



HAL
open science

Goniométrie parcimonieuse de sources radioélectriques : modèles, algorithmes et mises en œuvre robustes

Alice Delmer

► **To cite this version:**

Alice Delmer. Goniométrie parcimonieuse de sources radioélectriques : modèles, algorithmes et mises en œuvre robustes. Traitement du signal et de l'image [eess.SP]. Université Paris-Saclay, 2021. Français. NNT : 2021UPASG085 . tel-03518731

HAL Id: tel-03518731

<https://theses.hal.science/tel-03518731>

Submitted on 10 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Goniométrie parcimonieuse de sources radioélectriques : modèles, algorithmes et mises en œuvre robustes

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 580, Sciences et Technologies de
l'Information et de la Communication (STIC)
Spécialité de doctorat : Traitement du signal et des images
Unité de recherche : SATIE
Référent : École Normale Supérieure Paris-Saclay

**Thèse présentée et soutenue à Gif-Sur-Yvette, le 6 décembre 2021,
par**

Alice DELMER

Composition du jury:

Barbara Nicolas Directrice de recherche, CNRS / Université de Lyon	Présidente
David Brie Professeur, Université de Lorraine	Rapporteur
Yide Wang Professeur, Polytech Nantes - Université de Nantes	Rapporteur
Salah Bourenane Professeur, Ecole Centrale Marseille	Examineur
Caroline Chaux Chargée de recherche, HDR, CNRS / Institut de Mathématiques de Marseille	Examinatrice
Jean-Philippe Ovarlez Directeur de recherche, ONERA Paris-Saclay	Examineur
Pascal Larzabal Professeur, Université Paris-Saclay	Directeur
Anne Ferréol Ingénieur de recherche, HDR, Thales	Coencadrante

Remerciements

A l'issue de trois années de thèse, il est temps d'écrire la dernière page, peut-être la plus laborieuse, de ce manuscrit.

J'aimerais remercier en premier lieu mes deux encadrants, Pascal Larzabal et Anne Ferréol, pour ces trois années passionnantes et enrichissantes. Je les remercie pour leur grande disponibilité, leur bienveillance et la confiance qu'ils m'ont accordée. Complémentaires, ils ont su m'orienter et me guider tout en me laissant une liberté certaine.

Je souhaite ensuite remercier David Brie et Yide Wang pour avoir rapporté ce manuscrit, ainsi que tous les autres membres de mon jury de soutenance : Barbara Nicolas qui a présidé ce jury, Salah Bourennane, Caroline Chaux et Jean-Philippe Ovarlez. Je tiens à les remercier pour leur lecture attentive du manuscrit, leurs questions et remarques pertinentes et constructives, et leurs commentaires chaleureux qui m'ont beaucoup touchés.

Je remercie toutes les équipes du laboratoire SATIE et en particulier Béatrice, Catherine, Dominique, Françoise, Gwenaël et Sophie, qui ont toujours été disponibles pour m'aider ou m'aiguiller, ainsi que les collègues Alexiane, Eric, Jean-Pierre et Mehdi qui égayaient le couloir de l'IDA.

Un grand merci à Emmanuel De Gramont et William Hubert qui m'ont accueilli au sein de leur service à Thales, ainsi qu'à tous les ingénieurs et experts que j'ai pu y croiser pour les échanges scientifiques et le partage de leurs connaissances. Merci également à mes collègues et amis avec qui j'ai partagé un certain nombre de repas ou cafés, toujours dans la bonne humeur : Anaël, Antonio, Arnaud, Benoit, Bertrand, Cyril, Dorin, Elie, Elisa, Eric, Florian, François, Guillaume, Hélène, Laurent, Ludovic, Marc, Marie, Olivier, Philippe, Raphaël, Rémy, Romain, Serdar, Stéphane, Sylvain, Titouan, Vincent, Xavier, et tous les autres collègues du C2.

Je tiens également à remercier mes amis de l'ENS, les Kataclist, les Makia, la Bikok, ses encombrants et tous les autres gens qu'on aime bien, ainsi que tous mes amis lanceurs. Merci à mes amis de longue date, Jonathan, Mike, Thibaud, Thomas et Xavier pour leur présence et leur appui permanent et sans faille.

Enfin, merci à toute ma famille qui me suit et me soutient depuis toujours, à Lounès pour son aide au quotidien, son réconfort, son énergie et sa motivation contagieuse, et à notre petit lapin pour son accompagnement permanent ces derniers mois.

Résumé

Cette thèse porte sur la goniométrie d'émetteurs radioélectriques non coopératifs à partir de signaux reçus sur un réseau d'antennes. Les applications visées dans ce travail sont le scénario aéroporté, caractérisé par un nombre de sources supérieur au nombre de capteurs, et le scénario en environnement urbain, caractérisé par des multi-trajets cohérents. Les méthodes de goniométrie conventionnelles telles que formation de voies et Capon ou encore les méthodes à haute résolution telles que MUSIC ne sont pas performantes dans de tels scénarios. La méthode du maximum de vraisemblance souffre quant à elle d'une complexité de calcul incompatible avec les systèmes opérationnels actuels.

Pour palier ces limitations, la goniométrie est traitée ici avec un formalisme parcimonieux, parfaitement adapté à l'utilisation de tables de calibration dans les systèmes opérationnels. Après avoir montré les intérêts d'une approche reposant sur une régularisation par la norme L_0 , cette thèse s'attaque aux verrous que sont le paramètre de régularisation et la convergence globale des algorithmes d'optimisation. Pour cela, nous construisons et étudions statistiquement des représentations parcimonieuses adaptées i) aux scénarios aéroportés, et ii) aux environnements urbains. L'équivalence avec le maximum de vraisemblance donné par une formulation contrainte permet alors de déterminer un intervalle théorique admissible pour le paramètre de régularisation. Nous étudions également les minimiseurs et les surfaces d'erreurs de différents critères d'optimisation. Cela nous permet de proposer des schémas itératifs de minimisation augmentant la probabilité de convergence globale et ainsi moins sensibles à l'initialisation. De plus, l'algorithme ainsi proposé permet de séparer des sources proches.

Abstract

This thesis deals with the direction-of-arrival (DOA) estimation of non-cooperative radio transmitters from signals received on an antenna array. The applications targeted in this work are the airborne scenario, characterized by a number of sources higher than the number of sensors, and the urban environment scenario, characterized by coherent multipath. Conventional direction-of-arrival estimation methods such as beamforming and Capon method or high resolution methods such as MUSIC are not efficient in such scenarios. As for it, the maximum likelihood method suffers from a computational complexity incompatible with current operational systems.

In order to overcome these limitations, the problem of direction-of-arrival estimation is treated here with a sparse formalism, perfectly adapted to the use of calibration tables in operational systems. After having shown the interests of an approach based on a regularization by the L_0 norm, this thesis tackles the technical issues that are the regularization parameter and the global convergence of optimization algorithms. To this end, we construct and statistically study sparse representations adapted to i) airborne scenarios, and ii) urban environments. The equivalence with the maximum likelihood given by a constrained formulation then allows us to determine a theoretical admissible interval for the regularization parameter. We also study the minimizers and error surfaces of different optimization criteria. This allows us to propose iterative minimization schemes that increase the probability of global convergence and thus are less sensitive to initialization. In that respect, the proposed algorithm enables to separate close sources.

Table des matières

Liste des abréviations	xi
Notations	xiii
Préambule	1
1 Introduction	9
1.1 Modélisation du signal	9
1.1.1 Aperçu du système	9
1.1.2 Modélisation des signaux reçus	9
1.2 Estimation paramétrique des directions d'arrivées : table de calibration et performances	11
1.2.1 Généralités	11
1.2.2 La table de calibration	12
1.2.3 Sources de perturbations des algorithmes d'estimation et évaluation des performances	14
1.3 Méthodes conventionnelles de goniométrie	17
1.3.1 Méthodes basées sur le filtrage spatial	17
1.3.2 Méthodes haute résolution (HR)	18
1.3.3 Performances des méthodes HR et borne de Cramer-Rao	21
1.4 Motivations	22
2 État de l'art des méthodes parcimonieuses en goniométrie	23
2.1 Parcimonie et norme ℓ_0 : généralités	23
2.2 Représentation parcimonieuse du signal en goniométrie	25
2.2.1 Modèle SMV du signal reçu	25
2.2.2 Modèle MMV des signaux reçus	26
2.2.3 Réduction de la dimension pour la goniométrie parcimonieuse	27
2.3 Énoncés des problèmes parcimonieux : problèmes contraints et régularisés par la norme ℓ_0	28

2.4	Résolution des problèmes parcimonieux	29
2.4.1	Les algorithmes gloutons	30
2.4.2	Les algorithmes de descente à gradient-proximal	30
2.4.3	Résolution passant par des relaxations ou approximations de la norme ℓ_0	35
2.5	Conclusion	40
3	Modèles d'observation SMV parcimonieux	41
3.1	Enjeux	41
3.2	Modèle SMV pour le contexte aéroporté	42
3.2.1	Modèle du signal reçu	42
3.2.2	Observation SMV dans le cas aéroporté : la matrice de covariance vectorisée	43
3.2.3	Statistiques de la matrice de covariance vectorisée en présence d'erreurs de modèle	46
3.2.4	Statistiques de la matrice de covariance vectorisée pour un temps d'intégration fini	47
3.3	Modèle SMV pour le contexte urbain	48
3.3.1	Modèle du signal reçu	49
3.3.2	Observation SMV dans le cas de sources cohérentes : le premier vecteur propre de la matrice de covariance	49
3.3.3	Statistiques du vecteur propre en présence d'erreurs de modèle	51
3.4	Résumé des observations et statistiques en fonction du scénario	52
4	Paramètre de régularisation	53
4.1	Introduction et interprétation de la littérature existante pour la goniométrie	53
4.1.1	Réglage d'un paramètre basé sur le niveau de bruit	53
4.1.2	Relation entre problème régularisé et problème contraint : lien avec le maximum de vraisemblance	54
4.1.3	Chemin de régularisation et méthodes homotopiques	57
4.2	Comportement des algorithmes à gradient-proximal selon λ pour une observation donnée	58
4.2.1	Scénario de deux sources fortement corrélées à BT fini, avec un réseau circulaire uniforme à 7 antennes (scénario 1a)	59
4.2.2	Scénario de deux sources fortement corrélées à BT fini, avec un réseau linéaire uniforme à 10 antennes (scénario 1b)	60
4.2.3	Intérêt de l'intervalle pour la minimisation du critère J_{ℓ_0}	60
4.3	Intervalle pour des observations stochastiques d'un scénario bi-sources donné	62

4.3.1	Enjeux et motivations	62
4.3.2	Calcul des statistiques de λ_2^- et λ_2^+ en fonction des statistiques de l'observation \mathbf{y} pour un scénario donné	64
4.3.3	Existence d'un intervalle commun pour des observations stochastiques	74
4.4	Intervalle selon l'écart entre les sources et choix hors-ligne d'une valeur de λ	80
4.4.1	Intervalle pour différents niveaux d'erreurs de modèles	80
4.4.2	Intervalle pour différents RSB et normalisation	82
4.4.3	Performances pour un λ donné	83
4.5	Conclusion	86
5	Stratégie d'optimisation adaptative	89
5.1	Enjeux et démarche	89
5.2	Étude des minimiseurs de J_{ℓ_0}	90
5.2.1	Caractérisation des minimiseurs	90
5.2.2	Illustrations de quelques minimiseurs de J_{ℓ_0}	92
5.3	Étude des minimiseurs de J_{CEL0} comparativement à ceux de J_{ℓ_0}	95
5.3.1	Caractérisation des minimiseurs de J_{CEL0}	95
5.3.2	Étude des surfaces J_{CEL0} et J_{ℓ_0} pour des sources proches	97
5.4	Exploitation de la complémentarité des surfaces J_{ℓ_0} et J_{CEL0} dans un algo- rithme itératif adaptatif	99
5.4.1	Principe et algorithme	99
5.4.2	Évaluation des performances	100
5.4.3	Perspectives	102
	Conclusion et perspectives	105
	Annexes	109
	A Annexe relative au chapitre 5	109
	Bibliographie	111

Liste des abréviations

BP	Basis Pursuit
BPDN	Basis Pursuit De-Noising
CoSaMP	Compressive Sampling Matching Pursuit
CRB	Cramer-Rao Bound
CW	Constant Wave
CELO	Continuous Exact ℓ_0
EQM	Erreur Quadratique Moyenne
ERC	Exact Recovery Condition
FB	Forward-Backward
FBS	Forward-Backward Splitting
FISTA	Fast Iterative Soft Thresholding Algorithm
FNBW	First-Null Beam Width
GPS	Global Positioning System
GSM	Global System for Mobile Communications
HF	High Frequency (Hautes Fréquences)
HPBW	Half-Power Beam Width
HR	Haute Résolution
IHT	Iterative Hard Thresholding
IRL1	Iterative Reweighted L1
ISTA	Iterative Soft Thresholding Algorithm
KL	Kurdyka-Lojasiewicz
LOS	Line Of Sight
LTE	Long Term Evolution
MCP	Minimax-Concave Penalty
MMV	Multi Measurement Vector
MP	Matching Pursuit
MUSIC	MUltiple SIgnal Classification
MV	Maximum de Vraisemblance
MVD	Maximum de Vraisemblance Déterministe

OMP	Orthogonal Matching Pursuit
PGD	Projected Gradient Descent
PMR	Private Mobile Radiocommunications
REQM	Racine de l'Erreur Quadratique Moyenne
RIP	Restricted Isometry Property
RSB	Rapport Signal sur Bruit
SBR	Single Best Replacement
SCAD	Smoothly Clipped Absolute Deviation
SHF	Super High Frequency
SLO	Smoothed L0
SMV	Single Measurement Vector
UCA	Uniform Circular Array
UHF	Ultra High Frequency
ULA	Uniform Linear Array
VHF	Very High Frequency

Notations

Définitions et notations générales

Les lettres latines minuscules en gras représentent des vecteurs, les lettres latines majuscules en gras représentent des matrices.

\mathbb{I}_G	ensemble des $G \in \mathbb{N}_+$ premiers indices $\mathbb{I}_G = \{1, \dots, G\}$
$\text{card}(\omega) = \#\omega$	cardinalité (i.e. nombre d'éléments) de l'ensemble ω
$\Re(x)$	partie réelle de la variable x
$\Im(x)$	partie imaginaire de la variable x

Algèbre linéaire

\mathbf{I}_N	matrice identité carrée de dimension $N \times N$
$(\cdot)^T$	transposée d'un vecteur ou d'une matrice
$(\cdot)^H$	transposée conjuguée d'un vecteur ou d'une matrice
$(\cdot)^*$	conjugué d'un vecteur ou d'une matrice
$\mathbf{M}^\#$	pseudo-inverse de Moore-Penrose de la matrice \mathbf{M}
$\text{rank}(\mathbf{M})$	rang de la matrice \mathbf{M}
$\langle \cdot \cdot \rangle$	produit scalaire hermitien de l'espace \mathbb{C}^g

Définitions de fonctions

$\mathbb{1}$	fonction indicatrice : $\mathbb{1}_{condition} = \begin{cases} 1 & \text{si la condition est respectée} \\ 0 & \text{sinon} \end{cases}$
\mathbf{i}	fonction d'appartenance : $\mathbf{i}_{condition} = \begin{cases} 0 & \text{si la condition est respectée} \\ +\infty & \text{sinon} \end{cases}$
$(\cdot)_+$	fonction $(u)_+ = \max(0, u)$

Définitions en relation avec l'estimation parcimonieuse

$\text{prox}_{\beta f}$	opérateur proximal de la fonction βf , défini par l'équation (2.31)
Φ_{pen}	fonction de pénalité générique
Φ_{ℓ_0}	fonction de pénalité ℓ_0 définie par l'équation (2.38)
Φ_{CEL0}	fonction de pénalité CEL0 définie par l'équation (2.45)
J_{pen}	fonction coût régularisée par une fonction de pénalité Φ_{pen} (2.37)
J_{ℓ_0}	fonction coût régularisée par la fonction de pénalité ℓ_0
J_{CEL0}	fonction coût régularisée par la fonction de pénalité CEL0
ω^c	complémentaire de l'ensemble ω
\mathbf{D}_ω	sous-matrice de $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_G] \mid \mathbf{D}_\omega = [\mathbf{d}_{\omega_1}, \dots, \mathbf{d}_{\#\omega}]$
\mathbf{u}_ω	sous-vecteur de $\mathbf{u} = [u_1, \dots, u_G] \mid \mathbf{u}_\omega = [u_{\omega_1}, \dots, u_{\#\omega}]$
$C_\omega(\mathbf{u}_\omega)$	opérateur de complétion de zéros défini dans l'équation (2.10)

Préambule

Cette thèse a été proposée par THALES. Elle fait l'objet d'une convention CIFRE et s'inscrit dans le cadre d'une collaboration entre cette entreprise et le laboratoire des Systèmes et Applications des Technologies de l'Information et de l'Énergie (SATIE), placé sous la tutelle notamment de l'Université Paris-Saclay, de l'École Normale Supérieure Paris-Saclay et du CNRS. La thèse a démarré en octobre 2018.

Cadre général

La goniométrie est un problème classique en traitement d'antennes : il s'agit d'estimer les directions d'arrivées ou angles d'arrivées d'une ou plusieurs sources à partir des signaux reçus sur un réseau d'antennes. Ce problème a de multiples applications à la fois dans les secteurs civils et militaires. C'est en particulier une étape essentielle pour la localisation d'émetteurs radio-électriques ou sonores, ce qui en fait un enjeu majeur dans des domaines tels que le contrôle du spectre, les télécommunications, la détection sous-marine, la recherche de victimes enfouies après des catastrophes naturelles, le suivi et la protection d'espèces animales ou encore l'usine 4.0.

Les applications visées par ces travaux relèvent de systèmes de goniométrie par réseaux multicapteurs qui fonctionnent en écoute passive, pour des émetteurs non-coopératifs. Dans le cadre de cette thèse, on s'intéresse à des émetteurs radio-électriques, par exemple des téléphones mobiles, des radars, des stations de télécommunications ou tout autre source émettant des ondes électromagnétiques dans les gammes hautes fréquences (HF, 1-30 MHz), très hautes fréquences (VHF, 30-300 MHz), ultra hautes fréquences (UHF, 300 MHz-3 GHz) ou supra-hautes fréquences (SHF, 3 GHz-6 GHz). Les méthodes de goniométrie peuvent cependant aussi être appliquées à des émetteurs sonores tels que des voix, des sons etc.

Contexte et motivation des travaux

Les premières méthodes de goniométrie d'ondes radio-électriques apparaissent dans le domaine de la navigation, bien avant l'avènement du GPS. Elles permettent par exemple à des embarcations maritimes de rejoindre un port équipé d'un radiophare : il suffit de

suivre la direction d'arrivée estimée du signal reçu par le radio-goniomètre du bateau. Les premières méthodes telles que l'interférométrie ou la formation de voies sont ainsi développées pour des cas mono-sources. Capon [15] a ensuite amélioré ces méthodes avec un filtrage spatial s'adaptant aux interférences dans chaque direction pointée. Cependant, ces méthodes restent biaisées dans le cas multi-sources, et inefficaces en présence de sources de faibles puissances.

Les méthodes de goniométrie à haute résolution (HR) telles que MUSIC (pour *MU*ltiple *SI*gnal *C*lassification) [6, 7, 66] ou le maximum de vraisemblance, découvertes à la fin des années 1970, permettent d'estimer les directions d'arrivées de plusieurs sources incidentes de manière plus robuste que Capon.

La méthode MUSIC est une méthode dite à sous-espaces : elle exploite finement la structure algébrique du modèle d'observation, notamment la propriété d'orthogonalité entre les sous-espaces signal et bruit. Elle permet ainsi de surmonter les limitations en résolution des méthodes conventionnelles et super-résolutives comme Capon. La méthode présente l'avantage d'être asymptotiquement non biaisée - pour une durée d'observation infinie. Les bonnes performances de MUSIC reposent en effet sur la bonne estimation de la matrice de covariance des signaux reçus, dont la précision d'estimation est inversement proportionnelle à la durée d'observation : elle doit donc être estimée sur un grand nombre de réalisations indépendantes des signaux reçus. Les performances, et par conséquent la résolution, se détériorent également en présence de faibles rapport signal sur bruit (RSB), d'erreurs de modèle ou de sources cohérentes - où elles peuvent même être mises en défaut [30, 33, 74].

Les méthodes du maximum de vraisemblance sont sous certaines conditions asymptotiquement efficaces et non biaisées, et présentent de plus l'intérêt de pouvoir estimer les directions d'arrivées de sources fortement corrélées ou cohérentes. Mais elles sont cependant aussi sensibles aux erreurs de modèle. De plus, leur coût calculatoire est très important car toutes les directions doivent être estimées conjointement, et il est très rapidement prohibitif - dès lors que l'on dépasse quelques paramètres à estimer. C'est surtout pour cette raison que le maximum de vraisemblance a une utilisation limitée.

Il y a donc encore de nombreux obstacles à l'utilisation des méthodes HR dans certains contextes spécifiques et stratégiques pour Thales. Dans le cadre de cette thèse, nous nous focalisons, pour des applications opérationnelles, sur deux scénarios sensibles d'un point de vue scientifique car hors du champ d'application des méthodes conventionnelles ou HR.

Le premier contexte applicatif est un scénario aéroporté. Dans ce contexte, le porteur peut être un drone ou un avion : il est de dimension nécessairement réduite, avec une charge utile faible. Les réseaux d'antennes installés sont compacts, à la fois en terme de nombre d'antennes, de nombre de voies de réception et d'ouverture du réseau. Les algorithmes HR offrent alors de faibles résolutions et une faible robustesse aux ambiguïtés en contexte multi-sources, d'autant plus préjudiciable que le nombre d'émetteurs augmente avec l'altitude

du porteur : on peut alors avoir un nombre de sources supérieur au nombre d'antennes. De plus, le temps d'observation des signaux dans ce contexte est réduit, et les rapports signaux à bruit faibles (émissions distantes).

Ce scénario est équivalent au scénario inverse où l'on cherche à localiser des essais de drones à l'aide d'une station sol fixe.

Le second contexte présentant plusieurs difficultés opérationnelles auquel nous nous intéressons est le contexte urbain, par exemple pour des systèmes de communications GSM, LTE ou PMR. Le signal émis par un émetteur se réfléchit sur des bâtiments environnants, et il est alors reçu par le réseau antennaire sous forme de multi-trajets, chacun des trajets étant vu comme une source par le système. La mise en œuvre d'un algorithme à sous-espaces comme MUSIC nécessite la connaissance ou la bonne estimation du nombre de sources, ce qui est rendu compliqué dans ce contexte par la forte corrélation temporelle entre les multi-trajets. La forte corrélation dégrade très fortement les performances, et pour des trajets quasi-cohérents, la méthode MUSIC peut être mise en défaut car elle sous-estime le nombre de sources. La faible durée d'observation ou d'émission du signal, par exemple dans le cas d'émissions en évadion de fréquence, amène également une dégradation des performances.

Les dégradations de performances dues à des multi-trajets partiellement corrélés se retrouvent également dans un contexte de communications ionosphériques HF, par exemple pour des communications CW, où les sources sont de plus proches spatialement en élévation.

Objectifs de la thèse

Les méthodes conventionnelles ne fonctionnent quasiment pas dans les contextes opérationnels multi-sources qui nous intéressent. La méthode MUSIC présente également des faiblesses majeures dans la résolution des problématiques précédentes puisqu'elle ne peut traiter le cas de sources cohérentes, ou de l'estimation des directions d'arrivées de plus de sources que de capteurs dans le cas général. Certes, les méthodes du maximum de vraisemblance ont le potentiel théorique pour palier le problème des sources cohérentes, mais leur mise en œuvre s'avère impossible à l'heure actuelle en raison de la charge de calcul nécessaire en présence de plusieurs trajets. De plus, les contextes opérationnels entraînent des rapports signal à bruit faibles, un temps d'intégration réduit, et des erreurs de modèles inhérentes aux systèmes. Les performances de MUSIC sont alors dégradées et ne sont pas à la hauteur de celles attendues pour un cahier des charges donné. Les exigences portant sur la précision de localisation ou le pouvoir de résolution peuvent en effet ne pas être atteintes.

De nouvelles approches de goniométrie exploitant des méthodes parcimonieuses ont vu le jour au début des années 2000. L'intérêt pour ces méthodes s'est intensifié ces dernières

années devant leur potentiel et les avantages qu'elles apportent par rapport aux méthodes classiques : estimation des directions d'arrivées avec très peu d'échantillons voire avec une unique mesure instantanée, estimation dans des contextes avec plus de sources que de capteurs... Leurs performances et la résolution atteinte dans des scénarios de rapport signal à bruit faible sont également largement mises en avant.

Ces méthodes exploitent la parcimonie en angles d'arrivées et reposent sur une discrétisation des directions d'arrivées possibles de l'espace de recherche selon une grille. Ce formalisme est en totale adéquation avec l'utilisation, dans nos systèmes opérationnels, d'une table de calibration construite à partir d'enregistrements des réponses du réseau dans les directions du champ d'observation. Cette dernière remarque m'apparaît fondamentale pour justifier tous les espoirs que je vais mettre sur les méthodes parcimonieuses dans ce travail.

Il existe encore plusieurs verrous à l'utilisation de ces méthodes dans des systèmes opérationnels : la recherche sur les méthodes parcimonieuses nécessite encore de nombreux développements scientifiques. Le formalisme mathématique d'estimation parcimonieuse a notamment été construit autour de la pseudo-norme ℓ_0 , qui est discontinue, non dérivable, non convexe... Les difficultés d'optimisation sont ainsi nombreuses concernant la convergence et donc l'initialisation des algorithmes. Les difficultés ont poussé à l'utilisation de la norme ℓ_1 dans de nombreuses applications parcimonieuses. Cependant, pour les réseaux à peu de capteurs de nos systèmes opérationnels, les résolutions et les performances que nous attendons ne peuvent pas se satisfaire des méthodes basées sur la norme ℓ_1 et traditionnellement rencontrées dans la littérature. De plus, le formalisme parcimonieux fait appel à des paramètres souvent fixés de manière empirique, ce dont on ne peut se satisfaire dans un système opérationnel.

Cette thèse a pour objectif de lever les verrous actuels des méthodes parcimonieuses pour proposer des systèmes de goniométrie opérationnels fiables et qui correspondent aux réalités terrain. Les développements passent nécessairement par une approche pluridisciplinaire combinant la physique de réception, les statistiques des signaux et l'algorithmie d'optimisation mise en œuvre. La thèse présentée ici s'appuie sur ces différents champs pour permettre le développement de nouvelles méthodes parcimonieuses dans des systèmes opérationnels et dans les scénarios décrits ci-dessus : le scénario aéroporté, dans lequel le nombre de sources peut être supérieur au nombre d'antennes, et le scénario urbain avec des sources fortement corrélées ou cohérentes. Il s'agit d'approches en rupture avec les technologies actuellement utilisées par Thales.

Ce travail est en particulier consacré à la mise en œuvre d'une méthode régularisée reposant sur la norme ℓ_0 . Il se concentre sur les modèles d'observation, sur la détermination du paramètre de régularisation et sur la convergence de schémas itératifs.

Organisation du document

Le premier chapitre est une introduction générale sur la goniométrie. Après quelques éléments sur la propagation des signaux et sur les méthodes conventionnelles et à haute résolution, j'évoque leurs performances dans les contextes qui nous intéressent. Ce chapitre justifie l'utilisation de nouvelles méthodes pour un contexte opérationnel difficile.

Le deuxième chapitre présente l'état de l'art spécifique à l'estimation des directions d'arrivées par les méthodes parcimonieuses. Il permet de justifier le choix fait de travailler dans cette thèse sur la minimisation d'un critère des moindres carrés régularisé par la norme ℓ_0 .

Le troisième chapitre présente les deux modèles d'observation introduits pour répondre à nos problématiques. J'y définis des modèles d'observation spécifiques à nos scénarios afin d'inscrire ce travail dans le cadre formel de l'estimation des directions d'arrivée par des méthodes parcimonieuses. Dans le but de simplifier la complexité numérique tout en s'appuyant sur l'existence de preuves de convergence algorithmique, je présente deux modèles d'observation vectoriels qui utilisent les informations contenues dans la matrice de covariance. Ainsi, dans le contexte aéroporté, j'étudie le modèle de la matrice de covariance vectorisée et calcule ses statistiques à la fois dans le cas d'erreurs de modèles et pour un temps d'intégration fini. Pour le contexte urbain, je propose d'exploiter le premier vecteur propre de la matrice de covariance et l'étudie en présence d'erreurs de modèle. Pour des sources quasi-cohérentes, celui-ci contient en effet toute l'information utile. Ces deux modèles, à la dimension réduite et qui condensent toute l'information utile, peuvent se mettre sous une forme parcimonieuse grâce à un dictionnaire issu de la table de calibration. La connaissance des statistiques de ces modèles, développées dans ce chapitre, sera exploitée pour la détermination du paramètre de régularisation dans le quatrième chapitre.

Le quatrième chapitre porte sur le paramètre de régularisation. Le paramètre de régularisation détermine le poids relatif entre le terme d'attache aux données et la pénalité, ici la norme ℓ_0 , dans le critère régularisé qui fait l'objet d'une minimisation pour l'estimation parcimonieuse. Je propose une extension des travaux de Mila Nikolova dans le cas stochastique, et montre, pour des sources suffisamment éloignées, l'existence d'un intervalle éligible pour le paramètre de régularisation qui est indépendant de l'observation. L'analyse de cet intervalle me permet de définir le domaine de validité opérationnel de la méthode proposée. En particulier, l'analyse selon l'écart entre les sources me permet de déterminer une limite de résolution.

Le cinquième chapitre est consacré à la conception et au développement des algorithmes qui seront mis en œuvre. J'étudie d'abord les minimiseurs de la surface J_{ℓ_0} correspondant au critère régularisé par la norme ℓ_0 . Le nombre important de minimas locaux est à l'origine de difficultés de convergence et donc d'initialisation des algorithmes sous-optimaux utilisés, par exemple l'algorithme IHT (Iterative Hard Thresholding) dont la convergence n'est garantie que vers un point fixe. J'étudie en comparaison les minimiseurs du critère régularisé

par la pénalité CEL0, approximation continue récemment proposée de la norme ℓ_0 . Cette étude met en évidence, dans un schéma opérationnel, l'intérêt d'utiliser les deux surfaces d'erreur. Je propose un nouveau schéma adaptatif et montre sa pertinence par rapport à l'état de l'art et sa supériorité, tant dans les modalités de mise en œuvre que dans ses performances statistiques.

Le sixième chapitre conclut le travail proposé ici et offre quelques pistes de réflexions pour de futurs travaux.

Contributions

- Pour le scénario en milieu urbain, j'ai proposé d'utiliser comme observation le premier vecteur propre de la matrice de covariance car il porte toute l'information en présence de sources fortement corrélées voir cohérentes. Celui-ci peut s'obtenir très facilement par quelques itérations de la méthode des puissances. J'ai utilisé de manière originale cet unique vecteur propre pour effectuer la localisation de multi-trajets en goniométrie parcimonieuse. J'ai calculé les statistiques de cette observation en présence d'erreurs de modèle.
- Dans le cas aéroporté, j'ai proposé d'utiliser la matrice de covariance vectorisée. Cette technique permet de localiser plus de sources que de capteurs par la création d'un réseau virtuel comprenant beaucoup plus de capteurs que de capteurs physiques. J'ai mené les mêmes analyses statistiques que précédemment, et j'ai traité également le cas du temps d'intégration fini.
- Dans le cadre de ces deux scénarios difficiles, j'ai souligné la nécessité de travailler dans un cadre parcimonieux avec la norme ℓ_0 . J'ai de plus montré que l'estimation des directions d'arrivées par le problème contraint avec la norme ℓ_0 correspond à l'estimation du maximum de vraisemblance déterministe.
- Les deux modèles d'observation proposés m'ont permis d'exploiter la littérature parcimonieuse sur les modèles SMV (Single Measurement Vector), qui est plus fournie que celle sur les modèles MMV (Multi Measurement Vector). En particulier j'ai exploité la dualité problème contraint - problème régularisé, étudié les surfaces d'erreurs des critères et j'ai développé des algorithmes robustes à la complexité réduite.
- J'ai étendu les travaux de Mila Nikolova portant sur le paramètre de régularisation, consacrés à une observation déterministe, au cas d'observations stochastiques. J'ai ainsi obtenu, à l'aide d'approximations statistiques, un intervalle admissible pour le paramètre de régularisation et j'ai étudié son existence en fonction des scénarios rencontrés (RSB, niveau d'erreurs de modèle, écart entre les sources, temps d'intégration, corrélation...). Des simulations intensives ont confirmé le bien fondé de cette approche.
- J'ai montré que ces travaux sur le paramètre de régularisation permettent de déterminer des limites de fonctionnement théoriques des méthodes parcimonieuses consi-

dérées, par exemple la limite de résolution.

- J’ai analysé les surfaces d’erreurs des critères régularisés par la norme ℓ_0 et la pénalité CEL0. J’ai vérifié dans mes simulations que la pénalité CEL0 supprime certains minima locaux du critère ℓ_0 tout en préservant le minimum global. Ceci explique l’excellent comportement des algorithmes de minimisation du critère régularisé par la pénalité CEL0 quand les sources sont éloignées de plus d’une largeur de lobe. J’ai aussi mis en évidence, entre les sources, la présence d’un bassin d’attraction dont la taille augmente pour des sources à l’intérieur du lobe. Cela se traduit par une perte de résolution des méthodes basées sur la pénalité CEL0. Ces études sont essentielles pour proposer de nouveaux schémas d’optimisation améliorant la convergence globale pour tous types de scénarios.
- J’ai proposé un nouveau schéma d’optimisation adaptatif reposant sur les surfaces d’erreurs CEL0 et ℓ_0 . Le comportement adaptatif de cet algorithme est contrôlé par des critères d’arrêts évoluant selon les caractéristiques de l’estimée à l’itération courante. Les simulations intensives effectuées ont montré le grand intérêt de cet algorithme par rapport à l’existant.

Valorisations

Les travaux effectués au cours de cette thèse ont donné lieu aux valorisations suivantes.

Revue internationale

- J1.** A. Delmer, A. Ferréol et P. Larzabal. On the Complementarity of Sparse L0 and CEL0 Regularized Loss Landscapes for DOA Estimation, *Sensors*, 2021, 21(18)
- J2.** A. Delmer, A. Ferréol et P. Larzabal. L0 Regularization Parameter for Highly-Correlated Sources DOA Estimation with Modeling Errors, *en relecture finale pour soumission à Signal Processing*
- J3.** A. Delmer, A. Ferréol et P. Larzabal. A New Strategy for Regularization Parameter in L0-Sparse DOA Estimation, *en cours de rédaction pour soumission à IEEE Transactions on Signal Processing*

Conférences internationales

- C1.** A. Delmer, A. Ferréol et P. Larzabal. Selection of regularization parameter in sparse inverse problems for doa estimation. *Journal of Physics : Conference Series*, 2020, 1476
- C2.** A. Delmer, A. Ferréol et P. Larzabal. On Regularization Parameter for L0-Sparse Covariance Fitting Based DOA Estimation. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Amsterdam, 2020*

- C3.** A. Delmer, A. Ferréol et P. Larzabal. L0-Sparse DOA Estimation of Close Sources with Modeling Errors. *European Signal Processing Conference (EUSIPCO), Barcelona, 2020*

Conférences françaises

- C4.** A. Delmer, A. Ferréol et P. Larzabal. Quel paramètre de régularisation pour une radio-goniométrie parcimonieuse avec la norme L0? *GRETSI, Lille, 2019*

Brevet

- B1.** A. Delmer, A. Ferréol et P. Larzabal. Procédé et dispositif d'estimation des angles d'arrivée de signaux incidents sur un réseau d'antennes. Classifié par la DGA, 2021.

Logiciels

- L1.** Mise à disposition au sein du service d'une librairie de goniométrie parcimonieuse.
-

Chapitre 1

Introduction

1.1 Modélisation du signal

1.1.1 Aperçu du système

Un système de goniométrie comprend trois éléments : un réseau de capteurs, une unité de traitement permettant d'enregistrer et d'exploiter les signaux provenant d'une ou plusieurs sources avec des angles d'incidence différents, et, dans un système opérationnel, d'une interface homme-machine permettant de visualiser les résultats. Dans notre cas, les signaux proviennent de sources radio-électriques et les capteurs sont des antennes. La figure 1.1 présente la structure générale d'un tel système.

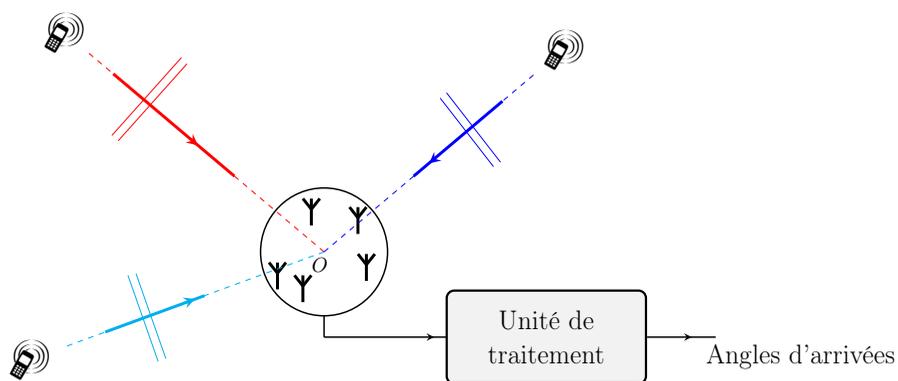


FIGURE 1.1 – Exemple de système de goniométrie à 5 antennes, avec 3 sources incidentes.

1.1.2 Modélisation des signaux reçus

Une source m émet un signal radioélectrique $s_m(t)$ d'amplitude complexe $\alpha_m(t)$, de bande B_m et de fréquence centrale f_m . Le signal s'exprime alors sous la forme suivante :

$$s_m(t) = \alpha_m(t) e^{j2\pi f_m t} \quad (1.1)$$

On considère un réseau composé de N antennes. L'expression du signal reçu par le n -ième capteur est :

$$s_m(t - \tau_n) = \alpha_m(t - \tau_n) e^{j2\pi f_m(t - \tau_n)}, \quad (1.2)$$

où τ_n est le temps mis par l'onde pour atteindre le n -ième élément depuis le centre du réseau. Notons $\Delta\tau = \max_{n,n'} |\tau_n - \tau_{n'}|$ le temps de parcours de l'onde à travers le réseau. Dans l'hypothèse bande étroite, on suppose

$$B_m \times \Delta\tau \ll 1, \quad (1.3)$$

où B_m est la largeur de bande du signal $s_m(t)$. Dans ce cas, on vérifie $\alpha_m(t - \tau_n) \approx \alpha_m(t)$. On a alors

$$s_m(t - \tau_n) \approx s_m(t) e^{-j2\pi f_m \tau_n}. \quad (1.4)$$

Dans l'hypothèse de front d'onde plan, le temps de parcours τ_n entre le capteur n et le centre du réseau est fonction de l'angle d'arrivée θ_m de la source sur le réseau et des coordonnées \mathbf{p}_n du n -ième capteur selon la formule :

$$\tau_n = \mathbf{k}^T(\theta_m) \mathbf{p}_n, \quad (1.5)$$

avec $\mathbf{k}(\theta_m)$ le vecteur d'onde. Celui-ci s'écrit, en 2D : $\mathbf{k}(\theta_m) = -\frac{1}{c} [\cos \theta_m \quad \sin \theta_m]^T$, avec $c = \lambda_0 f_m$ la célérité de l'onde. La figure 1.2 représente les différentes variables introduites.

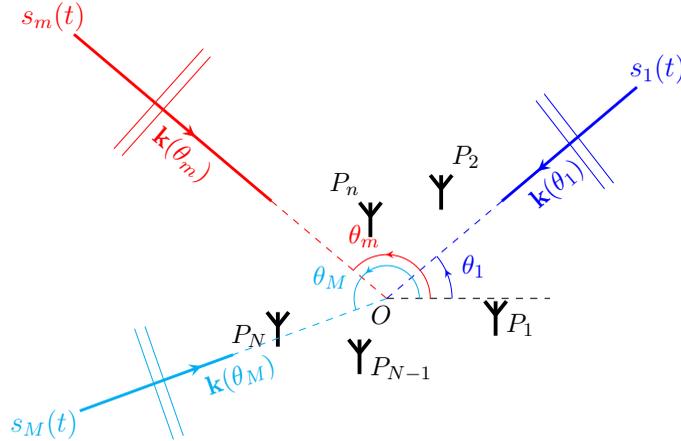


FIGURE 1.2 – Exemple de réseau d'antennes à N capteurs et M sources incidentes $s_m(t)$ d'angles d'arrivées θ_m , $m = \{1, \dots, M\}$.

Pour M sources ayant les incidences $\theta_1, \dots, \theta_M$ sur le réseau d'antennes, le signal mesuré par le n -ième capteur est le suivant :

$$x_n(t) = \frac{1}{\sqrt{N}} \sum_{m=1}^M e^{-j2\pi f_0 \mathbf{k}^T(\theta_m) \mathbf{p}_n} s_m(t) + n_n(t), \quad (1.6)$$

avec $n_n(t)$ un bruit additif. Le signal $\mathbf{x}(t)$ reçu en sortie du réseau se met alors sous la forme :

$$\mathbf{x}(t) = [x_1(t) \ \dots \ x_N(t)]^T = \sum_{m=1}^M \mathbf{a}(\theta_m) s_m(t) + \mathbf{n}(t), \quad (1.7)$$

où

$$\mathbf{a}(\theta_m) = \frac{1}{\sqrt{N}} \left[e^{-j2\pi f_0 \mathbf{k}^T(\theta_m) \mathbf{p}_1} \ \dots \ e^{-j2\pi f_0 \mathbf{k}^T(\theta_m) \mathbf{p}_N} \right]^T \quad (1.8)$$

est la réponse du réseau (ou vecteur directeur) à une source incidente d'angle d'arrivée θ_m pour des antennes à gains identiques et sans couplage mutuel entre les capteurs, et $\mathbf{n}(t) = [n_1(t), \dots, n_N(t)]^T$ est le vecteur de bruit. On définit l'ensemble Θ qui contient les directions d'arrivées des sources :

$$\Theta = \{\theta_1, \dots, \theta_M\}. \quad (1.9)$$

On introduit $\mathbf{A}_\Theta = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_M)]^T$ la matrice de mélange des ondes incidentes sur le réseau et $\mathbf{s}(t) = [s_1(t) \ \dots \ s_M(t)]^T$ le vecteur signal des sources. Le signal reçu s'écrit alors :

$$\mathbf{x}(t) = \mathbf{A}_\Theta \mathbf{s}(t) + \mathbf{n}(t). \quad (1.10)$$

1.2 Estimation paramétrique des directions d'arrivées : table de calibration et performances

1.2.1 Généralités

La goniométrie a pour objectif d'estimer les directions d'arrivées des sources, représentées par des angles par rapport à un axe de référence. L'estimation de paramètres d'un modèle constitue une partie importante du traitement du signal. L'estimation se fait à partir d'un ensemble de données mesurées et d'un modèle d'observation.

Dans le cadre de la goniométrie, on considère le modèle initial présenté ci-dessus :

$$\mathbf{x}(t) = \sum_{m=1}^M \mathbf{a}(\theta_m) s_m(t) + \mathbf{n}(t) = \mathbf{A}_\Theta \mathbf{s}(t) + \mathbf{n}(t). \quad (1.11)$$

- $\mathbf{x}(t) = [x_1(t) \ \dots \ x_N(t)]^T$ est le vecteur signal reçu par le réseau de N antennes, $x_n(t)$ étant le signal reçu par la n -ième antenne.
- $\Theta = \{\theta_1, \dots, \theta_M\}$ est l'ensemble constitué des M directions d'arrivées des sources : c'est le paramètre d'intérêt à estimer.
- $\mathbf{A}_\Theta = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_M)]$ est la matrice de mélange des sources, avec $\mathbf{a}(\theta_m)$ la réponse du réseau de capteurs ou vecteur directeur pour la direction θ_m , pour $m = 1 \dots M$.
- $\mathbf{s}(t) = [s_1(t) \ \dots \ s_M(t)]^T$ est le vecteur signal des sources, avec $s_m(t)$ le signal de la m -ième source. Le vecteur $\mathbf{s}(t)$ est inconnu.

- $\mathbf{n}(t) = [n_1(t) \dots n_N(t)]^T$ est le vecteur aléatoire modélisant le bruit sur chaque capteur : il est également inconnu.

L'objectif est donc d'estimer les directions d'arrivées θ_m à partir du signal $\mathbf{x}(t)$ reçu par le réseau d'antennes.

L'estimation est rendue possible par la connaissance de la réponse du réseau $\mathbf{a}(\theta)$ à une direction θ . Notons cependant que les méthodes développées dans cette thèse ne nécessitent pas de structure particulière de la fonction $\mathbf{a}(\theta)$. L'expression analytique donnée par l'équation (1.8) repose sur la connaissance de la position des capteurs \mathbf{p}_n , et suppose que les antennes sont identiques et qu'il n'y a pas de couplage mutuel entre capteurs. Or bien souvent, la position des capteurs peut être imprécise, et les hypothèses de gain ou de couplage sur les antennes ne sont pas entièrement vérifiées. Cela introduit ce qu'on appelle des erreurs de modèle : il y a un écart entre les réponses du réseau $\mathbf{a}(\theta)$ que l'on considère pour l'estimation dans les algorithmes, et les réponses réelles des capteurs $\tilde{\mathbf{a}}(\theta)$. Le modèle analytique peut également être amélioré pour mieux prendre en compte un éventuel couplage mutuel ou autres gains d'antennes. La bonne estimation de ces paramètres peut cependant nécessiter une étape de calibration.

1.2.2 La table de calibration

1.2.2.1 Définition

L'étape de calibration permet de prendre en compte les caractéristiques physiques du réseau de capteurs et d'obtenir ainsi de meilleures performances de goniométrie pour les systèmes opérationnels. La calibration s'effectue avec un signal pilote : on mesure la réponse du réseau de capteurs en fonction de l'incidence d'une source. La mesure peut être effectuée par exemple dans une chambre anéchoïque mais aussi lors d'essais en vol lorsque le réseau est sur un aéronef. La calibration nécessite une discrétisation de l'espace de recherche : on considère un nombre fini G de directions ψ_g , avec $\psi_g = (g-1)\delta\psi$, pour $g = 1 \dots G$, et $\delta\psi = \Delta\psi/G$, où $\Delta\psi$ est le champ de vision considéré. On parle de grille de discrétisation. On note Ψ l'ensemble des directions considérées qui composent la grille :

$$\Psi = \{\psi_1, \dots, \psi_G\} \quad (1.12)$$

Les vecteurs directeurs correspondants aux angles de la grille sont regroupés dans la table de calibration \mathbf{A} :

$$\mathbf{A} = [\mathbf{a}(\psi_1), \dots, \mathbf{a}(\psi_G)]. \quad (1.13)$$

La table de calibration peut servir à estimer la matrice de couplage. Elle peut également être semi-analytique, où des interpolations complètent des mesures parcellaires.

Dans le cadre de cette thèse, et en particulier parce que l'on s'intéresse aux méthodes parcimonieuses, on considère que la table de calibration \mathbf{A} contient l'ensemble des réponses

connues du réseau, qu'elles soient analytiques ou mesurées, pour les directions discrétisées de l'ensemble Ψ . La table agira donc en tant que dictionnaire. Ces caractéristiques algébriques sont très importantes pour les algorithmes de goniométrie.

On considère de plus que les directions d'arrivées θ_m font partie des directions candidates ψ_g de la grille : on pose $\Theta \subset \Psi$. La matrice de mélange des sources \mathbf{A}_Θ est ainsi une sous-matrice de la table de calibration \mathbf{A} , qui contient uniquement les colonnes correspondantes aux angles dans Θ .

1.2.2.2 Caractérisation

Corrélation On considère la table de calibration $\mathbf{A} = [\mathbf{a}(\psi_1), \dots, \mathbf{a}(\psi_G)]$. On définit la corrélation spatiale $r_{\mathbf{A}}(\psi_i, \psi_j)$ comme la valeur absolue du produit scalaire normalisé entre les deux vecteurs directeurs $\mathbf{a}(\psi_i)$ et $\mathbf{a}(\psi_j)$:

$$r_{\mathbf{A}}(\psi_i, \psi_j) = \frac{|\mathbf{a}^H(\psi_i)\mathbf{a}(\psi_j)|}{\sqrt{\mathbf{a}^H(\psi_i)\mathbf{a}(\psi_i)}\sqrt{\mathbf{a}^H(\psi_j)\mathbf{a}(\psi_j)}} \quad (1.14)$$

Elle vérifie $0 \leq r_{\mathbf{A}}(\psi_i, \psi_j) \leq 1$. La corrélation indique la ressemblance entre deux vecteurs directeurs : plus elle est proche de 1, plus les vecteurs sont colinéaires et donc proches.

Largeur de lobe On utilise la corrélation au carré pour définir les propriétés d'un réseau donné, en particulier la largeur de lobe. Un exemple de la courbe $r_{\mathbf{A}}^2(\psi_i, \psi_j)$ pour un angle donné $\psi_i = 180^\circ$ est représenté sur la figure 1.3 en fonction de ψ_j pour un réseau circulaire uniforme à $N = 7$ antennes de rayon une demi-longueur d'onde.

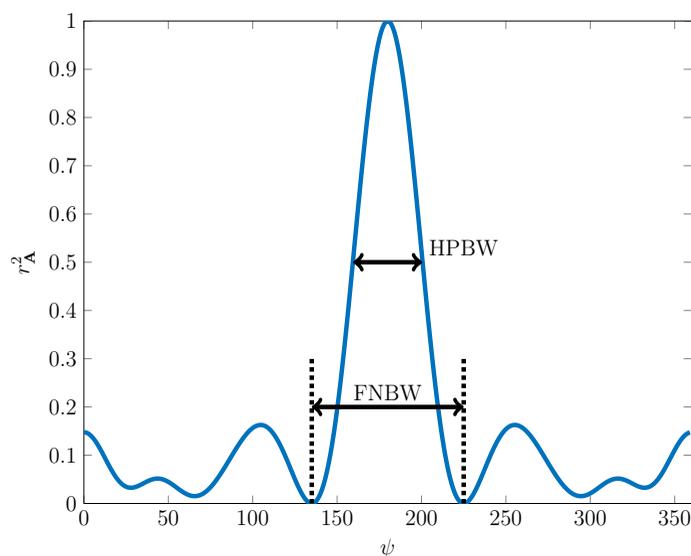


FIGURE 1.3 – Corrélation au carré $r_{\mathbf{A}}^2(180^\circ, \psi)$ en fonction de ψ pour un réseau circulaire uniforme à $N = 7$ antennes de rayon une demi-longueur d'onde.

Sont représentées sur cette courbe la largeur de lobe à mi-puissance (HPBW pour *Half-Power Beam Width*) et la largeur du lobe entre deux annulations de la corrélation (FNBW pour *First-Null Beam Width*). En goniométrie, deux signaux reçus dans le même lobe principal seront difficilement distinguables. Il est essentiel de connaître la largeur de lobe du réseau lorsque l'on étudie les performances d'une méthode de goniométrie en multi-sources.

Cohérence mutuelle La cohérence mutuelle est une mesure utilisée en estimation parcimonieuse pour caractériser un dictionnaire. Elle est définie comme le maximum de la corrélation entre deux colonnes du dictionnaire et donc, en goniométrie, de la table de calibration. On la note $\mu(\mathbf{A})$:

$$\mu(\mathbf{A}) = \max_{i,j} r_{\mathbf{A}}(\psi_i, \psi_j). \quad (1.15)$$

En goniométrie, le dictionnaire est construit à partir de la discrétisation de l'espace des angles d'arrivées, alors que la fonction $\mathbf{a}(\theta)$ est en réalité continue en θ . La cohérence mutuelle, en plus de dépendre de la géométrie du réseau, dépend ainsi notamment du pas de la grille. Pour un pas de grille de 1° , le réseau précédent donne une cohérence mutuelle de 0.99925, ce qui est donc très proche de 1. Un pas de grille de 10° résulte pour le même réseau en une cohérence mutuelle $\mu(\mathbf{A}) \approx 0.93$.

La cohérence mutuelle, en estimation parcimonieuse, permet de déterminer certaines des conditions ou garanties de convergence des méthodes ou algorithmes.

1.2.3 Sources de perturbations des algorithmes d'estimation et évaluation des performances

L'étude des performances d'un algorithme est une étape essentielle pour déterminer la qualité d'une estimation. On note $\hat{\Theta} = \{\hat{\theta}_1, \dots, \hat{\theta}_M\}$ l'estimée des paramètres de valeur théorique Θ en présence d'une perturbation du modèle. Pour un système opérationnel, les angles $\hat{\theta}_m$ estimés par les algorithmes diffèrent des valeurs réelles θ_m en raison de plusieurs sources d'erreurs qui jouent sur les résultats des algorithmes : on étudie en particulier les erreurs de modèle et le nombre d'échantillons K fini des signaux reçus.

On cherche donc à caractériser statistiquement l'écart entre la valeur estimée du paramètre et sa valeur réelle. Il s'agit alors dans notre cas de caractériser l'écart entre une direction estimée $\hat{\theta}_m$ et la direction d'arrivée correspondante θ_m . Celui-ci se mesure par un biais, une variance, une Erreur Quadratique Moyenne (EQM) ou Racine de l'Erreur Quadratique Moyenne (REQM). Avant de pouvoir caractériser l'écart entre une valeur estimée et le paramètre réel, d'autres indicateurs sont nécessaires pour caractériser les performances d'une méthode ou d'un algorithme : le pourcentage d'erreurs, et le pouvoir de résolution.

1.2.3.1 Les sources d'incertitudes des algorithmes de goniométrie

Le signal reçu en sortie du réseau d'antennes $\mathbf{x}(t)$ est bruité (1.11), ce qui entraîne donc des incertitudes dans l'estimation des directions d'arrivées si l'on se base directement sur le signal. Sous l'hypothèse de bruit blanc gaussien, l'utilisation de la matrice de covariance du signal $\mathbf{R}_{xx} = \mathbb{E}[\mathbf{x}(t)\mathbf{x}^H(t)]$, lorsqu'elle est parfaitement connue, permet de supprimer les incertitudes liées au bruit ; beaucoup de méthodes d'estimation reposent sur les propriétés de la matrice de covariance. Cependant, en pratique, la matrice de covariance exacte n'est pas connue, elle est estimée par un nombre fini d'échantillons. Celle-ci est généralement estimée sur K échantillons du signal reçu $\mathbf{x}(t)$ par :

$$\hat{\mathbf{R}}_{xx} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}(t_k)\mathbf{x}^H(t_k), \quad (1.16)$$

avec $t_k = kT_e$ où T_e est la période d'échantillonnage. En présence d'une bande de réception de bruit blanc vérifiant $t_{k+1} - t_k = 1/B$, les échantillons $\mathbf{x}(t_k)$ et $\mathbf{x}(t_{k+1})$ sont indépendants et on a $K = B \times T$. Les performances des estimateurs dépendent de la variance d'estimation de $\hat{\mathbf{R}}_{xx}$, qui dépend de K , encore appelé produit BT. Les performances idéales pour un temps d'observation infini seront appelées les performances asymptotiques dans ce travail : en effet, $\hat{\mathbf{R}}_{xx}$ tend vers \mathbf{R}_{xx} quand K tend vers l'infini.

L'estimation des directions d'arrivées θ_m à partir du signal reçu $\mathbf{x}(t)$ nécessite également la bonne connaissance des vecteurs directeurs $\mathbf{a}(\theta)$. Or les erreurs de modèle inhérentes à un système opérationnel entraînent une méconnaissance des vecteurs directeurs exacts, ce qui se répercute alors sur la précision de l'estimation. Les vecteurs directeurs $\tilde{\mathbf{a}}(\psi)$ du système de mesure ne correspondent pas exactement aux vecteurs directeurs $\mathbf{a}(\psi)$ utilisés dans les algorithmes, bien qu'ils puissent être issus d'une table de calibration expérimentale. Les erreurs peuvent provenir d'un certain nombre de facteurs, comme les erreurs dues aux gains des capteurs, à leurs phases ou à leurs positions, aux dérives mais aussi au couplage mutuel. On considère ainsi un modèle de bruit additif sur les vecteurs directeurs :

$$\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{E} \quad (1.17)$$

avec $\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}(\psi_1) \dots \tilde{\mathbf{a}}(\psi_G)]$ les vecteurs directeurs du système de mesure pour les angles candidats Ψ , $\mathbf{A} = [\mathbf{a}(\psi_1) \dots \mathbf{a}(\psi_G)]$ la table de calibration considérée dans les algorithmes, et $\mathbf{E} = [\mathbf{e}(\psi_1) \dots \mathbf{e}(\psi_G)]$ la matrice des erreurs de modèle. Les vecteurs $\mathbf{e}(\psi_g)$ sont supposés Gaussiens, circulaires, indépendants entre eux et entre capteurs :

$$\mathbb{E}[\mathbf{e}(\psi_i)] = \mathbf{0}, \quad (1.18)$$

$$\mathbb{E} \left[\frac{\mathbf{e}(\psi_i)\mathbf{e}(\psi_j)^H}{\sqrt{\mathbf{a}^H(\psi_i)\mathbf{a}(\psi_i)}\sqrt{\mathbf{a}^H(\psi_j)\mathbf{a}(\psi_j)}} \right] = \frac{\sigma_e^2}{N} \delta_{ij} \mathbf{I}_N, \quad (1.19)$$

$$\mathbb{E} [\mathbf{e}(\psi_i)\mathbf{e}(\psi_j)^T] = \mathbf{0}. \quad (1.20)$$

Ce modèle de perturbation est couramment utilisé [29, 30, 35] et est en général suffisant pour obtenir un bon aperçu des performances d'une approche. Dans la pratique, le niveau des erreurs de modèles σ_e peut être surestimé pour couvrir toutes sortes d'incertitudes. En effet, il est toujours préférable de surévaluer une erreur dans un contexte opérationnel.

1.2.3.2 Détection et pourcentage d'erreurs d'un algorithme

Le premier critère d'évaluation d'une méthode d'estimation est sa capacité à détecter les sources. En présence de M sources, un algorithme doit permettre d'estimer M angles d'arrivées qui puissent chacun s'associer à une source différente. Si une direction estimée $\hat{\theta}$ est située à plus d'une demi-largeur de lobe de toute vraie direction θ , on parle de source fantôme. Il peut s'agir d'un problème d'ambiguïtés, c'est à dire que l'algorithme ne peut estimer de façon unique les M vraies directions d'arrivées. Si $M' > M$ directions sont détectées, le problème peut aussi venir du mauvais réglage d'un seuil de détection. De même si l'on détecte moins de sources qu'il n'y en a en réalité.

Ces problèmes de détection seront considérés dans la suite comme des erreurs, en particulier pour l'évaluation des méthodes parcimonieuses. Ainsi, le pourcentage d'erreurs représente le nombre de tirages moyen, pour 100 tirages aléatoires, où un algorithme détecte soit le mauvais nombre de sources, soit si l'une des directions détectée ne peut être associée à une direction d'arrivée réelle à moins d'une demi-largeur de lobe.

1.2.3.3 Pouvoir ou limite de résolution

Une mauvaise détection et donc la prise en compte d'une erreur peut également être due à un problème de résolution. Le pouvoir de résolution est la capacité d'une méthode à détecter et estimer les directions d'arrivées de deux sources proches. La forte corrélation spatiale pour deux directions d'arrivées situées dans le même lobe d'antenne entraîne en effet des difficultés de détection pour certains algorithmes, d'autant plus que le niveau de bruit est élevé.

La limite de résolution d'une méthode est ainsi définie comme le plus petit écart angulaire entre deux sources pour lequel on est sûr de pouvoir distinguer les deux directions d'arrivées, c'est-à-dire le plus petit écart angulaire pour lequel on aura 0% d'erreurs dans un contexte avec $M = 2$ sources.

1.2.3.4 Variance, biais, et REQM

Une direction d'arrivée estimée $\hat{\theta}$ est une variable aléatoire qui dépend en particulier du bruit de mesure. La variance est une mesure de la dispersion des valeurs de $\hat{\theta}$. Elle est définie par :

$$\text{Var}(\hat{\theta}) = \mathbb{E} \left[\left(\hat{\theta} - \mathbb{E}[\hat{\theta}] \right)^2 \right] = \mathbb{E}[\hat{\theta}^2] - \left(\mathbb{E}[\hat{\theta}] \right)^2. \quad (1.21)$$

Le biais est l'écart entre l'espérance d'une valeur estimée $\hat{\theta}$ et la valeur exacte du même paramètre :

$$\text{Biais}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta. \quad (1.22)$$

L'Erreur Quadratique Moyenne (EQM) caractérise la précision d'un estimateur. Elle est définie par :

$$\text{EQM}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta)^2\right] = \text{Var}(\hat{\theta}) + \text{Biais}(\hat{\theta})^2. \quad (1.23)$$

La Racine de l'Erreur Quadratique Moyenne (REQM) est particulièrement utilisée pour caractériser les performances d'une méthode :

$$\text{REQM}(\hat{\theta}) = \sqrt{\text{EQM}(\hat{\theta})}. \quad (1.24)$$

1.3 Méthodes conventionnelles de goniométrie

Les premières approches historiques de type interférométrie ou goniométrie d'amplitude ne seront pas développées dans ce document, car adaptées uniquement au cas mono-source.

1.3.1 Méthodes basées sur le filtrage spatial

Les méthodes basées sur un filtrage spatial des signaux reçus $\mathbf{x}(t)$ ont été développées dans un second temps. Le signal est d'abord filtré dans la direction ψ avec un vecteur de pondération $\mathbf{w}(\psi)$: on note $x(t, \psi) = \mathbf{w}^H(\psi)\mathbf{x}(t)$ le signal filtré. Ces méthodes mesurent la puissance rayonnée dans toutes les directions de la zone utile. L'objectif est donc de maximiser l'énergie $P(\psi)$ en sortie du filtre selon la direction ψ , donnée par :

$$P(\psi) = \mathbb{E}[|x(t, \psi)|^2] \quad (1.25)$$

$$= \mathbf{w}^H(\psi)\mathbf{R}_{xx}\mathbf{w}(\psi) \quad (1.26)$$

où $\mathbf{R}_{xx} = \mathbb{E}[\mathbf{x}(t)\mathbf{x}^H(t)]$ est la matrice de covariance des signaux $\mathbf{x}(t)$ reçus en sortie du réseau. L'estimateur des directions d'arrivées est donné par les valeurs de ψ maximisant la puissance $P(\psi)$ pour une expression du vecteur de pondération $\mathbf{w}(\psi)$ donnée.

Le filtre de la formation de voies est défini par :

$$\mathbf{w}_{FV}(\psi) = \frac{\mathbf{a}(\psi)}{\mathbf{a}^H(\psi)\mathbf{a}(\psi)}. \quad (1.27)$$

La fonction coût à maximiser est ainsi :

$$P_{FV}(\psi) = \frac{\mathbf{a}^H(\psi)\mathbf{R}_{xx}\mathbf{a}(\psi)}{\mathbf{a}^H(\psi)\mathbf{a}(\psi)}. \quad (1.28)$$

L'estimateur est non biaisé en mono-source. En revanche, sa résolution multi-sources est limitée par l'ouverture d'antenne.

Le filtre de Capon [15] correspond au filtre spatial qui maximise le rapport signal à bruit plus interférences. Le vecteur de pondération est le suivant :

$$\mathbf{w}_C(\psi) = \frac{\mathbf{R}_{xx}^{-1}\mathbf{a}(\psi)}{\mathbf{a}^H(\psi)\mathbf{R}_{xx}^{-1}\mathbf{a}(\psi)}. \quad (1.29)$$

Il est calculé sous la contrainte $\mathbf{w}^H(\psi)\mathbf{a}(\psi) = 1$. La puissance à maximiser, appelée critère de Capon, est alors donnée par :

$$P_C(\psi) = \frac{1}{\mathbf{a}^H(\psi)\mathbf{R}_{xx}^{-1}\mathbf{a}(\psi)}. \quad (1.30)$$

La méthode a une capacité de résolution supérieure à la formation de voies. Les directions d'arrivées de deux sources proches resteront néanmoins biaisées, et ce même lorsque la covariance \mathbf{R}_{xx} est parfaitement estimée et en l'absence d'erreurs de modèle.

Les inconvénients de ces méthodes conventionnelles sont ainsi leur pouvoir de résolution limité et le biais induit en contexte multi-sources.

1.3.2 Méthodes haute résolution (HR)

Les algorithmes de goniométrie à haute résolution ont été découverts à la fin des années 70 et exploitent finement la structure du modèle d'observation. A la différence des méthodes basées de la filtrage spatial, ces méthodes sont asymptotiquement non biaisées en multi-sources, c'est-à-dire le pouvoir de résolution est théoriquement infini pour un nombre d'échantillons K infini [74]. Nous présentons ici quelques-unes des méthodes HR qui serviront comme références dans nos simulations.

1.3.2.1 La méthode MUSIC (Multiple Signal Classification)

La méthode MUSIC (Multiple Signal Classification) [6, 7, 66] est la méthode de référence ayant introduit la notion de sous-espaces. On fait l'hypothèse que le bruit additif $\mathbf{n}(t)$ est un bruit blanc dont la covariance vérifie :

$$\mathbb{E} [\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma_n^2\mathbf{I}_N. \quad (1.31)$$

On suppose de plus que le bruit est indépendant du signal $\mathbf{s}(t)$. La matrice de covariance du signal $\mathbf{x}(t)$ vérifie alors :

$$\mathbf{R}_{xx} = \mathbb{E} [\mathbf{x}(t)\mathbf{x}^H(t)] \quad (1.32)$$

$$= \mathbf{A}_\Theta\mathbf{R}_{ss}\mathbf{A}_\Theta^H + \sigma_n^2\mathbf{I}_N, \quad (1.33)$$

où $\mathbf{R}_{ss} = \mathbb{E} [\mathbf{s}(t)\mathbf{s}^H(t)]$ est la matrice de covariance des sources. La matrice \mathbf{R}_{xx} se décompose en éléments propres de la façon suivante :

$$\mathbf{R}_{xx} = \mathbf{V}\mathbf{\Upsilon}\mathbf{V}^H \quad (1.34)$$

$$= \mathbf{V}_s \mathbf{\Upsilon}_s \mathbf{V}_s^H + \mathbf{V}_n \mathbf{\Upsilon}_n \mathbf{V}_n^H, \quad (1.35)$$

avec $\mathbf{\Upsilon} = \text{diag}(\mathbf{\Upsilon}_s \ \mathbf{\Upsilon}_n)$ la matrice diagonale contenant les valeurs propres et $\mathbf{V} = [\mathbf{V}_s \ \mathbf{V}_n]$ la matrice contenant les vecteurs propres. Les valeurs propres vérifient $\mathbf{\Upsilon}_s = \text{diag}(\mu_1 + \sigma_n^2, \dots, \mu_M + \sigma_n^2)$, avec (μ_1, \dots, μ_M) les valeurs propres de la matrice $\mathbf{A}_\Theta \mathbf{R}_{ss} \mathbf{A}_\Theta^H$ classées par ordre décroissantes avec $\mu_M > 0$, et $\mathbf{\Upsilon}_n = \sigma_n^2 \mathbf{I}_{N-M}$. Les sous-espaces propres \mathbf{V}_s et \mathbf{V}_n sont orthogonaux entre eux puisque \mathbf{R}_{xx} est hermitienne : les vecteurs directeurs $\mathbf{a}(\theta_m)$ engendrant le sous-espace signal \mathbf{V}_s sont orthogonaux à tous les vecteurs qui composent le sous-espace bruit \mathbf{V}_n . On a alors :

$$\mathbf{a}^H(\theta_m) \mathbf{V}_n \mathbf{V}_n^H \mathbf{a}(\theta_m) = 0, \forall m \in 1, \dots, M. \quad (1.36)$$

Cette propriété est exploitée dans la méthode MUSIC, où l'estimation des paramètres θ_m repose sur la minimisation de la fonction coût suivante :

$$J_{MUSIC}(\psi) = \frac{\mathbf{a}^H(\psi) \mathbf{\Pi}_n \mathbf{a}(\psi)}{\mathbf{a}^H(\psi) \mathbf{a}(\psi)}. \quad (1.37)$$

avec

$$\mathbf{\Pi}_n = \mathbf{V}_n \mathbf{V}_n^H. \quad (1.38)$$

Les paramètres d'intérêt θ_m correspondent en effet au M zéros de cette fonction coût selon l'équation (1.36).

Le projecteur bruit $\mathbf{\Pi}_n$ est en pratique estimé par :

$$\hat{\mathbf{\Pi}}_n = \hat{\mathbf{V}}_n \hat{\mathbf{V}}_n^H \quad (1.39)$$

où $\hat{\mathbf{V}}_n$ est formé des $N - M$ vecteurs propres associés au plus petites valeurs propres de la matrice de covariance estimée $\hat{\mathbf{R}}_{xx}$ (1.16). Les angles d'arrivées θ_m sont estimés par les M minimas proches de zéro du pseudo-spectre $J_{MUSIC}(\psi)$.

La méthode MUSIC est très simple à mettre en œuvre, et ses bonnes performances dans la plupart des scénarios en ont ainsi fait un algorithme très utilisé. Cependant, la méthode voit ses performances fortement dégradées pour des sources fortement corrélées ou en présence d'erreurs de modèle [29, 33, 73]. Pour des sources cohérentes, le sous-espace signal est de dimension 1 et l'utilisation de MUSIC se heurte à des obstacles majeurs. De plus, la méthode requiert l'utilisation d'un certain nombre d'échantillons K pour estimer la matrice de covariance, sans quoi les performances se dégradent rapidement. La méthode ne permet pas d'estimer plus de sources que de capteurs dans sa formulation initiale, c'est-à-dire avec l'utilisation de la matrice de covariance. Des méthodes dérivées telles que 2q-

MUSIC permettent d'estimer les directions d'arrivées de plus de sources que de capteurs avec l'utilisation des statistiques d'ordres supérieurs : cela nécessite alors d'augmenter encore le nombre d'échantillons K pour obtenir de bonnes performances.

1.3.2.2 Le maximum de vraisemblance déterministe

La méthode du maximum de vraisemblance exploite la statistique des signaux $\mathbf{x}(t_k)$ reçus pour en maximiser la fonction de vraisemblance. Dans notre cas, nous étudierons uniquement le maximum de vraisemblance déterministe (MVD), qui considère le modèle (1.10) avec des signaux déterministes inconnus et un bruit aléatoire gaussien centré circulaire. On suppose ainsi que $\mathbf{n}(t) = \mathbf{x}(t) - \mathbf{A}_{\Theta}\mathbf{s}(t)$ suit une loi gaussienne centrée de matrice de covariance $\sigma_n^2\mathbf{I}_N$. Le nombre de sources M est supposé connu : les inconnues sont les signaux $\mathbf{s}(t)$, les angles Θ et la covariance du bruit σ_n^2 . On considère K échantillons temporels tels que les signaux reçus $\mathbf{x}(t_k)$ soient indépendants. La vraisemblance de l'observation $\bar{\mathbf{x}} = [\mathbf{x}^T(t_1) \dots \mathbf{x}^T(t_K)]^T$ est alors donnée par le produit des densités de chacune des observations $\mathbf{x}(t_k)$:

$$l(\Phi_M, \bar{\mathbf{s}}, \sigma_n; \bar{\mathbf{x}}) = \prod_{k=1}^K \frac{1}{(\pi\sigma_n^2)} e^{-\frac{1}{\sigma_n^2}(\mathbf{x}(t_k) - \mathbf{A}_{\Phi_M}\mathbf{s}(t_k))^H(\mathbf{x}(t_k) - \mathbf{A}_{\Phi_M}\mathbf{s}(t_k))}, \quad (1.40)$$

avec $\bar{\mathbf{s}} = [\mathbf{s}^T(t_1) \dots \mathbf{s}^T(t_K)]^T$, et $\Phi_M \subset \Psi$ tel que $\text{card}(\Phi_M) = M$ le nombre de sources. \mathbf{A}_{Φ_M} est la sous-matrice de la table de calibration \mathbf{A} qui contient les M vecteurs correspondants à l'ensemble Φ_M . On en déduit l'opposée de la log-vraisemblance :

$$-L(\Phi_M, \bar{\mathbf{s}}, \sigma_n; \bar{\mathbf{x}}) = NK \ln(\pi\sigma_n^2) + \frac{1}{\sigma_n^2} \sum_{k=1}^K (\mathbf{x}(t_k) - \mathbf{A}_{\Phi_M}\mathbf{s}(t_k))^H(\mathbf{x}(t_k) - \mathbf{A}_{\Phi_M}\mathbf{s}(t_k)). \quad (1.41)$$

En maximisant la log-vraisemblance par rapport à $\bar{\mathbf{s}}$, on obtient le filtre spatial suivant :

$$\hat{\mathbf{s}}_{MV}(t) = \mathbf{A}_{\Phi_M}^{\#} \mathbf{x}(t), \quad (1.42)$$

avec $\mathbf{A}_{\Phi_M}^{\#} = (\mathbf{A}_{\Phi_M}^H \mathbf{A}_{\Phi_M})^{-1} \mathbf{A}_{\Phi_M}^H$ la pseudo-inverse de Moore-Penrose. De même, l'estimée au sens du maximum de vraisemblance de σ_n^2 est donnée par :

$$\hat{\sigma}_{nMV}^2 = \frac{1}{N} \text{tr} \left(\mathbf{\Pi}_n(\Phi_M) \hat{\mathbf{R}}_{xx} \right) \quad (1.43)$$

avec $\mathbf{\Pi}_n(\Phi_M) = \mathbf{I}_N - \mathbf{A}_{\Phi_M} \mathbf{A}_{\Phi_M}^{\#}$. On obtient finalement l'estimateur des directions d'arrivées au sens du maximum de vraisemblance déterministe :

$$\hat{\Theta}_{MV} = \arg \min_{\Phi_M} \sum_{k=1}^K \mathbf{x}^H(t_k) \mathbf{\Pi}_n(\Phi_M) \mathbf{x}(t_k) \quad (1.44)$$

$$= \arg \min_{\Phi_M} \text{tr} \left(\mathbf{\Pi}_n(\Phi_M) \hat{\mathbf{R}}_{xx} \right). \quad (1.45)$$

On note $J_{MV}(\Phi_M)$ le critère à minimiser :

$$J_{MV}(\Phi_M) = \text{tr} \left(\Pi_n(\Phi_M) \hat{\mathbf{R}}_{xx} \right). \quad (1.46)$$

L'optimisation du critère J_{MV} du maximum de vraisemblance (1.46) pour l'estimation des directions d'arrivées est multi-dimensionnelle : on estime conjointement M paramètres. L'estimation est extrêmement coûteuse en temps de calcul dès lors que l'on dépasse $M = 2$ sources en 1D, ou à partir de $M = 2$ sources en 2D.

Le maximum de vraisemblance permet, contrairement à MUSIC, d'estimer les directions d'arrivées de sources cohérentes [73].

1.3.3 Performances des méthodes HR et borne de Cramer-Rao

La borne de Cramer Rao (CRB pour *Cramer Rao Bound*) exprime, pour un modèle statistique donné, une borne inférieure sur la variance d'un estimateur sans biais. Il s'agit d'un outil très employé pour évaluer les performances d'un estimateur en raison de sa simplicité de calcul. La borne de Cramer-Rao est basée sur l'information de Fisher.

L'inverse de l'information de Fisher $\mathcal{I}(\Theta)$ d'un paramètre Θ correspond à la borne de Cramer-Rao :

$$\text{Var} \left(\hat{\Theta} \right) \geq \mathcal{I}(\Theta)^{-1} = \mathbb{E} \left[\frac{\partial \ln l(\Theta, \bar{\mathbf{x}})}{\partial \Theta} \frac{\partial \ln l(\Theta, \bar{\mathbf{x}})}{\partial \Theta}^T \right]^{-1}. \quad (1.47)$$

Dans la suite, on écrira $\text{CRB}(\Theta) = \mathcal{I}(\Theta)^{-1}$. Un estimateur qui atteint la borne de Cramer-Rao est dit efficace.

Sans erreurs de modèle, il a été démontré que :

- L'estimée du maximum de vraisemblance déterministe est asymptotiquement Gaussienne mais non efficace lorsque K tend vers l'infini [72].
- L'estimée du maximum de vraisemblance déterministe est asymptotiquement Gaussienne et efficace lorsque le RSB tend vers l'infini [65].
- MUSIC n'est pas asymptotiquement efficace dans le cas général (matrice de covariance des sources quelconque) [72].
- MUSIC est asymptotiquement équivalent au maximum de vraisemblance déterministe si et seulement si la matrice de covariance des sources est diagonale, c'est-à-dire lorsque les sources sont décorrélatées entre elles [72].

Pour des temps d'intégration fini (BT fini), la borne de Cramer-Rao sert usuellement de référence pour l'évaluation des méthodes d'estimation. En présence d'erreurs de modèles, on utilisera les performances théoriques du maximum de vraisemblance déterministe en présence d'erreurs de modèles telles qu'établies dans l'article [31].

1.4 Motivations

Dans ce chapitre, nous avons décrit les méthodes MUSIC et du maximum de vraisemblance déterministe. Leurs performances asymptotiques en font des méthodes de référence pour la goniométrie.

Elles ont cependant chacune des défauts qui les rendent non utilisables dans les contextes opérationnels qui nous intéressent. La formulation classique de la méthode MUSIC présentée dans ce chapitre ne permet pas d'estimer les directions d'arrivées de plus de sources que de capteurs. De plus, elle ne fonctionne pas pour l'estimation des directions d'arrivées de sources cohérentes. Les méthodes du maximum de vraisemblance peuvent fonctionner dans ce cas, mais leur coût calculatoire est prohibitif dans un système opérationnel actuel.

Les méthodes parcimonieuses présentent un fort potentiel pour l'estimation des directions d'arrivées dans ces scénarios, et le formalisme parcimonieux est en totale adéquation avec l'utilisation, dans nos systèmes opérationnels, d'une table de calibration. Nous les présentons dans le prochain chapitre.

Chapitre 2

État de l'art des méthodes parcimonieuses en goniométrie

Le signal reçu par le réseau d'antennes est une combinaison linéaire des vecteurs directeurs $\mathbf{a}(\theta_m)$ correspondants aux directions d'arrivées θ_m des sources (1.11). Dans un système de goniométrie opérationnel, nous utilisons la connaissance d'une table de calibration \mathbf{A} qui contient les vecteurs directeurs d'un ensemble Ψ contenant les directions d'arrivées possibles. Ce cadre correspond donc tout à fait à celui d'un problème d'estimation parcimonieuse.

2.1 Parcimonie et norme ℓ_0 : généralités

Nous introduisons ici les notions de parcimonie et les notations utiles dans ce cadre.

Soit \mathbb{I}_G l'ensemble des indices ordonnés de 1 à G :

$$\mathbb{I}_G = \{1, \dots, G\}. \quad (2.1)$$

Soit ω un ensemble tel que $\omega \subseteq \mathbb{I}_G$. Le complément de ω dans \mathbb{I}_G est donné par :

$$\omega^c = \mathbb{I}_G \setminus \omega. \quad (2.2)$$

Il vérifie également $\omega^c \subseteq \mathbb{I}_G$. Pour un ensemble ω , on note

$$\#\omega = \text{card } \omega \quad (2.3)$$

le nombre d'éléments de l'ensemble ω .

On considère un vecteur $\mathbf{u} \in \mathbb{C}^G$, tel que $\mathbf{u} = [u_1, \dots, u_G]^T$. Ce vecteur est dit parcimonieux s'il contient un grand nombre de composantes nulles, et seulement quelques

composantes non nulles. Le support de \mathbf{u} est défini par

$$\text{supp}(\mathbf{u}) = \{1 \leq g \leq G \mid u_g \neq 0\}. \quad (2.4)$$

On dira que ce vecteur possède un degré de parcimonie M si seulement M de ses coefficients sont non nuls, avec $M \ll G$. La cardinalité du support de \mathbf{u} est alors M . On introduit ainsi la norme ℓ_0 , qui compte le nombre de composantes non nulles du vecteur :

$$\|\mathbf{u}\|_0 = \text{card}(\text{supp}(\mathbf{u})) = \#\text{supp}(\mathbf{u}). \quad (2.5)$$

Cette norme est la mesure naturelle ou intuitive de la parcimonie d'un vecteur. On notera que la fonction ℓ_0 n'est pas une norme par définition, puisqu'elle ne vérifie pas la propriété d'homogénéité : $\|\alpha\mathbf{u}\|_0 \neq |\alpha|\|\mathbf{u}\|_0$ dans le cas général, notamment pour tout α différent de 0,1 ou -1 . L'appellation norme ℓ_0 est donc un abus de langage que nous conservons également en concordance avec la littérature sur le sujet.

Certains signaux ne sont pas directement parcimonieux mais peuvent être exprimés sous forme parcimonieuse dans un espace transformé. On considère une observation $\mathbf{y} \in \mathbb{C}^{N'}$, et on suppose que celle-ci peut être modélisée comme combinaison linéaire parcimonieuse d'un dictionnaire donné $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_G] \in \mathbb{C}^{N' \times G}$ par un vecteur $\mathbf{u} \in \mathbb{C}^G$ tel que $\text{supp}(\mathbf{u}) = \omega$, avec $\omega \subseteq \mathbb{I}_G$ et $\#\omega = M$ le nombre d'éléments de l'ensemble ω :

$$\mathbf{y} = \mathbf{D}\mathbf{u} = \sum_{g=1}^G \mathbf{d}_g u_g = \sum_{m=1}^M \mathbf{d}_{\omega_m} u_{\omega_m}, \quad (2.6)$$

On a alors $\|\mathbf{u}\|_0 = M$. On note \mathbf{D}_ω et \mathbf{u}_ω la sous-matrice et le sous-vecteur définis par :

$$\mathbf{D}_\omega = [\mathbf{d}_{\omega_1}, \dots, \mathbf{d}_{\omega_{\#\omega}}] \in \mathbb{C}^{N' \times \#\omega} \quad (2.7)$$

et

$$\mathbf{u}_\omega = [u_{\omega_1}, \dots, u_{\omega_{\#\omega}}] \in \mathbb{C}^{\#\omega}. \quad (2.8)$$

L'équation (2.6) se réécrit :

$$\mathbf{y} = \mathbf{D}\mathbf{u} = \mathbf{D}_\omega \mathbf{u}_\omega. \quad (2.9)$$

On écrira alors $\mathbf{u} = C_\omega(\mathbf{u}_\omega)$ avec C_ω l'opérateur de complétion de zéros défini par :

$$\mathbf{u} = C_\omega(\mathbf{u}_\omega) \Leftrightarrow \forall g \in \mathbb{I}_G, u_g = \begin{cases} 0 & \text{si } g \notin \omega. \\ u_{\omega_m} & \text{pour l'unique } g \mid \omega_m = g \end{cases} \quad (2.10)$$

Dans le cadre de l'estimation parcimonieuse, la matrice \mathbf{D} est appelée dictionnaire, et ses colonnes sont appelées atomes. Dans cette thèse, on considère des dictionnaires sur-complets, ou "saturés", c'est-à-dire qui vérifient $N' \ll G$. Dans la pratique, le modèle contient un bruit additif \mathbf{w} : $\mathbf{y} = \mathbf{D}\mathbf{u} + \mathbf{w}$.

Le but de l'estimation parcimonieuse est l'estimation du vecteur \mathbf{u} à partir d'une observation bruitée \mathbf{y} : il s'agit ici d'un problème sous-déterminé qui n'a pas de solution unique. Les méthodes parcimonieuses visent à utiliser l'a priori de parcimonie du vecteur \mathbf{u} .

2.2 Représentation parcimonieuse du signal en goniométrie

Dans le cadre de l'estimation parcimonieuse en goniométrie, on suppose qu'un dictionnaire $\mathbf{D} \in \mathbb{C}^{N' \times G}$ se rapporte à l'ensemble des angles de la grille Ψ dont est issue la table de calibration \mathbf{A} , avec $\#\Psi = G$:

$$\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_G] = [\mathbf{d}(\psi_1), \dots, \mathbf{d}(\psi_G)]. \quad (2.11)$$

Pour tout ensemble $\Phi = \{\phi_1, \dots, \phi_s\}$ d'angles tel que $\#\Phi = s$ et $\Phi \subseteq \Psi$, on note \mathbf{D}_Φ la sous-matrice de \mathbf{D} contenant les s colonnes correspondantes aux angles dans Φ . De même, \mathbf{u}_Φ est le sous-vecteur de \mathbf{u} contenant uniquement les s composantes u_i dont les indices i vérifient : $\exists k \mid \psi_i = \phi_k$.

Soit $\mathbf{u} \in \mathbb{C}^G$ parcimonieux tel que $\text{supp}(\mathbf{u}) = \omega \subset \mathbb{I}_G$ et $\#\omega = s$. Soit $\Phi = \Psi_\omega = \{\psi_{\omega_1}, \dots, \psi_{\omega_s}\} \subset \Psi$. Dans ce cas, la définition de l'opérateur de complétion de zéros (2.10) est étendue aux ensembles d'angles dans Ψ :

$$\mathbf{u} = C_\omega(\mathbf{u}_\omega) = C_\Phi(\mathbf{u}_\Phi). \quad (2.12)$$

2.2.1 Modèle SMV du signal reçu

On considère un réseau de N antennes avec M sources incidentes. Dans l'hypothèse où les M angles d'arrivées Θ sont inclus dans l'ensemble Ψ des directions d'arrivées candidates sur la grille de dimension G , avec $M \ll G$, la matrice de mélange \mathbf{A}_Θ est une sous-matrice de la table de calibration \mathbf{A} . On a alors

$$\mathbf{A}_\Theta \mathbf{s}(t) = \mathbf{A} \bar{\mathbf{s}}(t) \quad (2.13)$$

où $\bar{\mathbf{s}}(t) = C_\Theta(\mathbf{s}(t))$. Cette relation est illustrée dans la figure 2.1. Le signal $\mathbf{x}(t)$ reçu par le réseau d'antennes devient alors :

$$\mathbf{x}(t) = \mathbf{A} \bar{\mathbf{s}}(t) + \mathbf{n}(t). \quad (2.14)$$

On notera également $\bar{\mathbf{s}}_\Theta = \mathbf{s}$, où $\bar{\mathbf{s}}_\Theta$ est le sous-vecteur de $\bar{\mathbf{s}}$ contenant seulement les composantes dont les indices g vérifient : $\exists m \mid \psi_g = \theta_m$.

Le signal reçu $\mathbf{x}(t)$ s'exprime comme combinaison linéaire parcimonieuse de la table de calibration \mathbf{A} connue, qui correspond donc à un dictionnaire. Les vecteurs directeurs des angles candidats $\mathbf{a}(\psi)$ correspondent donc à des atomes. Pour M sources, on aura

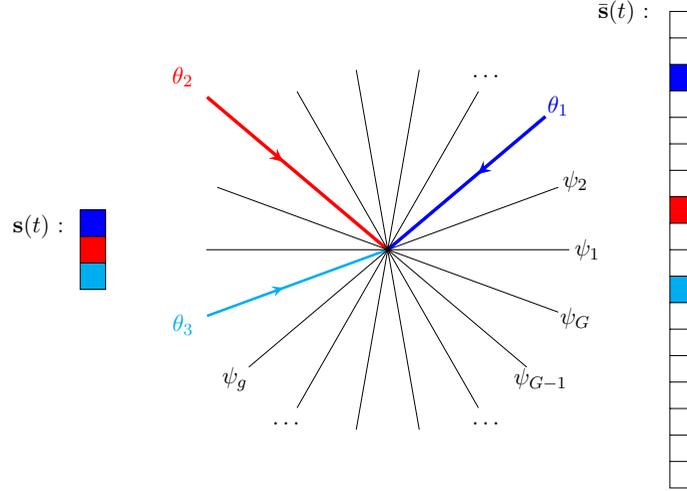


FIGURE 2.1 – Représentation des vecteurs $\mathbf{s}(t)$ et $\bar{\mathbf{s}}(t)$ (équation (2.13)), et de l'opérateur de complétion de zéros $\bar{\mathbf{s}}(t) = C_{\Theta}(\mathbf{s}(t))$, pour $M = 3$ sources.

$\|\bar{\mathbf{s}}(t)\|_0 = M$. Le but de l'estimation parcimonieuse est d'estimer le vecteur $\bar{\mathbf{s}}(t)$ à partir du signal reçu $\mathbf{x}(t)$. En particulier, en goniométrie parcimonieuse, seuls les indices des composantes non nulles de $\bar{\mathbf{s}}(t)$ nous suffisent pour estimer les directions d'arrivées θ_m : on aura en effet

$$\forall g \mid \bar{s}_g(t) \neq 0, \exists m \mid \theta_m = \psi_g. \quad (2.15)$$

L'équation (2.14) correspond, en estimation parcimonieuse, à un modèle SMV (pour *Single Measurement Vector*). Les méthodes d'estimation parcimonieuse peuvent alors s'appliquer directement sur le signal $\mathbf{x}(t)$ en sortie du réseau d'antennes à un unique instant t : c'est un des avantages mis en avant dans la littérature sur l'application des méthodes pour la goniométrie [4, 80].

Notons que l'estimation parcimonieuse de $\bar{\mathbf{s}}(t)$ effectuée directement à partir du signal reçu $\mathbf{x}(t)$ permet de faire également de la séparation de source : ce n'est cependant pas le propos de cette thèse.

2.2.2 Modèle MMV des signaux reçus

Afin de prendre en compte la variabilité des signaux au cours du temps, on considère plusieurs instants de mesure. Il est possible de considérer chaque instant de mesure séparément et de résoudre ainsi K problèmes d'estimation correspondants à du SMV pour K instants de mesure. Il faut ensuite combiner les résultats. Cependant, il n'y a ainsi pas de coopération entre les différents problèmes, et cette approche peut être très sensible au RSB. Il est ainsi plus intéressant de considérer K échantillons du signal $\mathbf{x}(t)$ simultanément pour la résolution du problème.

Pour K instants de mesure, la matrice des signaux reçus \mathbf{X} est donnée par :

$$\mathbf{X} = [\mathbf{x}(t_1), \dots, \mathbf{x}(t_K)] = \mathbf{A}\bar{\mathbf{S}} + \mathbf{N}, \quad (2.16)$$

avec $\bar{\mathbf{S}} = [\bar{\mathbf{s}}(t_1), \dots, \bar{\mathbf{s}}(t_K)]$ et $\mathbf{N} = [\mathbf{n}(t_1), \dots, \mathbf{n}(t_K)]$. On est alors dans le cadre d'un modèle MMV, pour *Multi Measurement Vectors*. Dans ce cas, la matrice $\bar{\mathbf{S}} \in \mathbb{C}^{G \times K}$ des inconnues est parcimonieuse par ligne. La parcimonie par ligne se mesure par la norme $\ell_{2,0}$, définie par :

$$\|\mathbf{S}\|_{2,0} = \text{card} \{g \mid \|\mathbf{S}_{g,:}\|_2 \neq 0\}. \quad (2.17)$$

L'estimation de $\bar{\mathbf{S}}$, et plus particulièrement les indices des lignes dont la norme est non nulle, permet donc également d'estimer les angles d'arrivées θ_m . Cependant, la dimension de cette matrice inconnue est bien supérieure à la dimension d'un unique vecteur de dimension G la taille de la grille de recherche, ce qui implique un coût calculatoire bien plus important. De plus, il est inutile pour l'estimation des directions d'arrivées d'estimer exactement les composantes des signaux dans $\bar{\mathbf{S}}$: la finalité de l'estimation, dans le cadre de la goniométrie, concerne les angles θ_m correspondants à des lignes dont les composantes sont non nulles. C'est pourquoi il a pu être proposé des méthodes de réduction de la dimension qui reposent sur une transformation des signaux reçus.

2.2.3 Réduction de la dimension pour la goniométrie parcimonieuse

Afin de traiter plusieurs instants de mesure de façon conjointe sans accroître significativement le coût de résolution du problème d'estimation des directions d'arrivées, Malioutov et al. [48] ont présenté une technique de réduction de la dimension basée sur une décomposition en éléments propres. L'avantage de cette technique est que la dimension ne croit pas avec le nombre d'instant de mesure. Elle est applicable pour tous types de signaux : signaux à moyenne nulle ou non nulle, sources corrélées temporellement ou non.

On considère la matrice des signaux reçus $\mathbf{X} = \mathbf{A}\bar{\mathbf{S}} + \mathbf{N}$. La décomposition en éléments propres de \mathbf{X} s'écrit :

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H. \quad (2.18)$$

Les vecteurs singuliers correspondants aux M plus grandes valeurs singulières correspondent à l'espace signal plus bruit, tandis que les autres sont uniquement dus au bruit. Il est ainsi proposé d'utiliser comme observation :

$$\mathbf{Y}_{SV} = \mathbf{U}\mathbf{A}\mathbf{D}_M = \mathbf{X}\mathbf{V}\mathbf{D}_M, \quad (2.19)$$

où $\mathbf{D}_M = [\mathbf{I}_M \ \mathbf{0}]^T$, avec \mathbf{I}_M la matrice identité de dimension $M \times M$ et $\mathbf{0}$ une matrice de dimension $M \times (K - M)$ contenant des zéros. $\mathbf{Y}_{SV} \in \mathbb{C}^{N \times M}$ est une matrice d'observation dont la dimension ne dépend pas du nombre d'échantillons. Soit $\bar{\mathbf{S}}_{SV} = \bar{\mathbf{S}}\mathbf{V}\mathbf{D}_M$ et $\mathbf{N}_{SV} =$

\mathbf{NVD}_M . Alors on obtient :

$$\mathbf{Y}_{SV} = \mathbf{A}\bar{\mathbf{S}}_{SV} + \mathbf{N}_{SV}. \quad (2.20)$$

La matrice inconnue $\bar{\mathbf{S}}_{SV} \in \mathbb{C}^{G \times M}$ est parcimonieuse par ligne. La dimension a ainsi été réduite de K à M : si le nombre d'instantants de mesure K peut parfois être de plusieurs centaines, le nombre de sources M est en général faible (de une à une dizaine dans nos applications, pour des réseaux composés de 4 à 8 antennes). La réduction de dimension est donc relativement importante. Cette technique a ainsi largement été reprise dans le domaine de l'estimation des directions d'arrivées [47, 68, 86]. Ce modèle correspond cependant toujours à un modèle MMV.

Or il existe encore de nombreux outils d'estimation parcimonieuse qui n'ont pas été étendus pour des modèles MMV et ne s'appliquent donc qu'à des modèles SMV. C'est pourquoi on fera dans cette thèse le choix de se rapporter à des modèles SMV tout en utilisant plusieurs instants de mesure. On utilise pour cela des transformations bien choisies de la matrice de covariance \mathbf{R}_{xx} , par exemple la vectorisation de la matrice de covariance, plus récemment utilisée [21, 42, 63]. Les transformations choisies dans cette thèse sont adaptées au scénario et font l'objet d'un chapitre entier de cette thèse : nous ne les détaillons donc pas ici.

On considère par la suite le modèle général suivant :

$$\mathbf{y} = \mathbf{D}\mathbf{u} + \mathbf{w}, \quad (2.21)$$

où \mathbf{y} est une observation vectorielle issue de la matrice de covariance \mathbf{R}_{xx} , \mathbf{D} est un dictionnaire connu calculé à partir de la table de calibration \mathbf{A} pour les G angles candidats de la grille dans l'ensemble Ψ , \mathbf{u} est le vecteur parcimonieux de dimension G à estimer pour retrouver les directions d'arrivées Θ , et \mathbf{w} est un bruit additif.

2.3 Énoncés des problèmes parcimonieux : problèmes contraints et régularisés par la norme ℓ_0

Le but des méthodes parcimonieuses est d'estimer le vecteur $\mathbf{u} \in \mathbb{C}^G$ à partir d'une observation $\mathbf{y} \in \mathbb{C}^{N'}$, connaissant le dictionnaire $\mathbf{D} \in \mathbb{C}^{N' \times G}$. Le problème est mal posé : il est sous-déterminé puisque $N' < G$. C'est pourquoi il est nécessaire d'utiliser le fait que le vecteur \mathbf{u} à estimer est parcimonieux. Les problèmes parcimonieux ont une formulation et des méthodes propres.

Pour un problème inverse classique d'estimation du vecteur \mathbf{u} correspondant à l'observation \mathbf{y} sous hypothèse de bruit blanc gaussien, on minimise un critère des moindres carrés. Dans le cas où l'on sait que le vecteur \mathbf{u} à estimer possède un degré de parcimonie au maximum égal à s , on ajoute une contrainte de parcimonie au problème des moindres

carrés :

$$\min_{\mathbf{u}} \|\mathbf{y} - \mathbf{D}\mathbf{u}\|_2^2 \quad \text{s.c.} \quad \|\mathbf{u}\|_0 \leq s. \quad (\mathcal{C}_s)$$

On l'appellera le problème contraint par la norme ℓ_0 . Dans certains contextes, le degré de parcimonie n'est pas connu, mais l'on a une connaissance portant sur la variance du bruit. Le problème contraint est alors mis sous la forme suivante :

$$\min_{\mathbf{u}} \|\mathbf{u}\|_0 \quad \text{s.c.} \quad \|\mathbf{y} - \mathbf{D}\mathbf{u}\|_2^2 \leq \epsilon. \quad (\mathcal{C}_\epsilon)$$

Dans le cas où l'on n'a aucune connaissance sur la variance du bruit ou la parcimonie, ou que l'on ne souhaite privilégier ni l'un ni l'autre, le problème peut également être formulé sous forme régularisée :

$$\min_{\mathbf{u}} \left\{ J_{\ell_0}(\lambda, \mathbf{u}) = \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{u}\|_2^2 + \lambda \|\mathbf{u}\|_0 \right\}, \quad (\mathcal{R}_\lambda)$$

où $J_{\ell_0}(\lambda, \cdot)$ est le critère régularisé par la norme ℓ_0 , et λ est le paramètre de régularisation qui permet d'équilibrer le poids relatif entre l'attache aux données (le terme des moindres carrés $1/2\|\mathbf{y} - \mathbf{D}\mathbf{u}\|_2^2$) et la pénalité qui favorise la parcimonie (la norme ℓ_0). Si le paramètre de régularisation λ est nul ou proche de zéro, on se ramène à un problème des moindres carrés et la solution obtenue n'est pas parcimonieuse. Si le paramètre de régularisation λ est élevé, les solutions obtenues seront très parcimonieuses : si λ est trop élevé, la solution peut être le vecteur nul. A noter que cette formulation, d'un point de vue Bayésien, peut être obtenue comme un maximum a posteriori où l'attache aux données correspond à une distribution de bruit gaussienne, et la distribution a priori favorisant les signaux parcimonieux est la limite d'une distribution Bernouilli-Gaussienne avec une variance gaussienne infinie.

2.4 Résolution des problèmes parcimonieux

Les problèmes ℓ_0 sous forme contrainte ou régularisée sont non convexes et NP-difficiles [53] : il n'existe pas d'algorithme optimal permettant de résoudre ces problèmes en un temps polynomial, sauf dans des situations spécifiques. Un algorithme d'optimisation optimal ou une résolution optimale s'entend comme une méthode qui permet d'obtenir le minimum global. La résolution optimale nécessite en effet une recherche exhaustive, ce qui est inenvisageable dans un système opérationnel en raison du coût calculatoire, même pour une dimension modérée. Les algorithmes utilisés seront ainsi dits sous-optimaux : ils vont permettre de chercher des solutions proches de la solution optimale, mais rien en général ne garantit qu'ils trouveront nécessairement le minimum global. Les algorithmes se divisent en plusieurs catégories. L'état de l'art proposé ici ne regroupe pas de façon exhaustive tous les algorithmes proposés dans la littérature, mais vise à en donner un bon aperçu.

2.4.1 Les algorithmes gloutons

Une première famille regroupe les algorithmes gloutons, développés pour résoudre les problèmes contraints (\mathcal{C}_s) ou (\mathcal{C}_ϵ) . Les algorithmes gloutons sont des méthodes itératives qui mettent à jour à chaque itération le support en ajoutant une nouvelle composante non nulle. Partant généralement d'un vecteur estimé $\hat{\mathbf{u}}^{(0)}$ nul, on ajoute à chaque itération i un coefficient non nul sur l'une des composantes de $\hat{\mathbf{u}}^{(i)}$ afin de diminuer le résidu $\|\mathbf{y} - \mathbf{D}\hat{\mathbf{u}}^{(i)}\|_2^2$, jusqu'à atteindre le degré de parcimonie s fixé pour (\mathcal{C}_s) , ou jusqu'à atteindre le niveau maximum de bruit ϵ désiré pour (\mathcal{C}_ϵ) . C'est le principe de *Matching Pursuit* (MP) [49], ensuite amélioré dans les algorithmes *Orthogonal Matching Pursuit* (OMP) [60] [79], *Orthogonal Least Square* (OLS) [16] ou CoSaMP [54]. Une revue des algorithmes gloutons est donnée dans [77]. Appliquées à la goniométrie, ces méthodes permettent d'obtenir de bons résultats en terme de précision pour une estimation en mono-source [8]. Cependant, la simplicité de ces algorithmes en fait aussi une faiblesse : une mauvaise détection d'un élément dans le support ne peut jamais être compensée ensuite par sa suppression. En particulier, la condition ERC (pour *Exact Recovery Condition*) [78] qui donne la garantie d'obtenir la solution exacte avec l'algorithme OMP n'est pas respectée dès lors que l'on est dans un cas multi-sources. En effet, il a été montré que $\mu(\mathbf{D}) < \frac{1}{2s-1}$ est une condition suffisante pour estimer exactement un vecteur parcimonieux de degré de parcimonie s dans le cas non bruité, avec $\mu(\mathbf{D})$ la cohérence mutuelle (1.15) du dictionnaire \mathbf{D} . Pour $M = 2$ sources, il faudrait ainsi une cohérence mutuelle inférieure à $1/3$: nous avons vu précédemment que la cohérence mutuelle de la table de calibration est généralement proche de 1, en particulier pour des réseaux de faible ouverture, si l'on veut être assez précis dans l'estimation angulaire. Bien que les transformations de la table de calibration proposées dans le chapitre suivant permettent de diminuer la cohérence mutuelle, cela ne sera toujours pas suffisant pour appliquer ces méthodes et conserver une précision acceptable dans le cas multi-sources.

Des algorithmes gloutons bidirectionnels comme *Single Best Replacement* (SBR) [26] ont par la suite été développés : l'algorithme SBR autorise à chaque itération l'ajout ou la suppression d'un élément du support afin de faire diminuer le critère J_{ℓ_0} du problème régularisé (\mathcal{R}_λ) . L'algorithme est cependant très coûteux en temps de calcul, puisqu'une itération de l'algorithme requiert la résolution de G problèmes des moindres carrés, G étant la taille de la grille de discrétisation.

2.4.2 Les algorithmes de descente à gradient-proximal

Les algorithmes de descente de gradient sont des algorithmes itératifs très utilisés dans le domaine de l'optimisation pour la minimisation de fonctions différentiables. Or la norme ℓ_0 utilisée comme régularisation parcimonieuse dans le critère J_{ℓ_0} est une fonction non-différentiable. Il est nécessaire d'introduire une nouvelle famille d'algorithme de descente pour les fonctions non dérivables et parfois même non continues.

2.4.2.1 Introduction générale sur les algorithmes de descente à gradient-proximal

Nous introduisons ici la notion de proximal et détaillons les algorithmes de descente à gradient-proximal développés pour la minimisation de fonctions coûts non dérivables. Nous détaillons et rappelons en premier lieu plusieurs définitions utiles pour la compréhension des notions introduites.

Soit f une fonction différentiable. On rappelle que

$$f \text{ est convexe} \Leftrightarrow \forall(\mathbf{u}, \mathbf{u}'), \quad f(\mathbf{u}') \geq f(\mathbf{u}) + \langle \nabla f(\mathbf{u}), \mathbf{u}' - \mathbf{u} \rangle. \quad (2.22)$$

Cela signifie que la tangente est sous la courbe en tous points.

Définition 2.1 (Fonction à gradient lipschitzien). *Soit f une fonction convexe et différentiable. On dit que f est une fonction à gradient lipschitzien de constante de Lipschitz L si :*

$$\forall(\mathbf{u}, \mathbf{u}') \quad |\nabla f(\mathbf{u}) - \nabla f(\mathbf{u}')| \leq L\|\mathbf{u} - \mathbf{u}'\|, \quad (2.23)$$

avec L un réel positif.

On considère maintenant une fonction f continue et convexe mais potentiellement non différentiable en tout point. On étend alors la notion de gradient pour ces points à l'aide de la sous-différentielle.

Définition 2.2 (Sous-différentielle). *Soit f une fonction convexe. La sous-différentielle de f , notée ∂f , est définie par :*

$$\partial f(\mathbf{u}) = \{ \mathbf{z} \in \mathbb{C}^G \mid \forall \mathbf{u}' \in \mathbb{C}^G, \quad f(\mathbf{u}') \geq f(\mathbf{u}) + \langle \mathbf{z}, \mathbf{u}' - \mathbf{u} \rangle \}. \quad (2.24)$$

Dans le cas non différentiable, $\partial f(\mathbf{u})$ est l'ensemble des pentes des droites passant par $(\mathbf{u}, f(\mathbf{u}))$ situées sous la courbe. Par exemple, prenons la fonction valeur absolue $f(u) = |u|$: elle n'est pas dérivable en 0. La sous-différentielle de f en 0 est l'ensemble $[-1, 1]$: toutes les droites de pentes entre -1 et 1 passant par le point de coordonnées $(0, f(0) = 0)$ sont situées sous la courbe $(u, f(u))$.

Si f est différentiable en \mathbf{u} , alors $\partial f(\mathbf{u}) = \nabla f(\mathbf{u})$.

Définition 2.3 (Fonction semi-continue inférieure). *Soit f une fonction de \mathbb{R}^G dans \mathbb{R} . f est semi-continue inférieure en \mathbf{u} si*

$$f(\mathbf{u}) \leq \liminf_{\mathbf{u}' \rightarrow \mathbf{u}} f(\mathbf{u}'). \quad (2.25)$$

Une fonction semi-continue inférieure en tout points est dite fonction semi-continue inférieure.

Définition 2.4 (Fonction propre). *Une fonction f est propre si elle ne prend pas la valeur $-\infty$ et n'est pas identiquement égale à $+\infty$.*

Théorème 2.1 (Condition nécessaire et suffisante pour définir un minimiseur dans le cas convexe non différentiable). *Soit f une fonction convexe, semi-continue inférieure et propre sur \mathbb{R}^G . Alors :*

$$\hat{\mathbf{u}} = \arg \min_{\mathbf{u} \in \mathbb{R}^G} f(\mathbf{u}) \Leftrightarrow 0 \in \partial f(\hat{\mathbf{u}}). \quad (2.26)$$

L'objectif est de trouver $\hat{\mathbf{u}}$ tel que $0 \in \partial f(\hat{\mathbf{u}})$. On peut appliquer une méthode itérative de sous-gradient : partant d'une initialisation $\hat{\mathbf{u}}^{(0)}$, on effectue les itérations $\hat{\mathbf{u}}^{(i+1)} = \hat{\mathbf{u}}^{(i)} - \beta_i \mathbf{g}_i$, avec $\mathbf{g}_i \in \partial f(\hat{\mathbf{u}}^{(i)})$. La version implicite de l'algorithme est :

$$\hat{\mathbf{u}}^{(i+1)} = \hat{\mathbf{u}}^{(i)} - \beta_i \mathbf{g}_{i+1} \text{ avec } \mathbf{g}_{i+1} \in \partial f(\hat{\mathbf{u}}^{(i+1)}). \quad (2.27)$$

On cherche $\hat{\mathbf{u}}^{(i+1)}$ tel que

$$\hat{\mathbf{u}}^{(i)} \in \hat{\mathbf{u}}^{(i+1)} + \beta_k \partial f(\hat{\mathbf{u}}^{(i+1)}) \quad (2.28)$$

$$\hat{\mathbf{u}}^{(i)} \in (I + \beta_k \partial f)(\hat{\mathbf{u}}^{(i+1)}) \quad (2.29)$$

$$\hat{\mathbf{u}}^{(i+1)} \in (I + \beta_k \partial f)^{-1}(\hat{\mathbf{u}}^{(i)}) = \text{prox}_{\beta_k f}(\hat{\mathbf{u}}^{(i)}) \quad (2.30)$$

Définition 2.5 (Opérateur proximal). *Soit f une fonction de \mathbb{R}^G dans \mathbb{R} , propre, convexe et semi-continue inférieure. On définit :*

$$\text{prox}_f(\mathbf{u}) = \arg \min_{\mathbf{u}' \in \mathbb{R}^G} \left\{ \frac{1}{2} \|\mathbf{u}' - \mathbf{u}\|_2^2 + f(\mathbf{u}') \right\}. \quad (2.31)$$

Le minimum est atteint en $\hat{\mathbf{u}}'$ tel que

$$0 \in \partial f(\hat{\mathbf{u}}') + \hat{\mathbf{u}}' - \mathbf{u} \quad (2.32)$$

$$\mathbf{u} \in \hat{\mathbf{u}}' + \partial f(\hat{\mathbf{u}}') \quad (2.33)$$

$$\hat{\mathbf{u}}' \in (I + \partial f)^{-1}(\mathbf{u}) \quad (2.34)$$

Les éléments de $\text{prox}_f(\mathbf{u})$ sont les éléments de $(I + \partial f)^{-1}(\mathbf{u})$.

Théorème 2.2 (Condition nécessaire et suffisante pour définir un minimiseur dans le cas convexe non différentiable). *Soit f une fonction convexe, semi-continue inférieure et propre sur \mathbb{R}^G . Alors :*

$$\hat{\mathbf{u}} = \arg \min_{\mathbf{u} \in \mathbb{R}^G} f(\mathbf{u}) \Leftrightarrow \hat{\mathbf{u}} = \text{prox}_f(\hat{\mathbf{u}}) \quad (2.35)$$

Autrement dit, un minimiseur de f est un point fixe de prox_f . Un algorithme proximal cherche de façon itérative une solution à $\hat{\mathbf{u}} = \text{prox}_f(\hat{\mathbf{u}})$.

Algorithme 2.1 Algorithme proximal

- Initialisation $\hat{\mathbf{u}}^{(0)}$
 - Itérations jusqu'à convergence : $\hat{\mathbf{u}}^{(i+1)} = \text{prox}_f(\hat{\mathbf{u}}^{(i)})$
-

Un algorithme proximal correspond donc à un algorithme de gradient pour les fonctions non différentiables.

Pour les besoins de l'optimisation non convexe, la notion de gradient (ou sous-différentielle) a été généralisée aux fonctions non-convexes par passage à la limite, dont la définition est disponible dans [3, 19]. On note ∂f le gradient généralisé (ou sous-différentielle limite) d'une fonction f non convexe.

Théorème 2.3 (Condition nécessaire pour les extrema locaux dans le cas non convexe non différentiable). f atteint un minimum ou maximum local en $\mathbf{u} \Rightarrow 0 \in \partial f(\mathbf{u})$

Les points $\mathbf{u} \in \mathbb{R}^G$ qui vérifient $0 \in \partial f(\mathbf{u})$ sont appelés points critiques de f .

L'opérateur proximal défini à l'équation (2.31) est étendu aux cas des fonctions f propres et semi-continues inférieures non nécessairement convexes. Dans ce cas, si f vérifie l'inégalité de Kurdyka-Lojasiewicz (KL) et que la restriction de f à son domaine est continue, alors l'algorithme proximal 2.1 converge vers un point critique de f [11, 58].

On considère à présent la minimisation de fonctions coûts composées de deux termes :

$$J(\mathbf{u}) = f(\mathbf{u}) + g(\mathbf{u}) \quad (2.36)$$

avec f différentiable à gradient lipschitzien de constante L , qui vérifie l'inégalité KL, et g semi-continue inférieure. L'algorithme à gradient-proximal, ou algorithme *Forward-Backward Splitting* (FBS) est donné par :

Algorithme 2.2 Algorithme à gradient-proximal

Entrée : $\beta_k \in]0, \frac{2}{L}[$

- Initialisation $\hat{\mathbf{u}}^{(0)}$
 - Itérations jusqu'à convergence : $\mathbf{u}^{(i+1)} \in \text{prox}_{\beta_k g}(\mathbf{u}^{(i)} - \beta_k \nabla f(\mathbf{u}^{(i)}))$
-

L'algorithme à gradient-proximal converge vers un point critique de J [3].

2.4.2.2 Application à la goniométrie parcimonieuse : algorithmes de seuillage dur pour la norme ℓ_0

Dans le cadre de la parcimonie, le premier terme $f(\mathbf{u})$ est un terme d'attache aux données, et le terme $g(\mathbf{u})$ un terme de pénalité qui impose ou favorise la parcimonie. Dans cette thèse, le terme d'attache aux données est exclusivement le résidu des moindres carrés $f(\mathbf{u}) = 1/2 \|\mathbf{y} - \mathbf{D}\mathbf{u}\|_2^2$. Le terme de pénalité est soit :

- un terme de régularisation qui favorise la parcimonie ; dans ce cas, on notera $g(\mathbf{u}) = \Phi_{pen}(\lambda, \mathbf{u})$. Le terme de régularisation est usuellement et a minima non dérivable en
-

zéro afin de favoriser la parcimonie [28]. Dans ce cas, on note

$$J_{pen}(\lambda, \mathbf{u}) = \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{u}\|_2^2 + \Phi_{pen}(\lambda, \mathbf{u}), \quad (2.37)$$

la fonction coût à minimiser. Dans le cas de la minimisation du critère régularisé par la norme ℓ_0 (\mathcal{R}_λ), on a ainsi :

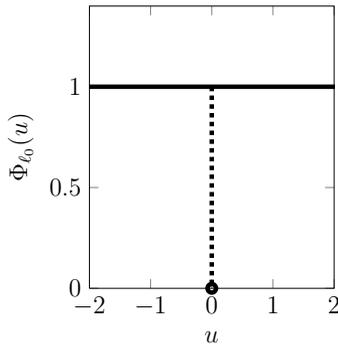
$$\Phi_{\ell_0}(\lambda, \mathbf{u}) = \lambda \|\mathbf{u}\|_0; \quad (2.38)$$

La pénalité ℓ_0 est représentée sur la figure 2.2(a).

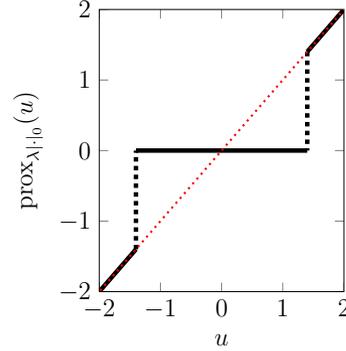
- une fonction qui impose une contrainte sur la solution. En effet, la contrainte de parcimonie $\|\mathbf{u}\|_0 \leq s$ dans le problème contraint (\mathcal{C}_s) correspond à une fonction de pénalité $\mathbb{1}_{\{\|\mathbf{u}\|_0 \leq s\}}$ définie comme suit :

$$\mathbb{1}_{\{\|\mathbf{u}\|_0 \leq s\}}(\mathbf{u}) = \begin{cases} 0 & \text{si } \|\mathbf{u}\|_0 \leq s \\ +\infty & \text{sinon} \end{cases} \quad (2.39)$$

Dans ces cas particuliers de problèmes contraints, l'algorithme à gradient-proximal est appelé algorithme à gradient projeté (PGD pour *Projected Gradient Descent*). L'opérateur proximal correspond en effet dans ce cas à une projection.



(a) Pénalité Φ_{ℓ_0} , $\lambda = 1$.



(b) Opérateur proximal $\text{prox}_{\lambda \|\cdot\|_0}$, $\lambda = 1$.

FIGURE 2.2 – Pénalité Φ_{ℓ_0} et opérateur proximal associé, pour $\lambda = 1$. Sur la figure (b), la droite pointillée rouge représente la droite $y = x$.

L'opérateur proximal de la norme ℓ_0 correspond à la fonction de seuillage dur (*Hard Threshold*). Le proximal de Φ_{ℓ_0} se calcule en effet composante par composante par :

$$\left[\text{prox}_{\lambda \|\cdot\|_0} \right]_g(\mathbf{u}) = u_g \mathbb{1}_{\{|u_g| \geq \sqrt{2\lambda}\}}, \quad (2.40)$$

avec $\mathbb{1}$ la fonction indicatrice :

$$\mathbb{1}_{\text{condition}} = \begin{cases} 1 & \text{si la condition est respectée} \\ 0 & \text{sinon} \end{cases} \quad (2.41)$$

Les plus petites composantes sont ainsi mises à zéro, tandis que les autres sont conservées telles qu'elles. Le proximal est ainsi discontinu. Il est représenté sur la figure 2.2(b).

L'algorithme à gradient-proximal pour la minimisation de J_{ℓ_0} (résolution de (\mathcal{R}_λ)) a été introduit par Blumensath et Davies [9, 10] sous le nom de *Iterative Hard Thresholding* (IHT).

Algorithme 2.3 Algorithme IHT

Entrée : Dictionnaire \mathbf{D} , observation \mathbf{y} , $\beta \leq 1/L$, $L = \|\mathbf{D}^H \mathbf{D}\|$

Initialisation : $\mathbf{u}^{(0)}$, $i = 0$

tant que critères d'arrêts non respectés **faire**

• $i = i + 1$

• $\mathbf{u}^{(i)} = \text{prox}_{\lambda\beta\|\cdot\|_0}(\mathbf{u}^{(i-1)} - \beta\mathbf{D}^H(\mathbf{D}\mathbf{u}^{(i-1)} - \mathbf{y}))$

fin tant que

En raison de la non convexité de la norme ℓ_0 , il n'y a aucune garantie que l'algorithme converge vers le minimum global du critère J_{ℓ_0} . De plus, le critère J_{ℓ_0} présente de nombreux minima locaux. Les résultats obtenus avec l'algorithme IHT dépendent donc fortement de l'initialisation.

La deuxième version de l'algorithme IHT est dédiée à la résolution du problème contraint (\mathcal{C}_s) . L'opérateur proximal associé à la fonction $\mathbb{1}_{\{\|\mathbf{u}\|_0 \leq M\}}$ correspond à la projection dans l'ensemble $\{\mathbf{u} \in \mathbb{C}^G \mid \|\mathbf{u}\|_0 \leq M\}$. L'ensemble n'étant pas convexe, la solution de l'opérateur proximal n'est pas unique. L'étape de projection dans l'algorithme IHT dédié à la minimisation du problème contraint ℓ_0 tel que décrit par Blumensath et Davies [9] consiste alors à ne garder que les M plus grandes composantes. Son application en goniométrie multi-sources ne peut alors fonctionner à cause de la trop grande corrélation entre colonnes proches du dictionnaire.

2.4.3 Résolution passant par des relaxations ou approximations de la norme ℓ_0

Devant les difficultés de résolution posées par les problèmes ℓ_0 , il a été proposé de remplacer la norme ℓ_0 par d'autres fonctions favorisant la parcimonie.

2.4.3.1 Relaxation convexe par la norme ℓ_1 : programmation linéaire ou quadratique et algorithmes de descente à seuillage doux

Une approche très commune pour pallier les difficultés de résolution des problèmes parcimonieux ℓ_0 combinatoires est de remplacer la norme ℓ_0 par sa relaxation convexe la norme ℓ_1 . Pour un vecteur $\mathbf{u} \in \mathbb{C}^G$, elle est définie par :

$$\|\mathbf{u}\|_1 = \sum_{g=1}^G |u_g| \quad (2.42)$$

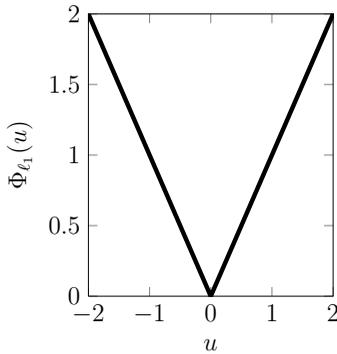
Bien que non différentiable en zéro, l'utilisation de la norme ℓ_1 permet de tirer parti de la littérature bien fournie en optimisation convexe.

Les méthodes *Basis Pursuit* (BP) ou *Basis Pursuit De-Noising* (BPDN) [17] permettent de résoudre les relaxations ℓ_1 des problèmes (\mathcal{C}_ϵ) et (\mathcal{R}_λ) en utilisant des algorithmes de programmation linéaire ou quadratique. Plusieurs solveurs ou bibliothèques tels que CVX [37] ou SeDuMi [75] sont disponibles pour résoudre de tels problèmes et leur utilisation est répandue en goniométrie [34, 48, 81, 84]. Le coût de calcul avec ces solveurs est cependant très important.

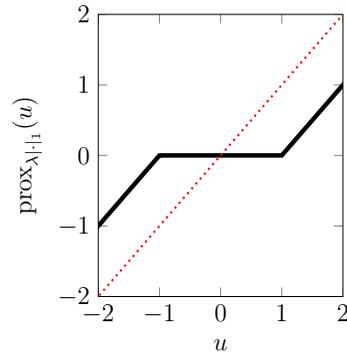
Les algorithmes itératifs de descente pour la minimisation du problème sous forme régularisée sont généralement moins coûteux. L'algorithme à gradient-proximal pour le critère régularisé par la norme ℓ_1 est appelé Iterative Soft Thresholding Algorithm (ISTA) [22]. On parle de seuillage doux, car le proximal de la régularisation $\Phi_{\ell_1}(\lambda, \mathbf{u}) = \lambda \|\mathbf{u}\|_1$ est donné composante par composante par :

$$\left[\text{prox}_{\lambda \|\cdot\|_1} \right]_g(\mathbf{u}) = \max \left\{ 0, 1 - \frac{\lambda}{|u_g|} \right\} u_g \mathbb{1}_{\{|u_g| \neq 0\}}. \quad (2.43)$$

Les plus petites composantes sont mises à zéro, tandis que les composantes supérieures à λ en valeur absolue sont biaisées d'un même facteur λ , ce qui assure la continuité du seuillage. La pénalité ℓ_1 est représentée sur la figure 2.3(a) tandis que son opérateur proximal est représenté sur la figure 2.3(b).



(a) Pénalité Φ_{ℓ_1} , $\lambda = 1$.



(b) Opérateur proximal $\text{prox}_{\lambda \|\cdot\|_1}$, $\lambda = 1$.

FIGURE 2.3 – Pénalité Φ_{ℓ_1} et opérateur proximal associé, pour $\lambda = 1$. Sur la figure (b), la droite pointillée rouge représente la droite $y = x$.

Il existe une version accélérée de l'algorithme ISTA, nommé FISTA (pour *Fast ISTA*) [5].

Les algorithmes ISTA et FISTA convergent vers le minimum de J_{ℓ_1} puisque la norme ℓ_1 est convexe. Ce sont donc des algorithmes très efficaces, d'autant plus qu'ils sont simples et peu coûteux ; ils sont en particulier utilisés en traitement d'image [20].

Algorithme 2.4 Algorithme FISTA

Entrée : Dictionnaire \mathbf{D} , observation \mathbf{y} , $\beta = 1/L$, $L = \|\mathbf{D}^*\mathbf{D}\|$
Initialisation : $\mathbf{u}^{(0)}$, $\mathbf{u}_{temp} = \mathbf{u}^{(0)}$, $T^{(0)} = 1$, $i = 0$
tant que critères d'arrêts non respectés **faire**

- $i = i + 1$

- $\mathbf{u}^{(i)} = \text{prox}_{\lambda\beta\|\cdot\|_1}(\mathbf{u}_{temp} - \beta\mathbf{D}^H(\mathbf{D}\mathbf{u}_{temp} - \mathbf{y}))$

- $T^{(i)} = \frac{1 + \sqrt{1 + (2T^{(i-1)})^2}}{2}$

- $\mathbf{u}_{temp} = \mathbf{u}^{(i)} + \frac{T^{(i-1)} - 1}{T^{(i)}}(\mathbf{u}^{(i)} - \mathbf{u}^{(i-1)})$

fin tant que

La facilité d'optimisation des critères avec la norme ℓ_1 ne doit cependant pas nous faire perdre de vue l'objectif principal qui est la minimisation des problèmes ℓ_0 (\mathcal{C}_s) ou (\mathcal{R}_λ). Or la norme ℓ_1 est bien différente de la norme ℓ_0 initiale. La norme ℓ_1 n'est d'ailleurs pas bornée, contrairement à la norme ℓ_0 , ce qui pénalise davantage les coefficients élevés. Sous certaines conditions suffisantes portant sur le dictionnaire \mathbf{D} , établies par plusieurs auteurs [12, 13, 23, 24, 78] sous le nom de conditions ERC ou RIP, des signaux suffisamment parcimonieux peuvent être estimés de façon exacte avec une minimisation ℓ_1 . Les conditions RIP ne sont en pratique souvent pas vérifiables, et on ne peut être certain de les remplir que si le dictionnaire \mathbf{D} est aléatoire : ce n'est évidemment pas le cas en goniométrie. La condition ERC, identique à celle de l'algorithme glouton OMP, n'est pas vérifiée. Les conditions sont en fait toutes liées à une cohérence mutuelle faible du dictionnaire : nous avons vu que la cohérence mutuelle d'une table de calibration est très facilement proche de 1 lorsque l'on considère un pas de grille suffisant pour l'estimation des directions d'arrivées avec une certaine précision. C'est pourquoi nous ne retenons pas les méthodes ℓ_1 dans le cadre de cette thèse.

2.4.3.2 Approximations non convexes de la norme ℓ_0 et algorithmes adaptés

De nombreuses approximations de la norme ℓ_0 ont été proposées dans la littérature afin, d'une part, de se rapprocher davantage de la régularisation ℓ_0 qu'avec la régularisation ℓ_1 , et d'autre part, de pallier les difficultés d'optimisation des problèmes ℓ_0 . Ces fonctions sont en général non convexes et non différentiables mais continues [67], et il a été montré dans de nombreuses applications qu'elles pouvaient permettre d'obtenir des performances bien meilleures qu'avec la norme ℓ_1 [82].

On peut ainsi mentionner les normes ℓ_p , pour $p \in]0; 1[$ [32, 64, 76, 83], ou la pénalité *Log-Sum* [14]. Les pénalités bornées comme *Capped- ℓ_1* [62], la pénalité *Smoothly Clipped Absolute Deviation* (SCAD) [28] ou *Minimax Concave Penalty* (MCP) [85] permettent d'approcher encore davantage la norme ℓ_0 . La pénalité *Smoothed L0* (SL0) [44, 51] est une des seules approximations différentiables proposée : il a cependant été montré que la non-différentiabilité d'une pénalité en 0 est ce qui permet de favoriser la parcimonie [28].

La pénalité *Continuous Exact ℓ_0* (CEL0) [69], correspondant à la limite d'une fonction

MCP, est particulièrement intéressante pour des modèles SMV puisqu'il a été démontré que le minimum global du critère régularisé J_{CEL0} coïncide avec le minimum global de J_{ℓ_0} , et que la fonction J_{CEL0} présente moins de minima locaux que J_{ℓ_0} . La fonction J_{CEL0} est définie par :

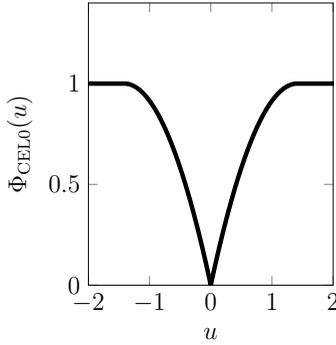
$$J_{\text{CEL0}}(\lambda, \mathbf{u}) = \frac{1}{2} \|\mathbf{D}\mathbf{u} - \mathbf{y}\|_2^2 + \Phi_{\text{CEL0}}(\lambda, \mathbf{u}), \quad (2.44)$$

avec

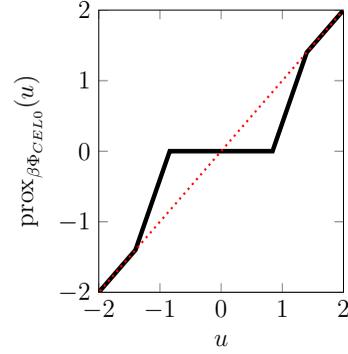
$$\Phi_{\text{CEL0}}(\lambda, \mathbf{u}) = \sum_{i \in \mathbb{I}_G} \phi_{\text{CEL0}}(\lambda, \alpha_i, u_i) \quad (2.45)$$

$$\phi_{\text{CEL0}}(\lambda, \alpha_i, u_i) = \lambda - \frac{(|u_i| - \sqrt{2\lambda\alpha_i})^2}{2\alpha_i} \mathbb{1}_{\{|u_i| \leq \sqrt{2\lambda\alpha_i}\}} \quad (2.46)$$

et $\alpha_i = 1/\|\mathbf{d}_i\|_2^2$, \mathbf{d}_i étant la i -ème colonne de \mathbf{D} . La pénalité Φ_{CEL0} pour un scalaire est représentée sur la figure 2.4(a).



(a) Pénalité Φ_{CEL0} , $\lambda = 1$.



(b) Opérateur proximal $\text{prox}_{\beta\Phi_{\text{CEL0}}}$, $\lambda = 1, \beta = 0.6$.

FIGURE 2.4 – Pénalité Φ_{CEL0} et opérateur proximal associé, pour $\lambda = 1$. Sur la figure (b), la droite pointillée rouge représente la droite $y = x$.

L'algorithme à gradient-proximal 2.2 peut être utilisé pour la minimisation d'un critère régularisé par la plupart de ces pénalités dès lors que le proximal a une expression simple. L'optimisation est dans ce cas sous-optimale puisque la convergence vers le minimum global n'est pas garantie. Pour la pénalité CEL0, on a :

$$\left[\text{prox}_{\beta\Phi_{\text{CEL0}}(\lambda, \cdot)} \right]_g(\mathbf{u}) = \begin{cases} 0 & \text{si } u_g = 0 \\ \frac{u_g}{|u_g|} \min \left(|u_g|, \frac{(|u_g| - \sqrt{2\lambda\beta})_+}{1-\beta} \right) & \text{sinon} \end{cases} \quad (2.47)$$

avec $(\cdot)_+ = \max(0, \cdot)$. Cet opérateur proximal est représenté sur la figure 2.4(b). L'algorithme à gradient-proximal 2.2 utilisant la pénalité CEL0 dont le proximal est donné par l'équation (2.47) est nommé FB-CEL0. Il a été utilisé avec succès pour la résolution de problèmes de goniométrie où il présente des meilleures performances que MUSIC pour des RSB faibles, et également de meilleures performances que IHT [18].

Un schéma plus élaboré d'algorithmes itératifs peut également être utilisé, à partir de fonctions convexes localement majorantes. En partant d'une initialisation, on calcule une surface convexe localement majorante à la surface J_{pen} à minimiser en ce point. On réalise plusieurs itérations de descente sur cette surface avec des algorithmes à gradient-proximal, avant de recalculer une nouvelle surface majorante et de répéter l'opération jusqu'à convergence. Ces algorithmes de majorisation-minimisation spécifiquement appliqués aux fonctions coûts contenant deux termes, dont un convexe et le second non convexe, sont appelés algorithmes à pondérations itératives. En particulier, pour une fonction J_{pen} , on définit l'algorithme *Iterative Reweighted ℓ_1* (IRL1) [59].

Algorithme 2.5 Algorithme IRL1 pour la minimisation de J_{pen}

Entrée : Dictionnaire \mathbf{D} , observation \mathbf{y} , $\beta \leq 1/L$, $L = \|\mathbf{D}^*\mathbf{D}\|$

Initialisation : $\mathbf{u}^{(0)}$, $i = 0$

tant que critères d'arrêts non respectés **faire**

Calcul du vecteur de pondération \mathbf{w} en fonction de $\mathbf{u}^{(i-1)}$

tant que critères d'arrêts non respectés **faire**

• Itérations de descente sur le critère $J_{\lambda\|\cdot\|_{\mathbf{w}\ell_1}} = \frac{1}{2}\|\mathbf{y} - \mathbf{D}\mathbf{u}\|_2^2 + \lambda\|\mathbf{u}\|_{\mathbf{w}\ell_1}$

• $i = i + 1$

fin tant que

fin tant que

Les itérations de descente peuvent être réalisées par un algorithme à gradient-proximal simple, ou sa version accélérée comme dans l'algorithme FISTA 2.4. $\|\cdot\|_{\mathbf{w}\ell_1}$ est la norme ℓ_1 -pondérée, définie par :

$$\|\mathbf{u}\|_{\mathbf{w}\ell_1} = \sum_{g=1}^G w_g |u_g| \quad (2.48)$$

dont le proximal se calcule par :

$$\left[\text{prox}_{\lambda\|\cdot\|_{\mathbf{w}\ell_1}} \right]_g (\mathbf{u}) = \max \left\{ 0, 1 - \frac{\lambda w_g}{|u_g|} \right\} u_g \mathbb{1}_{\{|u_g| \neq 0\}}. \quad (2.49)$$

Pour la pénalité CEL0, le vecteur de pondération \mathbf{w} est donné composante par composante par [36] :

$$w_g(u_g) = \left(\sqrt{2\lambda} - |u_g| \right) \mathbb{1}_{\{|u_g| \leq \sqrt{2\lambda}\}}. \quad (2.50)$$

L'algorithme IRL1 utilisant la pondération (2.50) visant à majorer itérativement la surface J_{CEL0} pour la minimisation du critère sera noté IRL1-CEL0. La convergence avec un algorithme de type IRL1 n'est garantie que vers un point stationnaire de J_{pen} .

Enfin des algorithmes optimaux ont été proposés très récemment [50] pour résoudre le problème avec une approximation continue non convexe de la norme ℓ_0 comme les fonctions Capped- ℓ_p , SCAD, MCP ou CEL0. La méthode a cependant encore un coût calculatoire trop élevé et ne permet donc pas de traiter le problème d'estimation des directions d'arrivées.

2.5 Conclusion

Nous avons montré que la formulation du problème parcimonieux et le choix d'une méthode de résolution dépendent de l'application considérée. Dans le cas d'une goniométrie parcimonieuse dans des contextes opérationnels multi-sources et en présence de peu de capteurs, nous avons notamment vu que la cohérence mutuelle du dictionnaire ne permet pas d'utiliser les méthodes gloutonnes et de relaxation par la norme ℓ_1 .

Ainsi, notre attention porte dans la suite sur la norme ℓ_0 et sur les algorithmes de descente à gradient-proximal, qui présentent l'avantage d'être peu coûteux en temps de calcul. Leur inconvénient majeur est toutefois l'absence de garantie de convergence globale. En particulier avec l'utilisation de la norme ℓ_0 , le critère J_{ℓ_0} présente de nombreux minima locaux et les performances de l'algorithme IHT dépendent donc fortement de l'initialisation. Le critère J_{CEL0} régularisé par la pénalité continue non-convexe CEL0 possède les mêmes minima globaux que le critère J_{ℓ_0} et présente un moindre nombre de minima locaux. La pénalité CEL0 semble alors très prometteuse, et son utilisation avec les algorithmes FB-CEL0 et IRL1-CEL0 sera donc étudiée et exploitée par la suite.

Chapitre 3

Modèles d'observation SMV parcimonieux

3.1 Enjeux

Comme nous l'avons vu précédemment, le signal reçu par le réseau d'antennes $\mathbf{x}(t)$ est naturellement parcimonieux quand on considère la table de calibration \mathbf{A} , sous l'hypothèse que les directions d'arrivées Θ vérifient $\Theta \subset \Psi$ l'ensemble des directions d'arrivées candidates de la grille. Il est cependant indispensable d'utiliser en goniométrie plusieurs instants de mesure afin de prendre en compte la variabilité des signaux au cours du temps et d'améliorer ainsi les performances d'estimation. L'utilisation directe de la matrice des signaux reçus \mathbf{X} (2.16) entraîne un coût calculatoire conséquent qu'il n'est pas nécessaire de supporter lorsque seuls les angles d'arrivées nous intéressent. De plus, de nombreuses méthodes parcimonieuses sont d'abord développées pour des modèles SMV et n'ont pas nécessairement d'équivalent - ou du moins les mêmes propriétés - pour des modèles MMV.

L'objectif est donc de construire un modèle SMV parcimonieux à partir des signaux reçus sur K instants de mesure indépendants : $\mathbf{X} = [\mathbf{x}(t_1), \dots, \mathbf{x}(t_K)]$. Le modèle construit dépendra de l'application opérationnelle visée. Le modèle d'observation parcimonieux obtenu est de la forme $\mathbf{y} = \mathbf{D}\mathbf{u} + \mathbf{w}$ (2.21). Le vecteur $\mathbf{y} \in \mathbb{C}^{N'}$ est l'observation calculée à partir des signaux reçus, $\mathbf{D} \in \mathbb{C}^{N' \times G}$ est un dictionnaire connu issu de la table de calibration \mathbf{A} dont chaque colonne \mathbf{d}_g correspond à une direction d'arrivée candidate $\psi_g \in \Psi$, le vecteur $\mathbf{u} \in \mathbb{C}^G$ est le vecteur parcimonieux à estimer, et $\mathbf{w} \in \mathbb{C}^{N'}$ est un vecteur de bruit additif.

L'observation \mathbf{y} doit conserver toute l'information utile pour la goniométrie malgré la réduction de la dimension. Cette observation doit donc s'adapter aux contextes opérationnels particuliers qui nous intéressent.

3.2 Modèle SMV pour le contexte aéroporté

On considère un réseau d'antennes embarqué dans un avion ou un drone, comme représenté sur la figure 3.1. La charge utile est alors limitée, et le réseau est contraint à la fois sur le nombre d'antennes et sur la dimension du réseau (ouverture). Dans ce contexte, le nombre de signaux incidents peut être très important : le nombre d'émetteurs visibles augmente avec l'altitude du porteur. De plus, les angles d'arrivées de deux sources peuvent alors être proches. Pour un scénario multi-sources avec un réseau de petite ouverture et un nombre d'antennes réduit, nous savons que les méthodes HR offrent une faible résolution et une faible robustesse aux ambiguïtés. De plus, elles ne permettent pas de localiser plus de sources que de capteurs à partir des statistiques d'ordre 2.

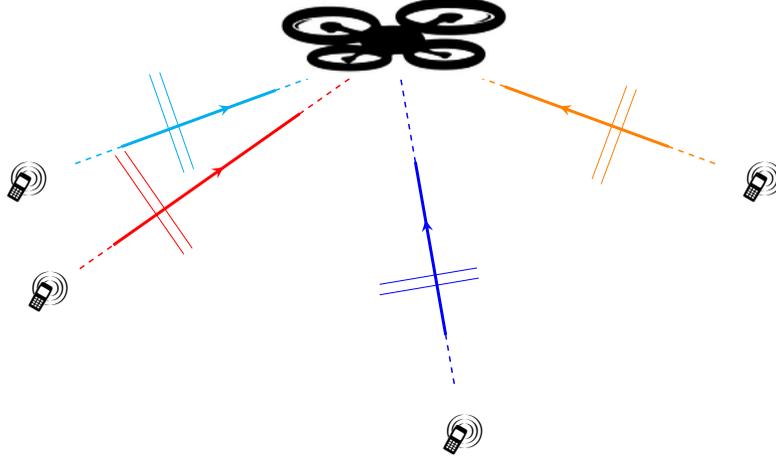


FIGURE 3.1 – Scénario aéroporté.

3.2.1 Modèle du signal reçu

Dans un contexte aéroporté, les signaux reçus sont indépendants : il n'y a généralement que des propagations en trajets directs. On parlera d'émetteurs LOS, pour *Line Of Sight*. On note $s_m(t)$ les signaux émis par les M émetteurs LOS. Pour des angles d'arrivée $\Theta = \{\theta_1, \dots, \theta_M\}$, le signal $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^T$ reçu en sortie du réseau de N antennes est alors :

$$\mathbf{x}(t) = \sum_{m=1}^M \mathbf{a}(\theta_m) s_m(t) + \mathbf{n}(t), \quad (3.1)$$

où $\mathbb{E}[s_m(t)s_{m'}^*(t)] = 0$ pour $m' \neq m$. On suppose de plus que $\mathbf{n}(t)$ est un bruit blanc indépendant du signal dont la covariance vérifie :

$$\mathbb{E}[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma_n^2 \mathbf{I}_N. \quad (3.2)$$

La matrice de covariance \mathbf{R}_{xx} (1.32) s'écrit selon (1.33) avec, dans le contexte de sources indépendantes,

$$\mathbf{R}_{ss} = \mathbb{E} [\mathbf{s}(t)\mathbf{s}^H(t)], \quad (3.3)$$

$$= \text{diag}(\gamma_{\theta_1}, \dots, \gamma_{\theta_M}) \quad (3.4)$$

avec, pour $m = 1 \dots M$, $\gamma_{\theta_m} = \mathbb{E} [|s_m(t)|^2]$ la puissance de la m -ième source de direction d'arrivée θ_m . On a alors :

$$\mathbf{R}_{xx} = \sum_{m=1}^M \mathbf{a}(\theta_m)\mathbf{a}^H(\theta_m)\gamma_{\theta_m} + \sigma_n^2\mathbf{I}_N. \quad (3.5)$$

3.2.2 Observation SMV dans le cas aéroporté : la matrice de covariance vectorisée

De manière à augmenter virtuellement le nombre d'antennes du réseau pour accroître la capacité de goniométrie en nombre de sources, on utilise la matrice de covariance vectorisée. Il s'agit des colonnes de la matrice \mathbf{R}_{xx} mises bout à bout dans une seule colonne. Pour une matrice \mathbf{R} , la vectorisation est définie par :

$$\text{vec}(\mathbf{R}) = [\mathbf{R}_{:,1}^T \ \mathbf{R}_{:,2}^T \ \dots \ \mathbf{R}_{:,N}^T]^T, \quad (3.6)$$

avec $\mathbf{R}_{:,i}$ la i -ème colonne de la matrice \mathbf{R} .

La matrice de covariance vectorisée $\mathbf{r} = \text{vec}(\mathbf{R}_{xx})$ est alors de la forme

$$\mathbf{r} = \sum_{m=1}^M \mathbf{b}(\theta_m)\gamma_{\theta_m} + \sigma_n^2\text{vec}(\mathbf{I}_N) \quad (3.7)$$

où

$$\mathbf{b}(\theta) = \mathbf{a}^*(\theta) \otimes \mathbf{a}(\theta), \quad (3.8)$$

\mathbf{a}^* est le conjugué de \mathbf{a} , et \otimes est le produit de Kronecker. On note $\mathbf{B}_{\Theta} = [\mathbf{b}(\theta_1), \dots, \mathbf{b}(\theta_M)]$ et $\boldsymbol{\gamma}_{\Theta} = [\gamma_{\theta_1}, \dots, \gamma_{\theta_M}]$. On a alors :

$$\mathbf{r} = \mathbf{B}_{\Theta}\boldsymbol{\gamma}_{\Theta} + \sigma_n^2\text{vec}(\mathbf{I}_N). \quad (3.9)$$

Le but est de construire une observation parcimonieuse à partir de cette matrice de covariance vectorisée. On choisit de soustraire le bruit moyen à la matrice de covariance vectorisée supposée exacte et sans erreurs de modèles :

$$\mathbf{y} = \mathbf{r} - \sigma_n^2\text{vec}(\mathbf{I}_N) = \mathbf{B}_{\Theta}\boldsymbol{\gamma}_{\Theta}. \quad (3.10)$$

Soit $\mathbf{B} \in \mathbf{C}^{N^2 \times G}$ un dictionnaire calculé comme le produit de Khatri-Rao par colonnes de

la table de calibration \mathbf{A} conjuguée avec elle-même :

$$\mathbf{B} = [\mathbf{b}(\psi_1), \dots, \mathbf{b}(\psi_G)] = \mathbf{A}^* \odot \mathbf{A}. \quad (3.11)$$

Sous l'hypothèse $\Theta \subset \Psi$, la matrice \mathbf{B}_Θ est une sous-matrice du dictionnaire \mathbf{B} . Soit $\gamma = C_\Theta(\gamma_\Theta)$. Alors :

$$\mathbf{B}_\Theta \gamma_\Theta = \mathbf{B} \gamma \quad (3.12)$$

et l'observation choisie s'écrit :

$$\mathbf{y} = \mathbf{r} - \sigma_n^2 \text{vec}(\mathbf{I}_N) = \mathbf{B} \gamma. \quad (3.13)$$

L'observation \mathbf{y} est de dimension N^2 , et le vecteur $\gamma \in \mathbb{R}^G$ est parcimonieux, de degré de parcimonie M , et les composantes non nulles correspondent aux puissances des sources γ_{θ_m} . Cette observation a déjà été utilisée avec succès en goniométrie parcimonieuse [42, 63].

L'augmentation de la dimension de l'observation \mathbf{y} par rapport à la dimension initiale N du signal reçu $\mathbf{x}(t)$, N correspondant au nombre de capteurs, permet théoriquement d'estimer davantage de sources que de capteurs. Utiliser la matrice de covariance vectorisée revient en effet à considérer un réseau virtuel qui contient jusqu'à $N^2 - N + 1$ capteurs différents. On représente sur la figure 3.2 un réseau circulaire uniforme à $N = 7$ antennes et les capteurs virtuels obtenus avec la vectorisation de la matrice de covariance.

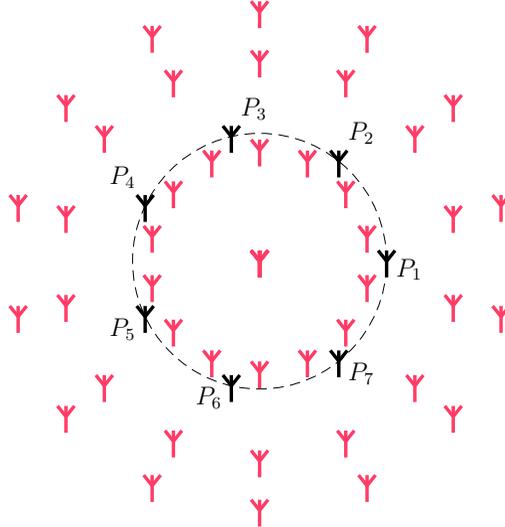


FIGURE 3.2 – Réseau circulaire uniforme à $N = 7$ antennes notées P_n , pour $N = \{1, \dots, N\}$, et les capteurs virtuels obtenus par la vectorisation de la matrice de covariance.

En prenant en compte les sources d'incertitudes telles que décrites dans la section 1.2.3.1, c'est-à-dire soit le nombre d'échantillons K fini, soit les erreurs de modèle, on

obtient dans la pratique une observation bruitée :

$$\mathbf{y} = \mathbf{B}\boldsymbol{\gamma} + \mathbf{w}. \quad (3.14)$$

L'observation est donc bien de la forme $\mathbf{y} = \mathbf{D}\mathbf{u} + \mathbf{w}$, avec $\mathbf{D} = \mathbf{B}$ calculé à partir de la table de calibration et $\mathbf{u} = \boldsymbol{\gamma}$ correspond à un spectre de puissance. L'expression du bruit \mathbf{w} et ses statistiques sont détaillées dans la suite.

3.2.2.1 Caractéristiques du dictionnaire \mathbf{B} par rapport à celles de la table de calibration \mathbf{A}

Le dictionnaire \mathbf{B} correspondant à une table de calibration d'un réseau virtuel à $N^2 - N + 1$ capteurs, les caractéristiques de celui-ci sont nécessairement différentes des caractéristiques de \mathbf{A} . Sur la figure 3.3, on représente les corrélations $r_{\mathbf{A}}(180^\circ, \psi)$ et $r_{\mathbf{B}}(180^\circ, \psi)$ en fonction de ψ . On constate que la largeur de lobe à mi-puissance est bien diminuée avec le nouveau dictionnaire \mathbf{B} , et que les lobes secondaires sont nettement atténués. Cette propriété explique le fait que l'utilisation de la matrice de covariance vectorisée en goniométrie parcimonieuse permet par exemple d'augmenter le pouvoir de résolution [42]. Pour un pas de grille fin (1° par exemple), la cohérence mutuelle de \mathbf{B} reste cependant très élevée, puisque dans ce cas $\mu(\mathbf{B}) = 0.9985$.

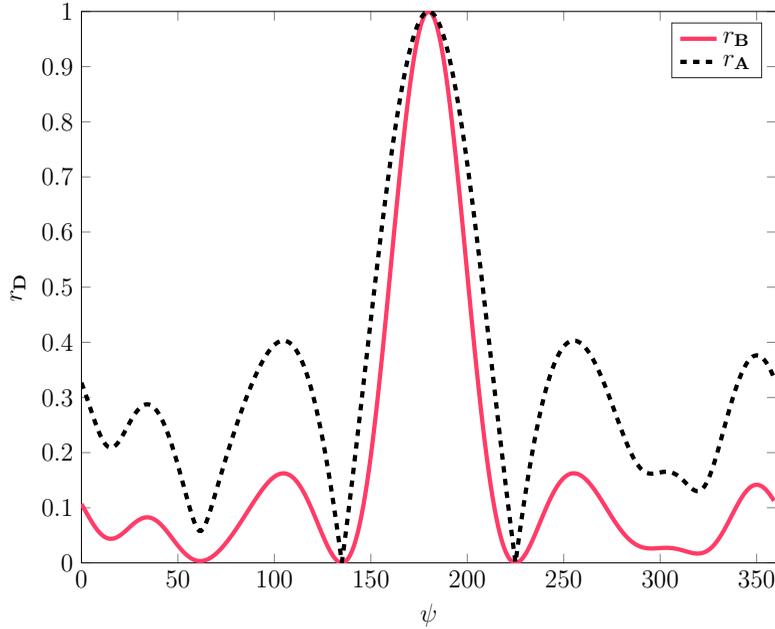


FIGURE 3.3 – Fonction de corrélation de la table de calibration \mathbf{A} pour un réseau circulaire uniforme à $N = 7$ antennes et de rayon une demi-longueur d'onde, et du dictionnaire $\mathbf{B} = \mathbf{A}^* \odot \mathbf{A}$ qui correspond au réseau virtuel construit par vectorisation de la matrice de covariance, en fonction de ψ , pour un premier vecteur fixe d'angle 180° .

3.2.3 Statistiques de la matrice de covariance vectorisée en présence d'erreurs de modèle

On considère dans un premier temps être en présence d'erreurs de modèles. On utilise le modèle décrit dans la section 1.2.3.1, avec un bruit additif sur les vecteurs directeurs donné par l'équation (1.17), et dont les statistiques sont décrites par les équations (1.19) et (1.20). On suppose par ailleurs que la matrice de covariance \mathbf{R}_{xx} est parfaitement connue ($K \rightarrow \infty$), ainsi que le niveau de bruit σ_n^2 : cela permet de se concentrer uniquement sur la dégradation des performances dues aux erreurs de modèles, sans les mélanger aux éventuelles erreurs induites par un nombre fini d'échantillons. La matrice \mathbf{R}_{xx} s'exprime ainsi par :

$$\mathbf{R}_{xx} = \sum_{m=1}^M \tilde{\mathbf{a}}(\theta_m) \tilde{\mathbf{a}}^H(\theta_m) \gamma_{\theta_m} + \sigma_n^2 \mathbf{I}_N \quad (3.15)$$

avec $\tilde{\mathbf{a}}(\theta_m)$ les vecteurs directeurs réels du système de goniométrie. La matrice de covariance vectorisée est donnée par :

$$\mathbf{r} = \tilde{\mathbf{B}}_{\Theta} \gamma_{\Theta} + \sigma_n^2 \text{vec}(\mathbf{I}_N) \quad (3.16)$$

avec $\tilde{\mathbf{B}}_{\Theta} = [\tilde{\mathbf{b}}(\theta_1), \dots, \tilde{\mathbf{b}}(\theta_M)]$ et $\tilde{\mathbf{b}}(\theta) = \tilde{\mathbf{a}}^*(\theta) \otimes \tilde{\mathbf{a}}(\theta)$. L'observation $\mathbf{y} \in \mathbb{C}^{N^2}$ considérée est ainsi :

$$\mathbf{y} = \mathbf{r} - \sigma_n^2 \text{vec}(\mathbf{I}_N) = \tilde{\mathbf{B}}_{\Theta} \gamma_{\Theta}. \quad (3.17)$$

La matrice $\tilde{\mathbf{B}}_{\Theta}$ n'est pas connue. Nous pouvons cependant calculer un dictionnaire $\mathbf{B} = [\mathbf{b}(\psi_1), \dots, \mathbf{b}(\psi_G)]$ à partir de la table de calibration \mathbf{A} connue. Pour $\mathbf{B} = \mathbf{A}^* \odot \mathbf{A}$, et $\gamma = C_{\Theta}(\gamma_{\Theta})$, on a alors :

$$\mathbf{y} = (\mathbf{A}^* \odot \mathbf{A}) \gamma + \mathbf{w} = \mathbf{B} \gamma + \mathbf{w}, \quad (3.18)$$

avec \mathbf{w} un vecteur de bruit additif issu des erreurs de modèles. D'après le modèle (1.17), on a en effet :

$$\tilde{\mathbf{b}}(\psi) = \tilde{\mathbf{a}}^*(\psi) \otimes \tilde{\mathbf{a}}(\psi) \quad (3.18a)$$

$$= (\mathbf{a}(\psi) + \mathbf{e}(\psi))^* \otimes (\mathbf{a}(\psi) + \mathbf{e}(\psi)) \quad (3.18b)$$

$$= \mathbf{a}^*(\psi) \otimes \mathbf{a}(\psi) + \mathbf{a}^*(\psi) \otimes \mathbf{e}(\psi) + \mathbf{e}^*(\psi) \otimes \mathbf{a}(\psi) + \mathbf{e}^*(\psi) \otimes \mathbf{e}(\psi), \quad (3.19)$$

d'où

$$\mathbf{w} = (\mathbf{A}^* \odot \mathbf{E} + \mathbf{E}^* \odot \mathbf{A} + \mathbf{E}^* \odot \mathbf{E}) \gamma. \quad (3.20)$$

Des calculs dont la seule difficulté est la longueur montrent que l'observation \mathbf{y} obtenue peut être considérée comme une observation gaussienne multivariée de loi $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{C})$, dont les caractéristiques $\boldsymbol{\mu} \triangleq \mathbb{E}[\mathbf{y}]$, $\boldsymbol{\Gamma} \triangleq \mathbb{E}[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^H]$ et $\mathbf{C} \triangleq \mathbb{E}[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T]$ sont

calculées, en négligeant les termes en σ_e^4 , par :

$$\boldsymbol{\mu} = \sum_{m=1}^M \mathbf{a}_m^* \otimes \mathbf{a}_m \gamma_{\theta_m} + \sigma_e^2 \sum_{m=1}^M \gamma_{\theta_m} \text{vec}(\mathbf{I}_N), \quad (3.21)$$

$$\boldsymbol{\Gamma} = \sum_{m=1}^M \gamma_{\theta_m}^2 \left([\mathbf{a}_m^* \mathbf{a}_m^T] \otimes \sigma_e^2 \mathbf{I}_N + \sigma_e^2 \mathbf{I}_N \otimes [\mathbf{a}_m \mathbf{a}_m^H] \right), \quad (3.22)$$

$$\mathbf{C} = \sum_{m=1}^M \gamma_{\theta_m}^2 \left([a_{m_1}^* \mathbf{I}_N \otimes \mathbf{a}_m, \dots, a_{m_N}^* \mathbf{I}_N \otimes \mathbf{a}_m] + [a_{m_1}^* \mathbf{I}_N \otimes \mathbf{a}_m, \dots, a_{m_N}^* \mathbf{I}_N \otimes \mathbf{a}_m]^T \right) \sigma_e^2. \quad (3.23)$$

3.2.4 Statistiques de la matrice de covariance vectorisée pour un temps d'intégration fini

On considère maintenant l'autre source d'erreurs décrite dans la section 1.2.3.1 : les erreurs dues à l'estimation de la matrice de covariance avec un nombre fini d'échantillons K , telle que donnée par l'équation (1.16). Notons $\hat{\mathbf{r}} = \text{vec}(\hat{\mathbf{R}}_{xx})$ l'estimée obtenue de la matrice de covariance vectorisée. Dans le scénario aéroporté et pour un temps d'intégration fini (BT fini), l'observation \mathbf{y} utilisée dans les algorithmes est définie par :

$$\mathbf{y} = \hat{\mathbf{r}} - \sigma_n^2 \text{vec}(\mathbf{I}_N) \quad (3.24)$$

Or il est connu que

$$\hat{\mathbf{R}}_{xx} = \mathbf{R}_{xx} + \Delta \mathbf{R}_{xx}, \quad (3.25)$$

où $K \Delta \mathbf{R}_{xx}$ suit une distribution de Wishart. L'observation s'écrit alors sous la forme :

$$\mathbf{y} = \mathbf{B} \boldsymbol{\gamma} + \delta \mathbf{r}, \quad (3.26)$$

où $\delta \mathbf{r}$ est le bruit du au nombre fini d'échantillons. L'observation \mathbf{y} (3.26) suit asymptotiquement une distribution normale multivariée complexe $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{C})$, avec

$$\boldsymbol{\mu} \hat{=} \mathbb{E}[\mathbf{y}] = \mathbf{r} - \sigma_n^2 \text{vec}(\mathbf{I}_N), \quad (3.27)$$

$$\boldsymbol{\Gamma} \hat{=} \mathbb{E}[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^H] = \frac{1}{K} \mathbf{R}_{xx}^T \otimes \mathbf{R}_{xx}, \quad (3.28)$$

$$\mathbf{C} \hat{=} \mathbb{E}[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T] = \mathbb{E}[\delta \mathbf{r} \delta \mathbf{r}^T]. \quad (3.29)$$

La matrice de pseudo-covariance \mathbf{C} est calculée bloc par bloc :

$$\mathbb{E}[\delta \mathbf{r}_i \delta \mathbf{r}_j^T] = 1/K \cdot \mathbf{r}_j \mathbf{r}_i^T \quad (3.30)$$

où $\delta \mathbf{r}_i$ est le i -ème sous-vecteur de dimension $(N \times 1)$ du vecteur $\delta \mathbf{r}$ de dimension $(N^2 \times 1)$, et \mathbf{r}_i le i -ème sous-vecteur de dimension $(N \times 1)$ du vecteur \mathbf{r} . Notons que $\mathbf{\Gamma}$ est hermitienne et \mathbf{C} est symétrique.

3.3 Modèle SMV pour le contexte urbain

Dans un contexte urbain, le signal émis par un émetteur se réfléchit sur les bâtiments environnants, ce qui donne lieu à des multi-trajets fortement corrélés vu chacun comme une source avec des angles d'arrivées différents d'un point de vue récepteur. Un tel scénario est représenté sur la figure 3.4, où l'on voit l'exemple d'un émetteur qui donne lieu à 4 trajets incidents sur le réseau d'antennes.

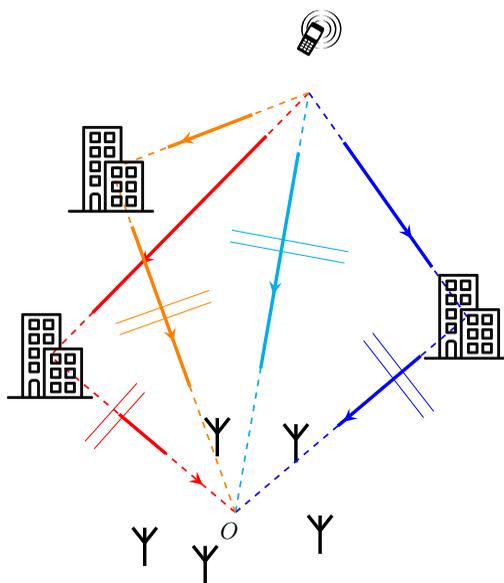


FIGURE 3.4 – Scénario urbain avec un émetteur et 4 multi-trajets incidents vus comme 4 sources d'un point de vue récepteur.

Dans ce contexte de forte corrélation entre les signaux reçus, les performances de la méthode MUSIC se dégradent fortement. Des trajets cohérents entraînent même des dysfonctionnements de celle-ci. Parmi les méthodes usuelles, la méthode du maximum de vraisemblance est donc la seule à permettre de traiter théoriquement ce scénario [73]. Sa complexité en terme de puissance de calcul ne permet cependant pas sa mise en œuvre lorsque l'on a plus de deux paramètres à estimer, soit au delà de deux trajets en 1D ou 1 trajet en 2D.

3.3.1 Modèle du signal reçu

On note $\tilde{s}(t)$ le signal complexe émis par l'émetteur, qui résulte en M multi-trajets. Le signal incident reçu du m -ième trajet s'écrit alors de la manière suivante :

$$s_m(t) = q_m \tilde{s}(t - \delta t_m) \quad (3.31)$$

où q_m est l'atténuation et $\delta t_m = \delta D_m/c$ le retard du m -ième trajet dû à la différence de marche δD_m . Dans cette thèse, on considère le contexte de multi-trajets cohérents, où les retards vérifient $\delta t_1 \approx \delta t_2 \approx \dots \approx \delta t_M$. Pour des angles d'arrivées $\{\theta_1, \dots, \theta_M\}$, le signal $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^T$ reçu en sortie du réseau de N antennes est alors :

$$\mathbf{x}(t) = \sum_{m=1}^M q_m \tilde{s}(t - \delta t_m) \mathbf{a}(\theta_m) + \mathbf{n}(t) \quad (3.32)$$

$$\approx \tilde{s}(t - \delta t) \sum_{m=1}^M q_m \mathbf{a}(\theta_m) + \mathbf{n}(t). \quad (3.33)$$

On note $\mathbf{s}(t) = [s_1(t), \dots, s_M(t)]^T$ et $\mathbf{q} = [q_1, \dots, q_M]^T$. On a alors

$$\mathbf{s}(t) = \mathbf{q} \tilde{s}(t - \delta t). \quad (3.34)$$

La matrice de covariance des sources $\mathbf{R}_{ss} = \mathbb{E}[\mathbf{s}(t)\mathbf{s}^H(t)]$ est de rang 1, et elle peut s'écrire sous la forme :

$$\mathbf{R}_{ss} = \gamma \mathbf{q} \mathbf{q}^H \quad (3.35)$$

avec $\gamma = \mathbb{E}[|\tilde{s}(t)|^2]$. Cette propriété nous permet d'utiliser comme observation vectorielle le premier vecteur propre de la matrice de covariance.

3.3.2 Observation SMV dans le cas de sources cohérentes : le premier vecteur propre de la matrice de covariance

3.3.2.1 Expression du premier vecteur propre

On note $\mathbf{R} = \mathbf{A}_\Theta \mathbf{R}_{ss} \mathbf{A}_\Theta^H$, pour \mathbf{A}_Θ la matrice de mélange des sources. D'après (3.35), on peut écrire

$$\mathbf{R} = \mu_1 \mathbf{v} \mathbf{v}^H \quad (3.36)$$

avec $\mu_1 = \gamma \mathbf{q}^H \mathbf{A}_\Theta^H \mathbf{A}_\Theta \mathbf{q}$ et

$$\mathbf{v} = \frac{1}{\sqrt{\mathbf{q}^H \mathbf{A}_\Theta^H \mathbf{A}_\Theta \mathbf{q}}} \mathbf{A}_\Theta \mathbf{q}. \quad (3.37)$$

\mathbf{v} est le vecteur propre normé associé à la plus grande (et unique non nulle) valeur propre μ_1 de \mathbf{R} . Toute l'information des paramètres des multi-trajets est ainsi contenue dans cet unique vecteur de dimension N le nombre d'antennes. Si l'on suppose que $\mathbf{n}(t)$ est un bruit

blanc indépendant du signal dont la covariance vérifie l'équation (3.2), alors

$$\mathbf{R}_{xx} = \mathbf{R} + \sigma_n^2 \mathbf{I}_N. \quad (3.38)$$

Dans ce cas, le premier vecteur propre \mathbf{v} de \mathbf{R} est aussi premier vecteur propre de \mathbf{R}_{xx} : seule la valeur propre est différente.

On pose

$$\boldsymbol{\rho}_\Theta = \frac{1}{\sqrt{\mathbf{q}^H \mathbf{A}_\Theta^H \mathbf{A}_\Theta \mathbf{q}}} \mathbf{q}. \quad (3.39)$$

D'après (3.37), le premier vecteur propre \mathbf{v} de \mathbf{R}_{xx} est combinaison linéaire des vecteurs directeurs des sources :

$$\mathbf{v} = \mathbf{A}_\Theta \boldsymbol{\rho}_\Theta. \quad (3.40)$$

Soit $\boldsymbol{\rho} = C_\Theta(\boldsymbol{\rho}_\Theta)$: $\boldsymbol{\rho}$ est le vecteur parcimonieux dont les composantes non nulles sont les composantes de $\boldsymbol{\rho}_\Theta$, aux indices correspondants aux angles de Θ dans Ψ . Le vecteur \mathbf{v} s'exprime ainsi sous forme parcimonieuse par :

$$\mathbf{v} = \mathbf{A} \boldsymbol{\rho}. \quad (3.41)$$

On propose ainsi de prendre comme observation \mathbf{y} , dans le cas de trajets cohérents, le premier vecteur propre \mathbf{v} de la matrice de covariance. On a dans ce cas $\mathbf{y} \in \mathbb{C}^N$. L'observation s'écrit sous forme parcimonieuse avec pour dictionnaire la table de calibration \mathbf{A} , et le vecteur parcimonieux $\boldsymbol{\rho}$ représente un pseudo-spectre d'amplitudes normalisées. En prenant en compte les différentes sources d'incertitudes, on obtient par la suite une observation bruitée :

$$\mathbf{y} = \mathbf{A} \boldsymbol{\rho} + \mathbf{w}. \quad (3.42)$$

L'observation est donc bien de la forme $\mathbf{y} = \mathbf{D} \mathbf{u} + \mathbf{w}$, avec $\mathbf{D} = \mathbf{A}$, soit directement la table de calibration et $\mathbf{u} = \boldsymbol{\rho}$ le pseudo-spectre en amplitude. L'expression du bruit \mathbf{w} et ses statistiques sont détaillées dans la suite.

3.3.2.2 Calcul du premier vecteur propre par la méthode des puissances

La méthode des puissances est une méthode itérative permettant de calculer de façon approchée la valeur propre dominante d'une matrice \mathbf{R} et le vecteur propre associé. On considère une initialisation $\mathbf{v}^{(0)}$. Pour une valeur propre dominante strictement positive, la suite de vecteurs définie par la relation de récurrence :

$$\mathbf{v}^{(i+1)} = \frac{1}{\|\mathbf{R} \mathbf{v}^{(i)}\|} \mathbf{R} \mathbf{v}^{(i)} \quad (3.43)$$

converge vers le vecteur unitaire de \mathbf{R} associé à la valeur propre μ_1 .

En pratique, on initialise par $\mathbf{v}^{(0)} = \mathbf{R}_{:,1}$ la première colonne de la matrice \mathbf{R} . On a alors :

$$\mathbf{R}\mathbf{v}^{(0)} = \mathbf{R} \cdot \mathbf{R}_{:,1} \quad (3.44)$$

$$= (\mathbf{A}_\Theta \mathbf{R}_{ss} \mathbf{A}_\Theta^H) \cdot (\mathbf{A}_\Theta \mathbf{R}_{ss} [\mathbf{A}_\Theta^H]_{:,1}) \quad (3.45)$$

$$= \gamma^2 \cdot (\mathbf{q}^H [\mathbf{A}_\Theta^H]_{:,1}) \cdot (\mathbf{q}^H \mathbf{A}_\Theta^H \mathbf{A}_\Theta \mathbf{q}) \cdot \mathbf{A}_\Theta \mathbf{q} \quad (3.46)$$

et

$$\|\mathbf{R}\mathbf{v}^{(0)}\| = \gamma^2 \cdot (\mathbf{q}^H [\mathbf{A}_\Theta^H]_{:,1}) \cdot (\mathbf{q}^H \mathbf{A}_\Theta^H \mathbf{A}_\Theta \mathbf{q}) \sqrt{\mathbf{q}^H \mathbf{A}_\Theta^H \mathbf{A}_\Theta \mathbf{q}} \quad (3.47)$$

La première itération donne alors :

$$\mathbf{v}^{(1)} = \frac{1}{\|\mathbf{R}\mathbf{v}^{(0)}\|} \mathbf{R}\mathbf{v}^{(0)} \quad (3.48)$$

$$= \frac{1}{\sqrt{\mathbf{q}^H \mathbf{A}_\Theta^H \mathbf{A}_\Theta \mathbf{q}}} \mathbf{A}_\Theta \mathbf{q} \quad (3.49)$$

On retrouve ainsi l'expression (3.37) : la méthode des puissances appliquée au cas de sources cohérentes permet d'obtenir en une seule itération le premier vecteur propre, et évite ainsi des calculs plus onéreux. Cette méthode est en pratique couramment utilisée pour déterminer le premier vecteur propre en présence de sources fortement corrélées.

3.3.3 Statistiques du vecteur propre en présence d'erreurs de modèle

On considère dans un premier temps la présence d'erreurs de modèles : les vecteurs directeurs $\tilde{\mathbf{a}}(\psi)$ du système de mesure ne correspondent pas exactement aux vecteurs directeurs $\mathbf{a}(\psi)$ utilisés dans les algorithmes. On considère le modèle de bruit additif donné par l'équation (1.17), détaillé par ses statistiques (1.19) et (1.20).

On suppose par ailleurs que la matrice de covariance \mathbf{R}_{xx} est parfaitement connue : cela permet de se concentrer uniquement sur la dégradation des performances due aux erreurs de modèles, sans les mélanger aux éventuelles erreurs induites par le nombre fini d'échantillons. L'observation \mathbf{y} considérée, correspondant au premier vecteur propre de \mathbf{R}_{xx} , est ainsi :

$$\mathbf{y} = \tilde{\mathbf{A}}_\Theta \boldsymbol{\rho}_\Theta \quad (3.50)$$

avec $\boldsymbol{\rho}_\Theta = \left(1/\sqrt{\mathbf{q}^H \tilde{\mathbf{A}}_\Theta^H \tilde{\mathbf{A}}_\Theta \mathbf{q}}\right) \mathbf{q}$. La vraie matrice de mélange des sources $\tilde{\mathbf{A}}_\Theta$ n'est pas connue. En faisant l'hypothèse que les directions d'arrivées Θ sont incluses dans l'ensemble des directions candidates Ψ , et en combinant (3.50) et (1.17), on montre par un développement limité à l'ordre 1 que l'observation peut s'écrire sous la forme parcimonieuse bruitée suivante :

$$\mathbf{y} = \mathbf{A}\boldsymbol{\rho} + \delta\mathbf{v} = \mathbf{v} + \delta\mathbf{v}, \quad (3.51)$$

avec $\boldsymbol{\rho} = C_{\Theta}(\boldsymbol{\rho}_{\Theta})$ avec C . l'opérateur de complétion de zéros, et

$$\delta \mathbf{v} = \frac{1}{2(\mathbf{q}^H \mathbf{A}_{\Theta}^H \mathbf{A}_{\Theta} \mathbf{q})^{3/2}} \left(2\mathbf{q}^H \mathbf{A}_{\Theta}^H \mathbf{A}_{\Theta} \mathbf{q} \mathbf{E}_{\Theta} \mathbf{q} - \mathbf{q}^H (\mathbf{E}_{\Theta}^H \mathbf{A}_{\Theta} + \mathbf{A}_{\Theta}^H \mathbf{E}_{\Theta}) \mathbf{q} \mathbf{A}_{\Theta} \mathbf{q} \right). \quad (3.52)$$

On établit alors que l'approximation d'ordre 1 de l'observation \mathbf{y} est gaussienne, de moyenne \mathbf{v} (3.37), et dont les statistiques du second ordre $\boldsymbol{\Gamma} = \mathbb{E}[\delta \mathbf{v} \delta \mathbf{v}^H]$ et $\mathbf{C} = \mathbb{E}[\delta \mathbf{v} \delta \mathbf{v}^T]$ sont données par :

$$\boldsymbol{\Gamma} = \frac{\sigma_e^2 \mathbf{q}^H \mathbf{q}}{2\mathbf{q}^H \mathbf{A}_{\Theta}^H \mathbf{A}_{\Theta} \mathbf{q}} \left[2\mathbf{I}_N - \frac{\mathbf{A}_{\Theta} \mathbf{q} \mathbf{q}^H \mathbf{A}_{\Theta}^H}{\mathbf{q}^H \mathbf{A}_{\Theta}^H \mathbf{A}_{\Theta} \mathbf{q}} \right] \quad (3.53)$$

$$\mathbf{C} = \frac{-\sigma_e^2 \mathbf{q}^H \mathbf{q}}{2\mathbf{q}^H \mathbf{A}_{\Theta}^H \mathbf{A}_{\Theta} \mathbf{q}} \times \frac{\mathbf{A}_{\Theta} \mathbf{q} \mathbf{q}^T \mathbf{A}_{\Theta}^T}{\mathbf{q}^H \mathbf{A}_{\Theta}^H \mathbf{A}_{\Theta} \mathbf{q}} \quad (3.54)$$

3.4 Résumé des observations et statistiques en fonction du scénario

Les différentes observations SMV proposées se mettent toutes sous la forme $\mathbf{y} = \mathbf{D}\mathbf{u} + \mathbf{w}$. Nous rappelons pour chacune les expressions du dictionnaire \mathbf{D} et du vecteur à estimer \mathbf{u} , et nous renvoyons vers les équations donnant les statistiques de l'observation.

	Scénario aéroporté (sources indépendantes)	Scénario urbain (sources cohérentes)
Observation	$\mathbf{y} = \text{vec}(\mathbf{R}_{xx}) - \sigma_n^2 \text{vec}(\mathbf{I}_N)$	$\mathbf{y} = \mathbf{v}$ (1er vecteur propre de \mathbf{R}_{xx})
Dictionnaire	$\mathbf{D} = \mathbf{B} = \mathbf{A}^* \odot \mathbf{A}$	$\mathbf{D} = \mathbf{A}$
Vecteur à estimer	$\mathbf{u} = \boldsymbol{\gamma}$ pseudo-spectre en puissance	$\mathbf{u} = \boldsymbol{\rho}$ pseudo-spectre en amplitude
Statistiques en présence d'erreurs de modèle	$\boldsymbol{\mu}$ (3.21) $\boldsymbol{\Gamma}$ (3.22) \mathbf{C} (3.23)	$\boldsymbol{\mu} = \mathbf{v}$ $\boldsymbol{\Gamma}$ (3.53) \mathbf{C} (3.54)
Statistiques à BT fini	$\boldsymbol{\mu}$ (3.27) $\boldsymbol{\Gamma}$ (3.28) \mathbf{C} (3.29)	non traité analytiquement

TABLE 3.1 – Expression des grandeurs statistiques d'intérêt en fonction du scénario.

Chapitre 4

Paramètre de régularisation

On considère dans ce chapitre la minimisation du problème régularisé par la norme ℓ_0 tel que définie par l'équation (\mathcal{R}_λ) . Le paramètre de régularisation λ vise à équilibrer le poids relatif entre le terme d'attache aux données $1/2\|\mathbf{y} - \mathbf{D}\mathbf{u}\|_2^2$ et le terme de régularisation $\|\mathbf{u}\|_0$ dans le critère J_{ℓ_0} . Le minimum de J_{ℓ_0} dépend donc à la fois du paramètre de régularisation λ et de \mathbf{u} .

On rappelle que $\mathbf{u} \in \mathbb{C}^G$ est le vecteur parcimonieux à estimer, $\mathbf{y} \in \mathbb{C}^{N'}$ est l'observation issue de la matrice de covariance, et $\mathbf{D} \in \mathbb{C}^{N' \times G}$ le dictionnaire issu de la table de calibration. Selon le scénario, les expressions des termes \mathbf{u} , \mathbf{y} et \mathbf{D} sont rappelées dans le tableau 3.1.

4.1 Introduction et interprétation de la littérature existante pour la goniométrie

Si certains articles de la littérature se contentent de présenter les résultats de simulations avec la valeur d'un paramètre de régularisation λ leur permettant d'obtenir les meilleures performances, certains articles se sont tout de même penchés davantage sur la question du réglage du paramètre de régularisation.

Il est cependant important de bien différencier les paramètres selon la formulation du problème. Bien que nous nous intéressions au paramètre de régularisation λ utilisé dans la formulation régularisée par la norme ℓ_0 (\mathcal{R}_λ) , nous faisons ici le choix de discuter plus largement des méthodes employées dans la littérature pour un certain nombre de problèmes parcimonieux.

4.1.1 Réglage d'un paramètre basé sur le niveau de bruit

La formulation sous contrainte (\mathcal{C}_ϵ) , où la contrainte porte sur le résidu $\|\mathbf{y} - \mathbf{D}\mathbf{u}\|_2^2$, peut être la plus appropriée dans le cas où le niveau de bruit est connu ou estimable. Dans cette formulation, la norme ℓ_0 a souvent été remplacée par sa relaxation convexe (la norme ℓ_1). Dans ce cas, le paramètre ϵ parfois également appelé paramètre de régularisation, est

réglé par rapport à un niveau de bruit ou d'incertitudes (erreurs de modèles) [42, 81]. Une transformation à l'aide du Lagrangien permet également de régler ce paramètre sur le niveau de bruit [2].

On notera également qu'une approche bayésienne avec un a priori selon une distribution laplacienne mène à la minimisation d'un critère régularisé par la norme ℓ_1 avec pour paramètre de régularisation $\lambda = \sigma_n^2/\gamma$, avec σ_n^2 la variance du bruit et γ la puissance des sources.

Enfin, les algorithmes de descente à gradient-proximal pour la résolution des problèmes régularisés font appel à des étapes de seuillage (avec les proximaux) : le seuillage dur pour la norme ℓ_0 , le seuillage doux pour la norme ℓ_1 , et des règles de seuillage plus lisses ou intermédiaires pour les autres normes ou approximations de la norme ℓ_0 , qui reposent tous en partie sur la valeur du paramètre de régularisation λ . Ainsi, en se basant sur les propriétés de seuillage et du bruit, il a été proposé la règle du $\alpha\sigma_n$ [25] : λ est choisi égal à $\alpha\sigma_n$, avec α un coefficient placé empiriquement entre 2 et 4, et σ_n l'écart-type du bruit. D'un point de vue statistique, cela revient donc à conserver uniquement les composantes supérieures à un niveau de bruit : si l'on considère une distribution gaussienne des coefficients, un seuil à 2σ permet de conserver 5% des plus grands coefficients, et pour 3σ , 0.3% des plus grands coefficients. Ce critère simple ne s'applique qu'aux problèmes de petite dimension. Pour les problèmes en grande dimension, il a été proposé un seuil universel $\lambda = \sigma\sqrt{2\log N'}$ [25], où N' est la dimension de l'observation. Cependant, les hypothèses sur la distribution sont très restrictives et loin de la réalité.

4.1.2 Relation entre problème régularisé et problème contraint : lien avec le maximum de vraisemblance

La littérature sur le réglage du paramètre λ fait apparaître des liens entre le niveau de bruit souhaité et le paramètre, à travers la règle de seuillage édictée par [25], ou à travers la formulation bayésienne. Bien souvent, le réglage fourni est valable pour des problèmes utilisant la norme ℓ_1 , ce qui implique une faible cohérence mutuelle du dictionnaire. Pour un problème d'estimation des directions d'arrivées avec un dictionnaire aux colonnes voisines très fortement corrélées entre elles, le réglage du paramètre sur le niveau de bruit seul ne semble pas suffisant pour obtenir un résultat parcimonieux d'où les angles d'arrivées pourront être extraits.

Nikolova [57] a travaillé sur l'équivalence entre le problème régularisé (\mathcal{R}_λ) et le problème contraint (\mathcal{C}_s). Le travail proposé par la suite établit de manière originale un lien plus étroit que celui la littérature actuelle entre niveau de bruit, degré de parcimonie, et paramètre de régularisation.

Dans la suite, les termes "valeur optimale" et "solutions optimales" font référence à la valeur de minima globaux et aux solutions globales. Nous faisons par ailleurs l'hypothèse que le dictionnaire $\mathbf{D} \in \mathbb{C}^{N' \times G}$ satisfait $\text{rank}(\mathbf{D}) = N' < G$. On suppose également $\mathbf{y} \neq \mathbf{0}$.

On considère dans un premier temps le problème contraint (\mathcal{C}_s). Soit $c_s(\mathbf{y})$ la valeur

optimale du résidu et $\hat{U}_c^s(\mathbf{y})$ l'ensemble des solutions optimales de (\mathcal{C}_s) pour l'observation \mathbf{y} donnée, correspondant à une réalisation de bruit :

$$c_s(\mathbf{y}) = \inf \{ \|\mathbf{D}\mathbf{u} - \mathbf{y}\|_2^2 \mid \mathbf{u} \in \mathbb{C}^G \text{ et } \|\mathbf{u}\|_0 \leq s \} \quad (4.1)$$

$$\hat{U}_c^s(\mathbf{y}) = \{ \mathbf{u} \in \mathbb{C}^G \text{ et } \|\mathbf{u}\|_0 \leq s \mid \|\mathbf{D}\mathbf{u} - \mathbf{y}\|_2^2 = c_s \} \quad (4.2)$$

On peut montrer que les valeurs optimales c_s s'écrivent comme les valeurs optimales du critère du maximum de vraisemblance $J_{MV}(\Phi_s)$ (1.46), considérant une observation \mathbf{y} et s potentielles sources :

$$c_s(\mathbf{y}) = \min_{\Phi_s} \mathbf{y}^H \mathbf{\Pi}_n(\Phi_s) \mathbf{y} = \min_{\Phi_s} J_{MV}(\Phi_s) \quad (4.3)$$

avec $\Phi_s = \{\varphi_1, \dots, \varphi_s\}$ un ensemble de s angles d'arrivées candidats inclus dans l'ensemble Ψ des directions candidates de la grille, et

$$\mathbf{\Pi}_n(\Phi_s) = \mathbf{I} - \mathbf{\Pi}(\Phi_s) \quad (4.4)$$

$$\mathbf{\Pi}(\Phi_s) = \mathbf{D}_{\Phi_s} [\mathbf{D}_{\Phi_s}^H \mathbf{D}_{\Phi_s}]^{-1} \mathbf{D}_{\Phi_s}^H. \quad (4.5)$$

En présence de M sources incidentes, l'ensemble $\hat{U}_c^M(\mathbf{y})$ correspond donc aux solutions du maximum de vraisemblance déterministe. Cet ensemble doit contenir un seul élément afin d'estimer \mathbf{u} sans ambiguïté.

Soit L le plus petit entier tel que $c_L(\mathbf{y})$ soit nul : on a $L \leq N'$, avec N' la dimension de l'observation \mathbf{y} . Pour M sources et en asymptotique, on vérifie $L = M$. Pour une observation bruitée, on aura de façon générique $L = N'$. La séquence c_s , pour $s = \{0, \dots, L\}$, vérifie

$$c_L(\mathbf{y}) = 0 \leq \dots \leq c_{M+1}(\mathbf{y}) \leq c_M(\mathbf{y}) \leq c_{M-1}(\mathbf{y}) \leq \dots \leq c_0(\mathbf{y}) \quad (4.6)$$

avec $c_0(\mathbf{y}) = \|\mathbf{y}\|_2^2$. Pour $\hat{\mathbf{u}}_s \in \hat{U}_c^s(\mathbf{y})$, avec $s \leq L$, on a par ailleurs $\|\hat{\mathbf{u}}_s\|_0 = s$.

Nous nous intéressons maintenant au problème ℓ_0 régularisé (\mathcal{R}_λ) . Pour un paramètre λ fixé, on définit la fonction $J_{\ell_0}^\lambda$ telle que :

$$J_{\ell_0}^\lambda : \begin{array}{ccc} \mathbb{C}^G & \longrightarrow & \mathbb{R} \\ \mathbf{u} & \longmapsto & J_{\ell_0}(\lambda, \mathbf{u}) \end{array} \quad (4.7)$$

Pour λ fixé et \mathbf{y} donné, l'ensemble $\hat{U}_r^\lambda(\mathbf{y})$ des solutions optimales de (\mathcal{R}_λ) est défini par :

$$\hat{U}_r^\lambda(\mathbf{y}) = \left\{ \mathbf{u} \in \mathbb{C}^G \mid J_{\ell_0}^\lambda(\mathbf{u}) = r_\lambda \right\}, \quad (4.8)$$

avec

$$r_\lambda = \inf \left\{ J_{\ell_0}^\lambda(\mathbf{u}) \mid \mathbf{u} \in \mathbb{C}^G \right\}. \quad (4.9)$$

Nous cherchons les valeurs du paramètre λ telles que les solutions optimales de (\mathcal{C}_s) et

(\mathcal{R}_λ) coïncident, c'est-à-dire $\hat{U}_r^\lambda(\mathbf{y}) = \hat{U}_c^s(\mathbf{y})$, en particulier pour $s = M$ dans le cas de M sources incidentes. Pour tout $s \leq L$ et $\hat{\mathbf{u}}^s \in \hat{U}_c^s(\mathbf{y})$, on a :

$$J_{\ell_0}(\lambda, \hat{\mathbf{u}}^s) = \frac{1}{2}c_s(\mathbf{y}) + \lambda s. \quad (4.10)$$

Pour tout \mathbf{u}^s tel que $\|\mathbf{u}^s\|_0 = s$, on a alors $J_{\ell_0}(\lambda, \hat{\mathbf{u}}^s) \leq J_{\ell_0}(\lambda, \mathbf{u}^s)$. Nous pouvons tracer les différentes fonctions affines $J_{\ell_0}(\lambda, \hat{\mathbf{u}}^s)$ (4.10) en fonction de λ pour $\hat{\mathbf{u}}^s \in \hat{U}_c^s(\mathbf{y})$ et $s = \{0 \dots, L\}$. Elles ont pour ordonnées à l'origine $1/2c_s(\mathbf{y})$, avec $c_s(\mathbf{y})$ décroissant (4.6), et pour pentes s . Un exemple est représenté sur la figure 4.1.

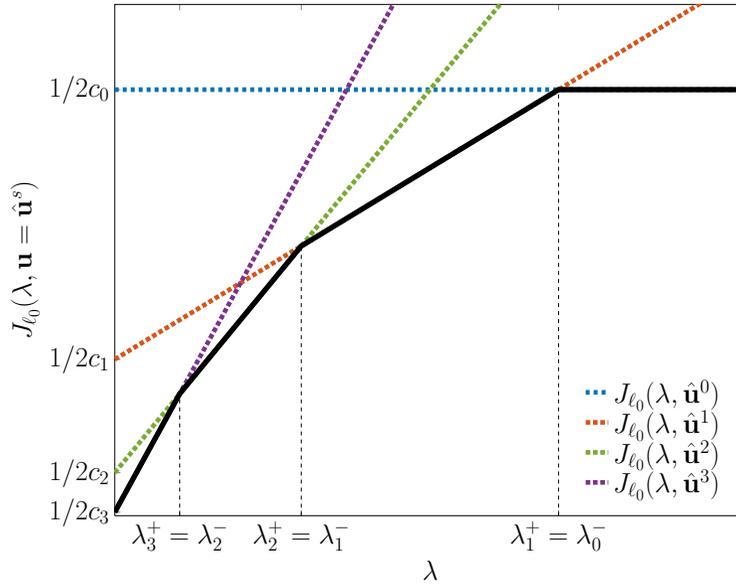


FIGURE 4.1 – Représentation des fonctions $J_{\ell_0}(\lambda, \hat{\mathbf{u}}^s)$ en fonction de λ pour $\hat{\mathbf{u}}^s \in \hat{U}_c^s(\mathbf{y})$ et $s = \{0 \dots, 3\}$ (traits pointillés), pour une réalisation de \mathbf{y} . En trait plein, la courbe $\lambda \mapsto \min_s J_{\ell_0}(\lambda, \hat{\mathbf{u}}^s)$.

Soit $\hat{\mathbf{u}}^M \in \hat{U}_c^M(\mathbf{y})$. L'objectif est de trouver λ tel que, pour tout $s \neq M$ et $\hat{\mathbf{u}}^s \in \hat{U}_c^s(\mathbf{y})$, $J_{\ell_0}(\lambda, \hat{\mathbf{u}}^M) < J_{\ell_0}(\lambda, \hat{\mathbf{u}}^s)$, soit :

$$\forall s \neq M, \quad \frac{1}{2}c_M(\mathbf{y}) + \lambda M < \frac{1}{2}c_s(\mathbf{y}) + \lambda s. \quad (4.11)$$

Autrement dit, on cherche les valeurs de λ pour que le minimum de J_{ℓ_0} soit atteint pour une valeur de \mathbf{u} dont seulement M composantes sont non nulles. L'équation (4.11) nous permet d'obtenir la relation suivante :

$$\hat{U}_r^\lambda(\mathbf{y}) = \hat{U}_c^M(\mathbf{y}) \quad \Leftrightarrow \quad \lambda_M^-(\mathbf{y}) < \lambda < \lambda_M^+(\mathbf{y}) \quad (4.12)$$

avec

$$\lambda_M^-(\mathbf{y}) = \max_s \left\{ \frac{c_M(\mathbf{y}) - c_s(\mathbf{y})}{2(s - M)} \mid M < s \leq L \right\}, \quad (4.13)$$

$$\lambda_M^+(\mathbf{y}) = \min_s \left\{ \frac{c_s(\mathbf{y}) - c_M(\mathbf{y})}{2(M - s)} \mid 0 \leq s < M \right\}. \quad (4.14)$$

On a également $\lambda_0^+ = +\infty$, et $\lambda_L^- = 0$.

La figure 4.1 représente les fonctions $J_{\ell_0}(\lambda, \hat{\mathbf{u}}^s)$ en fonction de λ pour $\hat{\mathbf{u}}^s \in \hat{U}_c^s(\mathbf{y})$ et $s = \{0 \dots, 3\}$. On notera que, selon les données et le problème, il peut arriver que $]\lambda_s^-(\mathbf{y}); \lambda_s^+(\mathbf{y})[= \emptyset$; cela peut notamment arriver dans des cas difficiles tels que faible RSB, limite de résolution etc. On ne considère ici que les valeurs de s pour lesquelles $\lambda_s^-(\mathbf{y}) < \lambda_s^+(\mathbf{y})$, ce qui résulte en des intervalles non vides, et on a alors $\lambda_s^+(\mathbf{y}) = \lambda_{s-1}^-(\mathbf{y})$. Les valeurs $\lambda_3^+ = \lambda_2^-$, $\lambda_2^+ = \lambda_1^-$ et $\lambda_1^+ = \lambda_0^-$ ont été représentées en abscisse de la figure 4.1 aux intersections entre les différentes courbes.

L'intervalle $]\lambda_M^-(\mathbf{y}); \lambda_M^+(\mathbf{y})[$ donné par l'équivalence (4.12) est d'une importance capitale : si λ est choisi dans l'intervalle, alors nous sommes sûrs que le minimum global de J_{ℓ_0} correspond à la solution du maximum de vraisemblance déterministe en présence de M sources.

4.1.3 Chemin de régularisation et méthodes homotopiques

Un chemin de régularisation consiste en l'étude des solutions du problème régularisé en fonction du paramètre de régularisation λ . Le paramètre de régularisation est choisi a posteriori parmi toutes les solutions obtenues selon différents critères.

4.1.3.1 La courbe en L

La courbe en L historiquement utilisée dans la littérature [1, 40, 41] est un critère graphique : elle représente, pour une pénalité basée sur une norme ℓ_p , la norme de la solution obtenue $\|\hat{\mathbf{u}}_\lambda\|_p$ en fonction du résidu $\|\mathbf{y} - \mathbf{D}\hat{\mathbf{u}}_\lambda\|_2$, pour différentes valeurs de λ . Pour des fonctions de régularisation continues, la courbe obtenue est plus ou moins proche de la lettre L. Le critère graphique de la courbe en L a été utilisé en particulier pour choisir le paramètre de régularisation pour une pénalité basée sur la norme ℓ_2 ou la norme ℓ_1 : il s'agit de prendre le paramètre λ pour lequel le rayon de courbure est le plus important, soit dans le coin de la courbe en L. Cependant, ce critère n'est pas applicable quand la pénalité est basée sur la norme ℓ_0 en raison des discontinuités introduites par celle-ci. De plus, ce critère est empirique.

4.1.3.2 Méthodes homotopiques basées sur la courbe ℓ_0

Duan et al. [26, 71] ont proposé un algorithme de continuation qui utilise les propriétés des minimiseurs de J_{ℓ_0} en fonction de λ . La courbe $\lambda \mapsto \min_s J_{\ell_0}(\lambda, \hat{\mathbf{u}}^s)$ représentée sur la figure 4.1 est linéaire par morceaux : l'ensemble des solutions optimales de J_{ℓ_0} par rapport

à λ est donc constant par morceaux, avec un nombre fini d'intervalles $]\lambda_s^-; \lambda_s^+[$. Il est donc inutile d'effectuer les minimisations de J_{ℓ_0} pour un grand nombre de paramètres de régularisation λ , mais seulement pour quelques valeurs bien choisies qui donneront chacune une solution $\hat{\mathbf{u}}^s$. La méthode proposée repose sur l'utilisation de l'algorithme SBR glouton bidirectionnel pour des valeurs décroissantes de λ .

L'étude complète des minimiseurs de la fonction J_{ℓ_0} qui sera proposée dans le chapitre 5 justifie, de notre point de vue, que les méthodes homotopiques basées sur une décroissance du paramètre de régularisation λ ne peuvent permettre d'estimer les directions d'arrivées de deux sources proches. Encore une fois, la cohérence mutuelle très élevée d'un dictionnaire basé sur une table de calibration en goniométrie est pénalisante pour un grand nombre de méthodes parcimonieuses.

4.2 Comportement des algorithmes à gradient-proximal selon λ pour une observation donnée

L'étude de Mila Nikolova [57] permet d'établir la relation entre les minimiseurs du problème contraint (\mathcal{C}_s) et du problème régularisé (\mathcal{R}_λ). Elle est détaillée dans le chapitre 4.1.2 et nous donne un intervalle théorique $]\lambda_M^-(\mathbf{y}); \lambda_M^+(\mathbf{y})[$ tel que la solution globale du problème régularisé soit un vecteur avec M composantes non nulles, et qui soit la solution du maximum de vraisemblance déterministe. Ces M composantes non nulles nous indiquent donc la position des M sources incidentes.

L'équivalence entre les problèmes ne permet cependant pas de dire que le minimum global sera nécessairement atteint par un algorithme de minimisation. En effet, comme je l'ai expliqué dans le chapitre 2.4, le critère J_{ℓ_0} est non convexe et présente de nombreux minimas locaux, et il n'existe pas d'algorithme optimal permettant de trouver la solution dans un temps polynomial : la seule façon d'obtenir cette solution optimale est de passer par un algorithme combinatoire, nécessitant un temps de calcul trop important. Les algorithmes de descente à gradient-proximal que nous choisissons d'utiliser pour leur faible coût de calcul sont sous-optimaux : il n'ont pas de garantie d'atteindre le minimum global. La solution peut notamment dépendre de l'initialisation. Afin d'augmenter les chances d'obtenir une solution proche du minimum global, on utilise dans ce paragraphe des algorithmes à gradient-proximal adaptés à la pénalité CEL0, qui préserve le minimum global de J_{ℓ_0} tout en supprimant des minimas locaux [69].

Nous étudions le comportement des algorithmes à gradient-proximal selon λ , pour une observation donnée, afin de valider l'utilité pratique de l'intervalle théorique $]\lambda_M^-(\mathbf{y}); \lambda_M^+(\mathbf{y})[$.

4.2.1 Scénario de deux sources fortement corrélées à BT fini, avec un réseau circulaire uniforme à 7 antennes (scénario 1a)

On considère dans un premier temps un réseau circulaire uniforme (UCA pour *Uniform Circular Array*) composé de $N = 7$ antennes, placées sur un cercle de rayon égal à 0.8 fois la longueur d'onde. On se place dans un scénario avec $M = 2$ sources fortement corrélées, et on utilise ainsi le premier vecteur propre de la matrice de covariance estimée comme observation. Les angles d'arrivées des deux sources sont $\theta_1 = 40^\circ$ et $\theta_2 = 120^\circ$. Le scénario est représenté sur la figure.

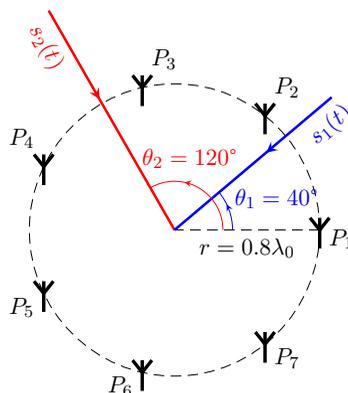


FIGURE 4.2 – Scénario avec un réseau UCA de 7 antennes.

Le coefficient de corrélation temporelle entre les sources est de 0.9, le RSB est égal à 0 dB, et le nombre d'échantillons pour l'estimation de la matrice de covariance est fixé à $K = 100$. Pour une réalisation de l'observation \mathbf{y} donnée, on calcule $\lambda_2^-(\mathbf{y})$ et $\lambda_2^+(\mathbf{y})$ selon les équations (4.13) et (4.14). On représente sur la figure 4.3 les pseudo-spectres représentant le vecteur $\hat{\mathbf{u}}$ obtenus avec l'algorithme FB-CEL0 selon λ . L'algorithme est initialisé par le pseudo-spectre obtenu d'un filtre adapté de Capon : $\hat{u}_g^{(ini)} = \mathbf{w}_g^H \mathbf{y}$, avec

$$\mathbf{w}_g = \frac{1}{\sqrt{N}} \cdot \frac{\hat{\mathbf{R}}_{xx}^{-1} \mathbf{d}_g}{\mathbf{d}_g^H \hat{\mathbf{R}}_{xx}^{-1} \mathbf{d}_g}. \quad (4.15)$$

Pour $\lambda < \lambda_2^-$, le pseudo-spectre comporte de larges lobes autour des deux directions d'arrivées, et de multiples lobes secondaires. Pour $\lambda = \lambda_2^-$, un bon nombre de composantes du vecteur obtenus sont nulles, et il n'y a plus que des larges lobes autour des vraies directions d'arrivées. Pour $\lambda = \lambda_2^+$, les deux seuls lobes deviennent très fins. Pour $\lambda > \lambda_2^+$, on ne retrouve plus qu'un seul pic correspondant à une des directions d'arrivées : l'information sur la deuxième direction d'arrivée a disparu.

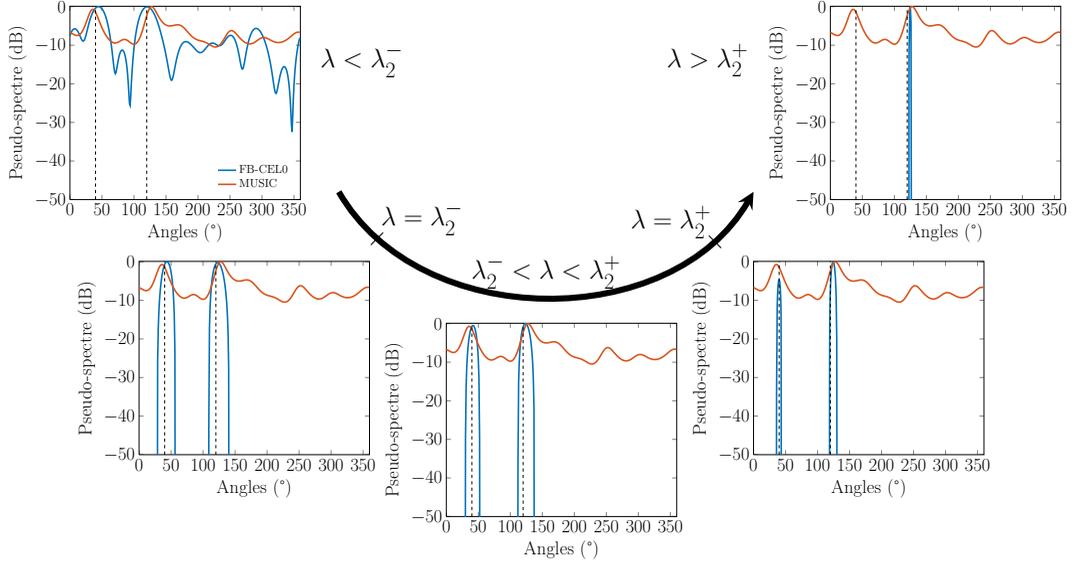


FIGURE 4.3 – Pseudo-spectres du vecteur $\hat{\mathbf{u}}$ obtenus par l’algorithme FB-CEL0 en fonction de λ , comparés au pseudo-spectre de MUSIC pour un réseau UCA à 7 antennes (scénario 1a). $\lambda_2^- = 0.0034$. $\lambda_2^+ = 0.16$.

4.2.2 Scénario de deux sources fortement corrélées à BT fini, avec un réseau linéaire uniforme à 10 antennes (scénario 1b)

On considère maintenant un réseau linéaire uniforme (ULA pour *Uniform Linear Array*) composé de $N = 10$ antennes, espacées chacune de 0.5 longueur d’onde. On se place dans un scénario avec $M = 2$ sources fortement corrélées, et on utilise donc le premier vecteur propre de la matrice de covariance estimée comme observation. Les angles d’arrivées des deux sources par rapport à l’axe perpendiculaire aux capteurs sont $\theta_1 = -6.9^\circ$ et $\theta_2 = 6.9^\circ$. Le coefficient de corrélation entre les sources est de 0.9, le RSB est égal à -14 dB, et le nombre d’échantillons pour l’estimation de la matrice de covariance est fixé à $K = 1000$. $\lambda_2^-(\mathbf{y})$ et $\lambda_2^+(\mathbf{y})$ sont calculés pour une réalisation de l’observation \mathbf{y} donnée. La figure 4.4 représente les pseudos-spectres du vecteur $\hat{\mathbf{u}}$ obtenus avec l’algorithme FB-CEL0 selon λ , initialisé selon le filtre adapté spatial donné en (4.15). Les observations sont similaires par rapport au scénario précédent. On notera pour $\lambda = \lambda_2^-$ la présence d’un 3ième pic qui ne correspond donc à aucune source : cependant, son amplitude à -40 dB ne représente aucun risque d’ambiguïté.

4.2.3 Intérêt de l’intervalle pour la minimisation du critère J_{ℓ_0}

L’étude présentée ici montre que l’intervalle théorique $\lambda_M^-(\mathbf{y})$ et $\lambda_M^+(\mathbf{y})$ donnant l’équivalence entre les problèmes régularisé et contraint permet d’obtenir en pratique des résultats très prometteurs pour le choix du paramètre de régularisation λ à utiliser lors de la minimisation du critère J_{ℓ_0} par des algorithmes sous-optimaux pour estimer les directions

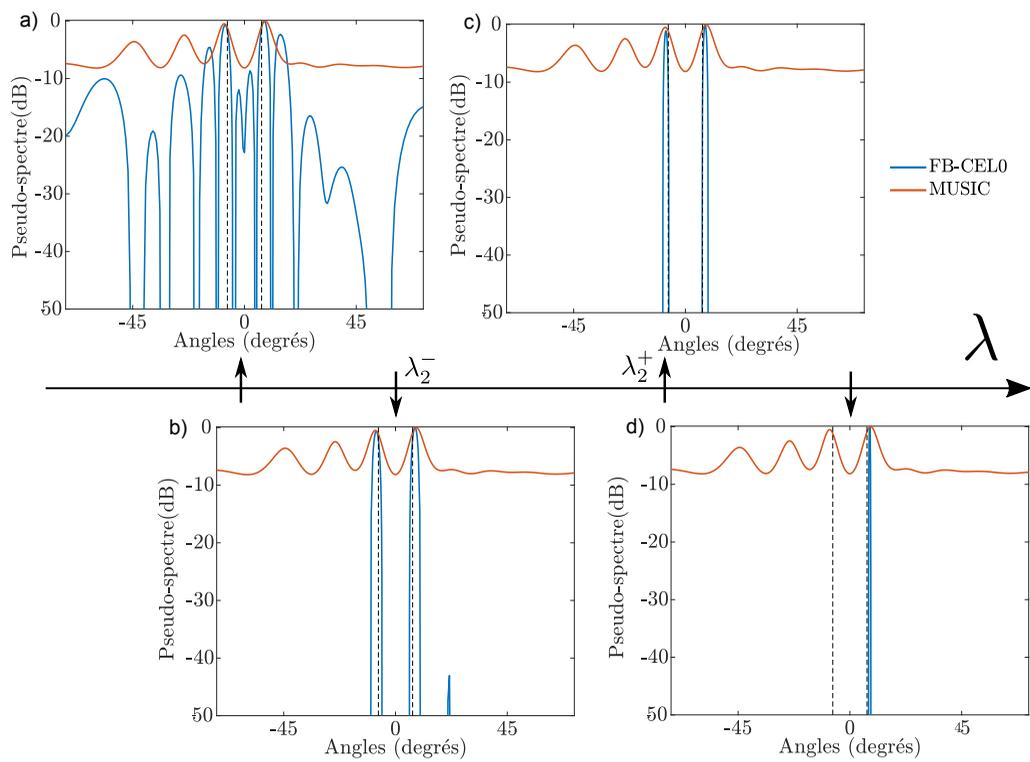


FIGURE 4.4 – Pseudo-spectres du vecteur $\hat{\mathbf{u}}$ obtenus par l'algorithme FB-CEL0 en fonction de λ , comparés au pseudo-spectre de MUSIC pour un réseau ULA à 10 antennes (scénario 1b).

d'arrivées des sources incidentes. En particulier, et comme attendu, si λ est plus faible que la borne basse de l'intervalle, la solution obtenue par les algorithmes se rapproche d'une solution pouvant être obtenue par les moindres carrés seuls : le vecteur solution n'est pas parcimonieux, et le pseudo-spectre présente de multiples lobes secondaires qui peuvent générer des ambiguïtés. Une valeur de λ située dans l'intervalle $] \lambda_2^-(\mathbf{y}); \lambda_2^+(\mathbf{y}) [$ en présence de 2 sources permet d'obtenir deux lobes plus ou moins larges selon l'endroit où l'on se situe dans l'intervalle. Pour une valeur de λ supérieure à la borne haute de l'intervalle, le vecteur solution obtenu est tellement parcimonieux que la contribution d'une des deux sources disparaît : on se situe en effet dans l'intervalle $] \lambda_1^-(\mathbf{y}); \lambda_1^+(\mathbf{y}) [$, donc la solution optimale n'a qu'une seule composante non nulle. Pour λ supérieur à un certain seuil donné par $\lambda_1^+(\mathbf{y}) = \lambda_0^-(\mathbf{y})$, la solution optimale est d'ailleurs le vecteur nul.

La connaissance des paramètres $\lambda_M^-(\mathbf{y})$ et $\lambda_M^+(\mathbf{y})$ apparait donc essentielle. Cependant, la valeur de ces paramètres dépend d'une réalisation bruitée de l'observation \mathbf{y} : ce sont donc des variables aléatoires. De plus, leur calcul est combinatoire : il nécessite de résoudre les problèmes contraints (\mathcal{C}_ϵ) pour chaque degré de parcimonie s jusqu'à avoir $c_s(\mathbf{y}) = 0$. Calculer les bornes de l'intervalle pour une réalisation donnée avant d'effectuer la minimisation de J_{ℓ_0} serait donc absurde. Comme il n'est pas envisageable de changer λ à chaque réalisation, il est nécessaire d'étudier les statistiques des bornes $\lambda_M^-(\mathbf{y})$ et $\lambda_M^+(\mathbf{y})$. Nous réalisons cette étude dans le cas $M = 2$ sources afin de trouver une plage commune, avec une probabilité élevée, à de multiples réalisations.

4.3 Intervalle pour des observations stochastiques d'un scénario bi-sources donné

4.3.1 Enjeux et motivations

Nous recherchons un intervalle admissible pour le paramètre de régularisation λ qui soit commun à de multiples réalisations. La recherche d'un intervalle commun doit passer par l'étude des distributions de λ_2^- et λ_2^+ . De plus, l'utilité pratique d'un tel intervalle doit être validée par une étude statistique du comportement des algorithmes en fonction de λ .

Dans un premier temps, nous étudions statistiquement le comportement de l'algorithme IRL1-CEL0 en fonction de λ pour le scénario 1a décrit précédemment en 4.2.1, pour des directions d'arrivées fixées $\theta_1 = 40^\circ$ et $\theta_2 = 100^\circ$.

Suite à l'étude du paragraphe 4.2, on note qu'il est difficile d'obtenir exactement deux uniques composantes non nulles même dans l'intervalle $] \lambda_2^-(\mathbf{y}); \lambda_2^+(\mathbf{y}) [$ sans doute à cause de difficultés de convergence des algorithmes sous-optimaux utilisés, d'autant plus que des colonnes voisines sont fortement corrélées. Dans la suite, l'estimation des directions d'arrivées de $M = 2$ sources à partir du vecteur estimé sera effectuée de la manière suivante :

- on s'intéresse aux différents pics, c'est à dire aux maxima locaux des différents lobes

dans le pseudo-spectre du vecteur estimé ;

- si le nombre de pics est trop important (on fixe ce nombre à 6 par exemple pour $M = 2$ sources), on considère qu'on ne peut décider quels pics correspondent à des sources et on échoue à détecter les directions d'arrivées ;
- s'il n'y a qu'un seul pic ou que la solution obtenue est le vecteur nul, alors on considère également que l'algorithme a échoué ;
- sinon, on retient les indices des deux maxima les plus élevés, et les directions d'arrivées estimées sont les angles correspondants dans Ψ .

Dans la pratique, ce principe simple de détection s'avère suffisant pour les réseaux et les scénarios bi-sources considérés. Ainsi, les pourcentages d'erreurs et REQM présentés dans l'ensemble de cette thèse utilisent ce post-traitement pour estimer les directions d'arrivées des sources à partir des vecteurs estimés.

Les performances obtenues pour l'algorithme IRL1-CEL0 initialisé par le vecteur nul sont mises en relation avec les histogrammes empiriques de λ_2^- et λ_2^+ obtenus pour les 1000 tirages réalisés. Les résultats sont représentés sur la figure 4.5. Les histogrammes de λ_2^- et λ_2^+ se rapprochent quand le RSB diminue, jusqu'à se chevaucher. La REQM obtenue par l'algorithme IRL1-CEL0 atteint un minimum sous l'histogramme de λ_2^- , et augmente dès que λ dépasse des valeurs potentielles de λ_2^+ .

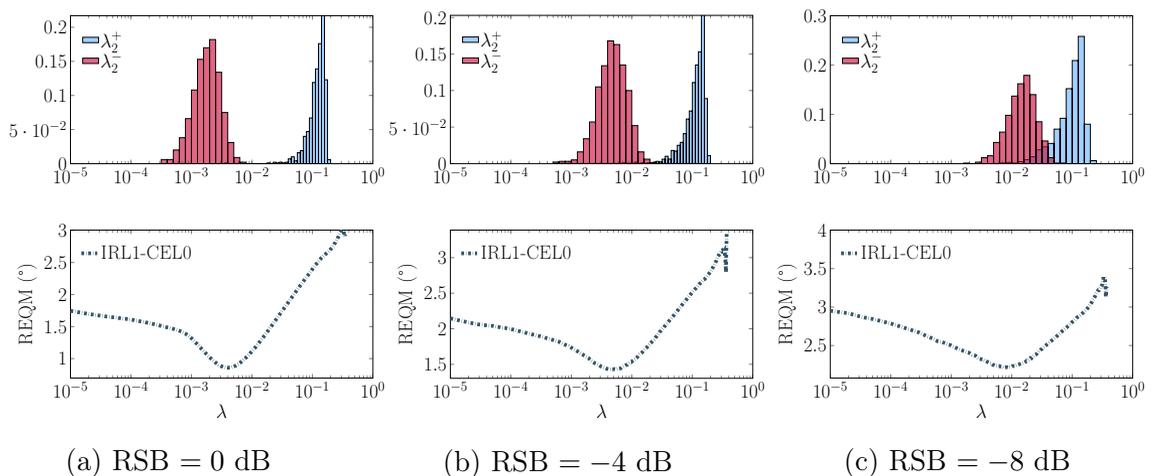


FIGURE 4.5 – Histogrammes empiriques de λ_2^- et λ_2^+ pour des observations stochastiques et REQM de l'algorithme IRL1-CEL0 (initialisation à 0) en fonction de λ pour deux sources fortement corrélées à BT fini.

Pour un RSB donné, nous choisissons de prendre comme valeur de λ à utiliser dans les algorithmes la valeur médiane des λ_2^- obtenus de façon combinatoire. Sur la figure 4.6, nous comparons les performances d'estimation des angles d'arrivées de l'algorithme IRL1-CEL0 initialisé par le vecteur nul pour cette valeur de λ avec les performances de l'algorithme MUSIC et la borne de Cramer-Rao en fonction du RSB, pour un écart entre les sources de 60° . Pour les RSB faibles considérés, la REQM obtenue avec l'algorithme

MUSIC est éloignée de la borne de Cramer-Rao. Le pourcentage d'erreurs de détection est d'ailleurs non nul. L'utilisation de l'algorithme IRL1-CEL0 permet de réduire sensiblement la REQM, avec un pourcentage d'erreurs qui reste très faible jusqu'à -10 dB.

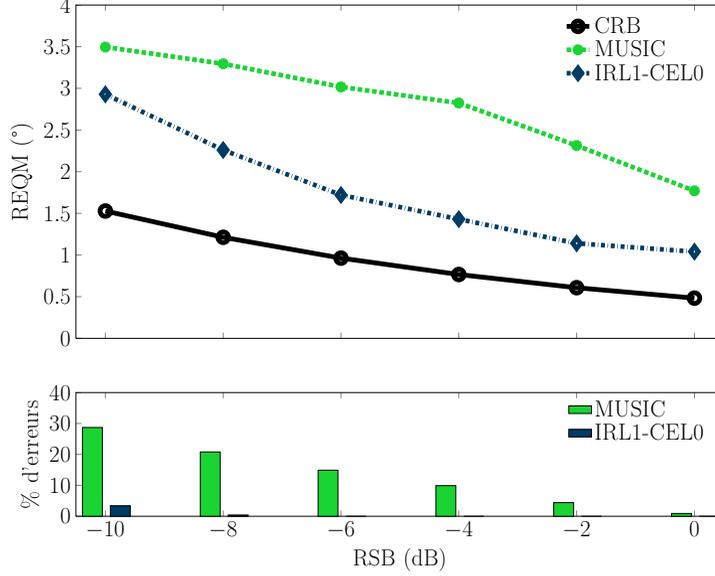


FIGURE 4.6 – Performances de l'algorithme IRL1-CEL0 initialisé par le vecteur nul en fonction du RSB dans le cadre du scénario 1a de sources fortement corrélées à BT fini : comparaison à MUSIC et à la borne de Cramer-Rao.

Cette étude préalable prouve à la fois l'intérêt qu'il faut porter aux distributions des bornes λ_2^- et λ_2^+ et plus globalement pour les méthodes parcimonieuses, et en particulier pour les algorithmes à gradient-proximal permettant de minimiser le critère régularisé J_{ℓ_0} .

Cependant, l'obtention de distributions à partir des équations (4.13) et (4.14), de façon empirique, est extrêmement lourd en temps de calcul, ce qui ne permet pas par exemple d'étudier l'intervalle stochastique pour n'importe quelles directions d'arrivées. C'est pourquoi j'ai étudié les statistiques de λ_2^- et λ_2^+ en fonction des statistiques de l'observation \mathbf{y} . Dans un premier temps, l'étude est réalisée pour un scénario de sources donné.

4.3.2 Calcul des statistiques de λ_2^- et λ_2^+ en fonction des statistiques de l'observation \mathbf{y} pour un scénario donné

4.3.2.1 Expressions approximées de $\lambda_2^-(\mathbf{y})$ et $\lambda_2^+(\mathbf{y})$

Dans le cas $M = 2$ sources, les expressions de $\lambda_2^-(\mathbf{y})$ et $\lambda_2^+(\mathbf{y})$ s'écrivent :

$$\lambda_2^-(\mathbf{y}) = \max \left\{ \frac{c_2(\mathbf{y}) - c_3(\mathbf{y})}{2}, \frac{c_2(\mathbf{y}) - c_4(\mathbf{y})}{4}, \dots, \frac{c_2(\mathbf{y}) - c_{N'}(\mathbf{y})}{2(N' - 2)} \right\}, \quad (4.16)$$

$$\lambda_2^+(\mathbf{y}) = \min \left\{ \frac{c_1(\mathbf{y}) - c_2(\mathbf{y})}{2}, \frac{c_0(\mathbf{y}) - c_2(\mathbf{y})}{4} \right\}. \quad (4.17)$$

On rappelle que

$$c_1(\mathbf{y}) = \min_{\psi_g} \mathbf{y}^H \mathbf{\Pi}_n(\{\psi_g\}) \mathbf{y}, \quad (4.18)$$

où $\psi_g \in \Psi$ est un point du maillage. On considère, en présence de $M = 2$ sources incidentes de directions d'arrivées θ_1 et θ_2 , que

$$c_2(\mathbf{y}) \approx \mathbf{y}^H \mathbf{\Pi}_n(\{\theta_1, \theta_2\}) \mathbf{y}. \quad (4.19)$$

On suppose par ailleurs

$$c_3(\mathbf{y}) \approx \min_{\psi_g \neq \theta_1, \theta_2} \mathbf{y}^H \mathbf{\Pi}_n(\{\theta_1, \theta_2, \psi_g\}) \mathbf{y}. \quad (4.20)$$

Ces deux hypothèses requièrent de faibles conditions sur le réseau de capteurs utilisé. L'ajout de composantes supplémentaires dans la projection ne fait pas diminuer significativement les valeurs $c_4(\mathbf{y})$, $c_5(\mathbf{y})$ etc. puisque l'on se situe au niveau du bruit. En première approximation, on peut donc faire l'hypothèse que :

$$\lambda_2^-(\mathbf{y}) \approx \frac{c_2(\mathbf{y}) - c_3(\mathbf{y})}{2}. \quad (4.21)$$

De même, on fait l'hypothèse que :

$$\lambda_2^+(\mathbf{y}) \approx \frac{c_1(\mathbf{y}) - c_2(\mathbf{y})}{2}. \quad (4.22)$$

Les deux bornes $\lambda_2^-(\mathbf{y})$ et $\lambda_2^+(\mathbf{y})$ peuvent alors s'exprimer comme le maximum et le minimum, respectivement, de formes quadratiques hermitiennes :

$$\lambda_2^-(\mathbf{y}) \approx \frac{1}{2} \max_{g|\psi_g \neq \theta_1, \theta_2} Q_g^-(\mathbf{y}) \quad (4.23)$$

$$\lambda_2^+(\mathbf{y}) \approx \frac{1}{2} \min_g Q_g^+(\mathbf{y}) \quad (4.24)$$

avec

$$Q_g^-(\mathbf{y}) = \mathbf{y}^H (\mathbf{\Pi}_n(\{\theta_1, \theta_2\}) - \mathbf{\Pi}_n(\{\theta_1, \theta_2, \psi_g\})) \mathbf{y} \quad (4.25)$$

$$Q_g^+(\mathbf{y}) = \mathbf{y}^H (\mathbf{\Pi}_n(\psi_g) - \mathbf{\Pi}_n(\{\theta_1, \theta_2\})) \mathbf{y}. \quad (4.26)$$

$Q_g^-(\mathbf{y})$ et $Q_g^+(\mathbf{y})$ sont des formes quadratiques hermitiennes de la variable aléatoire \mathbf{y} . On note qu'une forme quadratique hermitienne de vecteurs complexes a une valeur réelle. $Q_g^-(\mathbf{y})$ et $Q_g^+(\mathbf{y})$ sont donc elles-mêmes des variables aléatoires réelles. Afin d'exploiter les nombreux résultats sur les distributions de formes quadratiques de variables normales, il faut transformer les formes quadratiques hermitiennes en formes quadratiques réelles.

4.3.2.2 Transformation d'une forme quadratique hermitienne en forme quadratique réelle

Dans le chapitre 3, j'ai détaillé les différents modèles d'observations considérés et les statistiques associées. Pour chacune des observations \mathbf{y} selon le scénario, j'ai montré que \mathbf{y} suit une distribution normale multivariée complexe : $\mathbf{y} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{C})$.

On considère une forme quadratique hermitienne générique de la variable aléatoire \mathbf{y} notée

$$Q(\mathbf{y}) = \mathbf{y}^H \mathbf{P} \mathbf{y}, \quad (4.27)$$

avec \mathbf{P} une matrice hermitienne. On décompose la partie réelle et imaginaire de \mathbf{y} , et on définit ainsi le vecteur réel suivant :

$$\mathbf{y}_R = [\Re(\mathbf{y})^T, \Im(\mathbf{y})^T]^T, \quad (4.28)$$

où $\Re(\mathbf{y})$ est la partie réelle et $\Im(\mathbf{y})$ la partie imaginaire du vecteur \mathbf{y} . Pour $\mathbf{y} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{C})$, des calculs élémentaires montrent que \mathbf{y}_R suit une loi normale multivariée $\mathcal{N}(\boldsymbol{\mu}_R, \boldsymbol{\Gamma}_R)$, avec

$$\boldsymbol{\mu}_R = [\Re(\boldsymbol{\mu})^T, \Im(\boldsymbol{\mu})^T]^T \quad (4.29)$$

$$\boldsymbol{\Gamma}_R = \frac{1}{2} \begin{pmatrix} \Re(\boldsymbol{\Gamma} + \mathbf{C}) & \Im(-\boldsymbol{\Gamma} + \mathbf{C}) \\ \Im(\boldsymbol{\Gamma} + \mathbf{C}) & \Re(\boldsymbol{\Gamma} - \mathbf{C}) \end{pmatrix}. \quad (4.30)$$

La forme quadratique $Q(\mathbf{y})$ s'exprime alors comme

$$Q(\mathbf{y}) = \mathbf{y}_R^T \mathbf{P}_R \mathbf{y}_R \quad (4.31)$$

avec [52]

$$\mathbf{P}_R = \begin{pmatrix} (\mathbf{P} + \mathbf{P}^T)/2 & i(\mathbf{P} - \mathbf{P}^T)/2 \\ i(\mathbf{P}^T - \mathbf{P})/2 & (\mathbf{P} + \mathbf{P}^T)/2 \end{pmatrix}. \quad (4.32)$$

Pour \mathbf{P} hermitienne de dimension $N' \times N'$, la matrice \mathbf{P}_R de dimension $2N' \times 2N'$ est symétrique.

4.3.2.3 Distribution approchée de la forme quadratique d'une variable normale

On considère une forme quadratique $Q(\mathbf{y}) = \mathbf{y}^H \mathbf{P} \mathbf{y} = \mathbf{y}_R^T \mathbf{P}_R \mathbf{y}_R$, avec $\mathbf{y}_R \sim \mathcal{N}(\boldsymbol{\mu}_R, \boldsymbol{\Gamma}_R)$, et $\mathbf{P}_R \in \mathbb{R}^{2N' \times 2N'}$ une matrice symétrique. On suppose de plus que \mathbf{P}_R est définie non-négative.

Supposons que $\boldsymbol{\Gamma}_R$ est non singulière. On note alors $\boldsymbol{\Gamma}_R^{\frac{1}{2}}$ la matrice racine carrée telle que $\boldsymbol{\Gamma}_R^{\frac{1}{2}} \boldsymbol{\Gamma}_R^{\frac{1}{2}} = \boldsymbol{\Gamma}_R$. Soit \mathbf{M} une matrice orthogonale carrée de dimension $2N' \times 2N'$ qui

diagonalise $\mathbf{\Gamma}_R^{\frac{1}{2}} \mathbf{P}_R \mathbf{\Gamma}_R^{\frac{1}{2}}$. On a alors

$$\text{diag}(\nu_1, \dots, \nu_{2N'}) = \mathbf{M} \mathbf{\Gamma}_R^{\frac{1}{2}} \mathbf{P}_R \mathbf{\Gamma}_R^{\frac{1}{2}} \mathbf{M}^T \quad (4.33)$$

avec $\nu_1, \dots, \nu_{2N'}$ les valeurs propres de $\mathbf{\Gamma}_R^{\frac{1}{2}} \mathbf{P}_R \mathbf{\Gamma}_R^{\frac{1}{2}}$, ou, de façon équivalente, de $\mathbf{\Gamma}_R \mathbf{P}_R$.

Le vecteur $\mathbf{z} = \mathbf{M} \mathbf{\Gamma}_R^{-1/2} \mathbf{y}_R$ suit une loi de distribution normale de moyenne $\boldsymbol{\mu}_z = \mathbf{M} \mathbf{\Gamma}_R^{-1/2} \boldsymbol{\mu}_R$ et de variance $\mathbf{I}_{2N'}$. Soit $r_1 = \text{rank}(\mathbf{P}_R \mathbf{\Gamma}_R)$. On peut alors exprimer $Q(\mathbf{y})$ sous la forme

$$Q(\mathbf{y}) = \sum_{i=1}^{r_1} \nu_i \chi_{h_i}^2(\delta_i) \quad (4.34)$$

avec $h_i = 1$, $\delta_i = \mu_{zi}^2$ avec μ_{zi} la i -ème composante du vecteur $\boldsymbol{\mu}_z$, et où $\chi_{h_i}^2(\delta_i)$ représente des variables du χ^2 non centrés indépendantes ayant h_i degrés de liberté et un paramètre de décentralisation δ_i .

Le problème d'estimation de la queue de probabilité de $Q(\mathbf{y})$

$$\Pr(Q(\mathbf{y}) > t) \quad (4.35)$$

se pose dans de nombreuses applications. Plusieurs approximations basées sur les moments ont été proposées. Pearson [61] a initialement proposé d'approximer la distribution de $Q(\mathbf{y})$ en utilisant une distribution du χ^2 centrée en se basant sur les 3 premiers moments. Liu et al. [46] ont proposé beaucoup plus récemment d'utiliser une distribution du χ^2 non centrée $\chi_l^2(\delta)$, dont le nombre de degrés de liberté l et le paramètre de décentralisation δ sont déterminés par les quatre premiers moments de $Q(\mathbf{y})$. L'équation (4.34) permet d'obtenir $K(t)$ la fonction génératrice des moments de $Q(\mathbf{y})$, donnée par [45] :

$$K(t) = \frac{1}{2} \sum_{i=1}^{r_1} h_i \log(1 - 2t\nu_i) + \sum_{i=1}^{r_1} \frac{\delta_i \nu_i t}{1 - 2\nu_i t} \quad (4.36)$$

Le k -ième cumulante de $Q(\mathbf{y})$ se calcule par

$$\kappa_k = 2^{k-1} (k-1)! \left(\sum_{i=1}^{r_1} \nu_i^k h_i + k \sum_{i=1}^{r_1} \nu_i^k \delta_i \right) \quad (4.37)$$

Il n'est par ailleurs pas nécessaire de calculer les valeurs propres ν_i et les paramètres δ_i , puisque l'on a [46]

$$\begin{aligned} \sum_{i=1}^{r_1} \nu_i^k h_i &= \text{Tr}(\text{diag}(\nu_1, \dots, \nu_{2N'})^k) \\ &= \text{Tr}((\mathbf{M} \mathbf{\Gamma}_R^{\frac{1}{2}} \mathbf{P}_R \mathbf{\Gamma}_R^{\frac{1}{2}} \mathbf{M}^T)^k) \\ &= \text{Tr}((\mathbf{\Gamma}_R^{\frac{1}{2}} \mathbf{P}_R \mathbf{\Gamma}_R^{\frac{1}{2}})^k) \\ &= \text{Tr}((\mathbf{P}_R \mathbf{\Gamma}_R)^k) \end{aligned} \quad (4.38)$$

et

$$\begin{aligned}
\sum_{i=1}^{r_1} \nu_i^k \delta_i &= \boldsymbol{\mu}_z^T \text{diag}(\nu_1, \dots, \nu_{2N'})^k \boldsymbol{\mu}_z \\
&= \boldsymbol{\mu}_R^T \boldsymbol{\Gamma}_R^{-1/2} \mathbf{M}^T (\mathbf{M} \boldsymbol{\Gamma}_R^{\frac{1}{2}} \mathbf{P}_R \boldsymbol{\Gamma}_R^{\frac{1}{2}} \mathbf{M}^T)^k \mathbf{M} \boldsymbol{\Gamma}_R^{-1/2} \boldsymbol{\mu}_R \\
&= \boldsymbol{\mu}_R^T (\boldsymbol{\Gamma}_R \mathbf{P}_R)^{k-1} \mathbf{P}_R \boldsymbol{\mu}_R.
\end{aligned} \tag{4.39}$$

On note

$$\begin{aligned}
s_k &= \sum_{i=1}^{r_1} \nu_i^k h_i + k \sum_{i=1}^{r_1} \nu_i^k \delta_i \\
&= \text{Tr}((\mathbf{P}_R \boldsymbol{\Gamma}_R)^k) + k \boldsymbol{\mu}_R^T (\boldsymbol{\Gamma}_R \mathbf{P}_R)^{k-1} \mathbf{P}_R \boldsymbol{\mu}_R.
\end{aligned} \tag{4.40}$$

La moyenne, la variance, l'asymétrie et le kurtosis de $Q(\mathbf{y})$ sont alors donnés par :

$$\mu_Q = \kappa_1 = s_1 \tag{4.41}$$

$$\sigma_Q = \sqrt{\kappa_2} = \sqrt{2s_2} \tag{4.42}$$

$$\beta_1 = \frac{\kappa_3}{\kappa_2^{3/2}} = \sqrt{8} \frac{s_3}{s_2^{3/2}} = \sqrt{8} b_1 \tag{4.43}$$

$$\beta_2 = \frac{\kappa_4}{\kappa_2^2} = 12 \frac{s_4}{s_2^2} = 12b_2 \tag{4.44}$$

avec $b_1 = \frac{s_3}{s_2^{3/2}}$ et $b_2 = \frac{s_4}{s_2^2}$. On approxime ainsi la distribution de queue, par :

$$\Pr(Q(\mathbf{y}) > t) \approx \Pr\left(\frac{\chi_l^2(\delta) - \mu_\chi}{\sigma_\chi} > \frac{t - \mu_Q}{\sigma_Q}\right) \tag{4.45}$$

avec

$$\mu_\chi = \text{E}\{\chi_l^2(\delta)\} = l + \delta \tag{4.46}$$

$$\sigma_\chi = \sqrt{\text{var}(\chi_l^2(\delta))} = \sqrt{2}a \tag{4.47}$$

et

$$a = \sqrt{l + 2\delta} \tag{4.48}$$

Les paramètres δ et l sont déterminés de telle sorte que l'asymétrie de $Q(\mathbf{y})$ et $\chi_l^2(\delta)$ soit égale et que la différence entre le kurtosis de $Q(\mathbf{y})$ et $\chi_l^2(\delta)$ soit minimale. Les calculs mènent à [46] :

– si $b_1^2 > b_2$, alors

$$a = 1/\left(b_1 - \sqrt{b_1^2 - b_2}\right), \quad \delta = b_1 a^3 - a^2 \quad \text{et} \quad l = a^2 - 2\delta \tag{4.49}$$

– si $b_1^2 \leq b_2$, alors

$$a = 1/b_1, \quad \delta = b_1 a^3 - a^2 = 0 \quad \text{et} \quad l = a^2 - 2\delta = 1/b_1^2 \tag{4.50}$$

On utilise l'équation (4.45) pour calculer les fonctions de densité (marginales) des variables aléatoires $Q_g^-(\mathbf{y})$ et $Q_g^+(\mathbf{y})$, pour $g = \{1, \dots, G\}$. Dans la suite, les fonctions de

densité sont notées f_g^- et f_g^+ , respectivement. Les fonctions de répartition associées sont notées F_g^- et F_g^+ .

4.3.2.4 Utilisation de copules pour exprimer la dépendance entre des variables aléatoires non gaussiennes

Les paragraphes précédents 4.3.2.2 et 4.3.2.3 permettent de calculer les distributions des variables aléatoires $Q_g^-(\mathbf{y})$ et $Q_g^+(\mathbf{y})$ des équations (4.23) et (4.24) indépendamment les unes des autres pour $g = \{1, \dots, G\}$. Or ces variables ne sont pas indépendantes, puisqu'elles sont issues des mêmes vecteurs aléatoires \mathbf{y} .

De façon générique, notons $Q_i(\mathbf{y}) = \mathbf{y}^H \mathbf{P}_i \mathbf{y}$ et $Q_j(\mathbf{y}) = \mathbf{y}^H \mathbf{P}_j \mathbf{y}$, avec $\mathbf{y} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \mathbf{C})$ et \mathbf{P}_i et \mathbf{P}_j des matrices hermitiennes. La covariance entre les variables Q_i et Q_j se calcule par :

$$\begin{aligned} Cov(Q_i, Q_j) = & \text{tr} \left((\mathbf{P}_i \mathbf{C}) ((\mathbf{P}_j)^T \mathbf{C}^*) \right) + \text{tr} \left((\mathbf{P}_i \boldsymbol{\Gamma}) (\mathbf{P}_j \boldsymbol{\Gamma}) \right) + \boldsymbol{\mu}^H \mathbf{P}_i \boldsymbol{\Gamma} \mathbf{P}_j \boldsymbol{\mu} \\ & + \boldsymbol{\mu}^H \mathbf{P}_j \boldsymbol{\Gamma} \mathbf{P}_i \boldsymbol{\mu} + \boldsymbol{\mu}^H \mathbf{P}_i \mathbf{C} (\mathbf{P}_j)^T \boldsymbol{\mu}^* + \boldsymbol{\mu}^T (\mathbf{P}_j)^T \mathbf{C}^* \mathbf{P}_i \boldsymbol{\mu}. \end{aligned} \quad (4.51)$$

Afin de calculer les distribution de λ_2^- et λ_2^+ , il est nécessaire de calculer les distributions jointes $F_{\mathbf{q}^-}$ et $F_{\mathbf{q}^+}$ des variables multivariées $\mathbf{q}^-(\mathbf{y})$ et $\mathbf{q}^+(\mathbf{y})$ données par :

$$\mathbf{q}^-(\mathbf{y}) = [Q_1^-(\mathbf{y}), \dots, Q_g^-(\mathbf{y}), \dots, Q_G^-(\mathbf{y})]^T, \quad g \mid \psi_g \neq \theta_1, \theta_2 \quad (4.52)$$

$$\mathbf{q}^+(\mathbf{y}) = [Q_1^+(\mathbf{y}), Q_2^+(\mathbf{y}), \dots, Q_G^+(\mathbf{y})]^T \quad (4.53)$$

Les copules sont l'outil à ce jour le plus adapté à notre connaissance. Les copules sont des fonctions qui permettent de relier des fonctions de distribution multivariées aux fonctions de distributions marginales de leurs variables scalaires [55]. Elles permettent d'exprimer la dépendance entre des variables aléatoires de fonctions de distributions marginales variées. Soit \mathbf{c} une copule. La fonction de distribution multivariée $F_{\mathbf{q}^-}$ (respectivement $F_{\mathbf{q}^+}$) s'exprime en fonction des fonctions de répartition F_g^- (respectivement F_g^+) par :

$$F_{\mathbf{q}^-}(\varrho_1, \dots, \varrho_G) = \mathbf{c}(F_1^-(\varrho_1), \dots, F_G^-(\varrho_G)) \quad (4.54)$$

$$F_{\mathbf{q}^+}(\varrho_1, \dots, \varrho_G) = \mathbf{c}(F_1^+(\varrho_1), \dots, F_G^+(\varrho_G)). \quad (4.55)$$

La copule Gaussienne est particulièrement adaptée pour les grandes dimensions. En notant $F_g(\varrho_g) = F_g^-(\varrho_g)$ ou $F_g^+(\varrho_g)$, on a alors :

$$\mathbf{c}(F_1(\varrho_1), \dots, F_G(\varrho_G)) = \frac{1}{\sqrt{\det \boldsymbol{\Sigma}}} \exp \left(-\frac{1}{2} \boldsymbol{\beta}^T (\boldsymbol{\Sigma}^{-1} - \mathbf{I}) \boldsymbol{\beta} \right) \quad (4.56)$$

avec

$$\boldsymbol{\beta}^T = [\Phi^{-1}(F_1(\varrho_1)), \dots, \Phi^{-1}(F_G(\varrho_G))] \quad (4.57)$$

où Φ^{-1} est la fonction quantile réciproque de la fonction de répartition de la loi normale :

$$\Phi(x) = \frac{1}{2\pi} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt. \quad (4.58)$$

La dépendance entre les variables scalaires est modélisée par la matrice de corrélation symétrique et définie positive Σ , dont la ij -ième composante est donnée par :

$$\Sigma_{ij} = \frac{Cov(Q_i, Q_j)}{\sigma_{Q_i} \sigma_{Q_j}} \quad (4.59)$$

avec $\sigma_{Q_g} = \sqrt{Cov(Q_i, Q_i)}$, et la covariance entre deux variables Q_i et Q_j est donnée par l'équation (4.51).

4.3.2.5 Lois de distribution de λ_2^- et λ_2^+

A partir des fonctions de répartition $F_{\mathbf{q}^-}$ et $F_{\mathbf{q}^+}$ obtenues par les équations (4.54) et (4.55) et la définition de la copule Gaussienne (4.56), les fonctions de répartition des variables λ_2^- et λ_2^+ approximées selon les équations (4.23) et (4.24) s'expriment selon :

$$F_{\lambda_2^-}(t) = \Pr(\lambda_2^-(\mathbf{y}) < t) \quad (4.60)$$

$$= \Pr\left(\max_g \frac{1}{2} Q_g^-(\mathbf{y}) < t\right) \quad (4.61)$$

$$= \Pr\left(\frac{1}{2} Q_g^-(\mathbf{y}) < t, \forall g \mid \psi_g \in \{\Psi \setminus \theta_1, \theta_2\}\right) \quad (4.62)$$

$$= F_{\mathbf{q}^-}(2t, \dots, 2t) \quad (4.63)$$

et

$$F_{\lambda_2^+}(t) = \Pr(\lambda_2^+(\mathbf{y}) < t) \quad (4.64)$$

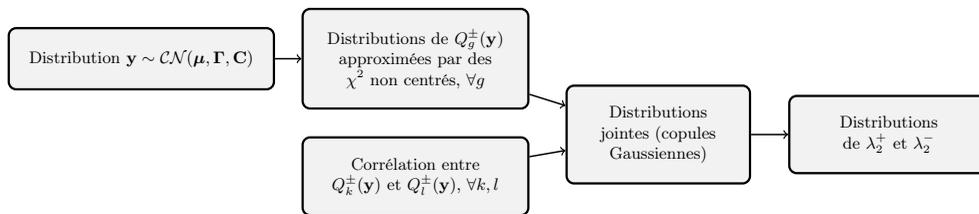
$$= \Pr\left(\min_g \frac{1}{2} Q_g^+(\mathbf{y}) < t\right) \quad (4.65)$$

$$= \Pr\left(\exists g, \psi_g \in \Psi \mid \frac{1}{2} Q_g^+(\mathbf{y}) < t\right) \quad (4.66)$$

$$= 1 - \Pr\left(\frac{1}{2} Q_g^+(\mathbf{y}) > t \mid \forall g, \psi_g \in \Psi\right) \quad (4.67)$$

La figure 4.7 résume les différentes étapes pour l'obtention des distributions.

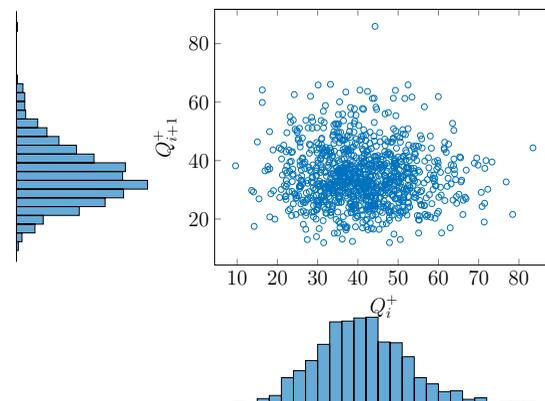
A noter qu'il n'existe pas d'expression analytique de la fonction de répartition Φ de la loi normale ni de sa fonction quantile réciproque Φ^{-1} . Les expressions (4.63) et (4.67) sont donc semi-analytiques. Il existe cependant de nombreuses méthodes d'approximations numériques très précises de Φ et Φ^{-1} , directement incluses dans les logiciels de calcul [27].

FIGURE 4.7 – Étapes de calcul des distributions approchées de λ_2^- et λ_2^+ .

4.3.2.6 Méthode de génération des lois de distribution et validation des approximations

Dans la pratique, j'ai généré des histogrammes et fonctions de répartition empiriques à l'aide de tirages aléatoires à partir des fonctions approchées proposées. Je détaille ici la méthodologie pour λ_2^+ : la méthodologie pour λ_2^- est similaire, à la dimension près puisque \mathbf{q}^- est de dimension $G - 2$ et non G .

1. On génère n_{tir} tirages indépendants de chaque variable aléatoire Q_g^+ pour $g = \{1, \dots, G\}$ selon des distributions du χ^2 non centrées selon l'équation (4.45). Ce tirage nécessite la connaissance des statistiques de \mathbf{y} . La figure 4.8 représente les histogrammes et le nuage de points 2D de 1000 tirages indépendants de deux variables Q_i^+ et Q_{i+1}^+ , donc pour deux directions d'arrivées voisines. La forme du nuage de points reflète l'absence évidente de corrélation.

FIGURE 4.8 – Nuage de points de 1000 tirages aléatoires indépendants de Q_i^+ et Q_{i+1}^+ selon des distributions du χ^2 non centrées et histogrammes associés.

2. On génère $n_{tir} \times G$ variables aléatoires corrélées en utilisant une copule Gaussienne, avec une matrice de corrélation $\boldsymbol{\Sigma}$ calculée selon l'équation (4.59). On note \mathbf{z}_g les G vecteurs aléatoires de dimension n_{tir} générés. La figure 4.9 représente les histogrammes et le nuage de points de 1000 tirages indépendants de deux variables z_i et z_{i+1} . Les histogrammes sont ceux de variables uniformes. Le nuage de points montre la forte corrélation entre ces deux variables.

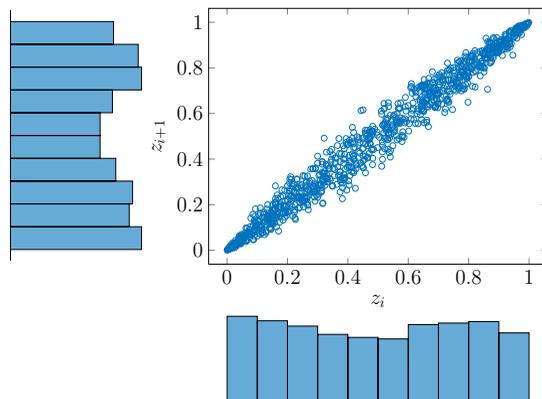


FIGURE 4.9 – Nuage de points de 1000 tirages aléatoires des variables z_i et z_{i+1} selon une copule Gaussienne.

3. On trie pour chaque G les variables générées par la copule : on récupère les vecteurs \mathbf{i}_g composés des indices 1 à G qui sont triés selon l'ordre croissant des composantes de \mathbf{z}_g .
4. Pour chaque g , dans un vecteur de dimension n_{tir} , on place les composantes des tirages de Q_g^+ par ordre croissant aux indices indiqués par \mathbf{i}_g . On obtient ainsi n_{tir} tirages aléatoires corrélés des variables Q_i^+ , donc du vecteur \mathbf{q}^+ . La figure 4.10 représente les histogrammes des deux variables Q_i^+ et Q_{i+1}^+ (qui sont identique à ceux de la figure 4.8) et le nuage de points des 1000 tirages maintenant corrélés.

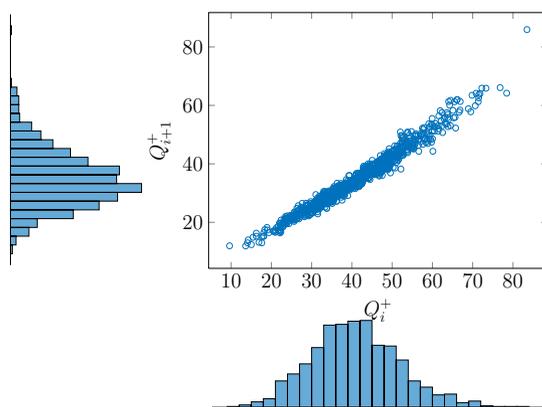


FIGURE 4.10 – Nuage de points des 1000 tirages aléatoires de Q_i^+ et Q_{i+1}^+ après étape de corrélation et histogrammes associés.

5. Pour chaque tirage, on retient la valeur minimale selon g , divisée par 2, pour obtenir n_{tir} tirages aléatoires de λ_2^+ .

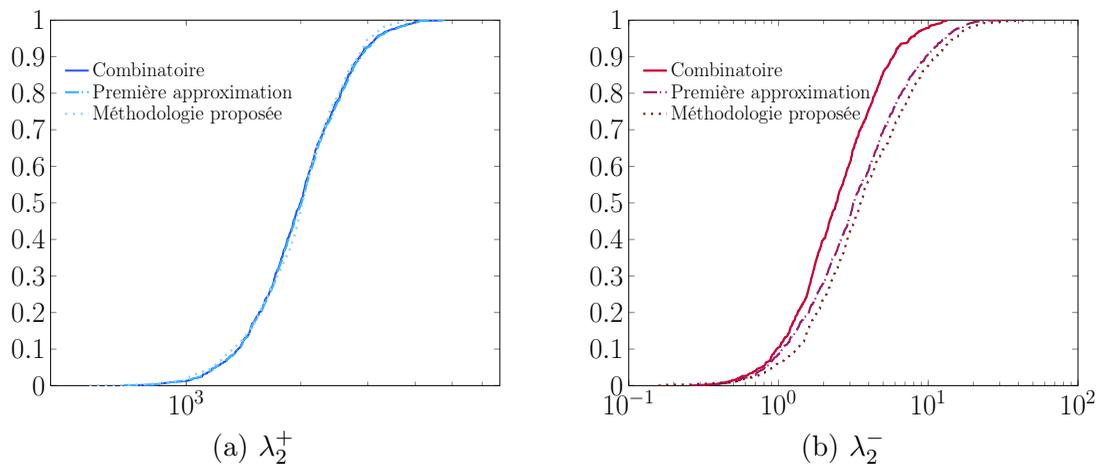


FIGURE 4.11 – Fonctions de répartition de λ_2^+ et λ_2^- . En traits pleins, les courbes obtenues de façon combinatoire. La ligne en tirets représente les fonctions obtenues en prenant en compte les premières approximations (4.21) et (4.22). La ligne en pointillées est celle obtenue avec les approximations successives du χ^2 non centré et des copules, selon la méthodologie proposée.

Cette méthodologie permet d'obtenir des distributions empiriques approchées de λ_2^- et λ_2^+ à partir des statistiques de l'observation \mathbf{y} , pour un scénario donné (puissance des sources et angles d'arrivées). La figure 4.11 représente les fonctions de répartition empiriques de λ_2^- et λ_2^+ obtenues selon trois méthodes différentes :

- la première courbe est obtenue de façon combinatoire selon les équations (4.16) et (4.17) à partir de tirages aléatoires du signal reçu $\mathbf{x}(t)$ (à BT fini), ou de la matrice de covariance \mathbf{R}_{xx} (en présence d'erreurs de modèles) ;
- la seconde courbe est obtenue en prenant en compte les approximations (4.21) et (4.22), à partir des mêmes tirages aléatoires du signal reçu $\mathbf{x}(t)$ (à BT fini), ou de la matrice de covariance \mathbf{R}_{xx} (en présence d'erreurs de modèles) ;
- la troisième courbe est obtenue selon la méthodologie proposée de tirages aléatoires de vecteurs corrélés \mathbf{q}^- et \mathbf{q}^+ .

Pour l'exemple de la figure 4.11, on utilise un réseau circulaire uniforme à $N = 7$ antennes avec $M = 2$ sources indépendantes d'angles incidents $\theta_1 = 32^\circ$ et $\theta_2 = 100^\circ$. L'observation utilisée est la matrice de covariance vectorisée à BT fini, pour $K = 50$ et un RSB égal à 10 dB. On remarque que les trois courbes pour λ_2^+ (figure 4.11(a)) sont presque confondues. La figure 4.11(b) laisse apparaître un écart un peu plus prononcé pour les courbes de λ_2^- , le principal écart étant du à la première approximation (4.21). On explique la différence entre λ_2^+ et λ_2^- par l'ordre des projections approximées.

L'avantage certain de la méthode proposée est le coût calculatoire très largement réduit par rapport à l'utilisation de la formule initiale. Le faible coût calculatoire permet de générer très facilement les distributions pour un grand nombre de scénarios (en particulier d'angles d'arrivées). On peut alors étudier l'existence d'un intervalle pour le paramètre

de régularisation λ qui soit commun à de multiples réalisations, d'abord sur un scénario d'angles donnés, puis potentiellement pour tous les angles d'arrivées possibles.

4.3.3 Existence d'un intervalle commun pour des observations stochastiques

Afin de montrer l'existence d'un intervalle commun pour des observations stochastiques, étudions le comportement et les performances des algorithmes en fonction de λ , en relation avec les statistiques de λ_2^- et λ_2^+ obtenues pour différents scénarios. Nous décrivons trois nouveaux scénarios de simulation, en plus des deux précédents décrits en 4.2.1 (scénario 1a) et 4.2.2 (scénario 1b). Pour ces scénarios, on s'intéresse en particulier à des réseaux circulaires uniformes (UCA). Bien que ces réseaux permettraient une estimation en deux dimensions, nous nous limitons dans le cadre de la thèse aux problèmes en une dimension (azimut uniquement), en faisant l'hypothèse que les sources et les capteurs sont coplanaires. Les propriétés d'invariance des réseaux UCA par rapport aux angles d'arrivées en mono-trajet nous permettent, pour $M = 2$ trajets, de nous concentrer sur l'écart entre les sources plutôt que sur les angles d'arrivées individuellement. Pour des sources d'angles d'arrivées θ_1 et θ_2 , nous nous intéressons donc au comportement des algorithmes selon l'écart $|\theta_2 - \theta_1|$.

Pour les scénarios 2 à 4 présentés dans ce paragraphe, on considère un réseau UCA à $N = 7$ antennes, de rayon une demi longueur d'onde. La largeur de lobe entre deux annulations de la corrélation est de 90° . Le champ de vision est l'intervalle $[0; 360]^\circ$, avec un pas de grille de 1° . On étudie les performances de différents algorithmes selon λ , en considérant 1000 tirages aléatoires. Les performances sont représentées en terme de REQM et de pourcentage d'erreurs tels qu'ils sont définis dans le paragraphe 1.2.3.

4.3.3.1 Sources cohérentes en présence d'erreurs de modèle (scénario 2)

On considère un scénario de sources cohérentes de même puissance : l'observation utilisée est le premier vecteur propre de la matrice de covariance en présence d'erreurs de modèles. Pour $\sigma_e = 0.15$, les histogrammes de λ_2^- et λ_2^+ sont mis en relation avec les performances de l'algorithme IRL1-CEL0 (initialisé par le vecteur nul) sur la figure 4.12 pour deux écarts différents : sur la sous-figure (a), $|\theta_2 - \theta_1| = 48^\circ$ (écart supérieur à une demi-largeur de lobe, corrélation spatiale $r_{\mathbf{D}}^2 \approx 0$) et sur la sous-figure (b) $|\theta_2 - \theta_1| = 28^\circ$ (écart inférieur à une demi-largeur de lobe, corrélation spatiale $r_{\mathbf{D}}^2 \approx 0.25$).

Pour un écart entre les sources de 48° - figure 4.12(a) - les histogrammes de λ_2^+ et λ_2^- sont bien séparés. Parallèlement, la courbe de pourcentage d'erreurs montre qu'il existe un intervalle pour λ , délimité par les traits pointillés verticaux, dans lequel l'estimation des deux directions d'arrivées est parfaitement réalisée puisqu'avec 0% d'erreurs. Cet intervalle correspond environ à l'intervalle entre les histogrammes : il commence légèrement en dessous de la borne supérieure de λ_2^- , et finit dès que l'on atteint la borne inférieure

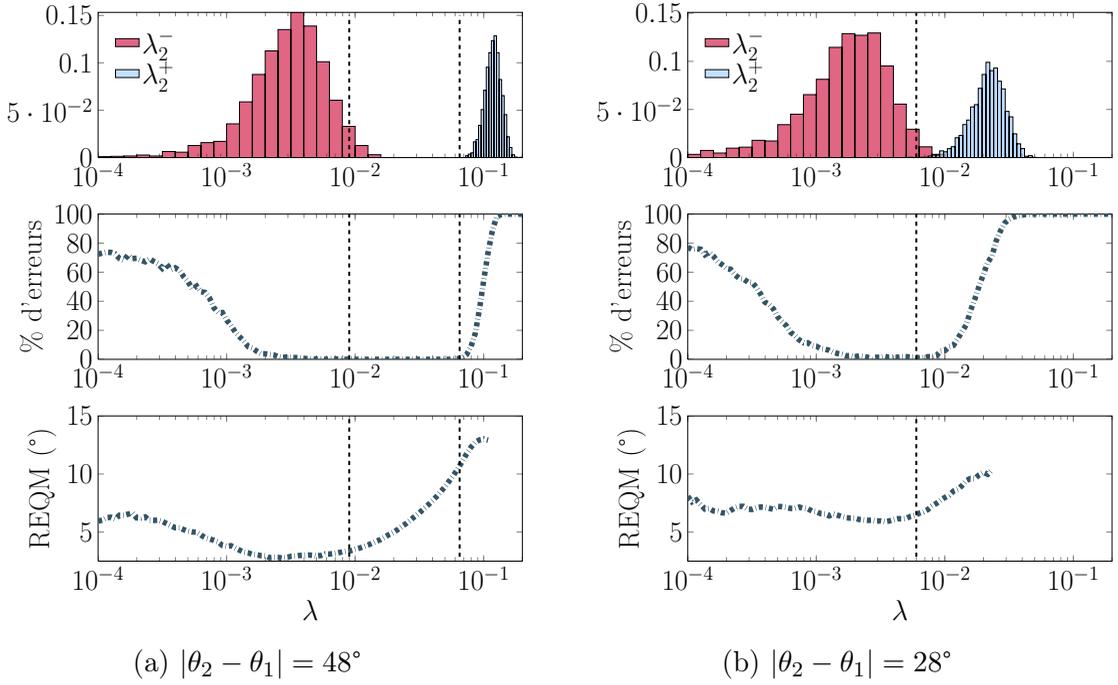


FIGURE 4.12 – Performances de l'algorithme IRL1-CEL0 (initialisé à 0) en fonction de λ pour deux sources cohérentes et avec erreurs de modèles ($\sigma_e = 0.15$). En haut : histogrammes de λ_2^- et λ_2^+ pour des observations stochastiques. Milieu et bas : pourcentage d'erreurs et REQM.

de λ_2^+ . Pour un écart de 28° - figure 4.12(b) - les histogrammes de λ_2^+ et λ_2^- commencent à se chevaucher. Cela signifie qu'on ne peut trouver une unique valeur de λ telle que la solution optimale de J_{ℓ_0} pour tous les tirages corresponde à la solution du maximum de vraisemblance avec deux sources. Il n'y a en parallèle plus d'intervalle de valeurs de λ pour lequel on a 0% d'erreurs : le trait vertical pointillé correspond au minimum de pourcentage d'erreurs qui est de 1%. Ce point correspond à la borne inférieure de λ_2^+ .

L'augmentation du pourcentage d'erreurs à partir de la borne inférieure de λ_2^+ correspond par ailleurs tout à fait avec l'allure de l'histogramme de λ_2^+ . Cela s'explique par le fait que lorsque λ est supérieur à $\lambda_2^+(\mathbf{y})$, le minimum global de J_{ℓ_0} ne correspond plus au minimum du maximum de vraisemblance considérant $M = 2$ sources. Le minimiseur global n'a en effet plus qu'une ou zéro composante non nulle. En revanche, le faible pourcentage d'erreurs sous l'histogramme de λ_2^- s'explique par l'étape d'estimation des angles à partir du vecteur estimé qui est décrite dans le paragraphe 4.3.1. On peut avoir plus de $M = 2$ composantes non nulles dans le vecteur solution tout en ayant que deux lobes, dont on extrait les pics pour estimer les deux directions d'arrivées. Si λ est trop faible en revanche, les lobes secondaires trop nombreux ou trop élevés ne nous permettent plus d'estimer avec certitude les deux directions d'arrivées.

4.3.3.2 Sources indépendantes pour un temps d'intégration fini (scénario 3)

Pour ce scénario de simulation, on considère $M = 2$ sources indépendantes, et on utilise donc comme observation la matrice de covariance vectorisée, estimée avec $K = 50$ échantillons du signal reçu. L'écart entre les sources est de 68° . On étudie le comportement de l'algorithme pour différents scénarios de puissance : dans un premier temps, les deux sources ont la même puissance avec un RSB = 10 dB. On augmente ensuite le RSB de la seconde source à 12 dB, 20 dB et 22 dB. Les résultats obtenus avec l'algorithme IRL1-CEL0 initialisé par le vecteur nul sont présentés sur la figure 4.13.

Les histogrammes de λ_2^- et λ_2^+ sont bien disjoints pour les deux sources de même puissance sur la sous-figure (a). Quand on augmente la puissance de la seconde source, l'histogramme de λ_2^- se rapproche de celui de λ_2^+ , jusqu'à ce que les deux histogrammes se chevauchent. Sur les figures, on trace en traits noirs pointillés verticaux les limites de l'intervalle de λ pour lequel on a 0% d'erreurs, quand il existe, ou la valeur pour laquelle le pourcentage d'erreurs est minimal. La relation entre pourcentage d'erreurs et histogrammes est identique à celle vue précédemment. Pour ce scénario, la valeur minimale de la REQM est également atteinte dans l'intervalle.

4.3.3.3 Sources indépendantes en présence d'erreurs de modèles (scénario 4)

Pour ce dernier scénario de simulation, on considère $M = 2$ sources indépendantes, et on utilise donc comme observation la matrice de covariance vectorisée mais en présence d'erreurs de modèles. Les deux sources ont la même puissance. On fixe $\sigma_e = 0.2$. On étudie le comportement de l'algorithme IRL1-CEL0, initialisé par le vecteur nul, et de l'algorithme IHT initialisé par la solution de l'algorithme FB-CEL0, lui-même initialisé par le vecteur nul. En effet, la trop grande vitesse de convergence de l'algorithme IRL1-CEL0 en fait un mauvais candidat pour l'initialisation de l'algorithme IHT puisque celui-ci n'arrive pas à sortir du minimum dans lequel il est alors déjà. Les performances obtenues sont représentées sur la figure 4.14 pour différents écarts entre les sources. Sur la sous-figure (a), $|\theta_2 - \theta_1| = 40^\circ$; les histogrammes de λ_2^- et λ_2^+ sont alors bien séparés. Les histogrammes se rapprochent quand l'écart entre les sources diminue : sur la sous-figure (b), $|\theta_2 - \theta_1| = 30^\circ$, et sur la sous-figure (c), $|\theta_2 - \theta_1| = 20^\circ$. Pour des sources et histogrammes bien séparés, on retrouve l'intervalle de λ pour lequel l'algorithme IRL1-CEL0 permet d'obtenir 0% d'erreurs dans la détection et l'estimation des deux sources. Alors que les histogrammes commencent à peine à se chevaucher pour un écart de 20° , le pourcentage d'erreurs minimum atteint est de 4.5%. Le comportement de l'algorithme IHT initialisé par la solution de FB-CEL0 est moins classique. Il n'y a plus une exacte correspondance entre l'intervalle entre les histogrammes et le pourcentage d'erreurs nul pour des sources séparées. En revanche, pour des sources proches, sur la sous-figure 4.14(c), malgré l'intersection entre les deux histogrammes de λ_2^- et λ_2^+ , on retrouve un intervalle de λ pour lequel le pourcentage d'erreurs est inférieur à 1, avec par ailleurs une REQM très faible.

4.3.3.4 Conclusion sur l'existence d'un intervalle commun

Pour chacun des trois scénarios et observations associées, le constat est identique. Si les sources sont de même puissance et suffisamment éloignées, les histogrammes de λ_2^- et λ_2^+ sont séparés. On constate alors que l'algorithme IRL1-CEL0 initialisé par le vecteur nul permet d'obtenir 0% d'erreurs dans l'estimation des deux directions d'arrivées pour λ dans l'intervalle situé entre les deux histogrammes. Quand on rapproche les deux sources, ou qu'on introduit un écart de puissance entre deux sources éloignées, les histogrammes se rapprochent jusqu'à se chevaucher. Il n'y a alors plus d'intervalle pour λ tel que le pourcentage d'erreurs soit nul. Le minimum du pourcentage d'erreurs est atteint pour λ situé au milieu du recouvrement des histogrammes. Par ailleurs, le comportement de l'algorithme IHT initialisé par la solution de l'algorithme FB-CEL0 pour des sources proches semble intéressant.

On définit un intervalle admissible pour λ situé entre les 95 à 98% de la fonction de répartition de λ_2^- et 2 à 5% de celle de λ_2^+ , selon un scénario donné. Dans la suite, on étudie cet intervalle en fonction de l'écart entre les sources afin de permettre la prédétermination hors-ligne d'une valeur de λ à utiliser dans les algorithmes.

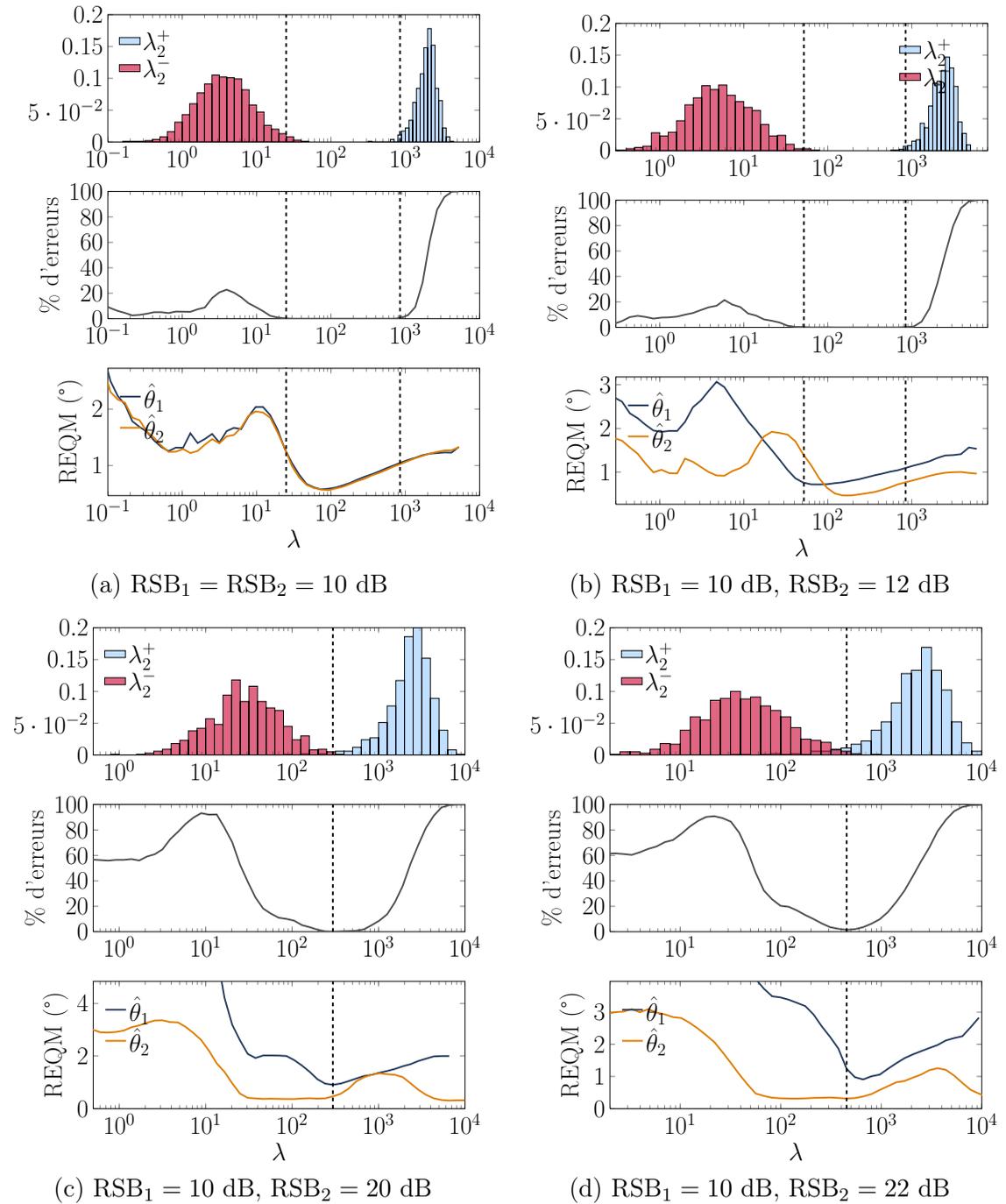


FIGURE 4.13 – Performances de l'algorithme IRL1-CEL0 (initialisation à 0) en fonction de λ pour deux sources indépendantes à BT fini ($K = 50$). En haut : histogrammes de λ_2^- et λ_2^+ pour des observations stochastiques. Milieu et bas : pourcentage d'erreurs et REQM.

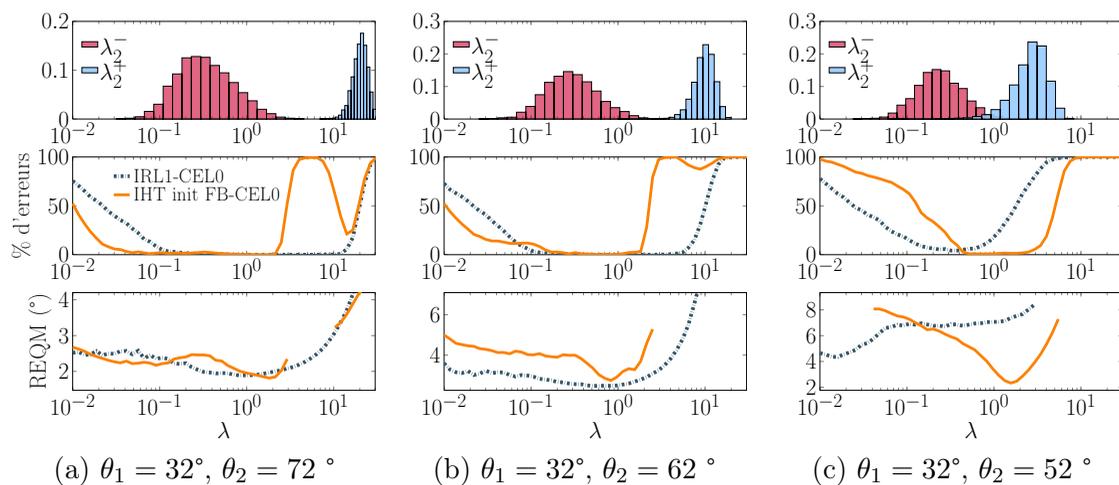


FIGURE 4.14 – Performances de l’algorithme IRL1-CEL0 (initialisation à 0) et IHT (initialisé par la solution de l’algorithme FB-CEL0) en fonction de λ pour deux sources indépendantes et avec erreurs de modèles ($\sigma_e = 0.2$). En haut : histogrammes de λ_2^- et λ_2^+ pour des observations stochastiques. Milieu et bas : pourcentage d’erreurs et REQM.

4.4 Intervalle selon l'écart entre les sources et choix hors-ligne d'une valeur de λ

Au vu du paragraphe précédent, on choisit d'étudier l'intervalle I situé entre les courbes $F_{\lambda_2^-}^{-1}(1 - \epsilon)$ et $F_{\lambda_2^+}^{-1}(\epsilon)$, pour $\epsilon \ll 1$. Cet intervalle est étudié ici pour deux sources de même puissance en fonction de l'écart entre les sources pour les différents scénarios précédents basés sur un réseau UCA comprenant $N = 7$ capteurs et de rayon une demi longueur d'onde. L'étude pourrait être étendue de la même façon aux écarts de puissance entre les sources pour deux directions d'arrivées fixées.

4.4.1 Intervalle pour différents niveaux d'erreurs de modèles

On représente sur la figure 4.15 les courbes de $F_{\lambda_2^-}^{-1}(0.95)$ et $F_{\lambda_2^+}^{-1}(0.05)$ en fonction de l'écart entre les sources pour le scénario 2 décrit dans le paragraphe 4.3.3.1. L'observation pour ce scénario est le premier vecteur propre de la matrice de covariance avec erreurs de modèles. On fixe $\sigma_e = 0.15$. La courbe représentant $F_{\lambda_2^-}^{-1}(0.95)$ est quasi constante; $F_{\lambda_2^+}^{-1}(0.05)$ diminue pour des sources proches dans le lobe, jusqu'à intersecter la courbe $F_{\lambda_2^-}^{-1}(0.95)$ pour un écart $|\theta_2 - \theta_1|$ égal à environ 25° .

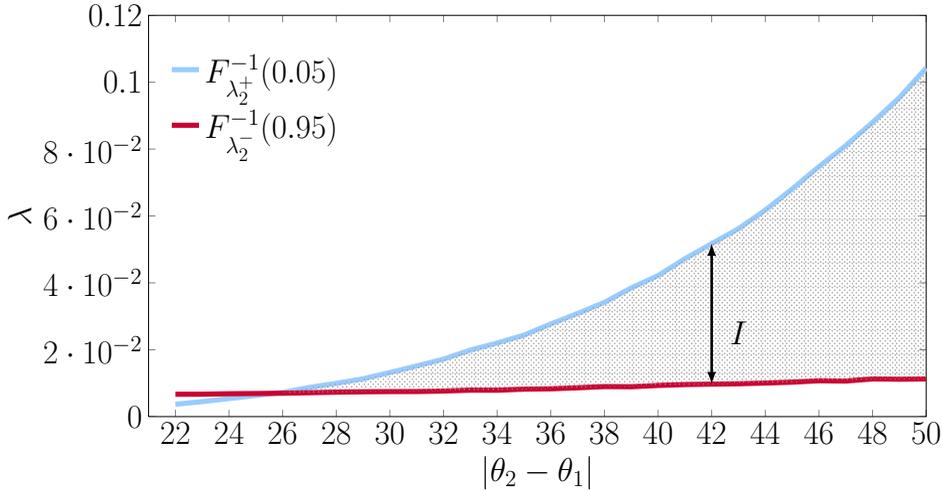


FIGURE 4.15 – Intervalle I situé entre les courbes $F_{\lambda_2^-}^{-1}(1 - \epsilon)$ et $F_{\lambda_2^+}^{-1}(\epsilon)$ tracées en fonction de l'écart $|\theta_2 - \theta_1|$, pour $\epsilon = 0.05$, dans le cadre du scénario 2 avec $\sigma_e = 0.15$.

Sur la figure 4.16 sont représentées les courbes $F_{\lambda_2^-}^{-1}(0.98)$ et $F_{\lambda_2^+}^{-1}(0.02)$ en fonction de l'écart entre les sources pour le scénario 4 décrit dans le paragraphe 4.3.3.3 : l'observation est cette fois-ci la matrice de covariance vectorisée, en présence d'erreurs de modèle. Les courbes sont représentées pour deux valeurs de σ_e : $\sigma_e = 0.2$ et $\sigma_e = 0.1$. L'intersection entre les deux courbes pour $\sigma_e = 0.2$ a lieu autour de l'écart $|\theta_2 - \theta_1| = 21^\circ$, tandis que les deux courbes pour $\sigma_e = 0.1$ ne se croisent pas avant le dernier écart de 15° représenté. On remarque par ailleurs que les courbes pour $\sigma_e = 0.1$ englobent les courbes pour $\sigma_e =$

0.2 : ainsi, l'intervalle I délimité par les courbes $F_{\lambda_2^-}^{-1}(0.98)$ et $F_{\lambda_2^+}^{-1}(0.02)$ pour $\sigma_e = 0.2$ est également un intervalle admissible pour $\sigma_e = 0.1$. Cette remarque prend toute son importance lorsque l'on considère un système opérationnel où le niveau des erreurs de modèles est inconnu. En effet, afin de prévoir les performances du système, on établit une borne supérieure du niveau d'erreurs de modèle σ_e afin de prendre en compte toutes sortes d'incertitudes. L'emboîtement des courbes pour différentes valeurs de σ_e nous permet de nous assurer qu'une valeur de λ choisie pour le pire cas fonctionnera également dans un cas plus favorable.

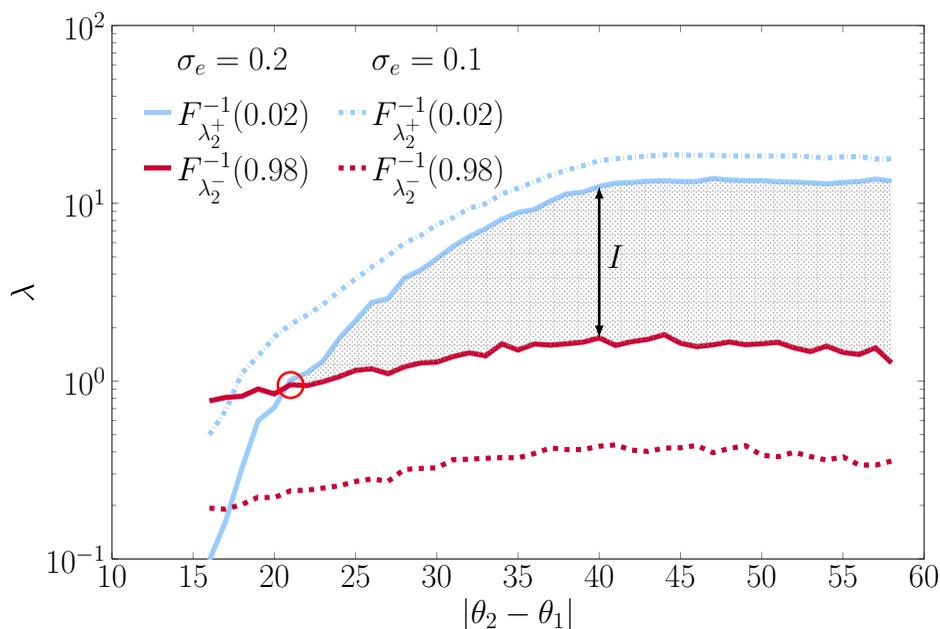


FIGURE 4.16 – Intervalle I situé entre les courbes $F_{\lambda_2^-}^{-1}(1 - \epsilon)$ et $F_{\lambda_2^+}^{-1}(\epsilon)$ tracées en fonction de l'écart $|\theta_2 - \theta_1|$, pour $\epsilon = 0.02$, dans le cadre du scénario 4 pour différentes valeurs de σ_e .

Cette remarque est également valable pour le scénario de la figure 4.15, mais où les différentes courbes en fonction de σ_e n'ont pas été représentées pour plus de clarté. On note également que pour les deux scénarios présentés, les courbes $F_{\lambda_2^-}^{-1}(1 - \epsilon)$ sont proportionnelles à σ_e^2 : le choix d'une valeur de λ à utiliser dans les algorithmes dépend bien du niveau du bruit.

Nous nous basons sur l'étude des performances des algorithmes en fonction du paramètre λ dans le paragraphe 4.3.3 et sur l'existence d'un intervalle I commun à plusieurs directions d'arrivées et plusieurs niveaux d'erreurs de modèles présentée dans ce paragraphe pour proposer un choix hors-ligne d'un unique paramètre de régularisation λ à utiliser dans les algorithmes alors que les directions d'arrivées des sources sont inconnues. On estime préalablement une borne supérieure du niveau d'erreurs de modèle du système. On choisit d'utiliser la valeur de λ située à l'intersection entre les courbes $F_{\lambda_2^-}^{-1}(1 - \epsilon)$ et $F_{\lambda_2^+}^{-1}(\epsilon)$ pour la valeur de σ_e maximale estimée. Par exemple pour le scénario de la figure 4.16, on choisit

le point d'intersection pour $\sigma_e = 0.2$, entouré par un cercle rose en $|\theta_2 - \theta_1| \approx 21^\circ$, pour $\lambda = 0.92$.

4.4.2 Intervalle pour différents RSB et normalisation

On représente sur la figure 4.17 les courbes de $F_{\lambda_2^-}^{-1}(0.98)$ et $F_{\lambda_2^+}^{-1}(0.02)$ en fonction de l'écart entre les sources pour le scénario 3 décrit dans le paragraphe 4.3.3.2 utilisant la matrice de covariance vectorisée à BT fini. Les courbes sont représentées pour 3 niveaux de RSB, pour deux sources de même puissance. Dans les simulations, on fixe le niveau du bruit $\sigma_n^2 = 1$, et on modifie les puissances des sources pour varier le RSB. Sur cette figure, nous pouvons faire plusieurs constatations. On voit directement que le niveau des courbes dépend du RSB : plus celui-ci est élevé, donc plus les puissances des sources sont élevées, plus les bornes de l'intervalle I entre les courbes ont des valeurs élevées : il faut un paramètre λ plus élevé pour une observation de plus grande norme. Les deux courbes $F_{\lambda_2^-}^{-1}(0.98)$ et $F_{\lambda_2^+}^{-1}(0.02)$ pour le RSB le plus faible (RSB = 0 dB) se croisent pour un écart entre les sources situé autour de 20° , alors qu'il n'y a pas d'intersection visible pour les RSB supérieurs au dessus de l'écart minimal de 18° représenté. Les courbes $F_{\lambda_2^-}^{-1}(0.98)$ présentent un motif semblable qui semble être relié à la corrélation spatiale. Ce motif est atténué pour des RSB plus faibles.

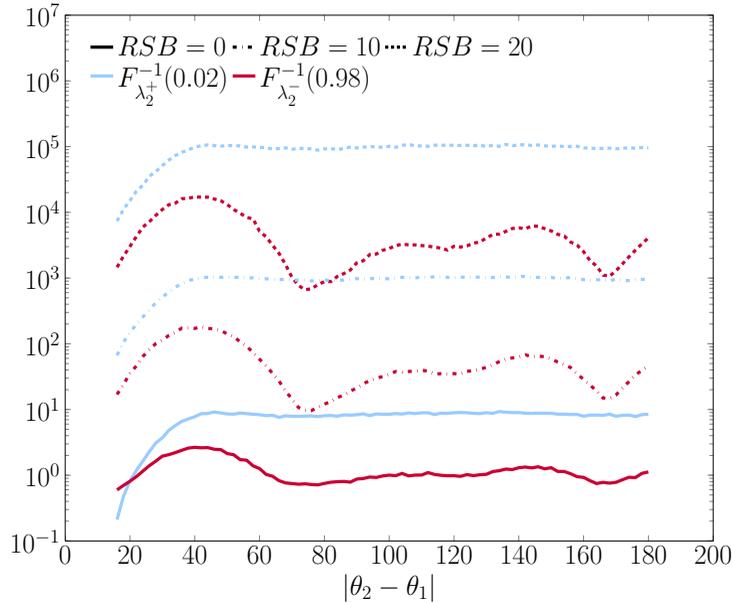


FIGURE 4.17 – Courbes $F_{\lambda_2^-}^{-1}(1 - \epsilon)$ et $F_{\lambda_2^+}^{-1}(\epsilon)$ tracées en fonction de l'écart $|\theta_2 - \theta_1|$ pour différents RSB dans le cadre du scénario 3.

Contrairement aux courbes présentées dans le paragraphe précédent pour différents niveaux d'erreurs de modèle mais un même RSB, les courbes de la figure 4.17 ne s'emboîtent pas les unes dans les autres. Un choix de paramètre λ dans l'intervalle I admissible pour un RSB donné ne sera ainsi pas valable pour un autre RSB. Le niveau des courbes $F_{\lambda_2^-}^{-1}(0.98)$

et $F_{\lambda_2^+}^{-1}(0.02)$ semble dépendre de la norme de l'observation.

On propose ainsi de normer les observations \mathbf{y} préalablement à la mise en œuvre des algorithmes d'estimation des directions d'arrivées. La normalisation d'une observation n'aura en effet pas d'influence sur l'estimation. Posons $\lambda = \alpha \|\mathbf{y}\|_2^2$. Alors

$$\arg \min_{\mathbf{u} \in \mathbb{C}^G} J_{\ell_0}(\lambda, \mathbf{u}) = \arg \min_{\mathbf{u} \in \mathbb{C}^G} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{u}\|_2^2 + \alpha \|\mathbf{y}\|_2^2 \cdot \|\mathbf{u}\|_0 \right\} \quad (4.68)$$

$$= \|\mathbf{y}\|_2 \arg \min_{\mathbf{u} \in \mathbb{C}^G} \left\{ \frac{1}{2} \left\| \frac{\mathbf{y}}{\|\mathbf{y}\|_2} - \mathbf{D}\mathbf{u} \right\|_2^2 + \alpha \|\mathbf{u}\|_0 \right\}. \quad (4.69)$$

L'étude effectuée pour calculer les statistiques de λ_2^- et λ_2^+ à partir des statistiques de l'observation \mathbf{y} doit être renouvelée dans le cas d'une observation normée $\mathbf{y}/\|\mathbf{y}\|_2$. Une telle étude fait appel aux résultats sur les ratios de formes quadratiques hermitiennes. Comme dans le paragraphe 4.3.2.2, les ratios peuvent être mis sous forme de ratios de formes quadratiques réelles. Les articles portant sur la distribution de ratios de formes quadratiques sont nombreux ; on peut citer [38, 39, 52]. Malheureusement, le problème se corse quand il s'agit de calculer la covariance entre deux ratios de formes quadratiques afin de pouvoir corrélérer les différentes distributions obtenues. Nous avons démarré des calculs fastidieux avec des développements limités à l'ordre 2.

Ceux-ci n'ayant pas encore abouti, en première approximation, on s'intéressera donc aux statistiques de α donné par :

$$\alpha = \frac{\lambda}{\mathbb{E} [\|\mathbf{y}\|_2^2]}, \quad (4.70)$$

et on trace sur la figure 4.18 les courbes $F_{\alpha_2^-}^{-1}(0.98)$ et $F_{\alpha_2^+}^{-1}(0.02)$. Il s'agit des mêmes courbes $F_{\lambda_2^-}^{-1}(0.98)$ et $F_{\lambda_2^+}^{-1}(0.02)$ que sur la figure 4.17 mais divisées par l'espérance de la norme de l'observation correspondante. Les intervalles I admissibles situés entre les courbes $F_{\alpha_2^-}^{-1}(0.98)$ et $F_{\alpha_2^+}^{-1}(0.02)$ pour les différents RSB sont chacun emboîtés les uns dans les autres, avec un intervalle plus faible pour le RSB le plus faible, et un intervalle plus grand pour le RSB le plus élevé.

On choisit un paramètre α_{lim} à l'intersection des courbes $F_{\alpha_2^-}^{-1}(1 - \epsilon)$ et $F_{\alpha_2^+}^{-1}(\epsilon)$ de la figure 4.18 pour le plus petit RSB considéré, et on utilise dans les algorithmes la valeur $\lambda = \alpha_{lim} \|\mathbf{y}\|_2^2$ pour paramètre de régularisation. On peut vérifier qu'un tel choix fonctionne relativement bien dans les algorithmes.

4.4.3 Performances pour un λ donné

L'ensemble de l'étude présentée dans ce chapitre sur le paramètre de régularisation a permis d'établir un choix théorique d'une valeur de λ à utiliser dans les algorithmes. Afin de valider définitivement le bien fondé de cette étude, nous établissons les performances obtenues par les algorithmes parcimonieux avec ce choix de paramètre.

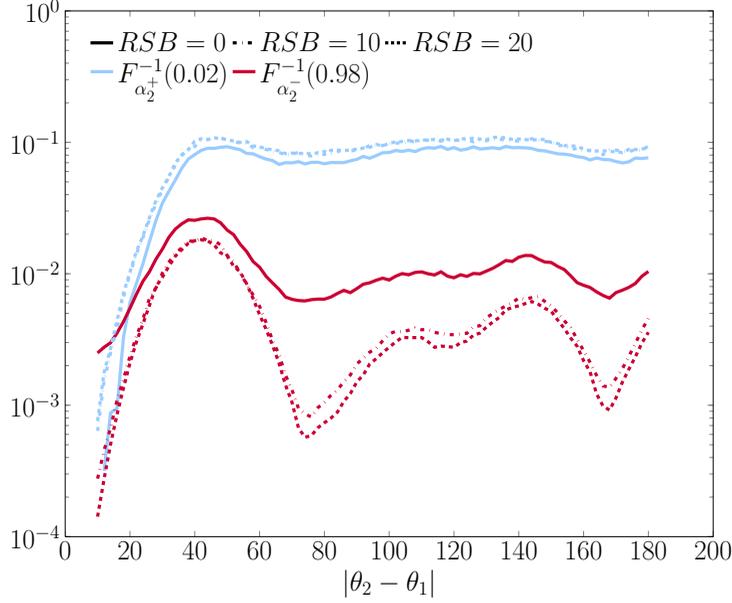


FIGURE 4.18 – Courbes $F_{\alpha_2}^{-1}(1 - \epsilon)$ et $F_{\alpha_2}^{-1}(\epsilon)$ faisant apparaître un intervalle I normalisé en fonction de $|\theta_2 - \theta_1|$ pour différents RSB dans le cadre du scénario 3.

Nous considérons le scénario 2 décrit en 4.3.3.1 : il s'agit du scénario de sources cohérentes en présence d'erreurs de modèles. Pour ce scénario, nous comparons les performances de l'algorithme IRL1-CEL0 aux performances théoriques du maximum de vraisemblance déterministe en présence d'erreurs de modèle telles qu'établies dans l'article [31]. Sur la figure 4.19, nous traçons la REQM en fonction de l'écart entre les sources pour un niveau d'erreurs de modèle fixé à $\sigma_e = 0.15$: le paramètre de régularisation est fixé à la valeur de l'intersection entre les courbes $F_{\lambda_2}^{-1}(1 - \epsilon)$ et $F_{\lambda_2}^{-1}(\epsilon)$ de la figure 4.15 correspondante à ce scénario. Sur la figure 4.20, on représente cette fois la REQM en fonction de l'erreur de modèle pour un écart fixé : on utilise pour chaque σ_e la valeur de λ déterminée pour $\sigma_e = 0.15$ et pondérée par le ratio $\sigma_e^2/0.15^2$. Pour ces deux figures, les performances de l'algorithme IRL1-CEL0 initialisé par le vecteur nul sont très proches des performances attendues par le maximum de vraisemblance déterministe. La méthode parcimonieuse, reposant sur des étapes itératives de gradient et de seuillage, est alors très prometteuse puisque très peu coûteuse en temps de calcul par rapport au maximum de vraisemblance, qui nécessite lui une recherche multi-dimensionnelle.

Nous considérons maintenant le scénario 4 décrit en 4.3.3.3 : il s'agit du scénario de sources indépendantes en présence d'erreurs de modèle, pour lequel l'intervalle I admissible est représenté sur la figure 4.16. Le paramètre de régularisation est fixé à $\lambda = 0.92$, ce qui correspond à l'intersection des bornes de l'intervalle pour $\sigma_e = 0.2$. Nous étudions les performances des algorithmes IRL1-CEL0 initialisé par le vecteur nul et IHT initialisé par la solution de FB-CEL0 en fonction de l'écart entre les sources. La figure 4.21 représente

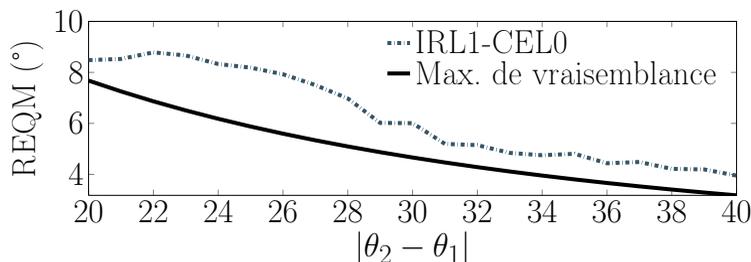


FIGURE 4.19 – Performances de l'algorithme IRL1-CEL0 initialisé par le vecteur nul en fonction de l'écart entre les sources dans le cadre du scénario 2 de sources cohérentes en présence d'erreurs de modèle : comparaison avec les performances théoriques atteignables par le maximum de vraisemblance. Le niveau d'erreurs de modèle est fixé à $\sigma_e = 0.15$.

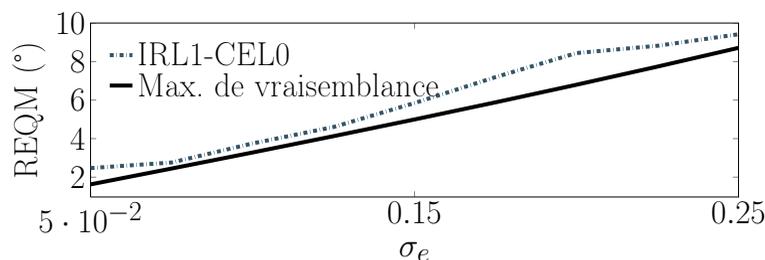


FIGURE 4.20 – Performances de l'algorithme IRL1-CEL0 initialisé par le vecteur nul en fonction de σ_e dans le cadre du scénario 2 de sources cohérentes en présence d'erreurs de modèle : comparaison avec les performances théoriques atteignables par le maximum de vraisemblance. L'écart entre les deux sources est fixé à $|\theta_2 - \theta_1| = 30^\circ$.

les performances pour $\sigma_e = 0.2$ tandis que la figure 4.22 représente les performances pour $\sigma_e = 0.1$. La courbe du pourcentage d'erreurs permet d'établir la limite de résolution atteinte par les différentes méthodes : c'est le plus petit écart angulaire entre les sources tel que le pourcentage d'erreurs soit nul. Pour $\sigma_e = 0.2$, on voit sur la figure 4.21 que la limite de résolution de la méthode MUSIC est d'environ 34° . L'algorithme IRL1-CEL0 permet d'obtenir une limite de résolution de 23° , tandis que l'algorithme IHT initialisé par FB-CEL0 atteint une résolution autour de 20° . Cette valeur de résolution obtenue correspond au point d'intersection entre les courbes $F_{\lambda_2^-}^{-1}(1 - \epsilon)$ et $F_{\lambda_2^+}^{-1}(\epsilon)$ de la figure 4.16, et qui a déterminé le choix de λ . Par ailleurs, la courbe de REQ \bar{M} montre que dès que l'écart angulaire est inférieur à une demi-largeur de lobe, soit 45° , alors la précision moyenne d'estimation pour les deux algorithmes parcimonieux est meilleure que celle de la méthode MUSIC.

Sur la figure 4.22, les constatations concernant l'amélioration de la limite de résolution et de la précision avec les méthodes parcimonieuses sont identiques. La résolution obtenue avec l'algorithme IHT initialisé par FB-CEL0 est autour de 17° , soit l'abscisse correspondante à l'ordonnée $\lambda = 0.92$ (valeur utilisée du paramètre de régularisation) de la courbe $F_{\lambda_2^+}^{-1}(\epsilon)$ pour $\sigma_e = 0.1$. Nous en déduisons que le choix de λ basé sur les courbes $F_{\lambda_2^-}^{-1}(1 - \epsilon)$ et $F_{\lambda_2^+}^{-1}(\epsilon)$ données en fonction de l'écart entre les sources détermine une limite de résolution

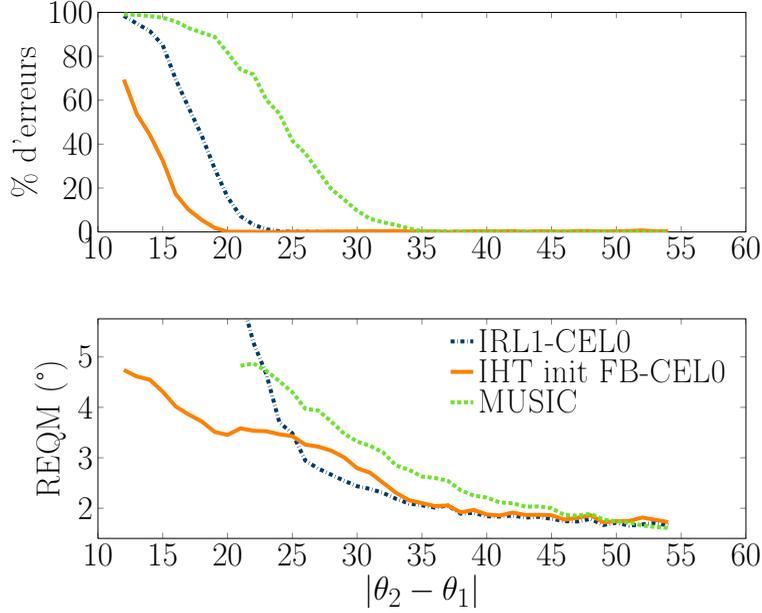


FIGURE 4.21 – Performances des algorithmes IRL1-CEL0 (initialisé par le vecteur nul) et IHT (initialisé par la solution de FB-CEL0) en fonction de l'écart entre les sources dans le cadre du scénario 4 de sources indépendantes en présence d'erreurs de modèle, pour $\sigma_e = 0.2$: comparaison à MUSIC.

théorique.

Par ailleurs, bien que l'algorithme IHT initialisé par la solution de FB-CEL0 permette d'obtenir la meilleure résolution, la précision pour un écart angulaire inférieur à la demi-largeur de lobe est un peu inférieure à celle de l'algorithme IRL1-CEL0. Sur sa plage de fonctionnement, la solution obtenue avec l'algorithme IRL1-CEL0 est donc plus proche du minimum global qu'avec l'algorithme IHT. Malheureusement, l'algorithme IRL1-CEL0 ne permet pas d'atteindre la limite de résolution théorique déterminée par le choix du paramètre de régularisation λ .

4.5 Conclusion

Nous avons mené une étude théorique complète pour déterminer un intervalle I admissible pour le paramètre de régularisation λ tel que la solution optimale de la minimisation du critère J_{ℓ_0} soit celle du maximum de vraisemblance pour $M = 2$ sources incidentes. Cette étude a nécessité le calcul des statistiques de λ_2^- et λ_2^+ , puis l'étude de leurs bornes respectivement supérieures et inférieures $F_{\lambda_2^-}^{-1}(1 - \epsilon)$ et $F_{\lambda_2^+}^{-1}(\epsilon)$, pour $\epsilon \ll 1$.

Nous avons validé par des simulations l'utilité pratique de cet intervalle avec l'utilisation d'algorithmes itératifs de descente à gradient-proximal, très peu coûteux en temps de calcul. Les performances de ces algorithmes parcimonieux sont très encourageantes par rapport aux performances des méthodes conventionnelles de référence telles que MUSIC ou le maximum de vraisemblance. Il subsiste cependant des difficultés sur la convergence et

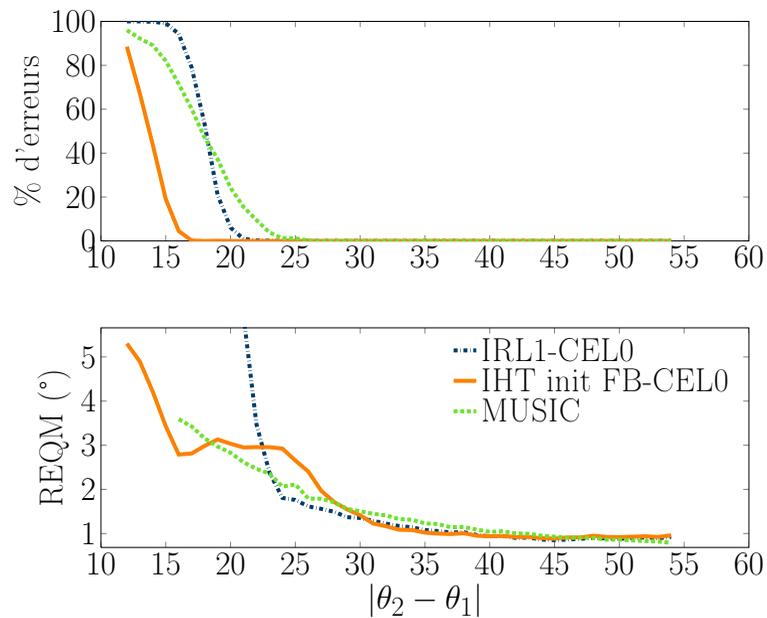


FIGURE 4.22 – Performances des algorithmes IRL1-CEL0 (initialisé par le vecteur nul) et IHT (initialisé par la solution de FB-CEL0) en fonction de l'écart entre les sources dans le cadre du scénario 4 de sources indépendantes en présence d'erreurs de modèle, pour $\sigma_e = 0.1$: comparaison à MUSIC.

l'initialisation de tels algorithmes parcimonieux, notamment dans le cas de sources proches.

Chapitre 5

Stratégie d'optimisation adaptative

5.1 Enjeux et démarche

Dans le chapitre précédent, j'ai montré l'existence d'un intervalle pour le paramètre de régularisation λ permettant de s'assurer que le minimum global de J_{ℓ_0} corresponde à la solution du maximum de vraisemblance déterministe (1.45). Cependant, il est connu que le critère J_{ℓ_0} présente un grand nombre de minima locaux, et les algorithmes de descente à gradient-proximal n'ont pas de garantie de convergence vers le minimum global mais seulement vers un point critique de J_{ℓ_0} . L'étude du comportement des algorithmes en fonction du paramètre λ dans le chapitre précédent a mis en évidence les avantages et inconvénients des deux pénalités ℓ_0 et CEL0, en lien avec leurs propriétés bien connues.

- Le nombre important de minima locaux de J_{ℓ_0} et l'absence de garantie de convergence des algorithmes de minimisation sous-optimaux utilisés rendent les résultats dépendants de l'initialisation. Dans certains scénarios, nous avons proposé une initialisation éclairée qui permet d'obtenir des performances encourageantes avec l'algorithme IHT, mais dans la plupart des cas les résultats restent aléatoires.
 - Le minimum global de J_{ℓ_0} est préservé dans le critère J_{CEL0} , ce qui n'est pas forcément le cas pour d'autres critères régularisés proposés dans la littérature. Nous avons donc utilisé les algorithmes à gradient-proximal dédiés à la minimisation de J_{CEL0} dans le chapitre précédent. Nous avons montré que ceux-ci permettent d'obtenir une solution proche de la solution globale, qui correspond à la solution du maximum de vraisemblance lorsque λ est choisi dans le bon intervalle.
 - Il est également connu qu'il y a moins de minima locaux sur J_{CEL0} que sur J_{ℓ_0} : c'est ce qui nous a permis notamment d'initialiser les algorithmes par le vecteur nul pour obtenir les bonnes performances évoquées ci-dessus.
 - Néanmoins, l'utilisation seule de J_{CEL0} et des algorithmes associés avec une initialisation par le vecteur nul ne permet pas d'atteindre la limite de résolution théorique donnée par l'étude sur le paramètre de régularisation λ .
-

Ces premiers résultats sont la preuve que les problèmes de convergence et donc d'initialisation dans les algorithmes de minimisation à faible complexité comme ceux à gradient-proximal ne sont pas entièrement résolus.

L'objectif de ce chapitre est de proposer un schéma d'optimisation avec des algorithmes à gradient-proximal qui augmente la probabilité d'obtenir la solution globale pour tous les scénarios et indépendamment de l'initialisation, y compris pour des sources proches. J'étudie ainsi préalablement en détail les minimiseurs de J_{ℓ_0} et J_{CEL0} pour comprendre les phénomènes qui entrent en jeu.

On rappelle que plusieurs des notations utilisées sont définies au début du chapitre 2.

5.2 Étude des minimiseurs de J_{ℓ_0}

5.2.1 Caractérisation des minimiseurs

On considère le problème (\mathcal{R}_λ) de minimisation de la fonction J_{ℓ_0} . Le minimiseur global de $J_{\ell_0}(\lambda, \cdot)$ a déjà été étudié dans le chapitre 4. On étudie ici l'ensemble des minimiseurs, à la fois locaux et globaux, de $J_{\ell_0}(\lambda, \cdot)$. Une étude a été réalisée par Mila Nikolova [56] dans le cas de vecteurs et matrices réels. Nous étendons les résultats au cas complexe et donnons leur interprétation pour la goniométrie.

Commençons par un résultat simple, qui a cependant toute son importance :

Lemme 5.1. *Pour tout $\mathbf{y} \in \mathbb{C}^{N'}$ et pour tout $\lambda > 0$, J_{ℓ_0} a un minimum (local) strict en $\hat{\mathbf{u}} = \mathbf{0}$.*

En effet, on a :

$$\|\mathbf{D}\mathbf{v} - \mathbf{y}\|_2^2 = \|\mathbf{D}\mathbf{v}\|_2^2 + \|\mathbf{y}\|_2^2 - 2\Re(\langle \mathbf{D}^H \mathbf{y}, \mathbf{v} \rangle), \quad (5.1)$$

et $J_{\ell_0}(\lambda, \mathbf{0}) = \|\mathbf{y}\|_2^2$, d'où

$$J_{\ell_0}(\lambda, \mathbf{v}) = J_{\ell_0}(\lambda, \mathbf{0}) + R_{\ell_0}(\lambda, \mathbf{v}) \quad (5.2)$$

avec

$$R_{\ell_0}(\lambda, \mathbf{v}) = \|\mathbf{D}\mathbf{v}\|_2^2 - 2\Re(\langle \mathbf{D}^H \mathbf{y}, \mathbf{v} \rangle) + \lambda \|\mathbf{v}\|_0. \quad (5.3)$$

Pour tout $\mathbf{v} \neq \mathbf{0}$ tel que $\|\mathbf{v}\|_2 < \frac{\lambda}{2\|\mathbf{D}^H \mathbf{y}\|_2 + 1}$, on a $R_{\ell_0}(\lambda, \mathbf{v}) > 0$, d'où le résultat. L'initialisation par $\mathbf{0}$ d'algorithmes sous-optimaux de descente pour la minimisation de J_{ℓ_0} , par exemple l'algorithme IHT 2.3, est alors généralement un mauvais choix.

Nous cherchons maintenant à identifier et caractériser les autres minima du critère J_{ℓ_0} dans le problème (\mathcal{R}_λ) . Soient $\mathbf{y} \in \mathbb{C}^{N'}$, $\mathbf{D} \in \mathbb{C}^{N' \times G}$ et un ensemble $\omega \subset \mathbb{I}_G$. Nous

définissons un nouveau problème contraint (\mathcal{C}_ω) :

$$\min_{\mathbf{u} \in \mathbb{C}^G} \|\mathbf{y} - \mathbf{D}\mathbf{u}\|_2^2 \quad \text{s.c.} \quad u_g = 0, \quad \forall g \in \omega^c. \quad (\mathcal{C}_\omega)$$

Si $\omega = \emptyset$, alors $\omega^c = \mathbb{I}_G$, et l'unique solution de (\mathcal{C}_ω) est $\hat{\mathbf{u}} = \mathbf{0}$. Pour $\omega \subset \mathbb{I}_G$ tel que $\#\omega \geq 1$, alors les solutions $\hat{\mathbf{u}}$ de (\mathcal{C}_ω) vérifient $\text{supp}(\hat{\mathbf{u}}) = \omega$: ce sont des vecteurs parcimonieux dont les indices des composantes non nulles sont dans ω . On a alors $\hat{\mathbf{u}} = C_\omega(\hat{\mathbf{u}}_\omega)$, en utilisant la définition de l'opérateur de complétion de zéros (2.10).

Résoudre le problème (\mathcal{C}_ω) revient à résoudre un problème quadratique (\mathcal{Q}_ω) donné par :

$$\min_{\mathbf{v} \in \mathbb{C}^{\#\omega}} \|\mathbf{D}_\omega \mathbf{v} - \mathbf{y}\|_2^2, \quad \#\omega \geq 1, \quad (\mathcal{Q}_\omega)$$

et qui admet toujours une solution. On a alors :

$$[\hat{\mathbf{u}} \in \mathbb{C}^G \text{ résout } (\mathcal{C}_\omega), \#\omega \geq 1] \Leftrightarrow [\hat{\mathbf{u}}_\omega \in \mathbb{C}^{\#\omega} \text{ résout } (\mathcal{Q}_\omega) \text{ et } \hat{\mathbf{u}} = C_\omega(\hat{\mathbf{u}}_\omega)] \quad (5.4)$$

$$\Leftrightarrow [\hat{\mathbf{u}}_\omega \text{ résout } \mathbf{D}_\omega^H \mathbf{D}_\omega \mathbf{v} = \mathbf{D}_\omega^H \mathbf{y} \text{ et } \hat{\mathbf{u}} = C_\omega(\hat{\mathbf{u}}_\omega)] \quad (5.5)$$

La valeur minimale du problème quadratique (\mathcal{Q}_ω) correspond à la valeur du critère du maximum de vraisemblance $J_{MV}(\Psi_\omega)$, avec $\Psi_\omega = \{\psi_{\omega_1}, \dots, \psi_{\omega_{\#\omega}}\} \subset \Psi$.

L'étude sur le paramètre de régularisation λ du chapitre 4 est basée sur l'équivalence entre le problème contraint (\mathcal{C}_s) , correspondant au maximum de vraisemblance pour M sources, et le problème régularisé (\mathcal{R}_λ) , c'est-à-dire entre minimiseurs globaux. Nous montrons ici qu'il y a de fortes connexions entre les minimiseurs du problème contraint (\mathcal{C}_ω) et les minimiseurs (locaux) du problème régularisé (\mathcal{R}_λ) .

Proposition 5.1. *Soit $\mathbf{y} \in \mathbb{C}^{N'}$. Pour un ensemble $\omega \subset \mathbb{I}_G$, soit $\hat{\mathbf{u}}$ une solution de (\mathcal{C}_ω) . Alors pour tout $\lambda > 0$, la fonction J_{ℓ_0} atteint un minimum (local) en $\hat{\mathbf{u}}$, et $\text{supp}(\hat{\mathbf{u}}) \subseteq \omega$.*

La preuve est similaire à celle du cas réel développée dans [56]. La réciproque, évidente, est vraie aussi :

Lemme 5.2. *Pour $\mathbf{y} \in \mathbb{C}^{N'}$ et $\lambda > 0$, soit $\hat{\mathbf{u}}$ un minimiseur (local) de J_{ℓ_0} . Alors $\hat{\mathbf{u}}$ est solution de (\mathcal{C}_ω) , pour $\omega = \text{supp}(\hat{\mathbf{u}})$.*

Le théorème suivant nous indique de plus la condition pour avoir un minimiseur strict.

Théorème 5.1. *Pour $\mathbf{y} \in \mathbb{C}^{N'}$ et $\lambda > 0$, soit $\hat{\mathbf{u}}$ un minimiseur (local) de J_{ℓ_0} , avec $\hat{\omega} = \text{supp}(\hat{\mathbf{u}})$. Alors $\hat{\mathbf{u}}$ est un minimiseur strict si et seulement si $\text{rank}(\mathbf{D}_{\hat{\omega}}) = \#\hat{\omega}$. Si $\hat{\mathbf{u}}$ est un minimiseur (local) strict de J_{ℓ_0} , alors on peut écrire :*

$$\hat{\mathbf{u}} = C_{\hat{\omega}}(\hat{\mathbf{u}}_{\hat{\omega}}), \quad \text{pour} \quad \hat{\mathbf{u}}_{\hat{\omega}} = (\mathbf{D}_{\hat{\omega}}^H \mathbf{D}_{\hat{\omega}})^{-1} \mathbf{D}_{\hat{\omega}}^H \mathbf{y}, \quad (5.6)$$

et $\#\hat{\omega} = \|\hat{\mathbf{u}}\|_0 \leq N'$.

Il est ainsi aisé de vérifier si un algorithme de minimisation a atteint un minimum (local) strict de J_{ℓ_0} . Notons que les minima $\hat{\mathbf{u}}^s \in \hat{U}_c^s(\mathbf{y})$ (4.2) du problème contraint (\mathcal{C}_s) pour $s \in \mathbb{I}_L$ sont inclus dans l'ensemble des minimiseurs stricts de J_{ℓ_0} puisque c_s défini par l'équation (4.1) vérifie, $\forall s > 0$:

$$c_s(\mathbf{y}) = \min_{\omega \subset \mathbb{I}_G, \#\omega=s} \left\{ \min_{\mathbf{v} \in \mathbb{C}^s} \|\mathbf{D}_\omega \mathbf{v} - \mathbf{y}\|_2^2 \right\} \quad (5.7)$$

$$= \min_{\omega \subset \mathbb{I}_G, \#\omega=s} J_{MV}(\Psi_\omega). \quad (5.8)$$

Définition 5.1. On considère une matrice $\mathbf{D} \in \mathbb{C}^{N' \times G}$. Pour tout $r \in \mathbb{I}_{N'}$, on définit Ω_r comme l'ensemble de tous les supports de taille r qui correspondent à des sous-matrices de dimension $N' \times r$ de rang plein de \mathbf{D} :

$$\Omega_r = \{\omega \subset \mathbb{I}_G : \#\omega = r = \text{rank}(\mathbf{D}_\omega)\}. \quad (5.9)$$

On pose $\Omega_0 = \emptyset$. On définit également :

$$\Omega_{\max} = \bigcup_{r=0}^{N'} \Omega_r. \quad (5.10)$$

On peut ainsi facilement calculer un minimiseur (local) strict $\hat{\mathbf{u}}$ de J_{ℓ_0} , sans même connaître la valeur de λ . Tous les supports $\omega \in \Omega_{\max}$ permettent en effet de calculer un minimiseur (local) strict de J_{ℓ_0} à partir uniquement de \mathbf{y} et \mathbf{D} , en utilisant (5.6). On peut ainsi avoir une idée du grand nombre de minima locaux du critère J_{ℓ_0} : le nombre de supports $\omega \in \Omega_{\max}$ qui donnent un minimiseur local strict est borné par $\sum_{r=0}^{N'} \binom{G}{r}$. On remarque alors que plus la grille de discrétisation de la table de calibration est fine, plus le nombre de minima locaux est important.

Il a également été montré, pour des conditions légères, que J_{ℓ_0} possède un unique minimiseur global strict [56]. Celui-ci dépend de la valeur de λ , comme nous l'avons montré dans le chapitre 4. De plus, nous avons la proposition suivante [56] :

Proposition 5.2. Pour $\mathbf{y} \in \mathbb{C}^{N'}$ et $\lambda > 0$, soit $\hat{\mathbf{u}}$ un minimiseur global de J_{ℓ_0} . Alors

$$g \in \text{supp}(\hat{\mathbf{u}}) \Rightarrow |\hat{u}_g| \geq \frac{\sqrt{2\lambda}}{\|\mathbf{d}_g\|_2}. \quad (5.11)$$

5.2.2 Illustrations de quelques minimiseurs de J_{ℓ_0}

On considère un réseau circulaire uniforme à $N = 7$ antennes et de rayon une demi longueur d'onde, avec $M = 2$ sources indépendantes. L'observation \mathbf{y} est donnée par la matrice de covariance vectorisée, dans un premier temps à durée d'intégration infinie et sans erreur de modèle. Les sources sont de même puissance, avec $RSB = 0$ dB. On considère

un pas de grille de 0.5° , pour des angles compris dans une zone utile réduite $[2, 100]^\circ$: on a ainsi $G = 196$. Les directions d'arrivées des sources sont fixées à $\theta_1 = 32^\circ$ et $\theta_2 = 62^\circ$. On rappelle que dans ce cas précis asymptotique, il est aisé de calculer précisément l'intervalle $]\lambda_2^-(\mathbf{y}); \lambda_2^+(\mathbf{y})[$. En effet, la valeur de $c_2(\mathbf{y})$ (4.1) est nulle, d'où $\lambda_2^-(\mathbf{y}) = 0$, et

$$\lambda_2^+(\mathbf{y}) = \min \left\{ \frac{c_1(\mathbf{y})}{2}, \frac{c_0(\mathbf{y})}{4} \right\}. \quad (5.12)$$

Pour ce scénario, on calcule $\lambda_2^+(\mathbf{y}) \approx 9.99$. On fixe le paramètre de régularisation à $\lambda = 9.5$.

Dans ce scénario, le vecteur inconnu \mathbf{u} correspond à un pseudo-spectre en puissance γ , où le vecteur γ est réel. Les composantes de ce vecteur sont ainsi à valeurs réelles : $\mathbf{u} \in \mathbb{R}^G$. Comme il est impossible de représenter le critère multidimensionnel J_{ℓ_0} selon la dimension de \mathbf{u} , on s'intéresse sur la figure 5.1 à des projections de J_{ℓ_0} pour des vecteurs ayant au plus deux composantes non nulles : $\mathbf{u} = C_\omega(\mathbf{u}_\omega)$ avec $\#\omega = 2$. Ces composantes non nulles correspondent à des directions données par les angles ψ_{ω_1} et ψ_{ω_2} . On fixe $\psi_{\omega_1} = \theta_1 = 32^\circ$. La deuxième direction ψ_{ω_2} varie sur les différentes sous-figures. Pour (a), $\psi_{\omega_2} = 47^\circ$, ce qui correspond à l'angle situé entre les deux sources $1/2(\theta_1 + \theta_2)$; sur la sous-figure (d), $\psi_{\omega_2} = \theta_2 = 62^\circ$. Entre ces deux valeurs, on fixe, pour (b) et (c), $\psi_{\omega_2} = 52^\circ$ et 57° . Les valeurs des deux composantes u_{ω_1} et u_{ω_2} , qui sont les seules composantes à pouvoir être non nulles, varient selon les deux axes.

Sur chaque sous-figure (a)-(d) de la figure 5.1, on retrouve quatre minima locaux (indiqués par les astérisques bleues) : le minimum local en $\mathbf{0}$, un minimum local sur chaque axe, correspondant à uniquement une composante non nulle selon l'une des directions ψ_{ω_1} ou ψ_{ω_2} , et le minimum local correspondant à strictement deux composantes non nulles. Sachant que λ est dans le bon intervalle, l'unique minimiseur global $\hat{\mathbf{u}}^2 \in \hat{U}_c^2(\mathbf{y}) = \hat{U}_r^\lambda(\mathbf{y})$ (4.2) correspond à $\psi_{\omega_1} = \theta_1$, $\psi_{\omega_2} = \theta_2$ et $u_{\omega_1} = u_{\omega_2} = 7$: il est représenté par un cercle noir plein sur la figure 5.1(d). La valeur du critère en ce point est donnée par :

$$J_{\ell_0}(\lambda, \hat{\mathbf{u}}^2) = \frac{1}{2}c_2 + 2\lambda = 2\lambda = 19 \quad (5.13)$$

dans ce cas précis asymptotique.

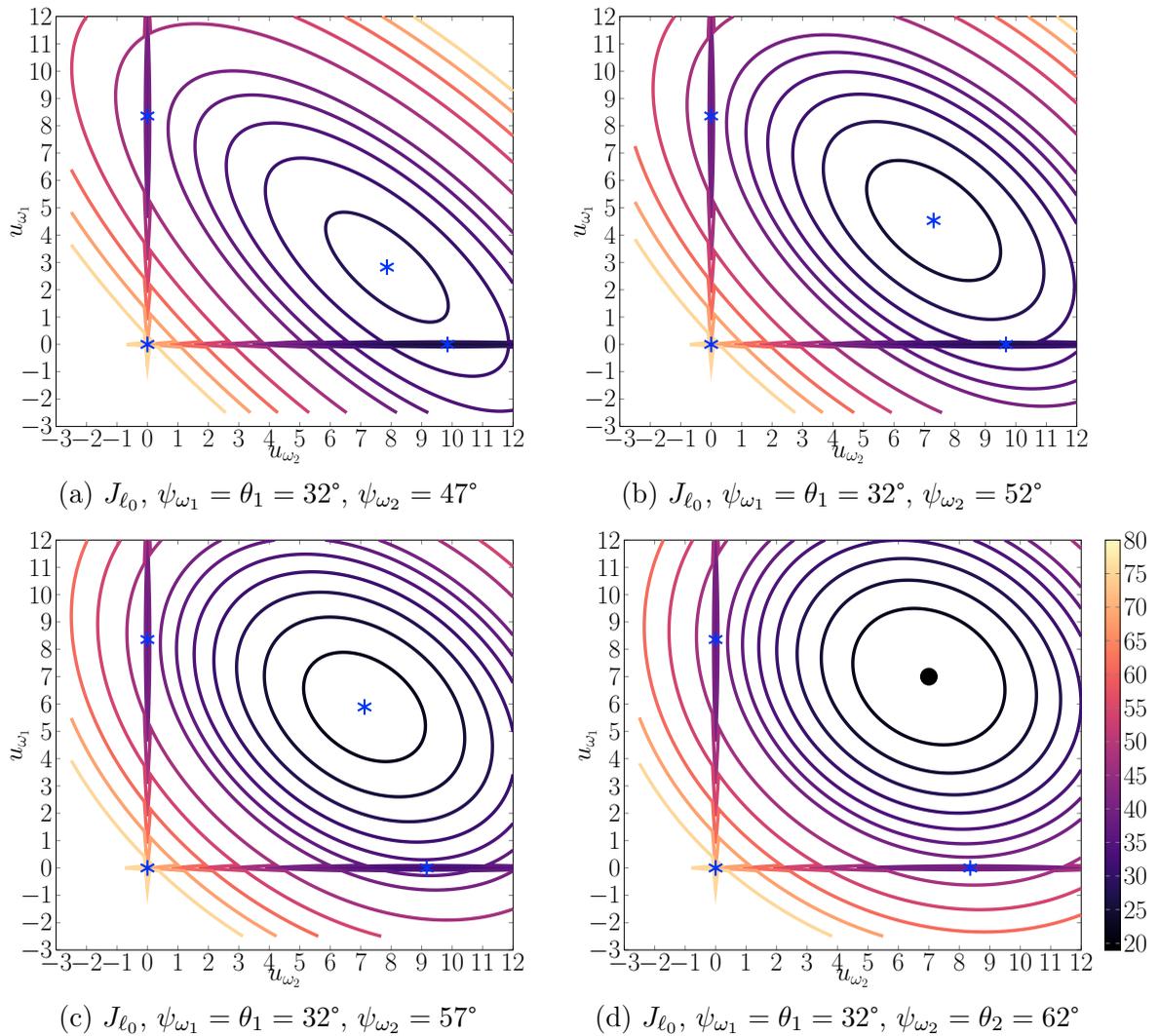


FIGURE 5.1 – Courbes de niveau de la surface $J_{\ell_0}(\lambda = 9.5, \mathbf{u})$, pour $\mathbf{u} = C_\omega(\mathbf{u}_\omega)$, avec $\#\omega = 2$, en fonction des valeurs des composantes u_{ω_1} et u_{ω_2} , correspondant à des directions d'arrivées ψ_{ω_1} et ψ_{ω_2} . Les vraies directions d'arrivées sont $\theta_1 = 32^\circ$, et $\theta_2 = 62^\circ$. Les minima locaux sont représentés par les astérisques bleues. Le minimum global est indiqué sur la sous-figure (d) par un cercle plein noir.

5.3 Étude des minimiseurs de J_{CEL0} comparativement à ceux de J_{ℓ_0}

5.3.1 Caractérisation des minimiseurs de J_{CEL0}

La pénalité CEL0 est une pénalité continue, non-convexe, approchant la norme ℓ_0 . Son atout par rapport à bien d'autres pénalités proposées dans la littérature réside dans les propriétés sur les minimiseurs de la fonctionnelle J_{CEL0} . Nous donnons ici les principaux résultats théoriques issus de [69, 70] avant d'illustrer ceux-ci.

Théorème 5.2 (Minimiseurs globaux de J_{CEL0} et J_{ℓ_0}). *L'ensemble des minimiseurs globaux de J_{ℓ_0} est inclus dans l'ensemble des minimiseurs globaux de J_{CEL0} :*

$$\arg \min_{\mathbf{u} \in \mathbb{C}^{\mathbf{G}}} J_{\ell_0}(\lambda, \mathbf{u}) \subseteq \arg \min_{\mathbf{u} \in \mathbb{C}^{\mathbf{G}}} J_{\text{CEL0}}(\lambda, \mathbf{u}) \quad (5.14)$$

Réciproquement, si $\hat{\mathbf{u}}$ est un minimiseur global de J_{CEL0} , alors $\hat{\mathbf{u}}_T$ défini par

$$\forall g \in \mathbb{I}_G, \hat{u}_{T,g} = \hat{u}_g \mathbb{1}_{\{\hat{u}_g \geq \sqrt{2\lambda} / \|\mathbf{d}_g\|_2\}}, \quad (5.15)$$

est un minimiseur global de J_{ℓ_0} , et

$$J_{\text{CEL0}}(\hat{\mathbf{u}}) = J_{\text{CEL0}}(\hat{\mathbf{u}}_T) = J_{\ell_0}(\hat{\mathbf{u}}_T). \quad (5.16)$$

Corolaire 5.3. *Si J_{ℓ_0} admet un unique minimiseur global, alors*

$$\arg \min_{\mathbf{u} \in \mathbb{C}^{\mathbf{G}}} J_{\ell_0}(\lambda, \mathbf{u}) = \arg \min_{\mathbf{u} \in \mathbb{C}^{\mathbf{G}}} J_{\text{CEL0}}(\lambda, \mathbf{u}). \quad (5.17)$$

Le nom de la pénalité CEL0 (*Continuous Exact ℓ_0*) a été donné en relation avec ce théorème : il est équivalent de minimiser les fonctions J_{CEL0} et J_{ℓ_0} . Cependant, J_{CEL0} est non-convexe, et les algorithmes itératifs à gradient-proximal utilisés pour l'optimisation n'ont pas de garantie de convergence globale. Il faut donc également s'intéresser aux minima locaux.

Théorème 5.3 (Minimiseurs locaux de J_{CEL0} et J_{ℓ_0}). *Soit $\hat{\mathbf{u}}$ un minimiseur local de J_{CEL0} , alors $\hat{\mathbf{u}}_T$ défini par (5.15) est un minimiseur local de J_{ℓ_0} et (5.16) est vérifiée.*

En d'autres termes, les deux théorèmes précédents nous indiquent que les minimiseurs globaux de J_{ℓ_0} sont préservés dans J_{CEL0} , et que de chaque minimiseur (local) de J_{CEL0} , on peut extraire un minimiseur (local) de J_{ℓ_0} . Notons qu'il n'y a pas de réciproque à cette dernière affirmation. Ainsi, et c'est une propriété qui nous intéresse particulièrement, on observe une diminution du nombre de minima locaux de J_{CEL0} par rapport à J_{ℓ_0} .

Lemme 5.4. Soit $\hat{\mathbf{u}}$ un minimiseur local de J_{CEL0} , alors [68]

$$\forall g \in \mathbb{I}_G, |\hat{u}_g| \in \{0\} \cup \left[\sqrt{2\lambda}; +\infty \right[. \quad (5.18)$$

Ainsi, si $\hat{\mathbf{u}}$ est un minimiseur local de J_{ℓ_0} tel que $|\hat{u}_g| \in]0, \sqrt{2\lambda}[$ pour un certain $g \in \mathbb{I}_G$, alors $\hat{\mathbf{u}}$ n'est pas un minimiseur local de J_{CEL0} .

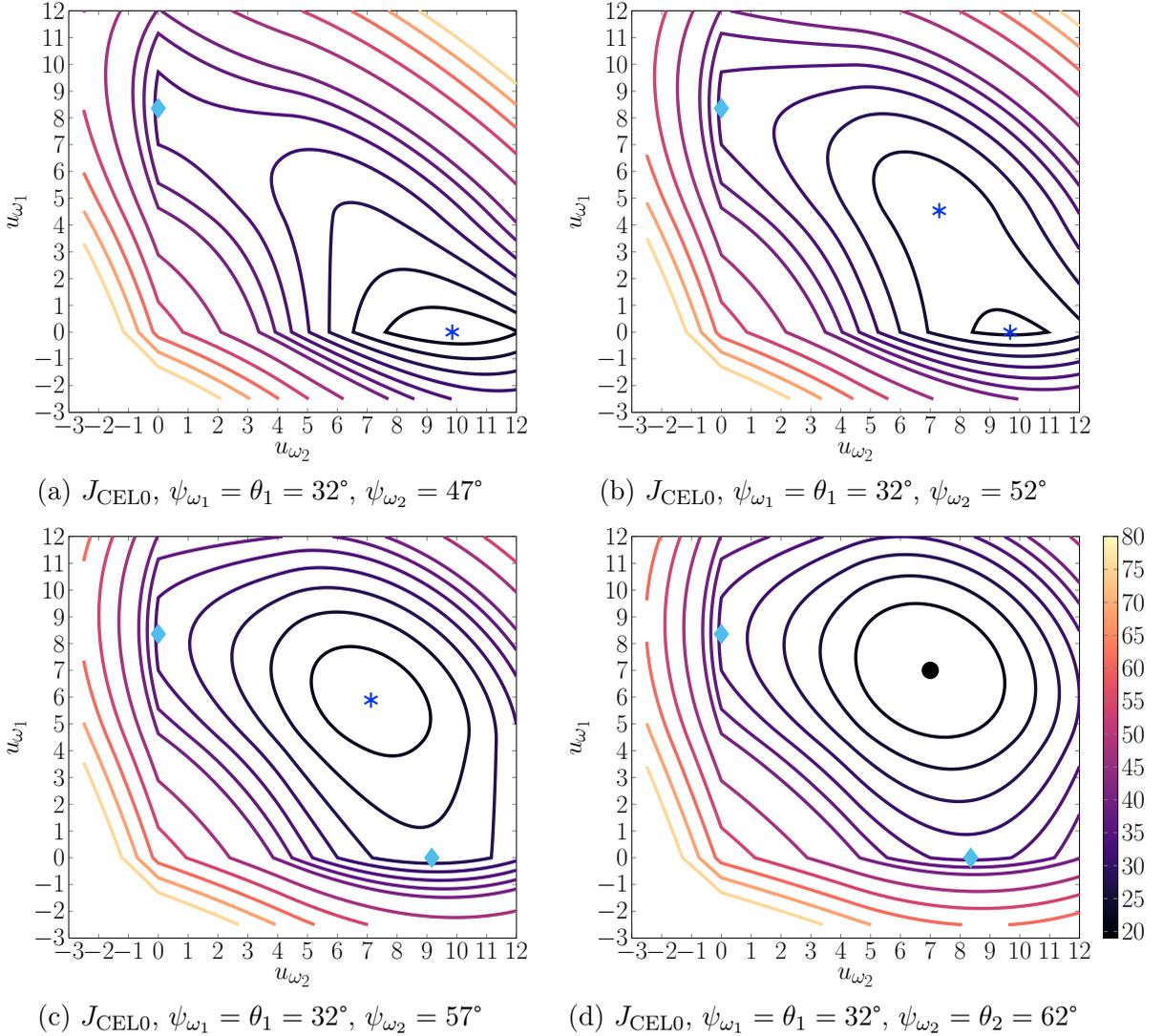


FIGURE 5.2 – Courbes de niveau de la surface $J_{\text{CEL0}}(\lambda = 9.5, \mathbf{u})$, pour $\mathbf{u} = C_\omega(\mathbf{u}_\omega)$, avec $\|\omega\| = 2$, en fonction des valeurs des composantes u_{ω_1} et u_{ω_2} , correspondant à des directions d'arrivées ψ_{ω_1} et ψ_{ω_2} . Les vraies directions d'arrivées sont $\theta_1 = 32^\circ$, et $\theta_2 = 62^\circ$. Les minima locaux sont représentés par les astérisques bleues. Les losanges bleus clairs sont des points critiques qui ne sont pas des minima locaux. Le minimum global est indiqué sur la sous-figure (d) par un cercle plein noir.

Ce comportement s'observe sur la figure 5.2 où sont représentées des projections de J_{CEL0} pour des vecteurs ayant au plus deux composantes non nulles : $\mathbf{u} = C_\omega(\mathbf{u}_\omega)$ avec

$\sharp\omega = 2$. On représente exactement le même scénario que sur la figure 5.1 afin de comparer certains des minimiseurs locaux de J_{CEL0} et J_{ℓ_0} . On rappelle que la valeur du paramètre de régularisation λ est fixée à $\lambda = 9.5$. On remarque dans un premier temps que le minimum local de la surface J_{ℓ_0} situé en $\mathbf{0}$ n'est plus un minimum local pour ce scénario pour J_{CEL0} : la suppression de ce minimiseur est particulièrement intéressante pour l'initialisation des algorithmes. Plusieurs minimiseurs locaux de J_{ℓ_0} sont des points critiques mais pas des minimiseurs locaux (indiqués par des losanges bleus clairs) de la surface J_{CEL0} . Enfin, sur la sous-figure 5.2(a), on observe la suppression du minimum local obtenu dans J_{ℓ_0} pour strictement deux composantes non nulles : en effet, ce minimiseur de J_{ℓ_0} vérifie $u_{\omega_1} < \sqrt{2\lambda} \approx 4.36$.

5.3.2 Étude des surfaces J_{CEL0} et J_{ℓ_0} pour des sources proches

La pénalité Φ_{CEL0} sous-estime la pénalité Φ_{ℓ_0} : en effet, chaque composante de Φ_{CEL0} donnée par (2.46) vérifie $0 \leq \phi_{\text{CEL0}}(\lambda, \alpha_i, u_i) \leq \lambda$, d'où $J_{\text{CEL0}} \leq J_{\ell_0}$ en tout point. En une dimension ou pour un dictionnaire orthogonal, la fonction J_{CEL0} est l'enveloppe convexe de J_{ℓ_0} [69], c'est-à-dire la plus grande fonction semi-continue inférieure convexe qui est plus petite que J_{ℓ_0} .

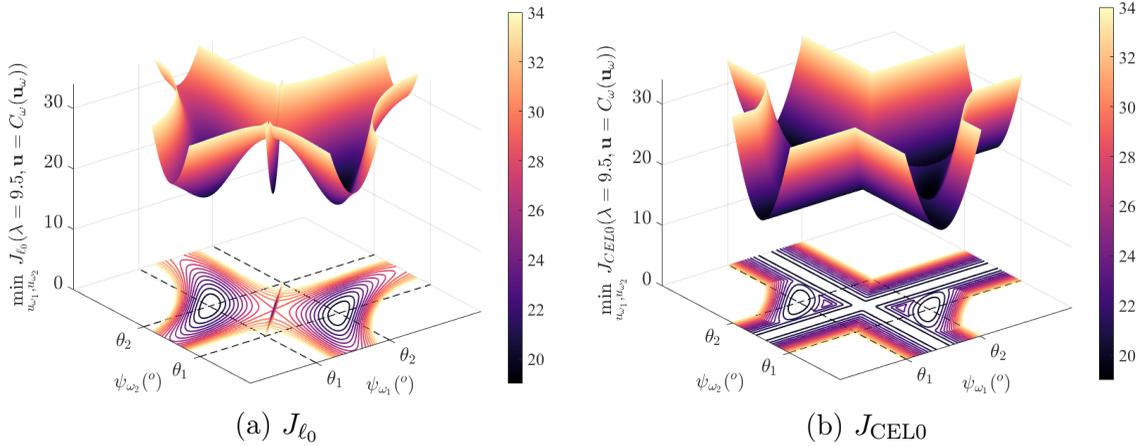


FIGURE 5.3 – Valeurs minimales des critères $J_{\ell_0}(\lambda = 9.5, \mathbf{u})$ et $J_{\text{CEL0}}(\lambda = 9.5, \mathbf{u})$, pour $\mathbf{u} = C_{\omega}(\mathbf{u}_{\omega})$, avec $\sharp\omega = 2$, en fonction des angles ψ_{ω_1} et ψ_{ω_2} . Les vraies directions d'arrivées sont $\theta_1 = 32^\circ$, et $\theta_2 = 62^\circ$.

La suppression de minima locaux observée dans le paragraphe précédent est ainsi liée à l'adoucissement de la discontinuité de la norme ℓ_0 par la pénalité CEL0. Il y a ainsi un étalement de ces minima, ce qui peut mener à des minima dits "plats" [43] : il s'agit d'une grande région de la surface d'erreur où celle-ci est presque constante.

On représente ce phénomène sur la figure 5.3. On reprend le scénario utilisé en 5.2.2. On trace les valeurs minimales des critères $J_{\ell_0}(\lambda = 9.5, \mathbf{u})$ et $J_{\text{CEL0}}(\lambda = 9.5, \mathbf{u})$, pour $\mathbf{u} = C_{\omega}(\mathbf{u}_{\omega})$, avec $\sharp\omega \leq 2$, en fonction des angles ψ_{ω_1} et ψ_{ω_2} . La diagonale correspond à $\sharp\omega = 1$; les surfaces d'erreur sont logiquement symétriques par rapport à cette diagonale, il

s'agit en fait des mêmes points représentés de part et d'autre. Notons que chaque point de la sous-figure 5.3(a) est un minimum local de J_{ℓ_0} : deux points représentés côte à côte dans cette représentation ne sont en général pas voisins dans l'espace \mathbb{C}^G . Pour la sous-figure 5.3(a), le minima global de J_{ℓ_0} apparait dans une large mais profonde cuvette quand le couple $(\psi_{\omega_1}, \psi_{\omega_2})$ correspond au couple (θ_1, θ_2) . La valeur de ce minimum est donnée par l'équation (5.13).

Pour des sources de même puissance et dont les directions sont à l'intérieur du même lobe d'antenne, on peut montrer que $\hat{\mathbf{u}}^1 \in \hat{U}_c^1$, correspondant au minimum du problème contraint pour une seule composante non nulle, est atteint pour $\omega = \{k\}$ tel que $\psi_k = \frac{\theta_1 + \theta_2}{2}$. Ce point est un minimum local de J_{ℓ_0} d'après le théorème 5.1. Ce minimum local de J_{ℓ_0} apparait comme un minimum local sur la surface représentée $\min_{u_{\omega_1}, u_{\omega_2}} J_{\ell_0}(\lambda = 9.5, \mathbf{u} = C_{\omega}(\mathbf{u}_{\omega}))$. Il apparait sur cette représentation dans une très fine ligne de minima situés sur la diagonale. L'étalement de la pénalité CEL0 autour de ce point particulier fait apparaitre une grande vallée sur la surface représentée sur la figure 5.3(b), avec un grand nombre de points approximativement au même niveau. Ce phénomène est d'autant plus gênant dans ce cas particulier où la valeur du minimum pour une seule composante non nulle $\min_{u_{\omega_1}} J_{\ell_0}$ en $\frac{\theta_1 + \theta_2}{2}$ est proche de la valeur $\min_{u_{\omega_1}, u_{\omega_2}} J_{\ell_0}$ correspondant au couple (θ_1, θ_2) . C'est le cas pour des sources proches (dont les directions sont à l'intérieur du même lobe d'antenne), et pour un paramètre de régularisation λ choisi proche de la borne supérieure λ_2^+ : d'après la figure 4.11, $J_{\ell_0}(\lambda, \hat{\mathbf{u}}^1) = \frac{1}{2}c_1 + \lambda$ est dans ce cas à peine supérieur à $J_{\ell_0}(\lambda, \hat{\mathbf{u}}^2)$.

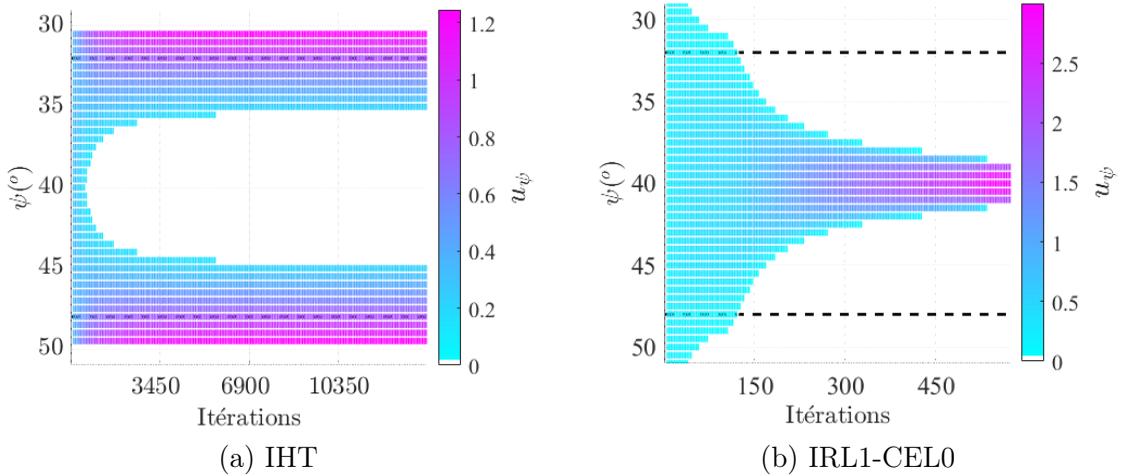


FIGURE 5.4 – Vecteurs solutions estimés au cours des itérations pour les algorithmes IHT et IRL1-CEL0, initialisés par le vecteur nul. Les vraies directions d'arrivées $\theta_1 = 32^\circ$ et $\theta_2 = 48^\circ$ sont indiquées par les lignes noires pointillées. Les composantes correspondantes de la solution optimale valent 7, les autres sont toutes nulles.

Plaçons nous dans un cas encore plus draconien et étudions le comportement des algorithmes de minimisation IHT (algorithme 2.3) et IRL1-CEL0 (algorithme 2.5 avec une

pondération calculée selon l'équation (2.50)). On considère un scénario similaire au scénario utilisé précédemment et décrit en 5.2.2. Les directions d'arrivées des sources sont cette fois fixées à $\theta_1 = 32^\circ$ et $\theta_2 = 48^\circ$, et le paramètre de régularisation est fixé à $\lambda = 0.78$ conformément au choix hors-ligne indépendant des directions d'arrivées dicté dans 4.4. On vérifie très rapidement que pour ce scénario asymptotique, $\lambda_2^+(\mathbf{y}) \approx 1.09$, donc on a bien $\lambda_2^-(\mathbf{y}) < \lambda < \lambda_2^+(\mathbf{y})$.

Les deux algorithmes sont initialisés à $\mathbf{0}$. Sur la figure 5.4, on représente les vecteurs solutions estimés au cours des itérations par les deux algorithmes. On note $\hat{\mathbf{u}}_{\ell_0}^{(i_{\text{end}})}$ et $\hat{\mathbf{u}}_{\text{CEL0}}^{(i_{\text{end}})}$ les vecteurs estimés finaux obtenus respectivement par IHT et IRL1-CEL0. L'estimée $\hat{\mathbf{u}}_{\ell_0}^{(i_{\text{end}})}$ contient un grand nombre de composantes non nulles qui forment deux lobes biaisés autour des vraies directions d'arrivées : on peut vérifier que le vecteur obtenu est un minimiseur local de J_{ℓ_0} à l'aide de l'équation (5.5). L'algorithme IHT est ainsi rapidement coincé dans un minimum local de J_{ℓ_0} ; l'ajout de bruit rend les résultats obtenus par IHT encore moins fiables. L'estimée $\hat{\mathbf{u}}_{\text{CEL0}}^{(i_{\text{end}})}$ contient elle seulement 5 composantes non nulles, formant un lobe situé au milieu des vraies directions d'arrivées. De même, en notant $\hat{\omega} = \text{supp}(\hat{\mathbf{u}}_{\text{CEL0}}^{(i_{\text{end}})})$, on vérifie que $\mathbf{D}_{\hat{\omega}}^H \mathbf{D}_{\hat{\omega}} \hat{\mathbf{u}}_{\text{CEL0}}^{(i_{\text{end}})} = \mathbf{D}_{\hat{\omega}}^H \mathbf{y}$ et $\text{rank}(\mathbf{D}_{\hat{\omega}}) = \#\hat{\omega} = 5$: il s'agit donc aussi d'un minimiseur local de J_{ℓ_0} . Pour des sources proches, l'algorithme IRL1-CEL0 est ainsi rapidement attiré dans le large bassin d'attraction du minimum local situé en $\hat{\mathbf{u}}^1 \in \hat{U}_c^1$.

5.4 Exploitation de la complémentarité des surfaces J_{ℓ_0} et J_{CEL0} dans un algorithme itératif adaptatif

5.4.1 Principe et algorithme

Pour des sources bien séparées, l'algorithme IRL1-CEL0 peut être utilisé pour minimiser la fonction régularisée J_{ℓ_0} en partant d'une initialisation par le vecteur nul, ce qui permet d'obtenir de bien meilleurs résultats statistiques qu'avec l'algorithme IHT. On exploite dans ce cas l'avantage principal de la surface J_{CEL0} , à savoir la réduction du nombre de minima locaux par rapport à J_{ℓ_0} et la préservation du minimum global. Pour des sources proches en revanche, la suppression de minima locaux entraîne l'élargissement de mauvais bassins d'attraction dans J_{CEL0} , et l'algorithme IRL1-CEL0 ne permet pas dans ce cas de séparer les deux sources.

J'ai ainsi proposé un nouveau schéma d'optimisation reposant sur la complémentarité entre les deux surfaces d'erreurs. La Direction Générale de l'Armement (DGA) ayant interdit la divulgation de celui-ci avec le niveau de protection "Diffusion Restreinte - Spécial France", ce schéma d'optimisation est détaillé en annexe A et ne peut être divulgué sur internet. Dans la suite, les résultats obtenus avec celui-ci seront donc donnés sous le nom "Algorithme proposé".

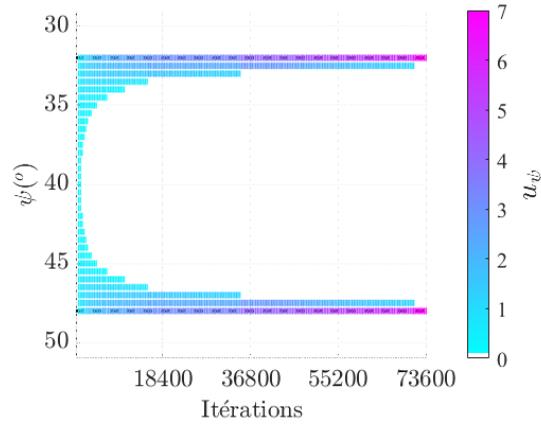


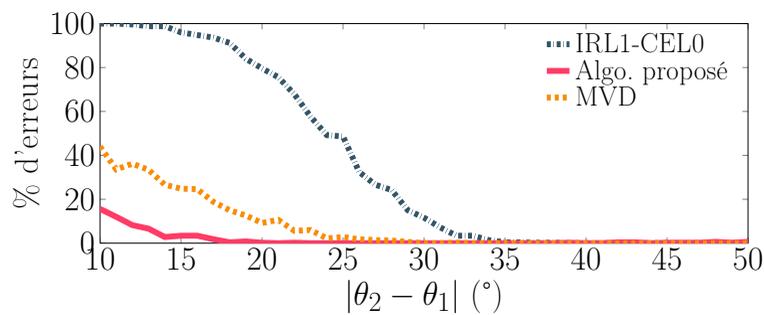
FIGURE 5.5 – Vecteur solution estimé au cours des itérations par l’algorithme proposé initialisé par le vecteur nul. Les vraies directions d’arrivées $\theta_1 = 32^\circ$ et $\theta_2 = 48^\circ$ sont indiquées par les lignes noires pointillées. Les composantes correspondantes de la solution optimale valent 7, les autres sont toutes nulles. Les paramètres de l’algorithme sont ceux donnés dans l’annexe A qui ne peut être divulguée en ligne.

On représente sur la figure 5.5 le vecteur solution estimé au cours des itérations par l’algorithme proposé pour un scénario similaire à celui décrit en 5.2.2 avec des directions d’arrivées $\theta_1 = 32^\circ$ et $\theta_2 = 48^\circ$, avec l’utilisation des paramètres donnés dans l’annexe A. On peut alors comparer aux solutions obtenues par IHT et IRL1-CEL0 et présentées dans la figure 5.4 : pour ce scénario asymptotique, l’algorithme proposé est le seul à atteindre la solution globale de J_{ℓ_0} , avec deux uniques composantes non nulles dont la valeur est 7. A noter cependant que la bonne détection des deux directions d’arrivées se fait au détriment d’un plus grand nombre d’itérations que dans les algorithmes IRL1-CEL0 et IHT. Cette augmentation du nombre d’itérations s’observe uniquement dans les scénarios les plus difficiles de sources proches : pour des sources éloignées, le nombre d’itérations effectuées par l’algorithme proposé est du même ordre de grandeur que le nombre d’itérations effectuées par IRL1-CEL0 ou IHT. Précisons également qu’un plus grand nombre d’itérations des algorithmes IRL1-CEL0 et IHT (en modifiant les critères d’arrêts) ne permet pas d’améliorer les solutions obtenues présentées sur la figure 5.4.

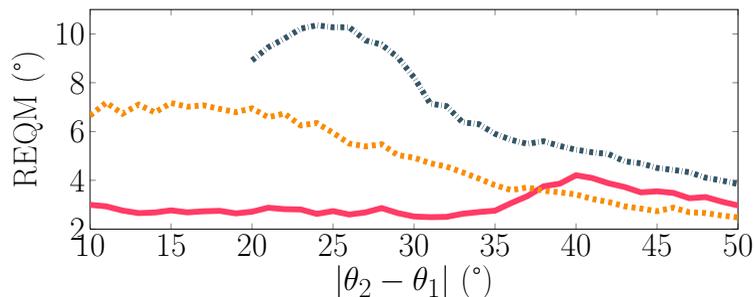
5.4.2 Évaluation des performances

Étudions maintenant les performances de l’algorithme proposé pour des observations stochastiques bruitées. Nous nous plaçons dans un premier temps dans le scénario 2 décrit en 4.3.3.1 avec des sources cohérentes en présence d’erreurs de modèle. Le paramètre de régularisation λ est fixé selon le choix effectué dans le paragraphe 4.4 pour ce scénario, avec $\sigma_e = 0.15$. Dans ce scénario, les performances des algorithmes parcimonieux sont comparées aux performances du maximum de vraisemblance déterministe. Sur la figure 5.6, on représente le pourcentage d’erreurs et la REQM obtenue par l’algorithme IRL1-CEL0 et l’algorithme proposé tous les deux initialisés par le vecteur nul, comparés aux

performances d'une méthode de Gauss-Newton utilisée pour la minimisation du critère du maximum de vraisemblance et initialisée aux vraies directions d'arrivées. L'algorithme IRL1-CEL0 a une limite de résolution d'approximativement 35° , alors que celle du maximum de vraisemblance est d'environ 25° . La résolution obtenue avec l'algorithme proposé est de 18° , soit une amélioration de 17° par rapport à l'algorithme IRL1-CEL0 qui utilise la même initialisation par le vecteur nul, et une amélioration de 7° par rapport à la résolution du maximum de vraisemblance par un algorithme de Gauss-Newton initialisé aux vraies directions d'arrivées. La précision obtenue avec l'algorithme proposé est également meilleure que celle obtenue avec l'algorithme IRL1-CEL0 pour l'ensemble des écarts angulaires considérés. Pour un écart compris entre 37 et 50° , c'est la méthode de Gauss-Newton appliquée au maximum de vraisemblance qui présente la meilleure précision : il faut néanmoins prendre en compte son initialisation, et le fait que l'estimation se fait dans le domaine continu alors que les angles estimés par les méthodes parcimonieuses utilisées appartiennent à la grille de discrétisation avec un pas de 1° .



(a) Pourcentage d'erreurs en fonction de l'écart entre les sources.

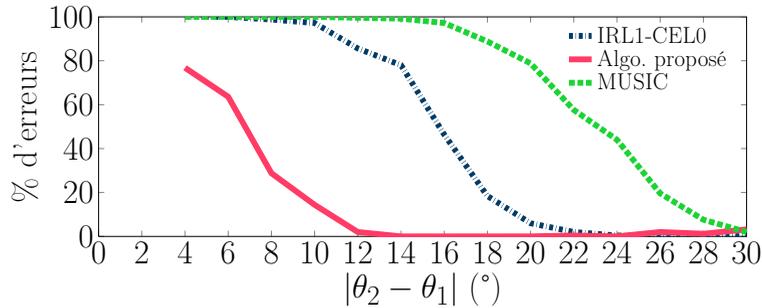


(b) REQM en fonction de l'écart entre les sources.

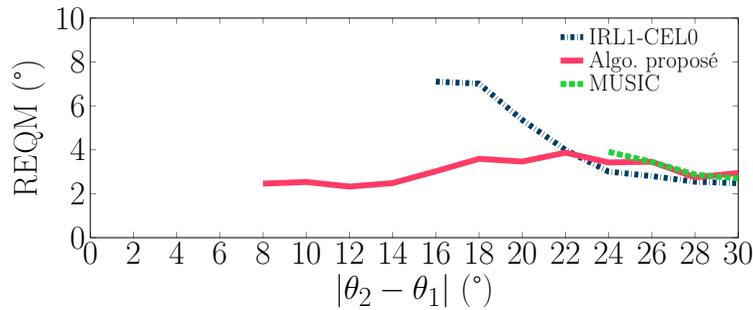
FIGURE 5.6 – Performances de l'algorithme proposé initialisé par le vecteur nul en fonction de l'écart entre les sources, comparées aux performances obtenues par IRL1-CEL0 (initialisé à 0) et une méthode de Gauss-Newton appliquée au maximum de vraisemblance déterministe (initialisée aux vraies directions), dans le cadre du scénario 2 de sources cohérentes en présence d'erreurs de modèle.

Nous étudions maintenant les performances de l'algorithme proposé pour le scénario 3 de deux sources indépendantes à BT fini, avec utilisation de la matrice de covariance vectorisée tel que décrit dans le paragraphe 4.3.3.2. Celles-ci sont comparées sur la figure

5.7 aux performances de l'algorithme IRL1-CEL0, initialisé au vecteur nul, et de la méthode MUSIC, pour un RSB égal à 0 dB. Le paramètre de régularisation est fixé par la valeur α_{lim} décrite dans le paragraphe 4.4, en correspondance avec la figure 4.18. Le gain en terme de résolution avec l'algorithme proposé est de 10° par rapport à l'algorithme IRL1-CEL0, et de 18° par rapport à la méthode MUSIC. La précision en terme de REQM est quasiment identique pour tous les algorithmes dans leurs plages de fonctionnement.



(a) Pourcentage d'erreurs en fonction de l'écart entre les sources.



(b) REQM en fonction de l'écart entre les sources.

FIGURE 5.7 – Performances de l'algorithme proposé initialisé par le vecteur nul en fonction de l'écart entre les sources, comparées aux performances obtenues par IRL1-CEL0 (initialisé à 0) et par la méthode MUSIC, dans le cadre du scénario 3 de sources indépendantes et à BT fini.

5.4.3 Perspectives

Les performances obtenues avec l'algorithme proposé à la fois pour un scénario de sources indépendantes et pour un scénario de sources cohérentes surpassent les performances obtenues par l'algorithme IRL1-CEL0 ou avec les méthodes conventionnelles de goniométrie utilisées pour ces scénarios, à savoir la méthode MUSIC et une mise en œuvre de la méthode du maximum de vraisemblance, respectivement. L'algorithme proposé repose sur des étapes peu coûteuses en temps de calcul, et pour lesquelles la complexité est fixe relativement au nombre de sources. Cet avantage est considérable par rapport aux méthodes optimales du maximum de vraisemblance, d'autant plus coûteuses en temps de calcul que le nombre de paramètres à estimer augmente.

Dans le cas de sources indépendantes, ces bonnes performances laissent envisager l'estimation de directions d'arrivées de plus de sources que de capteurs par l'utilisation de la matrice de covariance vectorisée. Cette perspective a été testée sur un scénario de $M = 10$ sources indépendantes, incidentes sur un réseau UCA de $N = 5$ capteurs et de rayon une demi-longueur d'onde. La méthode MUSIC ne peut fonctionner dans ce scénario où le nombre de sources est supérieur au nombre de capteurs. La matrice de covariance est estimée sur $K = 1500$ échantillons pour un RSB par source de 20dB. La figure 5.8 représente un pseudo-spectre représentatif d'une solution obtenue par l'algorithme proposé et IRL1-CEL0, initialisés tous les deux par le vecteur nul. L'algorithme proposé permet de détecter les 10 sources, alors que l'algorithme IRL1-CEL0 donne une solution avec seulement 9 composantes non nulles : la contribution de la 5ième source n'apparaît plus dans la solution. Cette source possède par ailleurs le plus petit écart angulaire avec ses voisins. Le pouvoir de résolution de l'algorithme proposé, supérieur à celui d'IRL1-CEL0 en bi-sources, semble également être un avantage pour la détection d'un grand nombre de sources. Les performances stochastiques et les limites de fonctionnement de l'algorithme restent cependant encore à établir pour un tel scénario. Notons que pour cette simulation (et pour plusieurs autres nombres de sources), la valeur du paramètre de régularisation λ utilisée est $\alpha_{\text{lim}} \cdot \|\mathbf{y}\|_2^2$, pour la valeur α_{lim} déterminée dans le cas bi-sources dans le paragraphe 4.4, ce qui amène une perspective également très intéressante pour le réglage du paramètre de régularisation.

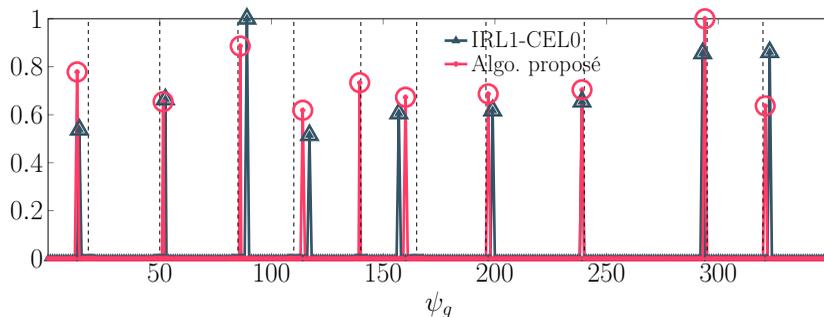


FIGURE 5.8 – Pseudo-spectre des vecteurs estimés par les algorithmes proposé et IRL1-CEL0 pour un scénario de 10 sources indépendantes. $K = 1500$, $RSB = 20$ dB.

Conclusion et perspectives

Dans cette thèse, j'ai étudié les méthodes parcimonieuses pour la goniométrie et proposé de nouvelles mises en œuvre dans des contextes opérationnels difficiles dans lesquels les méthodes de goniométrie usuelles ne fonctionnent pas. Je me suis intéressée à deux contextes en particulier. Le premier contexte est le contexte aéroporté, où l'on est en présence d'un grand nombre de sources indépendantes, nombre qui peut être supérieur au nombre de capteurs. Le second contexte est l'environnement urbain, où le réseau reçoit plusieurs multi-trajets fortement corrélés voir cohérents d'un même émetteur.

La thèse se consacre à la mise en œuvre d'une méthode parcimonieuse reposant sur un critère régularisé par la norme ℓ_0 . Les algorithmes considérés dans cette mise en œuvre sont des algorithmes à gradient-proximal, dont les opérations sont très peu coûteuses.

Dans le chapitre 3, j'ai proposé et étudié des modèles d'observation spécifiques à nos scénarios. L'objectif était de concentrer toute l'information utile de plusieurs instants de mesure, présente initialement dans la matrice de covariance, dans un unique vecteur qui s'exprime sous forme parcimonieuse. Pour des sources indépendantes dans un contexte aéroporté, l'utilisation de la matrice de covariance vectorisée équivaut à la formation d'un réseau virtuel comprenant beaucoup plus de capteurs que dans le réseau initial, ce qui donne la possibilité d'estimer les directions d'arrivées de plus de sources que de capteurs physiques. Pour des sources cohérentes ou même fortement corrélées en environnement urbain, le premier vecteur propre de la matrice de covariance est une combinaison linéaire des contributions des différents multi-trajets et suffit pour l'estimation des directions d'arrivées de ceux-ci. L'établissement des statistiques de ces observations était un pré-requis pour le développement du chapitre 4.

Le chapitre 4 est entièrement consacré au paramètre de régularisation, qui contrôle le poids relatif entre l'attache aux données et la régularisation par la norme ℓ_0 dans le critère régularisé J_{ℓ_0} . Ce paramètre est souvent empiriquement réglé dans la littérature, alors que son choix est pourtant déterminant. J'ai montré que la relation entre le problème régularisé et le problème contraint donné par Mila Nikolova correspondait à l'équivalence entre problème régularisé et critère du maximum de vraisemblance déterministe. Cette relation permet de déterminer un intervalle pour le paramètre de régularisation tel que le minimiseur global du critère J_{ℓ_0} corresponde au minimiseur du critère du maximum de vraisemblance, et donc à la solution recherchée. J'ai étendu cette relation, établie initialement dans le cas

d'une observation déterministe donnée, à des observations stochastiques. Les approximations proposées permettent de calculer pour un grand nombre de scénarios un intervalle dans lequel doit se trouver le paramètre de régularisation pour permettre l'estimation des $M = 2$ directions d'arrivées. Les simulations montrent l'importance pratique de cet intervalle théorique : les algorithmes à gradient-proximal tels que FB-CEL0, IRL1-CEL0 ou IHT permettent, selon leurs initialisations, d'obtenir de bonnes performances d'estimation pour des valeurs de λ dans l'intervalle. L'étude des bornes de l'intervalle admissible permet alors de choisir de façon hors-ligne et indépendamment du scénario de sources et de bruit une valeur de λ à utiliser dans les algorithmes. Les simulations réalisées en lien avec ce choix montrent que l'on peut déterminer des limites de fonctionnement théoriques (limite de résolution par exemple) des algorithmes parcimonieux à gradient-proximal à partir de l'étude sur le paramètre de régularisation.

Cependant, les algorithmes de descente à gradient-proximal utilisés garantissent au mieux de trouver un minimum local du critère J_{ℓ_0} . Les performances peuvent alors dépendre de leur initialisation. L'objectif du chapitre 5 est de proposer un schéma d'optimisation qui prenne en compte les avantages des critères J_{ℓ_0} et J_{CEL0} et qui permette alors d'utiliser une initialisation simple comme le vecteur nul dans tous les scénarios considérés. Dans ce chapitre, j'étudie les minimiseurs locaux du critère régularisé par la norme ℓ_0 puis par la pénalité CEL0 afin de comprendre les mécanismes en jeu. A partir de l'observation des deux surfaces d'erreurs, je propose un nouveau schéma d'optimisation itératif. L'algorithme mis en œuvre permet d'améliorer de façon significative le pouvoir de résolution des méthodes parcimonieuses par rapport aux méthodes conventionnelles.

Perspectives

Les approches utilisées dans cette thèse méritent encore des approfondissements, qui pourront faire l'objet de futurs travaux.

- Il reste à déterminer les statistiques du premier vecteur propre à BT fini pour en déduire les statistiques des bornes de l'intervalle pour λ dans ce scénario.
- Nous avons remarqué un lien empirique entre le niveau de bruit, la corrélation spatiale entre les sources, et la courbe $F_{\lambda_2}^{-1}(1-\epsilon)$ tracée en fonction de l'écart entre les sources. Il est nécessaire d'approfondir l'étude sur les bornes de l'intervalle admissible pour λ et de les relier à ces paramètres.
- L'étude statistique des bornes de l'intervalle, effectuée dans cette thèse en fonction de l'écart entre les sources, peut être étendue à l'écart de puissance entre sources, ce qui amènera à d'autres limites de fonctionnement des méthodes parcimonieuses.
- Les statistiques de λ_M^- et λ_M^+ sont développées dans le chapitre 4 uniquement dans le cas $M = 2$ sources. Il est nécessaire d'étendre l'étude à tout nombre de sources. La normalisation de l'observation proposée à la fin du chapitre 4 pour normaliser l'intervalle de λ semble être d'après les dernières simulations un bon point de départ

afin de pouvoir proposer un choix de λ valable pour n'importe quel nombre de sources.

- L'algorithme proposé nécessite le réglage de paramètres, pour l'instant sous forme de critères d'arrêts adaptatifs. Ceux-ci mériteraient une étude plus approfondie, notamment pour permettre de réduire le nombre d'itérations du schéma proposé.

Tous ces points permettraient de finaliser les travaux dans les contextes opérationnels de la goniométrie 1D en présence de trajets cohérents ou de la goniométrie d'un grand nombre de sources en aéroporté. Toutefois, il existe d'autres contextes opérationnels qui pourraient bénéficier des avantages de l'estimation parcimonieuse, à savoir la goniométrie 2D ou en diversité de polarisation, ou encore de multi-trajets partiellement corrélés.

- Dans le cadre des communications HF, les ondes se propagent par réflexion sur les couches ionosphériques : pour un même émetteur, plusieurs trajets corrélés voir cohérents arrivent suivant le même gisement mais avec des élévations différentes. Les études présentées dans cette thèse peuvent être étendus aux trajets 2D. Il est nécessaire de développer de nouveaux modèles d'observation parcimonieuse dans ce cas : pour des sources corrélées temporellement, il est nécessaire d'utiliser soit plusieurs vecteurs propres, soit la matrice de covariance vectorisée prenant en compte les termes croisés.
- En HF et également en V/UHF, les réseaux de goniométrie en diversité de polarisation sont de plus en plus courants pour mieux exploiter les caractéristiques électromagnétiques des ondes incidentes. Il faut en conséquence estimer la polarisation d'une onde incidente conjointement à sa direction d'arrivée. Cela met en jeu de nouveaux modèles d'observation des signaux.

Ces deux nouveaux scénarios évoqués peuvent alors faire appel à des modèles d'observation MMV. Les travaux sur le paramètre de régularisation devront alors être étendus à de telles observations.

Annexe A

Annexe relative au chapitre 5

Cette annexe décrit en détails le schéma d'optimisation proposé dans le chapitre 5 et les paramètres utilisés dans les simulations présentées. La DGA ayant interdit la divulgation de celui-ci, cette annexe ne peut être reproduite librement en ligne. Les informations contenues dans cette annexe ne doivent être communiquées qu'aux personnes qui ont besoin de les connaître pour nécessité du service.

Bibliographie

- [1] A. Antoniadis, I. Gijbels, and M. Nikolova, “Penalized Likelihood Regression for Generalized Linear Models with Nonquadratic Penalties,” **Annals of the Institute of Statistical Mathematics**, vol. 63, pp. 585–615, 2011.
 - [2] M. Atashbar and M. H. Kahaei, “Direction-of-arrival estimation using amlss method,” **IEEE Latin America Transactions**, vol. 10, no. 5, pp. 2053–2058, 2012.
 - [3] H. Attouch, J. Bolte, and B. Fux Svaiter, “Convergence of descent methods for semi-algebraic and tame problems : proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods,” **Mathematical Programming**, vol. 137, no. 1-2, pp. 91–129, 2013.
 - [4] G. Beaumont, R. Fablet, and A. Drémeau, “An approximate message passing approach for doa estimation in phase noisy environments,” in **Latent Variable Analysis and Signal Separation**, Y. Deville, S. Gannot, R. Mason, M. D. Plumbley, and D. Ward, Eds. Springer International Publishing, 2018, pp. 385–394.
 - [5] A. Beck and M. Teboulle, “A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems,” **SIAM Journal on Imaging Sciences**, vol. 2, no. 1, pp. 183–202, 2009.
 - [6] G. Bienvenu, “Propriétés haute résolution de la matrice de corrélation spatiale,” in **9ième GRETSI**, Nice, France, 1983.
 - [7] G. Bienvenu and L. Kopp, “Optimality of High Resolution Array Processing Using the Eigensystem Approach,” **IEEE Transactions on Acoustics, Speech, and Signal Processing**, 1983.
 - [8] A. Biswas and S. Reisenfeld, “New High Resolution Direction of Arrival Estimation using Compressive Sensing,” **2017 IEEE 22nd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)**, 2017.
 - [9] T. Blumensath and M. E. Davies, “Iterative Thresholding for Sparse Approximations,” **Journal of Fourier Analysis and Applications**, 2008.
 - [10] —, “Normalized iterative hard thresholding : Guaranteed stability and performance,” **IEEE Journal on Selected Topics in Signal Processing**, 2010.
 - [11] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet, “Characterizations of Lojasiewicz inequalities : sub-gradient flows, talweg, convexity,” **Transactions of the American Mathematical Society**, vol. 362, no. 6, pp. 3319–3363, 2010.
 - [12] E. J. Candes, “The Restricted Isometry Property and Its Implications for Compressed Sensing,” **Comptes Rendus Mathématiques**, vol. 346, no. 9, 2008.
-

-
- [13] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles : exact signal reconstruction from highly incomplete frequency information," **IEEE Transactions on Information Theory**, vol. 52, no. 52, 2006.
- [14] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted L1 minimization," **Journal of Fourier Analysis and Applications**, 2008.
- [15] J. Capon, "High-Resolution Frequency-Wavenumber Spectrum Analysis," **Proc. of the IEEE**, vol. 57, 1969.
- [16] S. Chen, S. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," **Int. J. Control**, vol. 50, no. 5, pp. 1873–1896, 1989.
- [17] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," **SIAM Journal on Scientific Computing**, 1998.
- [18] A. Chinatto, E. Soubies, C. Junqueira, J. M. Romano, P. Larzabal, J. P. Barbot, and L. Blanc-Féraud, "L0-optimization for channel and DOA sparse estimation," in **2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2015**, 2015.
- [19] F. H. Clarke, **Optimization and nonsmooth analysis**. siam, 1990, vol. 5.
- [20] P. L. Combettes and V. R. Wajs, "Signal Recovery by Proximal Forward-Backward Splitting," **Multiscale Modeling & Simulation**, 2005.
- [21] W. Cui, T. Qian, and J. Tian, "Enhanced covariances matrix sparse representation method for doa estimation," **Electronics Letters**, vol. 51, no. 16, pp. 1288–1290, 2015.
- [22] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," **Communications on Pure and Applied Mathematics**, vol. 57, no. 11, pp. 1413–1457, 2004.
- [23] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," **Proceedings of the National Academy of Sciences**, 2003.
- [24] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," **IEEE Transactions on Information Theory**, vol. 47, pp. 2845–2862, 2001.
- [25] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," **Biometrika**, vol. 81, no. 3, pp. 425–455, 1994.
- [26] J. Duan, C. Sossens, D. Brie, and J. Idier, "A continuation approach to estimate a solution path of mixed ℓ_2 - ℓ_0 minimization problems," in **Workshop : Signal Processing with Adaptive Sparse Structured Representations, SPARS'09**, Apr 2009.
- [27] M. Evans, N. Hastings, and B. Peacock, **Statistical Distributions**. John Wiley & Sons, Inc., 1993, vol. 2nd ed.
- [28] J. Fan and R. Li, "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties," pp. 1348–1360, 2001.
- [29] A. Ferreol, P. Larzabal, and M. Viberg, "On the asymptotic performance analysis of subspace doa estimation in the presence of modeling errors : case of music," **IEEE Transactions on Signal Processing**, vol. 54, no. 3, pp. 907–920, 2006.
-

-
- [30] —, “Modeling error sensitivity of the music algorithm conditioned on resolved sources,” in **16th European Signal Processing Conference (EUSIPCO 2008)**. EURASIP, 25–29 Aug 2008.
- [31] —, “Performance prediction of maximum-likelihood direction-of-arrival estimation in the presence of modeling errors,” **IEEE Transactions on Signal Processing**, vol. 56, no. 10, pp. 4785–4793, 2008.
- [32] S. Foucart and M.-J. Lai, “Sparsest solutions of underdetermined linear systems via ℓ_q -minimisation for $0 < q \leq 1$,” **Applied and Computational Harmonic Analysis**, vol. 3, no. 26, 2009.
- [33] B. Friedlander, “A Sensitivity Analysis of the MUSIC Algorithm,” **IEEE Transactions on Acoustics, Speech, and Signal Processing**, vol. 38, no. 10, 1990.
- [34] J. J. Fuchs, “Identification of real sinusoids in noise, the Global Matched Filter approach,” in **IFAC Proceedings Volumes (IFAC-PapersOnline)**, 2009.
- [35] J.-J. Fuchs, “Doa estimation in the presence of modeling errors, the global matched filter approach,” in **17th European Signal Processing Conference (EUSIPCO 2009)**. EURASIP, 24–28 Aug 2009, pp. 1963–1967.
- [36] S. Gazagnes, E. Soubies, and L. Blanc-Féraud, “High density molecule localization for super-resolution microscopy using cel0 based sparse approximation,” in **14th International Symposium on Biomedical Imaging (ISBI 2017)**. IEEE, 2017, pp. 28–31.
- [37] M. Grant and S. Boyd, “CVX : Matlab software for disciplined convex programming, version 2.1,” <http://cvxr.com/cvx>, mar 2014.
- [38] J. Gurland, “Distribution of ratios of quadratic forms,” **Cowles Commission Discussion Paper**, vol. 354, 1951.
- [39] E. Hannan, **Multiple Time Series**. John Wiley and Sons, New York, 1970.
- [40] C. Hansen and D. O’Leary, “The use of the L-curve in the regularization of discrete ill-posed problems,” **SIAM Journal on Scientific Computing**, vol. 14, pp. 1487–1503, 1993.
- [41] C. Hansen, “The l-curve and its use in the numerical treatment of inverse problems,” in **Computational Inverse Problems in Electrocardiology**, 2001, vol. 4, pp. 119–142.
- [42] Z. He, Q. Liu, L. Jin, and S. Ouyang, “Low complexity method for DOA estimation using array covariance matrix sparse representation,” **Electronics Letters**, vol. 49, pp. 228–230, 01 2013.
- [43] S. Hochreiter and J. Schmidhuber, “Flat minima,” **Neural computation**, vol. 9, pp. 1–42, 02 1997.
- [44] M. M. Hyder and K. Mahata, “Direction-of-Arrival Estimation Using a Mixed $L_2;0$ Norm Approximation,” **IEEE Transactions on Signal Processing**, vol. 58, no. 9, 2010.
- [45] J. Imhof, “Computing the distribution of quadratic forms in normal variables,” **Biometrika**, no. 48, pp. 419–426, 1961.
- [46] H. Liu, Y. Tang, and H. H. Zhang, “A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables,” **Computational Statistics and Data Analysis**, vol. 53, no. 4, pp. 853–856, feb 2009.
- [47] J. Liu, W. Zhou, and F. H. Juwono, “Joint smoothed l0-norm doa estimation algorithm for multiple measurement vectors in mimo radar,” **Sensors**, vol. 17, no. 5, 2017.
-

-
- [48] D. Malioutov, M. Çetin, and A. S. Willsky, “A Sparse Signal Reconstruction Perspective for Source Localization With Sensor Arrays,” **IEEE Transactions on Signal Processing**, vol. 53, no. 8, 2005.
- [49] S. G. Mallat and Z. Zhang, “Matching Pursuits With Time-Frequency Dictionaries,” **IEEE Transactions on Signal Processing**, vol. 41, no. 12, 1993.
- [50] A. Marmin, M. Castella, and J.-C. Pesquet, “How to Globally Solve Non-convex Optimization Problems Involving an Approximate L0 Penalization,” **ICASSP 2019, IEEE International Conference on Acoustics, Speech and Signal Processing**, pp. 5601–5605, apr 2019.
- [51] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, “A fast approach for overcomplete sparse decomposition based on smoothed L0norm,” **IEEE Transactions on Signal Processing**, 2009.
- [52] A. A. Mohsenipour and S. B. Provost, “On the Distribution of Quadratic Expressions in Various Types of Random Vectors,” Ph.D. dissertation, 2012.
- [53] B. Natarajan, “Sparse Approximate Solutions to Linear Systems,” **SIAM Journal on Computing**, vol. 24, no. 2, pp. 227–234, April 1995.
- [54] D. Needell and J. A. Tropp, “CoSaMP : Iterative signal recovery from incomplete and inaccurate samples,” **Applied and Computational Harmonic Analysis**, 2009.
- [55] R. Nelsen, **An Introduction to Copulas**, second edition ed., ser. Springer Series in Statistics. Springer, 2006.
- [56] M. Nikolova, “Description of the minimizers of least squares regularized with ℓ_0 -norm. Uniqueness of the global minimizer,” **SIAM J. Imaging Sciences**, vol. 6, no. 2, 2013.
- [57] —, “Relationship between the optimal solutions of least squares regularized with L0-norm and constrained by k-sparsity,” **Applied and Computational Harmonic Analysis**, vol. 41, no. 1, pp. 237–265, 2016.
- [58] D. Noll, “Convergence of Non-smooth Descent Methods Using the Kurdyka-Lojasiewicz Inequality,” **Journal of Optimization Theory and Applications**, 2013.
- [59] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, “On Iteratively Reweighted Algorithms for Nonsmooth Nonconvex Optimization in Computer Vision,” **SIAM Journal on Imaging Sciences**, 2015.
- [60] Y. Pati, R. Rezaifar, and P. Krishnaprasad, “Orthogonal matching pursuit : Recursive function approximation with applications to wavelet decomposition,” **Conference Record of the 27th Asilomar Conference on Signals, Systems and Computers**, pp. 40–44, 1993.
- [61] E. Pearson, “Note on an approximation to the distribution of non-central chi-square,” **Biometrika**, no. 46, p. 364, 1959.
- [62] D. Peleg and R. Meir, “A bilinear formulation for vector sparsity optimization,” **Signal Processing**, vol. 2, no. 88, 2008.
- [63] J. S. Picard and A. J. Weiss, “Direction finding of multiple emitters by spatial sparsity and linear programming,” in **9th International Symposium on Communications and Information Technology, ISCIT**, 2009.
- [64] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, “Subset selection in noise based on diversity measure minimization,” **IEEE Transactions on Signal Processing**, 2003.
-

-
- [65] A. Renaux, P. Forster, E. Chaumette, and P. Larzabal, "On the high-snr conditional maximum-likelihood estimator full statistical characterization," **IEEE Transactions on Signal Processing**, vol. 54, no. 12, pp. 4840–4843, 2006.
- [66] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," **IEEE Transactions on Antennas and Propagation**, 1986.
- [67] I. Selesnick, "Sparse Regularization via Convex Analysis," **IEEE Transactions on Signal Processing**, vol. 65, no. 17, pp. 4481–4494, 2017.
- [68] E. Soubies, A. Chinatto, P. Larzabal, J. M. T. Romano, and L. Blanc-Féraud, "Direction-of-arrival estimation through exact continuous $\ell_{2,0}$ -norm relaxation," **IEEE Signal Processing Letters**, vol. 28, pp. 16–20, 2021.
- [69] E. Soubies, L. Blanc-Féraud, and G. Aubert, "A Continuous Exact L0 penalty (CEL0) for least squares regularized problem," **SIAM J. Imaging Sciences**, vol. 8, 2015.
- [70] E. Soubies, L. Blanc-Féraud, and G. Aubert, "New Insights on the Optimality Conditions of the l2-l0 Minimization Problem," **Journal of Mathematical Imaging and Vision**, vol. 62, pp. 808–824, Jul. 2020.
- [71] C. Soussen, J. Idier, J. Duan, and D. Brie, "Homotopy based algorithms for L0-regularized least-squares," **IEEE Transactions on Signal Processing**, 2015.
- [72] P. Stoica and A. Nehorai, "Music, maximum likelihood, and cramer-rao bound : further results and comparisons," **IEEE Transactions on Acoustics, Speech, and Signal Processing**, vol. 38, no. 12, pp. 2140–2150, 1990.
- [73] P. Stoica and K. Sharman, "Maximum Likelihood Methods for Direction-Of-Arrival Estimation," **IEEE Transactions on Acoustics, Speech and Signal Processing**, no. 7, 1990.
- [74] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood, and Cramer-Rao Bound," **IEEE Transactions on Acoustics, Speech, and Signal Processing**, vol. 37, no. 5, 1989.
- [75] J. F. Sturm, "Using sedumi 1.02, a MATLAB toolbox for optimization over symmetric cones," **Optimization Methods and Software**, vol. 11, no. 1-4, pp. 625–653, 1999.
- [76] X. Tan, W. Roberts, J. Li, and P. Stoica, "Sparse learning via iterative minimization with application to MIMO radar imaging," **IEEE Transactions on Signal Processing**, 2011.
- [77] V. N. Temlyakov, "Greedy approximation," **Acta Numerica**, vol. 17, p. 235–409, 2008.
- [78] J. A. Tropp, "Greed is good : Algorithmic results for sparse approximation," **IEEE Transactions on Information Theory**, vol. 50, no. 10, pp. 2231–2242, 2004.
- [79] J. A. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," **IEEE Transactions on Information Theory**, vol. 53, 2007.
- [80] W. Wang and R. Wu, "High resolution direction of arrival (doa) estimation based on improved orthogonal matching pursuit (omp) algorithm by iterative local searching," **Sensors**, vol. 13, no. 9, pp. 11 167–11 183, 2013.
- [81] C. Weiss and A. M. Zoubir, "Doa estimation in the presence of array imperfections : A sparse regularization parameter selection problem," in **Proc. IEEE Workshop on Statistical Signal Processing (SSP)**, Gold Coast, Australia, June 2014.
-

- [82] F. Wen, L. Chu, P. Liu, and R. C. Qiu, "A Survey on Nonconvex Regularization-Based Sparse and Low-Rank Recovery in Signal Processing, Statistics, and Machine Learning," pp. 69 883–69 906, 2018.
 - [83] Z. Xu, X. Chang, F. Xu, and H. Zhang, "L1/2regularization : A thresholding representation theory and a fast solver," **IEEE Transactions on Neural Networks and Learning Systems**, 2012.
 - [84] J. Yin and T. Chen, "Direction-of-arrival estimation using a sparse representation of array covariance vectors," **IEEE Transactions on Signal Processing**, 2011.
 - [85] C. H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," **Annals of Statistics**, vol. 38, no. 2, pp. 894–942, 2010.
 - [86] Z. Zhang, X. Wu, C. Li, and W.-P. Zhu, "An ℓ_p -norm based method for off-grid doa estimation," **Circuits Syst. Signal Process.**, vol. 38, no. 2, p. 904–917, Feb. 2019.
-

Titre: Goniométrie parcimonieuse de sources radioélectriques : modèles, algorithmes et mises en œuvre robustes

Mots clés: Goniométrie, estimation parcimonieuse, régularisation L0, optimisation

Résumé: Cette thèse porte sur la goniométrie d'émetteurs radioélectriques non coopératifs à partir de signaux reçus sur un réseau d'antennes. Les applications visées dans ce travail sont le scénario aéroporté, caractérisé par un nombre de sources supérieur au nombre de capteurs, et le scénario en environnement urbain, caractérisé par des multi-trajets cohérents. Les méthodes de goniométrie conventionnelles telles que formation de voies et Capon ou encore les méthodes à haute résolution telles que MUSIC ne sont pas performantes dans de tels scénarios. La méthode du maximum de vraisemblance souffre d'une complexité de calcul incompatible avec les systèmes opérationnels actuels.

Pour palier ces limitations, la goniométrie est traitée ici avec un formalisme parcimonieux, parfaitement adapté à l'utilisation de tables de calibration dans les systèmes opérationnels. Après

avoir montré les intérêts d'une approche reposant sur une régularisation par la norme L0, cette thèse s'attaque aux verrous que sont le paramètre de régularisation et la convergence globale des algorithmes d'optimisation. Pour cela, nous construisons et étudions statistiquement des représentations parcimonieuses adaptées i) aux scénarios aéroportés, et ii) aux environnements urbains. L'équivalence avec le maximum de vraisemblance donné par une formulation contrainte permet alors de déterminer un intervalle théorique admissible pour le paramètre de régularisation. Nous étudions également les minimiseurs et les surfaces d'erreurs de différents critères d'optimisation. Cela nous permet de proposer des schémas itératifs de minimisation augmentant la probabilité de convergence globale et ainsi moins sensibles à l'initialisation. De plus, l'algorithme ainsi proposé permet de séparer des sources proches.

Title: Sparse bearing estimation of radio sources: models, algorithms and robust implementations

Keywords: DOA estimation, sparse estimation, L0 regularization, optimization

Abstract: This thesis deals with the direction-of-arrival (DOA) estimation of non-cooperative radio transmitters from signals received on an antenna array. The applications targeted in this work are the airborne scenario, characterized by a number of sources higher than the number of sensors, and the urban environment scenario, characterized by coherent multipath. Conventional direction-of-arrival estimation methods such as beamforming and Capon method or high resolution methods such as MUSIC are not efficient in such scenarios. The maximum likelihood method suffers from a computational complexity incompatible with current operational systems.

In order to overcome these limitations, the problem of direction-of-arrival estimation is treated here with a sparse formalism, perfectly adapted to the use of calibration tables in operational systems. Af-

ter having shown the interests of an approach based on a regularization by the L0 norm, this thesis tackles the technical issues that are the regularization parameter and the global convergence of optimization algorithms. To this end, we construct and statistically study sparse representations adapted to i) airborne scenarios, and ii) urban environments. The equivalence with the maximum likelihood given by a constrained formulation then allows us to determine a theoretical admissible interval for the regularization parameter. We also study the minimizers and error surfaces of different optimization criteria. This allows us to propose iterative minimization schemes that increase the probability of global convergence and thus are less sensitive to initialization. In that respect, the proposed algorithm enables to separate close sources.