaction on Signal Processing*, 58(1) :5693–5705, 2010.

[Nocedal06] Nocedal J. and Wright S.J. *Numerical optimization*. Springer Series on Operation Research and Financial Engineering, New York, USA, second edition, 2006.

[Okhovat89] Okhovat A. and Cruz J.R. Statistical analysis of the Tufts-Kumaresan and principal Hankel components methods for estimating damping factors of single complex exponentials. In *Proc. IEEE ICASSP*, pages 2286–2289. May 1989.

[Olshausen97] Olshausen B.A. and Field D.J. Sparse coding with an overcomplete basis set : A strategy employed bb V1 ? *Vision Research*, 37(23) :3311–3325, 1997.

[Parikh13] Parikh N. and Boyd S. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(13) :123–231, 2013.

[PellencST17] PellencST. Pellenc Selective Technology : Mistral Product. http://www.pellencst.com/products, 2017. Accessed : 2016-09-01.

[Pesavento04] Pesavento M., Mecklenbräuker C., and Böme J. Multidimensional rank reduction estimator for MIMO channel models. *EURASIP Journal on Applied Signal Processing*, 2004(9) :1354–1363, 2004.

[Porat87] Porat B. and Friedlander B. On the accuracy of the Kumaresan-Tufts method for estimating complex damped exponentials. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(2) :231–235, 1987.

[Rao88] Rao B.D. Perturbation analysis of an SVD-based linear prediction method for estimating the frequencies of multiple sinusoids. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(7) :1026–1035, 1988.

[Rao89] Rao B.D. and Hari K.V.S. Performance analysis of ESPRIT and TAM in determining the direction of arrival of plane waves in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12) :1990–1995, 1989.

[Rao92] Rao B.D. and Arun K.S. Model based processing of signals : A state space approach. *Proceedings of the IEEE*, 80(2) :283–309, 1992.

[Rouquette01] Rouquette S. and Najim M. Estimation of frequencies and damping factors by two-dimensional ESPRIT type methods. *IEEE Transactions on Signal Processing*, 49(1) :237–245, 2001.

[Roy86] Roy R., Paulraj A., and Kailath T. ESPRIT–a subspace rotation approach to estimation of parameters of cisoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5) :1340–1342, 1986.

[Rubinstein10a] Rubinstein R., Bruckstein A.M., and Elad M. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6) :1045–1057, 2010.

[Rubinstein10b] Rubinstein R., Zibulevsky M., and Elad M. Double sparsity : Learning sparse dictionaries for sparse signal approximation. *IEEE Transactions on Signal Processing*, 58(3) :1553–1564, 2010.

[Sacchini93] Sacchini J., Steedly W., and Moses R. Two-dimensional Prony modeling and parameter estimation. *IEEE Transactions on Signal Processing*, 41(11) :3127–3137, 1993.

[Sahnoun12a] Sahnoun S. *Développement de méthodes d'estimation modale de signaux multidimensionnels. Application à la spectroscopie RMN*. Ph.D. thesis, Université de Lorraine, 2012.

[Sahnoun12b] Sahnoun S., Djermoune E.H., Soussen C., and Brie D. Sparse multidimensional modal analysis using a multigrid dictionary refinement. *EURASIP Journal on Advances in Signal Processing*, 60, 2012.

[Sahnoun17a] Sahnoun S., Djermoune E.H., Brie D., and Comon P. A simultaneous sparse approximation method for multidimensional harmonic retrieval. *Signal Processing*, 131 :36–48, 2017.

[Sahnoun17b] Sahnoun S., Usevich K., and Comon P. Multidimensional ESPRIT for damped and undamped signals : Algorithm, computations and perturbation analysis. *Hal archive*, 2017. URL https://hal.archives-ouvertes.fr/hal-01360438v3/document.

[Schmidt79] Schmidt R.O. Multiple emitter location and signal parameter estimation. In *Proc. RADC Spectral Estimation Workshop*, pages 243–258. Rome, NY, 1979.

[Serranti12] Serranti S., Gargiulo A., and Bonifazi G. Classification of polyolefins from building and construction waste using NIR hyperspectral imaging system. *Resources, Conservation and Recycling*, 61 :52–58, 2012.

[Shan85] Shan T., Wax M., and Kailath T. On spatial smoothing for direction-of-arrival estimation of coherent signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(4) :806–811, 1985.

[Siesler02] Siesler H.W., Ozaki Y., Kawata S., and Heise H.M. *Near-infrared spectroscopy : Principles, instruments, applications*. Wiley-VCH, Weinheim, Germany, 2002.

[So10] So H., Chan F., Lau W., and Chan C. An efficient approach for two-dimensional parameter estimation of a single-tone. *IEEE Transactions on Signal Processing*, 58(4) :1999–2009, 2010. ISSN 1053-587X.

[Song16] Song Y., Brie D., Djermoune E.H., and Henrot S. Regularization parameter estimation for non-negative hyperspectral image deconvolution. *IEEE Transactions on Image Processing*, 25(11) :5316–5330, 2016.

[Soussen11] Soussen C., Idier J., Brie D., and Duan J. From Bernoulli-Gaussian deconvolution to sparse signal restoration. *IEEE Transactions on Signal Processing*, 59(10) :4572–4584, 2011.

[Soussen15] Soussen C., Idier J., Duan J., and Brie D. Homotopy based algorithms for ℓ_0-regularized least-squares. *IEEE Transactions on Signal Processing*, 63(13) :3301–3316, 2015.

[Steedly94] Steedly W.M., Ying C.H.J., and Moses R.L. Statistical analysis of TLS-based Prony techniques. *Automatica*, 30(1) :115–129, 1994.

[Stoica89a] Stoica P. and Nehorai A. MUSIC, maximum likelihood, and Cramér-Rao bound. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(5) :720–741, 1989.

[Stoica89b] Stoica P., Söderström T., and Ti F. Overdetermined Yule-Walker estimation of the frequencies of multiple sinusoids : Accuracy aspects. *Signal Processing*, 16 :155–174, 1989.

[Stoica11] Stoica P., Babu P., and Li J. SPICE : A sparse covariance-based estimation method for array processing. *IEEE Transactions Signal Processing*, 59(2) :629–638, 2011.

[Stuart05] Stuart B. *Infrared spectroscopy*. Wiley Online Library, New York, USA, 2005.

[Sun12] Sun W. and So H.C. Accurate and computationally efficient tensor-based subspace approach for multi-dimensional harmonic retrieval. *IEEE Transactions on Signal Processing*, 60(10) :5077–5088, 2012.

[Sward14] Sward J., Adalbjornsson S., and Jakobsson A. High resolution sparse estimation of exponentially decaying signals. In *Proc. IEEE ICASSP*, pages 7203–7207. IEEE, 2014.

[Swindlehurst92] Swindlehurst A.L. and Kailath T. A performance analysis of subspace-based methods in the presence of model errors, Part I : the MUSIC algorithm. *IEEE Transactions on Signal Processing*, 40(7) :1758–1774, 1992.

[Tang13] Tang G., Bhaskar B., Shah P., and Recht B. Compressed sensing off the grid. *IEEE Transactions on Information Theory*, 59(11) :7465–7490, 2013.

[Tatzer05] Tatzer P., Wolf M., and Panner T. Industrial application for inline material sorting using hyperspectral imaging in the NIR range. *Real-Time Imaging*, 11(2) :99–107, 2005.

[Thiebaut17] Thiebaut C., Chamard-Jovenin C., Chesnel A., Morel C., Djermoune E.H., Boukhobza T., and Dumond H. Mammary epithelial cell phenotype disruption in vitro and in vivo through ERalpha36 overexpression. *Plos One*, to appear, 2017.

[Tibshirani05] Tibshirani R., Saunders M., Rosset S., Zhu J., and Knight K. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 67(1) :91–108, 2005.

[Tibshirani11] Tibshirani R.J., Taylor J.E., Candès E.J., and Hastie T. *The solution path of the generalized lasso*. Ph.D. thesis, Stanford University, 2011.

[Toussaint16] Toussaint M. *Thérapies par rayonnements appliquées au cas du glioblastome : Intérêt du suivi par spectroscopie et imagerie de diffusion par résonance magnétique*. Ph.D. thesis, Université de Lorraine, 2016.

[Tropp06] Tropp J., Gilbert A., and Strauss M. Algorithms for simultaneous sparse approximation. Part I : Greedy pursuit. *Signal Processing*, 86 :572–588, 2006.

[Tufts82] Tufts D.W. and Kumaresan R. Estimation of frequencies of multiple sinusoids : Making linear prediction perform like maximum likelihood. *Proceedings of the IEEE*, 70(9) :975–989, 1982.

[Turlach05] Turlach B.A., Venables W.N., and Wright S.J. Simultaneous variable selection. *Technometrics*, 47(3) :349–363, 2005.

[Tylcz15] Tylcz J.B., El Alaoui-Lasmaili K., Djermoune E.H., Thomas N., Faivre B., and Bastogne T. Data-driven modeling and characterization of anti-angiogenic molecule effects on tumoral vascular density. *Biomedical Signal Processing and Control*, 20 :52–60, 2015. doi :10.1016/j.bspc.2015.04.008.

[Wilkinson65] Wilkinson J.H. *The algebraic eigenvalue problem*. Oxford University Press, Oxford, 1965.

[Willett14] Willett R.M., Duarte M.F., Davenport M.A., and Baraniuk R.G. Sparsity and structure in hyperspectral imaging : Sensing, reconstruction, and target detection. *IEEE Signal Processing Magazine*, 31(1) :116–126, 2014.

[Xin14] Xin B., Kawahara Y., Wang Y., and Gao W. Efficient generalized fused lasso and its application to the diagnosis of Alzheimer's disease. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 2163–2169. 2014.

[Xu14] Xu W., J.-F. C., Mishra K., Cho M., and Kruger A. Precise semidefinite programming formulation of atomic norm minimization for recovering d-dimensional ($d \geq 2$) off-the-grid frequencies. In *Information Theory and Applications Workshop (ITA)*. IEEE, 2014. URL https://arxiv.org/abs/1312.0485.

[Yaghoobi15] Yaghoobi M., Wu D., and Davies M.E. Fast non-negative orthogonal matching pursuit. *IEEE Signal Processing Letters*, 22(9) :1229–1233, 2015.

[Yang16a] Yang Z. and Xie L. On gridless sparse methods for multi-snapshot DOA estimation. In *Proc. IEEE ICASSP*. IEEE, 2016.

[Yang16b] Yang Z., Xie L., and Stoica P. Vandermonde decomposition of multilevel toeplitz matrices with application to multidimensional super-resolution. *IEEE Transactions on Information Theory*, 62(6) :3685–3701, 2016. URL http://arxiv.org/abs/1505.02510.

[Yao95] Yao Y.X. and Pandit S.M. Cramér-Rao lower bounds for a damped sinusoidal process. *IEEE Transactions on Signal Processing*, 43(4) :878–885, 1995.

[Yau93] Yau S.F. and Bresler Y. Maximum likelihood parameter estimation of superimposed signals by dynamic programming. *IEEE Transactions on Signal Processing*, 41(2) :804–820, 1993.

[Ying99] Ying C.J. and Potter L.C. Minimum variance linear estimation of amplitudes for exponential signal models. *IEEE Transactions on Signal Processing*, 47(9) :2522–2525, 1999.

[Yoon11] Yoon S.C., Park B., Lawrence K.C., Windham W.R., and Heitschmidt G.W. Line-scan hyperspectral imaging system for real-time inspection of poultry carcasses with fecal material and ingesta. *Computers and Electronics in Agriculture*, 79(2) :159–168, 2011.

[Zhou12] Zhou J., Liu J., Narayan V.A., and Ye J. Modeling disease progression via fused sparse group lasso. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1095–1103. ACM, 2012.

Quatrième partie

Annexes

Annexe A

Sélection de publications

Je présente dans cette annexe une sélection de publications en lien avec mes activités de recherche. Certains de ces travaux ne sont pas détaillés dans ce mémoire par souci de cohérence du manuscrit et de clarté de la présentation. Ainsi, l'article présenté en page 111 porte sur l'estimation des paramètres de régularisation en déconvolution d'images hyperspectrales[1] et celui de la page 147 concerne la modélisation du soudage MIG/MAG en mode *short-arc*[2].

Sommaire

[1] Thèse de Yingying Song.
[2] Thèse Cifre de Jean-Pierre Planckaert.

A.1 A Simultaneous Sparse Approximation Method for Multidimensional Harmonic Retrieval

S. Sahnoun, **E.-H. Djermoune**, D. Brie, P. Comon. *Signal Processing*, vol. 131, pp. 36–48, 2017.

Cet article présente une stratégie permettant d'estimer les paramètre d'un signal harmonique mutidimensionel en combinant un algorithme d'approximation parcimonieuse à une méthode multigrille de mise à jour des atomes du dictionnaire. Des preuves de convergence sont données. Le calcul des bornes de Cramér-Rao des paramètres du modèle exponentiel amorti est également présenté. Les résultats principaux sont décrits dans le chapitre 3.

Signal Processing 131 (2017) 36–48

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

A simultaneous sparse approximation method for multidimensional harmonic retrieval ☆

Souleymen Sahnoun [a,*], El-Hadi Djermoune [b], David Brie [b], Pierre Comon [a]

[a] CNRS, Gipsa-Lab, Univ. Grenoble Alpes, F-38000 Grenoble, France
[b] CRAN, Université de Lorraine, CNRS, Boulevard des Aiguillettes, BP 239, 54506 Vandoeuvre Cedex, France

ARTICLE INFO

Article history:
Received 26 February 2016
Received in revised form
31 May 2016
Accepted 25 July 2016
Available online 28 July 2016

Keywords:
Multidimensional harmonic retrieval
Frequency estimation
Simultaneous sparse approximation
Multigrid dictionary refinement
Cramér–Rao lower bound

ABSTRACT

In this paper, a new method for the estimation of the parameters of multidimensional (R-D) harmonic and damped complex signals in noise is presented. The problem is formulated as R simultaneous sparse approximations of multiple 1-D signals. To get a method able to handle large size signals while maintaining a sufficient resolution, a multigrid dictionary refinement technique is associated to the simultaneous sparse approximation. The refinement procedure is proved to converge in the single R-D mode case. Then, for the general multiple modes case, the signal tensor model is decomposed in order to handle each mode separately in an iterative scheme. The proposed method does not require an association step since the estimated modes are automatically "paired". We also derive the Cramér–Rao lower bounds of the parameters of modal R-D signals. The expressions are given in compact form in the single tone case. Finally, numerical simulations are conducted to demonstrate the effectiveness of the proposed method.

1. Introduction

The problem of estimating the parameters of sinusoidal signals from noisy measurements is an important topic in signal processing and several parametric and nonparameteric approaches have been developed for one-dimensional (1-D) signals [1]. Recently, this problem has received a renewed interest thanks to the emergence of multidimensional (R-D) applications. Indeed, parameter estimation from R-D signals is required in numerous applications in signal processing and communications such as nuclear magnetic resonance (NMR) spectroscopy, wireless communication channel estimation [2] and MIMO radar imaging [3]. In all these applications, signals are assumed to be a superposition of R-D sinusoids or, more generally, of exponentially decaying R-D complex exponentials (modal signals). As for the 1-D case, the crucial step is the estimation of the R-D modes (including frequencies and damping factors) because they are nonlinear functions of the data. In this paper, we consider the single snapshot R-D signal model described in [4].

In order to achieve high resolution estimates, parametric approaches are often preferred to nonparametric ones. Several parametric R-D methods ($R \geq 2$) have been proposed. They include linear prediction-based methods such as 2-D TLS-Prony [5], and subspace approaches such as matrix enhancement and matrix pencil (MEMP) [6], 2-D ESPRIT [7], multidimensional folding (MDF) [8], improved multidimensional folding (IMDF) [9,10], Tensor-ESPRIT [11], principal-singular-vector utilization for modal analysis (PUMA) [12,13] and the methods proposed in [14,15]. All these methods perform at various degrees but it is generally admitted that they yield accurate estimates at high SNR scenarios and/or when the frequencies are well separated. This is obtained at the expense of computational effort. In [12], tensor PUMA was proposed as an accurate and computationally efficient multidimensional harmonic retrieval method, which attains the Cramér–Rao lower bound (CRLB) and does not require to build large size matrix or tensor. However its performance degrades rapidly with the increase of the number of components in the R-D signal.

Recently, methods based on sparse approximations have been proposed to address the harmonic or modal retrieval problem [16–23]. For time-data spectral estimation, the dictionary is formed from a set of (normalized) complex exponentials potentially embedded in the data, which allows one to easily include some prior knowledge about the position of some known modes. More generally, the usual choice is a uniform spectral grid obtained by sampling the frequency and damping factor lines. Clearly, a fine

☆This work is funded by the European Research Council under the European Community's Seventh Framework Programme FP7/2007–2013 Grant Agreement no. 320594, DECODA project.

* Corresponding author.
E-mail addresses: souleymen.sahnoun@gipsa-lab.grenoble-inp.fr (S. Sahnoun), el-hadi.djermoune@univ-lorraine.fr (E.-H. Djermoune), david.brie@univ-lorraine.fr (D. Brie), pierre.comon@gipsa-lab.grenoble-inp.fr (P. Comon).

http://dx.doi.org/10.1016/j.sigpro.2016.07.029
0165-1684/

grid is required to get a good resolution but, on the other hand, it will result in a huge dictionary [16]. This complexity is further increased in the case of R-D signals in which we are confronted with $2R$-D grids. In order to reduce the computational burden, a multigrid scheme for sparse approximation was proposed in [20] to iteratively refine the dictionary starting from a coarse one. At each iteration, a sparse approximation is performed and then new grid points (called "atoms") are inserted in the vicinity of active ones leading to a multiresolution-like scheme. This algorithm, which refines jointly R2-D grids, is efficient but has mainly two drawbacks: (1) it does not have convergence guarantees, (2) the dictionary becomes intractable for large signals when $R \geq 2$. Recently, several studies have also focused on gridless sparse recovery methods based on continuous dictionaries [24,25]. However, the proposed algorithms demand a large computational burden even for 1-D signals.

The goal of the present paper is to propose a fast multidimensional modal estimation technique able to handle large signals and yielding a good estimation accuracy.

1. First, the proposed approach, as for some parametric methods for modal retrieval, is based on the idea of estimating the parameters independently along each dimension $r = 1, \ldots, R$. It will be shown that the *simultaneous* sparse approximation concept [26,21] is well-suited for R-D modal retrieval ($R \geq 2$).
2. The second contribution consists in the proposition of a new multigrid scheme which amounts to consider a two-step refinement of 1-D grids, the first step for frequencies and the second one for damping factors. One advantage of this procedure is that it reduces the computational time. The convergence of the proposed multigrid strategy is analyzed in the single tone case ($F = 1$), and convergence conditions are expressed in terms of atom positions in the initial dictionaries.
3. The extension of this result to the multiple tones case ($F > 1$) is not trivial because, not only it depends on the selected sparse approximation algorithm, but also on the coherence of the dictionary [26]. Indeed, due to the multigrid strategy, the columns of the refined dictionary are increasingly correlated, which may prevent convergence even in the noiseless case. Consequently, for $F > 1$, we exploit an alternative representation of the data model enabling the extraction of the R-D signal tones separately. Therefore, the third contribution of this paper is the derivation of a new algorithm for estimating parameters of R-D damped signals in which the results of the previous contribution apply. The effectiveness of the new algorithm for multiple R-D tones is also analyzed. One very interesting by-product of this approach is that the pairing of R-D parameters is achieved for free, without any further association stage.

The usual way to assess the performances of an estimation method is to compare the variance of the estimates to the CRLB. In [6] Hua derived the CRLB for 2-D frequencies, i.e., undamped 2-D exponentials; no damped signals are considered. Closed-form expressions of the CRLB for the general undamped R-D case are derived in [27]. CRLB for 2-D damped signals are derived in [28]. Therefore, to the best of our knowledge, no compact expressions of the CRLB's are available for the general R-D damped model. Thus, another contribution of the paper is the derivation of the CRLB's for the frequency, damping factor, amplitude and phase of this model.

The remainder of this paper is organized as follows. In Section 2, we introduce notation and present the R-D modal retrieval problem. In Section 3, we formulate the R-D modal estimation problem as R simultaneous sparse estimation problems, show how to construct a modal dictionary on a uniform grid and then describe the new fast multigrid strategy. In Section 4, we give sufficient conditions for convergence of the multigrid dictionary refinement in the case of single tone R-D signals. In light of these new results, we propose in Section 5 a new efficient algorithm for multiple tones R-D signals. In Section 6, we derive the expressions of the CRLB's for the parameters of R-D damped exponentials in Gaussian white noise. We then give the CRLB in the cases of single damped and undamped R-D cisoids. The effectiveness of the proposed method is demonstrated using simulation signals in Section 7. Finally, conclusions are drawn in Section 8.

2. Notation and problem statement

2.1. Notation

Scalars are denoted as lower-case letters (a, b, α), column vectors as lower-case bold-face letters ($\mathbf{a}$, $\mathbf{b}$), matrices as bold-face capitals ($\mathbf{A}$, $\mathbf{B}$), and tensors as calligraphic bold-face letters ($\mathcal{A}$, $\mathcal{B}$). Notations $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$ and $(\cdot)^{\dagger}$ stand for the transpose, the Hermitian transpose and the pseudo-inverse, respectively. The symbols "$\odot$" and "$\boxtimes$" will denote the Khatri–Rao product (column-wise Kronecker) and the Kronecker product, respectively. Both words "mode" and "tone" are used to refer to a component of the multidimensional signal. The tensor operations used here are consistent with [29]:

- the outer product of two tensors $\mathcal{A} \in \mathbb{C}^{M_1 \times \cdots \times M_R}$ and $\mathcal{B} \in \mathbb{C}^{K_1 \times \cdots \times K_N}$ is given by:

$$\mathcal{C} = \mathcal{A} \otimes \mathcal{B} \in \mathbb{C}^{M_1 \times \cdots \times M_R \times K_1 \times \cdots \times K_N},$$
$$c(m_1, \ldots, m_R, k_1, \ldots, k_N) = a(m_1, \ldots, m_R) b(k_1, \ldots, k_N) \quad (1)$$

- the contraction product acting on the r-th index of a tensor $\mathcal{A} \in \mathbb{C}^{M_1 \times \cdots \times M_R}$ and the second index of a matrix $\mathbf{U} \in \mathbb{C}^{K \times M_r}$ is:

$$\mathcal{B} = \mathcal{A} \underset{r}{\bullet} \mathbf{U} \in \mathbb{C}^{M_1 \times \cdots \times M_{r-1} \times K \times M_{r+1} \times \cdots \times M_R},$$
$$b(m_1, m_2, \ldots, m_{r-1}, k_r, m_{r+1}, \ldots, m_R) = \sum_{m_r=1}^{M_r} a(m_1, m_2, \ldots, m_R) u(k_r, m_r) \quad (2)$$

- the matrix $\mathbf{A}_{(r)} \in \mathbb{C}^{M_r \times (M_1 \cdots M_{r-1} M_{r+1} \cdots M_R)}$ represents the unfolding (dimension-r matricization) of the tensor $\mathcal{A}$ and corresponds to the arrangement of the dimension-r fibers of $\mathcal{A}$ in the columns of the resulting matrix.
- $\| \mathcal{A} \|$ denotes the Frobenius norm for tensors.
- The concatenation of two tensors $\mathcal{A}_1 \in \mathbb{C}^{M_1 \times \cdots \times M_{r-1} \times K_1 \times M_{r+1} \times \cdots \times M_R}$ and $\mathcal{A}_2 \in \mathbb{C}^{M_1 \times \cdots \times M_{r-1} \times K_2 \times M_{r+1} \times \cdots \times M_R}$ along the rth dimension is denoted by $\mathcal{A}_1 \sqcup_r \mathcal{A}_2$ and obtained by stacking $\mathcal{A}_1$ and $\mathcal{A}_2$ along the rth dimension.

Finally, throughout this paper, the tilde symbol ($\tilde{}$) denotes a noisy signal; e.g. $\tilde{y}(\cdot) = y(\cdot) + e(\cdot)$.

2.2. Problem formulation

An R-D modal signal is modeled as the superposition of F multidimensional damped complex sinusoids:

$$\tilde{y}(m_1, \ldots, m_R) = \sum_{f=1}^{F} c_f \prod_{r=1}^{R} a_{f,r}^{m_r - 1} + e(m_1, \ldots, m_R) \quad (3)$$

where $m_r = 1, \ldots, M_r$ for $r = 1, \ldots, R$. M_r denotes the sample support of the r-th dimension, $a_{f,r} = \exp(\alpha_{f,r} + j\omega_{f,r}) \in \mathbb{C}$ is the f-th mode in the r-th dimension, $\{\alpha_{f,r}\}_{f=1,r=1}^{F,R}$, $\alpha_{f,r} \in \mathbb{R}^-$, are the damping factors, $\{\omega_{f,r} = 2\pi\nu_{f,r}\}_{f=1,r=1}^{F,R}$ are the angular frequencies, and $c_f = \lambda_f \exp(j\phi_f)$ is the complex amplitude of the f-th mode where

$\lambda_f = |c_f|$ denotes the magnitude and ϕ_f the phase. $e(m_1, m_2, \ldots, m_R)$ is a zero-mean complex Gaussian white noise with variance σ^2 and mutually independent in all dimensions.

In a tensor form, the R-D signal in (3) may be written as

$$\widetilde{\mathcal{Y}} = \mathcal{Y} + \mathcal{E} \tag{4}$$

where $\{\widetilde{\mathcal{Y}}, \mathcal{Y}, \mathcal{E}\} \in \mathbb{C}^{M_1 \times M_2 \times \cdots \times M_R}$. The problem consists in estimating the set of parameters $\{a_{f,r}\}_{f=1,r=1}^{F,R}$ and $\{c_f\}_{f=1}^{F}$ from the R-D signal samples.

3. Simultaneous sparse approximation for R-D modal signals

3.1. Tensor formulation of the data model

The noise-free data tensor $\mathcal{Y}$ in (4) can be written in the following form:

$$\mathcal{Y} = \sum_{f=1}^{F} c_f \mathbf{a}_{f,1} \otimes \mathbf{a}_{f,2} \otimes \cdots \otimes \mathbf{a}_{f,R} \tag{5}$$

where $\mathbf{a}_{f,r} = [1, a_{f,r}, \ldots, a_{f,r}^{M_r-1}]^{\mathrm{T}}$, $r = 1, \ldots, R$. Eq. (5) is called the Canonical Polyadic (CP) decomposition form, or the Candecomp/Parafac decomposition of the tensor $\mathcal{Y}$ [29,30]. The CP model (5) can be concisely denoted by $\mathcal{Y} = [\![\mathbf{c}; \mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_R]\!]$ where $\mathbf{A}_r = [\mathbf{a}_{1,r}, \mathbf{a}_{2,r}, \ldots, \mathbf{a}_{F,r}]$, $r = 1, \ldots, R$, and $\mathbf{c} = [c_1, c_2, \ldots, c_F]^{\mathrm{T}}$ is the vector of complex amplitudes. Using these definitions, the matricized form of $\mathcal{Y}$ along the r-th dimension is given by

$$\mathbf{Y}_{(r)} = \mathbf{A}_r \mathbf{\Delta}_{\mathbf{c}} (\mathbf{A}_R \odot \cdots \odot \mathbf{A}_{r+1} \odot \mathbf{A}_{r-1} \odot \cdots \odot \mathbf{A}_1)^{\mathrm{T}} \tag{6}$$

where $\mathbf{\Delta}_{\mathbf{c}} = \mathrm{diag}(\mathbf{c})$. Then, we can write

$$\widetilde{\mathbf{Y}}_{(r)} = \mathbf{A}_r \mathbf{H}_r + \mathbf{E}_{(r)} \tag{7}$$

where $\mathbf{H}_r \in \mathbb{C}^{F \times M_r'}$ is

$$\mathbf{H}_r \mathrm{def} = \mathbf{\Delta}_{\mathbf{c}} (\mathbf{A}_R \odot \cdots \odot \mathbf{A}_{r+1} \odot \mathbf{A}_{r-1} \odot \cdots \odot \mathbf{A}_1)^{\mathrm{T}} \tag{8}$$

and $M_r' = \prod_{\substack{k=1 \\ k \neq r}}^{R} M_k$. Therefore

$$\mathbf{Y}_{(r)} \mathrm{def} = [\mathbf{y}_{(r),1}, \ldots, \mathbf{y}_{(r),M_r'}] = \left[\sum_{f=1}^{F} h_r(f, 1) \mathbf{a}_{f,r}, \ldots, \sum_{f=1}^{F} h_r(f, M_r') \mathbf{a}_{f,r} \right] \tag{9}$$

where $h_r(f, m_r')$ is the (f, m_r') entry of the matrix $\mathbf{H}_r$, for $f = 1, \ldots, F$ and $m_r' = 1, \ldots, M_r'$.

3.2. Simultaneous sparse approximation

Assuming[1] that $M_r > F, \forall\, r$, it is easy to see from (9) that the mode coordinates $\{a_{f,r}\}_{f=1}^{F_r}$ ($F_r \leq F$) in the r-th dimension are identifiable from any column of $\mathbf{Y}_{(r)}$. This process may be repeated on each dimension $r = 1, \ldots, R$ to get all the modes coordinates. In practice, we have to replace the matrix $\mathbf{Y}_{(r)}$ by its noisy counterpart $\widetilde{\mathbf{Y}}_{(r)}$ accounting for the additive white noise. In this case, (9) holds only approximately. Consequently, for each column $\tilde{\mathbf{y}}_{(r),m_r'}$, $m_r' = 1, \ldots, M_r'$, the modal estimation problem can be formulated as a sparse approximation problem corresponding to the following constrained optimization:

$$\mathbf{x}_{m_r'} = \arg\min_{\mathbf{x}} \| \mathbf{x} \|_0 \quad \text{subject to} \quad \| \tilde{\mathbf{y}}_{(r),m_r'} - \mathbf{Q}_r \mathbf{x} \|_2^2 \leq \epsilon \tag{10}$$

where $\mathbf{Q}_r \in \mathbb{C}^{M_r \times N}$, $N \gg M_r$, is a (known) modal dictionary, $\mathbf{x} \in \mathbb{C}^N$ is a (sparse) vector containing the coefficients of the activated columns in $\mathbf{Q}_r$, and ϵ is a small reconstruction error related to the noise variance. The pseudo-norm $\| \mathbf{x} \|_0$ counts the number of nonzero elements in $\mathbf{x}$. The design of $\mathbf{Q}_r$ is discussed in Section 3.3. The fact that each vector $\tilde{\mathbf{y}}_{(r),m_r'}$ corresponds to a 1-D signal generated by the same modes implies that the position of nonzero entries in $\mathbf{x}_{m_r'}$ should be the same for $m_r' = 1, 2, \ldots, M_R'$. Let $\mathbf{X}$ be the matrix defined by

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{M_r'}], \tag{11}$$

then the sparsity of $\mathbf{X}$ may be measured by computing the Euclidian norms of the rows: those providing a nonzero norm define the rows of active atoms (which are estimations of modes $a_{f,r}$ in the dimension r) in the dictionary $\mathbf{Q}_r$. Therefore, we are facing a simultaneous sparse approximation problem:

$$\mathbf{X}_r = \arg\min_{\mathbf{X}} \| \mathbf{X} \|_{0,2} \quad \text{subject to} \quad \| \widetilde{\mathbf{Y}}_{(r)} - \mathbf{Q}_r \mathbf{X} \|_F^2 \leq \epsilon \tag{12}$$

where $\| \mathbf{X} \|_{0,2}$ is the mixed ℓ_0/ℓ_2-norm of $\mathbf{X}$ (the number of rows with nonzero ℓ_2-norm). The simultaneous sparse representation models, called also "multiple measurement vectors" (MMV), have been studied from several angles of view, and different approaches have been proposed (see [31] and references therein). As the goal of the present paper is to develop a fast approach well adapted to large signals, we restrict our attention to the SOMP algorithm [26]. However, it is worth mentioning that, in more intricate cases and/or small size signals, much more efficient simultaneous sparse algorithms may be used at the price of an increased computational burden. A straightforward way to get the R-tuples $\{(a_{f,1}, \ldots, a_{f,R})\}_{f=1}^{F}$ consists in estimating the modes $a_{f,r}$ in the R dimensions using matrices $\widetilde{\mathbf{Y}}_{(r)}$, $r = 1, \ldots, R$, which requires a further pairing step to form the R-D modes. To get accurate estimates using the described scheme, two conditions have to be satisfied, (1) the dictionary should contain all modes in the signal, (2) the sparse approximation method should have sufficient guarantees for selecting the true atoms from the dictionary: this is known as "exact recovery guarantees". These problems are discussed in the following sections and an alternative representation of the data is used to avoid the pairing stage in the multiple tones case.

3.3. Modal dictionary design and multigrid strategy

3.3.1. Uniform modal dictionary

The dictionary $\mathbf{Q}_r \in \mathbb{C}^{M_r \times N}$ can be defined from a discretization of the (ν, α) plane. Each point of the grid corresponds to a hypothetic mode. Let N_μ be the number of points of a uniform grid covering the frequency interval $[0, 1)$. Similarly, let N_β be the number of points of a uniform grid covering the damping factor interval $(\beta_{\min}, 0]$, where $\beta_{\min}$ is a lower bound on $\{\alpha_{f,r}\}_{f=1}^{F}$. Then $\mathbf{Q}_r$ is given by

$$\begin{aligned} \mathbf{Q}_r = [&\mathbf{q}_r(0, 0), \ldots, \mathbf{q}_r((N_\mu - 1)\delta_\mu, 0), \mathbf{q}_r(0, \delta_\beta) \\ &, \ldots, \mathbf{q}_r((N_\mu - 1)\delta_\mu, \delta_\beta), \ldots, \mathbf{q}_r((N_\mu - 1)\delta_\mu, (N_\beta - 1)\delta_\beta)] \end{aligned} \tag{13}$$

where $\mathbf{q}_r(\mu, \beta) = \frac{\mathbf{a}_r(\mu, \beta)}{\| \mathbf{a}_r(\mu, \beta) \|_2}$, $\mathbf{a}_r(\mu, \beta) = [1, e^{(\beta + j2\pi\mu)}, \ldots, e^{(\beta + j2\pi\mu)(M_r - 1)}]^{\mathrm{T}}$, $\delta_\beta = \beta_{\min}/N_\beta$, and $\delta_\mu = 1/N_\mu$. The total number of columns in $\mathbf{Q}_r$ is $N = N_\mu N_\beta \gg F$, each of them is called atom. In the aim of reducing the computational complexity, we propose to estimate frequencies and then damping factors by calling twice the sparse approximation method. At the first step, the frequencies are estimated using a harmonic dictionary. In the second step, the damping factors are estimated using a modal dictionary formed by the already estimated frequencies and a damping factor grid. These two steps are explained in Section 4.

3.3.2. Multi-grid dictionary refinement

To achieve a high-resolution modal estimation, a possible way is

[1] Note that this assumption is considered only in this section.

to define uniform grids as before and selecting very small values for δ_μ and δ_β. However, the resulting dictionaries will lead to prohibitive calculation cost and memory capacities requested. Rather, we propose to start with a coarse grid (N_μ and N_β low) and to adaptively refine it through a multigrid scheme as sketched in Fig. 1. Let ℓ be the current grid level ($\ell = 0, \ldots, L-1$). At level ℓ, we first restore the signal $\mathbf{X}_r(\ell)$ related to the dictionary $\mathbf{Q}_r(\ell)$ by applying the SOMP method. Then we refine the dictionary by inserting atoms inbetween pairs of $\mathbf{Q}_r(\ell)$, in the neighborhood of each activated atom, and we apply again the SOMP method to restore $\mathbf{X}_r(\ell+1)$ with respect to the refined dictionary $\mathbf{Q}_r(\ell+1)$. This process is repeated until the desired level of resolution is reached. This procedure is applied for both frequencies and damping factors. Algorithm 1 presents the one-step dictionary refinement (DICREF), from level ℓ to $\ell+1$, where, for a and b reals, linspace(a, b, η) generates a set of η equispaced points in the interval $[a, b]$. The difference between the present framework and that in [20] is the following. In [20] the multigrid algorithm refines jointly R2-D grids, which leads to expensive computations when $R \geq 2$, without convergence guarantees. The present mutigrid scheme refines linear grids, which leads to low computational complexity with convergence guarantees as will be shown in the next section.

Finding the convergence conditions of the new multigrid strategy in the general case (multiple tones) is not easy and depends on the selected sparse approximation algorithm. By contrast, it is possible to show that, under mild conditions, the convergence may be guaranteed in the single tone case. This issue is discussed in the next section. In Section 5, we make use of an alternative representation of the data model in the case of multiple tones and we propose a method allowing one to retrieve the signal tones separately.

Algorithm 1. Dictionary refinement (DICREF).

input : A vector $\mathbf{d} \in \mathbb{R}^N$ of sorted frequencies or damping factors, an index set Ω of activated atoms, the number of atoms $\eta \in \mathbb{N}$ to add at each side of an activated one
output: Updated vector $\mathbf{d}_{\text{updated}}$

for $i = 1 : \text{numel}(\Omega)$ **do**
  $\mathbf{d}_{i,1} = \text{linspace}\,(\mathbf{d}(\Omega(i)-1), \mathbf{d}(\Omega(i)), \eta)$
  $\mathbf{d}_{i,2} = \text{linspace}(\mathbf{d}(\Omega(i)), \mathbf{d}(\Omega(i)+1), \eta)$
  $\mathbf{d}_i = [\mathbf{d}_{i,1}^{\mathsf{T}}, \mathbf{d}_{i,2}^{\mathsf{T}}(2:\eta)]^{\mathsf{T}}$
end
$\mathbf{d}_{\text{updated}} = \text{union}(\mathbf{d}_1, \ldots, \mathbf{d}_{\text{numel}(\Omega)})$
return $\mathbf{d}_{\text{updated}}$

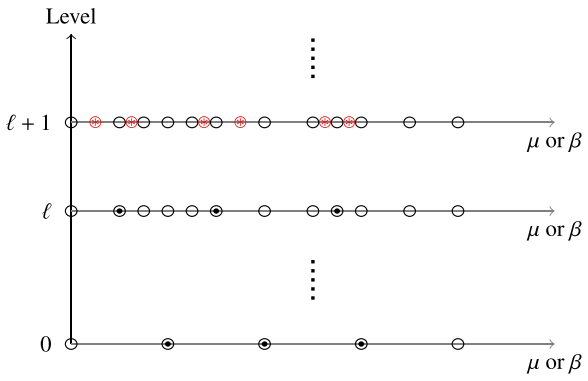


Fig. 1. The multigrid dictionary refinement procedure with $\eta = 1$. (○) atoms in the dictionary; (⊙) activated atoms; (⊛) new atoms.

4. Single *R*-D mode estimation

In the previous section, we have shown how the R-D modal retrieval problem may be tackled using a sparse approximation algorithm by estimating the set of parameters in each dimension $r = 1, \ldots, R$. Here, we give the sufficient conditions for convergence of the multigrid dictionary refinement for $F=1$. Without loss of generality, we set $R=1$. For notation simplicity, we omit reference to the dimension index r.

According to (3), the 1-D modal signal containing a single mode can be written as follows:

$$y(m) = c_1 a_1^{m-1} = c_1 e^{(\alpha_1 + j2\pi\nu_1)(m-1)}, \quad m = 1, \ldots, M. \tag{14}$$

Let $\mathbf{Q}$ be a normalized modal dictionary $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_N]$, with

$$\mathbf{q}_n = \frac{1}{\sqrt{\sum_m |q_n|^{2m}}}[1, q_n, \ldots, q_n^{M-1}]^{\mathsf{T}}, \tag{15}$$

$q_n = \exp(\beta_n + j2\pi\mu_n)$, $\mu_n \in [0, 1)$, $\beta_n \in (\beta_{\min}, 0]$, for $n = 1, \ldots, N$. The single tone sparse approximation of $\mathbf{y}$ with respect to $\mathbf{Q}$ is the solution of the criterion:

$$\min_{\mathbf{x}} J(\mathbf{x}) = \| \mathbf{y} - \mathbf{Q}\mathbf{x} \|^2 \quad \text{s. t.} \quad \| \mathbf{x} \|_0 = 1. \tag{16}$$

The optimal solution is given by

$$x_n^* = \mathbf{q}_n^{\mathrm{H}}\mathbf{y}, \quad \mathbf{x}^*_{\{1,\ldots,N\}\setminus n} = \mathbf{0}, \quad J(\mathbf{x}^*) = \| \mathbf{y} \|^2 - \mathbf{y}^{\mathrm{H}}\mathbf{q}_n\mathbf{q}_n^{\mathrm{H}}\mathbf{y} \tag{17}$$

where n is the selected column number in $\mathbf{Q}$. Finally, the minimum $J(\mathbf{x}^*)$ is reached for an atom $\mathbf{q}_n$ that maximizes $J'(\mathbf{q}_n) = \mathbf{y}^{\mathrm{H}}\mathbf{q}_n\mathbf{q}_n^{\mathrm{H}}\mathbf{y} = |\mathbf{q}_n^{\mathrm{H}}\mathbf{y}|^2$, $n = 1, \ldots, N$.

Remark 1. At this point we notice that, for 1-D single-tone harmonic signals, maximizing $J'(\mathbf{q}_n)$ leads to the well-known beamforming method (or more precisely to the periodogram in the single snapshot case). Hence, an estimate of the frequency ν_1 may be obtained from the peak of $J'(\mathbf{q}_n)$. Here, as in [32], we use the sparse approximation framework. In our case, it allows to estimate the frequency and then the damping factor from a unified point of view and using a unique algorithm based on SOMP and DICREF.

4.1. Estimating the frequency: the harmonic dictionary

First, we estimate frequency ν_1 using a harmonic dictionary (i.e. assuming $\beta_n = 0, \forall\, n$). In this case, we have:

$$J'(\mu_n) = \frac{|c_1|^2}{M}\left|\frac{1 - e^{\alpha_1 M + j2\pi(\nu_1 - \mu_n)M}}{1 - e^{\alpha_1 + j2\pi(\nu_1 - \mu_n)}}\right|^2. \tag{18}$$

The following theorem gives a sufficient condition for the multigrid dictionary refinement scheme to converge to the global maximum of J'.

Theorem 1. *Let $y(m)$ be a single tone ($F=1$) noiseless signal of length M and $\mathbf{Q}(\ell=0) = [\mathbf{q}_1\ \mathbf{q}_2\ \ldots\ \mathbf{q}_{N(0)}]^{\mathsf{T}}$ be the initial harmonic dictionary in which the columns are sorted in increasing order of $\mu_n(0)$, $n = 1, 2, \ldots, N(0)$ and covering the frequency interval [0,1): $\mu_1(0) = 0$ and $\mu_{N(0)}(0) = 1 - 1/M$. Then the refinement scheme is convergent (i.e. $\exists\, n \in \{1, \ldots, N(\ell)\}$ s.t. $\lim_{\ell\to\infty}\mu_n(\ell) = \nu_1$) if the following condition is satisfied:*

$$\max_{n\in\{1,\ldots,N(0)-1\}} |\mu_{n+1}(0) - \mu_n(0)| < 2\zeta_M \tag{19}$$

where ζ_M is a constant depending only on M.

Proof. It is easy to check that the global maximum of $J'(\mu_n)$ is reached for $\mu_n = \nu_1, \forall\, \alpha_1$. Fig. 2 shows the shape of $J'(\mu_n)$ for $\beta_n = 0$. For $\alpha_1 = 0$, $J'(\mu_n)$ reduces to a Fejér kernel of order M which has

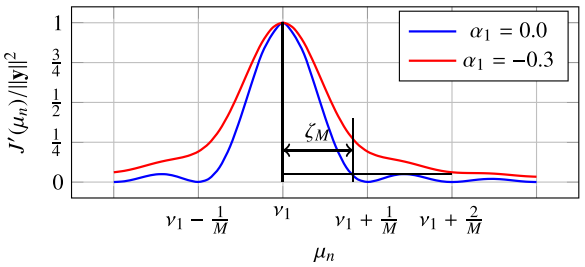


Fig. 2. $J'(\mu_n)$ in the single mode case with $\beta_n = 0$.

exactly one local maximum in the interval $[\nu_1 + k/M, \nu_1 + (k+1)/M]$, $k \neq 0$. Let J_1' be the amplitude of the first sidelobe and $\nu_1 + \zeta_M$ be the value of μ_n such that $J'(\mu_n = \nu_1 + \zeta_M) = J_1'$ in the interval $[\nu_1, \nu_1 + 1/M]$ (we assume[2] that $M > 2$). For the dictionary refinement strategy to converge to the global maximum, it is sufficient to the sparse approximation algorithm to select, at a given level ℓ, an atom whose frequency satisfies $|\mu_{n^*}(\ell) - \nu_1| < \zeta_M < 1/M$, where $\mu_{n^*}(\ell) = \arg\max_n J'(\mu_n)$. Indeed, if $\mu_{n^*}(\ell) \in (\nu_1 - \zeta_M, \nu_1 + \zeta_M)$ then adding two atoms whose frequencies are located on both sides of $\mu_{n^*}(\ell)$ will lead to the selection, at level $\ell + 1$, of an atom that satisfies $|\mu_{n^*}(\ell+1) - \nu_1| \leq |\mu_{n^*}(\ell) - \nu_1|$: the distance between the selected atom and the true frequency is a monotonically decreasing sequence. Finally, the convergence is guaranteed if the initial dictionary contains an atom n such that $|\mu_n(0) - \nu_1| < \zeta_M$, which is satisfied if

$$\max_{n \in \{1,\ldots,N(0)-1\}} |\mu_{n+1}(0) - \mu_n(0)| < 2\zeta_M. \tag{20}$$

given the fact that the sequence $\{\mu_n(0)\}$ covers the interval $[0, 1)$. For $\alpha_1 < 0$, the main lobe of $J'(\mu_n)$ becomes broader and ζ_M larger than for $\alpha_1 = 0$. Consequently, condition (20) is also sufficient for $\alpha_1 < 0$. □

Corollary 1. *In the single tone case, the harmonic dictionary refinement is convergent if the initial frequency grid ($\ell = 0$) is the Fourier grid.*

Proof. Fourier bins are obtained for $N{=}M$ and $\mu_n(0) = (n-1)/M$. Since $\zeta_M > 1/2M$, the proof is straightforward. □

It is important to note that condition (20) is sufficient but not necessary. Moreover, this condition is established when adding a single atom on both sides of the selected one (i.e. $\eta = 1$ in Algorithm 1). When $\eta \gg 1$, the condition may be relaxed and the rate of convergence is expected to be higher.

4.2. Estimating the damping factor: the modal dictionary

Assume that the previous sparse approximation using a harmonic dictionary has converged to select an atom with $\mu_n = \nu_1$. Now, we have to estimate the damping factor α_1. We form a modal dictionary using the damping factor grid and the frequency ν_1, i.e. $q_n = \exp(\beta_n + j2\pi\nu_1)$. Consequently,

$$J'(\beta_n) = \frac{|c_1|^2(1 - e^{2\beta_n})}{1 - e^{2\beta_n M}} \left(\frac{1 - e^{(\alpha_1+\beta_n)M}}{1 - e^{(\alpha_1+\beta_n)}} \right)^2. \tag{21}$$

Theorem 2. *Let y(m) be a single tone (F=1) noiseless signal of length M and* $\mathbf{Q}(0) = [\mathbf{q}_1\mathbf{q}_2\ldots\mathbf{q}_{N(0)}]^T$ *be the initial modal dictionary formed using the frequency* ν_1*, i.e.,* $q_n = \exp(\beta_n(0) + j2\pi\nu_1)$*, where* ν_1 *is the frequency of signal* $\mathbf{y}$*. The columns are sorted in increasing order of* $\beta_n(0)$, $n = 1, 2, \ldots, N$ *and covering the damping factor interval* $(\beta_{\min}, 0]$*. Then the refinement scheme is convergent (i.e.* $\exists\, n$ *s.t.* $\lim_{\ell\to\infty}\beta_n(\ell) = \alpha_1$*) if* $\alpha_1 \in (\beta_{\min}, 0]$.

Proof. Let $g(\beta_n)$ be the derivative of $J'(\beta_n)$ in (21) with respect to β_n. It is easy to check that $g(\beta_n) > 0$ for $\beta_n < \alpha_1$, $g(\beta_n) < 0$ for $\beta_n > \alpha_1$, and $g(\beta_n) = 0$ when $\beta_n = \alpha_1$. In other words, $J'(\beta_n)$ is monotonically increasing before the maximum reached at α_1 and monotonically decreasing after α_1. Therefore, the multigrid algorithm converges to α_1 if $\beta_{\min} < \alpha_1$. □

As a consequence of Theorem 2, the initial modal dictionary can be formed using only two points in the damping factor grid: $\beta_1(0) = \beta_{\min}$ and $\beta_2(0) = 0$.

We can now state that the multigrid algorithm based on two sparse approximations (for frequency and then damping factor) converges in the single tone case under some conditions. Note that in the noisy case when the SNR is sufficiently high, the convergence analysis is still valid as in the noiseless case, and the proposed multigrid sparse scheme for single tone converges to the global maximum of the Fejér kernel. The extension to the single tone *R*-D modal retrieval problem is straightforward and can be performed according to the formulation presented in Section 3.2. The details of this approach (*STSM: Single Tone Sparse Method*) are presented in Algorithm 2. The algorithm takes as input a noisy single tone *R*-D signal, and a couple of integers η_ν and η_α that correspond respectively to the number of frequency and damping factor atoms to be added on both sides of the corresponding selected ones. Next, for each dimension $r = 1, \ldots, R$, we run two tasks to estimate the frequency and then the damping factor: in each step we apply SOMP combined to DICREF using corresponding dictionaries and taking into account the convergence conditions discussed previously. Then parameters of a_r, i.e., ν_r and α_r, are given by the corresponding selected atoms.

Algorithm 2. Single tone sparse method (STSM) based on a multigrid refinement.

input : A tensor $\mathcal{Y} \in \mathbb{C}^{M_1\times\cdots\times M_R}$, $(\eta_\nu, \eta_\alpha) \in \mathbb{N}\times\mathbb{N}$
output: Parameters of the single *R*-D mode: $a_1, \ldots, a_R$
initialization: $(k_\nu, k_\alpha) = (0, 0)$
initialize $\mathbf{d}_\nu^{(0)}$ and $\mathbf{d}_\alpha^{(0)}$ using ζ
for $r = 1 : R$ **do**
  while *halting criterion false* **do**
    $k_\nu = k_\nu + 1$
    $\Omega_\nu^{(k_\nu)} = \text{SOMP}(\mathbf{Q}(\mathbf{d}_\nu^{(k_\nu)}, 0), \mathbf{Y}_{(r)}, \text{Iter} = 1)$
    $\mathbf{d}_\nu^{(k_\nu+1)} = \text{DICREF}(\mathbf{d}_\nu^{(k_\nu)}, \Omega_\nu^{(k)}, \eta_\nu)$
  end
  while *halting criterion false* **do**
    $k_\alpha = k_\alpha + 1$
    $\Omega_\alpha^{(k_\alpha)} = \text{SOMP}\left(\mathbf{Q}\left(\mathbf{d}_\nu^{(k_\nu)}(\Omega_\nu^{(k_\nu)}), \mathbf{d}_\alpha^{(k_\alpha)}\right), \mathbf{Y}_{(r)}\right)$
    $\mathbf{d}_\alpha^{(k_\alpha+1)} = \text{DICREF}(\mathbf{d}_\alpha^{(k_\alpha)}, \Omega_\alpha^{(k)}, \eta_\alpha)$
  end
  $a_r = \exp(\mathbf{d}_\alpha^{(k_\alpha)}(\Omega_\alpha^{(k_\alpha)}) + 2\pi\,\mathbf{d}_\nu^{(k_\nu)}(\Omega_\nu^{(k_\nu)}))$
end
return $a_1, \ldots, a_R$

5. Multiple *R*-D modes estimation

In the multiple tones case, sparse approximation algorithms yield suboptimal solutions when the coherence of the dictionary is high [33]. This is a crucial point because the refinement procedure

[2] The case of $M \leq 2$ is not of practical interest but the theorem is still valid by setting $\zeta_M = \frac{1}{2}$ because $J'(\mu_n)$ is a monotonically decreasing function in the interval $[\nu_1, \min\{\frac{1}{2}, \frac{1}{2} + \nu_1\}]$.

will increase the coherence with increasing ℓ, which may prevent convergence even in the noiseless case. In the following, we present a low complexity algorithm that is accurate and robust in the presence of noise. The idea is to begin by an initialization step where *F single tone* modal signals of order $R-1$ are extracted from the *multiple tones* R-D signal. Then an iterative technique is proposed to improve this decomposition and estimate more accurately the underlying parameters.

It is assumed that the frequencies are distinct in at least one dimension with $M_r > F$. Then dimensions are permuted such that the dimension with distinct frequencies becomes the first one ($r=1$).

5.1. From multiple tones to multiple single-tone signals

According to (5), $\mathcal{Y}$ can be written as [29,30]

$$\mathcal{Y} = \mathcal{I} \underset{1}{\bullet} \mathbf{A}_1 \underset{2}{\bullet} \cdots \underset{R}{\bullet} \mathbf{A}_R \underset{R+1}{\bullet} \mathbf{c}^{\mathsf{T}} \tag{22}$$

$$\mathcal{Y} = \left(\mathcal{I} \underset{2}{\bullet} \mathbf{A}_2 \underset{3}{\bullet} \cdots \underset{R}{\bullet} \mathbf{A}_R \underset{R+1}{\bullet} \mathbf{c}^{\mathsf{T}} \right) \underset{1}{\bullet} \mathbf{A}_1$$
$$\mathcal{Y} = \mathcal{S} \underset{1}{\bullet} \mathbf{A}_1 \tag{23}$$

where $\mathcal{I} \in \mathbb{C}^{F \times F \times \cdots \times F}$ is a tensor of order $R+1$ containing ones on its diagonal and zeros otherwise, and

$$\mathcal{S} = \mathcal{I} \underset{2}{\bullet} \mathbf{A}_2 \underset{3}{\bullet} \cdots \underset{R}{\bullet} \mathbf{A}_R \underset{R+1}{\bullet} \mathbf{c}^{\mathsf{T}} \tag{24}$$

is a complex tensor of order R and size $F \times M_2 \times \cdots \times M_R$. Similar expressions are evoked in, among others, [12]. The matricization of $\mathcal{S}$ along the first dimension is given by [30]

$$\mathbf{S}_{(1)} = \mathbf{I}_{(1)} (\mathbf{c}^{\mathsf{T}} \boxtimes \mathbf{A}_R \boxtimes \cdots \boxtimes \mathbf{A}_2)^{\mathsf{T}} \tag{25}$$

$$\mathbf{S}_{(1)} = \mathbf{I}_F (\mathbf{c}^{\mathsf{T}} \odot \mathbf{A}_R \odot \cdots \odot \mathbf{A}_2)^{\mathsf{T}} \tag{26}$$

where $\mathbf{I}_{(1)}$ and $\mathbf{I}_F$ denote respectively the matricization of $\mathcal{I}$ along the first dimension and the identity matrix of size $F \times F$. Then,

$$\mathbf{S}_{(1)} = (\mathbf{c}^{\mathsf{T}} \odot \mathbf{A}_R \odot \cdots \odot \mathbf{A}_2)^{\mathsf{T}} \tag{27}$$

$$\mathbf{S}_{(1)} = \begin{bmatrix} c_1 (\mathbf{a}_{1,R} \odot \cdots \odot \mathbf{a}_{1,2})^{\mathsf{T}} \\ c_2 (\mathbf{a}_{2,R} \odot \cdots \odot \mathbf{a}_{2,2})^{\mathsf{T}} \\ \vdots \\ c_F (\mathbf{a}_{F,R} \odot \cdots \odot \mathbf{a}_{F,2})^{\mathsf{T}} \end{bmatrix} \tag{28}$$

We can see that each row $f = 1, \ldots, F$ of $\mathbf{S}_{(1)}$ represents the matricization along the first dimension of a sub-tensor $\mathcal{S}_f$ of size $1 \times M_2 \times \cdots \times M_R$ that contains a single $(R-1)$-D tone, where:

$$\mathcal{S}_f = c_f \mathbf{a}_{f,2} \otimes \mathbf{a}_{f,3} \otimes \cdots \otimes \mathbf{a}_{f,R}. \tag{29}$$

The tensor $\mathcal{S}$ can then be written as the concatenation of the F sub-tensors $\mathcal{S}_f$ along the first dimension

$$\mathcal{S} = \mathcal{S}_1 \sqcup_1 \mathcal{S}_2 \sqcup_1 \cdots \sqcup_1 \mathcal{S}_F \tag{30}$$

This is the key property to transform the multiple tone R-D signal into F single tone signals, thus allowing to use the single-tone estimation of Section 4. To do so, we need firstly to estimate $\mathbf{A}_1$.

Now, we show how to estimate the modes of the first dimension, i.e. $\mathbf{A}_1$, using the matricized form of $\mathcal{Y}$ along the first dimension $\mathbf{Y}_{(1)}$ (Eq. (6)). The singular value decomposition (SVD) of $\mathbf{Y}_{(1)}$ yields

$$\mathbf{Y}_{(1)} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^{\mathrm{H}} \tag{31}$$

where matrices $\mathbf{U} \in \mathbb{C}^{M_1 \times L}$ and $\mathbf{V} \in \mathbb{C}^{M_1' \times L}$ are orthonormal and contain respectively the left and right singular vectors of $\mathbf{Y}_{(1)}$, with $L = \min\{M_1, M_1'\}$. $\mathbf{\Sigma}$ is a diagonal matrix containing the singular values σ_i, $i = 1, \ldots, L$, sorted in a decreasing order. As the number of components in $\mathcal{Y}$ is equal to F, we can decompose the SVD of $\mathbf{Y}_{(1)}$ in (31) as follows:

$$\mathbf{Y}_{(1)} = \mathbf{U}_F \mathbf{\Sigma}_F \mathbf{V}_F^{\mathrm{H}} + \mathbf{U}_n \mathbf{\Sigma}_n \mathbf{V}_n^{\mathrm{H}} \tag{32}$$

where $\mathbf{U}_F$ (resp. $\mathbf{V}_F$) stands for the matrix formed with the first F columns of $\mathbf{U}$ (resp. $\mathbf{V}$) and $\mathbf{\Sigma}_F$ contains nonzero singular values on its diagonal $\mathbf{\Sigma}_F = \mathrm{diag}(\sigma_1, \ldots, \sigma_F)$. $\mathbf{U}_n$ (resp. $\mathbf{V}_n$) is formed by the remaining columns associated with zero singular values $\mathbf{\Sigma}_n = \mathbf{0}$. It can be established from (7) and (32) that $\mathbf{A}_1$ and $\mathbf{U}_F$ span the same subspace, and thus there exists an unknown nonsingular matrix $\mathbf{T}$ that satisfies

$$\mathbf{A}_1 = \mathbf{U}_F \mathbf{T}. \tag{33}$$

Denote by $\underline{\mathbf{M}}$ (resp. $\overline{\mathbf{M}}$) the matrix obtained from $\mathbf{M}$ by deleting the first (resp. last) row. By harnessing the Vandermonde structure of $\mathbf{A}_1$, there exists a diagonal matrix

$$\mathbf{D} = \mathrm{diag}(a_{1,1}, \ldots, a_{F,1})$$

such that $\underline{\mathbf{A}}_1 = \overline{\mathbf{A}}_1 \mathbf{D}$. Since $\underline{\mathbf{A}}_1 = \underline{\mathbf{U}}_F \mathbf{T}$ and $\overline{\mathbf{A}}_1 = \overline{\mathbf{U}}_F \mathbf{T}$, then $\underline{\mathbf{U}}_F \mathbf{T} = \overline{\mathbf{U}}_F \mathbf{T} \mathbf{D}$, which proves that matrix $\mathbf{T}$ can be estimated by the eigenvectors of $\underline{\mathbf{U}}_F^{\dagger} \overline{\mathbf{U}}_F$. Therefore, using this estimate of $\mathbf{T}$, exponentials of $\mathbf{A}_1$ can be estimated by (33) where $\mathbf{U}_F$ is obtained from (32).

In the presence of noise, the SVD of $\widetilde{\mathbf{Y}}_{(1)}$ is given by

$$\widetilde{\mathbf{Y}}_{(1)} = \tilde{\mathbf{U}}_F \tilde{\mathbf{\Sigma}}_F \tilde{\mathbf{V}}_F^{\mathrm{H}} + \tilde{\mathbf{U}}_n \tilde{\mathbf{\Sigma}}_n \tilde{\mathbf{V}}_n^{\mathrm{H}}. \tag{34}$$

Due to the noise, all the quantities on the right-hand side (RHS) of (34) may be perturbed versions of those in the RHS of (32). In this case we may express $\tilde{\mathbf{U}}_F$ by $\tilde{\mathbf{U}}_F = \mathbf{U}_F + \Delta \mathbf{U}_F$. Then, neglecting $\tilde{\mathbf{\Sigma}}_n$, an approximation of $\mathbf{Y}_{(1)}$, denoted by $\hat{\mathbf{Y}}_{(1)}$, can be obtained using the first F principal components of the SVD of $\widetilde{\mathbf{Y}}_{(1)}$:

$$\hat{\mathbf{Y}}_{(1)} = \tilde{\mathbf{U}}_F \tilde{\mathbf{\Sigma}}_F \tilde{\mathbf{V}}_F^{\mathrm{H}}. \tag{35}$$

Thereby $\mathcal{S}$ can be estimated from the noisy data and $\hat{\mathbf{A}}_1$ using Eq. (23) as follows:

$$\hat{\mathcal{S}} = \widetilde{\mathcal{Y}} \underset{1}{\bullet} \hat{\mathbf{A}}_1^{\dagger}, \tag{36}$$

then $\hat{\mathcal{S}}_f$, $f = 1, \ldots, F$, are extracted from $\hat{\mathcal{S}}$ according to (30). Each $\mathcal{Y}_f = c_f \mathbf{a}_{f,1} \otimes \cdots \otimes \mathbf{a}_{f,R}$ can be estimated by $\bar{\mathcal{Y}}_f^{(0)} = \hat{\mathcal{S}}_f \underset{1}{\bullet} \hat{\mathbf{a}}_{f,1}$. The sparse multigrid algorithm for single tone (STSM) can be applied on each $\bar{\mathcal{Y}}_f^{(0)}$, $f = 1, \ldots, F$, to estimate the parameters. However, we propose in the following to improve the separated components using an iterative technique.

5.2. Improving the estimation accuracy

It is clear from (36) that, in the noisy case, the error in estimating $\mathcal{S}$ (due to the estimation of $\mathbf{A}_1$) will propagate when estimating the parameters $a_{f,2}, \ldots, a_{f,R}$. Hence, we propose to improve iteratively the mode estimates. The following procedure is used to update estimates at each iteration $i = 0, \ldots, K$

1. apply STSM to estimate $\mathbf{a}_{f,2}, \ldots, \mathbf{a}_{f,R}, f = 1, \ldots, F$

$$\{\hat{a}_{f,2}, \ldots, \hat{a}_{f,R}\} = \mathrm{STSM}(\bar{\mathcal{Y}}_f^{(i)}, \eta_\nu, \eta_\alpha, r = 2, \ldots, R) \tag{37}$$

2. estimate $c_f \mathbf{a}_{f,1}, f = 1, \ldots, F$ by least squares using the already

estimated $\mathbf{a}_{f,2}, \ldots, \mathbf{a}_{f,R}, f = 1, \ldots, F$

$$\widehat{c_f \mathbf{a}_{f,1}} = \bar{\mathbf{Y}}^{(i)}_{f_{(1)}} \left((\hat{\mathbf{a}}_{f,R} \boxtimes \cdots \boxtimes \hat{\mathbf{a}}_{f,2})^{\mathrm{T}} \right)^{\dagger} \tag{38}$$

3. compute $\hat{\mathcal{Y}}_f^{(i)}$

$$\hat{\mathcal{Y}}_f^{(i)} = \widehat{c_f \mathbf{a}_{f,1}} \otimes \hat{\mathbf{a}}_{f,2} \otimes \cdots \otimes \hat{\mathbf{a}}_{f,R} \tag{39}$$

where $\bar{\mathcal{Y}}_f^{(i)} = \hat{\mathcal{Y}}_f^{(i-1)} + \mathcal{R}_{f-1}^{(i)}$, $\mathcal{R}_f^{(i)} = \mathcal{R}_{f-1}^{(i)} + \hat{\mathcal{Y}}_f^{(i-1)} - \hat{\mathcal{Y}}_f^{(i)}$, $f = 1, \ldots, F$, $\mathcal{R}_0^{(i)} \text{def} = \mathcal{R}_F^{(i-1)}$, and $\mathcal{R}_F^{(0)} = \tilde{\mathcal{Y}} - \sum_{f=1}^{F} \hat{\mathcal{Y}}_f$. This iterative scheme will be analyzed in the next section.

Finally, the algorithm we propose (*MTM: Multiple Tones Method*) is summarized in Algorithm 3. Note that no association step of R-D modes is required. The initialization step consists in initializing: (i) $\hat{\mathbf{A}}_1$ and $\hat{\mathcal{S}}$ using (33), (36) and (30), (ii) the estimated single tones $\bar{\mathcal{Y}}_f^{(0)}, f = 1, \ldots, F$.

Note that the columns of $\hat{\mathbf{A}}_1$ are iteratively updated without extracting the related modes, whereas the modes of the other dimensions are extracted at each iteration using (37). Solely after the last iteration (i=K), the parameters of the first dimension are extracted using STSM algorithm. K denotes the maximum number of iterations, which is fixed to 2 in the simulations since no improvement was observed for $K > 2$.

Algorithm 3. Multiple tones method (MTM).

input : A tensor $\tilde{\mathcal{Y}} \in \mathbb{C}^{M_1 \times \cdots \times M_R}$, $(\eta_\nu, \eta_\alpha) \in \mathbb{N} \times \mathbb{N}$
output: Parameters of the multiple R-D modes : $\{a_{f,r}\}_{f=1,r=1}^{F,R}$
initialization:

1. Compute $\hat{\mathbf{A}}_1$ and $\hat{\mathcal{S}}_f$, $f = 1, \ldots, F$ using (33), (36) and (30)
2. $\bar{\mathcal{Y}}_f^{(0)} = \hat{\mathcal{S}}_f \underset{1}{\bullet} \hat{\mathbf{a}}_{f,1}$, $f = 1, \ldots, F$

For $f = 1, \ldots, F$, compute $\hat{\mathcal{Y}}_f^{(0)}$ using (37), (38) and (39)
$\mathcal{R}_F^{(0)} \overset{\text{def}}{=} \mathcal{R}_0^{(1)} = \tilde{\mathcal{Y}} - \sum_{f=1}^{F} \hat{\mathcal{Y}}_f^{(0)}$
for $i = 1 : K$ **do**
  for $f = 1 : F$ **do**
    $\bar{\mathcal{Y}}_f^{(i)} = \hat{\mathcal{Y}}_f^{(i-1)} + \mathcal{R}_{f-1}^{(i)}$
    compute $\hat{\mathcal{Y}}_f^{(i)}$ using (37), (38) and (39)
    $\mathcal{R}_f^{(i)} = \bar{\mathcal{Y}}_f^{(i)} - \hat{\mathcal{Y}}_f^{(i)}$, if $f = F$, then $\mathcal{R}_0^{(i+1)} \overset{\text{def}}{=} \mathcal{R}_F^{(i)}$
  end
end
For $f = 1, \ldots, F$, extract $a_{f,1}$ using
  $a_{f,1} = \text{STSM}(\hat{\mathcal{Y}}_f^{(K)} + \mathcal{R}_F^{(K)}, \eta_\nu, \eta_\alpha, r = 1)$
return $\{\hat{a}_{f,r}\}_{f=1,r=1}^{F,R}$

5.3. Analysis of the algorithm

Following the separation step described in (31)–(36), we can state that the algorithm yields the expected solution when the SNR is sufficiently high. We want now to prove that the second stage (next iterations), in addition to estimating the parameters from the single tones, is also improving the estimation accuracy. The general idea is inspired from greedy forward/backward sparse approximation, where the solution is refined by adding/removing atoms to/from the set of activated atoms. The improvement of the estimates is stated by the following theorem.

Theorem 3. *Assuming that the noise $\mathcal{E}$ is sufficiently small such that the ordering of the singular values in $\boldsymbol{\Sigma}$ in (31) is the same as the ordering of the corresponding singular values when $\mathcal{E} = 0$. Using the procedure expressed by (37), (38) and (39) to estimate $\mathcal{Y}_f$ at iteration $i = 0, \ldots, K$*

$$\hat{\mathcal{Y}}_f^{(i)} = \underset{\mathcal{X} \in \mathcal{H}}{\arg\min} \parallel \bar{\mathcal{Y}}_f^{(i)} - \mathcal{X} \parallel \tag{40}$$

where $\mathcal{H} = \{\mathcal{X} \in \mathbb{C}^{M_1 \times \cdots \times M_R} | \mathcal{X} = \mathbf{b}_1 \otimes \mathbf{b}_2 \otimes \cdots \otimes \mathbf{b}_R, \mathbf{b}_r \in \mathcal{P} \text{ for } r \neq 1\}$ *with* $\mathcal{P} = \{\mathbf{v} \in \mathbb{C}^{M_r} | \mathbf{v} = [1, \nu, \ldots, \nu^{M_r - 1}]^{\mathrm{T}}, \nu = \exp(\beta + j\omega), \beta \in \mathbb{R}^-, \omega \in [0, 2\pi)\}$. *Then, at each iteration i, the residual is decreased:*

$$\left\| \tilde{\mathcal{Y}} - \hat{\mathcal{Y}}^{(i)} \right\| \leq \left\| \tilde{\mathcal{Y}} - \hat{\mathcal{Y}}^{(i-1)} \right\| \tag{41}$$

where $\hat{\mathcal{Y}}^{(i)} = \sum_{f=1}^{F} \hat{\mathcal{Y}}_f^{(i)}$.

Proof. See Appendix A. □

5.4. Identifiability

Based on the assumptions under which Algorithm 3 is operating, the identifiability condition can be stated as $F < M_1$ and $\min\{M_2, \ldots, M_R\} \geq 2$. In [34], the condition is $M_r \geq 4$, $r = 1, \ldots, R$, and $F \leq \left\lfloor \frac{M_1}{2} \right\rfloor \prod_{r=1}^{R} \left\lceil \frac{M_r}{2} \right\rceil$.

We note that, when $M_r \geq 4$, $r = 1, \ldots, R$, the number of identifiable modes is slightly smaller than in [34], but the proposed algorithm is able to outperform the conventional methods in terms of computational complexity and accuracy. In addition, another advantage of the proposed algorithm is clear when the number of samples in one or more dimensions is less than 4 (i.e. $M_r < 4$), where identifiability in [34] is not satisfied. This latter case (i.e. $\exists\, r, M_r < 4$) can be encountered in signal processing applications when the size of one or more diversities (dimensions in our formulation problem) is less than 4.

6. Cramér–Rao lower bounds for R-D cisoids in noise

In this section, we derive the expressions of the CRLB for the parameters of R-D damped exponentials in Gaussian white noise. We then give the CRLB in the cases of single damped and undamped R-D cisoids. We consider the R-D sinusoidal model given in (3). Let

$$\boldsymbol{\theta} = [\omega_{1,1} \ldots \omega_{1,R}\ \omega_{2,1} \ldots \omega_{F,R}\ \alpha_{1,1} \ldots \alpha_{1,R} \alpha_{2,1} \ldots \alpha_{F,R}\ \lambda_1 \ldots \lambda_F\ \phi_1 \ldots \phi_F]^{\mathrm{T}}$$

be the unknown parameter vector. The aim here is to derive the CRLB of the parameters in $\boldsymbol{\theta}$.

The joint probability density function (pdf) of $\tilde{\mathbf{y}}$ is

$$p(\tilde{\mathbf{y}}; \boldsymbol{\theta}) = \frac{1}{(\sigma^2 \pi)^M} \exp\left\{ -\frac{1}{\sigma^2} (\tilde{\mathbf{y}} - \boldsymbol{\mu}(\boldsymbol{\theta}))^{\mathrm{H}} (\tilde{\mathbf{y}} - \boldsymbol{\mu}(\boldsymbol{\theta})) \right\} \tag{42}$$

where $\boldsymbol{\mu}(\boldsymbol{\theta})$ is the noise-free part of $\mathbf{y}$ and

$$\tilde{\mathbf{y}} = [\tilde{y}(1, \ldots, 1, 1), \ldots, \tilde{y}(1, \ldots, 1, M_R), \tilde{y}(1, \ldots, 2, 1), \ldots, \tilde{y}(1, \ldots, 2, M_R), \ldots, \tilde{y}(M_1, \ldots, M_R)]^{\mathrm{T}} \tag{43}$$

The i-th entry of $\boldsymbol{\mu}(\boldsymbol{\theta})$ can be written as:

$$\boldsymbol{\mu}(\boldsymbol{\theta})_i = \sum_{f=1}^{F} c_f \prod_{r=1}^{R} a_{f,r}^{t_{i,r}}, \tag{44}$$

for $i = 1, \ldots, M$, where

$$t_{i,r} = \left\lfloor \frac{i - 1}{\prod_{\ell=r+1}^{R} M_\ell} \right\rfloor \bmod M_r, \tag{45}$$

and $\lfloor \cdot \rfloor$ is the floor function. In the following, we derive the expressions of the CRLB in the general case ($F > 1$) and then we

deduce the result corresponding to a single R-D modal signal ($F=1$).

6.1. Derivation of the CRLB

Given the joint pdf in (42), the (k,l) entry of the Fisher information matrix is [35,36]:

$$[\mathbf{F}(\boldsymbol{\theta})]_{kl} = \frac{2}{\sigma^2}\mathrm{Re}\left\{\left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_k}\right]^{\mathrm{H}} \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_l}\right\}. \tag{46}$$

After some lengthy calculations, the $M \times (2RF + 2F)$ matrix $\partial\boldsymbol{\mu}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ can be expressed as

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \underbrace{[j\mathbf{Z}'\boldsymbol{\Phi}\ \mathbf{Z}'\boldsymbol{\Phi}\ \mathbf{Z}\boldsymbol{\phi}\ j\mathbf{Z}\boldsymbol{\phi}]}_{\mathbf{V}} \cdot \underbrace{\mathrm{blkdiag}(\boldsymbol{\Lambda}, \boldsymbol{\Lambda}, \mathbf{I}_F, \lambda)}_{\mathbf{S}} \tag{47}$$

where

$$\mathbf{Z}' = [\mathbf{Z}'_1, \ldots, \mathbf{Z}'_F] \in \mathbb{C}^{M\times RF}, \quad \text{with } \mathbf{Z}'_f(i,l) = t_{i,l}\prod_{r=1}^{R} a_{f,r}^{t_{i,r}}, \tag{48}$$

$$\boldsymbol{\Lambda} = \mathrm{blkdiag}(\lambda_1\mathbf{I}_R, \ldots, \lambda_F\mathbf{I}_R) \in \mathbb{R}^{RF\times RF}, \tag{49}$$

$$\boldsymbol{\Phi} = \mathrm{blkdiag}(e^{j\phi_1}\mathbf{I}_R, \ldots, e^{j\phi_F}\mathbf{I}_R) \in \mathbb{C}^{RF\times RF}, \tag{50}$$

$$\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_F] \in \mathbb{C}^{M\times F}, \quad \text{with } \mathbf{z}_f(i) = \prod_{r=1}^{R} a_{f,r}^{t_{i,r}}, \tag{51}$$

$$\lambda = \mathrm{diag}([\lambda_1, \ldots, \lambda_F]) \in \mathbb{R}^{F\times F}, \tag{52}$$

$$\boldsymbol{\phi} = \mathrm{diag}([e^{j\phi_1}, \ldots, e^{j\phi_F}]) \in \mathbb{C}^{F\times F}. \tag{53}$$

Finally, the inverse of the Fisher information matrix is

$$\mathbf{F}^{-1}(\boldsymbol{\theta}) = \frac{\sigma^2}{2}\mathbf{S}^{-1}\Big[\mathrm{Re}\{\mathbf{V}^{\mathrm{H}}\mathbf{V}\}\Big]^{-1}\mathbf{S}^{-1} = \frac{\sigma^2}{2}\mathbf{S}^{-1}\mathbf{W}\mathbf{S}^{-1} \tag{54}$$

where $\mathrm{Re}\{\cdot\}$ stands for the real part. The CRLB of θ_k is given by $[\mathbf{F}^{-1}(\boldsymbol{\theta})]_{kk}$. More explicitly, for $f = 1, \ldots, F$ and $r = 1, \ldots, R$:

$$\mathrm{CRLB}(\omega_{f,r}) = \frac{2\sigma^2\mathbf{W}_{R(f-1)+r,R(f-1)+r}}{\lambda_f^2} \tag{55}$$

$$\mathrm{CRLB}(\alpha_{f,r}) = \frac{2\sigma^2\mathbf{W}_{RF+R(f-1)+r,RF+R(f-1)+r}}{\lambda_f^2} \tag{56}$$

$$\mathrm{CRLB}(\lambda_f) = 2\sigma^2\mathbf{W}_{2RF+f,2RF+f} \tag{57}$$

$$\mathrm{CRLB}(\phi_f) = \frac{2\sigma^2\mathbf{W}_{2RF+F+f,2RF+F+f}}{\lambda_f^2} \tag{58}$$

Theorem 4. *For the general R-D exponential process, the CRLB's for $f = 1, \ldots, F$ and $r = 1, \ldots, R$ satisfy*

$$\mathrm{CRLB}(\omega_{f,r}) = \mathrm{CRLB}(\alpha_{f,r}) \tag{59}$$

$$\mathrm{CRLB}(\lambda_f) = \lambda^2\mathrm{CRLB}(\phi_f) \tag{60}$$

Proof. It is based on the special block structure of matrix $\mathrm{Re}\{\mathbf{V}^{\mathrm{H}}\mathbf{V}\}$ (see for instance [35]). □

6.2. Single mode case

In this section, the CRLB's will be simplified in the case of a single R-D modal signal ($F=1$) to obtain more precise details on their parameter dependency. For the sake of simplicity, the subscripts denoting the mode $f=1$ will be omitted. First, assume that $|a_r| = \exp(\alpha_r) < 1$. We shall express the products $\mathbf{Z}'^{\mathrm{H}}\mathbf{Z}'$, $\mathbf{Z}^{\mathrm{H}}\mathbf{Z}$ and $\mathbf{Z}'^{\mathrm{H}}\mathbf{Z}$. After some calculations, we get:

$$[\mathbf{Z}'^{\mathrm{H}}\mathbf{Z}']_{nk} = \prod_{\substack{r=1\\ r\neq n,k}}^{R}\left(\frac{1-|a_r|^{2M_r}}{1-|a_r|^2}\right)\times\begin{cases}\sum_{m=0}^{M_n-1} m|a_n|^{2m}\sum_{m=0}^{M_k-1} m|a_k|^{2m}, & \text{if } n\neq k\\ \sum_{m=0}^{M_n-1} m^2|a_n|^{2m}, & \text{if } n=k\end{cases} \tag{61}$$

$$\mathbf{Z}^{\mathrm{H}}\mathbf{Z} = \prod_{r=1}^{R}\left(\frac{1-|a_r|^{2M_r}}{1-|a_r|^2}\right) \tag{62}$$

$$\left[\mathbf{Z}'^{\mathrm{H}}\mathbf{Z}\right]_n = \prod_{\substack{r=1\\ r\neq n}}^{R}\left(\frac{1-|a_r|^{2M_r}}{1-|a_r|^2}\right)\times\sum_{m=0}^{M_n-1} m|a_n|^{2m}. \tag{63}$$

Denoting $M^{(\alpha)} = \prod_{r=1}^{R}(1-|a_r|^{2M_r})/(1-|a_r|^2)$, $q_1(n) = \sum_{m=0}^{M_n-1} m|a_n|^{2m}/\sum_{m=0}^{M_n-1}|a_n|^{2m}$ and $q_2(n) = \sum_{m=0}^{M_n-1} m^2|a_n|^{2m}/\sum_{m=0}^{M_n-1}|a_n|^{2m}$, we then obtain:

$$[\mathbf{P}]_{nk} = M^{(\alpha)}\times\begin{cases}q_1(n)q_1(k), & \text{if } n\neq k\\ q_2(n), & \text{if } n=k\end{cases} \tag{64}$$

$$\mathbf{G} = M^{(\alpha)} \tag{65}$$

$$[\mathbf{Q}]_n = M^{(\alpha)}q_1(n), \tag{66}$$

and

$$\mathrm{Re}\{\mathbf{V}^{\mathrm{H}}\mathbf{V}\} = \begin{bmatrix}\mathbf{P} & 0 & 0 & \mathbf{Q}\\ 0 & \mathbf{P} & \mathbf{Q} & 0\\ 0 & \mathbf{Q}^{\top} & \mathbf{G} & 0\\ \mathbf{Q}^{\top} & 0 & 0 & \mathbf{G}\end{bmatrix}. \tag{67}$$

The inversion of $\mathrm{Re}\{\mathbf{V}^{\mathrm{H}}\mathbf{V}\}$ yields the following expressions of the CRLB's:

$$\mathrm{CRLB}(\omega_r) = \mathrm{CRLB}(\alpha_r) = \frac{\sigma^2}{2\lambda^2M^{(\alpha)}}\times\frac{(1-|a_r|^2)^2(1-|a_r|^{2M_r})^2}{-M_r^2|a_r|^{2M_r}(1-|a_r|^2)^2+|a_r|^2(1-|a_r|^{2M_r})^2}, \tag{68}$$

$$\frac{\mathrm{CRLB}(\lambda)}{\lambda^2} = \mathrm{CRLB}(\phi) = \frac{\sigma^2}{2\lambda^2M^{(\alpha)}}\left(1+\sum_{r=1}^{R}\frac{q_1^2(r)}{q_2(r)-q_1^2(r)}\right). \tag{69}$$

Finally, for a single R-D purely harmonic signal ($\alpha_r = 0, \forall\, r$), we have $M^{(\alpha)} = \prod_{r=1}^{R} M_r = M$ and taking the limit of the CRLB's when $\alpha_r \to 0$ leads to:

$$\lim_{\alpha_r\to 0}\mathrm{CRLB}(\omega_r) = \frac{6\sigma^2}{\lambda^2M(M_r^2-1)} \tag{70}$$

$$\lim_{\alpha_r \to 0} \frac{\mathrm{CRLB}(\lambda)}{\lambda^2} = \frac{\sigma^2}{2\lambda^2 M}\left(1 + 3\sum_{r=1}^{R}\frac{M_r - 1}{M_r + 1}\right). \tag{71}$$

Hence, for the undamped case, our result in (70) is consistent with [12].

7. Simulation results

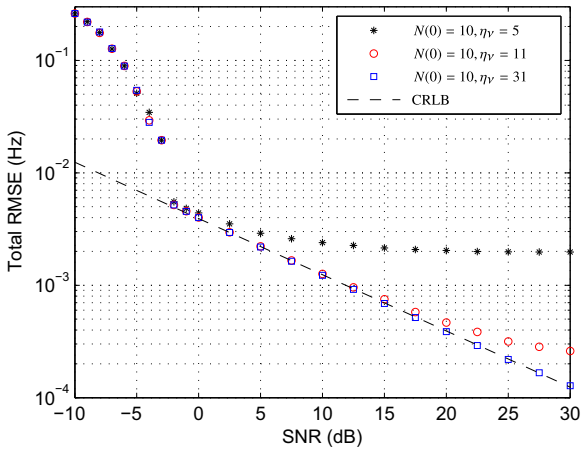
Numerical simulations have been carried out to assess the performances of the proposed method for 2-D and 3-D modal signals in the presence of white Gaussian noise. The performances are measured by the total root-mean square error (RMSE) on estimated parameters and the computational time. The total RMSE is defined as $\mathrm{RMSE}_{\mathrm{total}} = \sqrt{\frac{1}{RF}\mathbb{E}_p\left\{\sum_{r=1}^{R}\sum_{f=1}^{F}(\xi_{f,r} - \hat{\xi}_{f,r})^2\right\}}$ where $\hat{\xi}_{f,r}$ is an estimate of $\xi_{f,r}$, and $\mathbb{E}_p$ is the average on p Monte-Carlo trials. In our simulations, $\xi_{f,r}$ can be either a frequency or a damping factor.

7.1. RMSE for 2-D and 3-D signals

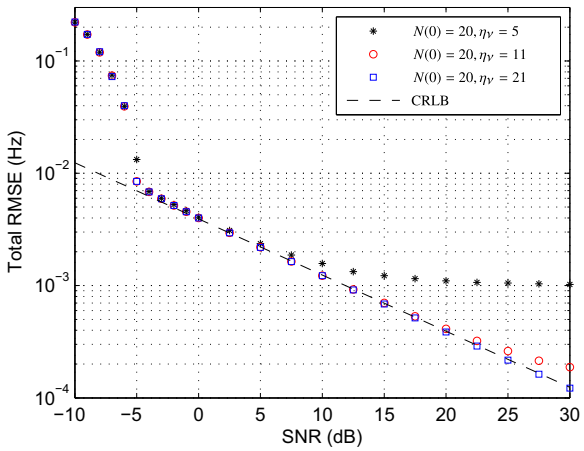
Experiment 1. To show the effectiveness of the multigrid scheme, this experiment presents the results obtained on Signal #1 (see Table 1) with different multigrid levels and different initial grids. Signal #1 is a single tone 2-D modal signal of size 10×10. The number of multigrid levels is fixed to $L=2$, i.e., $\ell = 0, 1, 2$. Then the results are presented as a function of the number of atoms in the initial dictionaries $N(0)$ and the number of atoms η_ν or η_α added at each level ℓ. The results we obtain for the first step, i.e., for the harmonic estimation, are presented in Fig. 3. We can observe that the frequency RMSE obtained with the R-D sparse algorithm can reach the CRLB using a uniform initial harmonic dictionary of 10 atoms and $\eta_\nu = 31$ (Fig. 3a). Fig. 3b shows that the frequency RMSE is improved at low SNR if the initial dictionary contains 20 atoms, and reaches the optimal estimates with $\eta_\nu = 21$. Fig. 4 shows the damping factor RMSE obtained by R-D sparse algorithm using different settings of the initial damping factor dictionary and η_α. We can observe that the damping factor RMSE depends on the number of atoms in the dictionary, the more atoms the better. At low SNR, the RMSE also depends $\beta_{\min}$. Therefore, it is better to choose $\beta_{\min}$ with small absolute value if we have a prior knowledge of the interval of damping factors in the signal. In general, the estimation error is of order $\frac{1}{N(0)\eta^2}$. For instance, in the frequency step estimation, we recommend to chose $N(0)$ to be greater than or equal to $\frac{3}{2}M_r$ if we want a good accuracy at lower SNR levels. Otherwise, we can set $N(0) = M_r$. Once $N(0)$ is set, η can be chosen with respect to the desired accuracy. Let ε be the desired estimation error, then $\varepsilon = \frac{1}{N(0)\eta^2}$ and we can set $\eta = \frac{1}{\sqrt{\varepsilon N(0)}}$.

Table 1
2-D and 3-D parameters of Signal #1 through #4.

Signal	$\nu_{f,1}$	$\alpha_{f,1}$	$\nu_{f,2}$	$\alpha_{f,2}$	$\nu_{f,3}$	$\alpha_{f,3}$	c_f
#1	0.22	−0.011	0.34	−0.015	–	–	1
#2	0.40	−0.01	0.1	−0.01	0.1	−0.01	1
	0.20	−0.01	0.3	−0.15	0.25	−0.01	1
#3	0.28	−0.01	0.31	−0.01	0.22	−0.01	1
	0.12	−0.01	0.45	−0.015	0.11	−0.01	1
	0.20	−0.01	0.31	−0.01	0.11	−0.01	1
#4	0.30	−0.01	0.1	−0.01	0.1	−0.01	1
	0.13	−0.01	0.45	−0.015	0.4	−0.01	1
	0.20	−0.01	0.31	−0.01	0.1	−0.01	1
	0.42	−0.012	0.22	−0.01	0.32	−0.01	1



(a) $N(0) = 10$



(b) $N(0) = 20$

Fig. 3. Frequency RMSE using R-D sparse algorithm with different η_ν. 2-D signal containing a single tone (Signal #1). $(M_1, M_2) = (10, 10)$. 1000 Monte-Carlo trials. (a) The initial harmonic dictionary contains $N(0) = 10$ atoms, (b) $N(0) = 20$ atoms.

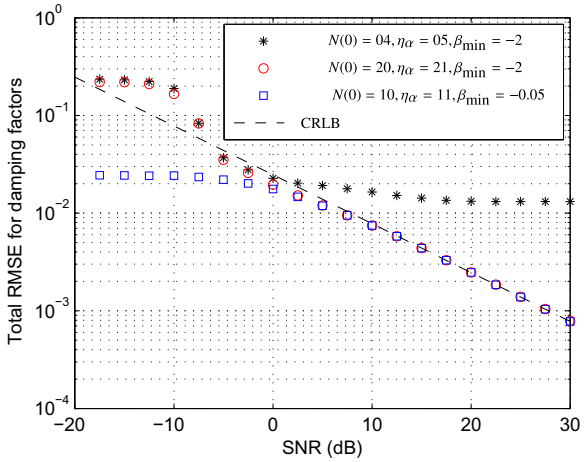


Fig. 4. Damping factor RMSE using R-D sparse algorithm with different η_α, $\beta_{\min}$ and $N(0)$. 2-D signal containing a single tone (Signal #1). $(M_1, M_2) = (10, 10)$. 1000 Monte-Carlo trials.

In the rest of this section, the proposed algorithms are compared

with 2-D ESPRIT [7], Tensor-ESPRIT [11], PUMA [13] and TPUMA [12]. If the R-D signal contains one tone then Algorithm 2 (STSM) is used, otherwise Algorithm 3 (MTM) is used. Thus, to facilitate notation, both proposed algorithms, Algorithm 2 and 3, will be called R-D sparse. For the proposed method, the initial grid used to build the harmonic dictionary is the same for all dimensions; it contains 50 frequency points uniformly distributed over the interval $[0, 1)$ and 10 damping factors $\beta \in [-0.05, 0]$. To simulate a random dictionary, at each run, the frequency grid is perturbed by a small random quantity. As a consequence of Experiment 1, we use the following settings $(L, \eta_\mu, \eta_\beta) = (2, 21, 11)$. The number of iterations in Algorithm 3 is set to $K=2$ because no improvement was observed for $K > 2$.

Since the proposed method is applied directly on data without using spatial smoothing, i.e., it does not require the construction of a large matrix or an augmented order tensor, then a relevant comparison will be with algorithms that do not use spatial smoothing. Thereby, in the next experiments, the proposed algorithm is compared to PUMA [13] and TPUMA [12], which are algorithms that do not require spatial smoothing. We also report comparisons with 2-D ESPRIT [7] and Tensor-ESPRIT [11], which need spatial smoothing.

7.1.1. Single tone R-D modal signal

Experiment 2. This experiment tends to show the efficiency of the proposed algorithm in estimating parameters of single tone R-D modal signals. We use the same signal as before (Signal #1). Our R-D sparse algorithm is compared to 2-D ESPRIT [7] and PUMA [13]. For each level of noise, 1000 Monte-Carlo trials are performed. Fig. 5 shows the obtained results. We can observe that: (i) the proposed algorithm and PUMA reach the CRLB and outperform 2-D ESPRIT, (ii) R-D sparse outperform PUMA in SNR less than 3 dB.

7.1.2. Multiple tones R-D modal signals

Several configurations are studied in the case of multiple tones to compare the proposed algorithm with Tensor-ESPRIT [11] and TPUMA [12]. These configurations (Experiments 3–4) are summarized in Table 2, in which the number of modes and the distance between frequencies in different dimensions are varied. Δ_{Fr} denotes the Rayleigh frequency resolution limit, which has the same value in all dimensions because $M_1 = M_2 = M_3$. In Experiment 5 we examine the case when the size of only one dimension is larger than 4, i.e., the identifiability condition of [34] is not satisfied. The parameters of the used signals are given in Table 1.

Experiment 3. In this experiment, we simulate a 3-D signal (Signal #2) of size $8 \times 8 \times 8$ and containing two modes whose frequencies in each dimension are well separated. Parameters of the signal are given in Table 1. Fig. 6 shows the obtained results. Here, the proposed method performs as TPUMA. Tensor-ESPRIT yields a slightly higher RMSE.

Experiment 4. 3-D signal of size $10 \times 10 \times 10$ containing three 3-D modes (Signal #3). Note that there exists identical modes in two dimensions and frequencies in the first dimension are separated by less than $1/M_1$. The results are shown in Fig. 7. In this experiment, the proposed R-D sparse approach performs better than TPUMA and Tensor-ESPRIT. Observe also that Tensor-ESPRIT outperforms TPUMA in this configuration (close frequencies and identical modes in dimensions 2 and 3).

Experiment 5. Results on Signal #4 of size $10 \times 3 \times 3$ containing 4 modes are given in Fig. 8. We observe that the proposed method

Table 2
Different configurations for experiments 3 and 4.

Experiments	F	Dim. 1	Dim. 2	Dim. 3
Experiment 3	2	$\Delta\nu > \Delta_{\text{Fr}}$	$\Delta\nu > \Delta_{\text{Fr}}$	$\Delta\nu > \Delta_{\text{Fr}}$
Experiment 4	3	$\Delta\nu < \Delta_{\text{Fr}}$	$\exists$ identical modes	$\exists$ identical modes

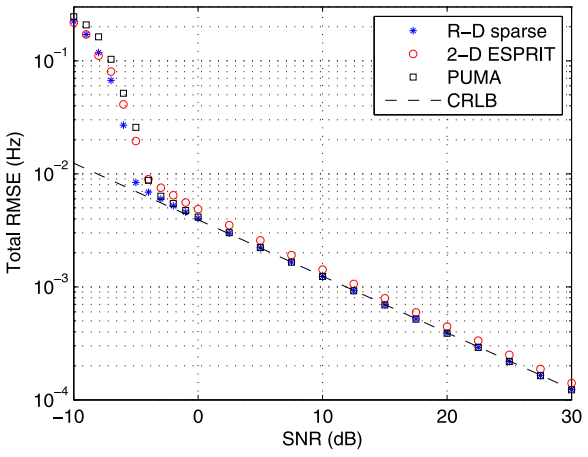


Fig. 5. Frequency total root-mean square error for a 2-D signal containing a single tone (Signal #1). $(M_1, M_2) = (10, 10)$. 1000 Monte-Carlo trials.

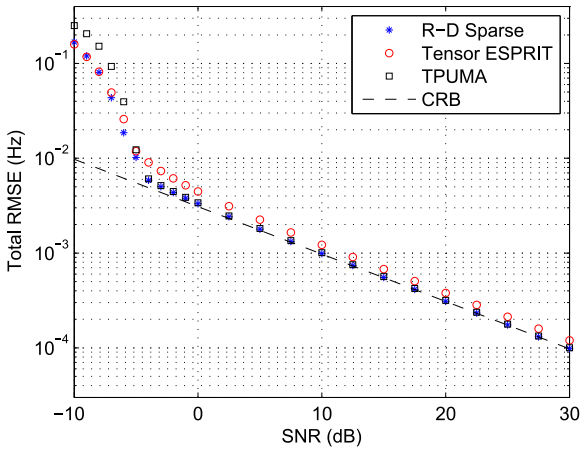


Fig. 6. Frequency total root-mean square error for a 3-D signal containing two 3-D modes (Signal #2). $(M_1, M_2, M_3) = (8, 8, 8)$. 1000 Monte-Carlo.

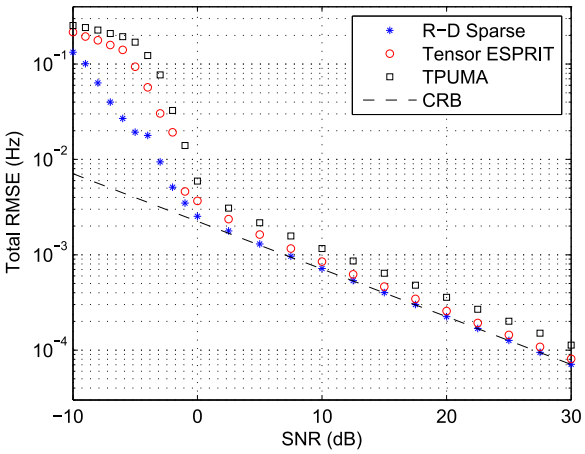


Fig. 7. Frequency total root-mean square error for a 3-D signal containing 3 modes with identical modes in two dimensions (Signal #3), with close modes in the first dimension (0.28,0.12,0.2). $(M_1, M_2, M_3) = (10, 10, 10)$. 200 Monte-Carlo.

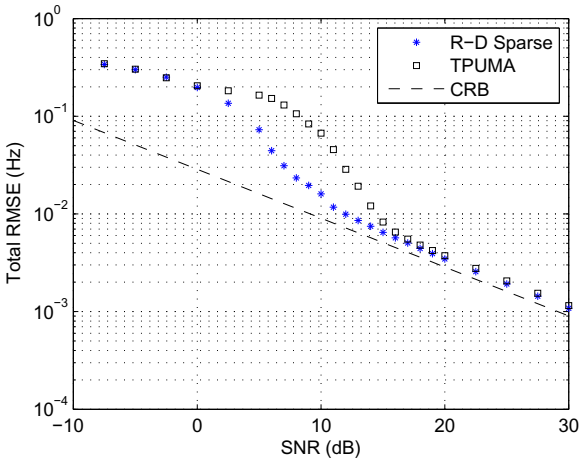


Fig. 8. Frequency total root-mean square error for a 3-D signal containing 4 modes with $(M_1, M_2, M_3) = (10, 3, 3)$ (Signal #4). 200 Monte-Carlo.

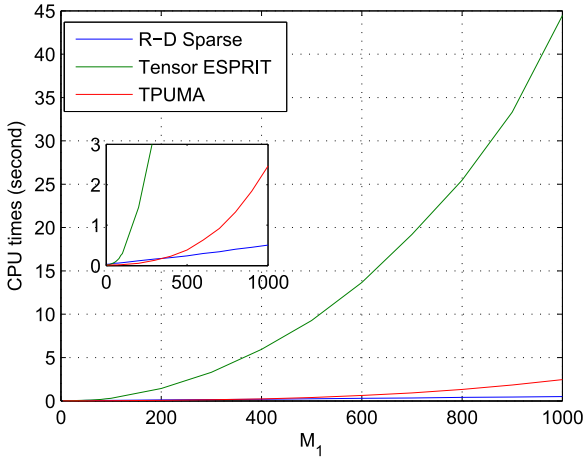


Fig. 9. Average CPU time for a single run under $M_2 = M_3 = 4$ and $F=2$.

outperforms TPUMA mainly at low SNR levels.

7.2. Numerical complexity

It is known that in the case of 1-D signals of size M, OMP costs $O(NFM)$ in terms of multiplications [37]; F is the sparsity (number of components) and N is the number of atoms in the dictionary. For a M-measurements R-D signal, the complexity of the STSM algorithm over a set of L multigrid levels is $O(MNLR)$, assuming that the number of dictionary atoms is maintained constant (equal to N) over all levels. Regarding the approach proposed in Algorithm 3, the main operations are the call of STSM and the update of $\widehat{c_f \mathbf{a}_{f,1}} = \bar{\mathbf{Y}}^{(i)}_{f_{(1)}} \left((\hat{\mathbf{a}}_{f,R} \boxtimes \cdots \boxtimes \hat{\mathbf{a}}_{f,2})^{\mathsf{T}} \right)^{\dagger}$ which has a complexity of $O(M)$ since $\left((\hat{\mathbf{a}}_{f,R} \boxtimes \cdots \boxtimes \hat{\mathbf{a}}_{f,2})^{\mathsf{T}} \right)^{\dagger}$ is a row vector of length $\prod_{r=2}^{R} M_r$ and $\bar{\mathbf{Y}}^{(i)}_{f_{(1)}}$ is a matrix of size $M_1 \times \prod_{r=2}^{R} M_r$. Therefore, the whole complexity of the proposed algorithm is $O\big((NL(F(R-1)K+1)+FK)M\big)$, which is linear in the number of measurements M. The complexity of the Tensor-ESPRIT algorithm with spatial smoothing is mainly related to that of the SVD which is at least $O(k_t F(R+1)PM)$ where k_t is a constant depending on the implementation of the SVD algorithm. Here $P = \prod_{r=1}^{R} P_r$ where $\{P_r\}_{r=1}^{R}$ are design parameters used to get smoothed measurements (see [11]). The accuracy of the estimates provided by ESPRIT depends on these parameters. Since the optimal value for P_r is a fraction of M_r (e.g. [38–40]), the complexity of the SVD step is, in fact, $O(M^2)$. The complexities of PUMA and TPUMA algorithms are $O(M^3)$ and $O(k_t M(R+F-1))+\sum_{r=1}^{R} O(K(F+1)M_r^3)$, respectively. Compared to PUMA and TPUMA, the proposed algorithm has an attractive computational complexity for large size signals. Fig. 9 shows the CPU time results of the proposed, Tensor-ESPRIT and TPUMA algorithms versus M_1 for a 3-D damped signal containing two modes with $M_2 = M_3 = 4$. We observe that the proposed method involves a low computational complexity compared to TPUMA and Tensor-ESPRIT when M_1 is large.

8. Conclusion

We presented an efficient sparse estimation approach for the analysis of multidimensional (R-D) damped or undamped modal signals. The idea consists in exploiting the simultaneous sparse approximation principle to separate this joint estimation problem into R multiple measurements problems. To be able to handle large size signals and yield accurate estimates, a multigrid dictionary refinement scheme is associated with the simultaneous orthogonal matching pursuit (SOMP) algorithm. We gave the convergence proof of the refinement procedure in the single tone case. Then, for the general multiple tones R-D case, the signal tensor model is decomposed in order to handle each tone separately in an iterative scheme so that the pairing of the R-D parameters is automatically achieved. Also, the CRLB of the R-D modal signal parameters were derived. The tests performed on simulated signals showed that the proposed algorithm attains the CRLB and outperforms state-of-the-art subspace algorithms. We also have shown that the complexity of the algorithm is linear with respect to the number of measurements, which allows the processing of large size signals. Finally, it is worth mentioning that this approach can be straightforwardly applied to other multidimensional array processing problems.

Acknowledgment

The authors would like to thank PhD Weize Sun for providing them with the TPUMA algorithm code.

Appendix A. Proof of Theorem 3

We begin the proof by introducing the following lemma.

Lemma 1. *Consider $\widetilde{\mathcal{Y}} = \mathcal{Y} + \Delta\mathcal{Y}$, where $\widetilde{\mathcal{Y}}$ is the perturbed version of the data tensor $\mathcal{Y}$ and $\Delta\mathcal{Y}$ is the perturbation. Assuming that $\Delta\mathcal{Y}$ is sufficiently small such that the ordering of the F singular values in $\boldsymbol{\Sigma}$ in (31) is the same as the ordering of the corresponding singular values when $\Delta\mathcal{Y} = 0$. Then the perturbation $\Delta\mathcal{Y}_f$ contains a linear combination of all $\mathcal{Y}_f$, $f = 1, \ldots, F$:*

$$\Delta\mathcal{Y}_f = \mathcal{D}_f + \sum_{i=1}^{F} v_{f,i}\mathcal{Y}_i$$

where $\mathbf{v}_f^{\mathsf{T}} = [v_{f,1}, \ldots, v_{f,F}] = \Delta\mathbf{A}_1^{\dagger}(f,:)\mathbf{A}_1$ *and* $\mathcal{D}_f = \Delta\mathcal{Y} \bullet_1 \mathbf{a}_{f,1}\mathbf{A}_1^{\dagger}(f,:) + \mathcal{Y}_{s,f} \bullet_1 \Delta\mathbf{a}_{f,1}$.

Proof. From (36) $\mathcal{S} = \mathcal{Y} \bullet_1 \mathbf{A}_1^{\dagger}$, we differentiate and obtain

$$\Delta\mathcal{S} = \Delta\mathcal{Y} \bullet_1 \mathbf{A}_1^{\dagger} + \mathcal{Y} \bullet_1 \Delta\mathbf{A}_1^{\dagger}$$

Then, we get $\Delta\mathcal{S}_f = \Delta\mathcal{Y} \bullet_1 \mathbf{A}_1^{\dagger}(f,:) + \sum_{p=1}^{F} v_{f,p}\mathcal{S}_p$. $\mathcal{Y}_f$ is estimated using

$\mathcal{Y}_f = \mathcal{S}_f \bullet_1 \mathbf{a}_{f,1}$, we differentiate and obtain $\Delta\mathcal{Y}_f = \sum_{p=1}^{F} v_{f,p}\mathcal{Y}_p + \mathcal{S}_f \bullet_1 \Delta\mathbf{a}_{f,1} + \Delta\mathcal{Y} \bullet_1 \mathbf{a}_{f,1}\mathbf{A}_1^{\dagger}(f,:)$. □

Using the previous lemma

$$\bar{\mathcal{Y}}_f^{(0)} = \mathcal{S}_f \bullet_1 (\mathbf{a}_{f,1} + v_{f,f}\mathbf{a}_{f,1} + \Delta\mathbf{a}_f) + \sum_{p=1,p\neq f}^{F} v_{f,p}\mathcal{Y}_p + \Delta\mathcal{Y} \bullet_1 \mathbf{a}_{f,1}\mathbf{A}_1^{\dagger}(f,:)$$

Therefore, $\mathbf{a}_{f,2}, \ldots, \mathbf{a}_{f,R}, f = 1, \ldots, F$ can be estimated using STSM algorithm since

$$\bar{\mathbf{Y}}_{f_{(r)}}^{(0)} = c_f \mathbf{a}_{f,r}(\mathbf{a}_{f,R} \boxtimes \cdots \boxtimes \mathbf{a}_{f,r+1} \boxtimes \mathbf{a}_{f,r-1} \boxtimes \cdots \boxtimes (\mathbf{a}_{f,1} + v_{f,f}\mathbf{a}_{f,1} + \Delta\mathbf{a}_{f,1})) + \left(\sum_{p=1,p\neq f}^{F} v_{f,p}\mathcal{Y}_p + \Delta\mathcal{Y} \bullet_1 \mathbf{a}_{f,1}\mathbf{A}_1^{\dagger}(f,:) \right)_{(r)}.$$

Since $\bar{\mathbf{Y}}_{f_{(1)}}^{(0)}$ has the following form

$$\bar{\mathbf{Y}}_{f_{(1)}}^{(0)} = c_f(\mathbf{a}_{f,1} + v_{f,f}\mathbf{a}_{f,1} + \Delta\mathbf{a}_{f,1})(\mathbf{a}_{f,R} \boxtimes \cdots \boxtimes \mathbf{a}_{f,2}) + \left(\sum_{p=1,p\neq f}^{F} v_{f,p}\mathcal{Y}_p + \Delta\mathcal{Y} \bullet_1 \mathbf{a}_{f,1}\mathbf{A}_1^{\dagger}(f,:) \right)_{(1)},$$

we estimate $c_f\mathbf{a}_{f,1}$ by least squares once $\mathbf{a}_{f,2}, \ldots, \mathbf{a}_{f,R}$ are estimated using STSM

$$\widehat{c_f\mathbf{a}_{f,1}} = \min_{\mathbf{a}} \| \bar{\mathcal{Y}}_f^{(0)} - \mathbf{a} \otimes \hat{\mathbf{a}}_{f,2} \otimes \cdots \otimes \hat{\mathbf{a}}_{f,R} \| = \bar{\mathbf{Y}}_{f_{(1)}}^{(0)} \left((\hat{\mathbf{a}}_{f,R} \boxtimes \cdots \boxtimes \hat{\mathbf{a}}_{f,2})^{\top} \right)^{\dagger}$$

So, we put $\hat{\mathcal{Y}}_f^{(0)} = \widehat{c_f\mathbf{a}_{f,1}} \otimes \hat{\mathbf{a}}_{f,2} \otimes \cdots \otimes \hat{\mathbf{a}}_{f,R}$ and $\mathcal{R}_f = \bar{\mathcal{Y}}_f^{(0)} - \hat{\mathcal{Y}}_f^{(0)}$. Therefore, the procedure to estimate $\mathcal{Y}_f$ at iteration $i = 0, \ldots, K$ can be summarized in (37), (38) and (39). Note that this procedure is optimal because STSM and the least squares are optimal when they are used to estimate $\mathbf{a}_{f,2}, \ldots, \mathbf{a}_{f,R}, f = 1, \ldots, F$ and $c_f\mathbf{a}_{f,1}, f = 1, \ldots, F$, respectively.

Now we present the technique for improving the estimation of $\mathcal{Y}_f$. Let $\mathcal{R}_F^{(0)} = \mathcal{R}_0^{(1)} = \tilde{\mathcal{Y}} - \sum_{f=1}^{F} \hat{\mathcal{Y}}_f^{(0)}$ and

$$\hat{\mathcal{Y}}_f^{(1)} = \arg\min_{\mathcal{X}\in\mathcal{H}} \| \hat{\mathcal{Y}}_f^{(0)} + \mathcal{R}_{f-1}^{(1)} - \mathcal{X} \| \tag{A.1}$$

where $\mathcal{R}_f^{(1)} = \hat{\mathcal{Y}}_f^{(0)} + \mathcal{R}_{f-1}^{(1)} - \hat{\mathcal{Y}}_f^{(1)}, f = 1, \ldots, F$, and $\hat{\mathcal{Y}}_f^{(i)}$ is an improved estimate of $\mathcal{Y}_f$. We follow the same procedure as described in Eqs. (37), (38) and (39) to calculate $\hat{\mathcal{Y}}_f^{(1)}$.

We can state that there is improvement in the estimation of $\mathcal{Y}_f$ if

$$\left\| \tilde{\mathcal{Y}} - \sum_{f=1}^{F} \hat{\mathcal{Y}}_f^{(1)} \right\| \leq \left\| \tilde{\mathcal{Y}} - \sum_{f=1}^{F} \hat{\mathcal{Y}}_f^{(0)} \right\| \tag{A.2}$$

We have $\| \mathcal{R}_0^{(1)} \| = \| \bar{\mathcal{Y}}_1^{(0)} - \hat{\mathcal{Y}}_1^{(0)} + \sum_{f=2}^{F} \mathcal{R}_f + \mathcal{V} \|$ where $\mathcal{V} = \tilde{\mathcal{Y}} - \bar{\mathcal{Y}}$ and $\bar{\mathcal{Y}} = \sum_{f=1}^{F} \bar{\mathcal{Y}}_f^{(0)}$. It can be verified that

$$\| \mathcal{R}_f^{(1)} \| = \left\| \left(\bar{\mathcal{Y}}_f^{(0)} + \sum_{p=1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(1)}) + \sum_{p=f+1}^{F} \mathcal{R}_p + \mathcal{V} \right) - \hat{\mathcal{Y}}_f^{(1)} \right\|$$

$$\| \mathcal{R}_{f-1}^{(1)} \| = \left\| \left(\bar{\mathcal{Y}}_f^{(0)} + \sum_{p=1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(1)}) + \sum_{p=f+1}^{F} \mathcal{R}_p + \mathcal{V} \right) - \hat{\mathcal{Y}}_f^{(0)} \right\|$$

However, from Eq. (A.1), $\hat{\mathcal{Y}}_f^{(1)}$ is the minimizer with respect to $\mathcal{X} \in \mathcal{H}$ of

$$\left\| \left(\bar{\mathcal{Y}}_f^{(0)} + \sum_{p=1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(1)}) + \sum_{p=f+1}^{F} \mathcal{R}_p \right) - \mathcal{X} \right\|$$

Therefore, $\| \mathcal{R}_f^{(1)} \| \leq \| \mathcal{R}_{f-1}^{(1)} \|, f = 1, \ldots, F$. As a consequence, $\| \mathcal{R}_F^{(1)} \| \leq \| \mathcal{R}_F^{(0)} \|$, which we are seeking in expression (A.2). Similarly, we can prove that $\| \mathcal{R}_F^{(i)} \| \leq \| \mathcal{R}_F^{(i-1)} \|, i > 1$, using the general forms of $\mathcal{R}_f^{(i)}$ and $\mathcal{R}_{f-1}^{(i)}$

$$\mathcal{R}_f^{(i)} = \left(\bar{\mathcal{Y}}_f^{(0)} + \sum_{p=1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(i)}) + \sum_{p=f+1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(i-1)}) + \mathcal{V} \right) - \hat{\mathcal{Y}}_f^{(i)}$$

$$\mathcal{R}_{f-1}^{(i)} = \left(\bar{\mathcal{Y}}_f^{(0)} + \sum_{p=1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(i)}) + \sum_{p=f+1}^{f-1} (\bar{\mathcal{Y}}_p^{(0)} - \hat{\mathcal{Y}}_p^{(i-1)}) + \mathcal{V} \right) - \hat{\mathcal{Y}}_f^{(i-1)}$$

which we are seeking in (41).

References

[1] P. Stoica, R. Moses, Introduction to Spectral Analysis, Prentice Hall, Upper Saddle River, NJ, 1997.
[2] A.B. Gershman, N.D. Sidiropoulos, Space–time Processing for MIMO Communications, Wiley Online Library (2005).
[3] D. Nion, S. Sidiropoulos, Tensor algebra and multidimensional retrieval in signal processing for MIMO radar, IEEE Trans. Signal Process. 58 (1) (2010) 5693–5705.
[4] Y. Li, J. Razavilar, K.J.R. Liu, A high-resolution technique from multidimensional NMR spectroscopy, IEEE Trans. Biomed. Eng. 45 (1) (1998) 78–86.
[5] J. Sacchini, W. Steedly, R. Moses, Two-dimensional prony modeling and parameter estimation, IEEE Trans. Signal Process. 41 (11) (1993) 3127–3137.
[6] Y. Hua, Estimating two-dimensional frequencies by matrix enhancement and matrix pencil, IEEE Trans. Signal Process. 40 (9) (1992) 2267–2280.
[7] S. Rouquette, M. Najim, Estimation of frequencies and damping factors by two-dimensional ESPRIT type methods, IEEE Trans. Signal Process. 49 (1) (2001) 237–245.
[8] K. Mokios, N. Sidiropoulos, M. Pesavento, C. Mecklenbrauker, On 3-D harmonic retrieval for wireless channel sounding, in: Proceedings of IEEE ICASSP, Montreal, Canada, 2004, pp. ii89–ii92.
[9] J. Liu, X. Liu, An eigenvector-based approach for multidimensional frequency estimation with improved identifiability, IEEE Trans. Signal Process. 54 (12) (2006) 4543–4556.
[10] J. Liu, X. Liu, X. Ma, Multidimensional frequency estimation with finite snapshots in the presence of identical frequencies, IEEE Trans. Signal Process. 55 (2007) 5179–5194.
[11] M. Haardt, F. Roemer, G. Del Galdo, Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems, IEEE Trans. Signal Process. 56 (7) (2008) 3198–3213.
[12] W. Sun, H.C. So, Accurate and computationally efficient tensor-based subspace approach for multi-dimensional harmonic retrieval, IEEE Trans. Signal Process. 60 (10) (2012) 5077–5088.
[13] H. So, F. Chan, W. Lau, C. Chan, An efficient approach for two-dimensional parameter estimation of a single-tone, IEEE Trans. Signal Process. 58 (4) (2010) 1999–2009.
[14] L. Huang, Y. Wu, H. So, Y. Zhang, L. Huang, Multidimensional sinusoidal frequency estimation using subspace and projection separation approaches, IEEE Trans. Signal Process. 60 (10) (2012) 5536–5543.
[15] C. Lin, W. Fang, Efficient multidimensional harmonic retrieval: a hierarchical signal separation framework, IEEE Signal Process. Lett. 20 (5) (2013) 427–430.
[16] S. Sahnoun, E.-H. Djermoune, D. Brie, Sparse multigrid modal estimation: initial grid selection, in: Proceedings of European Signal Processing Conference (EUSIPCO-2012), 2012, pp. 450–454.
[17] M. Goodwin, M. Vetterli, Matching pursuit and atomic signal models based on recursive filter banks, IEEE Trans. Signal Process. 47 (7) (1999) 1890–1902.
[18] D. Malioutov, M. Cetin, A. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process. 53 (8) (2005) 3010–3022.
[19] P. Stoica, P. Babu, J. Li, SPICE: a sparse covariance-based estimation method for array processing, IEEE Trans. Signal Process. 59 (2) (2011) 629–638.
[20] S. Sahnoun, E.H. Djermoune, C. Soussen, D. Brie, Sparse multidimensional modal analysis using a multigrid dictionary refinement, EURASIP J. Adv. Signal Process (1) (2012) 1–10, http://dx.doi.org/10.1186/1687-6180-2012-60.
[21] S. Sahnoun, E.-H. Djermoune, D. Brie, Sparse modal estimation of 2-D NMR signals, in: Proceedings of IEEE ICASSP, Vancouver, Canada, 2013, pp. 8751–8755.
[22] J. Sward, S.I. Adalbjornsson, A. Jakobsson, High resolution sparse estimation of exponentially decaying signals, in: Proceedings of IEEE ICASSP, Florence, Italy, 2014, pp. 7203–7207.
[23] S.I. Adalbjörnsson, J. Swärd, A. Jakobsson, High resolution sparse estimation of exponentially decaying two-dimensional signals, in: Proceedings of European Signal Processing Conference (EUSIPCO), EURASIP, 2014.
[24] G. Tang, B.N. Bhaskar, P. Shah, B. Recht, Compressed sensing off the grid, IEEE Trans. Inf. Theory 59 (11) (2013) 7465–7490.
[25] Z. Tan, Discrete and continuous sparse recovery methods and their applications (Ph.D. thesis), Washington University in St. Louis, 2015.
[26] J. Tropp, A. Gilbert, M. Strauss, Algorithms for simultaneous sparse

approximation: part i: greedy pursuit, Signal Process. 86 (2006) 572–588.
[27] R. Boyer, Deterministic asymptotic Cramér–Rao bound for the multi-dimensional harmonic model, Signal Process. 88 (12) (2008) 2869–2877.
[28] M. Clark, L. Scharf, Two-dimensional modal analysis based on maximum likelihood, IEEE Trans. Signal Process. 42 (6) (1994) 1443–1451.
[29] P. Comon, Tensors: a brief introduction, IEEE Signal Process. Mag. 31 (3) (special issue on BSS. hal-00923279) (2014) 44–53.
[30] T. Kolda, B. Bader, Tensor decompositions and applications, SIAM Rev. 51 (3) (2009) 455–500.
[31] A. Rakotomamonjy, Surveying and comparing simultaneous sparse approximation (or group-Lasso) algorithms, Signal Process. 91 (7) (2011) 1505–1526.
[32] M.F. Duarte, R.G. Baraniuk, Spectral compressive sensing, Appl. Comput. Harmon. Anal. 35 (1) (2013) 111–129.
[33] J. Tropp, Greed is good: algorithmic results for sparse approximation, IEEE Trans. Inf. Theory 50 (2004) 2231–2242.
[34] T. Jiang, N.D. Sidiropoulos, J.M. Ten Berge, Almost-sure identifiability of multidimensional harmonic retrieval, IEEE Trans. Signal Process. 49 (9) (2001) 1849–1859.
[35] Y.-X. Yao, S. Pandit, Cramér–Rao lower bounds for a damped sinusoidal process, IEEE Trans. Signal Process. 43 (4) (1995) 878–885.
[36] S. Kay, Fundamentals of Statistical Signal Processing. Estimation Theory, Prentice Hall International Editions, Englewood Cliffs, New Jersey, 1993.
[37] J. Tropp, S. Wright, Computational methods for sparse solution of linear inverse problems, Proc. IEEE 98 (6) (2010) 948–958.
[38] A. Lemma, A.-J. Van der Even, E. Deprettere, Analysis of joint angle-frequency estimation using ESPRIT, IEEE Trans. Signal Process. 51 (2003) 1264–1283.
[39] E.-H. Djermoune, M. Tomczak, Perturbation analysis of subspace-based methods in estimating a damped complex exponential, IEEE Trans. Signal Process. 57 (11) (2009) 4558–4563.
[40] Y. Hua, T. Sarkar, Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise, IEEE Trans. Acoust. Speech Signal Process. 38 (1990) 814–824.

A.2 Regularization Parameter Estimation for Non-Negative Hyperspectral Image Deconvolution

Y. Song, D. Brie, **E.-H. Djermoune**, S. Henrot. *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5316-5330, 2016.

Cet article est consacré à la présentation d'une méthode permettant de sélectionner automatiquement les coefficients de régularisation en déconvolution d'images hyperspectrales avec des pénalités quadratiques et une contrainte de positivité. Deux critères sont étudiés : le minumum distance criterion *(MDC) et le* maximum curvature criterion *(MCC).*

Regularization Parameter Estimation for Non-Negative Hyperspectral Image Deconvolution

Yingying Song, *Student Member, IEEE*, David Brie, *Member, IEEE*, El-Hadi Djermoune, *Member, IEEE*, and Simon Henrot

***Abstract*— This paper aims at studying a method to automatically estimate the regularization parameters of non-negative hyperspectral image deconvolution methods. The deconvolution problem is formulated as a multi-objective optimization problem and the properties of the corresponding response surface are studied. Based on these properties, the minimum distance criterion (MDC) and the maximum curvature criterion (MCC) are proposed to estimate regularization parameters especially for the non-negativity constrained deconvolution problem. MDC has good theoretical properties (convexity and uniqueness) but requires to choose a reference point. On the contrary, MCC does not need to choose any reference point but does not have interesting theoretical properties. A grid-search-based approach to minimize the computational cost of MDC and MCC is proposed. It results in fast approaches to estimate the regularization parameters. Based on simulated 2D images, the proposed approaches are compared with the state-of-the-art methods, confirming the effectiveness of the MDC and MCC for the non-negativity constrained image deconvolution problem. In the case of non-negative hyperpsectral image deconvolution, the fast MDC yields better performances than the fast MCC. An application to real-world hyperspectral fluorescence microscopy images is also provided; it confirms the superiority of MDC.**

***Index Terms*— Non-negative hyperspectral image deconvolution, multi-objective optimization, regularization parameter estimation.**

I. Introduction

A HYPERSPECTRAL image can be viewed as a stack of images obtained for different wavelengths. The observed images may suffer from degradation due to the measuring device, resulting in a convolution or blurring of the images. Hyperspectral image deconvolution consists in removing the blur to restore the original images at best. This problem arises in a number of applications including microscopy [1]–[3], astronomy [4]–[6] and industrial hyperspectral imaging systems [7], [8]. Actually, hyperspectral image deconvolution is required as soon as the spatial resolution has to be increased (super-resolution) [9]. Similar problems also arise in x-ray fluorescence tomography [10], [11], even if the problem at hand is not a deconvolution problem but rather a reconstruction problem.

The deconvolution of hyperspectral images is an ill-posed inverse problem which suffers from instability. To recover accurately the original images, it is necessary to resort to some regularization techniques. This can be done by formulating the problem as the minimization of a penalized criterion incorporating prior information enforcing the spatial and spectral regularity as well as the non-negativity of the image to recover. Different hyperspectral image deconvolution methods were proposed in [3], [5], and [12]–[14]. They all consider separable spatial and spectral regularization terms. The effective implementation of such methods is hampered by the choice of the regularization parameters. In general, this choice is made by successive trials which can be highly time consuming. Here, we focus on Tikhonov-like hyperspectral image deconvolution with non-negativity constraint proposed in [13].

A classical approach to estimate a single regularization parameter of Tikhonov-based deconvolution is the generalized cross-validation method [15]. It was used for choosing the regularization parameter in image deconvolution in [16]. The L-curve presented in [17] and [18] is also a method for selecting a single regularization parameter of the Tikhonov-based deconvolution. Plotting in a log-log scale the data fitting term versus the penalty term yields a curve which exhibits a corner. The curvature is expected to reach a maximum value yielding an estimated regularization parameter which provides an acceptable trade-off between these two terms. However, the L-curve approach has some undesirable properties discussed in [19] and [20]. In particular, it is not convex and the maximum curvature is not unique. The L-hypersurface as a multi-objective extension of the L-curve for selecting multiple regularization parameters was introduced in [21]. However, since the curvature is not uniquely defined, the maximum curvature approach is not an appropriate choice. Thus, in [21],

Manuscript received January 11, 2016; revised May 31, 2016 and July 13, 2016; accepted August 2, 2016. Date of publication August 18, 2016; date of current version September 16, 2016. This work was supported in part by the FUI AAP 2015 Trispirabois Project and in part by the Conseil Régional de Lorraine. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Guoliang Fan.

Y. Song, D. Brie, and E.-H. Djermoune are with the Centre de Recherche en Automatique de Nancy and the Centre National de la Recherche Scientifique, Université de Lorraine, F-54506 Vandœuvre-lès-Nancy, France (e-mail: yingying.song@univ-lorraine.fr; david.brie@univ-lorraine.fr; el-hadi.djermoune@univ-lorraine.fr).

S. Henrot is with the Grenoble Images Parole Signal Automatique Laboratory, Signal and Image Department, 38402 SaintMartind'Hères, France (e-mail: simon.henrot@gmail.com).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material includes a document that aims at evaluating the performances of the proposed MDC and MCC for different types of hyperspectral images. Also, it gives an application of these approaches on real data. The total size of the files is 756 kB. Contact yingying.song@univ-lorraine.fr for further questions about this work.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2601489

the minimum distance criterion (MDC), which was already introduced in [22] for the bi-objective case, is applied to the L-hypersurface for estimating the regularization parameters; this leads to a simple fixed-point iterative algorithm for computing regularization parameters in both bi-objective and multi-objective cases. But this approach can only be applied to the unconstrained Tikhonov-based deconvolution.

The goal of the present paper is to propose a general approach to estimate the regularization parameters of hyperspectral image deconvolution methods formulated as a convex multi-objective minimization problem. A key point is that it can be used indifferently for unconstrained and constrained problems. Addressing deconvolution as a multi-objective optimization problem is not very common. To the best of our knowledge, [23] is the first work mentioning the link between the L-curve and the multi-objective optimization. It is also mentioned that the use of the log-log scale results in a loss of convexity of the L-curve. Recently, [24] formulated the basis pursuit as a convex bi-objective optimization problem and proved that the corresponding Pareto front is convex and continuously differentiable over all points of interest. In fact, the Pareto front of basis pursuit is strongly connected to the regularization path for which a continuation-based approach allows fast calculation of the set of solutions when the regularization parameter is varying from 0 to $+\infty$ [25]. Two important results are proved in this paper. Firstly, the multi-objective criterion being composed of convex cost functions, its response surface is proved to be convex. This property holds for both the unconstrained and constrained cases. Secondly, as far as we know, no work is explicitly analyzing the impact of the non-negativity constraint on the regularization parameter estimation of the deconvolution algorithm. Here, we also prove that the non-negativity constraint results in a folding of the response surface. The beneficial consequences of these two properties on the regularization parameter estimation are discussed. This will be supported by extensive simulations aiming at evaluating the Mean Squares Error (MSE) as a function of the Signal to noise Ratio (SNR).

The paper is organized as follows: in section II, we present the non-negative hyperspectral image deconvolution problem. In section III, it is formulated as a multi-objective optimization problem and the properties of the corresponding response surface are studied. In section IV, to estimate the regularization parameters, the maximum curvature criterion (MCC) and the MDC (directly applied to the response surface) are proposed and their properties are studied. To reduce the computational burden of MCC and MDC, a grid-search strategy is proposed: it is proved to be convergent for the MDC but not for the MCC. In section V and in the supplementary material [26], numerical experiments allowing to assess the performances of the proposed approaches and to compare them with state-of-the-art methods are presented. Finally, these approaches are applied to hyperspectral fluorescence microscopy data.

II. Hyperspectral Image Deconvolution

Hyperspectral imaging consists in observing a spatial scene at several wavelengths. Physically, such an image can be obtained as a stack of two-dimensional (2D) images equipped with optical filters or as a collection of one-dimensional (1D) spectra acquired by a spectrometer. Hyperspectral imaging is used in a wide range of applications including remote sensing [27], chemistry [28], [29], food science [30], biology [31] and medical imaging [32]. Among the different spectroscopic techniques allowing to produce hyperspectral images, we can mention Infra Red (IR), Raman [33] and fluorescence [34] microscopies. The problem at hand aims at removing the blur affecting the observed images. Such a blurring arises, for example, when we want to increase the spatial resolution of the imaging spectrometer. To do that, it is necessary to choose a spatial sampling lower than the instrument resolution.

A. Discrete Representation of the Blurred Images

The unknown hyperspectral image is denoted by $\mathbf{X}$ and the observed image by $\mathbf{Y}$. Considering that the discrete image $\mathbf{X}$ has L wavelengths $\lambda_1, \ldots, \lambda_L$, it can be seen as a stack of images $\{\mathbf{X}_l, l = 1 \ldots L\}$. $\mathbf{X}_l$ is a matrix of size $N_1 \times N_2$. By concatenating the columns of each image $\mathbf{X}_l$, the hyperspectral image can be reorganized into L vectors $\{\mathbf{x}_l, l = 1 \ldots L,\}$ of length $N = N_1 \cdot N_2$ each, or a single vector $\mathbf{x}$ of length NL. We use similar notations for the observed image, substituting letter y to letter x.

The blurred image corresponds to the 2D (circular)[1] convolution of $\mathbf{X}_l$ with filter $\mathcal{H}_l$. An equivalent formulation of the 2D convolution is obtained by defining a circulant-block-circulant convolution matrix $\mathbf{H}_l$ of size $N \times N$. The discrete convolution can be written in matrix form as (see [35] for details):

$$\mathcal{H}_l \underset{(2D)}{*} \mathbf{X}_l = \mathbf{H}_l \mathbf{x}_l. \tag{1}$$

If we assume that the blurring affecting each spectral slice is different, then the global convolution matrix $\mathbf{H}$ yielding the (vectorized) hyperspectral spectral image $\mathbf{y}$ is block-diagonal; each block $\mathbf{H}_l$ is the convolution matrix corresponding to the wavelength λ_l:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \mathbf{H}_L \end{bmatrix}. \tag{2}$$

Finally, the blurred and noisy spectral image is obtained by adding a noise term $\mathbf{e}$ which results in the observation model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}. \tag{3}$$

The problem in hyperspectral image deconvolution is then the inverse problem which aims at finding the original image $\mathbf{x}$ from the observed one.

B. Hyperspectral Image Deconvolution

1) Spatial and Spectral Regularization: Henrot *et al.* [36] and Henrot [35] proposes to add a spectral regularization to

[1] In what follows, we will consider circular convolution which results in an exact discrete Fourier domain implementation of the convolution.

the traditional Tikhonov and Arsenin [37] approach yielding a criterion composed of three terms: the data fitting, the spatial regularization and the spectral regularization:

$$\min_{\mathbf{x}} J(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\mu_s}{2}\|\mathbf{D}_s\mathbf{x}\|_2^2 + \frac{\mu_\lambda}{2}\|\mathbf{D}_\lambda\mathbf{x}\|_2^2. \quad (4)$$

Here, μ_s and μ_λ are respectively the spatial and spectral regularization parameters. $\mathbf{D}_s$ corresponds to a Laplacian filter and $\mathbf{D}_\lambda$ corresponds to a first-order derivative filter along the spectral dimension.

The solution of problem (4) is given by

$$\mathbf{x}^* = (\mathbf{H}^T\mathbf{H} + \mu_s\mathbf{D}_s^T\mathbf{D}_s + \mu_\lambda\mathbf{D}_\lambda^T\mathbf{D}_\lambda)^{-1}\mathbf{H}^T\mathbf{y}. \quad (5)$$

Introducing the spectral regularization results in a coupling of both spatial and spectral dimensions: the slices of the hyperspectral images cannot be processed independently. Following [38], the Laplacian (second-order derivative) and first-order derivative are valid regularization differential operators. A rule of thumb to choose the differential operator is as follows: the identity operator will favor the reconstruction of null signals, first-order derivative will favor the reconstruction of constant signals, second-order derivative will favor the reconstruction of linear signals. It is worth noticing that these operators can be replaced by others. However, in practice, the choice of the differential operator is not so crucial since the criterion also includes the data fitting term and the trade-off between the data fitting and regularization terms is controlled by the regularization parameters. In fact, only the choice of the regularization parameters really matters.

2) Restoration With Non-Negativity Constraint: The solution expressed in (5) cannot guarantee the non-negativity of the restored image. As proposed in [13], we can add a non-negativity constraint resulting in the following constrained optimization problem:

$$\begin{aligned} &\min_{\mathbf{x}} \; J(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\mu_s}{2}\|\mathbf{D}_s\mathbf{x}\|_2^2 + \frac{\mu_\lambda}{2}\|\mathbf{D}_\lambda\mathbf{x}\|_2^2 \\ &\text{s.t. } \mathbf{x} \geqslant 0. \end{aligned} \quad (6)$$

To solve this problem, we use the quadratic penalty method proposed in [13] which consists in introducing a slack variable $\mathbf{p}$. The original problem is then replaced by a surrogate criterion expressed in (7)

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{p}} \; K(\mathbf{x}, \mathbf{p}; \xi) = J(\mathbf{x}) + \frac{\xi}{2}\|\mathbf{x} - \mathbf{p}\|_2^2 \\ &\text{s.t. } \mathbf{p} \geq 0. \end{aligned} \quad (7)$$

The solution is obtained iteratively. At each iteration, the following three steps are performed:

- unconstrained minimization of $K(\mathbf{x}, \mathbf{p}; \xi)$ with respect to $\mathbf{x}$;
- constrained minimization of $K(\mathbf{x}, \mathbf{p}; \xi)$ with respect to $\mathbf{p}$;
- increase of the penalty factor ξ.

These three steps are alternated until a maximum number of iterations N_{iter} is reached.

At each iteration $k = 1, \ldots, N_{iter}$, when $\mathbf{p}^k$ and ξ^k are fixed, $K(\mathbf{x}|\mathbf{p}^k, \xi^k)$ can be minimized explicitly

$$\begin{aligned} \mathbf{x}^{k+1} = &(\mathbf{H}^T\mathbf{H} + \mu_s\mathbf{D}_s^T\mathbf{D}_s + \mu_\lambda\mathbf{D}_\lambda^T\mathbf{D}_\lambda + \xi^k\mathbf{I}_{NL})^{-1} \\ &(\mathbf{H}^T\mathbf{y} + \xi^k\mathbf{p}^k). \end{aligned} \quad (8)$$

Once $\mathbf{x}^{k+1}$ is obtained, $\mathbf{p}$ can be updated according to

$$\mathbf{p}^{k+1} = \max(\mathbf{0}, \mathbf{x}^{k+1}). \quad (9)$$

A detailed analysis of the quadratic penalty method, including convergence, can be found in [39]. In practice, increasing the value of ξ will ensure the solution to converge to the minimum of the constrained problem. Following [39], the simplest choice is $\xi^{(k+1)} = \gamma\,\xi^{(k)}$ with $\gamma > 1$. The initial value of ξ should be large enough. Indeed, if it is too small, a large number of iterations may be required to reach the optimum. The choice of γ also influences the convergence rate. As the minimization of $K(\mathbf{x}|\mathbf{p}^k, \xi^k)$ yields the explicit solution (8), γ can be set to a large value. Here, the initial value of ξ is set to 1 and $\gamma = 10$.

III. Hyperspectral Image Deconvolution as a Multi-Objective Optimization

The starting point of our problem is the hyperspectral image deconvolution (HID) in (6). The optimal solution $\mathbf{x}^*$ depends on both μ_s and μ_λ. If the value of μ_s increases toward infinity, the term $\|\mathbf{D}_s\mathbf{x}\|_2^2$ will be minimized. Similarly, the value of μ_λ increasing toward infinity will minimize $\|\mathbf{D}_\lambda\mathbf{x}\|_2^2$. When μ_s and μ_λ approach zero, the solution tend to minimize the data fitting term $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$. When μ_s is very small, the resulting deconvolution is generally not satisfactory because the noise is not sufficiently rejected. On the other hand, for very small μ_λ, the intensities of two adjacent spectral bands are not similar enough. But when both are large, the error between the solution and the observed image increases. This means that we cannot improve one objective without deteriorating the others. In this section, by stating the problem as a convex multiple objective optimization problem, it is possible to estimate the response surface from which the Pareto front can be deduced: this gives a characterization of the set of solutions obtained by varying the values of $\boldsymbol{\mu} = (\mu_s, \mu_\lambda)$. While the notions presented here are mainly concerned with the HID, the problem is formulated in a much more general setting which is the minimization of cost functions consisting in the weighted sum of convex objectives.

A. Multi-Objective Optimization

A generic multi-objective optimization problem may be formulated as:

$$\min_{\mathbf{x}} \mathbf{J}(\mathbf{x}) = (J_1(\mathbf{x}), \ldots, J_z(\mathbf{x})). \quad (10)$$

Here, $\mathbf{J}$ is a (vector) criterion of z (which equals 3 in our case) objective functions which defines a multi-dimensional space.

1) Ideal Objective Vector: The ideal objective vector is defined as in [40]:

$$\mathbf{I} = (I_1, \ldots, I_z). \quad (11)$$

The i-th component of $\mathbf{I}$ is the minimum of the problem:

$$I_i = \min_{\mathbf{x}} J_i(\mathbf{x}), \quad i = 1, \ldots, z. \quad (12)$$

The ideal objective vector $\mathbf{I}$ corresponds to the point whose coordinates are minimum among each objective. It is termed as ideal because the value of each objective is the smallest

and that is exactly the goal of the minimization problem (10). Most of the time, it cannot be reached because the objectives are conflicting: decreasing one will increase the others.

2) Pareto Front: The notion of *domination* is defined for example in [40]. It is an important notion in multi-objective optimization since it allows to define a solution ordering, i.e. we can say that a solution is better than another one. Let $\mathbf{x^{(1)}}$ and $\mathbf{x^{(2)}}$ be two different solutions of the multi-objective minimization problem (10). The solution $\mathbf{x^{(1)}}$ is said to dominate $\mathbf{x^{(2)}}$ and we write $\mathbf{x^{(1)}} \preceq \mathbf{x^{(2)}}$, if the solution $\mathbf{x^{(1)}}$ is not worse than $\mathbf{x^{(2)}}$ in all objectives, and the solution $\mathbf{x^{(1)}}$ is strictly better than $\mathbf{x^{(2)}}$ in at least one objective:

$$\mathbf{x^{(1)}} \preceq \mathbf{x^{(2)}} \text{ iff } \begin{cases} J_i(\mathbf{x^{(1)}}) \leq J_i(\mathbf{x^{(2)}}), & \forall i \in [1,\ldots,z] \\ \exists j \in [1,\ldots,z] & \text{s.t. } J_j(\mathbf{x^{(1)}}) < J_j(\mathbf{x^{(2)}}). \end{cases} \tag{13}$$

Otherwise, the solution $\mathbf{x^{(1)}}$ does not dominate the solution $\mathbf{x^{(2)}}$. A solution is either dominated or non-dominated but cannot be both at the same time. The solution $\tilde{\mathbf{x}}$ is said to be non-dominated or Pareto-optimal for a multi-objective problem if all other vectors $\mathbf{x}$ in the set of all feasible points have a higher value for at least one of the objectives J_i with $i = 1, \cdots, z$. The set of all the non-dominated solutions is called Pareto front or Pareto curve or surface which means that each solution belonging to the Pareto front cannot be said better than another in the sense of domination. The shape of the Pareto surface reveals the nature of the trade-off between the different objective functions. In multi-objective optimization, the goal is to find the set of Pareto-optimal solutions rather than a single solution. Two different cases have to be distinguished:

- the case of convex criteria for which it is proved that any point of the Pareto front can be reached using the weighted sum approach;
- the case of non-convex criteria for which the weighted sum approach cannot find the non-convex part of the Pareto front [41]. A lot of attention was paid to this case (see for example [40], [42]–[44]).

In our case, convex criteria are considered, but we only search for a single solution being optimal according to a given criterion; the multi-objective optimization formalism is used to derive and analyze such a criterion.

3) Weighted-Sum Approach: The weighted-sum approach can solve the multi-objective problem by combining all of the objectives into a single one. With this method, the weights between objectives are assigned *a priori* before the optimization process is completed. With z objectives, the equivalent scalar objective $J(\mathbf{x_w})$ is given by:

$$\begin{aligned} J(\mathbf{x_w}) &= \sum_{j=1}^{z} w_j J_j(\mathbf{x}) \\ &= \boldsymbol{w}^T \mathbf{J}(\mathbf{x}). \end{aligned} \tag{14}$$

Here is an example of the weighted-sum method with two objectives for which (14) simplifies to:

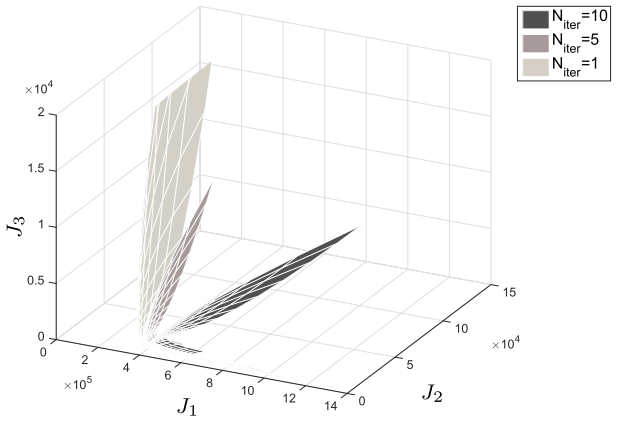


Fig. 1. Estimated response surface for different values of N_{iter}.

$$\begin{aligned} J(\mathbf{x}_{w_1,w_2}) &= w_1 J_1(\mathbf{x}) + w_2 J_2(\mathbf{x}) \\ \text{and} \quad w_1 + w_2 &= 1, \quad w_1 \geq 0, \ w_2 \geq 0. \end{aligned} \tag{15}$$

If the weight vector is parameterized by α, so that $w_1 = 1-\alpha$ and $w_2 = \alpha$, then the problem becomes:

$$\begin{aligned} J(\mathbf{x}_\alpha) &= (1-\alpha) J_1(\mathbf{x}) + \alpha J_2(\mathbf{x}) \\ \text{and } 0 &\leq \alpha < 1. \end{aligned} \tag{16}$$

In our minimization problem with the non-negativity constraint, the following two formulations are equivalent since $\mu = \alpha/(1-\alpha)$.

$$\min_{\mathbf{x}\geq 0} (1-\alpha) J_1(\mathbf{x}) + \alpha J_2(\mathbf{x}) \iff \min_{\mathbf{x}\geq 0} J_1(\mathbf{x}) + \mu J_2(\mathbf{x}). \tag{17}$$

In the case of three objectives, the criteria in (6) can also be written as:

$$\min_{\mathbf{x}\geq 0} J(\mathbf{x}) = J_1(\mathbf{x}) + \mu_s J_2(\mathbf{x}) + \mu_\lambda J_3(\mathbf{x}). \tag{18}$$

Each value of $\boldsymbol{\mu} = (\mu_s, \mu_\lambda)$ yields a solution:

$$\mathbf{x}_{\boldsymbol{\mu}} = \arg\min_{\mathbf{x}\geq 0} J(\mathbf{x}) \tag{19}$$

and gives a point in the response surface which will be denoted by $\Pi(\boldsymbol{\mu})$. Unlike the L-curve or the L-hypersurface, this response surface uses linear scales axes. For notation simplicity we will write $J_i(\mathbf{x}_{\boldsymbol{\mu}}) \triangleq J_i(\boldsymbol{\mu})$ and the same for $J(\mathbf{x}_{\boldsymbol{\mu}}) \triangleq J(\boldsymbol{\mu})$.

A necessary condition for the recovery of the Pareto front using the weighted-sum method is that all objectives are convex functions of $\mathbf{x}$ which is the case here. In the next section, an equivalent constrained minimization formulation will be used to prove that the response surface of a convex tri-objective is convex as well.

B. Properties of the Response Surface

1) Evaluating the Response Surface for HID Problem: To estimate the response surface of HID problem, we define (μ_s, μ_λ) on a 2D grid. Then, for each couple of parameters, the corresponding solution $\mathbf{x}^*$ is computed using the algorithm

presented in section II-B. To simplify notation, the dependence of $\mathbf{x}^*$ on μ_s and μ_λ is omitted.

Figure 1 shows three different empirical response surfaces estimated from the same simulated example (see section V) for $N_{iter} = 1, 5, 10$. For each response surface, the hyperparameters (μ_s, μ_λ) are sampled on a 20×20 regular logarithmic scale varying from 0.1 to 1000. The case $N_{iter} = 1$ corresponds to the response surface obtained with the unconstrained Tikhonov solution with spatial and spectral regularizations (section II-B.1). The two others correspond to the response surface obtained with the non-negative constrained Tikhonov solution of section II-B.2. For all cases, the penalty factor ξ is evolving similarly.

2) Convexity of the Response Surface: Because $J(\mathbf{x})$ is the sum of three convex objectives and the non-negative orthant is convex, problem (18) remains convex. We follow an approach similar to that of [24] to prove the convexity of the response surface of the multi-objective minimization problem (18).

Theorem 1: If $J(\mathbf{x})$ is convex, then the response surface of problem (18) is convex.

Proof: First, the tri-objective optimization problem (18) can be written in the following equivalent form:

$$\min_{\mathbf{x}\geq 0} J_1(\mathbf{x}) \quad \text{subject to} \quad \begin{cases} J_2(\mathbf{x}) \leq \tau_s, \\ J_3(\mathbf{x}) \leq \tau_\lambda. \end{cases} \tag{20}$$

Let $\mathbf{x}_{\tau_s,\tau_\lambda}$ be the optimal solution of (20) and $\Pi(\tau_s, \tau_\lambda)$ be the response surface, then for each $\tau_s \geq 0$, $\tau_\lambda \geq 0$:

$$\Pi(\tau_s, \tau_\lambda) = J_1(\mathbf{x}_{\tau_s,\tau_\lambda}). \tag{21}$$

The equivalence of problems (18) and (20) implies that there exists a unique hyperparameter (μ_s, μ_λ) yielding the same solution as (τ_s, τ_λ). In other words, there is a one-to-one correspondence between (μ_s, μ_λ) and (τ_s, τ_λ).

Equation (21) can be restated as:

$$\Pi(\tau_s, \tau_\lambda) = \inf_{\mathbf{x}\geq 0} f(\mathbf{x}, \tau_s, \tau_\lambda) \tag{22}$$

where

$$f(\mathbf{x}, \tau_s, \tau_\lambda) = J_1(\mathbf{x}) + \varphi_{\tau_s}(\mathbf{x}) + \varphi_{\tau_\lambda}(\mathbf{x}) \tag{23}$$

$$\varphi_{\tau_s}(\mathbf{x}) = \begin{cases} 0 & \text{if} \quad J_2(\mathbf{x}) \leq \tau_s \\ \infty & \text{otherwise} \end{cases} \tag{24}$$

$$\varphi_{\tau_\lambda}(\mathbf{x}) = \begin{cases} 0 & \text{if} \quad J_3(\mathbf{x}) \leq \tau_\lambda \\ \infty & \text{otherwise.} \end{cases} \tag{25}$$

Note that $\varphi_{\tau_s}(\mathbf{x})$ is convex in $(\mathbf{x}, \tau_s)$ [45] and the same for $\varphi_{\tau_\lambda}(\mathbf{x})$ which is convex in $(\mathbf{x}, \tau_\lambda)$. Since the objective $J_1(\mathbf{x})$ is convex in $\mathbf{x}$, f is then convex in $(\mathbf{x}, \boldsymbol{\tau})$, where $\boldsymbol{\tau} = (\tau_s, \tau_\lambda)$. Let $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$ be non-negative, $\mathbf{x}_1$ and $\mathbf{x}_2$ be the corresponding minimizers of (22) and $\alpha \in [0, 1]$. We then have:

$$\begin{aligned} \Pi((1-\alpha)\boldsymbol{\tau}_1 + \alpha\boldsymbol{\tau}_2) &= \inf_{\mathbf{x}\geq 0} f(\mathbf{x}, (1-\alpha)\boldsymbol{\tau}_1 + \alpha\boldsymbol{\tau}_2) \\ &\leq f((1-\alpha)\mathbf{x}_1 + \alpha\mathbf{x}_2, (1-\alpha)\boldsymbol{\tau}_1 + \alpha\boldsymbol{\tau}_2) \\ &\leq (1-\alpha) f(\mathbf{x}_1, \boldsymbol{\tau}_1) + \alpha f(\mathbf{x}_2, \boldsymbol{\tau}_2) \\ &= (1-\alpha)\Pi(\boldsymbol{\tau}_1) + \alpha\Pi(\boldsymbol{\tau}_2). \end{aligned} \tag{26}$$

Hence, the 2D response surface $\Pi(\boldsymbol{\tau})$ is convex. ■

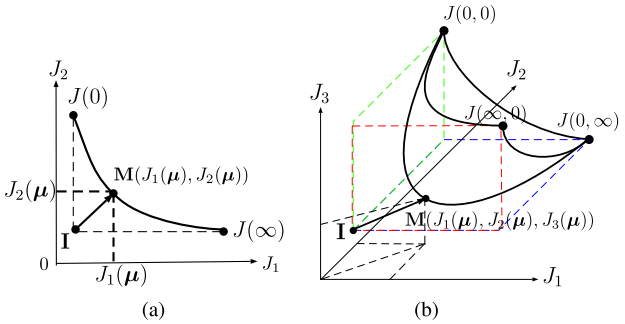


Fig. 2. Representation of the response surface for the unconstrained bi-objective and tri-objective cases: it corresponds to the Pareto front. The ideal point is denoted by $\mathbf{I}$. (a) bi-objective case. (b) tri-objective case.

Let us mention that the bi-objective unconstrained case can be written as follows:

$$\min J_1(\mathbf{x}) \quad \text{subject to } J_2(\mathbf{x}) \leq \tau. \tag{27}$$

In this case, [24] proved that Π is a nonincreasing[2] function of τ. This implies that the Pareto front exactly coincides with the response curve.[3] This is no longer true for the constrained problem at hand.

3) Shape of the Response Surface With Non-Negativity Constraint: Following [24], in the unconstrained bi-objective case, the response curve is convex and monotonically decreasing, as represented in Figure 2(a). This can be extended to the response surface corresponding to the unconstrained tri-objective case (Figure 2(b)). It is convex; its intersection with a plane parallel to either (J_1, J_2) or (J_1, J_3) or (J_2, J_3) also defines a monotonically decreasing function. In this case, the Pareto front coincides with the response surface since no point of the response surface is dominated by another one. This behavior is experimentally observed when we use the unconstrained deconvolution (case $N_{iter} = 1$ in Figure 1).

On the contrary, when a non-negativity constraint is enforced, the estimated response surface is no longer as in Figure 2(b). A folding of the response surface is observed ($N_{iter} = 5$ and $N_{iter} = 10$ in Figure 1). This results from the constrained data fitting term J_1 which is decreasing and then increasing as μ_s (or μ_λ) increases. In this case, only the set of non-dominated points of the response surface is corresponding to the Pareto front.

When comparing the estimated response surfaces obtained by using $N_{iter} = 5$ and $N_{iter} = 10$, it appears that the case $N_{iter} = 5$ gives an intermediate response between the unconstrained case ($N_{iter} = 1$) and the case $N_{iter} = 10$ for which the convergence of the algorithm is experimentally verified. Actually, this shows how the quadratic penalty approach is progressively modifying the response surface until it converges to the response surface with non-negativity constraint. The folding of the response surface is proved in the following theorem.

Theorem 2: Let us consider the following constrained

[2]In the bi-objective case, the regularization parameter is a scalar.

[3]In the bi-objective case, this is a curve.

optimization problem:

$$\min_{\mathbf{x}\geq 0}(1-\alpha)J_1(\mathbf{x})+\alpha J_2(\mathbf{x}) \quad s.t.\ J_3(\mathbf{x})=\tau_\lambda. \tag{28}$$

Then the data fitting J_1 has a unique minimum as α varies from 0 to 1.

Proof: When the non-negativity constraint is enforced, the data fitting can be written as in equation (29)

$$\|\mathbf{y}-\mathbf{H}\max(0,\mathbf{x}_\alpha)\|_2^2=\sum_{i\in\Omega_\alpha^+}(y_i-[\mathbf{H}\mathbf{x}_\alpha]_i)^2+\sum_{i\in\Omega_\alpha^-}y_i^2 \tag{29}$$

with $\Omega_\alpha^+=\{i \mid x_{\alpha,i}>0\}$ and $\Omega_\alpha^-=\{i \mid x_{\alpha,i}\leq 0\}$. Ω_α^+ corresponds to the set of points where the constraint is not active while Ω_α^- is the active constraint set. We have $\Omega_\alpha=\Omega_\alpha^+\cup\Omega_\alpha^-$ and $\varnothing=\Omega_\alpha^+\cap\Omega_\alpha^-$.

Let $n=|\Omega_\alpha|$ and $n^+=|\Omega_\alpha^+|$, $n^-=|\Omega_\alpha^-|$. By considering $n\to+\infty$, we can introduce the following probabilities:

$$\begin{aligned}\nu &= P_\alpha(x_i\in\Omega_\alpha^+)=\lim_{n\to+\infty}\frac{n^+}{n}\\ 1-\nu &= P_\alpha(x_i\in\Omega_\alpha^-)=\lim_{n\to+\infty}\frac{n^-}{n}.\end{aligned} \tag{30}$$

By taking the expectation, (29) can be rewritten as:

$$\begin{aligned}&\mathbb{E}\left[\|\mathbf{y}-\mathbf{H}\max(0,\mathbf{x}_\alpha)\|_2^2\right]\\ &\quad=\nu\mathbb{E}\left[\|\mathbf{y}-\mathbf{H}\mathbf{x}_\alpha\|_2^2\right]+(1-\nu)\mathbb{E}\left[\|\mathbf{y}\|_2^2\right].\end{aligned} \tag{31}$$

When $\alpha\to 1$, $\mathbf{x}_\alpha$ is highly regularized which means that it is very smooth and ν should be close to 1. On the contrary, when $\alpha\to 0$, $\mathbf{x}_\alpha$ is less regularized and ν is decreasing to a value $\nu_{\min}>0$. In other words, $\nu\in(0,1]$. At this point it is necessary to assume that there is a one-to-one correspondence between α and ν.

The term $\mathbb{E}\left[\|\mathbf{y}\|_2^2\right]$ is the norm of the data $\mathbf{y}$ which is constant with respect to ν. Thus, $(1-\nu)\mathbb{E}\left[\|\mathbf{y}\|_2^2\right]$ is a linearly decreasing function of ν. The term $\mathbb{E}\left[\|\mathbf{y}-\mathbf{H}\mathbf{x}_\alpha\|_2^2\right]$ is the estimation error with no non-negativity constraint. It is an increasing function of ν (or α). More precisely, as $\mathbb{E}\left[\|\mathbf{y}-\mathbf{H}\mathbf{x}_\alpha\|_2^2\right]$ is an increasing function of ν, $\nu\mathbb{E}\left[\|\mathbf{y}-\mathbf{H}\mathbf{x}_\alpha\|_2^2\right]=O(\nu^a)$ with $a>1$. Thus, there exists a value of ν for which the data fitting term is minimum. ∎

Remark 1: Instead of considering problem (28) which amounts to looking at a slice of the response surface parallel to (J_1, J_2), we could have considered

$$\min_{\mathbf{x}\geq 0}(1-\alpha)J_1(\mathbf{x})+\alpha J_3(\mathbf{x}) \quad s.t.\ J_2(\mathbf{x})=\tau_s$$

which corresponds to a slice of the response surface parallel to (J_1, J_2).

Remark 2: Due to the folding in the constrained convex multi-objective case, the Pareto front does not coincide with the response surface: actually, the Pareto front is the set of non-dominated points belonging to the response surface. This property has an important practical consequence: while the ideal point can be determined easily in the unconstrained case by setting the hyperparameters to particular values, this is no longer true in the constrained case. But this folding also has a very positive consequence since it will make the regularization parameter estimation easier (see section V-A)

IV. Choosing the Regularization Parameters

The response surface gives the set of all solutions of the convex multi-objective problem. The goal of this section is to choose among this set a particular solution which in turn consists in estimating the regularization parameters; here we will pay special attention to the non-negativity constrained multi-objective problem.

In the case of unconstrained bi-objective optimization problems, [18] proposes to find the point with maximum curvature on the L-curve which is a log-log plot of the norm of a regularized solution versus the norm of the corresponding fitting error. Actually, this is nothing but the response surface (which is also the Pareto front, see section III-B.3) plot in a log-log scale. However, this change of scale leads to a loss of convexity of the L-curve. In the case of multi-objective optimization problems, [21] extended the notion of L-curve to the L-hypersurface. Also, rather than using the maximum curvature, they proposed a minimum distance criterion to choose the optimal hyperparameters.

To preserve the convexity property, we will work directly on the response surface without resorting to a logarithmic scale. We propose the maximum curvature criterion (MCC) and the minimum distance criterion (MDC). An efficient algorithm to evaluate the MDC solution is also proposed.

A. Maximum Curvature Criterion

Since we have the regularization parameter $\boldsymbol{\mu}=(\mu_s,\mu_\lambda)$ for the tri-objective problem (18), the response surface $\Pi(\boldsymbol{\mu})$ is a two-dimensional manifold (surface) in $\mathbb{R}_+^3$. To calculate the curvature, we need to estimate the first and second derivatives of each objective $J_1=\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2$, $J_2=\|\mathbf{D}_s\mathbf{x}\|_2^2$ and $J_3=\|\mathbf{D}_\lambda\mathbf{x}\|_2^2$ with respect to μ_s and μ_λ. If f is a function of both variables $\{\mu_s,\mu_\lambda\}$, we can, for example, estimate the first partial derivative of f with respect to μ_s by:

$$f'_{\mu_s}(i,j)\approx\frac{f(\mu_{s,i},\mu_{\lambda,j})-f(\mu_{s,i-1},\mu_{\lambda,j})}{\mu_{s,i}-\mu_{s,i-1}} \tag{32}$$

where $(\mu_{s,i},\mu_{\lambda,j})$ is a discrete grid over which f is computed. Similarly, the second partial derivative of f is estimated by:

$$\begin{aligned}&f''_{\mu_s}(i,j)\\ &\quad\approx\frac{f(\mu_{s,i+1},\mu_{\lambda,j})+f(\mu_{s,i-1},\mu_{\lambda,j})-2f(\mu_{s,i},\mu_{\lambda,j})}{\mu_{s,i+1}-\mu_{s,i-1}}.\end{aligned} \tag{33}$$

Now we define the curvature as follows.

Definition 1 (Curvature): Let $x=J_1(\boldsymbol{\mu})$, $y=J_2(\boldsymbol{\mu})$ and $z=J_3(\boldsymbol{\mu})$, κ_{μ_s} and κ_{μ_λ} are the curvatures along μ_s and μ_λ respectively.

$$\begin{aligned}\kappa_{\mu_s}&=\frac{\sqrt{a_{\mu_s}^2+b_{\mu_s}^2+c_{\mu_s}^2}}{(x_{\mu_s}'^2+y_{\mu_s}'^2+z_{\mu_s}'^2)^{3/2}}\\ \kappa_{\mu_\lambda}&=\frac{\sqrt{a_{\mu_\lambda}^2+b_{\mu_\lambda}^2+c_{\mu_\lambda}^2}}{(x_{\mu_\lambda}'^2+y_{\mu_\lambda}'^2+z_{\mu_\lambda}'^2)^{3/2}}\end{aligned} \tag{34}$$

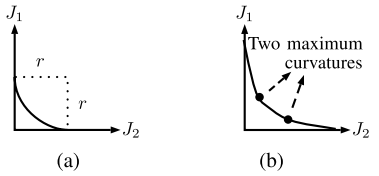


Fig. 3. Examples of response surfaces with non-unique maximum curvature.

where $a = z''y' - y''z', b = x''z' - z''x', c = y''x' - x''y'$. *The curvature of the surface is defined by*

$$\kappa = \kappa_{\mu_s} \cdot \kappa_{\mu_\lambda}. \tag{35}$$

Then we can formulate the maximum curvature criterion as follows:

Definition 2 (Maximum Curvature Criterion):

$$\boldsymbol{\mu}^* = \arg\max_{\boldsymbol{\mu}} \kappa(\boldsymbol{\mu}). \tag{36}$$

Note that the proposed definition of the curvature is a simplified version which does not correspond to the mean curvature as defined in differential geometry. Indeed, in our case, it is estimated as the average of two curvatures along two predefined directions while, in differential geometry, the mean curvature corresponds to the average of the principal curvatures which require the estimation of the curvatures along all possible directions.

The maximum curvature criterion suffers from two main shortcomings. One of them is related to the discrete derivative evaluation which is highly sensitive to noise. This noise comes from the use of an iterative solver which provides only approximate solutions thus yielding a noisy response surface. Another important issue is the non-uniqueness of the maximum curvature criterion. To illustrate this point, let us consider the two following bi-objective examples. The first one corresponds to a response curve having a convex quarter circle shape, as shown in Figure 3(a): the curvature is a constant and, thus, not unique. The second case corresponds to a convex response surface front whose curvature has two maxima highlighted by the two dots in Figure 3(b). Despite the fact that the non-negativity constraint increases the curvature of the response surface and makes the MCC more efficient, it cannot fully overcome these shortcomings. Instead, we propose in the next section the MDC.

B. Minimum Distance Criterion

As mentioned before, the MDC is applied directly to the response surface whose convexity is central in establishing the properties of the criterion. The ideal point as defined in (11) corresponds to the minimum of all objective functions. Even if it is a non-reachable solution, it can be considered as a reference point and the optimal point of the response surface will be the one having the minimum distance to this ideal point. However, this choice of the reference point is somewhat arbitrary; this will be discussed in remark 4 and in the experimental section V-A. Let us introduce the MDC by first defining the distance to the ideal point.

Definition 3 (Distance to Ideal Point): Let $\mathbf{I} = (I_1, \cdots, I_z)$ *denotes the coordinates of the ideal point. The function* $D(\boldsymbol{\mu})$ *is the squared distance from the ideal point* $\mathbf{I}$ *to the point* $\mathbf{M}(\boldsymbol{\mu}) = (J_1(\boldsymbol{\mu}), \cdots, J_z(\boldsymbol{\mu}))$ *on the response surface.*

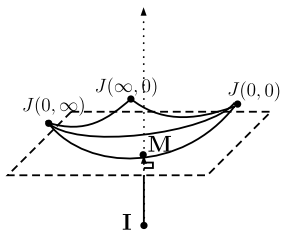


Fig. 4. Minimum distance criterion : the solution corresponds to $\mathbf{T} \perp \vec{\mathbf{IM}}$.

$$D(\boldsymbol{\mu}) = \sum_{i=1}^{z} (J_i(\boldsymbol{\mu}) - I_i)^2. \tag{37}$$

The MDC is defined as follows.

Definition 4 (Minimum Distance Criterion):

$$\boldsymbol{\mu}^* = \arg\min_{\boldsymbol{\mu}} D(\boldsymbol{\mu}). \tag{38}$$

The key property of the MDC is that it admits a unique minimum. Its proof relies on a geometrical interpretation of the MDC.

Theorem 3: If the response surface is convex, the MDC admits a unique minimum.

Proof: To find the minimum of $D(\boldsymbol{\mu})$, we have to find $\boldsymbol{\mu}$ such that:

$$\begin{aligned}
\frac{\partial D(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} &= \frac{\partial \left((J_1(\boldsymbol{\mu}) - I_1)^2 + \cdots + (J_z(\boldsymbol{\mu}) - I_z)^2 \right)}{\partial \boldsymbol{\mu}} \\
&= 2 \left(\frac{\partial J_1(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} (J_1(\boldsymbol{\mu}) - I_1) \right. \\
&\qquad \left. + \cdots + \frac{\partial J_z(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} (J_z(\boldsymbol{\mu}) - I_z) \right) \\
&= 2 \left[\frac{\partial J_1}{\partial \boldsymbol{\mu}}, \cdots, \frac{\partial J_z}{\partial \boldsymbol{\mu}} \right] \begin{bmatrix} J_1(\boldsymbol{\mu}) - I_1 \\ \vdots \\ J_z(\boldsymbol{\mu}) - I_z \end{bmatrix} \\
&= 0.
\end{aligned} \tag{39}$$

In (39), $\mathbf{T} = \left[\frac{\partial J_1}{\partial \boldsymbol{\mu}}, \cdots, \frac{\partial J_z}{\partial \boldsymbol{\mu}} \right]^T$ is the matrix whose columns span the tangent plane to the response surface at the point $\mathbf{M}$ and $\vec{\mathbf{IM}} = [J_1(\boldsymbol{\mu}) - I_1, \cdots, J_z(\boldsymbol{\mu}) - I_z]^T$. As (39) equals zero, we have:

$$\vec{\mathbf{IM}} \perp \mathbf{T} \tag{40}$$

which means that each column of $\mathbf{T}$ is orthogonal to $\vec{\mathbf{IM}}$. Any point satisfying the orthogonality condition is thus a critical point of $D(\boldsymbol{\mu})$. As the response surface is convex, the critical point is necessarily the unique minimum of $D(\boldsymbol{\mu})$ [45]. ■

This theorem is illustrated in Figure 4 showing the tangent plane orthogonal to $\vec{\mathbf{IM}}$.

Remark 3: The evaluation of the MDC requires the determination of the ideal point which is difficult to obtain when the non-negativity constraint is enforced. This is due to the folding discussed in section III-B.3. We propose to define it by determining the points of the response surface corresponding

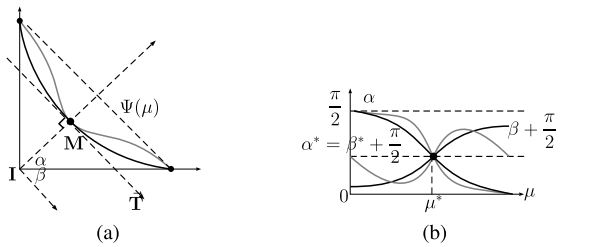


Fig. 5. Uniqueness of the MDC. (a) Definition of α and β. (b) Condition for the uniqueness.

to the unconstrained Tikhonov solution for three values of $\boldsymbol{\mu}$ *equal to* $(0,0), (0,\infty), (\infty,0)$.[4] *The ideal point coordinates are then obtained by finding the minimum coordinate of each of the three points (see Figure 2(b)). It is important to notice that determining the ideal point is not time consuming since the 3 points of the unconstrained response surface are computed using the fast frequency domain implementation of the unconstrained Tikhonov estimator.*

Remark 4: The convexity of the MDC is stated for a reference point chosen as the ideal point. The MDC remains convex for any other choice of the reference point, but the optimal point depends on this choice. More precisely, if the curvature of the response surface is low in the vicinity of the optimal point, then the estimated point will vary a lot with the chosen reference point while it does not if the curvature is large. The folding of the response surface resulting from the non-negativity constraint yields an increase of the response surface curvature, thus stabilizing the estimated point using MDC.

C. The Case of Unimodal MDC

Looking carefully at the shape of the estimated response surface in Figure 1 reveals that it may be slightly non-convex. In fact, the loss of convexity was experimentally observed when the deconvolution is possibly affected by numerical errors. We made a large number of experiments to identify the possible causes of this loss of convexity and we found that this phenomenon (sometimes but not always) arises in situations where the bandwidth of the signal to recover was much greater than the bandwidth of the convolution kernel. For the deconvolution problem at hand, it also depends on the conditioning of the convolution matrix. This phenomenon occurs mainly for small values of μ_s and μ_λ. A detailed analysis of this phenomenon is very complicated and is out of the scope of this paper. In all the considered cases, the loss of convexity resulted in a unimodal MDC. This motivates the study of the uniqueness of unimodal MDC. For simplicity reasons, only the bi-objective case is considered.

The minimum distance criterion consists in finding the vector $\vec{\mathbf{IM}}$ orthogonal to the tangent vector $\mathbf{T}$ at the point $\mathbf{M}$. Let α (resp. β) be the angle of $\vec{\mathbf{IM}}$ (resp. $\mathbf{T}$) with the horizontal axis, as shown in Figure 5(a). If the response surface Π is convex (as the black curves in Figures 5(a) and 5(b)), when $\vec{\mathbf{IM}}$ is moving from the vertical direction to the horizontal direction, the angle α is monotonically decreasing from $\frac{\pi}{2}$ to 0 and the angle $\beta \in [-\frac{\pi}{2}, 0]$ is monotonically increasing, as shown in Figure 5(b). There is only one point where $\alpha^* = \beta^* + \frac{\pi}{2}$. So there is only one point satisfying the orthogonality condition.

[4]In practice, the hyperparameter values cannot be set to ∞ but are fixed to large values

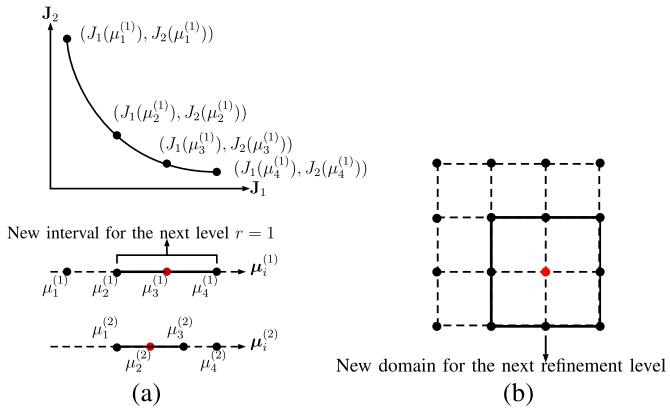


Fig. 6. Grid refinement method. (a) Bi-objective case. (b) Tri-objective case.

Let us examine what happens when the response surface Π is no longer convex, as highlighted by the gray curves in Figure 5(a) and 5(b). We assume that $D(\mu)$ is unimodal and admits the same minimum as in the convex case. This means that $D(\mu)$ is decreasing $\forall \mu < \mu^*$ and is increasing $\forall \mu > \mu^*$. There exists a line $\Psi(\mu)$ whose slope equals that of the tangent vector at $\mathbf{M}$, such that $\Pi(\mu) \leq \Psi(\mu)$. The slope of $\Psi(\mu)$ equals β^*. If $D(\mu)$ is unimodal, $\alpha(\mu)$ is monotonically decreasing from $\frac{\pi}{2}$ to 0 and we have the following inequalities:

$$\begin{cases} \beta(\mu) - \beta^* < \alpha(\mu) - \alpha^* & \text{if} \quad \mu < \mu^* \\ \beta(\mu) - \beta^* > \alpha(\mu) - \alpha^* & \text{if} \quad \mu > \mu^*. \end{cases} \tag{41}$$

So for the unimodal case, the unique intersection of α and $\beta(\mu) + \frac{\pi}{2}$ is in μ^*.

This proves that if the distance criterion is unimodal, it always admits a unique minimum even if the response surface is no longer convex. In that respect, the MDC can be said to be robust to the loss of convexity observed on the empirical response surface. We now turn our attention to the proposal of a fast method to estimate the MDC.

D. A Grid-Search Strategy for MDC

To begin, let us give one remark about the computational cost of the response surface estimation on a 2D grid. It can be very high if the grid size is high. To give some figures, the evaluation of the response surface of the non-negative Tikhonov solution on a 20×20 grid (black surface in Figure 1) takes more than 2 days with an Intel Core i7 2.2 Ghz processor. However, by properly exploiting the property of the response surface it is possible to design a fast approach aiming at finding a particular point on the response surface. We propose to use a grid-search method which is proved to be convergent for unimodal criteria [46].

Algorithm 1: Grid Minimization of MDC

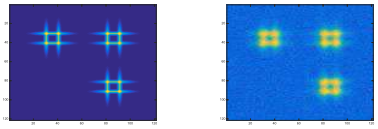
Data: Parameters of the initialized level $\boldsymbol{\mu} = (\mu_{1,1}^{(1)}, \cdots, \mu_{4,4}^{(1)})$; number of refinement levels R; image $\mathbf{y}_l$, matrix $\mathcal{H}_l$, matrix $\mathcal{D}_l$ for $l = 1, \cdots, L$; Δ_λ

Find the ideal point $\mathbf{I}$;
for $r = 1 : R$ **do**
  Π=**Procedure 2** ($\mathbf{y}_l$, $\mathcal{H}_l$, $\mathcal{D}_l$ for $l = 1, \cdots, L$; Δ_λ; $\boldsymbol{\mu}$);
  Calculate the distance $\mathbf{D}_r$ for each value of Π;
  Find the minimum distance d_{min} in $\mathbf{D}_r$ and the corresponding indexes m^*, n^* and $\boldsymbol{\mu}^*$;
  New domain for the grid refinement
  $\mu_{1,1}^{(r+1)} = \mu_{m^*-1,n^*-1}^{(r)}$;
  $\mu_{4,1}^{(r+1)} = \mu_{m^*+1,n^*-1}^{(r)}$;
  $\mu_{1,4}^{(r+1)} = \mu_{m^*-1,n^*+1}^{(r)}$;
  $\mu_{4,4}^{(r+1)} = \mu_{m^*+1,n^*+1}^{(r)}$;
  Calculate the grid points into the new domain
  $\boldsymbol{\mu} = (\mu_{1,1}^{(r+1)}, \cdots, \mu_{4,4}^{(r+1)})$;
end
Result: $\boldsymbol{\mu}^*$

Procedure 2: Evaluation of the Points on the Response Surface ()

Data: image $\mathbf{y}_l$, matrix $\mathcal{H}_l$, matrix $\mathcal{D}_l$ for $l = 1, \cdots, L$; matrix Δ_λ; $\boldsymbol{\mu}$

for $n = 2 : 3$ **do**
  for $m = 2 : 3$ **do**
    x=Deconvolution($\mathbf{y}_l$, $\mathcal{H}_l$, $\mathcal{D}_l$ for $l = 1, \cdots, L$; Δ_λ; (μ_s, μ_λ) corresponds to $\mu_{m,n}^{(r)}$);
    Calculate the response surface $\Pi(m, n) = (J_1(m, n), J_2(m, n), J_3(m, n))$;
  end
end
Result: Π



(a) (b)

Fig. 7. An example of simulated image. (a) Original image. (b) Blurred and noisy image (SNR = 10 dB).

Figure 6 illustrates the grid refinement. For the bi-objective case with one single regularization parameter, at the first level $r = 1$, we have only four points $\boldsymbol{\mu}_i^{(1)} (i = 1, \cdots, 4)$ on which the response surface is estimated. Then the grid is refined by defining a new search segment on which four new points $\boldsymbol{\mu}_i^{(2)} (i = 1, \cdots, 4)$ are defined. The procedure is repeated until a maximum number of levels is reached. In the tri-objective case with two regularization parameters, we define 4×4 points for the r-th level and choose an optimum point among them. Note that the response surface should be evaluated on the four central points of the grid. Then we select the points around it as the new domain. This new domain is refined to find a new optimum point. The procedure is repeated iteratively. The whole procedure is summarized in Algorithm 1.

In what follows, the number of levels is fixed to 6 which gives approximately the same resolution as the 20×20 grid of section III-B1 but requires only 24 evaluations of the response surface instead of 400. Note that such a grid strategy has also been used to maximize the MCC with a reasonable computation time. However, as the MCC is not necessarily unimodal, the procedure cannot be guaranteed to converge.

V. Examples and Experiments

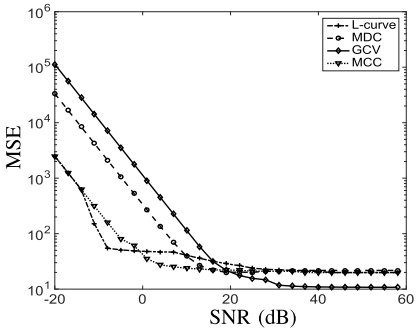
In this section, some numerical and experimental results will be presented to illustrate the effectiveness of proposed MCC and MDC for estimating the regularization parameters. First, a simple bi-objective image deconvolution problem is used to assess the performances of the proposed approaches and to compare them to state-of-the-art regularization parameter estimation methods. Then we address the tri-objective hyperspectral image deconvolution problem. We begin by giving an illustrative example and then we compare the performances of the two proposed approaches (MCC and MDC). The performance assessment is conducted by evaluating the mean square error (MSE) as a function of the signal-to-noise ratio (SNR): the lower the MSE, the better the performances.

A. Performances of MCC and MDC for 2D Image Deconvolution

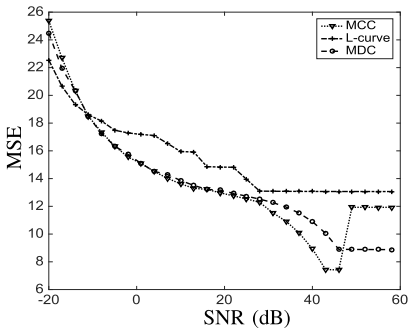
In this section, we consider a bi-objective 2D image deconvolution problem. We use the simulated image as shown in Figure 7 which corresponds to a single slice from the simulated hyperspectral cube in section V-B.

The first experiment aims at evaluating the performances of the proposed approaches using the unconstrained Tikhonov. They are compared to two state-of-the-art methods: the L-curve approach [18] and the generalized cross-validation (GCV) method [15]. The MSEs as a function of the SNR obtained for different methods are shown on Figure 8(a). The SNR is defined as follows: $\text{SNR} = 10 \log_{10} ||\mathbf{Hx}||_2^2 / ||\mathbf{e}||_2^2$. The second experiment aims at evaluating the performances of the criteria when using the non-negative Tikhonov approach. Note that in that case, GCV cannot be used since this algorithm cannot account for a non-negativity constraint. Thus, only the MSEs of MCC, MDC and L-curve are evaluated. The results are shown in Figure 8(b). To overcome the multiple maximum curvature problem which may occur in MCC and L-curve criteria, we follow the recommendation in [18]. Starting with a low value of the regularization parameter, the first maximum curvature is chosen to estimate the regularization parameter.

The MSE curve includes three main parts which are sketched on Figure 9. The non-efficiency zone corresponds to the part of the curve for which the MSE increases as fast as the noise level. The efficiency zone corresponds to the part of the curve for which the MSE increases at a lower rate than the noise: this is the zone where deconvolution is effective. Finally, the third horizontal part corresponds to the best performance of the regularized deconvolution method. The minimum value



(a)



(b)

Fig. 8. Performances of the 2D deconvolution with optimal parameter μ_s selected by different approaches. (a) Unconstrained case. (b) Non-negativity constrained case.

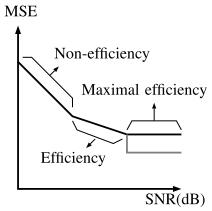
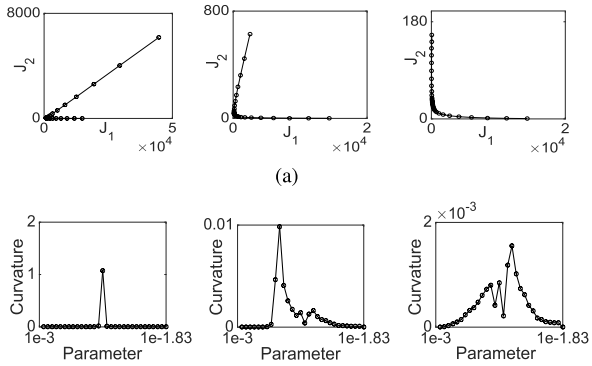


Fig. 9. Typical shape of the MSE.

of the MSE in this third part is also depending on both **H** and **x**. When the bandwidth of the filter **H** is lower than the bandwidth of **x**, even in noise-free situations, deconvolution cannot restore the signal **x** outside the frequency range (bandwidth) covered by the filter. In fact, this minimum MSE reflects the ill-conditioning of the matrix **H**. It decreases as the condition number decreases. For example, the gray curve in Figure 9 corresponds to a case where the conditioning is better than that of the black curve.

In the unconstrained case, no approach performs uniformly better than the others. GCV reaches the lowest minimum value of the MSE which is about 10^1. This shows that GCV works better than the other approaches for high SNR values. Both L-curve and MCC perform better than the other approaches when the SNR is low. For SNR ranging between $[0, 20]$ dB, the best performances are achieved by the MCC. Note that for SNR smaller than -14 dB (three first points), the maximum curvature of the L-curve is negative which is somewhat incoherent with the L-curve approach since the response curve no longer has the L-corner. The performances of the MDC (Figure 8(a)) are not very satisfying. As mentioned in



(a)

(b)

Fig. 10. Response curves and curvatures for different SNRs. (a) Response curves for SNR $= 25, 37, 52$ dB, respectively. (b) Curvatures for SNR $= 25, 37, 52$ dB, respectively.

remark 4, a strong folding of the response curve will decrease the sensitivity of the MDC to the choice of the reference point. In the unconstrained case, there is no folding of the response curve which results in the high sensivity of the MDC; this explains the poor performances of the MDC. On the contrary, the MCC appears to be less sensitive to the folding of the response curve.

When the non-negativity constraint is enforced, the L-curve does not yield satisfying results anymore. This can be attributed to the complex shape of the L-curve (which is reinforced by the non-negativity constraint) associated to the curvature maxima. At high SNR, the MSE reaches a minimum value similar to the unconstrained case. For SNR smaller than 28 dB, MCC and MDC behave similarly. MCC has the best performance for SNR in $[28, 46]$ dB. Indeed, the folding of the response curve decreases as the SNR increases since the constraint is active on fewer image points, see Figure 10(a). This explains why the MCC which is less sensitive to the folding, performs better than the MDC. However, the MCC may suffer from the non-uniqueness of the maximum curvature, see Figure 10(b). For example, after 46 dB, the first local maximum of the curvature no longer corresponds to the correct optimal point. This explains the step observed on the MSE of the MCC in Figure 8(b). Finally, it is worth mentioning that, in the non-negativity constrained case, both MCC and MDC reach a horizontal asymptote lower than the 10^1 reached by GCV which is the best performing method at high SNR values in the unconstrained case. This illustrates the regularizing effect of the non-negativity constraint.

B. An Illustrative Example of the Non-Negativity Constrained Hyperspectral Image Deconvolution

To simulate the blurred hyperspectral images, we first generate the unblurred image according to the instantaneous mixture model:

$$\mathbf{X} = \sum_k \mathbf{A}_k \circ \mathbf{s}_k. \tag{42}$$

Here, $\mathbf{A}_k$ represents the k-th abundance (spatial source) which is a function of the spatial variables, $\mathbf{s}_k$ represents the k-th endmember (spectral source) and $\circ$ is the outer (tensor)

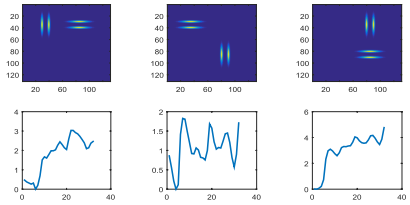
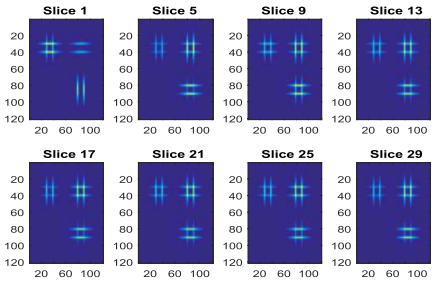
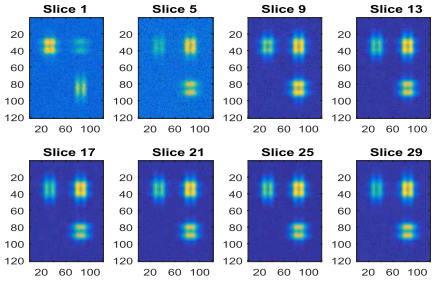


Fig. 11. Abundance maps and endmembers used to simulate the unblurred hyperspectral image.



(a)

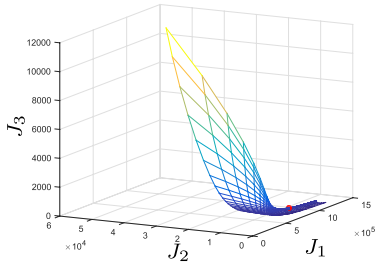


(b)

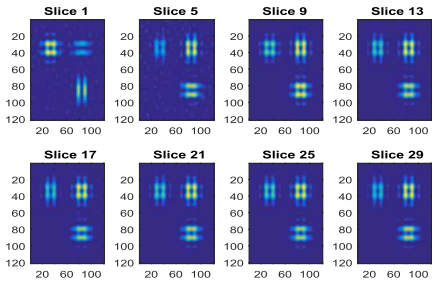
Fig. 12. Simulation of the hyperspectral image. (a) Unblurred hyperspectral image. (b) Blurred noisy hyperspectral image $\mathbf{y}$ (SNR= 20 dB).

product. In the example of Figure 11, an instantaneous mixture of 3 sources is considered. The abundance maps of size (120×120) are shown on the upper row while the endmembers, which include 32 spectral bands, are on the lower row. These endmembers correspond to NIR spectra of wood (raw, varnished and painted) samples. They were chosen because of their relative smoothness making the spectral smoothness penalty effective. However, in the supplementary material [26], we added other examples corresponding to different types of endmembers and abundance profiles. Eight slices of the resulting unblurred hyperspectral image are shown in Figure 12(a).

The convolution filter $\mathcal{H}_l$ is assumed to be a low-pass Gaussian filter of size (11×11) and its full width at half maximum is set to 5 points in both dimensions. It is invariant with respect to l. The blurring is implemented in the Fourier domain (circular convolution). Note that the filter invariance assumption adopted here for simplicity, is reasonable for applications such as Raman hyperspectral imaging systems, fluorescence confocal microscopy and industrial NIR spectro-imaging system. However, in some applications such as NIR microscopy, the variation of the filter with respect to l has to be taken into account (see [35] for details).

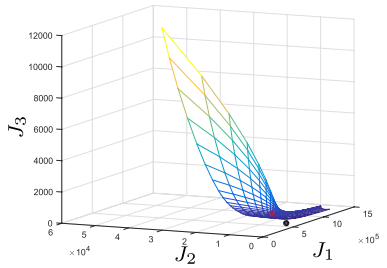


(a)

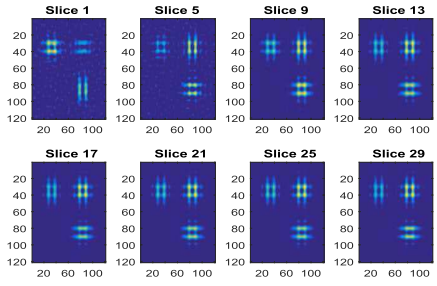


(b)

Fig. 13. Result of the non-negative deconvolution problem by using MCC. (a) Response surface with the point of maximum curvature (red point). (b) Deconvolution with parameters found by the MCC ($\mu_s = 88.5867$, $\mu_\lambda = 0.1624$).



(a)



(b)

Fig. 14. Result of the non-negative deconvolution problem by using MDC. (a) Response surface with the point at minimum distance (red one) from the ideal point (black one). (b) Deconvolution with parameters found by the MDC ($\mu_s = 2.9764$, $\mu_\lambda = 33.5982$).

A Gaussian noise is then added to the blurred image yielding the hyperspectral image of Figure 12(b). The noise level is the same for all bands. The simulated blurred hyperspectral image results in a difficult problem since the bandwidth of the unblurred image $\mathbf{x}$ is much larger than that of the filter $\mathbf{H}$. We also have to mention that the simulated hyperspectral image was chosen to favor non-negative deconvolution. This is because the simulated unblurred image includes a large amount of zero values.

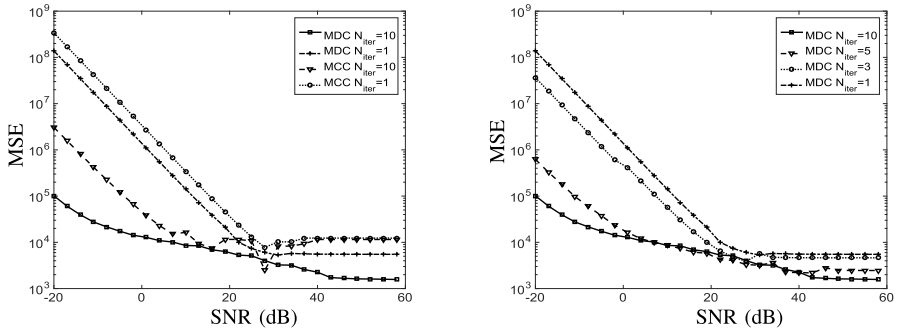


Fig. 15. Performances of the hyperspectral image deconvolution with optimal parameters (μ_s, μ_λ) selected by MCC and MDC. (a) Performances of the hyperspectral image deconvolution with optimal parameters (μ_s, μ_λ) selected by MCC and MDC. (b) Performances of the non-negative hyperspectral image deconvolution with optimal parameters (μ_s, μ_λ) selected by MDC.

We applied MCC and MDC to the non-negativity constrained deconvolution problem of the hyperspectral image shown in Figure 11. The response surface and the point corresponding to the MCC are shown in Figure 13(a). The result of the deconvolution with the parameters ($\mu_s = 88.5867$ and $\mu_\lambda = 0.1624$) found by this method is shown in Figure 13(b). The response surface and the point corresponding to MDC are shown in Figure 14(a). The corresponding parameters are $\mu_s = 2.9764$, $\mu_\lambda = 33.5982$. The corresponding restored images are shown in Figure 14(b). We can observe that for this deconvolution problem MDC works better than MCC. The poor performance of the MCC results from the multiple maximum curvatures. In fact, the point having the maximum curvature does not yield the best result.

The grid-search strategy with a number of levels fixed to 6 was applied to MDC. The estimated point ($\mu_s = 2.1274$, $\mu_\lambda = 31.3741$) is close to ($\mu_s = 2.9764$, $\mu_\lambda = 33.5982$), the point found on the whole response surface. The two points do not coincide exactly because the grids do not. The application of the grid search to MCC is highly sensitive to the choice of the initial grid. This is still a consequence of the already mentioned curvature multiple maxima. In the particular example considered here, using the same initial grid as for MCC yields a point ($\mu_s = 119.7262$, $\mu_\lambda = 7.2457$) which is completely different from ($\mu_s = 88.5867$, $\mu_\lambda = 0.1624$), the point found before on the whole response surface; the corresponding restored image (not shown) is not satisfying as well.

C. Performances of MCC and MDC for Non-Negative Hyperspectral Image Deconvolution

In the case of non-negative hyperspectral image deconvolution, as far as we know, no other approach than MCC and MDC can be used. The MSE of MCC and MDC corresponding to the unconstrained ($N_{iter} = 1$) and constrained ($N_{iter} = 10$) are shown in Figure 15(a). For MCC, the curves obtained with $N_{iter} = 10$ and $N_{iter} = 1$ tend to the same horizontal asymptote which is about 10^4 while it is about 10^3 for MDC. There is almost a factor 10 between the MSEs obtained with MCC and MDC. In fact, the poor behavior of the MCC associated to the grid search is only reflecting the already mentioned multiple maximum curvature problem.

Let us now examine the behavior of MDC. When $N_{iter} = 1$ (unconstrained Tikhonov approach), the efficiency zone is in the interval [20, 30] dB and MSE reaches its minimum when the SNR is greater than 30 dB. This is because MDC does not give good results in the unconstrained case. However, when $N_{iter} = 10$ (constrained Tikhonov approach), not only the minimum MSE is decreased but the efficiency zone (which is between -10 dB and 40 dB) increases significantly. This effect shows that MDC works better in the non-negativity constrained case and the non-negativity constraint improves the effectiveness of the deconvolution. This highlights the stabilizing property of the non-negativity constraint as proved in [47]. Once again, recall that the considered example favors non-negative deconvolution. This is because it does include large regions where the original hyperspectral image is null (or close to 0).

The Tikhonov approach with a non-negativity constraint is an iterative algorithm which converges to the optimal solution as the number of iterations increases. Figure 15(b) shows how the MSE obtained by MDC for different values of $N_{iter} = 1, 3, 5, 10$ is gradually changing from the unconstrained to the constrained case. In fact, the study of the non-negative deconvolution performance as a function of N_{iter} aims at evaluating how the convergence of the algorithm is affecting the performances of the MDC. Increasing the number of iteration allows to gradually increase the efficiency zone until the algorithm convergence is reached ($N_{iter} = 10$ for this example).

Extensive simulations investigating the behavior of MDC and MCC for different types of hyperspectral images can be found in [26]. The analysis of the results shows that MDC always performs better than MCC. Also, the corresponding MSEs are more stable (smooth) than those of MCC. This is due to the multiple maximum problem of MCC which renders the MSE behavior a bit erratic. The non-negativity constraint really matters when the image includes many zeros. Increasing the number of points on which the positivity constraint is active, will also increase the folding of the response surface resulting in an accurate regularization parameter estimation.

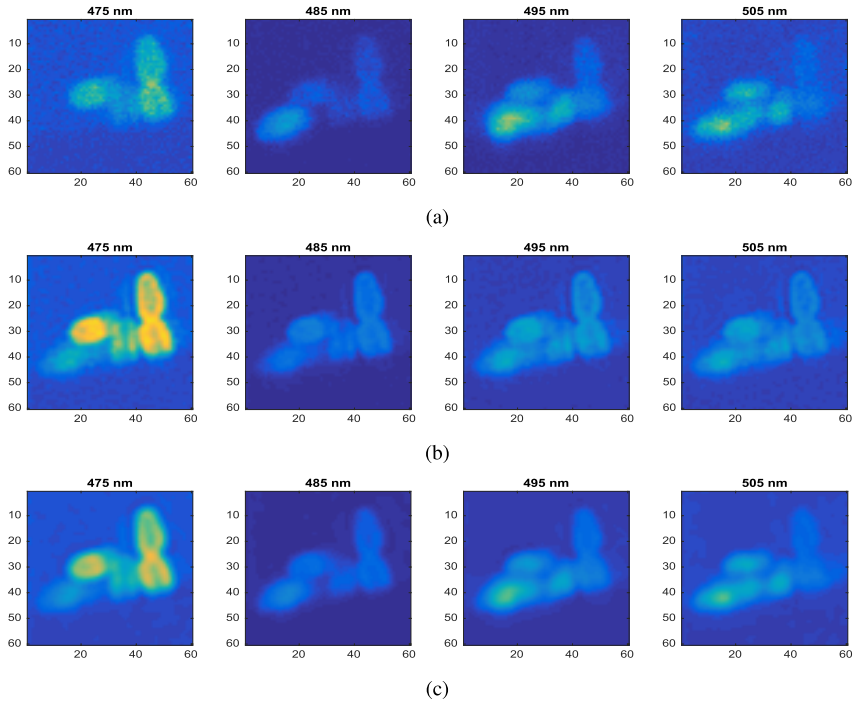


Fig. 16. Result of the non-negative deconvolution problem by using MCC and MDC. (a) Real hyperspectral image $\mathbf{y}$. (b) MCC ($\mu_s = 0.0047$, $\mu_\lambda = 4.3528$). (c) MDC ($\mu_s = 0.2505$, $\mu_\lambda = 0.6115$).

When the number of zeros is low, the non-negativity constraint is no longer relevant and both MDC and MCC are not very efficient. It may even happen that, for high SNR, the unconstrained deconvolution and associated MDC and MCC yields better solutions. See [26, Example 5] (which is a kind of worst-case scenario) for SNR $>$ 30 dB. Finally, the estimated regularization parameters with MDC (associated to non-negative deconvolution) is linked to the nature of the image to recover. Spatially (resp. spectrally) smooth images yield large values of μ_s (resp. μ_λ). Conversely, spatially (resp. spectrally) peaky images yield low values of μ_s (resp. μ_λ). It corresponds to what intuition suggests. This is another evidence of the interest of MDC.

D. Application to Hyperspectral Fluorescence Microscopy

A real-world example is included. It corresponds to an image of bacterial biosensors using hyperspectral fluorescence microscopy. A bacterial biosensor is a genetically modified bacteria which reacts to a stressing element (here iron, Fe) by producing a fluorescent protein (GFP). The hyperspectral fluorescence images will give indications of the Fe spatial concentration. This hyperspectral image size is $(512 \times 512 \times 16)$ and the pixel size is $0.117\,\mu$m along each dimension. The 16 wavelengths are ranging from 455nm to 605nm. It was obtained by Carl Zeiss Bio-Rad confocal microscope. The PSF of the microscope is evaluated according to [48] as a function of the imaging parameters (excitation wavelength, emission wavelength, numerical aperture and pixel size). This results in a 7×7 Gaussian approximation of the PSF.

Figure 16 is a part selected from the whole image. It shows raw data (upper row), restored data with the regularization parameters estimated by MCC (middle row) and restored data with the regularization parameters estimated by MDC (lower row). It should be noted that this image includes both peaky and smooth parts along the spectral dimension; this makes the choice of a global spectral regularization parameter not obvious. A large regularization parameter will over-smooth the peaky part while a low regularization parameter will under-regularize the smooth part.

Both results show an improved resolution. However, a closer look at the results of MCC reveals that the spectral regularization parameter is over-estimated. This results in a spectral over-smoothing which makes some high intensity patterns of bacteria remaining on adjacent spectral bands (see for example 16). This is less visible for MDC.

In fluorescence microscopy, the noise is typically modeled by a Poisson distribution due to photon counting in optical devices. Following [49], it includes two main contributions: the shot noise and the dark noise. This noise model is also well adapted to other types of hyperspectral images involving photon counting. The Poisson distribution is a non-negative support probability density function and, in that respect, it is well suited to the non-negative nature of hyperspectral images. When the SNR is high enough (large integration time), the Poisson noise can be well approximated by an additive Gaussian noise whose variance depends on the signal amplitude. For low SNR, the approximation is no longer valid. However, the Gaussian assumption is adopted in a large majority of works dealing with hyperspectral images. The application of the proposed methods to real hyperspectral data illustrates their relative insensitivity to the noise model.

VI. Conclusion

In this work, the estimation of the regularization parameters of non-negativity constrained hyperspectral image deconvolution algorithms is stated as a multi-objective optimization problem whose response surface is proved to be convex. A first contribution of this work is to show that the non-negativity constraint results in a folding of the response surface. A consequence is that, unlike the unconstrained case, the response surface does not coincide with the Pareto front. But this folding results in an increase of the curvature of the response surface which is making the regularization parameter estimation easier.

A second contribution of this work is the proposal of the MCC and MDC to estimate the optimal values of the regularization parameters μ_s and μ_λ for the non-negativity constrained tri-objective optimization problem. MCC aims at finding the point of the response surface with maximum curvature while MDC aims at finding the point of the response surface having the minimum distance from the ideal point. We also proved that this criterion admits a unique minimum even if the distance criterion is unimodal (and hence non-convex). A fast grid-search algorithm is proposed to estimate the point of the response surface maximizing MCC or minimizing MDC. Another very positive consequence of the response surface folding is that it decreases the sensitivity of both MDC and MCC. Finally, simulations were used to assess the performances of the proposed MCC and MDC. In addition, an application to a hyperspectral fluorescence microscopy is provided. In fact, MDC results is an efficient method to estimate the regularization parameters of non-negative hyperspectral image deconvolution.

Future works will focus on the extension of the proposed approaches to solve edge-preserving image deconvolution problems. We also intend to develop new approaches aiming at jointly performing the deconvolution and unmixing of hyperspectral images. The application of the MDC to real hyperspectral fluorescence data raises an interesting problem in image restoration when the convolution kernel is poorly known.

References

[1] P. Lasch and D. Naumann, "Spatial resolution in infrared microspectroscopic imaging of tissues," *Biochim. Biophys. Acta-Biomembranes*, vol. 1758, no. 7, pp. 814–829, 2006.

[2] S. Henrot, C. Soussen, M. Dossot, and D. Brie, "Does deblurring improve geometrical hyperspectral unmixing?" *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1169–1180, Mar. 2014.

[3] S. Henrot, C. Soussen, S. Moussaoui, and D. Brie, "Edge-preserving nonnegative deconvolution of hyperspectral fluorescence microscopy images," CRAN-IRCCyN, Univ. Lorraine Ecole Centrale Nantes, Res. Rep. hal01171524, Jul. 2015. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01171524

[4] J.-F. Giovannelli and A. Coulais, "Positive deconvolution for superimposed extended source and point sources," *Astron. Astrophys.*, vol. 439, no. 1, pp. 401–412, 2005.

[5] S. Bongard, F. Soulez, É. Thiébaut, and É. Pecontal, "3D deconvolution of hyper-spectral astronomical data," *Monthly Notices Roy. Astron. Soc.*, vol. 418, no. 1, pp. 258–270, 2011.

[6] X.-L. Zhao, F. Wang, T.-Z. Huang, M. K. Ng, and R. J. Plemmons, "Deblurring and sparse unmixing for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4045–4058, Jul. 2013.

[7] K. C. Lawrence, B. Park, W. R. Windham, and C. Mao, "Calibration of a pushbroom hyperspectral imaging system for agricultural inspection," *Trans. ASAE*, vol. 46, no. 2, pp. 513–521, 2003.

[8] Pellenc Selective Technology. *Mistral Product*, accessed on May 20, 2016. [Online]. Available: http://www.pellencst.com/products/

[9] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1860–1875, Nov. 2005.

[10] P. J. La Rivière and P. A. Vargas, "Monotonic penalized-likelihood image reconstruction for X-ray fluorescence computed tomography," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1117–1129, Sep. 2006.

[11] D. Gürsoy, T. Biçer, A. Lanzirotti, M. G. Newville, and F. De Carlo, "Hyperspectral image reconstruction for X-ray fluorescence tomography," *Opt. Exp.*, vol. 23, no. 7, pp. 9014–9023, 2015.

[12] N. P. Galatsanos, A. K. Katsaggelos, R. T. Chin, and A. D. Hillery, "Least squares restoration of multichannel images," *IEEE Trans. Signal Process.*, vol. 39, no. 10, pp. 2222–2236, Oct. 1991.

[13] S. Henrot, C. Soussen, and D. Brie, "Fast positive deconvolution of hyperspectral images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 828–833, Feb. 2013.

[14] S. Henrot, S. Moussaoui, C. Soussen, and D. Brie, "Edge-preserving nonnegative hyperspectral image restoration," in *Proc. 38th Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2013, pp. 1622–1625.

[15] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[16] 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 Trans. Image Process.*, vol. 1, no. 3, pp. 322–336, Jul. 1992.

[17] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Rev.*, vol. 34, no. 4, pp. 561–580, Dec. 1992.

[18] P. Johnston, *InviteComputational Inverse Problems in Electrocardiology* (Advances in Computational Bioengineering). Southampton, U.K.: WIT Press, 2000, pp. 119–142.

[19] C. R. Vogel, "Non-convergence of the L-curve regularization parameter selection method," *Inverse Problems*, vol. 12, no. 4, p. 535, 1996.

[20] M. Hanke, "Limitations of the L-curve method in ill-posed problems," *BIT Numer. Math.*, vol. 36, no. 2, pp. 287–301, 1996.

[21] M. Belge, M. E. Kilmer, and E. L. Miller, "Efficient determination of multiple regularization parameters in a generalized L-curve framework," *Inverse Problems*, vol. 18, no. 4, p. 1161, 2002.

[22] T. Regińska, "A regularization parameter in discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 17, no. 3, pp. 740–749, 1996.

[23] L. Kaufman and A. Neumaier, "Regularization of ill-posed problems by envelope guided conjugate gradients," *J. Comput. Graph. Statist.*, vol. 6, no. 4, pp. 451–463, 1997.

[24] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.

[25] D. L. Donoho and Y. Tsaig, "Fast solution of ℓ_1-norm minimization problems when the solution may be sparse," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.

[26] Y. Song, D. Brie, and E. Djermoune, "Simon Henrot. Regularization parameter estimation for nonnegative hyperspectral image deconvolution: Supplementary material," CRAN, Univ. Lorraine, Tech. Rep. hal01359058, 2016.

[27] A. F. H. Goetz, "Three decades of hyperspectral remote sensing of the earth: A personal view," *Remote Sens. Environ.*, vol. 113, pp. S5–S16, Sep. 2009.

[28] L. Duponchel, W. Elmi-Rayaleh, C. Ruckebusch, and J. P. Huvenne, "Multivariate curve resolution methods in imaging spectroscopy: Influence of extraction methods and instrumental perturbations," *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 6, pp. 2057–2067, 2003.

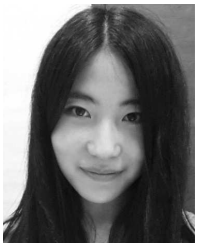
[29] S. Piqueras, L. Duponchel, R. Tauler, and A. de Juan, "Monitoring polymorphic transformations by using *in situ* Raman hyperspectral imaging and image multiset analysis," *Anal. Chim. Acta*, vol. 819, pp. 15–25, Mar. 2014.

[30] A. A. Gowen, C. P. O'Donnell, P. J. Cullen, G. Downey, and J. M. Frias, "Hyperspectral imaging—An emerging process analytical tool for food quality and safety control," *Trends Food Sci. Technol.*, vol. 18, no. 12, pp. 590–598, 2007.

[31] T. Zimmermann, J. Rietdorf, and R. Pepperkok, "Spectral imaging and its applications in live cell microscopy," *FEBS Lett.*, vol. 546, no. 1, pp. 87–92, 2003.

[32] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.*, vol. 19, no. 1, p. 010901, 2014.

[33] R. Salzer and H. W. Siesler, Eds., *Infrared and Raman Spectroscopic Imaging*. New York, NY, USA: Wiley, 2009.
[34] Y. Hiraoka, T. Shimi, and T. Haraguchi, "Multispectral imaging fluorescence microscopy for living cells," *Cell Struct. Funct.*, vol. 27, no. 5, pp. 367–374, 2002.
[35] S. Henrot, "Déconvolution et séparation d'images hyperspectrales en microscopie," Ph.D. dissertation, Centre Recherche Autom. Nancy, Univ. Lorraine, Nancy, France, 2013.
[36] S. Henrot, C. Soussen, and D. Brie, "Fast deconvolution of large fluorescence hyperspectral images," in *Proc. 3rd Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Lisbon, Portugal, Jun. 2011, pp. 1–4. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00593546
[37] A. N. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems* (Scripta Series in Mathematics). Washington, DC, USA: Winston, 1977.
[38] J. Idier, Ed., *Bayesian Approach to Inverse Problems*. New York, NY, USA: Wiley, 2008.
[39] J. Nocedal and S. Wright, *Numerical Optimization* (Springer Series in Operations Research and Financial Engineering). New York, NY, USA: Springer, 2006.
[40] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. New York, NY, USA: Wiley, 2001.
[41] I. Das and J. E. Dennis, "A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems," *Struct. Optim.*, vol. 14, no. 1, pp. 63–69, 1997.
[42] I. Y. Kim and O. L. de Weck, "Adaptive weighted-sum method for bi-objective optimization: Pareto front generation," *Struct. Multidiscipl. Optim.*, vol. 29, no. 2, pp. 149–158, 2005.
[43] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints," *IEEE Trans. Evol. Comput.*, vol. 18, no. 4, pp. 577–601, Aug. 2014.
[44] H. Jain and K. Deb, "An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: Handling constraints and extending to an adaptive approach," *IEEE Trans. Evol. Comput.*, vol. 18, no. 4, pp. 602–622, Aug. 2014.
[45] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[46] J. Kim, "Iterated grid search algorithm on unimodal criteria," Ph.D. dissertation, Dept. Statist., Virginia Polytech. Inst. State Univ., Blacksburg, VA, USA, 1997.
[47] J. M. Bardsley, J. K. Merikoski, and R. Vio, "The stabilizing properties of nonnegativity constraints in least-squares image reconstruction," *Int. J. Pure Appl. Math.*, vol. 43, no. 1, p. 95, 2008.
[48] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "Gaussian approximations of fluorescence microscope point-spread function models," *Appl. Opt.*, vol. 46, no. 10, pp. 1819–1829, 2007.
[49] P. Pankajakshan, "Blind deconvolution for confocal laser scanning microscopy," Ph.D. dissertation, INRIA, Univ. Nice Sophia Antipolis, Nice, France, 2009.



Yingying Song (S'15) was born in China in 1990. She received the Dipl.-Ing. degree in system, network and telecommunication engineering from the University of Technology of Troyes, France, in 2015. She is currently pursuing the Ph.D. degree with the Université de Lorraine, France. Her current research interests include hyperspectral image deconvolution, adaptive image processing, and hyperspectral image unmixing.

David Brie (M'13) received the Ph.D. and Habilitation Diriger des Recherches degrees from the Université de Lorraine, France, in 1992 and 2000, respectively. Since 1990, he has been with the Centre de Recherche en Automatique de Nancy, Université de Lorraine. He has been a Full Professor with the Université de Lorraine, since 2001. In 2013, he was a Visiting Researcher with the Department of Electronic and Information Engineering, Polytechnic University of Hong Kong. Since 2014, he has been serving as the Editor-in-Chief of the French journal *Traitement du Signal*. His research interests include statistical signal processing, inverse problems, and multidimensional signal processing.

El-Hadi Djermoune (M'12) was born in Amizour, Algeria, in 1974. He received the Dipl.-Ing. degree in electronics from the University of Bejaia, Algeria, in 1998, the M.S. degree in signal processing and control from the Université de Lorraine, France, in 1999, and the Ph.D. degree in 2003. Since 2004, he has been an Associate Professor with the Centre de Recherche en Automatique de Nancy, Université de Lorraine, CNRS. His research interests include statistical signal processing, inverse problems, and image processing.

Simon Henrot was born in France in 1986. He received the degree in electrical engineering from the École nationale de l'aviation civile, Toulouse, France, in 2009, and the Ph.D. degree in signal processing from the Université de Lorraine, Nancy, France, in 2013. He is currently a Post-Doctoral Student with the Grenoble Images Speech Signals and Automatics Laboratory, Université Joseph Fourier, Grenoble, France. His research interest include inverse problems in imaging and multimodal image processing.

A.3 Sparse Multidimensional Modal Analysis using a Multigrid Dictionary Refinement

S. Sahnoun, **E.-H. Djermoune**, C. Soussen, D. Brie. *EURASIP Journal on Advances in Signal Processing*, vol. 60, 2012.

Le travail présenté dans cet article est dédié à une stratégie combinant une approximation parcimonieuse et une mise à jour adaptative des atomes du dictionnaire pour estimer les paramètres d'un signal modal multidimensionnel.

EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

Sparse multidimensional modal analysis using a multigrid dictionary refinement

Souleymen Sahnoun*, El-Hadi Djermoune, Charles Soussen and David Brie

Abstract

We address the problem of multidimensional modal estimation using sparse estimation techniques coupled with an efficient multigrid approach. Modal dictionaries are obtained by discretizing modal functions (damped complex exponentials). To get a good resolution, it is necessary to choose a fine discretization grid resulting in intractable computational problems due to the huge size of the dictionaries. The idea behind the multigrid approach amounts to refine the dictionary over several levels of resolution. The algorithm starts from a coarse grid and adaptively improves the resolution in dependence of the active set provided by sparse approximation methods. The proposed method is quite general in the sense that it allows one to process in the same way mono-and multidimensional signals. We show through simulations that, as compared to high-resolution modal estimation methods, the proposed sparse modal method can greatly enhance the estimation accuracy for noisy signals and shows good robustness with respect to the choice of the number of components.

Keywords: modal estimation, multidimensional damped sinusoids, adaptive sparse approximation, multi grid

1 Introduction

The topic of sparse signal representation has received considerable attention in the last decades since it can find application in a variety of problems, including mono-and multidimensional deconvolution [1], statistical regression [2], and radar imaging [3]. Sparse approximation consists of finding a decomposition of a signal $\mathbf{y}$ as a linear combination of a limited number of elements from a dictionary $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$, i.e., finding a coefficient vector $\mathbf{x}$ that satisfies $\mathbf{y} \approx \mathbf{\Phi x}$, where $\mathbf{\Phi}$ is overcomplete ($M < N$). The sparsity condition on $\mathbf{x}$ ensures that the underdetermined problem does not have an infinite number of solutions. The dictionary $\mathbf{\Phi}$ can be chosen according to its ability to represent the signal with a limited number of coefficients or it can be imposed by the inverse problem at hand. In the latter case, we consider dictionaries whose atoms are function of some parameters. The different atoms of the dictionary are then formed by evaluating this function over a grid which has to be very fine to achieve a certain degree of resolution. This is the case for the modal estimation problem in which the atoms are formed by discretizing the frequency and damping factor axes. In this situation, the challenge is to get a good approximation without a prohibitive computational cost due to the huge size of the dictionary.

This study addresses the modal retrieval problem. This is an important topic in various applications including nuclear magnetic resonance (NMR) spectroscopy [4], wireless communications, radar, and sonar [5]. A modal signal is modeled as a sum of damped complex sinusoids. Several methods have been developed to address the modal estimation problem such as maximum likelihood [6,7] and subspace-based methods [5,8-12]. A special case of modal estimation is the harmonic retrieval problem (null damping factor) which has been formulated as a sparse approximation in a number of contributions. In the case of 1-D harmonic retrieval problem, we can cite FOCUSS [13], the method of Moal and Fuchs [14], basis pursuit [15], adaptive weighted norm extrapolation [16]. Some other contributions may be found in [17,18]. Nevertheless, only a few methods have been applied to the damped case. For instance, [19] presents a sparse estimation example of 1-D NMR (modal) data by using Lasso [20], LARS [21] and OMP [22]. Goodwin et al. [23] proposed a damped sinusoidal signal decomposition for 1-D signals using Matching Pursuit [24]. Similarly, regarding multigrid approaches associated with sparse approximation methods, only some

* Correspondence: Souleymen.Sahnoun@cran.uhp-nancy.fr
CRAN, Nancy-Université, CNRS, Boulevard des Aiguillettes, BP 70239, Vandoeuvre 54506, France

studies are considering the 1-D harmonic signals [25,26]. In the case of 2-D signals, an approach combining adaptive multigrid decomposition and TLS-Prony estimation was proposed in [27]. However, to authors knowledge, there is no study that deals with the problem of estimating parameters of multidimensional (*R*-D) damped sinusoidal signals by sparse approximation methods. This article provides a multidimensional generalization of the study presented in [28,29].

The goal of this article is to present an efficient approach that reduces the computational cost of sparse algorithms for *R*-D modal estimation problems. The main contributions of the article are as follows. (i) We propose a procedure which iteratively improves the set of atoms in the dictionary. The goal of this procedure is to improve resolution by avoiding computationally expensive operations due to the processing of large matrices; we refer to this procedure as the multigrid approach. (ii) We show how the 1-D modal retrieval problem can be addressed using sparse estimation approach by building a dictionary whose atoms are calculated by sampling the modal function over a 2-D grid (frequency and damping factor) in order to obtain all possible modes combinations. (iii) We show how to extend the sparse 1-D modal estimation problem to *R*-D modal problems.

The article is organized as follows. In Section 2, we provide background material and definitions for sparse signal representation. We present some known sparse methods and we recall the single best replacement (SBR) [30] algorithm and its advantages as compared to other algorithms such as OMP, OLS, and CoSaMP, to name a few. In Section 3, we present the multigrid dictionary refinement approach and we discuss its usefulness to accelerate computation and to improve resolution. In Section 4, we see how the 1-D modal retrieval problem may be addressed using sparse approximations and how the multigrid approach can be applied. In Section 5, we extend the sparse multigrid approach to the *R*-D modal estimation problem. In Section 6, experimental results are presented first to compare SBR to a greedy algorithm (OMP) and a solver to the basis pursuit problem. Then, the effectiveness of the multigrid approach will be illustrated on simulated 1-D and 2-D modal signals. Conclusions are drawn in section 7.

Notations: Upper and lower bold face letters will be used for matrices and column vectors, respectively. $\mathbf{A}^T$ denotes the transpose of $\mathbf{A}$. "$\odot$" will denote the Khatri-Rao product (column-wise Kronecker) and "$\otimes$" will denote the Kronecker product.

2 Sparse approximations

2.1 Key ideas of sparse approximations

Consider an observation vector $\mathbf{y} \in \mathbb{C}^M$ which has to be approximated by a sum of vectors from a matrix $\mathbf{\Phi}$ such that $\mathbf{y} \approx \mathbf{\Phi x}$, where $\mathbf{\Phi} = [\varphi_1..., \varphi_N] \in \mathbb{C}^{M \times N}$ and $\mathbf{x} \in \mathbb{C}^N$ contains coefficients that select and weight columns φ_n. We refer to $\mathbf{\Phi}$ as a *dictionary* and to $\mathbf{x}$ as a *representation* of the signal $\mathbf{y}$ with respect to the dictionary. To find an accurate approximation for any arbitrary signal $\mathbf{y}$, the dictionary has to be overcomplete, i.e., has to contain a large number of atoms. Therefore, we have to solve an underdetermined system when $M < N$. Clearly, there is an infinite number of solutions that can be used to represent $\mathbf{y}$. This is why additional conditions have to be imposed. Let us introduce the pseudo norm ℓ_0, $||\cdot||_0: \mathbb{C}^N \to \mathbb{N}$, which counts the number of non-zero components in its arguments. We say that a vector $\mathbf{x}$ is *s-sparse*, when $||\mathbf{x}||_0 = s$. In the case for an observed signal corrupted with noise, the problem of estimating the sparsest vector $\mathbf{x}$ such as $\mathbf{\Phi x}$ approximates $\mathbf{y}$ at best can be stated as an ℓ_2 - ℓ_0 minimization problem admitting two formulations:

- the constrained ℓ_2 - ℓ_0 problem whose goal is to seek the minimal error possible at a given level of sparsity $s \geq 1$:

$$\arg\min_{\|\mathbf{x}\|_0 \leq s} \{\mathcal{E}(\mathbf{x}) = \left\|\mathbf{y} - \mathbf{\Phi x}\right\|^2\} \tag{1}$$

- the penalized ℓ_2- ℓ_0 problem:

$$\arg\min_{\mathbf{x} \in \mathbb{C}^n} \{\, \mathcal{J}(\mathbf{x}, \lambda) = \mathcal{E}(\mathbf{x}) + \lambda \|\mathbf{x}\|_0\}. \tag{2}$$

The goal is to balance between the two objectives (fitting error and sparsity). Here, the solution sparsity level is controlled by the λ parameter.

The ℓ_2-ℓ_0 problem is known to yield an NP complete combinatorial problem which is usually handled by using suboptimal search algorithm. Restricting our attention to greedy algorithms, the main advantage of the ℓ_2-ℓ_0 penalized form is to allow both insertion and removal of elements in $\mathbf{x}$, while the constrained form only allows the insertion when optimization is carried through a descent approach [30,31].

A well known greedy method for sparse approximation is orthogonal matching pursuit (OMP) [22]. It minimizes iteratively the error $\mathcal{E}(\mathbf{x})$ until a stoping criterion is met. At each iteration the current estimate of the coefficient vector $\mathbf{x}$ is refined by selecting one more atom to yield a substantial improvement of the signal approximation.

There are other paradigms for solving sparse approximation problems by using ℓ_2- ℓ_p minimization for $p \leq 1$. One of these is basis pursuit (BP) [32], which is a principle for decomposing a signal into an "optimal" superposition of dictionary elements, where optimal means having the smallest ℓ_1 norm of coefficients among all such decompositions:

$$\min \|\mathbf{x}\|_1 \text{subject to } \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\| < \varepsilon. \quad (3)$$

This principle leads to approximation that can be sparse and this minimization problem can be solved via linear programming [32]. Instead of ℓ_2 - ℓ_1 penalized problem, FOCUSS algorithm [13] uses a ℓ_2 - $\ell_p$penalized criterion. For p <1, the cost function is nonconvex, and the convergence to global minima is not guaranteed. It is indicated in [33], that the best results are obtained for p close to 1, whereas the convergence is also slowest for p = 1.

In this article, we will use the SBR algorithm together with the multigrid approach. This algorithm has very interesting performance particularly in the case where the dictionary elements are strongly correlated [30], this is precisely the case with modal atoms. The algorithm is briefly recalled in the following paragraph.

2.2 SBR algorithm for penalized ℓ_2 - ℓ_0 problem

The heuristic SBR algorithm (see, Table 1) was proposed in [30] to minimize the mixed ℓ_2 - ℓ_0 cost function $\mathcal{J}(\mathbf{x}, \lambda)$ defined in (2) for a fixed parameter value λ. It is a forward-backward algorithm inspired by the SMLR method [34]. It is an iterative search algorithm that addresses the penalized ℓ_2-ℓ_0 problem for fixed λ. We denote by $\Omega \bullet n$ the insertion or removal of an index n into/from the active set Ω

$$\Omega \bullet n = \begin{cases} \Omega \cup \{n\} \text{ if } n \notin \Omega \\ \Omega \backslash \{n\} \quad \text{otherwise.} \end{cases} \quad (4)$$

At each iteration, the N possible single replacements $\Omega \bullet n$ (n = 1,..., N) are tested (i.e., N least square problems are solved to compute the minimum squared error $\mathcal{E}_{\Omega \bullet n}$ related to each support $\Omega \bullet n$), then the replacement yielding the minimal value of the cost function $\mathcal{J}(\mathbf{x}, \lambda)$, i.e., $J_{\Omega \bullet n}(\lambda) := \varepsilon_{\Omega \bullet n} + \lambda \text{Card}(\Omega \bullet n)$, is selected. In Table 1, the replacement rule is formulated by "n_k ∈ ..." in case several replacements yield the same value of $\mathcal{J}(\mathbf{x}, \lambda)$. However, this special case is not likely to occur when dealing with real data. A detailed analysis and performance evaluation can be found in [30] where it is shown that SBR performs very well in the case of highly correlated dictionary atoms (which is the case here). We note that unlike many algorithms which require to fix either a maximum number of iterations to be performed or a threshold on the squared error variation (OMP and OLS for instance), the SBR algorithm does not need any stopping condition since it stops when the cost function $\mathcal{J}(\mathbf{x}, \lambda)$ does not decrease anymore. However it requires to tune the parameter λ which is done empirically.

3 Multigrid dictionary refinement

As mentioned before, we restrict our attention to the case of modal dictionaries whose atoms are calculated by evaluating a function over a multidimensional grid, the grid dimension being equal to the number of unknown modal parameters. To achieve a high-resolution modal estimation, a possible way is to define a high resolution dictionary often resulting in a prohibitive computational burden. Rather than defining a highly resoluted dictionary, we propose to adaptively refine a coarse one through a multigrid scheme. This results in the algorithm sketched on Table 2, where the key step is the adaptation of the dictionary as a function of the previous dictionary and the estimated vector $\mathbf{x}$. The algorithm amounts to insert (resp, remove) atoms in (resp, from) $\mathbf{\Phi}$ and to re-run the sparse approximation algorithm. We propose two procedures to refine the dictionary. The first one consists in inserting new atoms in the $\mathbf{\Phi}$ matrix in the neighborhood of active ones. In other words, we first restore the signal $\mathbf{x}_{(l)}$ related to the dictionary $\mathbf{\Phi}_{(l)}$ by applying a sparse approximation method (SAM) at level l. Then we refine the dictionary by inserting atoms in between pairs of $\Phi_{(l)}$, in the neighborhood of each activated atom and we apply again the SAM at level l + 1 to restore $\mathbf{x}_{(l+1)}$ with respect to the refined dictionary $\Phi_{(l+1)}$. Thus we refine iteratively the dictionary until the maximum level $l = L$ - 1 is reached. This procedure is illustrated on Figure 1a where the dictionary atoms depend on two parameters, f and α. The disadvantage of this procedure is that the size of the dictionary is increasing as new atoms are constantly added between two resolution levels. Hence, the computational cost will be increasing. To cope with this limitation, we propose a second procedure consisting in adding new atoms as in the first procedure and deleting remote non-active ones (Figure 1b). The later multigrid approach may suffer from one main shortcoming. Indeed, removing non-active atoms excludes the possibility of further having active components in the

Table 1 SBR algorithm [30]

- **Input. A signal** $\mathbf{y} \in \mathbb{C}^M$, **a matrix** $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$ **and a scalar** λ
- **Output. A sparse coefficient vector** $\mathbf{x} \in \mathbb{C}^N$.

1. **Initialize**. Set the index set $\Omega_1 = \varnothing$, The coefficient vector $\mathbf{x}_1 = [0,..., 0]^T$ and set the counter to k = 1.
2. **Identify**. Find the replacement n_k of $\mathbf{\Phi}$ that most decreases the objective function:

$$n_k \in \arg\min\{\mathcal{J}_{\Omega_k \bullet n}(\lambda) := \mathcal{E}_{\Omega \bullet n} + \lambda \text{Card}(\Omega \bullet n)\}$$

$$\mathcal{J}_{\Omega_k \bullet n_k}(\lambda) < \mathcal{J}_{\Omega_k}(\lambda)$$

$$\Omega_{k+1} = \Omega_k \bullet n_k.$$

Table 2 Sparse multigrid algorithm

- Input. A signal $\mathbf{y} \in \mathbb{C}^M$, a matrix $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$, a scaler λ and an integer L
- Output. A sparse coefficient vector $\mathbf{x}_{L-1} \in \mathbb{C}^N$

For $l = 0$ up to $l = L - 1$

$$\mathbf{x}_l = \mathrm{SAM}\left(\mathbf{\Phi}_l, \mathbf{y}, \lambda\right)$$
$$\mathbf{\Phi}_{l+1} = \mathrm{ADAPT}\left(\mathbf{\Phi}_l, \mathbf{x}_l\right),$$

End For.

neighborhood of already suppressed atoms. A possible way to overcome this problem consists in maintaining all the atoms from the initial dictionary in all the $\Phi_{(l)}$'s.

The multigrid dictionary refinement is proposed in the context of modal analysis. However, it is worth noticing that this idea can be straightforwardly extended to any dictionary obtained by sampling a continuous function over a grid.

4 Monodimensional modal estimation using sparse approximation and multigrid

4.1 1-D data model

A 1-D complex modal signal containing F modes can be written as:

$$y(m) = \sum_{i=1}^{F} c_i a_i^m + e(m) \tag{5}$$

for $m = 0,...,M - 1$, where $(a_i = e^{(-\alpha_i + j2\pi f_i)})$, with $\{\alpha_i\}_{i=1}^F$ the damping factors and $\{f_i\}_{i=1}^F$ the frequencies. $\{c_i\}_{i=1}^F$ are complex amplitudes and $e(m)$ is an additive noise. The problem is to estimate the set of parameters $\{a_i, c_i\}_{i=1}^F$ from the observed sequence $y(m)$. Equation (5) can be written under a matrix form as:

$$\mathbf{y} = \mathbf{A}\mathbf{c} + \mathbf{e} \tag{6}$$

$$\mathbf{\Phi} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 & 1 & \cdots & 1 & \cdots & 1 \\ \phi_{1,1}(1) & \phi_{1,2}(1) & \cdots & \phi_{1,K}(1) & \phi_{2,1}(1) & \cdots & \phi_{2,K}(1) & \cdots & \phi_{P,K}(1) \\ \phi_{1,1}(2) & \phi_{1,2}(2) & \cdots & \phi_{1,K}(2) & \phi_{2,1}(2) & \cdots & \phi_{2,K}(2) & \cdots & \phi_{P,K}(2) \\ \vdots & & & \vdots & & & & & \\ \phi_{1,1}(M-1) & \phi_{1,2}(M-1) & \cdots & \phi_{1,K}(M-1) & \phi_{2,1}(M-1) & \cdots & \phi_{2,K}(M-1) & \cdots & \phi_{P,K}(M-1) \end{bmatrix}}_{N=PK \text{ atoms}} \tag{7}$$

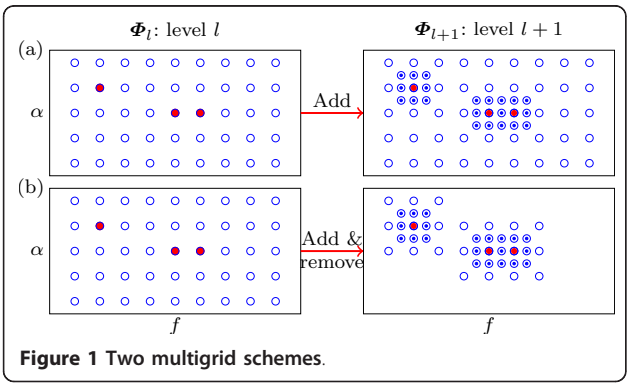


Figure 1 Two multigrid schemes.

where $\mathbf{A}$ is an $M \times F$ Vandermonde matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_F \\ a_1^2 & a_2^2 & \cdots & a_F^2 \\ \vdots & \vdots & & \vdots \\ a_1^{M-1} & a_2^{M-1} & \cdots & a_F^{M-1} \end{bmatrix}$$

and $\mathbf{c} = [c_1,...,c_F]^T$.

4.2 1-D sparse modal estimation

The problem of modal estimation is an inverse problem since $\mathbf{y}$ is given and $\mathbf{A}$,$\mathbf{c}$, and F are unknown. It can be formulated as a sparse signal estimation problem by defining the dictionary $\mathbf{\Phi}$ gathering all the possible modes obtained by sampling α (P samples) and f (K samples) on a 2-D grid. $\mathbf{\Phi}$ is expressed in (7) with $\phi_{p,k}(m) = e^{(-\alpha_p + j2\pi f_k)m}$ and $N = PK$ Provided that α and f are finely sampled, we can assume that $\mathbf{A}$ is a submatrix of $\mathbf{\Phi}$ so that $\mathbf{c}$ correspond to the nonzero elements in $\mathbf{x}$. Then the modal estimation problem can be formulated as a penalized ℓ_2 - ℓ_0 sparse signal estimation problem (2). The multigrid approach presented before can be used to that end.

5 Multidimensional modal estimation using sparse approximation and multigrid

5.1 *R*-D data model

A multidimensional complex modal signal containing F modes can be written as:

$$y(m_1, \ldots, m_R) = \sum_{i=1}^{F} c_i \prod_{r=1}^{R} a_{i,r}^{m_r} + e(m_1, \ldots, m_R) \tag{8}$$

where $m_r = 0,...,M_r - 1$ for $r = 1,...,R$. M_r denotes the sample support of the rth dimension, $a_{i,r} = e^{(-\alpha_{i,r} + j2\pi f_{i,r})}$ is the ith mode in the rth dimension, with $\{\alpha_{i,r}\}_{i=1,r=1}^{F,R}$ the damping factors and $\{f_{i,r}\}_{i=1,r=1}^{F,R}$ the frequencies, $\{c_i\}_{i=1}^F$ the complex amplitudes, and $e(m_1m_2...,m_R)$ stands for an additive observation noise. The problem is to estimate the set of parameters $\{\alpha_{i,r}\}_{i=1,r=1}^{F,R}$ and $\{c_i\}_{i=1}^F$ from the samples $y(m_1,...,m_R)$.

In order to facilitate the presentation, we rewrite the data model using the Khatri-Rao product. Given (8), we define the vector $\mathbf{y}$ as:

$$\mathbf{y} = \begin{bmatrix} y(0,0,\ldots,0) \\ y(0,0,\ldots,1) \\ \vdots \\ y(0,0,\ldots,M_R-1) \\ y(0,0,\ldots,1,0) \\ \vdots \\ y(M_1-1,M_2-1,\ldots,M_R-1) \end{bmatrix}.$$

Then, we define R Vandermonde matrices $\mathbf{A}_r \in \mathbb{C}^{M_r \times F}$ with generators $\{a_{i,r}\}_{i=1}^{F}$ such that

$$\mathbf{A}_r = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ a_{1,r} & a_{2,r} & \cdots & a_{F,r} \\ a_{1,r}^2 & a_{2,r}^2 & \cdots & a_{F,r}^2 \\ \vdots & \vdots & & \vdots \\ a_{1,r}^{M_r-1} & a_{2,r}^{M_r-1} & \cdots & a_{F,r}^{M_r-1} \end{bmatrix},$$

with $r = 1,..., R$. It can be checked that

$$\mathbf{y} = (\mathbf{A}_1 \odot \mathbf{A}_2 \odot \cdots \odot \mathbf{A}_R)\mathbf{c} + \mathbf{e} \tag{9}$$

where $\mathbf{c} = [c_1, c_2, ..., c_F]^T$ gathers the complex amplitudes and e is the noise vector.

5.2 E-D sparse modal estimation

Similar to the 1-D case, the R-D modal retrieval problem can be formulated as a sparse signal estimation problem by defining a dictionary that gathers all possible combinations of 1-D modes obtained by sampling damping factors and frequencies for each dimension on 2-D grids. Let P_r be the number of damping factors $\alpha_{1,r}$, $\alpha 2,r,...,\alpha P_{r,r}$ and K_r the number of frequencies $f_{1,r}, f_{2,r},..., f_{Kr,r}$ resulting from the sampling of the r^{th} dimension, then the corresponding dictionary is given by

$$\mathbf{\Phi}^{(r)} = \left[\phi_{1,1}^{(r)}, \ldots, \phi_{1,K_r}^{(r)}, \phi_{2,1}^{(r)}, \ldots, \phi_{2,K_r}^{(r)}, \ldots, \phi_{p_r,K_r}^{(r)}\right],$$

where $\phi_{p,k}^{(r)} = \left[\phi_{p,k}^{(r)}(0), \ldots, \phi_{p,k}^{(r)}(M_r - 1)\right]^T$ and $\phi_{p,k}^{(r)}(m_r) = e^{(-\alpha_{p,r}+j2\pi f_{k,r})m_r}$ for $p = 1,...,P_r$ and $k = 1,...,K_r$. Finally, the dictionary involved in the R-D sparse modal approximation is defined by:

$$\mathbf{\Phi} = \mathbf{\Phi}^{(1)} \otimes \mathbf{\Phi}^{2} \otimes \ldots \mathbf{\Phi}^{(R)}, \tag{20}$$

where the number of atoms is $N = \Pi_{r=1}^{R} N_r$, with $N_r = P_r K_r$. Note that the dictionary $\mathbf{\Phi}$ can be seen as a $2R$-dimensional sampling of the R-dimensional modal function. Then the R-D modal retrieval problem can be formulated as a penalized ℓ_2 - ℓ_0 sparse signal estimation problem (2).

5.3 Multigrid approach for R-D modal estimation

According to (10), the dictionary is obtained by doing the Kronecker product of R 1-D modal dictionaries. Thus, we can still use the multigrid approach presented in section 3 to adapt each 1-D dictionary to form the R-D dictionary. This results in the algorithm sketched in Table 3.

6 Experimental results

In this section, we present some experimental results for the multigrid sparse modal estimation. First, we present two examples on 1-D simulated modal signals. Next, we present and discuss results on a 2-D simulated signal and we compare them with those obtained by the subspace method "2-D ESPRIT" [5]. We chose the 2-D ESPRIT method because a comparative performance study [35] has shown that among different subspace-based high resolution modal estimation techniques, it was the one which was giving the best results.

6.1 1-D modal estimation results

First, we compare the results achieved by SBR, OMP, and the primal-dual logarithmic barrier (log-barrier) algorithm for solving the BP problem [15]. Here we used the `SparseLab`[a] implementations of OMP and BP (`SolveOMP` and `SolveBP`). Then, we present the results achieved using the multigrid SBR approach.

The first dataset is a noise-free 1-D modal signal $\mathbf{y}$ composed of $M = 30$ samples and made up of three 1-D superimposed damped complex sinusoids having the same amplitude. The 1-D modes are:

$$(f_1, \alpha_1) = (0.2, 0.050);$$
$$(f_2, \alpha_2) = (0.3, 0.025);$$
$$(f_3, \alpha_3) = (0.9, 0.050).$$

The dictionary is constructed using 20 equally spaced frequency points in the interval [0 1], where each frequency point is coupled with 20 points of damping factors in [0 0.5] and each atom represents a 1-D complex sinusoid of M samples. Thus, the dictionary $\mathbf{\Phi}$ is of size 30 × 400. We notice that the simulated 1-D modes belong to the dictionary. Thus, in the noise free case, it is possible to have an exact representation of the signal.

We estimate the parameters of $\mathbf{y}$ using SBR, OMP, and log-barrier; the results are shown in Figure 2. The representation given at the bottom of Figure 2 plots the active modes in the frequency-magnitude plane: the vertical lines are located at the frequencies of the active set Ω and their heights represent the corresponding estimated magnitudes $|\mathbf{x}_\Omega|$. The horizontal segments represent the damping factors. Clearly, the results obtained by SBR and OMP are more sparse than those achieved by the BP solver because BP detects much more than three modes. This is due to the fact that BP is an l_2- l_1 solver and thus tends to detect many atoms having low amplitudes, while OMP and SBR do not impose any l_1 penalty on the amplitudes allowing for the detection of a small number of atoms possibly having large amplitudes. SBR exactly yields the three modes (exact recovery) whereas OMP gives the true frequencies but leads to a wrong α_2. The Fourier transform of signal $\mathbf{y}$ and its estimates obtained by SBR, OMP and log-barrier algorithms are given on Figure 2 (top). We observe that unlike OMP and log-barrier, SBR correctly estimates the modal parameters of $\mathbf{y}$. Although log-barrier algorithm estimates correctly the frequencies for harmonic signals, it

Table 3 *R-D* sparse multigrid algorithm

- **Input. A signal** $\mathbf{y} \in \mathbb{C}^{M_1 M_2 \ldots M_R}$ ***R*** **matrices** $\mathbf{\Phi}_0^{(r)} \in \mathbb{C}^{M_r \times N_r}$, **a scaler** λ **and an integer** ***L***
- **Output. A sparse coefficient vector** $\mathbf{x}_{L\text{-}1} \in \mathbb{C}^N$

For $l = 0$ up to $l = L - 1$

$$\mathbf{\Phi}_l = \mathbf{\Phi}_l^{(1)} \otimes \mathbf{\Phi}_l^{(2)} \otimes \cdots \otimes \mathbf{\Phi}_l^{(R)}$$

$$\mathbf{x}_l = \mathrm{SAM}(\mathbf{\Phi}_l, \mathbf{y}, \lambda)$$

For $r = 1$ up to $r = R$

$$\mathbf{\Phi}_{l+1}^{(r)} = \mathrm{ADAPT}(\mathbf{\Phi}_l^{(r)}, \mathbf{x}_l)$$

End For
End For.

does not estimate correctly the parameters of 1-D modal signals and the solution is less sparse than the solutions provided by SBR and OMP.

In the second example, SBR algorithm coupled to multigrid approach is applied to estimate the 1-D modes from a simulated 1-D modal signal expressed in (5) with 30 samples embedded in additive Gaussian white noise such that the SNR is 23 dB. We start restoration using the same dictionary described in the first example, then we refined it with the multigrid approach. The simulated modes are:

$$(f_1, \alpha_1) = (0.19, 0.025),$$
$$(f_2, \alpha_2) = (0.23, 0.050).$$

These modes are chosen in such a way that they cannot be separated by the Fourier transform (Figure 3)

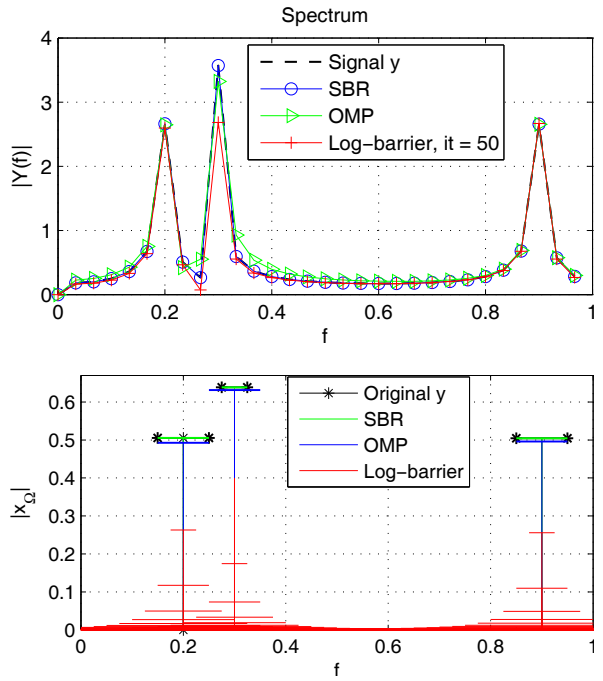
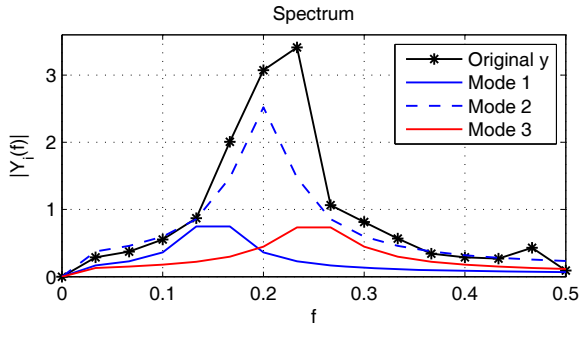
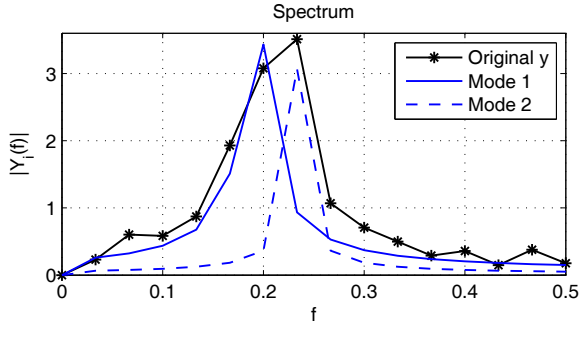


Figure 2 Spectrum of y and its estimates by SBR, OMP, and log-barrier (top), frequencies and damping factors of estimated modes (bottom).



(a) Level 1: $f_i = 0.15, 0.2, 0.25$ and $\alpha_i = 0.125, 0.15, 0.225$ resp.



(b) Level 7: $f_i = 0.1917, 0.2302$ and $\alpha_i = 0.05, 0.025$ resp.

Figure 3 Spectrum of y and estimated modes by SBR and multigrid, (a) level 1; (b) level 7, SNR = 23 dB.

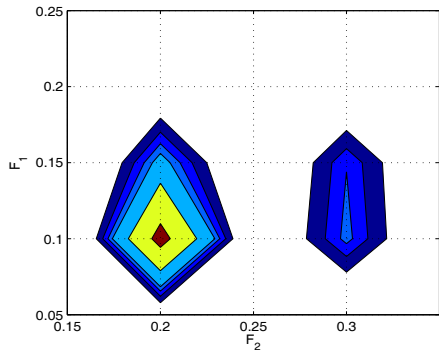
and they are not initially in the dictionary $\mathbf{\Phi}$. Figure 3a shows the spectrum of each sinusoid activated in the first level. Using the second multigrid procedure presented before, we see on Figure 3b that the two 1-D modes have been well separated in level 7, which proves the effectiveness of the approach. To give some figures about the efficiency of the multigrid approach, it is interesting to compare the size of the dictionary at the 7th degree of resolution to the uniform dictionary allowing the same resolution. For our example, $\mathbf{\Phi}_{(7)}$ is of dimension 30 × 520 while the uniform one achieving the same resolution would require a dictionary of size 30 × 6553600. This dramatic increase in the number of atoms is due to the bidimensional nature of the dictionary. Obviously, this complexity becomes huge for bi-and multidimensional modal signals.

6.2 2-D modal estimation results

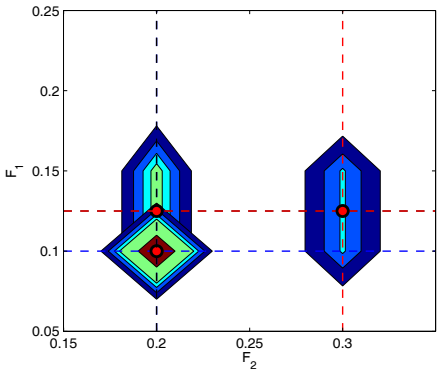
First, SBR is used in combination with the multigrid approach to estimate parameters of a 2-D simulated signal ($\mathbf{y}_{sim}$) of dimensions 20 × 20 which contain three modes with parameters:

$$(f_{1,1}, \alpha_{1,1}; f_{1,2}, \alpha_{1,2}) = (0.100, 0.00; 0.2, 0.0),$$
$$(f_{2,1}, \alpha_{2,1}; f_{2,2}, \alpha_{2,2}) = (0.125, 0.00; 0.2, 0.0),$$
$$(f_{3,1}, \alpha_{3,1}; f_{3,2}, \alpha_{3,2}) = (0.125, 0.05; 0.3, 0.1).$$

Note that the first two modes are not separated by 2-D Fourier transform. Amplitudes are (c_1, c_2, c_3) = (1,1,3) and the additive white noise variance is such that the SNR of the first mode is 7 dB (SNR_1 = 7). In the following simulations we use this simulated 2-D signal ($\mathbf{y}_{sim}$) with the same modes and amplitudes, we only change the SNR value. The spectrum of the simulated signal is represented by contour lines in Figure 4a where it is verified that the first two peaks are not separable by Fourier transform. The SBR method coupled with the proposed multigrid approach detects well the three components at the third resolution level. Their respective spectras are shown in Figure 4b. To give an idea about the gain in computational cost, the size of the dictionary at the third level is equal to

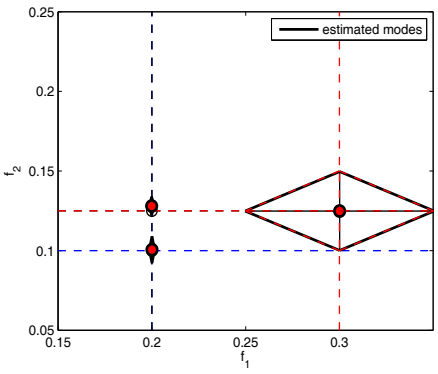


(a) Spectrum of simulated signal

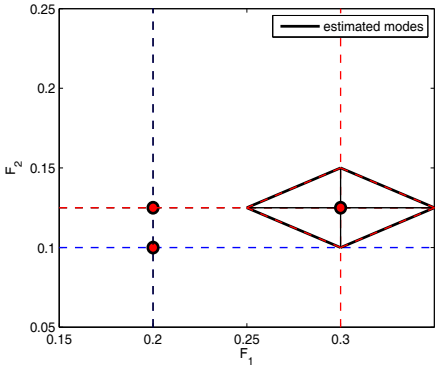


(b) Spectrum of estimated modes

Figure 4 Frequency resolution of the proposed method with SNR_1 = 7 dB. The position of simulated modes is presented by a circle.



(a) Modes estimated by 2-D ESPRIT



(b) Modes estimated by proposed algorithm

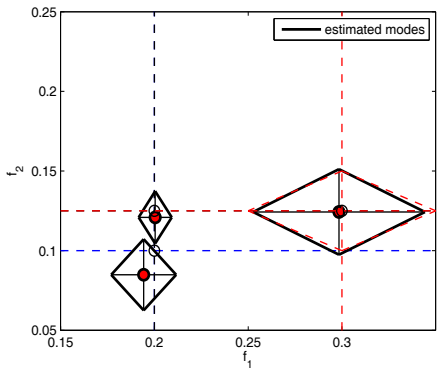
Figure 5 Comparison between 2D-ESPRIT and the proposed method with SNR_1 = 20 dB. (○) Frequencies of simulated 2-D modes; (•) Frequencies of estimated 2-D modes; (⋄) The diagonals of the diamonds represent the 2-D damping factors.

400 × 3136. The size of the uniform dictionary achieving the same resolution is 400 × 409600; the gain in term of size is a multiplicative factor 130.

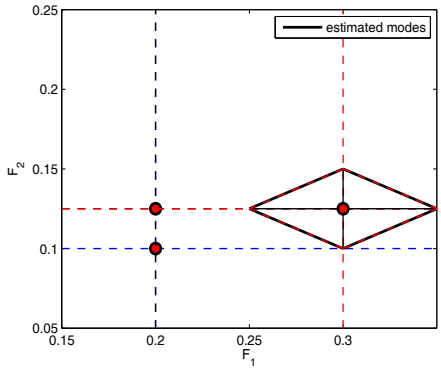
In Figure 5, we compare the estimated modes obtained by the 2-D ESPRIT [5] and our proposed technique. We use the 2-D simulated signal ($\mathbf{y}_{sim}$) with the SNR set to 20 dB. Both our technique and 2-D ESPRIT are able to separate the three modes, whereas there is a slight error made by 2-D ESPRIT on the first and second modes. In Figure 6, we decrease the SNR to 7 dB, and only the proposed algorithm is still able to estimate the three modes with an accuracy similar to what was obtained when the SNR equals 20 dB. In that case, the 2-D ESPRIT performance decreases and the modal parameters are biased.

In Figure 7, we test the sensitivity of our technique to the correct determination of the number of modes in the signal. In the previous examples, the parameter λ of the penalized cost function in SBR algorithm was fixed to 0.01 and we did not give any constraint on the number of modes to be estimated. However, in the example presented in Figure 7, we use the 2-D simulated signal with SNR equal to 20 dB, and we force 2-D ESPRIT and the proposed algorithm to estimate 5 modes (while the actual number of modes is 3). We observe that the proposed algorithm is not very sensitive to the correct determination of the number of existing modes in the sense that the true modes are activated and the other active atoms lies in the neighborhood of the true modes. On the contrary, the 2-D ESPRIT yields spurious modes located very far from the true ones.

In Figure 8, we analyze the sensitivity of the multigrid algorithm to noise power. We use the same signal $\mathbf{y}_{\mathrm{sim}}$ with different noise levels SNR_1. For each noise level we do 20 Monte Carlo trials and then we calculate the percentage of successful estimations obtained after three multigrid levels. We can see that the proposed algorithm reconstruct exactly the signal with a rate upper than

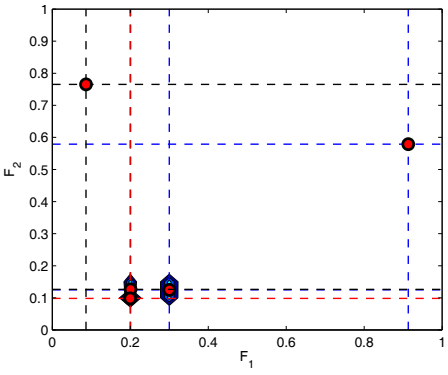


(a) Modes estimated by 2-D ESPRIT

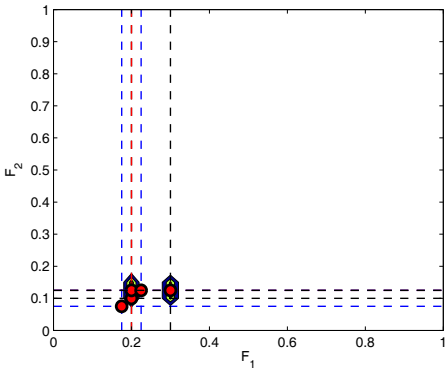


(b) Modes estimated by proposed algorithm

Figure 6 Comparison between 2-D ESPRIT and the proposed method with SNR_1 = 7 dB. (○) Frequencies of simulated 2-D modes; (•) Frequencies of estimated 2-D modes; (⋄) The diagonals of the diamonds represent the 2-D damping factors.



(a) Spectrum of modes estimated by 2-D ESPRIT



(b) Spectrum of modes estimated by proposed algorithm

Figure 7 Simulation results with number of estimated modes forced to be 5 and SNR_1 = 20 dB (on the first mode). Estimated modes are presented by red circles.

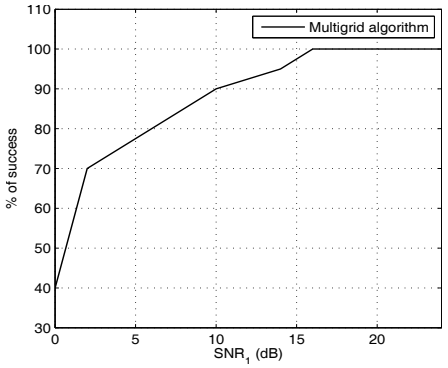


Figure 8 Rate of successful estimations with different noise levels of SNR_1.

80% for an SNR_1 more than 6 dB; and the rate of success is 100% with an SNR_1 upper than 15 dB.

7 Conclusion

We presented a multigrid technique that adaptively refines ordered dictionaries for sparse approximation. The algorithm may be associated with any sparse method, but clearly the accuracy of the final results will depend on the accuracy of the sparse approximation. Then sparse approximation associated to multigrid are used to tackle mono-and multidimensional modal (damped sinusoids) estimation problem. Thus, we applied the SBR algorithm which is shown, using simulation results, to perform better than OMP and Basis Pursuit for modal approximation. Finally, we examined performances of our proposed algorithm over existing *R*-modal estimation algorithms. It allows one to separate modes that the Fourier transform cannot resolve without a huge increase in the computational cost, improves robustness to noise and does not require initialization. As perspectives, we will study possible improvements for the sparse multigrid approach in the case of multidimensional modal signals. In particular, we can envisage to used multiple 1-D modal estimation to get a low dimension initial dictionary for *R*-D modal estimation. We also are planning to study the convergence properties of the multigrid approach and we will apply the method to the modal estimation of real NMR signals.

Endnote

[a]http://sparselab.stanford.edu.

Authors' contributions

All authors contributed to all aspects of the article. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 15 September 2011 Accepted: 8 March 2012
Published: 8 March 2012

References

1. FX Dupé, JM Fadili, JL Starck, A proximal iteration for deconvolving poisson noisy images using sparse representations. IEEE Trans Image Process. **18**(2), 310–321 (2009)
2. AJ Miller, *Subset Selection in Regression*, 2nd edn. (Chapman and Hall, London, UK, 2002)
3. M Cetin, W Karl, Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization. IEEE Trans Image Process. **10**(4), 623–631 (2001). doi:10.1109/83.913596
4. JC Hoch, A Stern, *NMR Data Processing* (Wiley-Liss, NY, 1996)
5. S Rouquette, M Najim, Estimation of frequencies and damping factors by two-dimensional ESPRIT type methods. IEEE Trans Signal Process. **49**(1), 237–245 (2001). doi:10.1109/78.890367
6. Y Bresler, A Makovski, Exact maximum likelihood parameter estimation of superimposed exponential in noise. IEEE Trans Acoustics Speech Signal Process. **35**(5), 1081–1089 (1986)
7. MP Clark, LL Scharf, Two-dimensional modal anlysis based on maximum likelihood. IEEE Trans Signal Process. **42**(6), 1443–1451 (1994). doi:10.1109/78.286959
8. R Kumaresan, DW Tufts, Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. IEEE Trans Acoustics Speech Signal Process. **30**, 833–840 (1982). doi:10.1109/TASSP.1982.1163974
9. P Stoica, A Nehorai, MUSIC, maximum likelihood, and cramer-rao bound. IEEE Trans Acoustic Speech Signal Process. **37**(5), 720–741 (1989). doi:10.1109/29.17564
10. R Roy, T Kailath, ESPRIT: Estimation of signal parameters via rotational invariance. IEEE Trans Acoustics Speech Signal Process. **37**(7), 984–995 (1989). doi:10.1109/29.32276
11. JJ Sacchini, WM Steedly, RL Moses, Two-dimensional Prony modeling and parameter estimation. IEEE Trans Signal Process. **41**(11), 3127–3137 (1993). doi:10.1109/78.257242
12. J Liu, X Liu, X Ma, Multidimensional frequency estimation with finite snapshots in the presence of identical frequencies. IEEE Trans Signal Process. **55**, 5179–5194 (2007)
13. IF Gorodnitsky, BD Rao, Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. IEEE Trans Signal Process. **45**(3), 600–616 (1997). doi:10.1109/78.558475
14. N Moal, J Fuchs, Sinusoids in white noise: a quadratic programming approach, in *IEEE Proc ICASSP*, (Seattle, WA, USA), **4**, 2221–2224 (1998)
15. SS Chen, DL Donoho, Application of basis pursuit in spectrum estimation, in *IEEE Proc ICASSP*, (Seattle, WA, USA), **3**, 1865–1868 (1998)
16. S Cabrera, T Boonsri, AE Brito, Principal component separation in sparse signal recovery for harmonic retrieval, in *Proc of the IEEE SAM Workshop*, 249–253 (2002)
17. S Bourguignon, H Carfantan, J Idier, A sparsity-based method for the estimation of spectral lines from irregularly sampled data. IEEE J Sel Topics Signal Process. **1**(4), 575–585 (2007)
18. J Fuchs, Convergence of a sparse representations algorithm applicable to real or complex data. IEEE J Sel Topics Signal Process. **1**(4), 598–605 (2007)
19. DL Donoho, Y Tsaig, Fast solution of ℓ_1-norm minimization problems when the solution may be sparse. IEEE Trans Inf. Theory **54**(11), 4789–4812 (2008)
20. R Tibshirani, Regression shrinkage and selection via the lasso. J Royal Stat Soc, Series B (Methodol.) **58**(1), 267–288
21. B Efron, T Hastie, I Johnstone, R Tibshirani, Least angle regression. Ann Statist. **32**(2), 407–499 (2004). doi:10.1214/009053604000000067
22. YC Pati, R Rezaiifar, PS Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in *1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA, USA), **1**, 40–44 (1993)
23. M Goodwin, M Vetterli, Matching pursuit and atomic signal models based on recursive filter banks. IEEE Trans Signal Process. **47**(7), 1890–1902 (1999). doi:10.1109/78.771038
24. SG Mallat, Zhang Zhifeng, Matching pursuits with time-frequency dictionaries. IEEE Trans Signal Process. **41**(12), 3397–3415 (1993). doi:10.1109/78.258082

25. S Cabrera, S Malladi, R Mulpuri, A Brito, Adaptive refinement in maximally sparse harmonic signal retrieval, in *IEEE Digital Signal Processing Workshop* 231–235 (2004)
26. D Malioutov, M Cetin, AS Willsky, A sparse signal recontruction perspective for source localization with sensor arrays. IEEE Trans Signal Process. **53**(8), 3010–3022 (2005)
27. EH Djermoune, G Kasalica, D Brie, Estimation of the parameters of two-dimensional NMR spectroscopy signals using an adapted subband decomposition, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-2008)*, Las Vegas, USA), 3641–3644 (2008)
28. S Sahnoun, E Djermoune, C Soussen, D Brie, Analyse modale bidimensionnelle par approximation parcimonieuse et multirésolution, in *GRETSI*, Bordeaux, France, (2011)
29. S Sahnoun, E Djermoune, C Soussen, D Brie, Sparse multiresolution modal estimation, in *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP-2011)*, France, 309–312 (2011)
30. C Soussen, J Idier, D Brie, J Duan, From Bernoulli-Gaussian deconvolution to sparse signal restoration. IEEE Trans Signal Process. **56**(10), 4572–4584 (2011)
31. C Herzet, A Drémeau, Bayesian pursuit algorithms, in *Proceedings of the European Signal Processing Conference (EUSIPCO-2010)*, (Aalborg, Denmark, 2010), pp. 1474–1478
32. SS Chen, D Donoho, M Saunders, Atomic decomposition by basis pursuit. SIAM J SIAM J Sci Comput. **20**(1), 33–61 (1998). doi:10.1137/S1064827596304010
33. B Rao, K Kreutz-Delgado, An affine scaling methodology for best basis selection. IEEE Trans Signal Process. **47**(1), 187–200 (1999). doi:10.1109/78.738251
34. J Kormylo, J Mendel, Maximum likelihood detection and estimation of Bernoulli-Gaussian processes. IEEE Trans Inf. Theory **28**(3), 482–488 (1982). doi:10.1109/TIT.1982.1056496
35. S Sahnoun, EH Djermoune, D Brie, A comparative study of subspace-based methods for 2-D nuclear magnetic resonance spectroscopy signals, Tech rep, CRAN, (2010)

doi:10.1186/1687-6180-2012-60
Cite this article as: Sahnoun *et al.*: **Sparse multidimensional modal analysis using a multigrid dictionary refinement.** *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:60.

A.4 Perturbation Analysis of Subspace-Based Methods in Estimating a Damped Complex Exponential

E.-H. Djermoune, M. Tomczak. *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4558–4563, 2009.

Cet article est consacré à l'analyse de variance d'algorithmes d'estimation de paramètres de signaux exponentiels complexes amortis 1–D. Les résultats principaux sont synthétisés dans le chapitre 2 de ce manuscrit.

Perturbation Analysis of Subspace-Based Methods in Estimating a Damped Complex Exponential

El-Hadi Djermoune and Marc Tomczak

***Abstract*—We present a study of mode variance statistics for three SVD-based estimation methods in the case of a single-mode damped exponential. The methods considered are namely Kumaresan–Tufts, matrix pencil and Kung's direct data approximation. Through first-order perturbation analysis, we derive closed-form expressions of the variance of the complex mode, frequency and damping factor estimates. These expressions are used to compare the different methods and to determine the optimal prediction order for matrix pencil and direct data approximation methods. Application to the undamped case shows the coherence of the results with those already stated in the literature. It is also found that the variances converge linearly towards the Cramér–Rao bound. Finally, the theoretical results are verified using Monte Carlo simulations.**

***Index Terms*—Damped exponential model, direct data approximation, linear prediction, matrix pencil, perturbation analysis.**

I. Introduction

The question of estimating model parameters of exponential signals in noise is a fundamental problem in signal processing. It has applications in several areas, including array processing, radar scattering, and nuclear magnetic resonance spectroscopy. In this context, several algorithms have been developed, including maximum likelihood approaches [1], [2] and subspace-based methods such as MUSIC [3], [4], backward linear prediction (BLP) [5], state-space [6], ESPRIT [7], and matrix pencil (MP) [8]. Statistical performances of these methods, at high signal-to-noise ratio (SNR), have also been extensively studied in the case of pure sinusoids [4], [9]–[14] and damped ones [15]–[18]. Most of these analyses are based on perturbation theory. For instance, Okhovat *et al.* [16] have studied BLP and direct data approximation (DDA) [6] methods, in the case of a single damped mode. The achieved expressions of variance come in the form of multiple sums, which is not very convenient. In [17], the authors consider the multimodal damped case using BLP. The resulting matrix expression is compact but does not give much insight about the actual performances. Finally, in [8], the MP method is studied in the multiple mode case. Here again, the variances come in the form of matrix expressions. However, the performances of the method have been clearly stated as closed-form expressions in the case of a single undamped exponential.

In a manner similar to [16] and [17], the present work uses Wilkinson's approach [19] to derive the expressions of the mode variance. The three methods discussed previously are studied in the case of a single noisy damped complex exponential. The first technique considered is the popular Kumaresan–Tufts method [5]. It performs a reduced-rank pseudoinverse of a data matrix to get backward linear prediction coefficients, from which the signal modes are obtained through polynomial rooting. The matrix pencil method, introduced by Hua and Sarkar [8], is based on a matrix prediction equation in which the data matrices have a Hankel structure similar to that found in BLP. The last method considered here is Kung's state-space direct data approximation method [6]. As will be seen in Section II, subspace-based methods that operate directly on data share a common step which amounts to find a reduced rank pseudoinverse of a data matrix. So the three aforementioned methods are studied in Section III, starting from the first-order perturbation analysis of the singular values and vectors, assuming a high SNR. Then, in Section IV, it is shown that these estimators are unbiased. Furthermore, closed-form expressions of the variance of the complex mode and the corresponding frequency and damping factor are derived. This enables us to establish the expression of the optimal tuning parameter of MP and DDA. In order to check the consistency of our results with those stated in the literature, the known equivalence between MP and DDA, for a first-order approximation, is shown again using the approach chosen. In the same manner, the frequency variance expression for an undamped exponential is given. In Section V, we demonstrate the superiority of MP and DDA over BLP in the single damped/undamped mode case, and we prove the convergence of the variances towards the Cramér–Rao bound. Finally, in Section VI, simulation results are presented to verify the theoretical expressions.

Manuscript received May 19, 2008; accepted April 29, 2009. First published May 27, 2009; current version published October 14, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Brian M. Sadler.

The authors are with the Centre de Recherche en Automatique de Nancy, Nancy-Université, CNRS, BP 239, 54506 Vandoeuvre Cedex, France (e-mail: el-hadi.djermoune@cran.uhp-nancy.fr; marc.tomczak@cran.uhp-nancy.fr).

Color versions of one or more of the figures in this correspondence are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2024030

II. Estimation Methods

The noise-perturbed exponential signal model is given by

$$\tilde{x}(n) = x(n) + e(n) = \sum_{i=1}^{M} a_i p_i^n + e(n) \tag{1}$$

for $n = 0, \ldots, N-1$. Here $p_i = \exp(\alpha_i + j\omega_i) = r_i \exp(j\omega_i)$, $i = 1, \ldots, M$ are the signal modes ($\alpha_i < 0$ and $r_i < 1$) with complex amplitudes $a_i = A_i \exp(j\phi_i)$. The term $e(n)$ is a zero-mean complex white Gaussian noise with variance σ_e^2; so the real and imaginary parts of $e(n)$ are assumed to be independent and of equal variances. Model (1) is used in this section to present the principles of the estimation techniques considered. Then, for perturbation analysis, we consider only the single-mode case, i.e., $M = 1$, and we use the following signal model instead of model (1):

$$\tilde{x}(n) = x(n) + e(n) = ap^n + e(n). \tag{2}$$

Throughout this correspondence, the notation $\tilde{X}$ refers to the noisy or perturbed version of the quantity X, i.e., $\tilde{X} = X + \Delta X$, where X is a scalar or matrix and ΔX the error term (or noise). Matrices are denoted by bold capital letters and vectors by bold lowercase letters.

A. BLP Method

The BLP method [5] is based on backward linear prediction and uses a reduced rank approximation of the data matrix in order to decrease noise influence. It is made up of the following steps.

1) Using the available data, form the system of equations

$$\tilde{\mathbf{X}}_1 \tilde{\mathbf{b}} \approx -\tilde{\mathbf{x}}_0 \tag{3}$$

where $\tilde{\mathbf{X}}_1 = [\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_L]$ and $\tilde{\mathbf{x}}_k = [\tilde{x}(k), \tilde{x}(k+1), \ldots, \tilde{x}(N-L-1+k)]^T$ for $k = 0, \ldots, L$. The vector $\tilde{\mathbf{b}}$ contains the prediction coefficients $\{\tilde{b}_i\}_{i=1}^{L}$, and $L \geq M$ is the prediction order.

2) Perform the singular value decomposition (SVD) of matrix $\tilde{\mathbf{X}}_1$ and set to 0 all but the first M largest singular values. The resulting matrix, denoted by $\hat{\mathbf{X}}_1$, is the best rank M approximation of $\tilde{\mathbf{X}}_1$ in the Frobenius norm sense.

3) Compute the prediction vector estimate $\tilde{\mathbf{b}}$ using the reduced rank pseudoinverse of $\tilde{\mathbf{X}}_1$:

$$\tilde{\mathbf{b}} = -\hat{\mathbf{X}}_1^{+}\tilde{\mathbf{x}}_0 \tag{4}$$

where the superscript "+" denotes the Moore-Penrose pseudoinverse.

4) Obtain the roots $\{\tilde{z}_i\}_{i=1}^{L}$ of the polynomial $\tilde{B}(z) = 1 + \sum_{i=1}^{L}\tilde{b}_i z^{-i} = \prod_{i=1}^{L}(1 - \tilde{z}_i z^{-1})$, and then select those located outside the unit circle. They correspond to the inverse of signal modes (i.e., $\tilde{p}_i = 1/\tilde{z}_i$ for $i = 1, \ldots, M$).

B. Matrix Pencil Method

For damped sinusoids, Hua and Sarkar's matrix pencil method [8] consists of the following steps.

1) Form two matrices $\tilde{\mathbf{X}}_0$ and $\tilde{\mathbf{X}}_1$. The matrix $\tilde{\mathbf{X}}_1$ is the same as before and $\tilde{\mathbf{X}}_0$ is a shifted version of the latter: $\tilde{\mathbf{X}}_0 = [\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_{L-1}]$.
2) Compute the low rank approximation $\hat{\mathbf{X}}_1$ of $\tilde{\mathbf{X}}_1$ using, as before, the SVD.
3) As for BLP, compute the reduced rank pseudoinverse of $\tilde{\mathbf{X}}_1$ to obtain the matrix estimate $\tilde{\mathbf{Z}}$:

$$\tilde{\mathbf{Z}} = \hat{\mathbf{X}}_1^{+}\tilde{\mathbf{X}}_0. \tag{5}$$

4) The estimates of the modes are the inverse of the M eigenvalues of $\tilde{\mathbf{Z}}$ lying outside the unit circle.

C. DDA Method

The DDA method by Kung *et al.* [6] is based on state-space formalism. The signal $x(n)$ is seen as the free response of a linear system with transition matrix $\mathbf{F}$ having eigenvalues which are the signal modes. So, the problem is to estimate the matrix $\mathbf{F}$, which can be done as follows.

1) Form the data matrix $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_L]$.
2) Obtain its SVD which is partitioned as follows:

$$\tilde{\mathbf{X}} = \tilde{\mathbf{U}}'\tilde{\mathbf{S}}'\tilde{\mathbf{V}}'^{H} = [\,\tilde{\mathbf{U}}'_1 \quad \tilde{\mathbf{U}}'_2\,]\begin{bmatrix}\tilde{\mathbf{S}}'_1 & \mathbf{0}\\ \mathbf{0} & \tilde{\mathbf{S}}'_2\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{V}}'^{H}_1\\ \tilde{\mathbf{V}}'^{H}_2\end{bmatrix} \tag{6}$$

where $\tilde{\mathbf{S}}'$ is a diagonal matrix containing the singular values of $\tilde{\mathbf{X}}$ ordered in non-increasing fashion, and $\tilde{\mathbf{S}}'_1$ is an M-by-M diagonal matrix. The superscript "H" denotes the conjugate transpose.

3) Estimate the observability matrix

$$\tilde{\boldsymbol{\Theta}} = \tilde{\mathbf{U}}'_1\tilde{\mathbf{S}}'^{1/2}_1. \tag{7}$$

An estimate of $\mathbf{F}$ is then given by

$$\tilde{\mathbf{F}} = \tilde{\boldsymbol{\Theta}}_1^{+}\tilde{\boldsymbol{\Theta}}_2 \tag{8}$$

where $\tilde{\boldsymbol{\Theta}}_1$ (respectively, $\tilde{\boldsymbol{\Theta}}_2$) is deduced from $\tilde{\boldsymbol{\Theta}}$ by eliminating the last (respectively, the first) row.

4) The M eigenvalues of $\tilde{\mathbf{F}}$ are the estimated modes.

III. Singular Value and Singular Vector Perturbations

All the methods presented before use the reduced rank pseudoinverse of data matrices ($\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{X}}$). So we start our study with the perturbation in the singular values and vectors of $\tilde{\mathbf{X}}_1$ and $\tilde{\mathbf{X}}$ for $M = 1$. We denote $\tilde{\mathbf{X}}_1 = \mathbf{X}_1 + \mathbf{E}_1$, where $\mathbf{X}_1$ and $\mathbf{E}_1$ are constructed as $\tilde{\mathbf{X}}_1$ using $x(n)$ and $e(n)$ respectively. The notations $\tilde{\mathbf{X}}_0 = \mathbf{X}_0 + \mathbf{E}_0$, $\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{E}$, and $\tilde{\mathbf{x}}_0 = \mathbf{x}_0 + \mathbf{e}_0$ are defined similarly.

A. Matrix $\tilde{\mathbf{X}}_1$

It can be shown that the noiseless data matrix $\mathbf{X}_1$ is rank 1 and its SVD is $\mathbf{X}_1 = \sigma_1\mathbf{u}_1\mathbf{v}_1^H$, where $\sigma_1 = Ar\sqrt{k_v k_u}$, $\mathbf{u}_1 = (e^{j\phi}/\sqrt{k_u})[1, p, \ldots, p^{N-L-1}]^T$, $\mathbf{v}_1 = (e^{-j\omega}/\sqrt{k_v})[1, p^*, \ldots, p^{*L-1}]^T$, $k_v = \sum_{i=0}^{L-1} r^{2i}$, and $k_u = \sum_{i=0}^{N-L-1} r^{2i}$. The singular value σ_1 is the square-root of the unique non-zero eigenvalue of the matrix $\mathbf{X}_1^H\mathbf{X}_1$, associated to the eigenvector $\mathbf{v}_1$. In the noisy case, we have

$$\tilde{\mathbf{X}}_1^H\tilde{\mathbf{X}}_1 = \mathbf{X}_1^H\mathbf{X}_1 + (\mathbf{E}_1^H\mathbf{E}_1 + \mathbf{E}_1^H\mathbf{X}_1 + \mathbf{X}_1^H\mathbf{E}_1). \tag{9}$$

At high SNR, the first-order perturbation of the eigenvalue σ_1^2 of $\mathbf{X}_1^H\mathbf{X}_1$ is given by [19]

$$\Delta\sigma_1^2 = \mathbf{v}_1^H(\mathbf{E}_1^H\mathbf{X}_1 + \mathbf{X}_1^H\mathbf{E}_1)\mathbf{v}_1 = \sigma_1(\mathbf{v}_1^H\mathbf{E}_1^H\mathbf{u}_1 + \mathbf{u}_1^H\mathbf{E}_1\mathbf{v}_1). \tag{10}$$

B. Matrix $\tilde{\mathbf{X}}$

The singular value decomposition of $\mathbf{X}$ is given by $\mathbf{X} = \sigma'_1\mathbf{u}'_1\mathbf{v}'^H_1$, where $\sigma'_1 = Ar\sqrt{k_{v'}k_{u'}}$, $\mathbf{u}'_1 = (e^{j\phi}/\sqrt{k_{u'}})[1, p, \ldots, p^L]^T$, $\mathbf{v}'_1 = (e^{-j\omega}/\sqrt{k_{v'}})[1, p^*, \ldots, p^{*N-L-1}]^T$, $k_{v'} = \sum_{i=0}^{N-L-1} r^{2i}$, and $k_{u'} = \sum_{i=0}^{L} r^{2i}$. In the case of DDA, we also need the expression of the first-order perturbation of the singular vector $\mathbf{u}'_1$. Let $\tilde{\mathbf{u}}'_1 = \mathbf{u}'_1 + \Delta\mathbf{u}'_1$, then [19]:

$$\Delta\mathbf{u}'_1 = \frac{1}{\sigma'_1}\sum_{i=2}^{L+1}(\mathbf{u}'^H_i\mathbf{E}\mathbf{v}'_1)\mathbf{u}'_i = \frac{1}{\sigma'_1}\sum_{i=2}^{L+1}\gamma_i\mathbf{u}'_i \tag{11}$$

where $\{\mathbf{u}'_i\}_{i=1}^{L+1}$ is the set of left-singular vectors of the data matrix $\mathbf{X}$, whose noisy counterparts are given by the columns of matrix $\tilde{\mathbf{U}}'$ in (6).

IV. Analysis of the Estimation Methods

A. BLP Method

At high SNR, matrix $\tilde{\mathbf{X}}_1$ is approximately rank 1, thus [16]

$$\hat{\mathbf{X}}_1^{+} \approx \frac{\hat{\mathbf{X}}_1^H}{\tilde{\sigma}_1^2}. \tag{12}$$

From (4), we can deduce that

$$\tilde{\mathbf{b}} \approx -\frac{1}{\tilde{\sigma}_1^2}\hat{\mathbf{X}}_1^H\tilde{\mathbf{x}}_0. \tag{13}$$

Since $\tilde{\mathbf{b}} = \mathbf{b} + \Delta\mathbf{b}$, and $\mathbf{b} = (-1/r\sqrt{k_v})\mathbf{v}_1$, the first-order perturbation of $\mathbf{b}$ is

$$\Delta\mathbf{b} = -\frac{1}{\sigma_1^2}(\mathbf{X}_1^H\mathbf{e}_0 + \mathbf{E}_1^H\mathbf{x}_0 + \mathbf{b}\Delta\sigma_1^2). \tag{14}$$

The error $\Delta\mathbf{b}$ in the prediction coefficients induces a shifting of the root $z_1 = 1/p$ of the polynomial $B(z)$ towards a new position $\tilde{z}_1 = z_1 + \Delta z_1$, where [20]

$$\Delta z_1 = -\sum_{k=1}^{L}\frac{z_1^{L-k}}{\prod_{i=2}^{L}(z_1 - z_i)}\Delta b_k \tag{15}$$

and $\{z_i\}_{i=2}^{L}$ are the zeros of the polynomial $B(z)$ which are different from z_1. Let $\beta_1 = \prod_{i=2}^{L}(z_1 - z_i) = (1/p^{L-1})((k_v - Lr^{2L})/(1 - r^{2L}))$ and $\mathbf{g}_1 = [z_1^{L-1}, z_1^{L-2}, \ldots, 1]^H = (r\sqrt{k_v}/p^{*L})\mathbf{v}_1$, then (15) may be rewritten as

$$\Delta z_1 = -\frac{1}{\beta_1}\mathbf{g}_1^H\Delta\mathbf{b} = -\frac{r\sqrt{k_v}}{\beta_1 p^L}\mathbf{v}_1^H\Delta\mathbf{b}. \tag{16}$$

After some straightforward calculations using (16) and (14), we finally obtain:

$$\Delta z_1 = \frac{1}{\sigma_1 \beta_1 p^L} \mathbf{u}_1^H (r\sqrt{k_v}\mathbf{e}_0 - \mathbf{E}_1 \mathbf{v}_1). \quad (17)$$

It can be seen that the estimate $\tilde{z}_1$ is unbiased since $\mathbb{E}\{\Delta z_1\} = 0$. Since $\tilde{z}_1 = 1/\tilde{p} = (\tilde{\alpha} + j\tilde{\omega})^{-1}$, we obtain using a first-order series expansion $\Delta z_1 \approx -(\Delta\alpha + j\Delta\omega)/p$. Hence, the estimates $\tilde{\alpha}$ and $\tilde{\omega}$ are also unbiased: $\mathbb{E}\{\Delta\alpha\} = \mathbb{E}\{\Delta\omega\} = 0$. Now, using the fact that $e(n)$ is zero-mean uncorrelated complex noise and after some lengthy calculations, it can be shown that $\mathbb{E}\left\{(\Delta z_1)^2\right\} = 0$, and

$$\mathbb{E}\left\{|\Delta z_1|^2\right\} = \frac{A^2\sigma_e^2}{\sigma_1^4 r^2}\left(\frac{1 - r^{2L}}{k_v - Lr^{2L}}\right)^2 (r^4 k_v^2 k_u + s_2 - 2r^2 k_v s_1) \quad (18)$$

with $s_1 = \sum_{i=0}^{m-1} i r^{2i} + m\sum_{i=m}^{N-L-1} r^{2i}$, $s_2 = \sum_{i=0}^{m-1} i^2 r^{2i} + m^2 \sum_{i=m}^{N-m-1} r^{2i} + \sum_{i=N-m}^{N-1} (N-i)^2 r^{2i}$, and $m = \min\{L, N-L\}$. Moreover, $\mathrm{var}(\Delta\alpha) = \mathrm{var}(\Delta\omega) = (r^2/2)\mathbb{E}\left\{|\Delta z_1|^2\right\}$, which implies

$$\mathrm{var}(\Delta\omega) = \frac{\sigma_e^2}{2A^2 k_v^2 k_u^2 r^4}\left(\frac{1 - r^{2L}}{k_v - Lr^{2L}}\right)^2 \left(r^4 k_v^2 k_u + s_2 - 2r^2 k_v s_1\right). \quad (19)$$

Finally, replacing all the sums leads to expression (23) given at the bottom of the page. This new result is interesting since it is easily exploitable. Namely, it may be used to compare the performances of BLP to MP and DDA under the assumption of high SNR. This will be done in Section V. For the particular case of an undamped sinusoid ($r = 1$), the frequency variance reduces to: $\mathrm{var}(\Delta\omega) = (\sigma_e^2/A^2)(2(2L+1)/3L(L+1)(N-L)^2)$ if $L \leq N/2$ and

$$\mathrm{var}(\Delta\omega) = \frac{\sigma_e^2}{A^2}\frac{2[-(N-L)^2 + 3L^2 + 3L + 1]}{3L^2(L+1)^2(N-L)}$$

if $L \geq N/2$, which is consistent with the results in [8] and [14].

B. Matrix Pencil Method

Starting from (5) and (12), and using the fact that $\tilde{\mathbf{X}}_0 = \mathbf{X}_0 + \mathbf{E}_0$, we obtain the following expression:

$$\Delta\mathbf{Z} = \tilde{\mathbf{Z}} - \mathbf{Z} \approx \frac{1}{\sigma_1^2}(-\mathbf{Z}\Delta\sigma_1^2 + \mathbf{X}_1^H\mathbf{E}_0 + \mathbf{E}_1^H\mathbf{X}_0). \quad (20)$$

The first-order perturbation of the eigenvalue z_1 of the matrix $\mathbf{Z}$ is then [19]

$$\Delta z_1 = \frac{1}{\sigma_1^2}\mathbf{v}_1^H(-\mathbf{Z}\Delta\sigma_1^2 + \mathbf{X}_1^H\mathbf{E}_0 + \mathbf{E}_1^H\mathbf{X}_0)\mathbf{v}_1. \quad (21)$$

As $\mathbf{Z} = (1/p)\mathbf{v}_1\mathbf{v}_1^H$, replacing $\Delta\sigma_1^2$ by its expression in (10), we get

$$\Delta z_1 = \frac{1}{\sigma_1 p}\mathbf{u}_1^H(p\mathbf{E}_0 - \mathbf{E}_1)\mathbf{v}_1. \quad (22)$$

Applying mathematical expectation, it leads to $\mathbb{E}\{\Delta z_1\} = 0$, which implies that the estimates $\tilde{\alpha}$ and $\tilde{\omega}$ are unbiased: $\mathbb{E}\{\Delta\alpha\} = \mathbb{E}\{\Delta\omega\} = 0$. As for BLP, it can also be shown that $\mathbb{E}\left\{(\Delta z_1)^2\right\} = 0$, and

$$\mathbb{E}\left\{|\Delta z_1|^2\right\} = \frac{A^2\sigma_e^2}{\sigma_1^4 r^2}\left((1 + r^2)s_2 - 2r^2 s_2'\right) \quad (24)$$

where $s_2' = s_2 + \sum_{i=0}^{m-1} i r^{2i} - \sum_{i=N-m}^{N-1}(N-i)r^{2i}$. Thus, $\mathrm{var}(\Delta\alpha) = \mathrm{var}(\Delta\omega)$, with

$$\mathrm{var}(\Delta\omega) = \frac{\sigma_e^2}{2A^2 k_v^2 k_u^2 r^4}\left((1 + r^2)s_2 - 2r^2 s_2'\right). \quad (25)$$

Again, replacing all the sums leads to

$$\mathrm{var}(\Delta\omega) = \frac{\sigma_e^2(1 - r^2)^3}{2A^2 r^2} \times \begin{cases} \frac{1 + r^{2N-2L}}{(1 - r^{2N-2L})^2(1 - r^{2L})}, & \text{if } L \leq \frac{N}{2} \\ \frac{1 + r^{2L}}{(1 - r^{2N-2L})(1 - r^{2L})^2}, & \text{if } L \geq \frac{N}{2}. \end{cases} \quad (26)$$

We observe that $\mathrm{var}(\Delta\omega)$ is a rational function in r^{2L}, so it is possible here to obtain the optimal value of L for which the variance is minimized. For instance, for $L \leq N/2$, the first derivative of $\mathrm{var}(\Delta\omega)$ with respect to $t = r^{2L}$ is zero when

$$t^3 + 3r^{2N}t^2 - 3r^{2N}t - r^{4N} = 0. \quad (27)$$

This is a cubic equation in t which may be solved analytically using, for example, Cardano's formula [21]. The value of t is found to be $t = r^N/\tan((\pi - \tan^{-1} r^{-N})/3)$, which implies

$$L_{\min} = \frac{N}{2} - \frac{1}{2\ln r}\ln\left(\tan\frac{\pi - \tan^{-1} r^{-N}}{3}\right). \quad (28)$$

One special case of interest is when $r = 1$, for which we obtain the well-known optimal value for an undamped sinusoid: $L_{\min} = N/3$. On the contrary, when r tends towards zero, we are confronted with a damped wave with a strong damping. In the latter case, $L_{\min}$ tends towards $N/2$. This value is also reached asymptotically (as $N \to \infty$), for any $r < 1$. So the optimal value of L lies between $N/3$ and $N/2$ and approaches $N/2$ as the damping factor increases. Of course, since $\mathrm{var}(\Delta\omega)$ is symmetric about $L = N/2$ and assuming N even, the variance reaches the same minimum at $N - L_{\min}$.

Finally, the variance in the undamped case may be derived easily from (26), and corresponds to the one presented in [8]: $\mathrm{var}(\Delta\omega) = (\sigma_e^2/A^2)(1/(N-L)^2 L)$ if $L \leq N/2$ and $\mathrm{var}(\Delta\omega) = (\sigma_e^2/A^2)(1/(N-L)L^2)$ if $L \geq N/2$.

C. DDA Method

In the single mode case, the matrices $\tilde{\mathbf{\Theta}}_1$ and $\tilde{\mathbf{\Theta}}_2$ in (8) are vectors. The pseudoinverse of $\tilde{\mathbf{\Theta}}_1$ is then $\tilde{\mathbf{\Theta}}_1^+ = (1/\tilde{\kappa})\tilde{\mathbf{\Theta}}_1^H$, in which $\tilde{\kappa} = \|\tilde{\mathbf{\Theta}}_1\|^2$. Let $\tilde{\kappa} = \kappa + \Delta\kappa$, the perturbation of κ is then

$$\Delta\kappa = \mathbf{\Theta}_1^H\Delta\mathbf{\Theta}_1 + \Delta\mathbf{\Theta}_1^H\mathbf{\Theta}_1 \quad (29)$$

from which we get

$$\tilde{\mathbf{\Theta}}_1^+ \approx \mathbf{\Theta}_1^+ + \frac{1}{\kappa}\left(\Delta\mathbf{\Theta}_1^H - \frac{1}{\kappa}\mathbf{\Theta}_1^H\Delta\kappa\right). \quad (30)$$

$$\mathrm{var}(\Delta\omega) = \frac{\sigma_e^2(1 - r^2)^3}{2A^2 r^2} \begin{cases} \frac{(1 + r^{2N-2L})[1 - (2L+1)(1 - r^2)r^{2L} - r^{4L+2}]}{(1 - r^{2N-2L})^2[1 - r^{2L} - L(1 - r^2)r^{2L}]^2}, & \text{if } L \leq \frac{N}{2} \\ \frac{(1 - r^{2N-2L})(1 + r^{2L})(1 + r^{2L+2}) - 2(N-L)(1 - r^2)(1 + r^{2N-2L})r^{2L}}{(1 - r^{2N-2L})^2[1 - r^{2L} - L(1 - r^2)r^{2L}]^2}, & \text{if } L \geq \frac{N}{2}. \end{cases} \quad (23)$$

Then, using (8) and according to (29), the following expression of the perturbation of $\mathbf{F}$ (which in our case equals the scalar p) can be derived:

$$\Delta p = \frac{1}{\kappa}\boldsymbol{\Theta}_1^H(\Delta\boldsymbol{\Theta}_2 - p\Delta\boldsymbol{\Theta}_1). \tag{31}$$

Since $\tilde{\mathbf{S}}_1'$ in (7) is a scalar, it simplifies itself in (8). Consequently, one may simply choose $\tilde{\boldsymbol{\Theta}} = \tilde{\mathbf{u}}_1'$ (and $\boldsymbol{\Theta} = \mathbf{u}_1'$), which leads to $\mathbb{E}\{\Delta p\} = 0$ in view of (11). Therefore, the estimates of the frequency and the damping factor are also unbiased. After some simplifications, we get $\mathbb{E}\left\{(\Delta p)^2\right\} = 0$, and

$$\mathbb{E}\left\{|\Delta p|^2\right\} = \frac{\sigma_e^2}{\kappa^2 A^2 k_{v'}^2 k_{u'}^2 r^2}\left((1+r^2)s_2 - 2r^2 s_2'\right). \tag{32}$$

Finally, using the fact that $\mathrm{var}(\Delta\alpha) = \mathrm{var}(\Delta\omega) = (1/2r^2)\mathbb{E}\left\{|\Delta p|^2\right\}$, $\kappa = k_v/k_{u'}$ and $k_{v'} = k_u$, we obtain

$$\mathrm{var}(\Delta\omega) = \frac{\sigma_e^2}{2A^2 k_v^2 k_u^2 r^4}\left((1+r^2)s_2 - 2r^2 s_2'\right). \tag{33}$$

From (33) and (25), it appears that DDA and MP are equivalent for a first-order approximation. Of course, this result had already been established in [22] for the general case of $M \geq 1$ exponentials, using a different approach. So all the properties demonstrated before on MP are also valid for DDA.

V. Performance Comparison and Convergence to the Cramér–Rao Bound

Here we demonstrate the superiority of MP and DDA over BLP, for a first-order approximation. Then, we discuss the asymptotic convergence of all the variances towards the Cramér–Rao lower bound (CRB).

Proposition 1: For all $r \in (0,1]$ and $1 \leq L \leq N-1$, the frequency variance achieved by BLP is at least equal to that obtained by MP and DDA for a first-order approximation, that is

$$\mathrm{var}(\Delta\omega)^{\mathrm{BLP}} \geq \mathrm{var}(\Delta\omega)^{\mathrm{MP}} = \mathrm{var}(\Delta\omega)^{\mathrm{DDA}}. \tag{34}$$

Proof: See Appendix I. ∎

This result was already established in the case of a single pure sinusoid (see, e.g., [8]) and is now proven in the damped case.

For any unbiased estimator of the frequency and the damping factor, under the assumption of a single exponential model, the CRB is given by [23]:

$$\mathrm{CRB}(\alpha) = \mathrm{CRB}(\omega) = \frac{\sigma_e^2}{2A^2}\frac{(1-r^2)^3(1-r^{2N})}{r^2(1-r^{2N})^2 - N^2 r^{2N}(1-r^2)^2}. \tag{35}$$

As $N \to \infty$, the convergence of all the previous estimation variances towards the CRB is obvious in view of (23), (26), and (35), for any given value of L such that $L = \mu N$, $\mu \in (0,1)$, provided that $r < 1$. The following proposition gives the rate of convergence.

Proposition 2: Let ε_N be the deviation of one of the variances in (23), (26) or (33) from the CRB: $\varepsilon_N = \mathrm{var}(\Delta\omega) - \mathrm{CRB}(\omega)$. Then, $\forall r \in (0,1)$ and $L = \mu N$, $\mu \in (0,1)$, ε_N converges linearly towards zero with the following rates:

$$\lim_{N\to\infty}\frac{\varepsilon_{N+1}}{\varepsilon_N} = \begin{cases} r^{2\mu} & \text{if } \mu \in \left(0, \frac{1}{2}\right] \\ r^{2(1-\mu)} & \text{if } \mu \in \left[\frac{1}{2}, 1\right). \end{cases} \tag{36}$$

Proof: See Appendix II. ∎

Proposition 2 gives the rate of convergence to the CRB if L is chosen such that $L = \mu N$. For instance, the rate of convergence is r for $L = N/2$, and $r^{2/3}$ for $L \in \{N/3, 2N/3\}$. In the same manner, it can be shown that, for the particular case of MP and DDA, the minimum variance corresponding to $L_{\min}$ in (28) converges to the CRB with rate r.

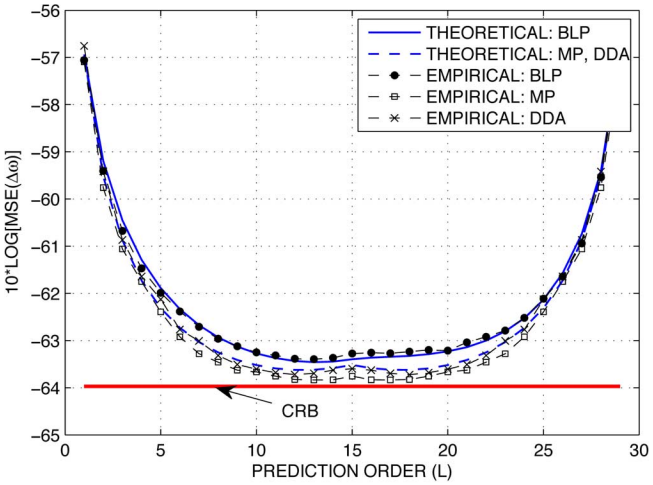


Fig. 1. Theoretical and empirical MSEs for BLP, MP and DDA versus prediction order (SNR = 40 dB).

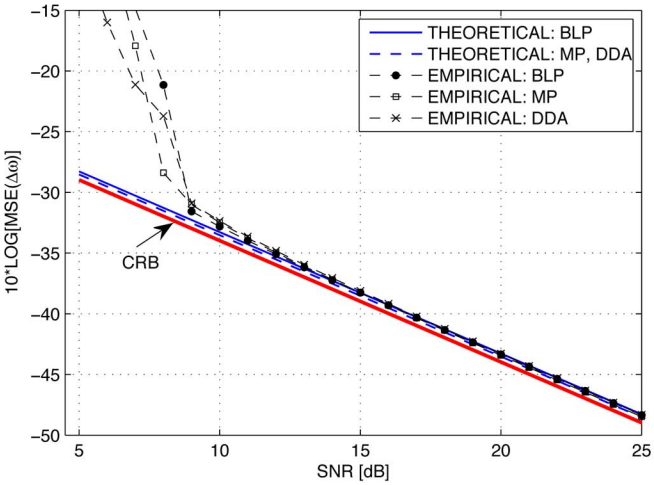


Fig. 2. Theoretical and empirical MSEs for BLP, MP and DDA versus SNR ($L = 10$).

VI. Numerical Simulations

We consider a signal composed of one damped exponential with parameters $\alpha = -0.1$, $\omega = 0.2$, and $N = 30$. The peak SNR is fixed to 40 dB: $10\log(A^2/\sigma_e^2) = 40$. Fig. 1 shows the theoretical and sample mean square errors (MSEs) obtained from 1000 realizations of additive noise for BLP, MP and DDA, together with the CRB. It can be seen that the theoretical MSEs are close to the estimated ones. Moreover, it appears clearly, as demonstrated analytically before, that MP and DDA perform better than BLP. Note that this result is valid whatever the value of the damping, assuming a sufficiently high SNR. We also observe that the minimum MSE for MP and DDA is attained at $L = 12$ and $L = 18$, which correspond to the values obtained from expression (28): $L_{\min} = 12.44$ and $N - L_{\min} = 17.66$.

For the second example, the same signal is used but the SNR is now varying. The prediction order is set to $L = 10$. The results achieved are given in Fig. 2. Here we observe that the theoretical expressions of the variance are valid beyond a threshold SNR, which in this case is about 8 dB. Of course, this is not a rule of thumb because, in fact, it also depends on the damping factor.

VII. Conclusion

In this correspondence, we have presented a first-order perturbation analysis of three subspace-based techniques operating directly on the data: Kumaresan–Tufts, matrix pencil and direct data approximation, in the case of a single damped exponential. We have derived the analytical closed-form expressions of the mode, frequency and damping factor variances. Thanks to these expressions, we have shown that MP and DDA perform better than BLP. Moreover, we have found the optimal prediction order for MP and DDA. In fact, this order depends not only on the number of samples but also on the damping, which is unknown. So, in practice, an appropriate value will lie between $N/3$ and $N/2$, the latter being preferable for a strongly damped sinusoid. Note that, unless some restrictive hypotheses on the signal parameters are stated, the expression of the optimal prediction order cannot be found for BLP, due to the nonlinear aspect of the underlying problem. The extension of the results to the multimodal case is possible provided that the modes are separated enough and the damping factors are of the same order of magnitude.

Appendix I
Proof of Proposition 1

Here we prove (34) for $r \in (0,1)$. The case $r = 1$ is simple and will not be considered. For $L \leq N/2$, the ratio of BLP and MP variances is

$$\frac{\mathrm{var}\,(\Delta\omega)^{\mathrm{BLP}}}{\mathrm{var}\,(\Delta\omega)^{\mathrm{MP}}} - 1 = r^{2L}\frac{r^2(1-r^{2L})^2 - L^2(1-r^2)^2 r^{2L}}{[1-r^{2L}-L(1-r^2)r^{2L}]^2}. \quad (37)$$

Since $\sum_{i=0}^{L-1} r^{2i} = (1-r^{2L})/(1-r^2)$ and $\prod_{i=0}^{L-1} r^{2i} = r^{L(L-1)}$, using the arithmetic mean-geometric mean inequality, we obtain $(1-r^{2L}) \geq L(1-r^2)r^{L-1}$. Thus, $r^2(1-r^{2L})^2 \geq L^2(1-r^2)^2 r^{2L}$.

For $L \geq N/2$, the ratio of BLP and MP variances is

$$\begin{aligned}\frac{\mathrm{var}\,(\Delta\omega)^{\mathrm{BLP}}}{\mathrm{var}\,(\Delta\omega)^{\mathrm{MP}}} - 1 &= \frac{r^{2L}(1+r^{2L})[r^2(1-r^{2L})^2 + 2L(1-r^2)(1-r^{2L}) - L^2(1-r^2)^2 r^{2L}]}{(1+r^{2L})[1-r^{2L}-L(1-r^2)r^{2L}]^2} \\ &\quad - \frac{2r^{2L}(N-L)(1-r^2)(1+r^{2N-2L})(1-r^{2L})^2}{(1-r^{2N-2L})(1+r^{2L})[1-r^{2L}-L(1-r^2)r^{2L}]^2}. \quad (38)\end{aligned}$$

Using, once again, the arithmetic mean-geometric mean inequality on the sequence $\{r^{2n}\}_{n=0}^{L-1}$, it yields $L^2(1-r^2)^2 r^{2L} \leq r^2(1-r^{2L})^2$, so

$$\begin{aligned}\frac{\mathrm{var}\,(\Delta\omega)^{\mathrm{BLP}}}{\mathrm{var}\,(\Delta\omega)^{\mathrm{MP}}} - 1 &\geq \frac{2r^{2L}(1-r^2)(1-r^{2L})^2}{(1+r^{2L})[1-r^{2L}-L(1-r^2)r^{2L}]^2} \\ &\quad \times \left[\frac{L(1+r^{2L})}{1-r^{2L}} - \frac{(N-L)(1+r^{2N-2L})}{1-r^{2N-2L}}\right]. \quad (39)\end{aligned}$$

It can be shown that the sequence $w_n(r) = n(1+r^{2n})/(1-r^{2n})$, $n \geq 1$, increases for all $r \in (0,1)$. So, using the fact that $L \geq N-L$, we obtain $w_L(r) \geq w_{N-L}(r)$, which completes the proof of Proposition 1.

Appendix II
Proof of Proposition 2

In this appendix, we demonstrate the linear convergence of the variance towards the CRB for the MP method with parameter $L = \mu N$ such that $0 < \mu \leq 1/2$. For $L \leq N/2$, we have

$$\varepsilon_N = \frac{\sigma_e^2(1-r^2)^3}{2A^2r^2} \times \frac{W_N}{(1-r^{2(1-\mu)N})^2(1-r^{2\mu N})[r^2(1-r^{2N})^2 - N^2r^{2N}(1-r^2)^2]} \quad (40)$$

where $\varepsilon_N = \mathrm{var}\,(\Delta\omega) - \mathrm{CRB}(\omega)$, and

$$W_N = r^{2\mu N}\left\{r^2(1-r^{2N})\left[1 - r^{2(1-2\mu)N}(3 + 3r^{2\mu N} + r^{2(1-\mu)N}\right] - N^2 r^{2(1-\mu)N}(1+r^{2(1-\mu)N})\right\}. \quad (41)$$

Since $\mu \leq 1/2$, it is easy to see that $\varepsilon_N \to 0$ as $N \to \infty$, $\forall r \in (0,1)$. The rate of convergence is then

$$\rho = \lim_{N\to\infty}\frac{\varepsilon_{N+1}}{\varepsilon_N} = \lim_{N\to\infty}\frac{W_{N+1}}{W_N} = \lim_{N\to\infty}\frac{r^{2\mu(N+1)}}{r^{2\mu N}} = r^{2\mu}. \quad (42)$$

This shows that the convergence is linear since $\rho \in (0,1)$. The cases of BLP and $\mu > 1/2$ may be proved similarly.

References

[1] Y. Bresler and A. Makovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1081–1089, Oct. 1986.

[2] S. F. Yau and Y. Bresler, "Maximum likelihood parameter estimation of superimposed signals by dynamic programming," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 804–820, Feb. 1993.

[3] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectral Estimation Workshop*, Rome, NY, 1979, pp. 243–258.

[4] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.

[5] R. Kumaresan and D. W. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 30, no. 6, pp. 833–840, Dec. 1982.

[6] R. Kung, K. S. Arun, and D. V. B. Rao, "State-space and singular value decomposition-based approximation methods for the harmonic retrieval problem," *J. Opt. Soc. Amer.*, vol. 73, no. 12, pp. 1799–1811, 1983.

[7] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT—A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1340–1342, Oct. 1986.

[8] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.

[9] Y. Hua and T. K. Sarkar, "Perturbation analysis of TK method for harmonic retrieval problem," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 228–240, Feb. 1988.

[10] B. D. Rao, "Perturbation analysis of an SVD-based linear prediction method for estimating the frequencies of multiple sinusoids," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 7, pp. 1026–1035, Jul. 1988.

[11] P. Stoica, T. Söderström, and F. Ti, "Overdetermined Yule–Walker estimation of the frequencies of multiple sinusoids: Accuracy aspects," *Signal Process.*, vol. 16, pp. 155–174, 1989.

[12] B. D. Rao and K. V. S. Hari, "Performance analysis of ESPRIT and TAM in determining the direction of arrival of plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1990–1995, Dec. 1989.

[13] A. L. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors, Part I: The MUSIC algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1758–1774, Jul. 1992.

[14] B. D. Rao and K. S. Arun, "Model based processing of signals: A state space approach," *Proc. IEEE*, vol. 80, pp. 283–309, 1992.

[15] B. Porat and B. Friedlander, "On the accuracy of the Kumaresan–Tufts method for estimating complex damped exponentials," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, pp. 231–235, Feb. 1987.

[16] A. Okhovat and J. R. Cruz, "Statistical analysis of the Tufts–Kumaresan and principal Hankel components methods for estimating damping factors of single complex exponentials," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, May 1989, pp. 2286–2289.

[17] A. C. Kot, D. W. Tufts, and R. J. Vaccaro, "Analysis of linear prediction by matrix approximation," *IEEE Trans. Signal Process.*, vol. 41, no. 11, pp. 3174–3177, Nov. 1993.
[18] W. M. Steedly, C.-H. J. Ying, and R. L. Moses, "Statistical analysis of TLS-based Prony techniques," *Automatica*, vol. 30, no. 1, pp. 115–129, 1994.
[19] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. Oxford, U.K.: Oxford Univ. Press, 1965.
[20] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
[21] E. W. Weisstein, *The CRC Concise Encyclopedia of Mathematics*. Boca Raton, FL: Chapman & Hall, 1999.
[22] Y. Hua and T. K. Sarkar, "On SVD for estimating generalized eigenvalues of singular matrix pencil in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 39, no. 4, pp. 892–900, Apr. 1991.
[23] Y.-X. Yao and S. M. Pandit, "Cramér-Rao lower bounds for a damped sinusoidal process," *IEEE Trans. Signal Process.*, vol. 43, no. 4, pp. 878–885, Apr. 1995.

Noisy FIR Identification as a Quadratic Eigenvalue Problem

Roberto Diversi

***Abstract*—This correspondence describes a method for identifying FIR models in the presence of input and output noise. The proposed algorithm takes advantage of both the bias compensation principle and the instrumental variable method. It is based on a nonlinear system of equations whose unkowns are the FIR coefficients and the input noise variance. This system allows mapping the noisy FIR identification problem into a quadratic eigenvalue problem. The identification problem is thus solved without requiring the use of iterative least-squares algorithms. The performance of the proposed approach has been tested and compared with that of other identification methods by means of Monte Carlo simulations.**

***Index Terms*—Finite-impulse-response (FIR) models, noisy input-output data, quadratic eigenvalue problem, system identification.**

I. Introduction

Many system engineering and signal processing problems involve the identification of finite-impulse-response (FIR) models [1]–[4]. The estimation of FIR models from noise-corrupted output measurements has been extensively treated in the literature. Nevertheless, the presence of noise on the input arises in many practical situations, as described in [5]. In particular, the sampling and quantization of the input often lead to a broadband noise that can be considered as independent of the quantized signal [5].

It is well known that the least squares (LS) approach gives biased parameter estimates when the input is affected by noise. On the contrary, the total least squares (TLS) method, that is based on a generalized eigenvalue problem, leads to consistent estimations but requires the *a priori* knowledge of the ratio between the input and output noise variances [5]–[10].

Manuscript received January 16, 2009; accepted May 27, 2009. First published June 23, 2009; current version published October 14, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hideaki Sakai.

The author is with the Department of Electronics, Computer Science and Systems, University of Bologna, 40136 Bologna, Italy (e-mail: roberto.diversi@unibo.it).

Digital Object Identifier 10.1109/TSP.2009.2026069

To overcome the aforementioned drawbacks, it is possible to rely on the bias compensation principle [11], [12]. It consists in estimating the noise variances in order to remove the effect of the noise-induced bias in the standard LS estimate. It is worth to stress that the particular structure of the FIR models allows to remove the LS bias by using only an estimate of the input noise variance.

Some bias-compensation based approaches have been recently proposed for identifying noisy input-output FIR models [13]–[17]. In particular, the methods described in [13], [15] and [17] rely on iterative least-squares algorithms where the current estimate of the input noise variance is used to improve the estimate of the FIR parameters and *vice versa*. An eigen analysis based technique is proposed in [14] while the approach introduced in [16] consists in searching for the solution of the identification problem inside a set of solutions compatible with the second order statistics of the data.

Another approach for identifying systems with noisy input is the instrumental variable (IV) method [18], [19]. Most of IV algorithms use delayed input and output as instruments and lead to a set of high-order Yule–Walker equations. IV methods are applicable under fairly general noise conditions and are computationally very efficient. However, the accuracy of the obtained parameter estimates is often poor since high-lag auto and cross covariances must be estimated [20].

This correspondence proposes a method that takes advantage of both the bias compensation principle and the instrumental variable approach. The method is based on a nonlinear system of equations whose unkowns are the FIR coefficients and the input noise variance. This system is similar to the set of noise-compensated Yule-Walker equations considered by Davila in [21] with reference to the identification of autoregressive models in the presence of additive noise. This set allows mapping the noisy FIR identification problem into a quadratic eigenvalue problem that, in turn, can be mapped into a linear generalized eigenvalue problem. The FIR parameters are thus estimated without requiring the use of iterative least-squares algorithms. The performance of the proposed approach has been tested and compared with that of other identification methods by means of Monte Carlo simulations.

The correspondence is organized as follows. Section II defines the noisy FIR identification problem. Section III describes a set of noise-compensated Yule-Walker equations where the unknowns are the FIR coefficients and the input noise variance. Section IV shows that this set leads to a quadratic eigenvalue problem and proposes an algorithm for solving it. In Section V the effectiveness of the proposed algorithm is compared with that of other methods by means of Monte Carlo simulations. Some conclusions are finally reported in Section VI.

II. Problem Statement

Consider the linear FIR system described by the difference equation

$$d(k) = H(q^{-1})\, x(k) \tag{1}$$

where $x(k)$ and $d(k)$ denote the noise-free input and output while $H(q^{-1})$ is the following polynomial in the backward shift operator q^{-1} ($q^{-1}\, x(k) = x(k-1)$)

$$H(q^{-1}) = h_0 + h_1\, q^{-1} + \cdots + h_{M-1}\, q^{1-M}. \tag{2}$$

The observations of $x(k)$ and $d(k)$ are both affected by additive noise so that only the signals

$$u(k) = x(k) + n_i(k) \tag{3}$$

$$y(k) = d(k) + n_o(k) \tag{4}$$

are available (see Fig. 1).

A.5 Modeling of MIG/MAG Welding with Experimental Validation using an Active Contour Algorithm Applied on High Speed Movies

J.-P. Planckaert, **E.-H. Djermoune**, D. Brie, F. Briand, F. Richard. *Applied Mathematical Modelling*, vol. 34, no. 4, pp. 1004–1020, 2010.

*Cet article présente un modèle physique simplifié d'un procédé de soudage MIG/MAG en mode court-circuit (*short-arc*). Le modèle est constitué de deux états continus : un état d'arc pendant lequel le métal est fondu et un état de court-circuit pendant lequel ce métal est transféré dans le bain. La transition entre les deux états est contrôlée par une variable de commutation. Le modèle hybride proposé est validé par des séquences de vidéo rapides à partir desquels la géométrie de la goutte est estimée par un algorithme de contour actif.*

Applied Mathematical Modelling 34 (2010) 1004–1020

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Modeling of MIG/MAG welding with experimental validation using an active contour algorithm applied on high speed movies

Jean-Pierre Planckaert [a,b], El-Hadi Djermoune [a,*], David Brie [a], Francis Briand [b], Frédéric Richard [b]

[a] Centre de Recherche en Automatique de Nancy, Nancy-Université, CNRS, Boulevard des Aiguillettes, B.P. 239, 54506 Vandoeuvre-lès-Nancy Cedex, France
[b] Air Liquide–CTAS, 13 rue des Epluches, St-Ouen l'Aumône, 95315 Cergy-Pontoise Cedex, France

ARTICLE INFO

Article history:
Received 2 June 2008
Received in revised form 18 June 2009
Accepted 23 July 2009
Available online 30 July 2009

Keywords:
MIG/MAG welding
Physical modeling
Dynamic active contour

ABSTRACT

This paper investigates some issues in physical modeling of metal inert gas/metal active gas (MIG/MAG) welding process in the short arc mode. In this mode, a metal supply is molten in the arc state and then transferred to the weld pool during the short-circuit state. A hybrid model having two distinct continuous states whose switchings are controlled by two guard conditions is proposed. Due to the complexity of the physical phenomena involved in the welding process, simplifications are used to obtain a model accounting for the main physical contributions but simple enough to yield an efficient, fast and numerically tractable simulator which can be used intensively for evaluating different control strategies. In an attempt to validate the proposed model, different measurements have been made including supply voltage and current sampled synchronously with high speed digital video. In order to extract some relevant quantities representative of the metal transfer from image sequences, an active contour algorithm is developed and tested. The effectiveness of the proposed model in the prediction of major tendencies of a welding process, especially in the arc state, is shown using experimental data. Some limitations of the model during the metal transfer are also stressed and possible remedies are then proposed.

1. Introduction

The metal inert gas/metal active gas (MIG/MAG) welding is an arc welding process in which additional metal is brought by a roll of wire line and is molten by Joule effect and an electric arc [1]. In the short arc mode, the weld is made by successive drops. An inert gas, generally argon based gas (MIG welding), or active gas, generally CO_2 based gas (MAG welding) is used as plasma for electric arc outbreak and as protective atmosphere for metal at high temperature, avoiding contamination of the metal by oxygen and nitrogen. The welding generator supplies the electric energy required for melting and for arc outbreak and sustainment between wire and workpieces to weld. It works according to two distinct control modes: (1) the arc mode in which the voltage delivered by the generator is controlled to reach a set point chosen by the welder; (2) the short-circuit mode in which the current follows a pre-defined law.

The general framework of this study concerns the modeling of the MIG/MAG welding process for a metal transfer in short arc mode. It aims at providing a model of the whole system including wire, gas, arc, sheet metal and control. Not only such a model is expected to give some insights into the understanding of the welding process but its implementation in the simulator is intended to assess the effectiveness of different control strategies as well as the effects of the experimental setup on

* Corresponding author.
E-mail addresses: jean-pierre.planckaert@airliquide.com (J.-P. Planckaert), el-hadi.djermoune@cran.uhp-nancy.fr (E.-H. Djermoune), david.brie@cran.uhp-nancy.fr (D. Brie), francis.briand@airliquide.com (F. Briand), frederic-p.richard@airliquide.com (F. Richard).

0307-904X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.apm.2009.07.011

the control performances. Hence, the model should be precise enough to produce the major tendencies but simple enough to result in a fast simulator which can be used intensively.

The molten metal drop detachment from an electrode in gas metal arc welding involves complex interactions between different physical phenomena. Some authors have studied the electromagnetic effects [2–5], and others the thermal effects [6] and the fluid dynamics [7]. As a consequence, sophisticated mathematical models have been developed, including arc plasma generation, electrode melting, droplet formation, detachment and impingement onto the base metal. Generally speaking, the most used numerical techniques belong to the class of Eulerian methods such as volume of fluid or level set methods. These methods enable one to take into account all physical phenomena with limited assumptions on the droplet shape, stream lines, etc. For instance, Choi et al. used the volume of fluid to simulate the spray and globular metal transfer modes [7], as well as the short-circuit mode [8]. The model have been extended by Zhu et al. [9], in the spray and globular modes, to account for the thermal phenomena in both the droplet and the molten pool. Although these models generally achieve a precise simulation of the MIG/MAG process, they suffer from some practical difficulties which are closely linked to the calculation time/precision trade-off and may suffer from instability problems. Hence, some simplified models have been also proposed assuming, for instance, a flat weld-pool surface, truncated spherical droplets, stable transfer, etc. [10–12]. These models are clearly less precise but account for the main tendencies. Inspired by these articles, we propose a hybrid model accounting for the switching from the arc mode to the short-circuit mode [13] taking into account the main forces acting on the molten metal during a complete cycle, and we validate it by comparison with experimental data. The recordings are realised on a test bed equipped with an acquisition system measuring voltage, current and wire feed speed. All these measurements are sampled synchronously with a high speed video system operating at 10,000 frames per second. The video images are processed to measure some relevant quantities for model validation, such as molten metal bridge volume, bridge minimal diameter or surface of contact between bridge and weld pool. In that respect, it is necessary to perform a segmentation of the images. However our images are too complex to expect local, low-level operations to generate perfect primitives. Higher level features have to be used to get a better delineation of objects. So, we chose to use deformable models which adapt to the data, which is the case of active contour algorithms [14]. This step enables one to measure indirectly the geometry of the drops that is inaccessible by the acquisition system. The obtained variables are used here to validate the proposed model, but they may be also exploited to build a database from which model identification, with physical prior, can be achieved.

The paper is organised as follows: In Section 2, a description of the physical aspect of Gas Metal Arc Welding (GMAW) in the short arc mode is given. After having introduced the main forces acting in the process, we chose the variables of interest in order to implement a model of the system. Next, in Section 3, we introduce the active contour methodology allowing us to observe these variables. Experimental results are given in Section 4. They show the approach effectiveness (as well as its limitation) to give access to the dynamic of the chosen variables. Finally, the conclusion raises some issues to be developed in future work.

2. GMAW system in short arc mode

2.1. General description

Short-circuiting metal transfer is a metal transfer whereby a continuously fed wire electrode is deposited during electrical short-circuits. The transfer of a single molten droplet of electrode occurs during the shorting phase of the transfer cycle when physical contact with the molten weld pool occurs.

Fig. 1 illustrates the time evolution of a droplet together with the corresponding arc voltage and welding current. Metal transfer goes through five steps:

- A: The electrode makes physical contact with the molten pool. The arc voltage approaches zero and the current level increases. The rate of rise to the peak current is affected by the amount of applied inductance.
- B: This point demonstrates the effect of electromagnetic forces that are applied around the electrode. This force pinches the electrode.
- C: This is the point where the bridge of molten metal explodes. The droplet is forced from the tip of the electrode to the welding pool.
- D: The molten droplet reforms while current is at its background level.
- E: The electrode is once again making contact with the pool, preparing for the transfer of another droplet.

The area of the welding arc, sketched in Fig. 2, is a region of high complexity that is comprised of physical forces and chemical reactions. The interaction of the components of the arc affects metal transfer and the quality of the finished weld. The behavior of the arc is influenced by: the type and the diameter of the filler metal, the base metal conditions, the shielding gas, the welding parameters (voltage and current) and the interaction of physical forces.

The phenomenon of metal transfer in short arc welding can be seen as a hybrid system as sketched in Fig. 3. Indeed it presents two continuous states: a first one during which the droplet grows and a second one during which the electrode is in physical contact with the workpiece. The jump between the two states is linked to the fulfillment of a guard condition:

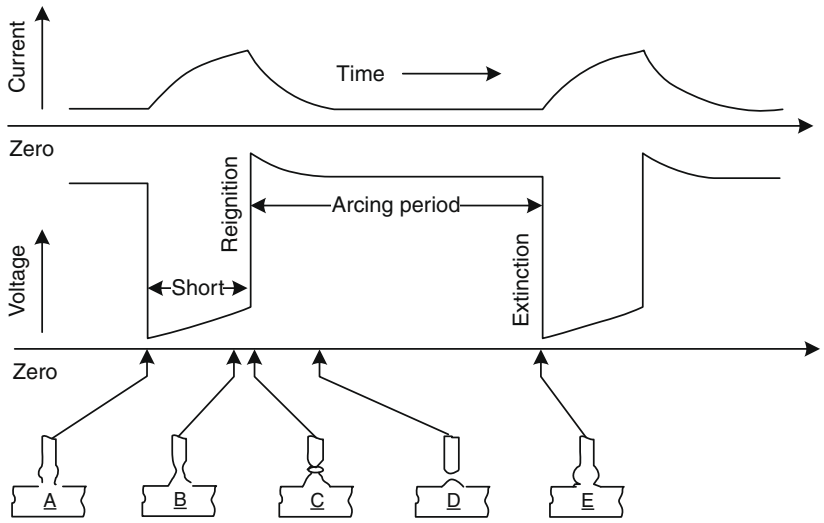


Fig. 1. Oscillograms and sketches of short-circuiting transfer.

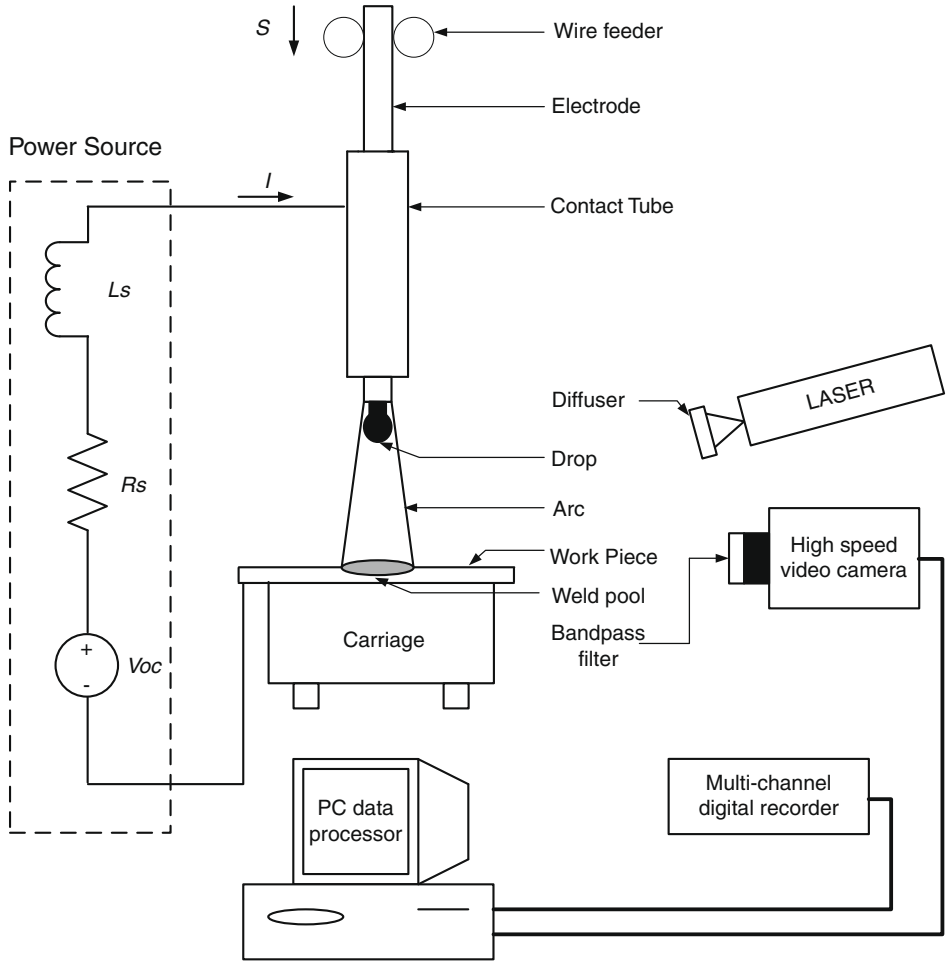


Fig. 2. Sketch of the experimental apparatus.

- Cond 1: The contact-tube to workpiece distance is less than the electrode extension plus droplet length.
- Cond 2: The molten metal bridge diameter is less than a threshold fixed by electrical and material laws.

In each state, a set of differential equations drives the process behaviour. These equations stem from power source characteristics, the set of forces acting on a droplet and fluid dynamics for the metal transfer.

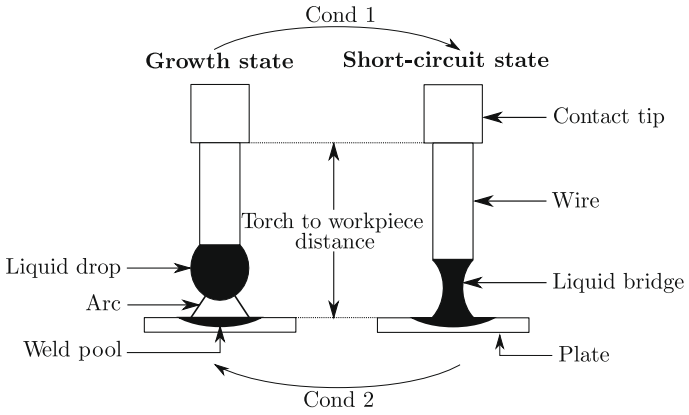


Fig. 3. States of the hybrid system.

2.2. Model of droplet growth

Two categories of forces can arise in our process: detaching ones and retaining ones. Three forces have been identified as relevant:

(1) The gravitational force due to the mass of the drop and acts as a detaching force when welding in the flat position:

$$F_g = \frac{4}{3}\pi r_d^3 \rho_w g, \tag{1}$$

where r_d denotes the droplet radius, ρ_w the mass density of molten metal, and g the gravity acceleration.

(2) The surface tension is a property of all liquids which comes from the surface of a liquid has an energy (more precisely the interface between the liquid and the phase surrounding it). This energy is reduced by minimizing the area for any volume. This is realized when the liquid takes a spherical shape as the ratio area/volume is minimal for a sphere. In 1864, Tate observed the direct proportionality between the maximum mass of a drop of water detached from a pipe and the radius of the pipe. This relation is known as Tate's law and is given by

$$F_S = 2\pi r_w \gamma \tag{2}$$

with F_S the weight of the drop, r_w the radius of the tip (in MIG/MAG welding, the radius of the electrode) and γ the liquid surface tension. In their study of 1919 Harkins and Brown [15] proposed a modification of (2) based on the observation that with each drop forming at the end of the tip, all of the liquid does not detach and some remains on the tip. Defining F'_S as the weight of liquid actually detached from the tip, (2) becomes

$$F'_S = 2\pi r_w \gamma f, \tag{3}$$

where f is an empirical correction function which will take the unity value for an ideal drop detaching totally from the pipe.

(3) The electromagnetic force which results from diverging or converging current flow within the electrode. As shown in Fig. 5, if the current lines diverge in the drop, the Lorentz force, which acts at right angle to these current lines, creates a detaching force; on the contrary, if the current lines converge, the Lorentz force opposes drop detachment. This force is calculated thanks to the Lorentz's law:

$$\vec{F}_{em} = \vec{J} \times \vec{B}, \tag{4}$$

where $\vec{J}$ is the current density and $\vec{B}$ the magnetic flux. The total electromagnetic force can be obtained by integrating (4) over the current conducting surface of the drop. By assuming that the current density on the drop is uniform, Amson [5] obtained:

$$F_{emz} = \frac{\mu_0 I^2}{4\pi} f_z \tag{5}$$

and

$$f_z = -\left[\frac{1}{4} - \ln\left(\frac{r_d \sin\theta}{r_w}\right) + \frac{1}{1-\cos\theta} - \frac{2}{(1-\cos\theta)^2}\ln\left(\frac{2}{1+\cos\theta}\right)\right], \tag{6}$$

where I is the welding current and μ_0 is the permeability of free space. The geometry used in (6) is shown in Fig. 4 where θ is the arc hanging angle. The black surface indicates the area of the drop allowing the current to go through

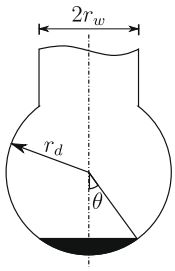


Fig. 4. Geometry of a drop.

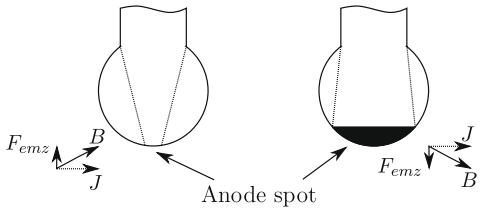


Fig. 5. Current path and Lorentz force.

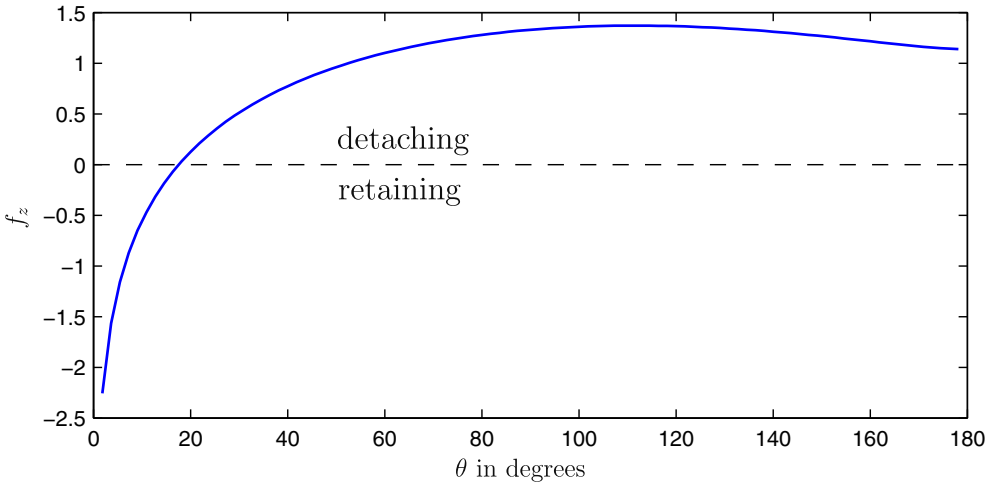


Fig. 6. Variation of f_z as a function of θ.

the plasma (this area is called the anode spot) . A graph of f_z as a function of the conduction zone angle is given in Fig. 6. As it can be seen in Fig. 5, when the conduction zone is small such that the current lines converge (the current lines are included between the dotted lines), f_z becomes negative so the electromagnetic force acts as a repulsive force. When the conduction zone is large enough so that the current lines diverge, f_z becomes positive and the electromagnetic force becomes a detaching force.

In MIG/MAG welding metal, transfer only occurs after the solid is molten and thus the coupling between mass flow and heat transfer is very strong. The main heat sources for melting the electrode are Joule effect and electron condensation [16]. Joule heat is the heat generated by the electrical resistance of the consumable electrode. The heat of electrode condensation results from the plasma high energy electrons condensing on the metal surface and releasing most of their energy.

The Joule heat can be calculated by

$$Q_{joule} = \int_V \rho(T) I^2 dV = \bar{\rho} A L I^2, \tag{7}$$

where $\bar{\rho}$ is the average electrical resistivity, L the electrode extension and A the area of the electrode. The electrode condensation heat can be obtained by

$$Q_{cond} = \left(\frac{3}{2}\frac{kT}{e} + V_a + \Phi\right) I \tag{8}$$

with k denotes the Boltzmann constant. The first term of the right hand-side of (8) represents the kinetic energy of the electrons in the plasma, the second term is the acceleration energy of the electrons in the anode drop region and the third term is the work function of the electrode material.

If one assumes that the whole heat is consumed in melting the consumable electrode, the melting rate can be obtained by dividing the total heat input into the system by the heat required to melt a unit mass of the material:

$$\frac{dm}{dt} = \frac{\overline{\rho} A L I^2 + \left(\frac{3}{2}\frac{kT}{e} + V_a + \Phi\right) I}{\int_{T_i}^{T_m} C_p dT + \Delta H_{trans} + \Delta H_m} \tag{9}$$

where ΔH_{tans} is the heat of crystalline transition and ΔH_m is the heat of fusion. Since the coefficients are constant the melting rate can be expressed as a second order function of the welding current:

$$\frac{dm}{dt} = (C_1 I + C_2 \rho l_s I^2)\rho_w, \tag{10}$$

where C_1 and C_2 are constant parameters fixed to strike a balance between Joule effect and arc heat fusions, $\rho = \bar{\rho} A$ denotes the linear resistivity of the electrode and l_s represents the solid wire extension (stick-out).

The power supply being converted into the equivalent *RL* circuit [12], we write the Kirchoff's voltage law for the welding system represented in Fig. 2 as:

$$L_S \frac{dI}{dt} + (R_S + R)I + U_{arc} = U_{oc} \tag{11}$$

with L_S and R_S are the inductance and resistance of the welding system, R is the wire-droplet system resistance, U_{oc} is the equivalent open-circuit voltage and U_{arc} is the arc voltage. The latter is expressed by

$$U_{arc} = U_0 + R_a I + E_a(CT - l_s), \tag{12}$$

where U_0 is the arc voltage constant, E_a the arc length factor, CT the contact-tube to workpiece distance and R_a the arc resistance.

The stick-out evolution is controlled by

$$\frac{dl_s}{dt} = S - \frac{1}{\pi r_w^2}\frac{dm}{dt} \tag{13}$$

with S the wire feed speed.

We have used these expressions of forces, of melting rate and of evolution of the welding current to write a state-space representation of the governing differential equations of droplet growth. Five state variables have been retained: droplet displacement x_1, droplet velocity x_2, current x_3, stick-out x_4 and droplet mass x_5. State equations are:

$$\dot{x}_1 = x_2, \tag{14}$$

$$\dot{x}_2 = \frac{1}{x_5}(F_g + F_{em} - F_S), \tag{15}$$

$$\dot{x}_3 = \frac{1}{L_S}\left[U_{oc} - (R_a + R_S)x_3 - U_0 - E_a(CT - x_4) - \left[x_4 + x_1 + \left(\frac{3x_5}{4\pi\rho_w}\right)^{\frac{1}{3}}\right]\rho x_3\right], \tag{16}$$

$$\dot{x}_4 = S - \frac{1}{\pi r_w^2}(C_1 x_3 + C_2 \rho x_3^2 x_4), \tag{17}$$

$$\dot{x}_5 = (C_1 x_3 + C_2 \rho x_3^2 x_4)\rho_w. \tag{18}$$

In order to implement the model of the system shown in Fig. 3, offering a good agreement with the practical work of the process, it is necessary to supervise some variables of interest during a normal operation of the testing bed. Indeed, we need the dynamic behaviour of the volume of the drop, the center of gravity and the liquid–solid border synchronously with arc voltage and welding current. So an experimental database has been built with electrical measurements and high speed videos. The work presented in Section 3 aims at extracting information concerning the droplet from videos. In this purpose we decided to use active contours.

2.3. Model of metal transfer

We present here a simple model of metal transfer during short-circuiting period. Through this work we aim at checking some ideas often encountered when dealing with welding.

The power supply being converted into the equivalent *RL* circuit [12], we write the Kirchoff's voltage law for the welding system represented in Fig. 2 as:

$$L_S \frac{dI}{dt} + (R_S + R)I = U_{oc}. \tag{19}$$

The metal transfer is controlled by three relevant forces:

(1) The gravitational force.
(2) The electromagnetic force which results from diverging or converging current flow within the electrode. When the current path diverges in the drop, the Lorentz force presents an axial component acting, normal to the current path and creating a detaching force, and an orthoradial component which squeezes the molten metal bridge.
(3) The surface tension which is normal to the surface of a molten droplet. It serves to support the form of the molten metal bridge.

An exhaustive welding simulation would solve a set of coupled partial differential equations governing the fluid, thermal and electromagnetic fields. These equations would include the continuity and Navier–Stokes equations. The following assumptions [11] are made to simplify the complex behavior of short-circuiting transfer:

(1) The initial bridge shape is spherical and the contact diameter within the weld-pool surface is equal to the wire diameter.
(2) Flow velocity within the bridge and pressure within the weld pool are neglected.
(3) The bridge shape is described by two principal radii.
(4) The pool surface remains flat and metal transfer is stable.

Thanks to these assumptions and taking into account only the three forces mentioned above we expect to model the sequence of events represented in Fig. 7.

The average pressure on the cross section of the bridge center with a configuration of principal radii as in Fig. 8 is [11]:

$$P_{avg} = \frac{\mu_0 I^2}{8\pi^2 R_1^2} + \gamma\left(\frac{1}{R_1} + \frac{1}{R_2}\right). \tag{20}$$

Applying the Bernoulli equation with assumption 2, the flow velocity at the contact between the bridge and pool surface can be calculated using the average pressure and bridge height as:

$$v = \sqrt{\frac{2}{\rho_w}(P_{avg} + \rho_w g h)}, \tag{21}$$

where h is the distance between the pool surface and bridge center. Finally, the state-space model corresponding to the metal transfer mode is given by

$$\dot{x}_1 = x_2, \tag{22}$$

$$\dot{x}_2 = \frac{1}{x_5}(F_g + F_{em} - F_S), \tag{23}$$

$$\dot{x}_3 = \frac{1}{L_S}\left[U_{oc} - R_S x_3 - \left[x_4 + x_1 + \left(\frac{3x_5}{4\pi\rho_w}\right)^{\frac{1}{3}}\right]\rho x_3\right], \tag{24}$$

$$\dot{x}_4 = S - \frac{C_2 \rho x_3^2 x_4}{\pi r_w^2}, \tag{25}$$

$$\dot{x}_5 = \left(C_2 \rho x_3^2 x_4 - \sqrt{\frac{2}{\rho_w}(P_{avg} + \rho_w g h)A_c}\right)\rho_w, \tag{26}$$

where A_c is the exchange surface between the liquid metal bridge and the weld pool.

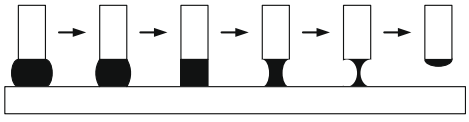


Fig. 7. Short-circuit modeling.

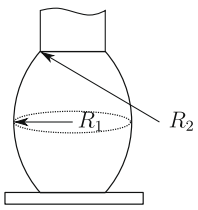


Fig. 8. Bridge configuration with principal radii.

2.4. Implementation of the hybrid model

The hybrid model for MIG/MAG welding has been implemented in the Matlab language. The simulation of the continuous state is performed by the Runge–Kutta solver and the main difficulty was to handle with the switching between the two continuous states. The values of the state variables obtained at the end of one continuous state are used as initial conditions for the other continuous state.

The switching from the droplet growth state to the metal transfer state is achieved when the distance between the drop and the weld pool equals zero. This consists in checking whether the sum of the stick-out l_s (state variable x_4) and the drop height, obtained thanks to its mass and geometry (truncated sphere, see Fig. 4), equals the contact tip to workpiece distance CT.

The switching from the metal transfer state to the droplet growth state is controlled by the molten metal bridge radius (R_1). It is obtained from the volume of the bridge using once again its geometry (see Fig. 8).

3. Active contour segmentation

In order to implement the model of the system shown in Fig. 3, offering a good agreement with the practical work of the process, it is necessary to supervise some variables of interest during a normal operation of the testing bed. Indeed, we need the dynamic behaviour of the volume of the bridge, the minimal diameter of the bridge and the surface of contact between the bridge and the pool synchronously with arc voltage and welding current. So an experimental database has been built with electrical measurements and high speed videos. The work presented in Section 3 aims at extracting information concerning the drop from videos. In this purpose we decided to use active contours.

Contour extraction is often used in technical and scientific fields. It can be employed for shape recognition in industry applications or to process images in astronomy to make them clearer. Our study requires the use of such techniques in order to observe the dynamical behaviour of molten metal at the end of the electrode.

Standard segmentation methods can be seen as low-level methods. They offer the detection of elements in the picture. They create a division of the image in separated homogeneous regions by considering contours as places of significant variation of information. Nevertheless they do not execute the last step we need. Indeed the metrology is not allowed by these methods and we are looking for quantitative information about the drop geometry during the arc period and about the liquid bridge shape during the short-circuit period. That is why we have chosen active contours as a high-level method performing this last step.

3.1. Theory

Active contour models introduced by Kass and Witkins [14] have rapidly seduced the scientific community by their ability to efficiently combine both the available *a priori* knowledge about the structure of interest and local correspondence with the image features [17–19]. Their principle is to evolve from an initial contour to equilibrium which corresponds to the edges of the object to detect.

Active contour models are also known as *snakes* or *minimizing curves*. This is a semi-interactive method in which the user puts an initial contour in the image (near the shape to find) and this one will be deformed by the action of several energies. The contour will evolve searching the position of lowest local energy.

In its continuous form, the snake is described by a parametric curve $v(s) = [x(s), y(s)]^t$ with $s \in [0, 1]$ and $v(0) = v(1)$ for a closed contour. Its total energy can be written as:

$$E_{snake} = \int_0^1 [E_{int}(v(s)) + E_{im}(v(s)) + E_{ext}(v(s))]ds, \tag{27}$$

where E_{int}, E_{im} and E_{ext} denote internal, image and external energies, respectively. The internal energy depends only on the contour shape. It is given by

$$E_{int}(v(s)) = \alpha E_{cont}(v(s)) + \beta E_{curv}(v(s)), \tag{28}$$

$$= \alpha\left(\frac{dv}{ds}\right)^2 + \beta\left(\frac{d^2v}{d^2s}\right)^2. \tag{29}$$

The first order term drives continuity while the second order term controls curvature.

The potential energy related to the image can be written as:

$$E_{im}(v(s)) = -\lambda(s)|\nabla I(v(s))|^2. \tag{30}$$

This energy qualifies the elements towards which we want to attract the contour on the image. In the segmentation framework it corresponds to the lines of high gradient.

The external energy is user defined in accordance with the problem specificities. For instance we can impose a minimal distance between two consecutive points of the snake. Without an external energy and when high gradient lines are sought we can write:

$$E_{snake}(v(s)) = \alpha E_{cont}(v(s)) + \beta E_{curv}(v(s)) + \sigma E_{im}(v(s)). \tag{31}$$

The triplet (α, β, σ) enables to strike a balance between different energies. In the continuous domain, the energy equation of a contour C can be expressed by

$$E_{snake} = \int_C \left[-\sigma(s)|\nabla I(v(s))|^2 + \alpha\left(\frac{dv}{ds}\right)^2 + \beta\left(\frac{d^2v}{d^2s}\right)^2 \right] ds. \tag{32}$$

In a discrete approach, the contour is a list of points M_i, $i \in [1, n]$. The energy of the snake is assimilated to the sum of the energies associated to the n points defining it:

$$E_{snake} = \sum_{i=1}^{n} [\alpha E_{cont}(M_i) + \beta E_{curv}(M_i) + \sigma E_{im}(M_i)]. \tag{33}$$

3.2. Algorithm

In order to minimize E_{snake}, we determine the list of n points constituting C. To achieve this goal we have chosen to implement the *greedy* algorithm [20–22]. In this approach the derivatives in (32) have to be approximated by finite differences. If $v_i = [x_i, y_i]^t$ is a point of the contour, we write:

$$\left|\frac{dv_i}{ds}\right|^2 \approx |v_i - v_{i-1}|^2 = (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 \tag{34}$$

and

$$\left|\frac{d^2v_i}{ds^2}\right|^2 \approx |v_{i-1} - 2v_i + v_{i+1}|^2 = (x_{i-1} - 2x_i + x_{i+1})^2 + (y_{i-1} - 2y_i + y_{i+1})^2. \tag{35}$$

Two assumptions have been made here:

- The points are spaced at unit intervals. If the points are evenly spaced, then (34) should be divided by d^2, where d is the distance between points and (35) by d^4.
- If the points are not evenly spaced, the first derivative term will be divided by d_i^2 where d_i is the distance between points i and $i-1$.

The third term in (31), E_{im}, is the image force, which is the gradient magnitude at each point in the image, intensity being coded as an eight bit integer.

The *greedy* algorithm is iterative. During each iteration, a neighborhood of each point is examined and the point in the neighborhood giving the smallest value for the energy term is chosen as the new location of the point.

Our final aim is to get information on the drop evolution. To achieve it, we have acquired a set of movies which are lists of images. Consequently, the idea is to use the detected contour at image $(k-1)$ to initialize the one at image k.

This method has been employed with success during the arcing period [23]. Nevertheless it does not work during the short-circuiting period sketched in Fig. 9. Indeed there is no line of high gradient between the molten metal bridge and the weld pool. As a consequence some points of the snake can either go in the bridge or be lost in the pool. A solution is to force these points on an estimated edge.

In this purpose we defined three zones of interest: two for the pool and one for the bridge. The chain of ideas is:

(1) to follow the weld-pool surface and approximate it by a polynomial,
(2) to project the points of the initial bridge contour on the polynomial curve if $y_i \geqslant y_{pool}(x_i) + \epsilon$ where $y_{pool}(x)$ is the equation of the weld-pool surface and ϵ is a safety margin ($\epsilon = 2$ pixels),
(3) to launch the *greedy* algorithm on the bridge with the projected points forced in a tube around the estimated pool surface.

4. Experimental validation

In this section, our main concern is the experimental validation of the proposed model. This will be made in two steps: firstly, we focus on the fitting of the simulated current and voltage to the measurements; secondly, we aim at checking the ability of the model to give accurate prediction of variables which are not easily accessible. Here, we focus on the droplet volume (or equivalently its surface assuming a revolution symmetry).

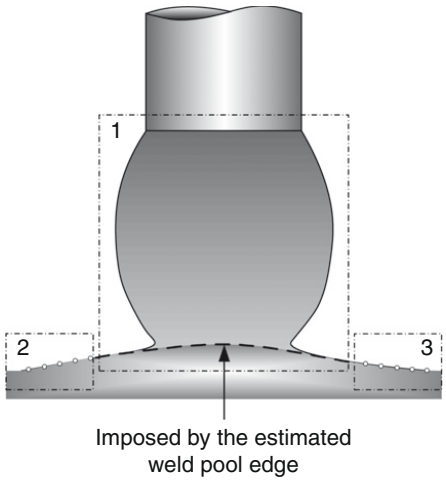


Fig. 9. Short-circuit snake principle.

Table 1
List of parameters used for simulation.

Constant	Value
System resistance, R_S	4 mΩ
System inductance, L_S	0.14 mH
Wire radius, r_w	0.5 mm
Linear resistivity, ρ	8.64×10^{-14} Ω/m
Mass density, ρ_w	7800 kg/m^3
Wire feed speed, S	2 or 3 m/min
Contact tip to workpiece distance, CT	15 mm
Permeability, μ_0	1.25×10^{-6} H/m
Surface tension, γ	1.2 N/m
U_0	16 V
R_a	0.036 Ω
E_a	1500 V/m
Arc hanging angle, θ	50°
Constant for arc heating, C_1	2.96×10^{-14} m^3/As
Constant for Joule heating, C_2	0.0537 m^3/VAs

4.1. Simulator

4.1.1. Preliminary results

The state space representation written from equations presented in Section 2 is used in a simulator implemented in the Matlab language. The physical parameters of the welding process are given in Table 1. The dynamics of the measured and simulated welding currents and voltages, for $U_{oc} = 22$ V and a wire feed rate $S = 2$ m/min, are shown in Fig. 10. The curves correspond to a whole period with the creation of a droplet at the end of an electrode followed by the transfer of the metal to the weld pool.

During the arcing period, molten metal is added to the drop at each time step because of the physical law driving the melting rate in (10). We notice in Fig. 10b that, in the arc period, the calculated variables present an order of magnitude close to experimental ones in Fig. 10a. Consequently we can assume that the main influences represented by the three forces acting on the drop are sufficient for modeling this period. Nevertheless the short-circuit period seems too short when compared with the experiments. The analysis of the state equations (22)–(26) indicates that the short-circuit duration is mainly influenced by the current dynamic and the metal flow velocity (21). Recalling that, in fact, the current is controlled during the short-circuit time, so (19) can be modified (through L_S for example) to adjust the simulated current dynamic to experimental data. Also, the derivation of (21) relies on simplifying assumptions which can be questioned. So we propose to modify this equation by introducing a multiplicative factor allowing the simulated current and voltage to be fitted to the experimental ones. The modified expression for the flow velocity is:

$$v = \kappa \sqrt{\frac{2}{\rho_w}(P_{avg} + \rho_w g h)}, \tag{36}$$

where κ is an adjustment term. We found experimentally that the values $L_S = 0.7$ mH and $\kappa = 0.25$ yield the results shown in Fig. 10c. Clearly, the short-circuit duration is now longer and the current dynamic is smaller; they are in better agreement with those observed experimentally in Fig. 10a.

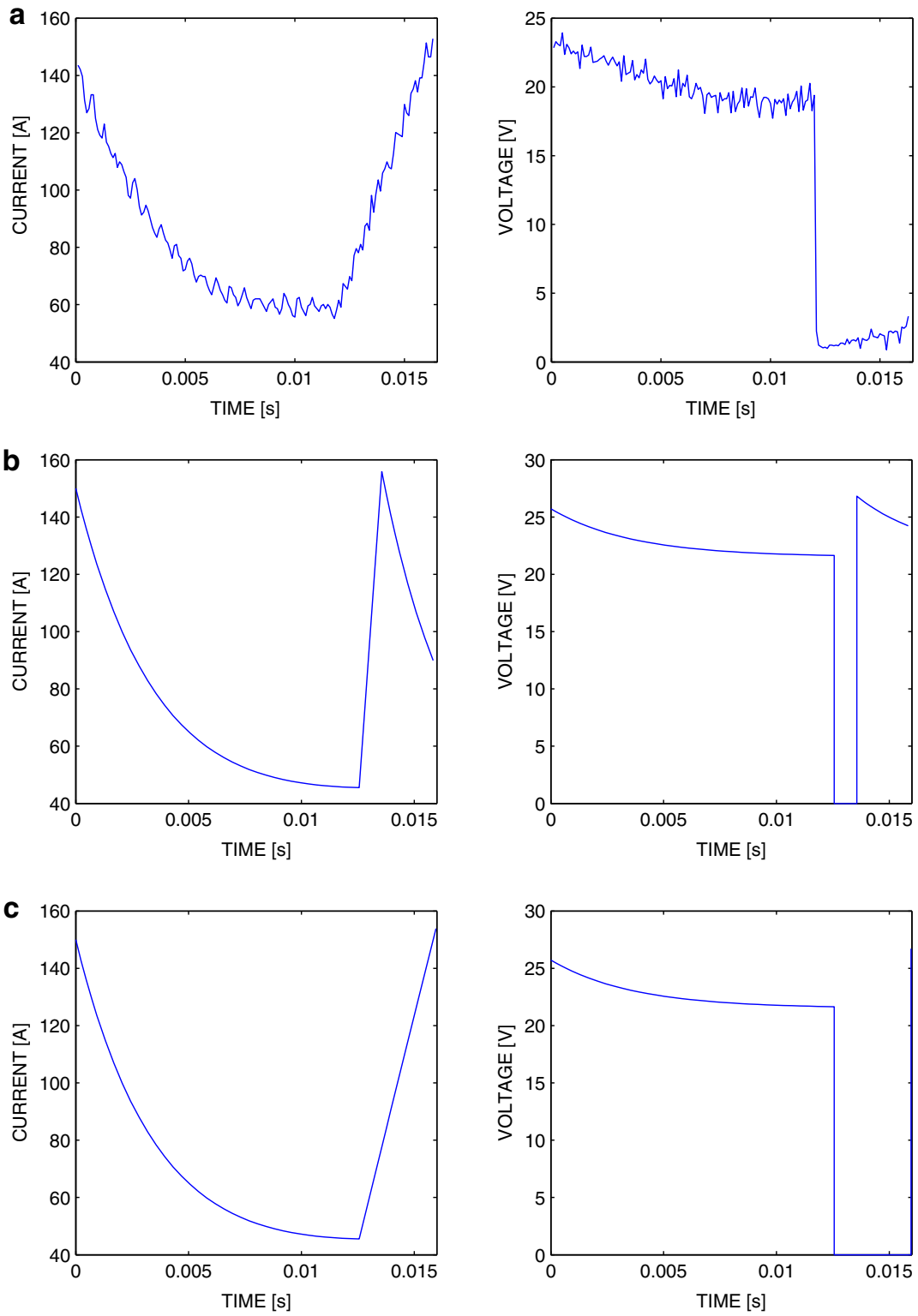


Fig. 10. Comparison between (a) measured current and voltage; (b) simulated current and voltage; (c) simulated current and voltage with the modified flow expression.

4.1.2. Adjustment to experiments

Weld-pool oscillation is a dominant feature of short-circuiting welding and a major factor in regards to the welding process stability [24]. The weld pool when excited into motion will exhibit natural frequencies of oscillation [25]. This excitation results from the transfer of the droplet's momentum at the molten bridge rupture [26] and the generation of arc pressure following the ignition of a new arc. According to the authors of [27] the maximal process stability is achieved when the

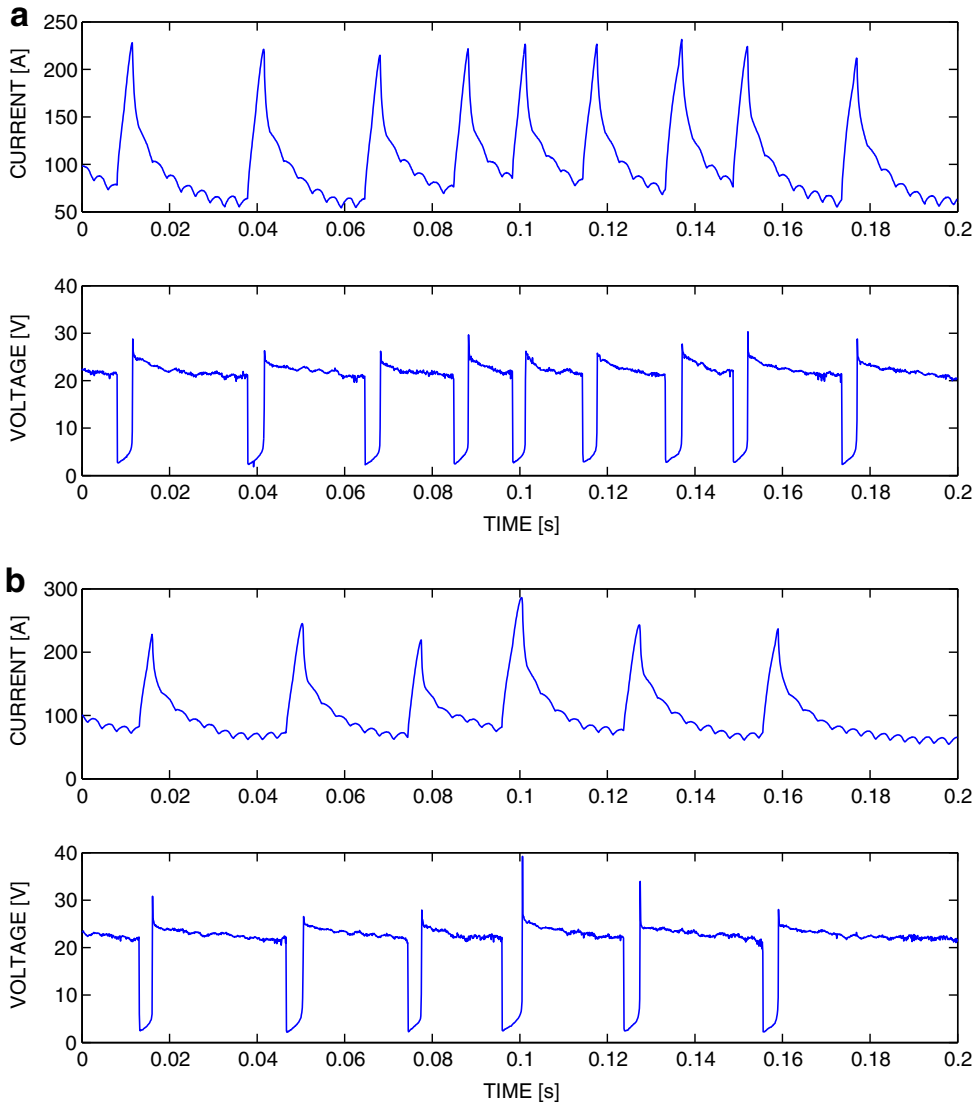


Fig. 11. Oscillogramms of current and voltage for (a) $U_{oc} = 22$ V, and (b) $U_{oc} = 23$ V.

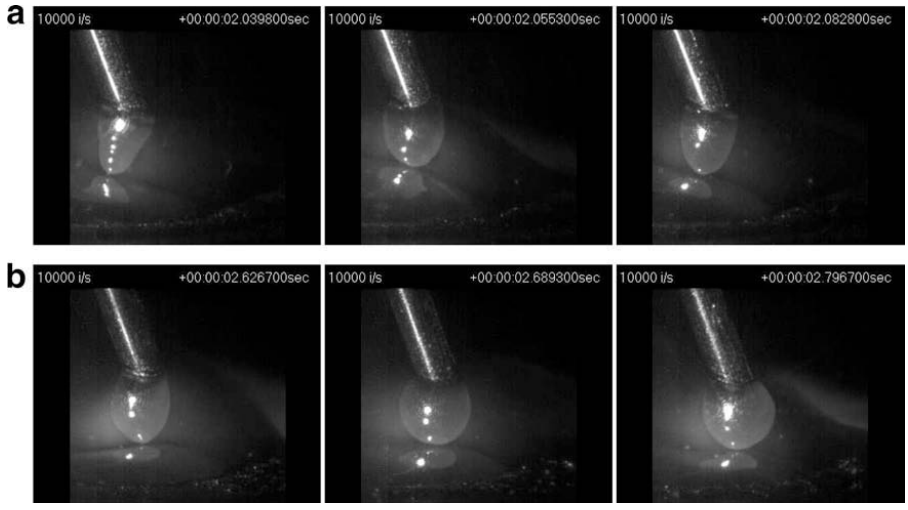


Fig. 12. Images before short-circuit with $S = 3$ m/min and (a) $U_{oc} = 22$ V; (b) $U_{oc} = 23$ V.

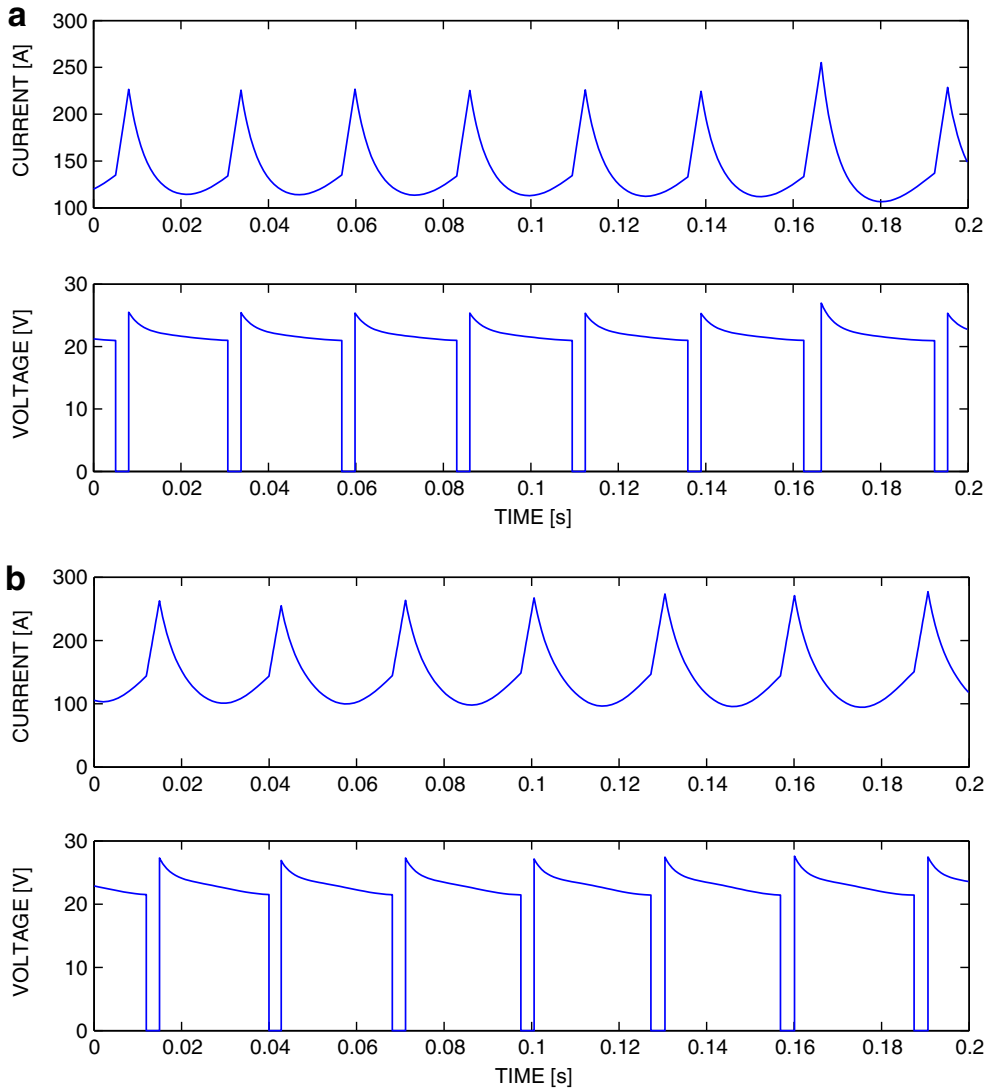


Fig. 13. Simulated current and voltage for (a) $U_{oc} = 22$ V, and (b) $U_{oc} = 23$ V.

short-circuiting frequency and the oscillation frequency of the weld pool are synchronised such that the short-circuiting frequency equals the oscillation frequency of the pool. Consequently we have chosen to add a motion on the weld-pool surface to give a more precise simulation of MIG/MAG welding. The height of molten metal in the pool is controlled through a function of the type $A\sin(2\pi ft)$. Here we use $A = 1$ mm and $f = 25$ Hz.

4.1.3. Major tendencies

In this paragraph we want to explain some behaviours of our industrial process thanks to the theoretical equations written previously. We have fixed the wire feed speed to 3 m/min, the shielding atmosphere and the metal. The changed variable is the open-circuit voltage of the power source. The experimental electrical recordings show that increasing U_{oc} has a great consequence on the length of the arcing period (T_{arc}) and a smaller impact on the length of the short-circuit period (T_{sc}).

In Fig. 11, we have plotted the oscillograms of measured welding voltage and current for $U_{oc} = 22$ V and $U_{oc} = 23$ V. On these short sequences the lowest U_{oc} allows 9 whole periods and the highest one allows 6 periods. By practicing an analysis on a longer sequence, we estimated the averages of T_{arc} and T_{sc} as:

- with $U_{oc} = 22$ V, $\overline{T}_{arc} = 13.75$ ms and $\overline{T}_{sc} = 2.81$ ms,
- with $U_{oc} = 23$ V, $\overline{T}_{arc} = 21.75$ ms and $\overline{T}_{sc} = 3.06$ ms.

Practically speaking, U_{oc} is linked to the so-called *arc length* L_{arc}: when you increase U_{oc} you increase L_{arc}. Arc voltage and arc length are terms that are often used interchangeably. It should be pointed out, however, that they are different even though they are related. Arc length is a critical variable that must be carefully controlled. Should the arc be too long, it tends to wander, affecting both the penetration and surface bead profiles. A long arc can also disrupt the gas shield. Arc length is

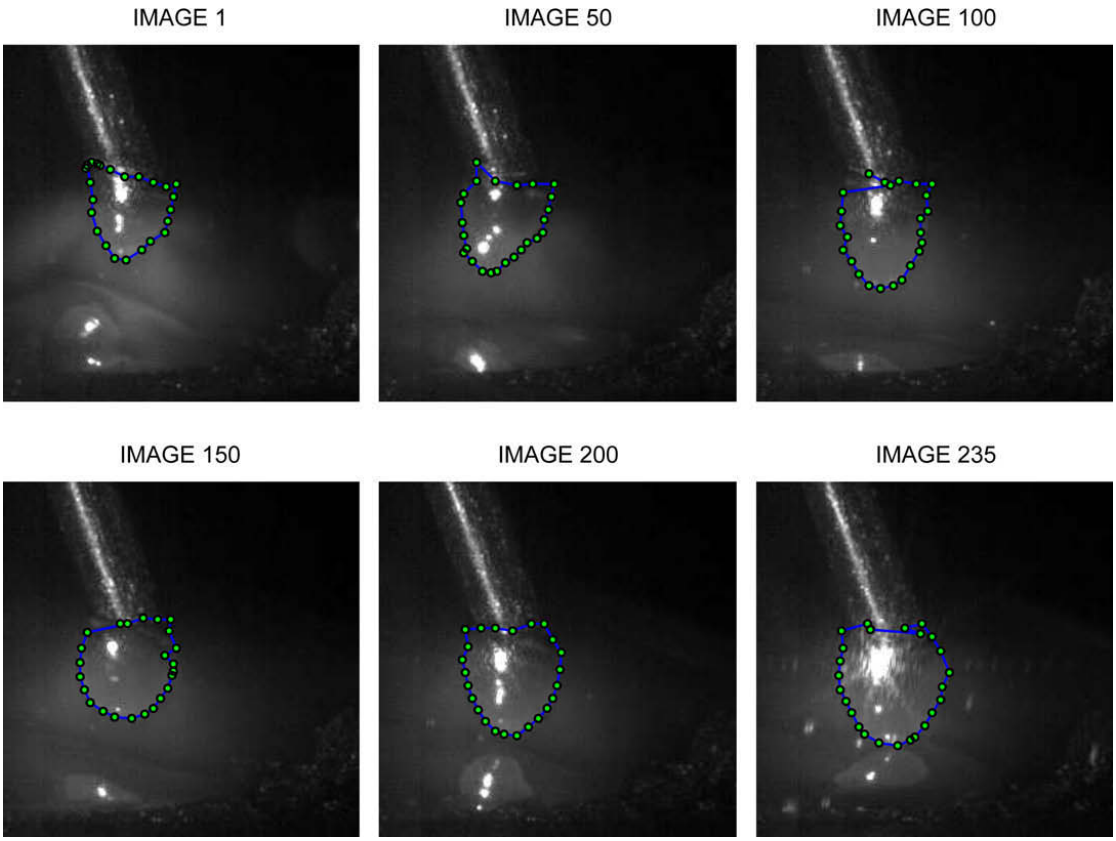


Fig. 14. Converging active contour on a set of experimental images in the arc period.

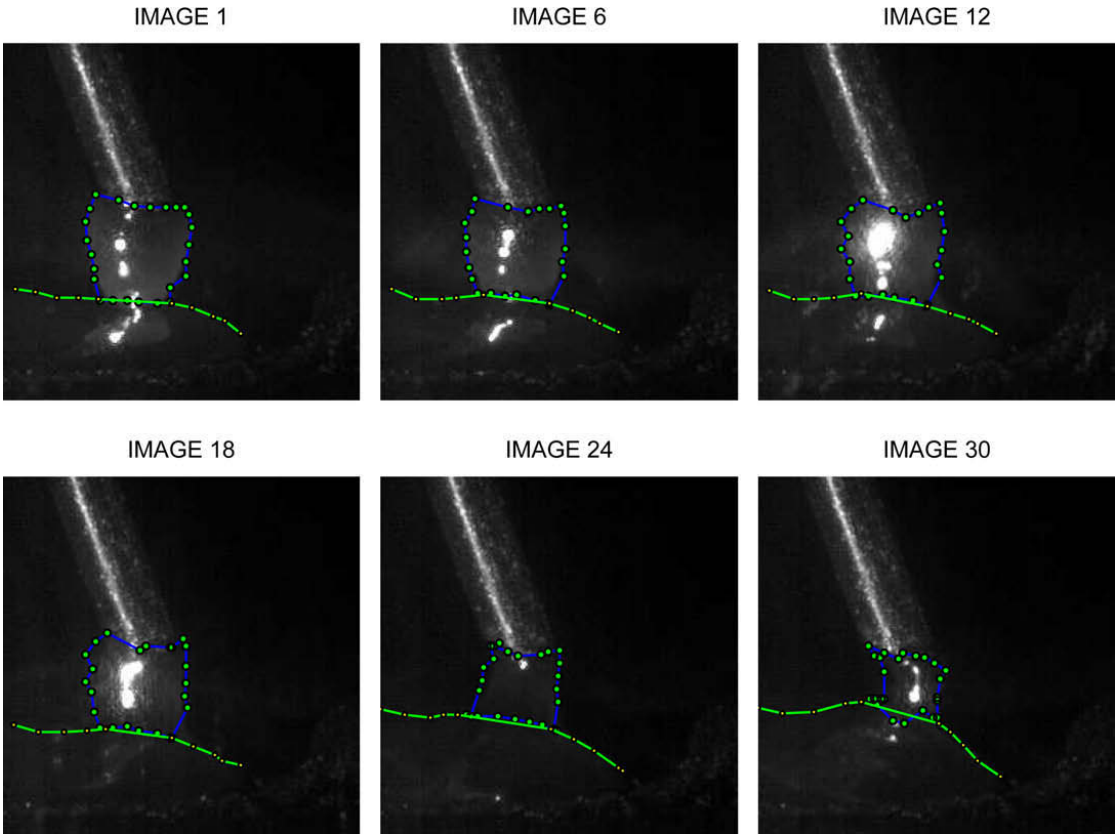


Fig. 15. Converging active contour on a set of experimental images in the short-circuit period.

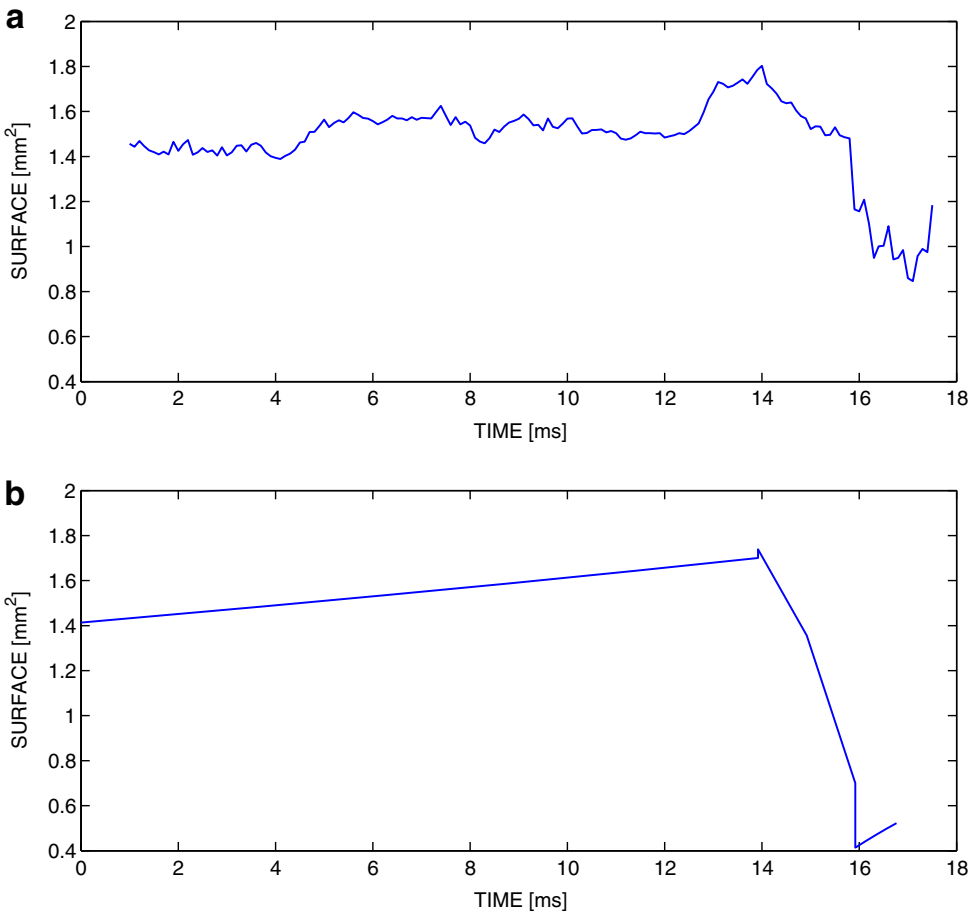


Fig. 16. Droplet surfaces for $U_{oc} = 22$ V. (a) Measured surface and (b) simulated surface.

the independent variable. Arc voltage depends on the arc length as well as many other variables, such as the electrode dimensions and composition and the shield gas for instance. Arc voltage is an approximate means of stating the physical arc length in electrical terms, although the arc voltage also includes the voltage drop in the electrode extension beyond the contact-tube. Welds made with arc voltages that are too low show a depression in the centre of the reinforcement. Arc voltages that are too high create heavy spatter conditions.

Moreover the energy of the arc is higher so the pressure applied on the weld-pool surface is stronger. As a consequence the distance between the drop and the weld pool is subjected to an increment when U_{oc} passes from 22 V to 23 V. All these assertions can be checked in Fig. 12a and b. A bigger drop has to be created at the end of the electrode. If the distance is increased with a welding current of the same order of magnitude, which means a corresponding melting rate, the time to form this bigger drop is longer.

We can test our simulator and see if this tendency of major significance for an industrial application is well represented. We have used the same parameters and we show in Fig. 13 the oscillograms of simulated welding voltage and current for $U_{oc} = 22$ V and $U_{oc} = 23$ V. On these short sequences the lowest U_{oc} allows 9 whole periods and the highest one allows 7 periods. An analysis made on a longer sequence gives:

- with $U_{oc} = 22$ V, $\overline{T}_{arc} = 22.18$ ms and $\overline{T}_{sc} = 3$ ms,
- with $U_{oc} = 23$ V, $\overline{T}_{arc} = 25$ ms and $\overline{T}_{sc} = 3.05$ ms.

The mean of T_{arc} is increased too by the increment on U_{oc}. This means that our simulator is capable of giving access to information about the frequency of the metal transfer phenomenon. This can be of high interest if such an information can be linked to notions of stability or efficiency.

4.2. Active contours

4.2.1. Preliminary results

Our experimental data are produced thanks to the method exposed in Section 3. Figs. 14 and 15 are taken from a movie acquired at 10,000 frames/s. We have used a triplet [1,1,1] for the active contour in charge of the weld-pool surface and a

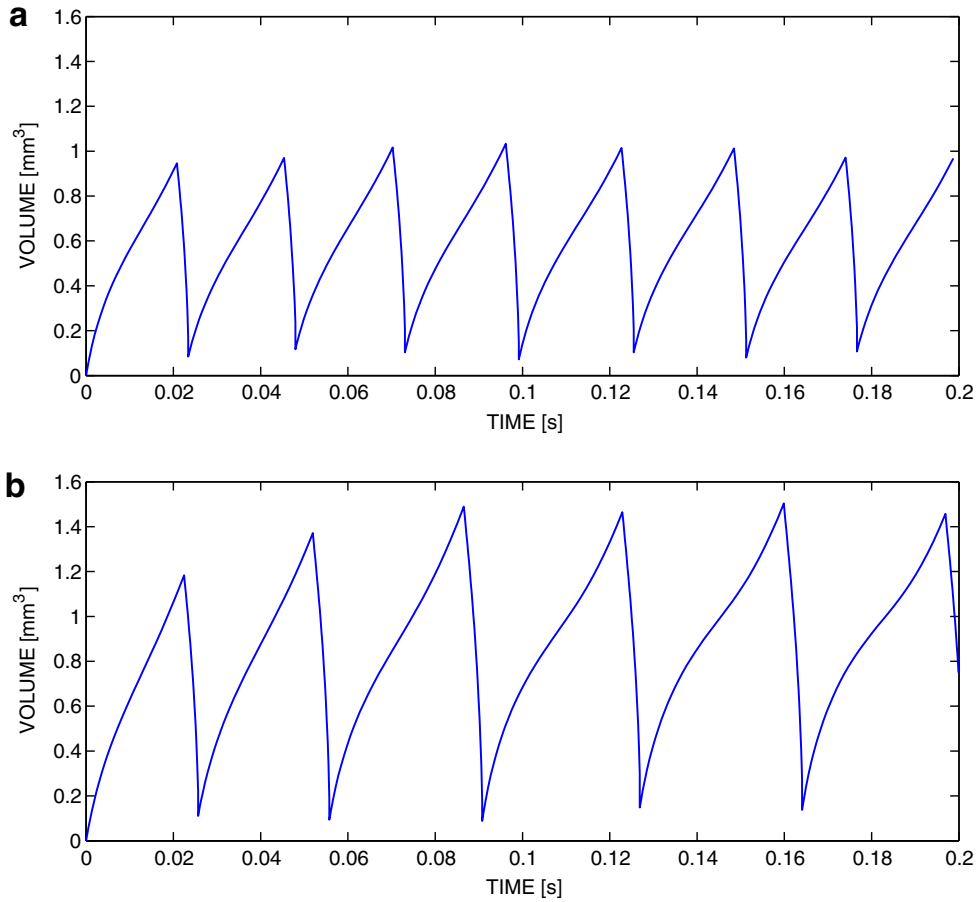


Fig. 17. Simulated volume for (a) $U_{oc} = 22$, and (b) $U_{oc} = 23$ V.

triplet [0.6, 0.7, 1] for the one in charge of the molten bridge. The weight given to the gradient magnitude is increased here because of the lack of luminosity due to the use of an extension tube. The edges are well detected so we will be able to extract the variables of interest for our study, especially the volume of metal included in the drop during the arc period and then the amount transferred to the weld pool during the short-circuit period.

4.2.2. Major tendencies

In Fig. 16 we have plotted the measured and simulated surfaces of the droplet during the arc period and those of the molten metal bridge during the short-circuit period with an open-circuit voltage $U_{oc} = 22$ V. The measured surface is obtained from the estimated contour of the droplet and then transformed from pixels to millimeters, while the simulated one is computed directly from the geometry of the droplet. We can observe that the surface of the drop at the end of the arc period increases with the open-circuit voltage. This tendency can be seen in Fig. 17. Indeed, when U_{oc} passes from 22 V to 23 V, the volume of the droplet at the end of the electrode passes from 0.95 mm^3 to 1.45 mm^3. This means that our explanation for the phenomenon observed and simulated is correct: a bigger drop has to be created to allow the physical contact with the weld pool. This has to be used in accordance with the practical knowledge of the welders. The weld diameter at the interface and the reinforcement are the two characteristics of a GMAW weld which determine whether the weld will satisfy the intended service. Two major process variables (weld current and arc voltage) affect one or both of these characteristics. An increase in welding current will result in the following: an increase in the depth and width of the weld penetration; an increase in the deposition rate; an increase in the size of the weld bead. From any specific value of arc voltage, a voltage increase tends to flatten the weld bead and increase the width of the fusion zone. Excessively high voltage may cause porosity, spatter and undercut. Reduction in voltage results in a narrower weld bead with a higher crown and deeper penetration. Excessively low voltage may cause stubbing of the electrode.

5. Conclusion

In this paper, we presented the design and testing of a physical model for short arc MIG/MAG welding. The proposed hybrid model have two distinct continuous states: the arc state during which a drop forms at the end of the electrode, and a short-circuit state when the liquid drop transfers to the weld pool. Due to the complex physical interactions involved in the

welding process, simplifications have been made to get a model accounting for the main physical contributions. In addition, a discrete contour approach have been developed to follow the dynamical behaviour of the molten drops in image sequences acquired during the welding process. We have shown that model predictions are consistent with experimental data in both arc and short-circuit states. The work will be extended in order to test some control strategies in simulation, and also to help developing fault detection methods in MIG/MAG welding.

References

[1] S. Adolfsson, A. Bahrami, G. Bolmsj, I. Claesson, On-line quality monitoring in short-circuit gas metal arc welding, Weld. Res. Suppl. 78 (2) (1999) 59–73.
[2] Y.-S. Kim, Metal Transfer in Gas Metal Arc Welding, Ph.D. Thesis, Massachusetts Institute of Technology, June 1989.
[3] L. Jones, T. Eagar, J. Lang, A dynamic model of drops detaching from a gas metal arc welding electrode, Appl. Phys. 31 (1998) 107–123.
[4] L. Jones, T. Eagar, J. Lang, Magnetic forces acting on molten drops in gas metal arc welding, Appl. Phys. 31 (1998) 93–106.
[5] J.C. Amson, Lorentz force in the molten tip of an arc electrode, Brit. J. Appl. Phys. 16 (1965) 1169–1179.
[6] Y.-S. Kim, D. McElliot, T. Eagar, Analyses of electrode heat transfer in gas metal arc welding, Weld. Res. Suppl. (1991) 20–31.
[7] S. Choi, C. Yoo, Y.-S. Kim, Dynamic simulation of metal transfer in GMAW. Part I: Globular and spray transfer modes, Weld. Res. Suppl. (1998) 38–44.
[8] S. Choi, C. Yoo, Y.-S. Kim, Dynamic simulation of metal transfer in GMAW. Part II: Short-circuit transfer mode, Weld. Res. Suppl. (1998) 45–51.
[9] F. Zhu, H. Tsai, S. Marin, P. Wang, A comprehensive model on the transport phenomena during gas metal arc welding process, Prog. Comput. Fluid Dynam. 4 (2) (2004) 99–117.
[10] T. Quinn, R. Madigan, T. Siewert, An electrode extension model for gas metal arc welding, Weld. J. 73 (10) (1994) 241–248.
[11] J. Choi, J. Lee, C. Yoo, Simulation of dynamic behavior in a GMAW system, Weld. Res. Suppl. (2001) 239–245.
[12] H. Terasaki, S. Simpson, Circuit simulation for gas metal arc welding system, in: 47th IEEE International Midwest Symposium on Circuits and Systems, 2004, pp. 387–390.
[13] J.-P. Planckaert, E.-H. Djermoune, D. Brie, F. Briand, F.-P. Richard, Modélisation du soudage MIG/MAG en mode short-arc, Tech. rep., Centre de Recherche en Automatique (CRAN)/Centre Technique des Applications du Soudage (CTAS), 35 pages, June 2005.
[14] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, Int. J. Comput. Vis. 55 (1988) 321–331.
[15] W. Harkins, R. Brown, The determination of surface tension (free surface energy), and the weight of falling drops: the surface tension of water and benzene by the capillary height method, J. Am. Chem. Soc. 41 (1919) 499–524.
[16] A. Lesnewich, Control of melting rate and metal transfer in gas-shielded metal-arc welding. Part I: Control of electrode melting rate, Weld. J. 37 (1958) 343s.
[17] H. Delingette, J. Montagnat, Topology and shape constraints on parametric active contours, Report 3880, INRIA, Sophia Antipolis, January 2000.
[18] L.D. Cohen, On active contour models and balloons, Comput Vis. Graph. Image Process.: Image Understand. 53 (2) (1991) 211–218.
[19] H.T. Nguyen, M. Worring, R. van den Boomgaard, A.W.M. Smeulders, Tracking nonparameterized object contours in video, IEEE Trans. Image Process. 11 (9) (2002) 1–11.
[20] D.J. Williams, M. Shah, A fast algorithm for active contours and curvature estimation, Comput. Vis. Graph. Image Process.: Image Understand. 55 (1) (1992) 14–26.
[21] P. Ladret, B. Latombe, F. Granada, Active contour algorithm: an attractive tool for snow avalanche analysis, Signal Process. 79 (1999) 197–204.
[22] S. Lobregt, M. Viergever, A discrete dynamic contour model, IEEE Trans. Med. Imag. 14 (1) (1995) 12–24.
[23] J.-P. Planckaert, E.-H. Djermoune, D. Brie, F. Briand, F.-P. Richard, Droplet features extraction with a dynamic active contour for MIG/MAG welding modelling, in: 18th International Conference on Systems Engineering, Coventry, England, 2006, pp. 365–370.
[24] M. Hermans, M. Spikes, G. den Ouden, Characteristics features of the short-circuiting arc welding process, Weld. Rev. Int. (1993) 80–86.
[25] R. Renwick, R. Richardson, Experimental investigation of GTA weld pool oscillations, Weld. J. 62 (2) (1983) 29–35.
[26] J. Lancaster, Metallurgy of Welding, fifth ed., Chapman & Hall, London, 1993.
[27] M. Hermans, G. Ouden, Process behavior and stability in short-circuit gas metal arc welding, Weld. J. 78 (4) (1999) 137–141.

Résumé

Ce mémoire d'habilitation à diriger des recherches synthétise mon activité de recherche au Centre de Recherche en Automatique de Nancy (CRAN) entre 2004 et 2017. Les travaux que j'y ai développés s'articulent autour de la résolution de problèmes inverses en traitement du signal et des images, de l'approximation parcimonieuse en analyse spectrale multidimensionnelle et en sélection de variables, et l'analyse d'images biologiques. La première partie du manuscrit présente un bilan synthétique de mes activités d'enseignement et de recherche menées respectivement à Faculté des Sciences et Technologies de l'Université de Lorraine, et au CRAN. Dans la seconde partie, je présente mes contributions méthodologiques, algorithmiques et appliquées sur trois sujets. Dans le premier chapitre, je présente une partie théorique portant sur l'analyse de perturbation de trois algorithmes haute-résolution d'estimation modale 1-D de signaux exponentiellement amortis. L'objectif est d'obtenir des expressions compactes sur la variance des paramètres du modèle qui soient facilement exploitables pour, par exemple, régler les hyperparamètres des algorithmes. Le deuxième chapitre est consacré à l'estimation modale multidimensionnelle avec comme objectif principal la réduction du coût de calcul. Dans un premier temps, un algorithme d'estimation utilisant une approximation parcimonieuse avec une mise à jour adaptative du dictionnaire est présenté. Par la suite, je décris une autre contribution concernant le calcul des bornes de Cramér-Rao des paramètres du modèle multidimensionnel. Le troisième chapitre est une partie applicative en spectroscopie infrarouge où le problème consiste à sélectionner des bandes spectrales dans un but de classification. Dans la troisième partie, je dresse quelques perspectives dans les domaines de l'approximation parcimonieuse, du traitement en ligne d'images hyperspectrales, et la modélisation de la croissance tumorale.

Mots-clés : estimation modale, analyse spectrale, analyse de perturbation, approximation parcimonieuse, sélection de variables, spectroscopie

Abstract

This manuscript is a synthesis of my research activity at CRAN lab between 2004 and 2017 where my projects dealt with inverse problems in signal and image processing, sparse approximation for harmonic retrieval and simultaneous variable selection, and biological image analysis. The first part of the manuscript is a review of my teaching and research activities conducted respectively at Faculté des Sciences et Technologies (Lorraine University), and CRAN. In the second part, I give details of some theoretical, algorithmic and applied contributions on three subject. The first chapter is a theoretical first-order perturbation analysis of three high-resolution algorithms usually used for 1-D harmonic and modal retrieval. The objective was to derive compact expressions of the parameter variances allowing one to select the optimal tuning parameters in a simple way. The second chapter is devoted to the multidimensional modal retrieval problem where the main objective is to propose a computationally efficient method to estimate the model parameters. First, a simultaneous sparse algorithm with multigrid dictionary refinement is presented. Then, the Cramér-Rao bounds of the multidimensional model parameters are derived. The third chapter is an applicative part. A method for simultaneous variable selection for classification of infrared spectrometry measurements is described. The third part sketches some perspectives in the fields of sparse approximation, on-line hyperspectral image processing, and cancer growth modelling.

Keywords: harmonic retrieval, spectral analysis, perturbation analysis, sparse approximation, variable selection, spectroscopy