HAL
open science

Caractérisation des performances minimales d'estimation pour des modèles d'observations non-standards

Chengfang Ren

To cite this version:

Chengfang Ren. Caractérisation des performances minimales d'estimation pour des modèles d'observations non-standards. Traitement du signal et de l'image [eess.SP]. Universite paris sud, 2015. Français. NNT : . tel-01237145v1

HAL Id: tel-01237145

https://hal.science/tel-01237145v1

Submitted on 2 Dec 2015 (v1), last revised 1 Feb 2016 (v2)

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.

Université Paris-Sud

Ecole Doctorale S.T.I.T.S.
Laboratoire des Signaux et Systèmes

Discipline : Physique

Thèse de doctorat

Soutenue le 28 septembre 2015 par

Chengfang Ren

Caractérisation des performances minimales d'estimation pour des modèles d'observations non-standards

Composition du jury :

Président du jury :	Cédric Richard	Professeur des universités (Université Nice Sophia-Antipolis)
Rapporteurs :	Jean-Yves Tourneret	Professeur des universités (INP - ENSEEIHT Toulouse)
	Philippe Forster	Professeur des universités (Université Paris-Ouest)
Examinateur :	Karim Abed-Meraim	Professeur des universités (Université d'Orléans)
Directeur de thèse :	Alexandre Renaux	Maître de Conférence (Université Paris Sud)
Encadrant de thèse :	Éric Chaumette	Professeur ISAE (ISAE)
Encadrant de thèse :	Jérôme Galy	Maître de Conférence (Université Montpellier)

Laboratoire des Signaux et Systèmes
Université Paris-Sud/ Centrale-Supélec/CNRS/UMR 8506
3 rue Joliot-Curie, Plateau du Moulon, 91192 Gif-sur-Yvette, France

À mes parents, à mes amis, à mes profs
et à M. Bornes professeur de mathématiques au lycée Voltaire.

Remerciements

Ces trois années de thèse ont été riches en découvertes tant sur le plan scientifique que sur le plan humain. Maintenant que cette page se tourne, j'exprime mes profonds remerciements à toutes les personnes qui m'ont accompagnée dans cette aventure.

Je tiens d'abord à remercier les personnes qui m'ont fait l'honneur d'accepter de faire partie de mon jury de thèse. Je remercie Jean-Yves Tourneret et Philippe Forster qui ont accepté de rapporter cette thèse. Leurs commentaires, critiques et suggestions ont contribué à améliorer la qualité du manuscrit. Je remercie Cédric Richard de m'avoir fait l'honneur de présider le jury de thèse. Je remercie Karim Abed-Meraim de m'avoir fait l'honneur et le plaisir de participer activement à mon jury de thèse.

La qualité de l'encadrement et l'environnement dont j'ai bénéficié ont été remarquable. C'est pourquoi il est important pour moi d'adresser mes sincères remerciements à toutes les personnes qui m'ont aidé à soutenir cette thèse. En tout premier, merci à Alex, mon directeur de thèse, de me faire partager ses recherches. Sa disponibilité et ses nombreuses explications très pédagogiques et m'ont rapidement permis d'apprendre et de comprendre les notions de bornes. Merci également pour les réunions « hors labo » très instructives sur le plan professionnel et sur le plan humain. Merci à mes encadrant Éric et Jérôme pour m'avoir appris à allier intuition, rigueur et ténacité dans le travail de recherche. Leurs idées et leurs nombreux conseils m'ont permis de publier autant. Je suis toujours admiratif devant les compétences de programmation d'Éric. Merci à vous trois de m'avoir permis de voyager malgré le petit budget que l'on dispose.

Je remercie les chercheurs que j'ai eus l'occasion de collaborer durant ma thèse. Merci à Nabil El Korso pour sa vision de la recherche. Merci à Pascal Larzabal pour ses conseils. Merci à Julien Le Kernec pour les trois mois de collaboration fructueuse passée à l'université de Nottingham de Ningbo en Chine. Je tiens à remercier les personnels du laboratoire pour leur disponibilité et leur enthousiasme. Merci à Frédéric Desprez, l'informaticien, pour le développement et la maintenance du Cluster qui m'a été grandement utile pendant les trois années de thèse et notamment durant mon séjour à Ningbo. Merci à Christophe Vignat, professeur exceptionnel, pour m'avoir partagé "sa boîte à outil" pour calculer des intégrales et sa passion autour des mathématiques.

Je remercie mes collègues doctorants et post-doctorants, François, Pierre, José, Amadou, Li, Long, Yuling, Benjamin, Vineeth, Mathieu, Jean-François, Zeina, Etienne, Diane, Maxime, Laurie et Johanna pour tous les moments que nous avons partagés à la pause, au CESFO et aux conférences. Merci Elsa pour les parties de Badminton à midi, merci à Achal, avec qui j'ai partagé le rôle de représentant des doctorants, merci à Olivier pour les balades en vélo à Lyon et merci à Stéphanie pour m'avoir appris Tikz. Une petite pensée à mes camarades, Adrien, Batou, Victor, Kevin, Arnaud, Alice et Cyril, qui vont bientôt soutenir. Et d'autres, Virginie, Lucien et Eugénie, qui viennent de commencer. Je leur souhaite tous bon courage !

Enfin merci à ma famille. Merci à mes parents pour leur amour et leurs encouragements.

Merci de m'avoir laissée libre de mes choix, tout en me permettant de le faire dans les meilleures conditions. Tout cela n'aurait pas été possible sans vous et sans votre soutien.

Chengfang
18 octobre 2015

Rien n'est impossible, seules les limites de nos esprits
définissent certaines choses comme inconcevable.
Marc Levy
Et si c'était vrai...

Résumé

Dans le contexte de l'estimation paramétrique, les performances d'un estimateur peuvent être caractérisées, entre autre, par son erreur quadratique moyenne (EQM) et sa résolution limite. La première quantifie la précision des valeurs estimées et la seconde définit la capacité de l'estimateur à séparer plusieurs paramètres. Cette thèse s'intéresse d'abord à la prédiction de l'EQM "optimale" à l'aide des bornes inférieures pour des problèmes d'estimation simultanée de paramètres aléatoires et non-aléatoires (estimation hybride), puis à l'extension des bornes de Cramér-Rao pour des modèles d'observation moins standards. Enfin, la caractérisation des estimateurs en termes de résolution limite est également étudiée. Ce manuscrit est donc divisé en trois parties :

— Premièrement, nous complétons les résultats de littérature sur les bornes hybrides en utilisant deux bornes bayésiennes : la borne de Weiss-Weinstein et une forme particulière de la famille de bornes de Ziv-Zakaï. Nous montrons que ces bornes "étendues" sont plus précises pour la prédiction de l'EQM optimale par rapport à celles existantes dans la littérature.

— Deuxièmement, nous proposons des bornes de type Cramér-Rao pour des contextes d'estimation moins usuels, c'est-à-dire : (i) Lorsque les paramètres non-aléatoires sont soumis à des contraintes d'égalité linéaires ou non-linéaires (estimation sous contraintes). (ii) Pour des problèmes de filtrage à temps discret où l'évolution des états (paramètres) est régit par une chaîne de Markov. (iii) Lorsque la loi des observations est différente de la distribution réelle des données.

— Enfin, nous étudions la résolution et la précision des estimateurs en proposant un critère basé directement sur la distribution des estimées. Cette approche est une extension des travaux de Oh et Kashyap et de Clark pour des problèmes d'estimation de paramètres multidimensionnels.

Mots clés : Estimation paramétrique, estimateurs au sens du maximum de vraisemblance, estimateurs au sens du maximum *a posteriori*, estimation hybride, analyse de performance, bornes inférieures de l'erreur quadratique moyenne, résolution limite statistique.

Abstract

In the parametric estimation context, estimator performances can be characterized, *inter alia*, by the mean square error and the resolution limit. The first quantifies the accuracy of estimated values and the second defines the ability of the estimator to allow a correct resolvability. This thesis deals first with the prediction the "optimal" MSE by using lower bounds in the hybrid estimation context (i.e. when the parameter vector contains both random and non-random parameters), second with the extension of Cramér-Rao bounds for non-standard estimation problems and finally to the characterization of estimators resolution. This manuscript is then divided into three parts :

— First, we fill some lacks of hybrid lower bound on the MSE by using two existing Bayesian lower bounds : the Weiss-Weinstein bound and a particular form of Ziv-Zakai family lower bounds. We show that these extended lower bounds are tighter than the existing hybrid lower bounds in order to predict the optimal MSE.

— Second, we extend Cramer-Rao lower bounds for uncommon estimation contexts. Precisely : (i) Where the non-random parameters are subject to equality constraints (linear or nonlinear). (ii) For discrete-time filtering problems when the evolution of states are defined by a Markov chain. (iii) When the observation model differs to the real data distribution.

— Finally, we study the resolution of the estimators when their probability distributions are known. This approach is an extension of the work of Oh and Kashyap and the work of Clark to multi-dimensional parameters estimation problems.

Keywords : Parametric estimation, maximum likelihood estimator, maximum *a posteriori* estimator, performance analysis, hybrid estimation, lower bounds on the mean square error, statistical resolution limit.

Table des matières

Résumés **ix**

1 Introduction **1**
1.1 Généralités . . . 1
1.2 But de la thèse et résultats . . . 4

2 Bornes hybrides **9**
2.1 Introduction . . . 9
2.2 Inégalité de covariance . . . 10
2.3 Borne de Cramér-Rao hybride . . . 14
2.3.1 Forme générale . . . 14
2.3.2 Formule de Slepian-Bangs hybride dans le cas gaussien . . . 16
2.3.3 Efficacité du maximum de vraisemblance maximum *a posteriori* pour les modèles linéaires gaussiens . . . 16
2.4 Borne de Barankin hybride . . . 18
2.4.1 Introduction . . . 18
2.4.2 Expression générale . . . 19
2.4.3 Relaxation des contraintes . . . 21
2.4.4 Compression par transformation du noyau . . . 22
2.5 Borne de Barankin-Weiss-Weinstein hybride . . . 23
2.5.1 Combinaison de la borne de Barankin et de la borne de Weiss-Weinstein . 23
2.5.2 Modèle d'observation gaussien à moyenne paramétrée . . . 25
2.5.2.1 Analyse de $\xi(k, l, \mathbf{h}_a, \mathbf{h}_b)$ . . . 26
2.5.2.2 Indépendance de $A(\boldsymbol{\theta}, \mathbf{h}_a, \mathbf{h}_b)$ par rapport à $\boldsymbol{\theta}_r$ . . . 26
2.5.3 Application et simulation . . . 26
2.5.3.1 Analyse spectrale . . . 26
2.5.3.2 Simulation . . . 27
2.6 Borne de Barankin-Ziv-Zakaï hybride . . . 28
2.6.1 La famille de bornes de Ziv-Zakaï . . . 29
2.6.2 La borne proposée . . . 29
2.6.3 Expression analytique pour le modèle gaussien à moyenne paramétrée . . 30
2.6.3.1 Expression de $\beta(.,.)$ . . . 31
2.6.3.2 Expression de $\alpha(.,.)$ . . . 32
2.6.4 Simulation . . . 33

3 Bornes de Cramér-Rao non standards **35**
3.1 Borne de Cramér-Rao hybride sous contraintes . . . 35
3.1.1 Contexte de l'étude et notations . . . 35
3.1.2 Borne proposée . . . 36
3.1.3 Comparaison avec la borne de Cramér-Rao hybride sans contraintes . . . 37
3.1.4 Comparaison avec la borne de Cramér-Rao marginale . . . 38
3.1.5 Application à l'estimation de fréquence Doppler . . . 39
3.2 La borne de Cramér-Rao hybride récursive . . . 40
3.2.1 Problématique . . . 42
3.2.2 Borne récursive hybride . . . 43
3.3 La borne de Cramér-Rao déterministe avec erreurs de modèle : application à la problématique du Radar MIMO . . . 45
3.3.1 Rappel sur les bornes de Cramér-Rao désadaptées . . . 46
3.3.2 Application à l'estimation de la direction de départ et d'arrivée pour un système radar MIMO . . . 48
3.3.2.1 Expression générale de la borne de Cramér-Rao désadaptée pour les erreurs de positions des antennes . . . 49
3.3.2.2 Simulation . . . 50

4 Caractérisation des estimateurs en termes de résolution et de précision **53**
4.1 Introduction . . . 53
4.2 Rappels et définitions . . . 55
4.3 Une approche pour la résolution multidimensionnelle . . . 56
4.3.1 Approximation d'ensembles disjoints . . . 56
4.3.2 Précision d'estimation . . . 57
4.3.3 Probabilité de résolution . . . 59
4.3.4 Cas de distributions gaussiennes . . . 59
4.4 Application à la localisation de sources . . . 60
4.4.1 Performance asymptotique du maximum de vraisemblance conditionnel . . . 60
4.4.2 Simulations . . . 61

5 Conclusion et Perspectives **65**

A Etude de la matrice C **69**
A.1 La borne de Barankin-Weiss-Weinstein hybride . . . 69
A.2 La borne de Barankin Ziv-Zakaï hybride . . . 70

B Recursive hybrid Cramér-Rao bound for discrete-time Markovian dynamic systems **73**

C On the Accuracy and Resolvability of Vector Parameter Estimates **79**

Liste des figures **93**

Bibliographie **95**

Notations

Acronymes

- BCR : borne de Cramér-Rao.
- BCRH : borne de Cramér-Rao hybride
- BBH : borne de Barankin hybride
- BBWWH : borne de Barankin Weiss-Weinstein hybride
- BBZZH : borne de Barankin Ziv-Zakaï hybride
- BCRHC : borne de Cramér-Rao hybride sous contrainte
- CCLRB : borne de Barankin hybride par compression du noyau
- EQM : erreur quadratique moyenne.
- F_D : matrice d'information de Fisher déterministe.
- F_H : matrice d'information de Fisher hybride.
- MV : maximum de vraisemblance.
- MVMAP : maximum de vraisemblance / maximum *a postériori*
- RSB : rapport signal sur bruit.

Symboles mathématiques généraux

- $\mathbb{C}$ indique le corps des complexes.
- $\mathbb{R}$ indique le corps des réels.
- $\mathrm{Re}\,(z)$ indique la partie réelle d'un nombre complexe z.
- $\mathrm{Im}\,(z)$ indique la partie imaginaire d'un nombre complexe z.
- Le symbole * indique l'opérateur complexe conjugué.
- $\min\limits_{\boldsymbol{\eta}} f\,(\boldsymbol{\eta})$, $\max\limits_{\boldsymbol{\eta}} f\,(\boldsymbol{\eta})$ indique la valeur minimale ou maximale de l'expression $f\,(\boldsymbol{\eta})$.
- $\arg\min\limits_{\boldsymbol{\eta}} f\,(\boldsymbol{\eta})$, $\arg\max\limits_{\boldsymbol{\eta}} f\,(\boldsymbol{\eta})$, indique la valeur du vecteur $\boldsymbol{\eta}$ minimisant ou maximisant l'expression $f\,(\boldsymbol{\eta})$.
- $\triangleq$ indique une égalité par définition

Symboles et opérateurs matriciels

- a, A, les lettres en italiques représentent une quantité scalaire.
- $\mathbf{a}$, les lettres minuscules en gras représentent une quantité vectorielle (vecteur colonne).
- $\mathbf{A}$, les lettre majuscules en gras représentent une quantité matricielle.
- $\mathbf{A}^T$, le symbole ^{T} indique l'opérateur transposé.
- $\mathbf{A}^H$, le symbole ^{H} indique l'opérateur Hermitien (transposé conjugué).

- $\mathbf{I}_N$ indique la matrice identité de taille $N \times N$.
- $|\mathbf{A}|$ indique le déterminant de la matrice $\mathbf{A}$.
- $\|.\|$ indique la norme L^2.
- $\{A\}_{i,j}$ est l'élément correspondant à la $i^{\text{ième}}$ ligne et $j^{\text{ième}}$ colonne de la matrice $\mathbf{A}$.
- $\{A\}_{i,:}$ désigne le $i^{\text{ième}}$ vecteur ligne de la matrice $\mathbf{A}$.
- $\{A\}_{:,j}$ désigne le $j^{\text{ième}}$ vecteur colonne de la matrice $\mathbf{A}$.
- $Tr\left(\mathbf{A}\right) = \sum\limits_{i=1}^{N} \{A\}_{i,i}$ est la trace de la matrice carrée $\mathbf{A}$ de taille $N \times N$.
- $\otimes$ est le produit de Kronecker.
- Soient deux matrices $\mathbf{A}$ et $\mathbf{B} \in \mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$), alors la relation d'ordre

$$\mathbf{A} \succeq \mathbf{B},$$

  signifie que la matrice $\mathbf{A} - \mathbf{B}$ est définie non-négative *i.e.* $\forall \mathbf{y} \in \mathbb{K}^n$, $\mathbf{y}^T \left(\mathbf{A} - \mathbf{B}\right) \mathbf{y} \geq 0$.
- Soit $\mathbf{A} \in \mathbb{K}^n$, $\mathbf{A} \prec \infty$ signifie que $\forall \mathbf{y} \in \mathbb{R}^n$, $\mathbf{y}^T \mathbf{A} \mathbf{y} < \infty$.
- $Diag\left(\mathbf{a}\right)$ désigne une matrice diagonale dont les éléments diagonaux sont les éléments de la matrice $\mathbf{a}$.

Symboles relatifs aux probabilités

- $\mathbb{C}\mathcal{N}\left(\mathbf{m}, \mathbf{C}\right)$ indique une loi gaussienne complexe circulaire multivariée de moyenne $\mathbf{m}$ et de matrice de covariance $\mathbf{C}$.
- $\Pr\left(.\right)$ indique une probabilité.
- $f_X\left(.\right)$ indique la densité de probabilité de X.
- $f_{X,Y}\left(.\right)$ indique la densité de probabilité jointe de X et de Y.
- $f_{X|Y}\left(.\right)$ indique la densité de probabilité de X conditionnelle à Y.
- $\sim$ indique l'opérateur suivre en loi.
- $\mathbb{E}\left[.\right]$ indique l'opérateur d'espérance mathématique.
- $\mathbb{E}_{\mathbf{x},\theta}\left[.\right]$ indique l'opérateur d'espérance mathématique par rapport à la loi jointe $f_{\mathbf{x},\theta}\left(.\right)$.
- $\mathbb{E}_{\mathbf{x}|\theta}\left[.\right]$ indique l'opérateur d'espérance mathématique par rapport à la loi conditionnelle $f_{\mathbf{x}|\theta}\left(.\right)$.

Chapitre 1

Introduction

1.1 Généralités

En traitement statistique du signal, une modélisation probabiliste des données est effectuée afin de tenir compte des phénomènes aléatoires comme les bruits de mesures. Cette modélisation consiste à définir, en fonction des paramètres inconnus, dont ceux d'intérêt, la densité de probabilité des données. Elle peut être établie en se basant sur des considérations physiques du processus de mesure et/ou à travers l'histogramme des données. De manière générale, le problème d'estimation est défini par le schéma suivant : soient $\mathbf{x}$ un vecteur d'observation correspondant aux données mesurées et $\boldsymbol{\theta} \in \Theta$ les paramètres à estimer**.** Nous supposons que le modèle d'observation est paramétrique c'est-à-dire qu'il existe une fonction déterministe $\mathbf{m}(.)$ permettant de relier les observations $\mathbf{x}$**,** les paramètres $\boldsymbol{\theta}$ et un bruit aléatoire $\mathbf{b}$ dont on connaît les caractéristiques statistiques. Puisque $\mathbf{b}$ est une variable aléatoire, nous devons définir en toute rigueur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et faire la distinction entre la variable aléatoire $\mathbf{b}$ et une réalisation du bruit $\mathbf{b}(\omega)$ avec $\omega \in \Omega$. Ainsi les observations mesurées peuvent être exprimées en fonction des paramètres par la relation $\mathbf{x}(\boldsymbol{\theta}, \omega) = \mathbf{m}(\boldsymbol{\theta}, \mathbf{b}(\omega))$. Par conséquent, nous devons introduire également un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P}_{\boldsymbol{\theta}})$ pour les observations $\mathbf{x}$. De plus, ce problème doit être identifiable c'est-à-dire $\forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$, si $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, alors $\exists A \subset \Omega$ de mesure non nulle tel que $\forall \omega \in A$, $\mathbf{x}(\boldsymbol{\theta}_1, \omega) \neq \mathbf{x}(\boldsymbol{\theta}_2, \omega)$. Si cette condition n'est pas satisfaite alors il est impossible d'estimer de manière distincte ces valeurs $\boldsymbol{\theta}_1$ et $\boldsymbol{\theta}_2$ même dans les conditions idéales c'est-à-dire sans bruit. Dans toute la suite, nous supposerons que les conditions susmentionnées sont satisfaites et nous nous affranchirons de ces formalismes mathématiques quand il n'y a pas d'ambiguïté. Nous considérons simplement que le bruit $\mathbf{b}$ est un vecteur aléatoire (absolument) continue [1] et donc que le vecteur d'observations $\mathbf{x}$ est également un vecteur aléatoire continue qui admet une densité de probabilité conditionnelle $f_{\mathbf{x}|\boldsymbol{\theta}}$. Une fois que le modèle est bien défini, une estimation des paramètres est déduite à partir des données mesurées, du modèle d'observation et des informations *a priori* dont on dispose. L'algorithme d'estimation $\widehat{\boldsymbol{\theta}}$ est donc construit à partir de la densité de probabilité conditionnelle $f_{\mathbf{x}|\boldsymbol{\theta}}$ et de la densité a priori $f_{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}$ si $\boldsymbol{\theta}$ est également considéré comme un vecteur aléatoire (estimation bayésienne). Cet estimateur $\widehat{\boldsymbol{\theta}}$ dépend des observations $\mathbf{x}$ et on note parfois $\widehat{\boldsymbol{\theta}}(\mathbf{x})$ pour souligner cette dépendance. Cependant, même si on connaît parfaitement le modèle d'observation, des erreurs d'estimation se produisent à cause

1. Une variable aléatoire admet une densité de probabilité si elle est absolument continue. Par abus de language, on dit parfois qu'elle est seulement continue.

du caractère aléatoire du bruit. En effet, pour une même valeur de paramètre, plusieurs jeux de données différents peuvent être mesurés et conduisent à des valeurs d'estimation différentes. L'estimateur $\widehat{\boldsymbol{\theta}}(\mathbf{x})$ est par conséquent un vecteur aléatoire distribué suivant une loi de probabilité. Pour évaluer la précision des estimateurs, il est donc nécessaire d'introduire des critères statistiques qui prennent en compte cette distribution de données. Idéalement, la connaissance de la loi de probabilité de $\widehat{\boldsymbol{\theta}}(\mathbf{x})$ permettrait de caractériser ses performances d'estimation. En effet, par cette loi, nous pouvons quantifier les probabilités d'avoir une estimation proche de la vraie valeur et obtenir, sous conditions d'existence, tous les moments statistiques de $\widehat{\boldsymbol{\theta}}(\mathbf{x})$. Malheureusement, nous ne connaissons pas en générale cette loi à l'exception de certains cas simples où on peut obtenir une expression analytique explicite de l'estimateur $\widehat{\boldsymbol{\theta}}(\mathbf{x})$ en fonction des données $\mathbf{x}$.

Néanmoins, deux critères sont fréquemment utilisés pour évaluer les performances d'un estimateur : le biais, la variance et l'EQM de l'estimateur. Le biais quantifie l'écart moyen entre les estimées et les vraies valeurs des paramètres

$$\mathbf{biais}_{\widehat{\boldsymbol{\theta}}} = \mathbb{E}\left[\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right].$$

et la variance mesure la dispersion des estimées autour de leurs valeurs moyennes

$$\mathbf{VAR}_{\widehat{\boldsymbol{\theta}}} = \mathbb{E}\left[\left(\widehat{\boldsymbol{\theta}} - \mathbb{E}\left[\widehat{\boldsymbol{\theta}}\right]\right)\left(\widehat{\boldsymbol{\theta}} - \mathbb{E}\left[\widehat{\boldsymbol{\theta}}\right]\right)^T\right].$$

L'EQM mesure la dispersion des estimés autour de la vraie valeur du paramètre et permet de regrouper le biais et la variance en un seul critère

$$\mathbf{EQM}_{\widehat{\boldsymbol{\theta}}} = \mathbb{E}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^T\right] = \mathbf{VAR}_{\widehat{\boldsymbol{\theta}}} + \mathbf{biais}_{\widehat{\boldsymbol{\theta}}}\mathbf{biais}_{\widehat{\boldsymbol{\theta}}}^T.$$

Une justification concernant l'utilisation de ces critères est que certains estimateurs suivent asymptotiquement une loi gaussienne (par exemple la normalité asymptotique du maximum de vraisemblance (MV) [LC03]). Donc la connaissance du biais et de la variance suffisent à reconstruire la densité de probabilité de ces estimateurs. Cependant, ces critères de performance n'ont pas en générale d'expression analytique hormis les cas d'estimation linéaire en les paramètres avec un bruit additif gaussien [Kay93]. Pour des problèmes d'estimation non-linéaire, une valeur approchée de ces termes peut être obtenue par des simulations de type Monte-Carlo. Cette méthode consiste d'abord à générer indépendamment des séries de données $\{\mathbf{x}_i\}_{i=1\cdots MC}$ en se basant sur le modèle d'observation, puis de faire une estimation $\left\{\widehat{\boldsymbol{\theta}}(\mathbf{x}_i)\right\}_{i=1\cdots MC}$ pour chaque série de données et enfin une évaluation de l'EQM est donnée par

$$\mathbf{EQM}_{\widehat{\boldsymbol{\theta}}} \simeq \frac{1}{MC}\sum_{i=1}^{MC}\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}_i) - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}_i) - \boldsymbol{\theta}\right)^T, \tag{1.1}$$

où MC est le nombre de simulations de Monte-Carlo réalisées. D'après la loi des grands nombres, le membre de droite de (1.1), communément appelé l'EQM empirique, tend vers l'EQM de l'estimateur quand MC tend vers l'infini. De plus pour un nombre d'échantillons MC donné, l'erreur de quantification est proportionnelle à $\frac{1}{\sqrt{MC}}$. La vitesse de convergence est donc relativement lente, il faut un nombre de tirages conséquent pour avoir un bonne estimation de l'EQM. Par ailleurs, cette méthode fait appel à l'estimateur $\widehat{\boldsymbol{\theta}}$ pour chaque série de données $\{\mathbf{x}_i\}_{i=1\cdots MC}$. Son

coût de calcul est généralement lourd selon la complexité de l'algorithme $\widehat{\boldsymbol{\theta}}$. Pour remédier à ce problème, des bornes inférieures [2] de l'EQM ont été proposées dans la littérature afin de prédire l'EQM ou la variance des estimateurs sans biais. Ces bornes présentent plusieurs avantages : premièrement, elles sont intrinsèques au problème d'estimation et donc l'évaluation de ces bornes ne dépend pas de l'estimateur en question. Deuxièmement, la plupart de ces bornes ont une expression analytique ou semi-analytique (c'est-à-dire sous forme d'intégrale ou de maximisation suivant un ou plusieurs paramètres) pour les observations à distribution gaussienne [Kay93], elliptiques [PR10] [GFG14a] [BA13] et autres [TRB+14]. Donc l'évaluation de ces expressions est rapide numériquement. Troisièmement, certaines de ces bornes minimales sont atteignables à fort rapport signal sur bruit (RSB) [RFCL06] ou à grand nombre d'observations [LC03], donc elles décrivent les performances optimales d'estimation en régime asymptotique, c'est-à-dire que ces bornes donnent l'EQM minimale que l'on peut espérer d'atteindre parmi un ensemble d'estimateurs réalisables (non clairvoyants).

Par ailleurs, lorsque le problème d'estimation est non-linéaire par rapport aux paramètres d'intérêt, on peut distinguer trois zones de fonctionnement pour l'EQM des estimateurs [RB74] (voir la figure 1.1) : une zone dite asymptotique où une faible variation du RSB ou du nombre d'observation induit une faible variation de l'EQM ; ce comportement s'explique par le fait que les valeurs estimées sont généralement proches de la vraie valeur du paramètre. Puis une zone de décrochement apparaît lorsque l'EQM croît considérablement pour une faible variation du RSB ou du nombre d'observations, ce phénomène intervient lorsque le maximum global du critère n'est plus autour de la vraie valeur. Enfin, l'EQM devient quasi constante dans la zone de non information car l'information est complètement noyée dans le bruit. Le seuil de décrochement délimite donc la zone optimale de fonctionnement de l'estimateur. La connaissance de ce seuil est donc fondamentale, c'est pourquoi un ensemble de bornes communément appelées "large error bounds" sont proposées dans la littérature pour la prédiction de ce seuil. Enfin, les bornes inférieures de l'EQM peuvent servir d'étalon (benchmark) pour verifier en amont, c'est-à-dire avant de mettre en place l'estimateur, si les performances exigées par le cahier des charges sont réalistes ou non.

Dans la littérature, les bornes sont classées en fonction de la nature des paramètres à estimer. Nous distinguons essentiellement trois familles de bornes :

1. Les bornes dites "déterministes" sont appliquées aux problèmes d'estimation de paramètres non-aléatoires. [Cra46,Rao45,Bar49,CR51,MS69,Gla72,Abe93,FL01,FL02,Cha04, Rendf]
2. Les bornes dites "bayésiennes" sont utilisées lorsque les paramètres d'estimation sont aléatoires de densité de probabilité *a priori* connue. [Van68,BMWZ87,ZZ69,CZZ75,BZ76, WW85,WW88,BEV95,QCL06]
3. Les bornes dites "hybrides" sont des minorants de l'EQM dans le contexte d'estimation de paramètres dont une partie est aléatoire de densité de probabilité *a priori* connue et dont le reste est non-aléatoire. [RS87,RM97,Mes06,TT09]

La majorité des bornes existantes s'applique uniquement dans le contexte de l'estimation classique, c'est-à-dire estimer les paramètres $\boldsymbol{\theta}$ à l'aide des données $\mathbf{x}$ et du modèle d'observation

2. Ces bornes sont aussi appelés bornes minimales ou bornes minorantes. La dénomination de ces bornes divergent par le fait qu'elles ne correspondent pas toujours à la définition mathématique de "borne inférieure" à savoir le plus grand des minorants.

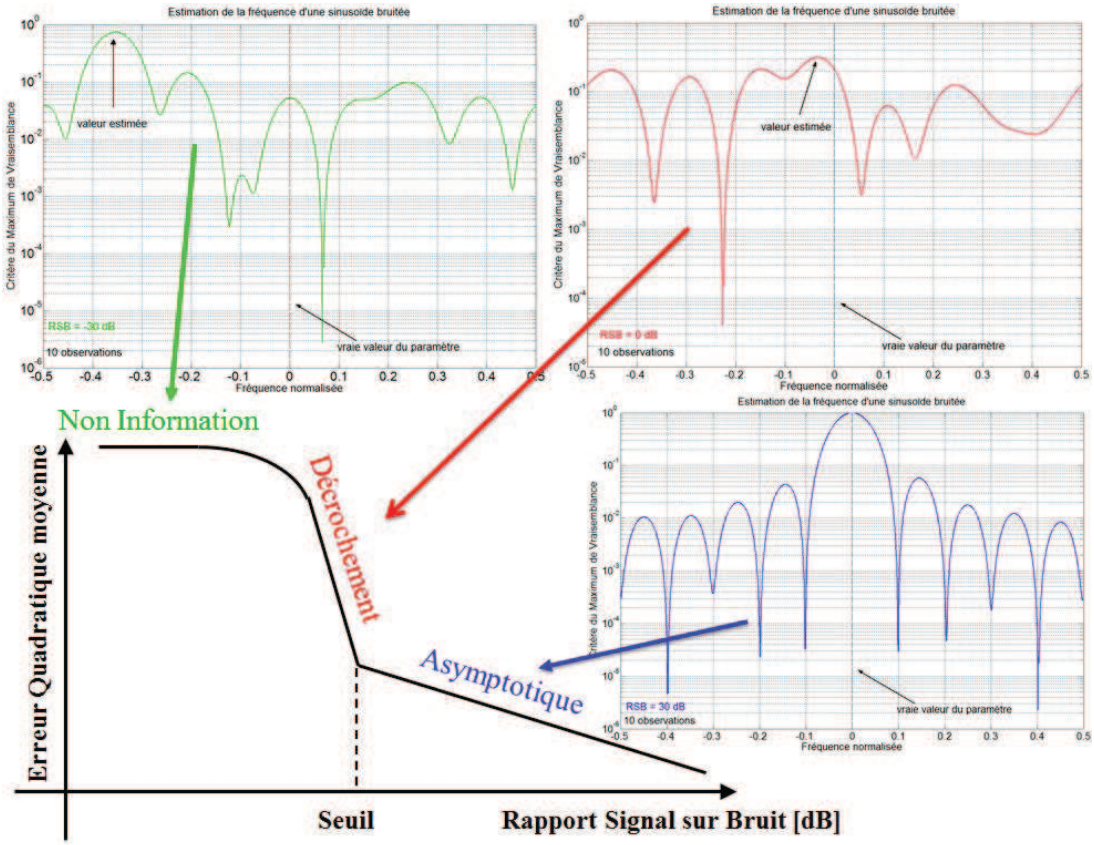

FIGURE 1.1 – EQM du maximum de vraisemblance pour l'estimation de la fréquence d'une sinusoïde

$\mathbf{x}(\boldsymbol{\theta}, \omega) = \mathbf{m}(\boldsymbol{\theta}, \mathbf{b}(\omega))$. Or, ce contexte peut être irréaliste pour des problèmes concrets. Notamment, certains scénarios d'estimation peuvent inclure implicitement des contraintes sur les paramètres, des paramètres qui évoluent dans le temps ou encore une modélisation incorrecte des données. On peut se demander si les bornes "classiques" s'appliquent dans de telles conditions. Sinon peut-on obtenir des bornes qui permettent de prédire l'EQM optimale dans ces situations ? En plus des erreurs d'estimation, les estimateurs sont limités par leur résolution. En effet, bien que le modèle d'observation soit parfaitement connu, les estimateurs ne sont pas en mesure de donner des estimations distinctes lorsque l'écart entre les paramètres sont en deçà d'une distance limite. Nous nous sommes intéressés à l'étude d'un critère unique qui permet de caractériser à la fois la résolution et la précision d'un estimateur.

1.2 But de la thèse et résultats

Le travail proposé dans cette thèse a été initialement motivé par les résultats obtenus par Eric Chaumette [Cha04] et Alexandre Renaux [Rendf] qui traitent respectivement des bornes deterministes et bayésiennes. L'idée de départ est d'étendre ces bornes dans le contexte de l'estimation hybride, c'est-à-dire l'estimation simultanée de paramètres aléatoires et non-aléatoires.

Notamment, nous nous sommes aperçus que l'inégalité de covariance [LC03,Mes06] est la pierre angulaire pour unifier à la fois des bornes déterministes, bayésiennes et hybrides. Bien qu'en estimation bayésienne, la borne de Weiss-Weinstein [WW88] et les bornes de Ziv-Zakaï [ZZ69] soient parmis les bornes les plus précises, il n'existait pas de version "hybride" de ces bornes. Puis, nous avons étudié des bornes pour des contextes d'estimation moins classiques. En effet, peu de bornes dans la littérature traitent les cas où les paramètres d'estimation sont soumis à des contraintes d'égalité [GO91,SN98] ou s'attaquent aux problèmes de filtrages à temps discret [BZ75,TMN98]. On notera que l'estimation simultanée de paramètres aléatoires et non-aléatoires dans de tels contextes n'a jamais été étudiée. Une analyse de performances a également été menée pour des problèmes d'estimation en présence d'erreur de modélisation. Cette dernière reprend les travaux de [Hub67,Whi82,RH13,GFG14a] que nous appliquons au problème de la localisation de sources. Et enfin, lorsqu'on veut estimer plusieurs paramètres de même nature, l'EQM devient insuffisante pour caractériser les performances d'un estimateur. En effet, pour des problèmes de localisation de sources [EBRM12,EBRM11], il semble aussi important de savoir si nous arriverons à résoudre ces sources que de connaître les erreurs d'estimation. Pour cette raison, nous cherchons s'il existe des critères qui permettrait à la fois de caractériser la résolution et la précision d'un estimateur.

Les résultats menés tentent de répondre aux problèmes sus-mentionnés. Cette thèse est donc organisée en trois parties :

— Premièrement, nous avons proposé deux nouvelles bornes inférieures de l'EQM pour l'estimation hybride. La première est basée sur une combinaison entre la borne de Barankin (version McAulay-Seidman [MS69]) pour les paramètres déterministes et la borne de Weiss-Weinstein [WW88] pour les paramètres aléatoires tandis que la deuxième est basée sur une combinaison entre la borne de Barankin (version McAulay-Seidman [MS69]) pour les paramètres déterministes et la borne de Ziv-Zakaï (version proposée par Kristine Bell dans sa thèse [Bel95]) pour les paramètres aléatoires. Pour ces deux bornes, nous avons établi la classe d'estimateurs pour laquelle chaque borne est valide et nous avons proposé des expressions analytiques pour les modèles d'observations gaussiens à moyenne ou à variance paramétrés. Les résultats de simulation montrent très clairement un gain par rapport à la borne de Barankin hybride [RM97], connue à ce jour comme la borne hybride la plus précise. Ces résultats sont décrits dans le chapitre 2.

— La seconde partie de ce travail est dédiée à l'étude de bornes inférieures de l'EQM dans un contexte non-standard au sens où les paramètres évoluent dans le temps ou sont sous contraintes entre eux et où le modèle d'observation n'est plus correctement spécifié. Nous avons étendu les résultats de Gorman et Hero [GO90] concernant la borne de Cramér-Rao sous contraintes d'égalité (c'est-à-dire lorsque les paramètres sont reliés par des équations linéaire ou non-linéaire) au cas hybride. Lorsque les paramètres ne sont plus statiques (c'est-à-dire lorsqu'ils sont régis par une dépendance temporelle), on parle alors de problème de filtrage et d'estimation d'état. Dans ce contexte, nous avons proposée une borne hybride récursive de type Cramér-Rao qui prend en compte les cas où l'évolution des états du système depend des paramètres déterministes. Elle étend ainsi les résultats de [TMN98]. Enfin, lorsque les observations mesurées ne suivent pas le modèle supposé, on dit que le modèle est malspécifié (ou désadapté). Récemment des bornes sont proposées pour prédire les estimateurs opérant dans de tels conditions [RH15]. Nous avons appliqué la borne de Cramér-Rao "malspécifiée" (appellée également limite de Huber) au cas de l'estimation de directions d'arrivée et de départ dans un contexte de Radar MIMO. Ces résultats sont présentés dans le chapitre 3.

— La troisième et dernière partie concerne la résolution limite statistique. Nous avons proposé une extension de l'approche de Oh & Kashyap [OK91] et de Clark [Cla95] pour des problèmes d'estimation de paramètres multidimensionnels. Basée sur la densité de probabilité des valeurs estimées, le critère proposé permet de caractériser à la fois la résolution et la précision d'un estimateur. Une illustration de ce critère a été appliquée à travers un problème de localisation de source dans lequel nous avons fait une comparaison avec le critère de Lee-Yau-Bressler [Lee93,YB92], de Smith [Smi05], de Shahram-Milanfar [SM93] et de Amar-Weiss [AW08]. Parmi tous ces critères sus-mentionnés, le nôtre semble être le plus exigeant par le fait qu'il contraint les valeurs estimées à appartenir à un domaine restreint. Ces résultats sont décrits dans le chapitre 4.

Les travaux exposés dans ce document ont donné lieu aux articles et communications suivants (les deux revues internationales acceptées sont jointes en annexes B et C)

Revues internationales

— **C. Ren**, M. N. El Korso, J. Galy, E. Chaumette, P. Larzabal and A. Renaux, "On the accuracy and resolvability of vector parameter estimates", *IEEE* Transactions on Signal Processing, Volume : 62, Issue : 14, Jul. 2014, pp. 3682-3694.
— **C. Ren**, J. Galy, E. Chaumette, F. Vincent, P. Larzabal and A. Renaux, "Recursive hybrid Cramér-Rao bound for discrete-time Markovian dynamic systems", *IEEE* Signal Processing Letters, Volume : 22, Issue : 10, Oct. 2015, pp. 1543-1547.
— **C. Ren**, J. Galy, E. Chaumette, F. Vincent, P. Larzabal and A. Renaux, "Hybrid Barankin-Weiss-Weinstein Bounds", en révision mineure pour *IEEE* Signal Processing Letters.

Congrés avec comité de lecture et actes

— **C. Ren**, J. Galy, E. Chaumette, P. Larzabal and A. Renaux, "Hybrid lower bound on the MSE based on the Barankin and Weiss-Weinstein bounds", in Proc. of *IEEE* International Conference on Acoustics, Speech, and Signal Processing, ICASSP-2013, Vancouver, Canada.
— **C. Ren**, J. Galy, E. Chaumette, P. Larzabal and A. Renaux, "Une borne inférieure de l'erreur quadratique moyenne pour l'estimation simultanée de paramètres aléatoires et non-aléatoires", in Proc. of Colloque GRETSI 2013, Brest, France.
— **C. Ren**, J. Galy, E. Chaumette, P. Larzabal and A. Renaux, "High resolution techniques for Radar : Myth or Reality ?", in Proc. of the European Signal Processing Conference, EUSIPCO-2013, Marrakech, Morocco.
— **C. Ren**, J. Galy, E. Chaumette, P. Larzabal and A. Renaux, "A Ziv-Zakai type bound for hybrid parameter estimation", in Proc. of *IEEE* International Conference on Acoustics, Speech, and Signal Processing, ICASSP-2014, Firenze, Italy.
— **C. Ren**, J. Le Kernec J. Galy, E. Chaumette, P. Larzabal and A. Renaux, "A constrained hybrid Cramér-Rao bound for parameter estimation", in Proc. of *IEEE* International Conference on Acoustics, Speech, and Signal Processing, ICASSP-2015, Brisbane, Australia.
— **C. Ren**, M. N. El Korso, J. Galy, E. Chaumette, P. Larzabal and A. Renaux, "Performances bounds under misspecification model for MIMO Radar application", in Proc. of

the European Signal Processing Conference, EUSIPCO-2015, Nice, France.

— **C. Ren**, J. Le Kernec J. Galy, E. Chaumette, P. Larzabal and A. Renaux, "Borne de Cramér-Rao sous contraintes pour l'estimation simultanée de paramétres aléatoires et non aléatoires", in Proc. of Colloque GRETSI 2015, Lyon, France.

— **C. Ren**, M. N. El Korso, J. Galy, E. Chaumette, P. Larzabal and A. Renaux, "Une nouvelle approche de résolution limite dans le cadre d'estimation paramétrique multidimensionnelle", in Proc. of Colloque GRETSI 2015, Lyon, France.

Chapitre 2

Bornes hybrides

2.1 Introduction

L'estimation hybride consiste à estimer un vecteur composé à la fois de paramètres aléatoires et de paramètres non-aléatoires à partir des observations. Contrairement aux estimations déterministes et bayésiennes où le modèle d'observation est respectivement paramétrisé par des inconnues non-aléatoires et par des inconnues aléatoires, le contexte hybride s'intérresse aux modèles dépendant simultanément de ces deux types d'inconnues. Ce contexte modèles dépendant simultanément de ces deux types d'inconnues. Ce contexte d'estimation intervient dans de nombreuses applications. On peut citer, par exemple, le modèle linéaire gaussien généralisé [Van02], la calibration d'un réseau d'antennes [RS87], l'estimation de retard pour les signaux radar [RM97], l'estimation de la phase dans une transmission binaire à modulation de phase sans signal de référence [BGR+08], l'estimation de phase dans les modulations QAM [YGB11], l'analyse spectrale [TT09], l'estimation conjointe de la fréquence Doppler et du retard dans un récepteur numérique [VVRB12], l'estimation de paramétres dans les systèmes DS/CDMA [BLS05], l'estimation de la direction d'arrivée pour des antennes déformées dans un contexte de mécanique des fluides [TW07]. Malheureusement, le cadre d'estimation hybride n'est généralement pas une simple concaténation d'un problème d'estimation déterministe et d'un problème d'estimation bayésienne. En effet, l'estimation hybride n'est pas séparable, dans le sens où nous ne pouvons pas utiliser les estimateurs déterministes pour les paramètres non-aléatoires et les estimateurs bayésiens pour les paramètres aléatoires puisque les paramètres déterministes et bayésiens peuvent être statistiquement dépendants. Par conséquent, les limites de performance des estimateurs hybrides ne peuvent pas, en général, être comparées avec les bornes déterministes et les bornes bayésiennes. C'est pourquoi, les bornes hybrides ont été proposées pour prédire les performances en terme d'EQM de tels estimateurs. Historiquement, la première borne hybride est la borne de Cramér-Rao hybride (BCRH) introduite par Rockah et Schultheiss dans le contexte d'estimation de directions d'arrivée et de l'auto-calibration d'un réseau d'antennes [RS87] où la densité de probabilité *a priori* des paramètres aléatoires est indépendante des paramètres déterministes. Cependant la BCRH est une borne locale ("small error bound"), elle ne permet pas de prédire les phénomènes de décrochement. Afin de remédier à ce problème, Reuven et Messer ont introduit la borne de Barankin hybride (BBH) [RM97] qui est une extension de la borne de Barankin [Bar49] connue pour son aptitude à prédire les décrochements. Récemment, une forme plus générale de la BCRH a été proposée par Bay *et. al.* [BGR+08] lorsque la loi *a priori* des paramètres aléatoires dépendent des paramètres déterministes. Une étude approfondie de la BCRH

a été faite par Noam and Messer [NM09] en proposant une condition suffisante pour que la sous-matrice concernant les paramètres déterministes de la BCRH soit égale à la BCR déterministe en considérant les paramètres aléatoires comme des paramètres de nuisance. Ces travaux ont notamment montré que la BCRH prédit bien l'EQM asymptotique à fort RSB ou à fort nombre d'observations, comme les BCRs déterministes et bayésiennes. Parallèlement, la BBH a été étendue par Todros et Tabrikian [TT09] en appliquant une transformation de noyau (basée sur des fonctions noyaux appelées "rapport de vraisemblance centré"). Cette approche a permi d'obtenir une BBH plus précise notamment pour la prédiction du seuil de décrochement. Plus récemment, nous avons montré qu'une combinaison de la BBH avec la borne Weiss-Weinstein [WW88] d'une part et avec la borne de Ziv-Zakaï [BEV95] d'autre part améliorent la prédiction du seuil de décrochement.

Dans ce chapitre, nous rappellerons tout d'abord l'inégalité de covariance qui permet de minorer l'EQM des estimateurs et donc d'obtenir des bornes hybrides. Puis nous présenterons les deux bornes hybrides les plus connues et utilisées dans la littérature, à savoir la borne de Cramér-Rao hybride et la borne de Barankin hybride. Le reste de ce chapitre sera consacré à la présentation de deux résultats originaux que nous avons obtenus. Il s'agit dans les deux cas d'une nouvelle borne hybride plus précise ("tight") que la BBH. La première est basée sur un mélange entre la borne de Barankin classique (déterministe) et la borne de Weiss-Weinstein (bayésienne). La deuxième est une combinaison entre la borne de Barankin classique (déterministe) et la borne de Ziv-Zakaï (bayésienne).

2.2 Inégalité de covariance

Il existe dans la littérature de nombreuses inégalités mathématiques qui permettent d'établir des minorants de l'EQM. Les plus utilisées sont :

— L'inégalité de Cauchy-Schwarz pour la génération des bornes déterministes et bayésiennes de la famille Weiss-Weinstein (généralisable par l'inégalité de Hölder aux moments absolus d'ordre > 1) ,

— L'inégalité de Kotelnikov pour les bornes bayésiennes de la famille Ziv-Zakaï.

L'inégalité de Cauchy-Schwarz admet une forme duale : l'inégalité de covariance et la minimisation d'une matrice de Gram sous contraintes linéaires. Cette dualité à été illustrée dans le cas des bornes déterministes [Cha04] et les bornes bayésiennes de la famille Weiss-Weistein [WW88]. En estimation hybride, l'approche classique tend à privilégier l'utilisation de l'inégalité de covariance. Nous conservons cette approche afin de faciliter les comparaisons avec les travaux antérieurs.

Théorème *Inégalité de covariance*

Soient $\mathbf{x}$ *un vecteur d'observation de* $\Omega \subset \mathbb{C}^P$ *et* $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^M$ *un vecteur de paramètres à estimer. On suppose que* $\mathbf{x}$ *est un vecteur aléatoire de densité de probabilité conditionnelle* $f_{\mathbf{x}|\boldsymbol{\theta}}(.)$ *et* $\boldsymbol{\theta}$ *est un vecteur de paramètres soit déterministe, soit aléatoire de densité de probabilité a priori connue* $f_{\boldsymbol{\theta}}(.)$*, soit hybride c'est-à-dire* $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_d^T \ \boldsymbol{\theta}_r^T\right]^T$ *où* $\boldsymbol{\theta}_d$ *est un vecteur de paramètres déterministes et* $\boldsymbol{\theta}_r$ *est un vecteur de paramètres aléatoires de densité de probabilité a priori connue* $f_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(.)$*. Si* $\widehat{\boldsymbol{\theta}}(\mathbf{x})$*, appelée estimateur de* $\boldsymbol{\theta}$*, une fonction mesurable de* Ω *dans* Θ *et* $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ *une fonction réelle mesurable de* $\Omega \times \Theta$ *dans* $\mathbb{R}^K$*,* $K \in \mathbb{N}^*$*, telle que la matrice de*

covariance $\mathbb{E}\left[\mathbf{v}\mathbf{v}^H\right]$ *est définie positive,* $\mathbb{E}\left[\widehat{\boldsymbol{\theta}}(\mathbf{x})\,\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta})\right] \prec \infty$ *et* $\mathbb{E}\left[\boldsymbol{\theta}\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta})\right] \prec \infty$ *alors*

$$\mathbb{E}\left[\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)^T\right] \succeq \mathbf{C}\mathbf{V}^{-1}\mathbf{C}^T \tag{2.1}$$

avec $\mathbf{C} = \mathbb{E}\left[\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta})\right]$ *et* $\mathbf{V} = \mathbb{E}\left[\mathbf{v}(\mathbf{x},\boldsymbol{\theta})\,\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta})\right]$.

Remarque

1) Le membre de droite de (2.1) n'est pas en général une borne minorante de l'EQM car la matrice $\mathbf{C}$ *dépend de* $\widehat{\boldsymbol{\theta}}$*. Cependant un choix judicieux du vecteur* $\mathbf{v}(\mathbf{x},\boldsymbol{\theta})$ *appliqué pour une classe d'estimateurs* $\widehat{\boldsymbol{\theta}}$ *bien précise nous permettra d'obtenir une expression de la matrice* $\mathbf{C}$ *indépendante de* $\widehat{\boldsymbol{\theta}}$ *et donc une borne.*

2) On appelle parfois $\mathbf{C}\mathbf{V}^{-1}\mathbf{C}^T$ *la matrice d'information, elle est définie positive si* $K \geq M$*.*

3) Quelle que soit la nature des paramètres $\boldsymbol{\theta}$ *(déterministes, aléatoires ou hybrides), l'inégalité (2.1) reste valable. Autrement dit, l'opérateur espérance* $\mathbb{E}[.]$ *peut moyenner suivant la distribution* $f_{\mathbf{x};\boldsymbol{\theta}}$*,* $f_{\mathbf{x},\boldsymbol{\theta}}$ *ou* $f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}$*.*

Il existe plusieurs démonstrations de cette inégalité. Nous allons reformuler ci-dessous la démonstration proposée par [LC03] à base de l'inégalité de Cauchy-Schwarz. Une autre démonstration est proposée dans [Mes06] en utilisant la propriété du complément de Schur d'une matrice définie positive.

Preuve Soit l'ensemble $\mathcal{L}^2(\mathbb{P}) \triangleq \left\{ f \text{ réelle mesurable} \mid \left(\int |f|^2 d\mathbb{P}\right)^{\frac{1}{2}} < \infty \right\}$ [1], muni du produit scalaire suivant : pour toutes fonctions réelles mesurables $g(\mathbf{x},\boldsymbol{\theta})$ et $h(\mathbf{x},\boldsymbol{\theta})$

$$\langle g,h\rangle = \int g(\mathbf{x},\boldsymbol{\theta})\,h(\mathbf{x},\boldsymbol{\theta})\,d\mathbb{P} = \mathbb{E}\left[g(\mathbf{x},\boldsymbol{\theta})\,h(\mathbf{x},\boldsymbol{\theta})\right], \tag{2.2}$$

et soit la norme induite

$$\|g\| = \sqrt{\langle g,g\rangle}.$$

On remarque que l'opérateur espérance entre deux fonctions peut être vu comme le produit scalaire de ces deux fonctions. Soit $\psi(\mathbf{x},\boldsymbol{\theta})$, $e(\mathbf{x},\boldsymbol{\theta})$ des fonctions réelles mesurables telles que $\psi(\mathbf{x},\boldsymbol{\theta}) \in \mathcal{L}^2(\mathbb{P})$ et $e(\mathbf{x},\boldsymbol{\theta}) \in \mathcal{L}^2(\mathbb{P})$. Par l'inégalité de Cauchy-Schwarz, nous avons

$$|\langle e(\mathbf{x},\boldsymbol{\theta}),\psi(\mathbf{x},\boldsymbol{\theta})\rangle| \leq \|e(\mathbf{x},\boldsymbol{\theta})\|\,\|\psi(\mathbf{x},\boldsymbol{\theta})\|,$$

(rappel $|\langle e(\mathbf{x},\boldsymbol{\theta}),\psi(\mathbf{x},\boldsymbol{\theta})\rangle| \triangleq \int |e(\mathbf{x},\boldsymbol{\theta})\,\psi(\mathbf{x},\boldsymbol{\theta})|\,d\mathbb{P}$.) ce qui est équivalent à

$$\|e(\mathbf{x},\boldsymbol{\theta})\|^2 \geq \frac{|\langle e(\mathbf{x},\boldsymbol{\theta}),\psi(\mathbf{x},\boldsymbol{\theta})\rangle|^2}{\|\psi(\mathbf{x},\boldsymbol{\theta})\|^2}. \tag{2.3}$$

Soient $\mathbf{a} \in \mathbb{R}^K$ et $\mathbf{b} \in \mathbb{R}^M$ quelconques tels que

$$\psi(\mathbf{x},\boldsymbol{\theta}) = \mathbf{a}^T\mathbf{v}(\mathbf{x},\boldsymbol{\theta}) \text{ et } e(\mathbf{x},\boldsymbol{\theta}) = \mathbf{b}^T\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right).$$

1. Admettons que cet espace $\mathcal{L}^2(\mathbb{P})$ est bien défini, nous préciserons plus tard la mesure $\mathbb{P}$ et l'espace mesurable associé.

L'inégalité (2.3) s'écrit donc

$$\left\|\mathbf{b}^T\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)\right\|^2 \geq \frac{\left|\left\langle \mathbf{b}^T\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right), \mathbf{a}^T\mathbf{v}(\mathbf{x},\boldsymbol{\theta})\right\rangle\right|^2}{\left\|\mathbf{a}^T\mathbf{v}(\mathbf{x},\boldsymbol{\theta})\right\|^2}.$$

Or le produit scalaire est une forme bilinéaire et

$$\int\left|\mathbf{b}^T\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta})\,\mathbf{a}\right|d\mathbb{P} \geq \left|\int \mathbf{b}^T\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta})\,\mathbf{a}d\mathbb{P}\right|, \tag{2.4}$$

on obtient donc

$$\mathbf{b}^T\left\langle\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right),\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)^T\right\rangle\mathbf{b} \geq \frac{\left|\mathbf{b}^T\left\langle\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta},\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta})\right\rangle\mathbf{a}\right|^2}{\mathbf{a}^T\left\langle\mathbf{v}(\mathbf{x},\boldsymbol{\theta}),\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta})\right\rangle\mathbf{a}}. \tag{2.5}$$

Notons $\mathbf{C}=\left\langle\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta},\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta})\right\rangle$ et $\mathbf{V}=\left\langle\mathbf{v}(\mathbf{x},\boldsymbol{\theta}),\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta})\right\rangle$, nous avons

$$\mathbf{b}^T\left\langle\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right),\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)^T\right\rangle\mathbf{b} \geq \frac{\left|\mathbf{b}^T\mathbf{C}\mathbf{a}\right|^2}{\mathbf{a}^T\mathbf{V}\mathbf{a}} \tag{2.6}$$

Puisque l'inégalité (2.6) est vraie quelque soit $\mathbf{a}\in\mathbb{R}^K$,

$$\mathbf{b}^T\left\langle\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right),\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)^T\right\rangle\mathbf{b} \geq \sup_{\mathbf{a}\in\mathbb{R}^K}\frac{\left|\mathbf{b}^T\mathbf{C}\mathbf{a}\right|^2}{\mathbf{a}^T\mathbf{V}\mathbf{a}} \tag{2.7}$$

Comme $\mathbf{V}$ est une matrice définie positive, on peut définir le produit scalaire suivant $\forall\mathbf{a}$ et $\mathbf{b}$

$$\langle\mathbf{a},\mathbf{b}\rangle_{\mathbf{V}}=\mathbf{a}^T\mathbf{V}\mathbf{b},$$

donc

$$\mathbf{b}^T\mathbf{C}\mathbf{a}=\left\langle\mathbf{V}^{-1}\mathbf{C}^T\mathbf{b},\mathbf{a}\right\rangle_{\mathbf{V}}.$$

D'après l'inégalité de Cauchy-Schwarz

$$\left|\left\langle\mathbf{V}^{-1}\mathbf{C}^T\mathbf{b},\mathbf{a}\right\rangle_{\mathbf{V}}\right|^2 \leq \left\|\mathbf{V}^{-1}\mathbf{C}^T\mathbf{b}\right\|_{\mathbf{V}}^2\left\|\mathbf{a}\right\|_{\mathbf{V}}^2, \tag{2.8}$$

où

$$\left\|\mathbf{a}\right\|_{\mathbf{V}}^2=\langle\mathbf{a},\mathbf{a}\rangle_{\mathbf{V}}=\mathbf{a}^T\mathbf{V}\mathbf{a}.$$

et l'égalité (2.8) est atteinte si

$$\mathbf{a}=k\mathbf{V}^{-1}\mathbf{C}^T\mathbf{b},\ \forall k\in\mathbb{R}.$$

On a finalement

$$\mathbf{b}^T\left\langle\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right),\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)^T\right\rangle\mathbf{b} \geq \left\|\mathbf{V}^{-1}\mathbf{C}^T\mathbf{b}\right\|_{\mathbf{V}}^2,$$

$\Leftrightarrow$

$$\mathbf{b}^T\left\langle\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right),\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)^T\right\rangle\mathbf{b} \geq \mathbf{b}^T\mathbf{C}\mathbf{V}^{-1}\mathbf{C}^T\mathbf{b}. \tag{2.9}$$

Puisque (2.9) est vrai pour tout $\mathbf{b} \in \mathbb{R}^M$ et comme $\langle g, h \rangle = \mathbb{E}\left[g\left(\mathbf{x}, \boldsymbol{\theta}\right) h\left(\mathbf{x}, \boldsymbol{\theta}\right)\right]$ donc

$$\mathbb{E}\left[\left(\widehat{\boldsymbol{\theta}}\left(\mathbf{x}\right) - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}\left(\mathbf{x}\right) - \boldsymbol{\theta}\right)^T\right] \succeq \mathbf{C}\mathbf{V}^{-1}\mathbf{C}^T. \tag{2.10}$$

Nous retrouvons l'inégalité de covariance. Revenons maintenant sur la mesure $\mathbb{P}$ et la tribu associée : $\mathbb{P}$ est une mesure de probabilité déterminant la fréquence d'apparition des évènements issue d'un processus aléatoire. L'expression $\mathbb{P}$ dépend du problème d'estimation considéré. En estimation paramétrique, les observations sont aléatoires, donc nous devons définir pour les observations l'espace probabilisé $\left(\Omega, \mathcal{F}, \mathbb{P}_{\mathbf{x}|\boldsymbol{\theta}}\right)$ où Ω est l'espace des observations $\mathbf{x}$, $\mathcal{F}$ est la tribu associée et $\mathbb{P}_{\mathbf{x}|\boldsymbol{\theta}}$ est la mesure de probabilité de $\mathbf{x}$ sur $\mathcal{F}$. Cette dernière est définie par l'application suivante

$$\mathbb{P}_{\mathbf{x}|\boldsymbol{\theta}} : \begin{cases} \mathcal{F} \to [0;1] \\ A \to \mathbb{P}_{\mathbf{x};\boldsymbol{\theta}}\left(\mathbf{x} \in A\right) \end{cases}.$$

Si la variable aléatoire conditionnelle $\mathbf{x}|\boldsymbol{\theta}$ admet une densité de probabilité continue $f_{\mathbf{x}|\boldsymbol{\theta}}$ alors

$$\mathbb{P}_{\mathbf{x}|\boldsymbol{\theta}}\left(\mathbf{x} \in A\right) = \int_{\mathbf{x} \in A} d\mathbb{P}_{\mathbf{x};\boldsymbol{\theta}} = \int_{\mathbf{x} \in A} f_{\mathbf{x}|\boldsymbol{\theta}}\left(\mathbf{x}\right) d\mathbf{x}.$$

Elle dépend évidemment de $\boldsymbol{\theta}$ puisque nous voulons estimer $\boldsymbol{\theta}$ à partir de $\mathbf{x}$. Par ailleurs, le critère d'identifiabilité est défini par la condition suivante : l'application $\boldsymbol{\theta} \to \mathbb{P}_{\mathbf{x};\boldsymbol{\theta}}$ doit être injective $\forall \boldsymbol{\theta} \in \Theta$.

On distingue en général trois types d'estimation : l'estimation est dite déterministe si les paramètres à estimer sont non-aléatoires, l'estimation est dite bayésienne s'ils sont tous aléatoires et l'estimation est dite hybride si une partie des paramètres sont aléatoires et l'autre partie est non aléatoire.

Dans le contexte d'estimation déterministe, il existe une vraie valeur des paramètres notée $\boldsymbol{\theta}_0$ et le produit scalaire mentionné précédemment est le suivant : $\forall g\left(\mathbf{x}\right)$ et $h\left(\mathbf{x}\right)$ fonctions réelles mesurables de $\left(\Omega, \mathcal{F}\right)$ dans $\left(\mathbb{R}, \mathcal{B}\left(\mathbb{R}\right)\right)$

$$\langle g, h \rangle = \int_{\Omega} g\left(\mathbf{x}\right) h\left(\mathbf{x}\right) d\mathbb{P}_{\mathbf{x};\boldsymbol{\theta}_0} = \mathbb{E}_{\mathbf{x};\boldsymbol{\theta}_0}\left[g\left(\mathbf{x}\right) h\left(\mathbf{x}\right)\right].$$

La notation $\mathbf{x};\boldsymbol{\theta}_0$ (parfois $\mathbf{x}|\boldsymbol{\theta}_0$) est utilisée ici pour souligner la dépendance de $\mathbf{x}$ par rapport à la vraie valeur $\boldsymbol{\theta}_0$. L'expression

$$\mathbf{EQM}_D = \mathbb{E}_{\mathbf{x};\boldsymbol{\theta}_0}\left[\left(\widehat{\boldsymbol{\theta}}\left(\mathbf{x}\right) - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}\left(\mathbf{x}\right) - \boldsymbol{\theta}\right)^T\right] \tag{2.11}$$

est appelé l'erreur quadratique moyenne locale au point $\boldsymbol{\theta}_0$.

Dans le contexte d'estimation bayésienne, nous définissons l'espace probabilisé $\left(\Theta, \mathcal{T}, \mathbb{P}_{\boldsymbol{\theta}}\right)$ pour le paramètre $\boldsymbol{\theta}$ avec une mesure de probabilité *a priori* $\mathbb{P}_{\boldsymbol{\theta}}$. Par conséquent, la variable jointe $\mathbf{x}, \boldsymbol{\theta}$ est définie sur un espace produit $\left(\Omega \times \Theta, \mathcal{J}, \mathbb{P}_{\mathbf{x},\boldsymbol{\theta}}\right)$ où $\mathcal{J} = \left\{(A, B) \,|A \in \mathcal{F} \text{ et } B \in \mathcal{T}\right\}$ et $\mathbb{P}_{\mathbf{x},\boldsymbol{\theta}}$ est la mesure telle que $\forall A \times B \in \mathcal{J}$, $\mathbb{P}_{\mathbf{x},\boldsymbol{\theta}}\left(A, B\right) = \mathbb{P}_{\mathbf{x}|\boldsymbol{\theta}}\left(A\right)\mathbb{P}_{\boldsymbol{\theta}}\left(B\right)$. Et le produit scalaire est défini par : $\forall g\left(\mathbf{x}, \boldsymbol{\theta}\right)$ et $h\left(\mathbf{x}, \boldsymbol{\theta}\right)$ fonctions réelles mesurables de $\left(\Omega \times \Theta, \mathcal{J}\right)$ dans $\left(\mathbb{R}, \mathcal{B}\left(\mathbb{R}\right)\right)$

$$\langle g, h \rangle = \int_{\Theta}\int_{\Omega} g\left(\mathbf{x}, \boldsymbol{\theta}\right) h\left(\mathbf{x}, \boldsymbol{\theta}\right) d\mathbb{P}_{\mathbf{x},\boldsymbol{\theta}} = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[g\left(\mathbf{x}, \boldsymbol{\theta}\right) h\left(\mathbf{x}, \boldsymbol{\theta}\right)\right].$$

L'erreur quadratique moyenne, définie par

$$\mathbf{EQM}_B = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)^T\right], \tag{2.12}$$

est dite globale puisqu'on moyenne par rapport à $\boldsymbol{\theta}$.

Si les paramètres à estimer sont hybrides $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_d^T\ \boldsymbol{\theta}_r^T\right)^T$ alors nous définissons l'espace probabilisé $(\Theta_r, \mathcal{T}_r, \mathbb{P}_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d})$ pour le paramètre $\boldsymbol{\theta}_r$ avec une mesure de probabilité a priori $\mathbb{P}_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}$ qui peut dépendre du paramètre déterministe. De même, la variable jointe $\mathbf{x}, \boldsymbol{\theta}_r; \boldsymbol{\theta}_d$ est définie sur un espace produit $\left(\Omega \times \Theta, \mathcal{J}, \mathbb{P}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\right)$ où $\mathcal{J} = \{(A,B)\,|A \in \mathcal{F} \text{ et } B \in \mathcal{T}_r\ \}$ et $\mathbb{P}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}$ est la mesure telle que $\forall\,(A,B) \in \mathcal{J}$, $\mathbb{P}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(A,B) = \mathbb{P}_{\mathbf{x}|\boldsymbol{\theta}}(A)\,\mathbb{P}_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(B)$. Le produit scalaire est défini par : $\forall g(\mathbf{x},\boldsymbol{\theta})$ et $h(\mathbf{x},\boldsymbol{\theta})$ fonctions réelles mesurables de $(\Omega \times \Theta, \mathcal{J})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

$$\langle g, h\rangle = \int_{\Theta_r}\int_{\Omega} g(\mathbf{x},\boldsymbol{\theta})\,h(\mathbf{x},\boldsymbol{\theta})\,d\mathbb{P}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d} = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[g(\mathbf{x},\boldsymbol{\theta})\,h(\mathbf{x},\boldsymbol{\theta})\right].$$

L'erreur quadratique moyenne, définie par

$$\mathbf{EQM}_H = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)^T\right], \tag{2.13}$$

est à la fois moyennée par rapport à la variable $\boldsymbol{\theta}_r$ et locale au point $\boldsymbol{\theta}_d$. ■

2.3 Borne de Cramér-Rao hybride

La borne la plus connue dans la littérature est la borne de Cramér-Rao. C'est historiquement la première borne qui a été publiée dans la littérature. Cette borne a été beaucoup étudiée et appliquée pour son expression analytique simple, son lien avec l'information de Fisher et sa relation avec l'EQM asymptotique du maximum de vraisemblance.

2.3.1 Forme générale

Soit $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_d^T\ \boldsymbol{\theta}_r^T\right]^T \in \Pi_d \times \Pi_r$ un vecteur de paramètres à estimer à partir des observations $\mathbf{x} \in \boldsymbol{\Omega}$. Si la densité de probabilité conditionnelle des observations $f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(.)$ est dérivable par rapport au vecteur de paramètres $\boldsymbol{\theta}$, alors la borne de Cramér-Rao hybride s'obtient en choisissant le vecteur

$$\mathbf{v}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d) = \frac{\partial \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}}, \tag{2.14}$$

dans l'inégalité de covariance (2.1). Nous obtenons alors

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)^T\right] \succeq \mathbf{C}\mathbf{V}^{-1}\mathbf{C}^T,$$

où la matrice $\mathbf{V}$ est donnée par

$$\mathbf{V} = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}}\frac{\partial \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}^T}\right], \tag{2.15}$$

et la matrice $\mathbf{C}$ est

$$\mathbf{C} = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)\frac{\partial \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}^T}\right]. \tag{2.16}$$

Si $f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}$ est deux fois dérivable par rapport au vecteur $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_d^T \ \boldsymbol{\theta}_r^T\right]^T$ alors la matrice $\mathbf{V}$ s'écrit également

$$\mathbf{V} = -\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial^2 \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]. \tag{2.17}$$

Cependant, la matrice $\mathbf{C}$ dépend *a priori* de l'estimateur sans hypothèses supplémentaires. Il existe deux formes de la borne de Cramér-Rao selon la classe d'estimateurs considérée : les estimateurs sans biais et les estimateurs biaisés. Nous nous intéressons ici à l'expression de cette borne pour une classe d'estimateur $\widehat{\boldsymbol{\theta}}$ sans biais, c'est-à-dire :

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right] = \mathbf{0}. \tag{2.18}$$

Cette condition (2.18) est généralement satisfaite par les estimateurs opérants dans les zones asymptotiques (à fort RSB ou à grand nombre d'observations). Dans ce cas, si on note θ_i le i^{ème} élément du vecteur $\boldsymbol{\theta}$, alors la i^{ème} colonne de la matrice $\mathbf{C}$ s'écrit

$$\begin{aligned}\mathbf{c}_i &= \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)\frac{\partial \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\theta_i}\right] \\ &= \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right] - \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial}{\partial\theta_i}\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)\right].\end{aligned}$$

Puisque l'estimateur est non biaisé (2.18) et ne dépend que des données $\left(\frac{\partial\widehat{\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_i} = 0\right)$ donc $\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial}{\partial\theta_i}\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)\right] = -\mathbf{e}_i$ où $\mathbf{e}_i$ est un vecteur unitaire qui vaut 1 à la i^{ème} composante et 0 ailleurs. Par conséquent

$$\forall i,\ \mathbf{c}_i = \mathbf{e}_i \Leftrightarrow \mathbf{C} = \mathbf{I}.$$

Finalement la borne de Cramér-Rao hybride s'écrit simplement

$$\mathbf{BCRH} = \mathbb{E}^{-1}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}}\frac{\partial \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}^T}\right]. \tag{2.19}$$

Si la densité de probabilité conditionnelle est deux fois dérivable par rapport à $\boldsymbol{\theta}$, alors

$$\mathbf{BCRH} = -\mathbb{E}^{-1}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial^2 \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]. \tag{2.20}$$

Par analogie avec la borne de Cramér-Rao classique, on appelle la matrice d'information de Fisher la quantité

$$\mathbf{F}_H \triangleq \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}}\frac{\partial \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}^T}\right] \tag{2.21}$$

ou

$$\mathbf{F}_H = -\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial^2 \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]. \tag{2.22}$$

Plus les valeurs diagonales de la matrice $\mathbf{F}_H$ sont grandes et plus l'EQM minimale pour chacun des paramètres $\boldsymbol{\theta}$ sera petite. Cependant, il y a très peu de résultats sur les conditions d'atteignabilité de cette borne. A notre connaissance, un seul papier [NM09] compare la borne de Cramér-Rao hybride à la borne de Cramér-Rao déterministe lorsqu'on considère que les paramètres aléatoires sont des paramètres de nuisances. Pour le cas des estimateurs biaisés, le raisonnement est similaire et son expression est disponible dans [Eld06].

2.3.2 Formule de Slepian-Bangs hybride dans le cas gaussien

Il existe une expression plus simple de la borne de Cramér-Rao losque le modèle d'observation est gaussien complexe, c'est à dire que les observations complexe $\mathbf{x}$ vérifient

$$\mathbf{x} = \mathbf{m}(\boldsymbol{\theta}) + \mathbf{b},$$

où $\mathbf{m}(.)$ est une fonction déterministe dérivable par rapport aux paramètres réels $\boldsymbol{\theta}$ et $\mathbf{b}$ est un vecteur aléatoire gaussien complexe circulaire de moyenne nulle et de matrice de covariance $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. On le note aussi $\mathbf{x} \sim \mathbb{C}\mathcal{N}(\mathbf{m}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. Pour obtenir cette expression, remarquons d'abord que la densité de probabilité des données peut s'écrire comme le produit suivant

$$f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x}, \boldsymbol{\theta}_r; \boldsymbol{\theta}_d) = f_{\mathbf{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x}|\boldsymbol{\theta}_r; \boldsymbol{\theta}_d) f_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\boldsymbol{\theta}_r; \boldsymbol{\theta}_d).$$

Puisqu'une loi gaussienne est indéfiniment dérivable, on peut utiliser l'expression (2.22) pour obtenir l'information de Fisher hybride :

$$\begin{aligned}
\mathbf{F}_H &= -\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial^2 \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x}, \boldsymbol{\theta}_r; \boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right] \\
&= -\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial^2 \ln f_{\mathbf{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x}|\boldsymbol{\theta}_r; \boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} + \frac{\partial^2 \ln f_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\boldsymbol{\theta}_r; \boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right] \\
&= -\mathbb{E}_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial^2 \ln f_{\mathbf{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x}|\boldsymbol{\theta}_r; \boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]\right] - \mathbb{E}_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial^2 \ln f_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\boldsymbol{\theta}_r; \boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right].
\end{aligned}$$

Or $\mathbf{F}_D \triangleq \mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial^2 \ln f_{\mathbf{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]$ est la matrice d'information de Fisher déterministe lorsqu'on considère les paramètres $\boldsymbol{\theta}_r$ et $\boldsymbol{\theta}_d$ certains. D'après la formule de Slepian-Bangs [Sle54, Ban71], l'élément (i, j) de matrice $\mathbf{F}_D$ est donnée par

$$\{\mathbf{F}_D\}_{i,j} = 2\,\mathrm{Re}\left(\frac{\partial \mathbf{m}^H(\boldsymbol{\theta})}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial\theta_j}\right) + Tr\left(\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial\theta_j}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right)$$

donc la matrice d'information de Fisher hybride s'écrit

$$\mathbf{F}_H = -\mathbb{E}_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}[\mathbf{F}_D] - \mathbb{E}_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{\partial^2 \ln f_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\boldsymbol{\theta}_r; \boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right] \tag{2.23}$$

Cette expression est beaucoup plus rapide à évaluer car il ne reste plus que des espérances sur la loi *a priori* $f_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}$.

2.3.3 Efficacité du maximum de vraisemblance maximum *a posteriori* pour les modèles linéaires gaussiens

Il existe des propriétés de consistence et de convergence du maximum de vraisemblance qui atteint la borne de Cramér-Rao sous certaines conditions [LC03]. On dit alors que le maximum de vraisemblance est efficace. D'après nos connaissances, une telle étude n'a jamais été faite dans le contexte de l'estimation hybride. Néanmoins, nous pouvons montrer qu'il existe au moins un estimateur atteignant la borne de Cramér-Rao hybride pour les modèles linéaires gaussiens.

Soient $\mathbf{x} \in \mathbb{R}^M$ un vecteur d'observations avec $M \geq 2$ et $\boldsymbol{\theta} = [\theta_d \ \theta_r]^T \in \mathbb{R}^2$ un vecteur de paramètres hybrides où θ_d est un paramètre déterministe et θ_r est un paramètre aléatoire de densité de probabilité *a priori* gaussienne, $\theta_r \sim \mathcal{N}\left(m_{\theta_r}, \sigma^2_{\theta_r}\right)$ avec $m_{\theta_r} \in \mathbb{R}$ et $\sigma_{\theta_r} > 0$. Le modèle d'observation est décrit par la relation

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{b},$$

où $\mathbf{H}$ est une matrice réelle de taille $M \times 2$ de rang 2 et $\mathbf{b}$ représente le bruit de mesure supposé aléatoire gaussien, $\mathbf{b} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_N\right)$, et indépendant de θ_r. Ainsi la densité de probabilité jointe de $\mathbf{x}$,θ_r;θ_d est donnée par

$$f_{\mathbf{x},\theta_r;\theta_d}\left(\mathbf{x},\theta_r;\theta_d\right) = \frac{e^{\frac{-1}{2\sigma^2}\|\mathbf{x}-\mathbf{H}\boldsymbol{\theta}\|^2} e^{\frac{-1}{2\sigma^2_{\theta_r}}\|\theta_r - m_{\theta_r}\|^2}}{\left(2\pi\sigma^2\right)^{\frac{M}{2}}\left(2\pi\sigma^2_{\theta_r}\right)^{\frac{1}{2}}}. \tag{2.24}$$

En utilisant la formule de Slepian-bangs (2.23), l'information de Fisher hybride s'écrit

$$\mathbf{F}_H = \frac{1}{\sigma^2}\mathbf{H}^T\mathbf{H} + \frac{1}{\sigma^2_{\theta_r}}\mathbf{e}_2\mathbf{e}_2^T \tag{2.25}$$

avec $\mathbf{e}_2 = [0 \ 1]^T$. Il est légitime de se demander si cette performance prédite est atteignable par un estimateur. En estimation hybride, peu d'algorithmes ont été développés dans la littérature, nous pouvons souligner l'estimateur joint du maximum de vraisemblance maximum *a posteriori* (MVMAP) [TT09] qui consiste à maximiser le critère suivant :

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{MVMAP} &= \arg\max_{\boldsymbol{\theta}\in\mathbb{R}^2} f_{\mathbf{x},\theta_r;\theta_d}\left(\mathbf{x},\theta_r;\theta_d\right) \\ &= \arg\min_{\theta_d,\theta_r\in\mathbb{R}} \frac{\|\mathbf{x}-\mathbf{H}\boldsymbol{\theta}\|^2}{2\sigma^2} + \frac{\|\theta_r - m_{\theta_r}\|^2}{2\sigma^2_{\theta_r}}\end{aligned} \tag{2.26}$$

Le critère (2.26) est convexe, le minimum est atteint pour

$$\left.\frac{\partial}{\partial\boldsymbol{\theta}}\left(\frac{\|\mathbf{x}-\mathbf{H}\boldsymbol{\theta}\|^2}{2\sigma^2} + \frac{\|\theta_r - m_{\theta_r}\|^2}{2\sigma^2_{\theta_r}}\right)\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{MVMAP}} = 0,$$

$\Leftrightarrow$

$$\left.\frac{-\mathbf{H}^T\left(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}\right)}{\sigma^2} + \frac{\mathbf{e}_2\mathbf{e}_2^T\left(\boldsymbol{\theta} - m_{\theta_r}\mathbf{e}_2\right)}{\sigma^2_{\theta_r}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_{MVMAP}} = 0.$$

Donc, on obtient

$$\widehat{\boldsymbol{\theta}}_{MVMAP} = \mathbf{F}_H^{-1}\left(\frac{1}{\sigma^2}\mathbf{H}^T\mathbf{x} + \frac{1}{\sigma^2_{\theta_r}}m_{\theta_r}\mathbf{e}_2\right), \tag{2.27}$$

avec $\mathbf{F}_H$ la matrice définie par (2.25). Pour savoir si la borne de Cramér-Rao s'applique à l'estimateur MVMAP, il faut montrer d'abord que cet estimateur est non biaisé. On a,

$$\begin{aligned}\mathbb{E}_{\mathbf{x},\theta_r;\theta_d}\left[\widehat{\boldsymbol{\theta}}_{MVMAP} - \boldsymbol{\theta}\right] &= \mathbf{F}_H^{-1}\left(\mathbb{E}_{\mathbf{x},\theta_r;\theta_d}\left[\frac{1}{\sigma^2}\mathbf{H}^T\mathbf{x} - \mathbf{F}_H\boldsymbol{\theta}\right] + \frac{1}{\sigma^2_{\theta_r}}m_{\theta_r}\mathbf{e}_2\right) \\ &= \mathbf{F}_H^{-1}\left(\frac{1}{\sigma^2}\mathbf{H}^T\mathbb{E}_{\mathbf{x},\theta_r;\theta_d}\left[\mathbf{x}-\mathbf{H}\boldsymbol{\theta}\right] + \frac{\mathbf{e}_2\mathbf{e}_2^T}{\sigma^2_{\theta_r}}\mathbb{E}_{\mathbf{x},\theta_r;\theta_d}\left[m_{\theta_r}\mathbf{e}_2 - \boldsymbol{\theta}\right]\right) \\ &= \mathbf{0}.\end{aligned} \tag{2.28}$$

Ce qui signifie que l'estimateur θ_{MVMAP} est un estimateur non biaisé de θ. Donc la borne de Cramér-Rao est applicable pour cet estimateur et sa variance s'écrit

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x},\theta_r;\theta_d}\left[\left(\widehat{\boldsymbol{\theta}}_{MVMAP}-\boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}_{MVMAP}-\boldsymbol{\theta}\right)^T\right]\\
=\;&\mathbf{F}_H^{-1}\mathbb{E}_{\mathbf{x},\theta_r;\theta_d}\left[\left(\frac{1}{\sigma^2}\mathbf{H}^T\mathbf{x}+\frac{1}{\sigma_{\theta_r}^2}m_{\theta_r}\mathbf{e}_2-\mathbf{F}_H\boldsymbol{\theta}\right)\left(\frac{1}{\sigma^2}\mathbf{H}^T\mathbf{x}+\frac{1}{\sigma_{\theta_r}^2}m_{\theta_r}\mathbf{e}_2-\mathbf{F}_H\boldsymbol{\theta}\right)^T\right]\mathbf{F}_H^{-1}\\
=\;&\mathbf{F}_H^{-1}\mathbb{E}_{\mathbf{x},\theta_r;\theta_d}\left[\left(\frac{1}{\sigma^2}\mathbf{H}^T\mathbf{b}+\frac{\mathbf{e}_2\mathbf{e}_2^T}{\sigma_{\theta_r}^2}\left(m_{\theta_r}\mathbf{e}_2-\boldsymbol{\theta}\right)\right)\left(\frac{1}{\sigma^2}\mathbf{H}^T\mathbf{b}+\frac{\mathbf{e}_2\mathbf{e}_2^T}{\sigma_{\theta_r}^2}\left(m_{\theta_r}\mathbf{e}_2-\boldsymbol{\theta}\right)\right)^T\right]\mathbf{F}_H^{-1}.
\end{aligned}$$

Puisque $\mathbf{b}$ est indépendante de θ_r, nous obtenons

$$\begin{aligned}
\mathbb{E}_{\mathbf{x},\theta_r;\theta_d}\left[\left(\widehat{\boldsymbol{\theta}}_{MVMAP}-\boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}_{MVMAP}-\boldsymbol{\theta}\right)^T\right] &= \mathbf{F}_H^{-1}\left(\frac{1}{\sigma^4}\mathbf{H}^T\mathbb{E}_{\mathbf{x},\theta_r;\theta_d}\left[\mathbf{b}\mathbf{b}^T\right]\mathbf{H}+\frac{\mathbf{e}_2\mathbf{e}_2^T}{\sigma_{\theta_r}^4}\mathbb{E}_{\mathbf{x},\theta_r;\theta_d}\left[\theta_r^2\right]\right)\mathbf{F}_H^{-1}\\
&= \mathbf{F}_H^{-1}=\mathbf{BCRH}.
\end{aligned}$$

Nous déduisons donc que l'estimateur MVMAP est efficace pour toute estimation linéaire hybride avec un bruit additif gaussien quel que soit le nombre de mesure M ou la puissance du bruit σ^2. Ce résultat étend, en un certain sens, l'efficacité du maximum de vraisemblance pour les problèmes linéaires [Kay93]. Cependant pour les problèmes d'estimation non-linéaire, nous ne pouvons pas faire le même raisonnement car il est en général impossible d'obtenir une forme explicite de $\widehat{\boldsymbol{\theta}}_{MVMAP}$. Par ailleurs, l'évaluation de l'EQM de l'estimateur MVMAP exhibe souvent un phénomène de décrochement que la borne de Cramér-Rao est incapable de prédire. Or la connaissance du seuil de décrochement est fondamentale pour délimiter la zone de fonctionnement optimale des estimateurs. C'est pourquoi nous verrons dans les sections suivantes, un ensemble de bornes permettant de prédire ce seuil.

2.4 Borne de Barankin hybride

2.4.1 Introduction

Dans la littérature, nous séparons les bornes en deux catégories : les bornes de faible erreur ("small-error bounds") et les bornes de large erreur ("large-error bounds"). Cette séparation fait la distinction entre les bornes qui exhibent le phénomène de décrochement, en l'occurence les bornes de large erreur, et celles qui ne le prédisent pas, en l'occurence les bornes de faible erreur. Les représentants types de ces deux catégories sont respectivement la borne de Cramér-Rao pour les bornes de faible erreur et la borne de Barankin pour les bornes de large erreur.

Rappelons nous que l'inégalité de covariance (2.1) permet d'obtenir une borne si la matrice $\mathbf{C}$ est indépendante de $\widehat{\boldsymbol{\theta}}$. Pour cela, nous devons restreindre l'applicabilité des bornes à une classe d'estimateurs satisfaisant un certain nombre de contraintes. Par exemple, la borne de Cramér-Rao hybride, présenté à la section (2.3), nécessite que l'estimateur soit localement non biaisé (eq. 2.18). Si nous voulons obtenir une borne plus précise (dans le sens où elle majore la BCRH), il nous faut *a priori* restreindre davantage la classe d'estimateurs. En effet, Barankin [Bar49] s'est intéressé, dans le contexte d'estimation déterministe, à des estimateurs localement non biaisés

quelles que soient les vraies valeurs des paramètres, c'est-à-dire

$$\mathbb{E}_{\mathbf{x};\boldsymbol{\theta}_d}\left[\widehat{\boldsymbol{\theta}}_d(\mathbf{x})-\boldsymbol{\theta}_d\right]=\mathbf{0},\ \forall\boldsymbol{\theta}_d\in\Pi_d, \tag{2.29}$$

où Π_d est l'espace de définition des paramètres $\boldsymbol{\theta}_d$. Cette condition est différente de celle imposée par la borne de Cramér-Rao qui considère un ensemble d'estimateurs ayant un biais nul à une valeur précise de $\boldsymbol{\theta}_d$. Cependant, la forme générale de la borne de Barankin est la solution d'une équation intégrale qui n'a pas de solution analytique. Néanmoins des approximations de cette contrainte ont été proposées par McAulay, Seidman and Hofstetter [MS69, MH71, Abe93] puis étendues par [CGQL08, TT10a, CRL09]. Tous ces travaux proposent des bornes calculables en pratique et dont la précision varie selon la qualité d'approximation de la contrainte (2.29).

Pour ne pas alourdir le manuscript, nous ne présenterons pas les bornes de Barankin déterministes et ses extensions. Les lecteurs intéressés peuvent consulter les références sus-mentionnées. Nous nous focalisons uniquement sur la borne Barankin hybride proposée par Reuven et Messer [RM97] qui sera utile plus tard pour dériver de nouvelles bornes plus précises (voir section (2.5) et (2.6)).

2.4.2 Expression générale

Soit $\boldsymbol{\theta}=\left[\boldsymbol{\theta}_d^T\ \boldsymbol{\theta}_r^T\right]^T\in\Pi_d\times\Pi_r\subset\mathbb{R}^D\times\mathbb{R}^R$ un vecteur de paramètres à estimer à partir des observations $\mathbf{x}\in\boldsymbol{\Omega}$. Dans ce contexte d'estimation hybride, l'extension "naturelle" de la contrainte proposée par Barankin (2.29) est de se restreindre à la classe d'estimateurs $\widehat{\boldsymbol{\theta}}$ satisfaisant :

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right]=\mathbf{0},\ \forall\boldsymbol{\theta}\in\Pi_d\times\Pi_r. \tag{2.30}$$

Pour les mêmes raisons qu'en déterministe, cette contrainte n'aboutit pas à une forme explicite de la borne de Barankin hybride. Une version "discrétisée" de cette contrainte a été proposée dans [RM97] en considérant que l'estimateur $\widehat{\boldsymbol{\theta}}$ est de biais nul au point $\boldsymbol{\theta}\in\Pi_d\times\Pi_r$ et à un ensemble de points tests $\{\boldsymbol{\theta}+\mathbf{h}_i|\boldsymbol{\theta}+\mathbf{h}_i\in\Pi_d\times\Pi_r\}_{i=1\cdots N}$ où N est le nombre de points test considérés. Par souci de simplicité, notons la variable jointe $\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d\triangleq\mathbf{x},\boldsymbol{\theta}$. Les contraintes considérées sont donc

$$\begin{cases}\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right]=\mathbf{0}\\ \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_i}\left[\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right]=\mathbf{h}_i,\ i=1\cdots N.\end{cases} \tag{2.31}$$

Pour obtenir cette borne, la fonction $\mathbf{v}(\mathbf{x},\boldsymbol{\theta})=\left[v_1(\mathbf{x},\boldsymbol{\theta})\ v_2(\mathbf{x},\boldsymbol{\theta})\ \cdots\ v_N(\mathbf{x},\boldsymbol{\theta})\right]^T$ est donnée par : pour tout $i=1\cdots N$

$$v_i(\mathbf{x},\boldsymbol{\theta})=\begin{cases}\frac{f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_i)}{f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta})}-1,\ \boldsymbol{\theta}\in\boldsymbol{\Theta},\\ 0,\text{ sinon},\end{cases}$$

où $\boldsymbol{\Theta}$ est un ensemble défini par

$$\boldsymbol{\Theta}=\left\{\boldsymbol{\theta}|f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta})>0,\ \mathbf{x}\in\boldsymbol{\Omega}\right\}. \tag{2.32}$$

En appliquant l'inégalité de covariance (2.1), nous obtenons

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)^T\right]\succeq\mathbf{C}\mathbf{V}^{-1}\mathbf{C}^T$$

où l'élément (i,j) de la matrice $\mathbf{V}$ est défini par

$$\begin{aligned} V_{ij} &= \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\left(\frac{f_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_i\right)}{f_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}\right)}-1\right)\left(\frac{f_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_j\right)}{f_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}\right)}-1\right)\right] \\ &= \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\frac{f_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_i\right)f_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_j\right)}{f^2_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}\right)}\right]-1, \end{aligned} \tag{2.33}$$

et le $i^{ème}$ vecteur colonne de la matrice $\mathbf{C}$ est donné par

$$\begin{aligned} \mathbf{c}_i &= \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}}\left(\mathbf{x}\right)-\boldsymbol{\theta}\right)\left(\frac{f_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_i\right)}{f_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}\right)}-1\right)\right] \\ &= \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}}\left(\mathbf{x}\right)-\boldsymbol{\theta}\right)\frac{f_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_i\right)}{f_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}\right)}\right]-\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}}\left(\mathbf{x}\right)-\boldsymbol{\theta}\right] \\ &= \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_i}\left[\widehat{\boldsymbol{\theta}}\left(\mathbf{x}\right)-\boldsymbol{\theta}\right]-\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}}\left(\mathbf{x}\right)-\boldsymbol{\theta}\right]. \end{aligned} \tag{2.34}$$

D'après les conditions de biais (2.31), nous avons

$$\mathbf{c}_i = \mathbf{h}_i. \tag{2.35}$$

Remarquons que les contraintes (2.31) ne sont pas nécessaires pour obtenir le résultat (2.35). Un relâchement des contraintes est possible et sera discuté dans la section (2.4.3).

La borne de Barankin hybride avec N points test pour un ensemble de points test $\{\boldsymbol{\theta}+\mathbf{h}_i \in \Theta\}_{i=1\cdots N}$ s'exprime donc sous la forme suivante :

$$\mathbf{BBH}\left(\mathbf{H}\right) = \mathbf{H}\mathbf{Q}^{-1}\mathbf{H}^T,$$

où $\mathbf{H} = [\mathbf{h}_1\ \mathbf{h}_2\ \cdots\ \mathbf{h}_N]$ et les éléments (i,j) de la matrice $\mathbf{Q}$ sont

$$Q_{i,j} = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\frac{f_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_i\right)f_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_j\right)}{f^2_{\mathbf{x},\boldsymbol{\theta}}\left(\mathbf{x},\boldsymbol{\theta}\right)}\right]-1. \tag{2.36}$$

Sous réserve d'existence, la meilleure borne de Barankin avec N points test est donnée par

$$\mathbf{BBH} = \max_{\mathbf{h}_1\ \mathbf{h}_2\ \cdots\ \mathbf{h}_N} \mathbf{H}\left(\mathbf{Q}\right)^{-1}\mathbf{H}^T. \tag{2.37}$$

Nous remarquons que la meilleure borne de Barankin avec N points test nécessite la maximisation d'un critère suivant N vecteurs de même dimension que le vecteur de paramètre $\boldsymbol{\theta}$. Notons que l'espérance jointe peut être décomposée en deux espérances imbriquées $\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[.\right] = \mathbb{E}_{\boldsymbol{\theta}}\left[\mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}}\left[.\right]\right]$. Récemment, des expressions analytiques de l'espérance conditionnelle suivante

$$\eta\left(\mathbf{h}_i,\mathbf{h}_j\right) \triangleq \mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}}\left[\frac{f_{\mathbf{x}|\boldsymbol{\theta}}\left(\mathbf{x}|\boldsymbol{\theta}+\mathbf{h}_i\right)f_{\mathbf{x}|\boldsymbol{\theta}}\left(\mathbf{x}|\boldsymbol{\theta}+\mathbf{h}_j\right)}{f^2_{\mathbf{x}|\boldsymbol{\theta}}\left(\mathbf{x}|\boldsymbol{\theta}\right)}\right],$$

ont été proposées par [VRBM14, eq. (13) et eq. (15)] pour les modèles gaussiens à moyenne paramétrée ou à covariance paramétrée. Ces résultats ont permis de réduire considérablement le coût de calcul de cette borne.

2.4.3 Relaxation des contraintes

En reprenant les calculs du vecteur $\mathbf{c}_i$ (2.34), il suffit de se restreindre à la classe d'estimateurs satisfaisant

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_i}\left[\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right]-\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right]=\mathbf{h}_i. \tag{2.38}$$

pour obtenir la borne de Barankin (2.37). La condition susmentionnée (2.38) est bien nécessaire et suffisante. Nous essayons dans la suite d'interpréter les conséquences de cette contrainte sur $\widehat{\boldsymbol{\theta}}$. Si nous décomposons l'estimateur hybride en deux parties $\widehat{\boldsymbol{\theta}}(\mathbf{x})=\left[\widehat{\boldsymbol{\theta}}_d^T(\mathbf{x})\ \widehat{\boldsymbol{\theta}}_r^T(\mathbf{x})\right]^T$ où $\widehat{\boldsymbol{\theta}}_d$ est un estimateur de $\boldsymbol{\theta}_d$ et $\widehat{\boldsymbol{\theta}}_r$ est un estimateur de $\boldsymbol{\theta}_r$ alors ces deux estimateurs sont soumis aux contraintes

$$\begin{aligned}\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r+\mathbf{h}_{ri};\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_d(\mathbf{x})-(\boldsymbol{\theta}_d+\mathbf{h}_{di})\right] &= \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widehat{\boldsymbol{\theta}}_d(\mathbf{x})-\boldsymbol{\theta}_d\right] &(2.39)\\ \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r+\mathbf{h}_{ri};\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_r(\mathbf{x})-(\boldsymbol{\theta}_r+\mathbf{h}_{ri})\right] &= \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widehat{\boldsymbol{\theta}}_d(\mathbf{x})-\boldsymbol{\theta}_r\right] &(2.40)\end{aligned}$$

avec $\mathbf{h}_i=\left[\mathbf{h}_{di}^T\ \mathbf{h}_{ri}^T\right]^T$. Etudions d'abord la deuxième égalité (2.40). Si nous notons $1_{\Pi_r}(\boldsymbol{\theta}_r)$ la fonction indicatrice qui vaut 1 si $\boldsymbol{\theta}_r\in\Pi_r$ et 0 sinon, alors :

$$\begin{aligned}&\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r+\mathbf{h}_{ri};\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_r(\mathbf{x})-(\boldsymbol{\theta}_r+\mathbf{h}_{ri})\right]\\ =\ &\mathbb{E}_{\boldsymbol{\theta}_r+\mathbf{h}_{ri};\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}_r+\mathbf{h}_{ri};\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_r(\mathbf{x})-(\boldsymbol{\theta}_r+\mathbf{h}_{ri})\right]\right]\\ =\ &\int_{\mathbb{R}^r}\mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}_r+\mathbf{h}_{ri};\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_r(\mathbf{x})-(\boldsymbol{\theta}_r+\mathbf{h}_{ri})\right]f_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\boldsymbol{\theta}_r+\mathbf{h}_{ri};\boldsymbol{\theta}_d+\mathbf{h}_{di})\,1_{\Pi_r}(\boldsymbol{\theta}_r)\,d\boldsymbol{\theta}_r\\ =\ &\int_{\mathbb{R}^r}\mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}'_r;\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_r(\mathbf{x})-\boldsymbol{\theta}'_r\right]f_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\boldsymbol{\theta}'_r;\boldsymbol{\theta}_d+\mathbf{h}_{di})\,1_{\Pi_r}(\boldsymbol{\theta}'_r-\mathbf{h}_{ri})\,d\boldsymbol{\theta}'_r,\end{aligned}$$

en utilisant un changement de variable $\boldsymbol{\theta}'_r=\boldsymbol{\theta}_r+\mathbf{h}_{ri}$. Si $1_{\Pi_r}(\boldsymbol{\theta}_r)=1_{\Pi_r}(\boldsymbol{\theta}_r-\mathbf{h}_{ri})$ pour tout $\boldsymbol{\theta}_r\in\Pi_r$, impliquant que le domaine Π_r n'est pas compact, alors

$$\begin{aligned}\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r+\mathbf{h}_{ri};\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_r(\mathbf{x})-(\boldsymbol{\theta}_r+\mathbf{h}_{ri})\right] &= \mathbb{E}_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}_r;\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_r(\mathbf{x})-\boldsymbol{\theta}_r\right]\right]\\ &= \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_r(\mathbf{x})-\boldsymbol{\theta}_r\right]. &(2.41)\end{aligned}$$

D'après (2.40) et (2.41), la condition sous-jacente que doit vérifier $\widehat{\boldsymbol{\theta}}_r(\mathbf{x})$ est

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_r(\mathbf{x})-\boldsymbol{\theta}_r\right]=\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widehat{\boldsymbol{\theta}}_d(\mathbf{x})-\boldsymbol{\theta}_r\right], \tag{2.42}$$

autrement dit $\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widehat{\boldsymbol{\theta}}_r(\mathbf{x})-\boldsymbol{\theta}_r\right]=\boldsymbol{\mu}_r\in\mathbb{R}^r$, $\forall\boldsymbol{\theta}_d\in\Pi_d$. En faisant la même étude pour $\widehat{\boldsymbol{\theta}}_d(\mathbf{x})$, nous avons

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r+\mathbf{h}_{ri};\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_d(\mathbf{x})-(\boldsymbol{\theta}_d+\mathbf{h}_{di})\right]=\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_d(\mathbf{x})-(\boldsymbol{\theta}_d+\mathbf{h}_{di})\right],$$

et par l'égalité (2.39), nous obtenons

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}_d(\mathbf{x})-(\boldsymbol{\theta}_d+\mathbf{h}_{di})\right]=\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widehat{\boldsymbol{\theta}}_d(\mathbf{x})-\boldsymbol{\theta}_d\right]. \tag{2.43}$$

Ce qui signifie $\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widehat{\boldsymbol{\theta}}_d(\mathbf{x})-\boldsymbol{\theta}_d\right]=\boldsymbol{\mu}_d\in\mathbb{R}^D, \forall\boldsymbol{\theta}_d\in\Pi_d$. Nous pouvons résumer les conditions de validité des bornes de Barankin hybrides "usuelles" par trois contraintes fondamentales :

- (R1) : $\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right]=\boldsymbol{\mu}\in\mathbb{R}^{D+R}, \forall\boldsymbol{\theta}_d\in\Pi_d$.
- (R2) : $\mathbf{h}_{di}\in\mathbb{R}^D$ tel que $\boldsymbol{\theta}_d+\mathbf{h}_{di}\in\Pi_d$,
- (R3) : $\mathbf{h}_{ri}\in\mathbb{R}^R$ tel que $1_{\Pi_r}(\boldsymbol{\theta}_r-\mathbf{h}_{ri})=1_{\Pi_r}(\boldsymbol{\theta}_r),\ \forall\boldsymbol{\theta}_r\in\Pi_r$.

Une autre interprétation de la condition (R3) est donnée par

$$\forall\boldsymbol{\theta}_r\in\mathbb{R}^R,\ 1_{\Pi_r^0}(\boldsymbol{\theta}_r)=\begin{cases}0 \text{ si } \sum\limits_{\mathbf{h}_r\in A}\left(\sum\limits_{l\in\mathbb{Z}}1_{\Pi_r}(\boldsymbol{\theta}_r+l\mathbf{h}_r)\right)=0,\\ 1, \text{ sinon,}\end{cases}. \tag{2.44}$$

où A et Π_r^0 sont des sous-ensemble de $\mathbb{R}^R$, ce qui signifie en général que Π_r peut être un sous ensemble discret de $\mathbb{R}^R$ ou un sous-ensemble d'intervalles de $\mathbb{R}^R$. Cependant, pour que la borne de Cramér-Rao hybride existe, $\mathbf{h}_r$ doit être infinitésimal [RM97, III.A], ce qui impose que $\Pi_r=\mathbb{R}^R$. Enfin, notons que, sous la contrainte (R3), nous avons $\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\frac{f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_i)}{f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta})}-1\right]=0$, la borne de Barankin hybride est alors compressible (dans le sens où on peut réduire le nombre de points test) en utilisant les fonctions noyaux mentionnées dans [TT09].

2.4.4 Compression par transformation du noyau

L'augmentation du nombre de points test et la recherche précise des valeurs optimales de ces points s'avèrent coûteuses en temps de calcul pour la borne de Barankin hybride (eq. 2.37). Par conséquent, deux stratégies d'optimisation ont été proposées dans la littérature : la première est de prendre un nombre de points test raisonnables et d'effectuer le processus d'optimisation pour ces points test [CGQL08]. La seconde est de prendre beaucoup de points test répartis de manière uniforme sur l'espace des paramètres puis d'effectuer une transformation linéaire, basée sur les particularités de la densité de probabilité des observations, qui permet de réduire le nombre de points tests par élimination de redondances [TT10a]. Cette méthode s'affranchit de l'étape d'optimisation par le fait que le nombre de points test au départ est conséquent. Le papier [TT09] montre qu'une transformation linéaire, sous forme discrète ou intégrale, par des fonctions noyaux basées sur le rapport de vraisemblance centré permet de générer une large famille de bornes de Barankin hybrides lorsque la loi *a priori* des paramètres aléatoires est indépendante des paramètres déterministes, c'est à dire $f_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}=f_{\boldsymbol{\theta}_r}\ \forall\boldsymbol{\theta}_d\in\Pi_d$. Nous montrons ici que ces résultats peuvent être étendus au cas général où la densité *a priori* des paramètres aléatoires peut dépendre des paramètres déterministes pour la classe d'estimateurs satisfaisant (R1). Notons dans la suite $v(\mathbf{x},\boldsymbol{\theta},\mathbf{h})\triangleq v(\mathbf{x},\boldsymbol{\theta},\mathbf{h}_r,\mathbf{h}_d)$ et considérons le cas simple d'un vecteur de paramètre hybride composé d'un seul paramètre déterministe et d'un seul paramètre aléatoire, c'est à dire $\boldsymbol{\theta}=[\theta_d,\theta_r]^T$. De ce fait, la contrainte de biais (2.34) devient

$$\begin{pmatrix}h_{di}\\ h_{ri}\end{pmatrix}=\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)v(\mathbf{x},\boldsymbol{\theta},h_{ri},h_{di})\right]. \tag{2.45}$$

Soit $\alpha(\tau,\mathbf{h})=\alpha(\tau,h_r,h_d),\tau\in\Phi\subset\mathbb{R}$, alors sous les conditions (R1)(R2) et (R3), $\widehat{\boldsymbol{\theta}}$ vérifie :

$$\sum_{i=1}^{N}\alpha(\tau,h_{ri},h_{di})\begin{pmatrix}h_{di}\\ h_{ri}\end{pmatrix}=\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\left(\widehat{\boldsymbol{\theta}}(\mathbf{x})-\boldsymbol{\theta}\right)\sum_{i=1}^{N}\alpha(\tau,h_{ri},h_{di})\,v(\mathbf{x},\boldsymbol{\theta},h_{ri},h_{di})\right]. \tag{2.46}$$

Si de plus la fonction $\alpha(\tau, \mathbf{h})$ est intégrable par rapport à $\mathbf{h}$ sur $\Lambda_{\boldsymbol{\theta}} \triangleq \{\mathbf{h} | \boldsymbol{\theta} + \mathbf{h} \in \Pi_d \times \Pi_r\}, \forall \tau \in \Phi$, alors l'équation (2.46) est, à une constante de normalisation près, une approximation numérique de l'équation intégrale suivante :

$$\int_{\Lambda_{\boldsymbol{\theta}}} \alpha(\tau, \mathbf{h}) \mathbf{h} d\mathbf{h} = \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}_r | \boldsymbol{\theta}_d} \left[\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta} \right) \int_{\Lambda_{\boldsymbol{\theta}}} \alpha(\tau, \mathbf{h}) v(\mathbf{x}, \boldsymbol{\theta}, \mathbf{h}) d\mathbf{h} \right], \tag{2.47}$$

avec l'élément différentiel $d\mathbf{h} = dh_r dh_d$. Soit

$$\eta_\alpha(\mathbf{x}, \boldsymbol{\theta}, \tau) = \int_{\Lambda_{\boldsymbol{\theta}}} \alpha(\tau, \mathbf{h}) v(\mathbf{x}, \boldsymbol{\theta}, \mathbf{h}) d\mathbf{h} \tag{2.48}$$

et

$$\boldsymbol{\Gamma}_\alpha(\tau) = \int_{\Lambda_{\boldsymbol{\theta}}} \alpha(\tau, \mathbf{h}) \mathbf{h} d\mathbf{h}, \tag{2.49}$$

alors sous la contrainte (2.47), les matrices $\mathbf{V}$ (2.33) et $\mathbf{C}$ (2.35) deviennent

$$\mathbf{V} = \mathbb{E}_{\mathbf{x}, \theta_r; \theta_d} \left[\eta_\alpha(\mathbf{x}, \boldsymbol{\theta}, \tau)^2 \right] = \int_{\Lambda_{\boldsymbol{\theta}}^2} \alpha(\tau, \mathbf{h}) K(\mathbf{h}, \mathbf{h}') \alpha(\tau, \mathbf{h}') d\mathbf{h} d\mathbf{h}' \tag{2.50}$$

et

$$\mathbf{C} = \boldsymbol{\Gamma}_\alpha(\tau), \tag{2.51}$$

où $K(\mathbf{h}, \mathbf{h}') = \mathbb{E}_{\mathbf{x}, \theta_r; \theta_d} \left[v(\mathbf{x}, \boldsymbol{\theta}, \mathbf{h}) v^T(\mathbf{x}, \boldsymbol{\theta}, \mathbf{h}') \right]$, ce qui correspond à [TT09, eq (18) et (19)] dans le cas d'une unique fonction $\alpha(,)$. Par conséquent, la borne de Barankin hybride est compressible par transformation du noyau $\eta_\alpha(\mathbf{x}, \boldsymbol{\theta}, \tau)$.

2.5 Borne de Barankin-Weiss-Weinstein hybride

Parmi toutes les bornes bayésiennes [RFL+08, TT10b], la borne de Weiss-Weinstein [WW88] est connue pour être l'une des bornes les plus précises. Cependant, nous ne pouvons pas obtenir cette borne dans le contexte hybride car il n'existe pas de borne de Weiss-Weinstein déterministe dans le cas général. En effet, l'obtention de cette dernière aboutisse à des contraintes sur les estimateurs qui sont difficilement interprétables [Rendf, CRL10]. L'idée principale est donc de combiner de manière astucieuse la BBH et la borne de Weiss-Weinstein pour s'affranchir de cette problématique. A l'aide de l'inégalité de covariance, nous présentons ici une nouvelle famille de bornes combinant une généralisation de la BBH introduite dans [TT09] et la borne de Weiss-Weinstein.

2.5.1 Combinaison de la borne de Barankin et de la borne de Weiss-Weinstein

Soit $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_d^T \ \boldsymbol{\theta}_r^T \right]^T \in \Pi_d \times \Pi_r \subset \mathbb{R}^{D+R}$ un vecteur de paramètres à estimer à partir des observations $\mathbf{x} \in \boldsymbol{\Omega}$. Soient $\widehat{\boldsymbol{\theta}}(\mathbf{x})$ un estimateur de $\boldsymbol{\theta}$ et $\{\mathbf{h}_i\}_{i=1 \cdots N}$ un ensemble de points tests tels que les relations (R1), (R2) et (R3) sont vérifiées. Pour obtenir la borne de Barankin Weiss-Weinstein hybride (BBWWH), nous séparons le vecteur $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ en deux parties de tailles respectives $I \geq 1$ et $J \geq 1$, $I + J = N$ où

$$v_i(\mathbf{x}, \boldsymbol{\theta}) = \begin{cases} \frac{f_{\mathbf{x}, \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_i)}{f_{\mathbf{x}, \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta})} - 1, \ \boldsymbol{\theta} \in \boldsymbol{\Theta}, \\ 0, \text{ sinon}, \end{cases}, \ 1 \leq i \leq I \tag{2.52}$$

et

$$v_{I+j}(\mathbf{x},\boldsymbol{\theta}) = \begin{cases} \frac{f_{\mathbf{x},\boldsymbol{\theta}}^{m_j}(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_{I+j})}{f_{\mathbf{x},\boldsymbol{\theta}}^{m_j}(\mathbf{x},\boldsymbol{\theta})} - \frac{f_{\mathbf{x},\boldsymbol{\theta}}^{1-m_j}(\mathbf{x},\boldsymbol{\theta}-\mathbf{h}_{I+j})}{f_{\mathbf{x},\boldsymbol{\theta}}^{1-m_j}(\mathbf{x},\boldsymbol{\theta})}, \boldsymbol{\theta} \in \boldsymbol{\Theta}, \\ 0, \text{ sinon}, \end{cases}, \ 1 \leq j \leq J \tag{2.53}$$

où $v_n(\mathbf{x},\boldsymbol{\theta})$ désigne le n^{ième} élément du vecteur $\mathbf{v}(\mathbf{x},\boldsymbol{\theta})$, $\boldsymbol{\Theta}$ est défini par (2.32) et les puissances $\{m_j\}_{j=1\cdots J}$ sont strictement comprises [2] entre 0 et 1. Si nous restreignons les D premières composantes des vecteurs $\mathbf{h}_{I+j}$, $j = 1\cdots J$ à zéro, c'est-à-dire $\mathbf{h}_{I+j} = \left[\mathbf{0}^T \ \mathbf{h}_{r(I+j)}^T\right]^T$, alors la nouvelle borne hybride notée BBWWH est obtenue en appliquant l'inégalité de covariance (2.1). Son expression est donnée par

$$\mathbf{BBWWH} = \sup_{\substack{\mathbf{h}_1,\cdots,\mathbf{h}_I \in \Lambda_{\boldsymbol{\theta}} \\ \mathbf{h}_{I+1},\cdots,\mathbf{h}_N \in \Delta_{\boldsymbol{\theta}} \\ 0<m_{I+1},\cdots,m_N<1}} \left\{\mathbf{C}\mathbf{V}^{-1}\mathbf{C}^T\right\}, \tag{2.54}$$

où $\Lambda_{\boldsymbol{\theta}} = \{\mathbf{h}|\boldsymbol{\theta}+\mathbf{h} \in \Pi_d \times \Pi_r\}$ et $\Delta_{\boldsymbol{\theta}} = \left\{\mathbf{h} = \left[\mathbf{h}_d^T \ \mathbf{h}_r^T\right]^T | \mathbf{h}_d = \mathbf{0} \text{ et } \boldsymbol{\theta}_r + \mathbf{h}_r \in \Pi_r\right\}$ et

$$\mathbf{C} = [\mathbf{h}_1 \cdots \mathbf{h}_I \ \mu(m_{I+1},\mathbf{h}_{I+1})\,\mathbf{h}_{I+1} \cdots \mu(m_N,\mathbf{h}_N)\,\mathbf{h}_N], \tag{2.55}$$

avec

$$\mu(m,\mathbf{h}) = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r|\boldsymbol{\theta}_d}\left[\frac{f^m(\mathbf{x},\boldsymbol{\theta}+\mathbf{h})}{f^m(\mathbf{x},\boldsymbol{\theta})}\right]. \tag{2.56}$$

L'obtention de la matrice $\mathbf{C}$ est détaillée en annexe A.1. Et les éléments de la matrice $\mathbf{V}$ sont données par

- $\forall (i,i') \in [|1,I|]^2$,

$$V_{i,i'} = Q_{i,j} \text{ défini par l'équation (2.36)} \tag{2.57}$$

- $\forall (i,j) \in [|1;I|] \times [|I+1,N|]$,

$$\begin{aligned} V_{i,j} &= \xi(1,m_j,\mathbf{h}_i,\mathbf{h}_j) - \xi(1,1-m_j,\mathbf{h}_i,-\mathbf{h}_j) - \mu(m_j,\mathbf{h}_j) + \mu(1-m_j,-\mathbf{h}_j), \\ V_{j,i} &= V_{i,j}. \end{aligned} \tag{2.58}$$

- $\forall (j,j') \in [|I+1,N|]^2$,

$$\begin{aligned} V_{j,j'} &= \xi\left(m_j,m_{j'},\mathbf{h}_j,\mathbf{h}_{j'}\right) - \xi\left(1-m_j,m_{j'},-\mathbf{h}_j,\mathbf{h}_{j'}\right) \\ &\quad -\xi\left(m_j,1-m_{j'},\mathbf{h}_j,-\mathbf{h}_{j'}\right) + \xi\left(1-m_j,1-m_{j'},-\mathbf{h}_j,-\mathbf{h}_{j'}\right). \end{aligned} \tag{2.59}$$

avec

$$\xi(k,l,\mathbf{h}_a,\mathbf{h}_b) = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\frac{f_{\mathbf{x},\boldsymbol{\theta}}^k(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_a)\,f_{\mathbf{x},\boldsymbol{\theta}}^l(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_b)}{f_{\mathbf{x},\boldsymbol{\theta}}^{k+l}(\mathbf{x},\boldsymbol{\theta})}\right]. \tag{2.60}$$

Notons que l'obtention d'une expression analytique de la BBWWH nécessite uniquement la formulation analytique de $\xi(k,l,\mathbf{h}_a,\mathbf{h}_b)$ puisque $\mu(m,\mathbf{h}) = \xi(m,0,\mathbf{h},\mathbf{0})$. Nous verrons dans la section suivante (2.5.2) qu'il est possible de simplifier l'expression (2.60) dans le cas gaussien à moyenne paramétrée. Par ailleurs, la BBWWH (2.54) permet de déterminer une nouvelle famille de bornes basée sur la même transformation linéaire (discrète ou intégrale) que pour la borne de

2. En pratique, la valeur optimale de m_j, qui permet d'obtenir la BBWWH la plus précise, est souvent égale à $\frac{1}{2}$.

Barankin hybride (voir la section 2.4.4). De plus, nous pouvons retrouver certaines bornes de la littérature en jouant sur le nombre des points test I et J à savoir :
• Si $I > 0$, $J = 0$, $\boldsymbol{\theta} = \boldsymbol{\theta}_d$, nous retrouvons toutes les approximation de la borne de Barankin déterministe [TT10a, CGQL08, CRL09]
• Si $I > 0$, $J = 0$, $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_d^T \ \boldsymbol{\theta}_r^T\right]^T$, nous avons les bornes de Barankin hybrides comprimées par une transformation du noyau [TT09],
• Si $I > 0$, $J > 0$, $\boldsymbol{\theta} = \boldsymbol{\theta}_r$, nous retrouvons toute la famille de bornes de Weiss-Weinstein bayésiennes [TT10b, (38)],
• Si $I > 0$, $J > 0$, $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_d^T \ \boldsymbol{\theta}_r^T\right]^T$, nous obtenons l'extension de la famille de bornes de Weiss-Weinstein [TT10b, (38)] aux cas d'estimation hybride pour une classe d'estimateurs hybrides vérifiant (R1) (R2) et (R3) quel que soit le lien statistique entre les paramètres déterministes et bayésiens.

2.5.2 Modèle d'observation gaussien à moyenne paramétrée

Bien que l'évaluation numérique des bornes est moins lourde que celle de l'EQM des estimateurs, cela peut prendre du temps si nous ne réussissons pas simplifier les espérances (2.60). Les expressions analytiques de Slepian-Bangs [Sle54, Ban71] ont grandement simplifié la borne de Cramér-Rao. C'est pourquoi, nous proposons une analyse des expressions analytiques de la BBWWH dans le cas général du modèle d'observation gaussien à moyenne paramétrée :

$$\mathbf{x} = \mathbf{m}(\boldsymbol{\theta}) + \mathbf{b}, \tag{2.61}$$

où $\mathbf{x} \in \mathbb{C}^P$, $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_d^T \ \boldsymbol{\theta}_r^T\right]^T$ est le vecteur de paramètre à estimer composé de $\boldsymbol{\theta}_d \in \mathbb{R}^D$ un vecteur de paramètres déterministes et $\boldsymbol{\theta}_r \in \mathbb{R}^R$ un vecteur de paramètres aléatoires de densité de probabilité *a priori* connue. La fonction complexe $\mathbf{m}(.)$ est défini sur $\mathbb{R}^{D+R}$ et est à valeur dans $\mathbb{C}^P$. Le vecteur $\mathbf{b}$ est un bruit aléatoire gaussien complexe circulaire centrée et de matrice de covariance $\sigma_P^2\mathbf{I}$ que l'on note $\mathbf{b} \sim \mathbb{C}\mathcal{N}\left(\mathbf{0},\sigma_P^2\mathbf{I}\right)$. Les données observées sont également aléatoires de densité de probabilité conditionnelle $\mathbf{x}|\boldsymbol{\theta} \sim \mathbb{C}\mathcal{N}\left(\mathbf{m}(\boldsymbol{\theta}), \sigma_P^2\mathbf{I}\right)$. Ce modèle d'observation est largement utilisé dans une pléthore de problèmes d'estimation en traitement du signal tel que : l'analyse spectrale [SM05], le traitement d'antenne [Van02], les communications numériques [PS08], etc.
L'objectif de cette section est donc d'analyser les conditions requises pour obtenir une expression analytique des matrices $\mathbf{V}$ et $\mathbf{C}$. Pour cela, une condition suffisante est de trouver une expression analytique de $\xi(k, l, \mathbf{h}_a, \mathbf{h}_b)$ dont la forme générale est donnée par (2.60). Si nécessaire, nous décomposons de la manière suivante $\mathbf{h}_a = \left[\mathbf{h}_{da}^T \ \mathbf{h}_{ra}^T\right]^T$ et $\mathbf{h}_b = \left[\mathbf{h}_{db}^T \ \mathbf{h}_{rb}^T\right]^T$ où les vecteurs $\mathbf{h}_{da}$ et $\mathbf{h}_{db}$ sont des points test pour les paramètres déterministes et $\mathbf{h}_{ra}$ et $\mathbf{h}_{rb}$ sont les points test pour les paramètres aléatoires. Pour simplifier les expressions, notons les variables $\boldsymbol{\theta} \triangleq \boldsymbol{\theta}_d; \boldsymbol{\theta}_r$ et $\mathbf{x}|\boldsymbol{\theta} \triangleq \mathbf{x}|\boldsymbol{\theta}_r; \boldsymbol{\theta}_d$. Nous pouvons exprimer $\xi(k, l, \mathbf{h}_a, \mathbf{h}_b)$ par la relation suivante

$$\xi(k, l, \mathbf{h}_a, \mathbf{h}_b) = \mathbb{E}_{\boldsymbol{\theta}}\left[A(\boldsymbol{\theta}, \mathbf{h}_a, \mathbf{h}_b)\frac{f_{\boldsymbol{\theta}}^k(\boldsymbol{\theta} + \mathbf{h}_a) f_{\boldsymbol{\theta}}^l(\boldsymbol{\theta} + \mathbf{h}_b)}{f_{\boldsymbol{\theta}}^{k+l}(\boldsymbol{\theta})}\right], \tag{2.62}$$

avec

$$A(\boldsymbol{\theta}, \mathbf{h}_a, \mathbf{h}_b) = \mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}}\left[\frac{f_{\mathbf{x}|\boldsymbol{\theta}}^k(\mathbf{x}|\boldsymbol{\theta} + \mathbf{h}_a) f_{\mathbf{x}|\boldsymbol{\theta}}^l(\mathbf{x}|\boldsymbol{\theta} + \mathbf{h}_b)}{f_{\mathbf{x}|\boldsymbol{\theta}}^{k+l}(\mathbf{x}|\boldsymbol{\theta})}\right]. \tag{2.63}$$

Nous allons calculer $\xi(k, l, \mathbf{h}_a, \mathbf{h}_b)$ en deux étapes : la première étape est de déterminer une expression analytique de $A(\boldsymbol{\theta}, \mathbf{h}_a, \mathbf{h}_b)$ pour le modèle d'observation gaussien. La seconde étape consiste à analyser les scénarios pour lesquels $\xi(k, l, \mathbf{h}_a, \mathbf{h}_b)$ possède une forme analytique.

2.5.2.1 Analyse de $\xi(k,l,\mathbf{h}_a,\mathbf{h}_b)$

D'après les résultats de [VRBM14, (15)], l'expression de $A(\boldsymbol{\theta},\mathbf{h}_a,\mathbf{h}_b)$ vérifie

$$\begin{aligned}\sigma_n^2 \ln A(\boldsymbol{\theta},\mathbf{h}_a,\mathbf{h}_b) &= \|k\mathbf{m}(\boldsymbol{\theta}+\mathbf{h}_a)+l\mathbf{m}(\boldsymbol{\theta}+\mathbf{h}_b)-(k+l-1)\mathbf{m}(\boldsymbol{\theta})\|^2 \\ &\quad -k\|\mathbf{m}(\boldsymbol{\theta}+\mathbf{h}_a)\|^2 - l\|\mathbf{m}(\boldsymbol{\theta}+\mathbf{h}_b)\|^2 + (k+l-1)\|\mathbf{m}(\boldsymbol{\theta})\|^2 .\end{aligned} \tag{2.64}$$

Nous remarquons que $A(\boldsymbol{\theta},\mathbf{h}_a,\mathbf{h}_b)$ dépend en général de $\boldsymbol{\theta}$ conduisant à une expression de $\mathbb{E}_{\boldsymbol{\theta}}\left[A(\boldsymbol{\theta},\mathbf{h}_a,\mathbf{h}_b)\frac{f^k(\boldsymbol{\theta}+\mathbf{h}_a)f^l(\boldsymbol{\theta}+\mathbf{h}_b)}{f^{k+l}(\boldsymbol{\theta})}\right]$ difficile à simplifier analytiquement. Cependant, si $A(\boldsymbol{\theta},\mathbf{h}_a,\mathbf{h}_b)$ est indépendante de $\boldsymbol{\theta}_r$, un cas largement rencontré en traitement du signal ([VB07] et [VRBM14]), une forme analytique de $\xi(k,l,\mathbf{h}_a,\mathbf{h}_b)$ est obtenue lorsque la loi *a priori* est gaussienne de type $\mathcal{N}\left(\boldsymbol{\theta}_{r0},\sigma_{\theta_r}^2\mathbf{I}\right)$. Dans ce cas, nous avons $A(\boldsymbol{\theta},\mathbf{h}_a,\mathbf{h}_b)$ indépendante de $\boldsymbol{\theta}_r$, c'est-à-dire $A(\boldsymbol{\theta},\mathbf{h}_a,\mathbf{h}_b) = A(\boldsymbol{\theta}_d,\mathbf{h}_a,\mathbf{h}_b)\ \forall\boldsymbol{\theta}_d\in\Pi_d$, et l'expression de $\xi(k,l,\mathbf{h}_a,\mathbf{h}_b)$ est alors donnée par

$$\xi(k,l,\mathbf{h}_a,\mathbf{h}_b) = A(\boldsymbol{\theta}_d,\mathbf{h}_a,\mathbf{h}_b)\, e^{-\frac{1}{2\sigma_{\theta_r}^2}\left[k\|\mathbf{h}_{ar}\|^2+l\|\mathbf{h}_{br}\|^2-\|k\mathbf{h}_{ar}+l\mathbf{h}_{br}\|^2\right]} . \tag{2.65}$$

2.5.2.2 Indépendance de $A(\boldsymbol{\theta},\mathbf{h}_a,\mathbf{h}_b)$ par rapport à $\boldsymbol{\theta}_r$

Notons qu'une condition suffisante pour obtenir $A(\boldsymbol{\theta},\mathbf{h}_a,\mathbf{h}_b)$ indépendant de $\boldsymbol{\theta}_r$ est la suivante : si $\forall\mathbf{h}$, $\exists\mathbf{P}$ une matrice carrée et $\mathbf{q}$ un vecteur tel que :

$$\mathbf{m}(\boldsymbol{\theta}+\mathbf{h}) = \mathbf{P}(\boldsymbol{\theta}_r)\,\mathbf{q}(\boldsymbol{\theta}_d,\mathbf{h}) \ \text{ et } \ \mathbf{P}^H(\boldsymbol{\theta}_r)\,\mathbf{P}(\boldsymbol{\theta}_r) = \mathbf{S}, \tag{2.66}$$

où $\mathbf{S}$ est une matrice carrée indépendant de $\boldsymbol{\theta}_r$, alors
- $\|\mathbf{m}(\boldsymbol{\theta}+\mathbf{h})\|^2 = \mathbf{q}^H(\boldsymbol{\theta}_d,\mathbf{h})\,\mathbf{S}\mathbf{q}(\boldsymbol{\theta}_d,\mathbf{h})$, ne dépend pas de $\boldsymbol{\theta}_r$, $\forall\mathbf{h}$,
- $\|k\mathbf{m}(\boldsymbol{\theta}+\mathbf{h}_a)+l\mathbf{m}(\boldsymbol{\theta}+\mathbf{h}_b)-(k+l-1)\mathbf{m}(\boldsymbol{\theta})\|^2$ ne dépend pas de $\boldsymbol{\theta}_r$, $\forall k,l,\mathbf{h}_a,\mathbf{h}_b$.

En effet, à l'aide de l'hypothèse (2.66), $A(\boldsymbol{\theta}_d,\mathbf{h}_a,\mathbf{h}_b)$ vérifie la relation suivante :

$$\begin{aligned}\sigma_n^2\ln A(\boldsymbol{\theta}_d,\mathbf{h}_a,\mathbf{h}_b) &= k(k-1)\,\mathbf{q}^H(\boldsymbol{\theta}_d,\mathbf{h}_a)\,\mathbf{S}\mathbf{q}(\boldsymbol{\theta}_d,\mathbf{h}_a) + l(l-1)\,\mathbf{q}^H(\boldsymbol{\theta}_d,\mathbf{h}_b)\,\mathbf{S}\mathbf{q}(\boldsymbol{\theta}_d,\mathbf{h}_b) \\ &\quad + (k+l)(k+l-1)\,\mathbf{q}^H(\boldsymbol{\theta}_d,\mathbf{0})\,\mathbf{S}\mathbf{q}(\boldsymbol{\theta}_d,\mathbf{0}) + 2kl\,\mathrm{Re}\left(\mathbf{q}^H(\boldsymbol{\theta}_d,\mathbf{h}_b)\,\mathbf{S}\mathbf{q}(\boldsymbol{\theta}_d,\mathbf{h}_a)\right) \\ &\quad -2(k+l-1)\,\mathrm{Re}\left(\left(k\mathbf{q}^H(\boldsymbol{\theta}_d,\mathbf{h}_a)+l\mathbf{q}^H(\boldsymbol{\theta}_d,\mathbf{h}_b)\right)\mathbf{S}\mathbf{q}(\boldsymbol{\theta}_d,\mathbf{0})\right).\end{aligned}$$

2.5.3 Application et simulation

2.5.3.1 Analyse spectrale

Nous considérons ici le problème d'estimation de fréquence abordé dans [TT09]. Il s'agit d'un problème de référence dans l'étude du phénomène de décrochement. Le modèle d'observation est le suivant :

$$\mathbf{x} = se^{j\varphi}\mathbf{c}(\omega) + \mathbf{b}, \tag{2.67}$$

où $\mathbf{x}\in\mathbb{C}^P$ est le vecteur d'observation, $\mathbf{c}(\omega) = \left[1\ e^{j\omega}\ \cdots\ e^{j(P-1)\omega}\right]^T$ est une cisoïde normalisée et le vecteur bruit $\mathbf{b}$ est supposé aléatoire gaussien complexe circulaire centré et de matrice de covariance $\sigma_n^2\mathbf{I}$. La phase $\varphi\in\,]-\pi;\pi]$ et l'amplitude $s\in\mathbb{R}^+$ sont des paramètres supposés inconnus déterministes et la pulsation normalisée ω est un paramètre inconnu aléatoire de densité de probabilité *a priori* gaussienne centrée de matrice de covariance σ_ω^2. Par conséquent le vecteur de paramètres à estimer est $\boldsymbol{\theta} = [s\ \varphi\ \omega]^T$. Pour évaluer la nouvelle borne BBWWH, il suffit d'obtenir une expression analytique de $\xi(k,l,\mathbf{h}_a,\mathbf{h}_b)$ donnée par la relation (2.60). Puisque les

observations sont distribuées suivant une loi gaussienne, il suffit de reprendre le modèle donné dans la section (2.5.2) et de poser $\mathbf{m}(\boldsymbol{\theta}) = se^{j\varphi}\mathbf{c}(\omega)$. En effet, notons que pour tout $\mathbf{h} = [h_s \; h_\varphi \; h_\omega]^T$,

$$\mathbf{m}(\boldsymbol{\theta}+\mathbf{h}) = Diag(\mathbf{c}(\omega))(s+h_s)\, e^{j(\varphi+h_\varphi)}\mathbf{c}(h_\omega), \tag{2.68}$$

satisfait la condition (2.66) où $Diag(\mathbf{c}(\omega))$ est une matrice carrée dont les éléments diagonaux sont les éléments du vecteur $\mathbf{c}(\omega)$ et les autres éléments sont nuls. Par conséquent, nous pouvons appliquer la relation (2.64) et nous obtenons pour tout $\mathbf{h}_a = [h_{as} \; h_{a\varphi} \; h_{a\omega}]^T$ et $\mathbf{h}_b = [h_{bs} \; h_{b\varphi} \; h_{b\omega}]^T$

$$\begin{aligned} A(\boldsymbol{\theta}_d, \mathbf{h}_a, \mathbf{h}_b) &= e^{\frac{1}{\sigma_n^2}P\left(k(k-1)(s+h_{as})^2+l(l-1)(s+h_{bs})^2+(k+l)(k+l-1)s^2\right)} \\ &\times e^{\frac{2}{\sigma_n^2}\left[kl(s+h_{as})(s+h_{bs})T\left(h_{a\omega}-h_{b\omega},h_{a\varphi}-h_{b\varphi}\right)-(k+l-1)s\left(k(s+h_{as})T(h_{a\omega},h_{a\varphi})+l(s+h_{bs})T\left(h_{b\omega},h_{b\varphi}\right)\right)\right]}, \end{aligned}$$

avec $T(h_\omega, h_\varphi) = \cos\left(\frac{(P-1)}{2}h_\omega + h_\varphi\right)\frac{\sin\left(\frac{P}{2}h_\omega\right)}{\sin\left(\frac{h_\omega}{2}\right)}$.

Finalement, l'expression $\xi(k, l, \mathbf{h}_a, \mathbf{h}_b)$ est donnée par :

$$\xi(k, l, \mathbf{h}_a, \mathbf{h}_b) = A(\boldsymbol{\theta}_d, \mathbf{h}_a, \mathbf{h}_b)\, e^{\frac{1}{2\sigma_\omega^2}\left[k(k-1)h_{a\omega}^2+l(l-1)h_{b\omega}^2+2klh_{a\omega}h_{b\omega}\right]}. \tag{2.69}$$

En conséquence, une expression analytique de la BBWWH est obtenue en substituant (2.69) dans l'expression (2.54).

2.5.3.2 Simulation

Nous allons comparer numériquement la BBWWH avec la BCRH, la BBH, la CCLRB et l'EQM du MVMAP. Les conditions de simulations sont les suivantes : $s = 1$, $\varphi = \frac{\pi}{4}$, $\sigma_\omega^2 = \frac{1}{2}$ et $N = 2^5$. Les expressions de la BCRH, la BBH, la CCLRB sont données dans [TT09, section 5]. La BBH et la BBWWH, sont calculées numériquement en maximisant les points tests $\mathbf{h}_1 \in [-1;1] \times \{0\} \times \{0\}$, $\mathbf{h}_2 \in \{0\} \times]-\pi;\pi] \times \{0\}$ et $\mathbf{h}_3 \in \{0\} \times \{0\} \times]-\pi;\pi]$ sur leur ensemble de définition respectif avec un pas de discrétisation de $\delta h_s = 0.01$ pour la première composante, de $\delta h_\varphi = \frac{2\pi}{2^8}$ pour la seconde composante et de $\delta h_\omega = \frac{\pi}{2^8}$ pour la dernière composante. La CCLRB est obtenue numériquement avec 2^{10} points tests choisis suivant la même configuration proposée par [TT09]. Finalement, l'estimateur MVMAP consiste à chercher les valeurs optimales de $s \in [0;2]$, $\varphi \in]-\pi;\pi]$ et $\omega \in]-\pi;\pi]$ qui maximisent la densité de probabilité jointe

$$f_{\mathbf{x},\omega|s,\varphi}(\mathbf{x}, \omega|s, \varphi) = \frac{e^{-\frac{1}{\sigma_n^2}\left\|\mathbf{x}-se^{j\varphi}\mathbf{b}(\omega)\right\|^2-\frac{\omega^2}{2\sigma_\omega^2}}}{(\pi\sigma_n^2)^P\sqrt{2\pi\sigma_\omega^2}}, \tag{2.70}$$

c'est à dire

$$\widehat{\boldsymbol{\theta}}_{MVMAP} = \arg\max_{s,\varphi,\omega} f_{\mathbf{x},\omega|s,\varphi}(\mathbf{x}, \omega|s, \varphi). \tag{2.71}$$

L'EQM empirique de cet estimateur est obtenue avec 1000 tirages de Monte-Carlo.

Nous superposons sur la figure (2.1) la BCRH, la BBH, la CCRLB, la BBWWH et l'EQM empirique du MVMAP pour l'estimation du paramètre aléatoire ω seulement car il est connu qu'un phénomène de décrochement apparaît pour l'estimation de ce paramètre. Nous remarquons d'après cette figure que toutes les bornes prédisent bien le comportement asymptotique à fort RSB. Cependant la BBWWH proposée est plus efficace à prédire le seuil de décrochement par rapport aux bornes existantes.

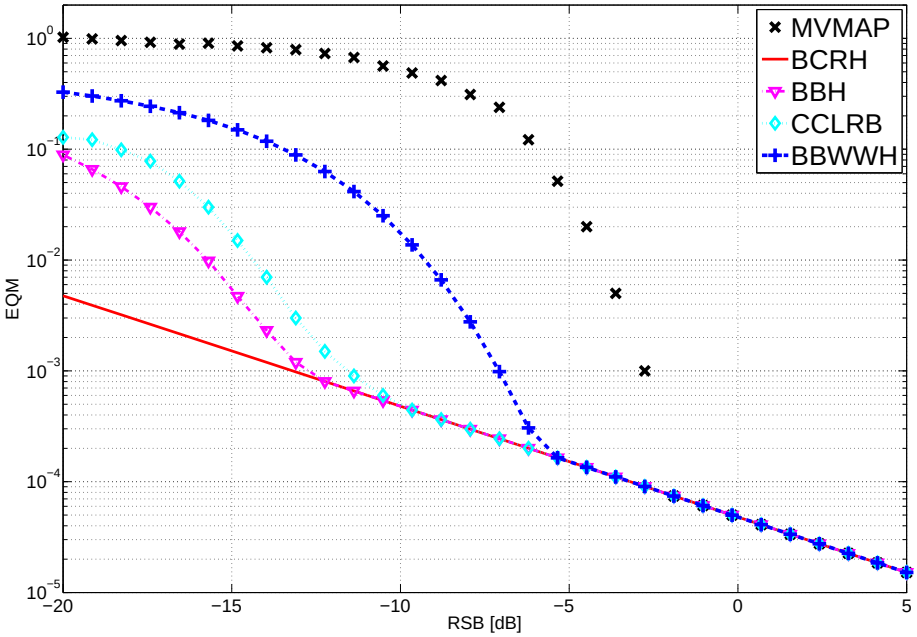

FIGURE 2.1 – Comparaison des bornes hybrides et de l'EQM du MVMAP en fonction du RSB.

2.6 Borne de Barankin-Ziv-Zakaï hybride

A travers l'obtention de la borne de Barankin hybride [RM97] et la borne de Barankin Weiss-Weinstein hybride [RGC+13], nous avons illustré le principe d'une combinaison astucieuse des bornes déterministes et bayésiennes. Dans le contexte des bornes bayésiennes (estimation de paramètres aléatoires seulement), deux grandes familles de bornes sont connues pour la qualité de leur prédiction de l'EQM optimale, à savoir l'EQM de l'estimateur moyenne *a posteriori* : la famille de bornes de Weiss-Weinstein et la famille de bornes de Ziv-Zakaï. La première famille est basée sur l'inégalité de covariance conduisant à une extension naturelle de ces bornes dans le contexte hybride [RGC+13] (section précedente). La seconde famille [ZZ69, CZZ75, BSEV97] est basée sur un problème de test d'hypothèses qui n'est pas extensible, à notre connaissance, aux cas d'estimation hybride. En effet, la comparaison entre ces deux familles de bornes est souvent faite par des simulations et il n'existe pas de relation générale entre ces deux familles sauf pour une forme particulière de la borne de Ziv-Zakaï, mentionnée dans [Bel95, chap 4], qui est obtenable à la fois par l'inégalité de covariance et par l'inégalité de Kotelnikov [Kot59] (la base des bornes de Ziv-Zakaï). Cette relation nous servira à établir une borne hybride de type Ziv-Zakaï.

2.6.1 La famille de bornes de Ziv-Zakaï

Soient $\mathbf{x} \in \Omega$ les observations, $\theta \in \mathbb{R}$ le paramètre à estimer et $\widehat{\theta}(\mathbf{x})$ un estimateur de θ de loi *a priori* f_θ. On a l'égalité suivante :

$$\mathbb{E}_{\mathbf{x},\theta}\left[\left(\widehat{\theta}-\theta\right)^2\right] = \int_0^\infty \frac{\Delta}{2}\Pr\left(\left|\widehat{\theta}-\theta\right| \geq \frac{\Delta}{2}\right) d\Delta, \tag{2.72}$$

où $\Pr\left(\left|\widehat{\theta}-\theta\right| \geq \frac{\Delta}{2}\right)$ désigne la probabilité que l'erreur absolue soit supérieure à $\frac{\Delta}{2}$. L'idée est de minorer la quantité $\Pr\left(\left|\widehat{\theta}-\theta\right| \geq \frac{\Delta}{2}\right)$ pour obtenir une borne. En utilisant l'inégalité de Koteknikov [Kot59], l'EQM est bornée par

$$\mathbb{E}_{\mathbf{x},\theta}\left[\left(\widehat{\theta}-\theta\right)^2\right] \geq \int_0^\infty \frac{\Delta}{2}\int_{-\infty}^{+\infty}\left(f(\theta)+f(\theta+\Delta)\right)P_{\min}(\theta,\theta+\Delta)\,d\theta d\Delta \tag{2.73}$$

où $P_{\min}(\theta,\theta+\Delta)$ est la probabilité d'erreur minimale pour un problème de détection binaire suivant :

$$\begin{cases} H_0 : \theta; \Pr(H_0) = \frac{f(\theta)}{f(\theta)+f(\theta+\Delta)}; \mathbf{x} \sim f(\mathbf{x}|\,\theta) \\ H_1 : \theta+\Delta; \Pr(H_1) = \frac{f(\theta+\Delta)}{f(\theta)+f(\theta+\Delta)}; \mathbf{x} \sim f(\mathbf{x}|\,\theta+\Delta) \end{cases}.$$

Le membre de droite de (2.73) est la borne de Bellini-Tartara étendue [BT74,BESV]. Cependant elle présente un coût de calcul élevé à cause de l'intégrale par rapport à Δ et à l'évaluation de $P_{\min}$. Il existe des méthodes d'encadrement de $P_{\min}$ qui sont exposés dans la thèse de Kristine Bell [Bel95]. Pour s'affranchir de la double intégrale, elle a proposée une version simplifiée moins coûteuse en temps de calcul

$$\mathbb{E}_{\mathbf{x},\theta}\left[\left(\widehat{\theta}-\theta\right)^2\right] \geq \max_{\Delta}\frac{\Delta^2}{2}\int_{-\infty}^{+\infty}\left(f(\theta)+f(\theta+\Delta)\right)P_{\min}(\theta,\theta+\Delta)\,d\theta. \tag{2.74}$$

Bien que cette dernière (2.74) soit moins précise que la borne (2.73), elle est néanmoins unifiable par l'inégalité de covariance dans le cas scalaire

$$\mathbb{E}_{\mathbf{x},\theta}\left[\left(\widehat{\theta}-\theta\right)^2\right] \geq \frac{\mathbb{E}^2_{\mathbf{x},\theta}\left[\left(\widehat{\theta}-\theta\right)v(\mathbf{x},\theta)\right]}{\mathbb{E}_{\mathbf{x},\theta}\left[v^2(\mathbf{x},\theta)\right]}, \tag{2.75}$$

avec le choix de $v(\mathbf{x},\theta)$ défini par

$$v(\mathbf{x},\theta) = \min_{\Delta}\left(1, \frac{f(\mathbf{x},\theta+\Delta)}{f(\mathbf{x},\theta)}\right) - \min_{\Delta}\left(1, \frac{f(\mathbf{x},\theta-\Delta)}{f(\mathbf{x},\theta)}\right). \tag{2.76}$$

La démonstration de cette relation est exposée dans [Bel95, chapitre 4]. Notre but est donc étendre cette borne dans le contexte de l'estimation hybride.

2.6.2 La borne proposée

Soit $\mathbf{x} \in \Omega$ un point d'observation et soit $\boldsymbol{\theta} = [\theta_d\ \theta_r]^T$ un vecteur de paramètres hybrides composé d'un paramètre déterministe et d'un paramètre aléatoire de densité de probabilité jointe $f(\mathbf{x},\boldsymbol{\theta}) = f(\mathbf{x},\theta_r;\theta_d)$. Soit un estimateur $\widehat{\boldsymbol{\theta}} = \left[\widehat{\theta}_d\ \widehat{\theta}_r\right]^T$, et des points tests $[h_d\ h_r]$ tels que

$\theta_d + h_d \in \Pi_d$ et $\theta_r + h_r \in \Pi_r$ satisfaisant les hypothèses (R1), (R2) et (R3) mentionnées à la section (2.4.3). De même que pour la BBWWH, nous séparons en deux le vecteur $\mathbf{v} = [v_d \; v_r]^T$ où

$$v_d = \begin{cases} \frac{f(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_1)}{f(\mathbf{x},\boldsymbol{\theta})} - 1 \text{ si } \boldsymbol{\theta} \in \Theta \\ 0 \text{ sinon} \end{cases} \tag{2.77}$$

et

$$v_r = \begin{cases} \min\limits_{\mathbf{h}_2} \left(\frac{f(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_2)}{f(\mathbf{x},\boldsymbol{\theta})}, 1 \right) - \min\limits_{\mathbf{h}_2} \left(\frac{f(\mathbf{x},\boldsymbol{\theta}-\mathbf{h}_2)}{f(\mathbf{x},\boldsymbol{\theta})}, 1 \right) \text{ si } \boldsymbol{\theta} \in \Theta \\ 0 \text{ sinon} \end{cases}, \tag{2.78}$$

pour tout $\mathbf{h}_1 = [h_{1d} \; h_{1r}]^T$ et $\mathbf{h}_2 = [0 \; h_{2r}]^T$. Alors en appliquant l'inégalité de covariance (2.1), nous obtenons la borne suivante

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}} \left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right)^T \right] \succeq \mathbf{C}\mathbf{V}^{-1}\mathbf{C}^T, \tag{2.79}$$

où la matrice $\mathbf{C}$ est donnée par (voir le détail des calculs en annexe A.2)

$$\mathbf{C} = \begin{pmatrix} h_{1d} & 0 \\ h_{1r} & h_{2r}\alpha\left(\mathbf{0}, \mathbf{h}_2\right) \end{pmatrix}, \tag{2.80}$$

et les éléments de la matrice $\mathbf{V}$ s'expriment de la manière suivante :

$$\{\mathbf{V}\}_{1,1} = \mu\left(\mathbf{h}_1\right) - 1, \tag{2.81}$$
$$\{\mathbf{V}\}_{2,2} = \beta\left(\mathbf{h}_2, \mathbf{h}_2\right) + \beta\left(-\mathbf{h}_2, -\mathbf{h}_2\right) - 2\beta\left(\mathbf{h}_2, -\mathbf{h}_2\right), \tag{2.82}$$
$$\{\mathbf{V}\}_{1,2} = \{\mathbf{V}\}_{2,1} = \alpha\left(\mathbf{h}_1, \mathbf{h}_2\right) - \alpha\left(\mathbf{h}_1, -\mathbf{h}_2\right). \tag{2.83}$$

Les fonctions $\mu(.)$, $\alpha(.,.)$ et $\beta(.,.)$ sont définies par

$$\mu\left(\mathbf{h}\right) = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}} \left[\frac{f^2\left(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}\right)}{f^2\left(\mathbf{x}, \boldsymbol{\theta}\right)} \right], \tag{2.84}$$

$$\alpha\left(\mathbf{h}_1, \mathbf{h}_2\right) = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}} \left[\frac{f\left(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_1\right)}{f\left(\mathbf{x}, \boldsymbol{\theta}\right)} \min\left(\frac{f\left(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_2\right)}{f\left(\mathbf{x}, \boldsymbol{\theta}\right)}, 1 \right) \right] \tag{2.85}$$

et

$$\beta\left(\mathbf{h}_1, \mathbf{h}_2\right) = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}} \left[\min\left(\frac{f\left(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_1\right)}{f\left(\mathbf{x}, \boldsymbol{\theta}\right)}, 1 \right) \min\left(\frac{f\left(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_2\right)}{f\left(\mathbf{x}, \boldsymbol{\theta}\right)}, 1 \right) \right]. \tag{2.86}$$

Nous remarquons que si $\boldsymbol{\theta} = \theta_r$ (c'est-à-dire lorsque θ_d est connue), alors nous retrouvons la borne donnée par l'expression (2.74).

2.6.3 Expression analytique pour le modèle gaussien à moyenne paramétrée

Considérons le modèle d'observation suivant :

$$\mathbf{x} = \mathbf{m}\left(\theta_d, \theta_r\right) + \mathbf{b}, \tag{2.87}$$

où $\mathbf{x} \in \mathbb{C}^N$ est un vecteur de taille N, $\mathbf{m}(.,.)$ est une fonction définie sur $\mathbb{R}^2$ à valeur dans $\mathbb{C}^N$ modélisant le processus de mesure et $\mathbf{b}$ est le bruit de mesure supposé aléatoire de densité de probabilité gaussienne complexe circulaire centrée et de matrice de covariance $\sigma_b^2\mathbf{I}_N$. Nous

faisons l'hypothèse que la densité de probabilité *a priori* de θ_r est également gaussienne centrée de variance $\sigma^2_{\theta_r}$ et indépendante du bruit $\mathbf{b}$.

En pratique, la difficulté majeure pour programmer cette borne est l'évaluation des expressions de $\alpha(.,.)$ (2.85) et $\beta(.,.)$ (2.86). Concernant $\mu(.)$ (2.84), il s'agit d'une expression largement étudiée dans [RFL+08] qui est donnée par :

$$\mu(\mathbf{h}) = \mathbb{E}_{\theta_r}\left(e^{\frac{2}{\sigma_b^2}\|\mathbf{m}(\boldsymbol{\theta}+\mathbf{h})-\mathbf{m}(\boldsymbol{\theta})\|^2}\right). \tag{2.88}$$

Nous allons d'abord étudier $\beta(.,.)$ car son expression découle directement d'un résultat publié récemment [Org12].

2.6.3.1 Expression de $\beta(.,.)$

Pour tout $\boldsymbol{\delta}_1 = [0\ \delta_1]^T$ et $\boldsymbol{\delta}_2 = [0\ \delta_2]^T$ avec $\delta_1 \neq \delta_2 \in \mathbb{R}$, $\beta(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$ est similaire à l'expression (2) de [Org12], donnée par

$$\beta(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = \mathbb{E}_{\theta_r}\left(I_{\beta_1} + I_{\beta_2} + I_{\beta_3} + I_{\beta_4}\right), \tag{2.89}$$

avec

$$I_{\beta_1} = F_{\mathcal{N}\left(\mathbf{0}, \frac{\sigma_b^2}{2}\boldsymbol{\Gamma}\right)}\left(-\frac{\sigma_b^2 b(\delta_1)}{2}, -\frac{\sigma_b^2 b(\delta_2)}{2}\right), \tag{2.90}$$

$$I_{\beta_2} = e^{-\frac{\delta_2^2+2\theta_r\delta_2}{2\sigma_{\theta_r}^2}} F_{\mathcal{N}\left(\mathbf{m}_{\beta_2}, \frac{\sigma_b^2}{2}\boldsymbol{\Gamma}'\right)}\left(-\frac{\sigma_b^2 b(\delta_1)}{2}, \frac{\sigma_b^2 b(\delta_2)}{2}\right), \tag{2.91}$$

$$I_{\beta_3} = e^{-\frac{\delta_1^2+2\theta_r\delta_1}{2\sigma_{\theta_r}^2}} F_{\mathcal{N}\left(\mathbf{m}_{\beta_3}, \frac{\sigma_b^2}{2}\boldsymbol{\Gamma}'\right)}\left(\frac{\sigma_b^2 b(\delta_1)}{2}, -\frac{\sigma_b^2 b(\delta_2)}{2}\right), \tag{2.92}$$

et

$$I_{\beta_4} = e^{\frac{2\,\mathrm{Re}\left(\mathbf{d}^H(\delta_1)\mathbf{d}(\delta_2)\right)}{\sigma_b^2} - \frac{\delta_1^2+\delta_2^2+2\theta_r(\delta_1+\delta_2)}{2\sigma_{\theta_r}^2}} F_{\mathcal{N}\left(\mathbf{m}_{\beta_4}, \frac{\sigma_b^2}{2}\boldsymbol{\Gamma}\right)}\left(\frac{\sigma_n^2 b(\delta_1)}{2}, \frac{\sigma_n^2 b(\delta_2)}{2}\right).$$

où $F_{\mathcal{N}(\mathbf{m},\boldsymbol{\Sigma})}(\mathbf{a})$ est la fonction de répartition d'un vecteur gaussien de moyenne $\mathbf{m}$ et de matrice de covariance $\boldsymbol{\Sigma}$ évaluée au point $\mathbf{a}$ et

$$\begin{aligned}
\mathbf{d}(\delta_1) &= \mathbf{g}(\theta_d, \theta_r + \delta_1) - \mathbf{g}(\theta_d, \theta_r), && (2.93)\\
b(\delta_1) &= \frac{1}{\sigma_b^2}\|\mathbf{d}(\delta_1)\|^2 + \frac{\delta_1^2 + 2\theta_r\delta_1}{2\sigma_{\theta_r}^2}, && (2.94)\\
\mathbf{m}_{\beta_2} &= \left(-\mathrm{Re}\left(\mathbf{d}^H(\delta_1)\,\mathbf{d}(\delta_2)\right)\quad \|\mathbf{d}(\delta_2)\|^2\right)^T, && (2.95)\\
\mathbf{m}_{\beta_3} &= \left(\|\mathbf{d}(\delta_1)\|^2\quad -\mathrm{Re}\left(\mathbf{d}^H(\delta_1)\,\mathbf{d}(\delta_2)\right)\right)^T, && (2.96)\\
\mathbf{m}_{\beta_4} &= \begin{pmatrix} \|\mathbf{d}(\delta_1)\|^2 + \mathrm{Re}\left(\mathbf{d}^H(\delta_1)\,\mathbf{d}(\delta_2)\right) \\ \|\mathbf{d}(\delta_2)\|^2 + \mathrm{Re}\left(\mathbf{d}^H(\delta_1)\,\mathbf{d}(\delta_2)\right) \end{pmatrix}, && (2.97)\\
\boldsymbol{\Gamma} &= \begin{pmatrix} \|\mathbf{d}(\delta_1)\|^2 & \mathrm{Re}\left(\mathbf{d}^H(\delta_1)\,\mathbf{d}(\delta_2)\right) \\ \mathrm{Re}\left(\mathbf{d}^H(\delta_1)\,\mathbf{d}(\delta_2)\right) & \|\mathbf{d}(\delta_2)\|^2 \end{pmatrix}, && (2.98)\\
\boldsymbol{\Gamma}' &= \begin{pmatrix} \|\mathbf{d}(\delta_1)\|^2 & -\mathrm{Re}\left(\mathbf{d}^H(\delta_1)\,\mathbf{d}(\delta_2)\right) \\ -\mathrm{Re}\left(\mathbf{d}^H(\delta_1)\,\mathbf{d}(\delta_2)\right) & \|\mathbf{d}(\delta_2)\|^2 \end{pmatrix}. && (2.99)
\end{aligned}$$

Notons que si $\delta_1 = \delta_2$ alors les matrices $\mathbf{\Gamma}$ et $\mathbf{\Gamma}'$ devient singulières car une des valeurs propres de $\mathbf{\Gamma}$ et de $\mathbf{\Gamma}'$ est nulle. Néanmoins, il existe bien une expression analytique de cette limite [Org12]

$$I_{\beta_1} = F_{\mathcal{N}\left(0,\frac{\sigma_b^2\|\mathbf{d}(\delta_1)\|^2}{2}\right)}\left(-\frac{\sigma_n^2 b\left(\delta_1\right)}{2}\right), \tag{2.100}$$

$$I_{\beta_2} = I_{\beta_3} = 0, \tag{2.101}$$

$$\text{et } I_{\beta_4} = e^{\frac{2\|\mathbf{d}(\delta_1)\|^2}{\sigma_b^2} - \frac{\delta_1^2 + 2\theta_r\delta_1}{\sigma_{\theta_r}^2}} F_{\mathcal{N}\left(2\|\mathbf{d}(\delta_1)\|^2,\frac{\sigma_b^2\|\mathbf{d}(\delta_1)\|^2}{2}\right)}\left(\frac{\sigma_b^2 b\left(\delta_1\right)}{2}\right). \tag{2.102}$$

2.6.3.2 Expression de $\alpha\left(.,.\right)$

Afin d'obtenir une forme de $\alpha\left(.,.\right)$, on peut séparer le domaine d'intégration en deux parties dans lesquelles l'opérateur min peut être substitué par $\frac{f(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_2)}{f(\mathbf{x},\boldsymbol{\theta})}$ ou 1. Nous allons donc séparer $\alpha\left(\mathbf{h}_1,\mathbf{h}_2\right)$ en la somme de deux termes, pour tout $\mathbf{h}_1 = [h_{1d}\ h_{1r}]^T$ et $\mathbf{h}_2 = [0\ h_{2r}]^T$ nous avons :

$$\alpha\left(\mathbf{h}_1,\mathbf{h}_2\right) = \mathbb{E}_{\theta_r}\left(I_{\alpha_1} + I_{\alpha_2}\right) \tag{2.103}$$

avec

$$I_{\alpha_1} = \int_{\mathcal{V}_1} \frac{f\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_1\right)}{f\left(\mathbf{x},\boldsymbol{\theta}\right)} f\left(\mathbf{x}|\,\boldsymbol{\theta}\right) d\mathbf{x}, \tag{2.104}$$

et

$$I_{\alpha_2} = \int_{\mathcal{V}_2} \frac{f\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_1\right) f\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_2\right)}{f^2\left(\mathbf{x},\boldsymbol{\theta}\right)} f\left(\mathbf{x}|\,\boldsymbol{\theta}\right) d\mathbf{x}, \tag{2.105}$$

où les domaines $\mathcal{V}_1$ et $\mathcal{V}_2$ sont définis de la manière suivante

$$\mathcal{V}_1 = \{\mathbf{x}\in\Omega|\ \frac{f\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_2\right)}{f\left(\mathbf{x},\boldsymbol{\theta}\right)} \geq 1\}, \tag{2.106}$$

et

$$\mathcal{V}_2 = \{\mathbf{x}\in\Omega|\ \frac{f\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_2\right)}{f\left(\mathbf{x},\boldsymbol{\theta}\right)} < 1\}. \tag{2.107}$$

En suivant le même processus de calcul proposé par [Org12], nous obtenons les expression suivantes :

$$I_{\alpha_1} = e^{-\frac{h_{1r}^2 + 2\theta_r h_{1r}}{2\sigma_{\theta_r}^2}} F_{\mathcal{N}\left(m_{\alpha_1},\frac{\sigma_n^2\|\mathbf{d}(h_{2r})\|^2}{2}\right)}\left(-\frac{\sigma_b^2 b\left(h_{2r}\right)}{2}\right) \tag{2.108}$$

et

$$I_{\alpha_2} = e^{\frac{2\,\mathrm{Re}\left(\tilde{\mathbf{d}}^H\left(h_{1d},h_{1r}\right)\mathbf{d}(h_{2r})\right)}{\sigma_b^2} - \frac{h_{1r}^2 + h_{2r}^2 + 2\theta_r(h_{1r}+h_{2r})}{2\sigma_{\theta_r}^2}} F_{\mathcal{N}\left(m_{\alpha_2},\frac{\sigma_n^2\|\mathbf{d}(h_{2r})\|^2}{2}\right)}\left(\frac{\sigma_b^2 b\left(h_{2r}\right)}{2}\right),$$

avec

$$m_{\alpha_1} = -\,\mathrm{Re}\left(\tilde{\mathbf{d}}^H\left(h_{1d},h_{1r}\right)\mathbf{d}\left(h_{2r}\right)\right), \tag{2.109}$$

$$m_{\alpha_2} = \mathrm{Re}\left(\tilde{\mathbf{d}}^H\left(h_{1d},h_{1r}\right)\mathbf{d}\left(h_{2r}\right)\right) + \|\mathbf{d}\left(h_{2r}\right)\|^2, \tag{2.110}$$

et

$$\tilde{\mathbf{d}}\left(h_{1d},h_{1r}\right) = \mathbf{g}\left(\theta_d + h_{1d},\theta_r + h_{1r}\right) - \mathbf{g}\left(\theta_d,\theta_r\right). \tag{2.111}$$

2.6.4 Simulation

Pour comparer la borne de Barankin Ziv-Zakai hybride (BBZZH) aux autres bornes, nous allons considérer le même modèle d'observation (2.67) que celui proposé dans la section précédente. Cependant, nous avons proposé une version scalaire de la BBZZH. Nous allons supposer que la phase est connue. Le modèle s'écrit donc $\mathbf{m}(\theta_d, \theta_r) = \theta_d \mathbf{c}(\theta_r)$ où θ_d est l'amplitude du signal et $\mathbf{c}(\theta_r) = \begin{bmatrix} 1 & e^{j\theta_r} & e^{j2\theta_r} & \cdots & e^{j(N-1)\theta_r} \end{bmatrix}$ est une cisoïde normalisée de fréquence angulaire θ_r. Les configurations de simulation sont les suivantes $N = 32$, $\theta_d = 1$ and $\sigma^2_{\theta_r} = \frac{1}{2}$. A l'aide des formules de Slepian-Bangs [Sle54] [Ban71], la borne de Cramér-Rao hybride est une matrice diagonale 2×2 donnée par

$$\mathbf{BCRH} = \begin{pmatrix} \frac{\sigma_b^2}{2N} & 0 \\ 0 & \left(\frac{2\theta_d^2}{\sigma_b^2} \left(\frac{N(N+1)(2N+1)}{6} - N^2 \right) + \frac{1}{\sigma^2_{\theta_r}} \right)^{-1} \end{pmatrix}.$$

La BBH, la BBWWH et la BBZZH proposée ici sont programmées avec deux points test $\mathbf{h}_1 \in [-1; 1] \times \{0\}$ et $\mathbf{h}_2 \in \{0\} \times \left[-\frac{3}{2}; \frac{3}{2}\right]$. Etant donné que ces trois bornes sont obtenues par maximisation des points test sur leur intervalle respectif, nous choisissons $\delta h_{1d} = 0.01$ le pas de discrétisation pour l'intervalle de recherche de $\mathbf{h}_1$ et $\delta h_{2r} = \frac{3}{2^8}$ celui de $\mathbf{h}_2$. Enfin, l'estimateur MVMAP consiste à obtenir le meilleur candidat $\theta_d \in [0; 2]$ et $\theta_r \in \left[-\frac{3}{2}; \frac{3}{2}\right]$ qui maximise la densité de probabilité jointe $f_{\mathbf{x},\theta_r;\theta_d}(\mathbf{x}, \theta_r; \theta_d)$. L'EQM empirique du MVMAP est obtenue à l'aide de 1000 tirages de Monte-Carlo. Nous illustrons sur la figure (2.2), la BCRH, la BBH, la BBWWH, la BBZZH et l'EQM empirique du MVMAP pour l'estimation du paramètre aléatoire θ_r. Etant donné que le problème d'estimation est non-linéaire par rapport à θ_r, un phénomène de décrochement est attendu pour l'EQM du MVMAP. D'après la figure (2.2), nous remarquons que la borne de Ziv-Zakaï hybride est aussi précise que la BBWWH pour la prédiction du seuil de décrochement. Une telle comparaison n'avait été faite que dans le cas bayésien (voir [BEV96] et [VRBM14]).

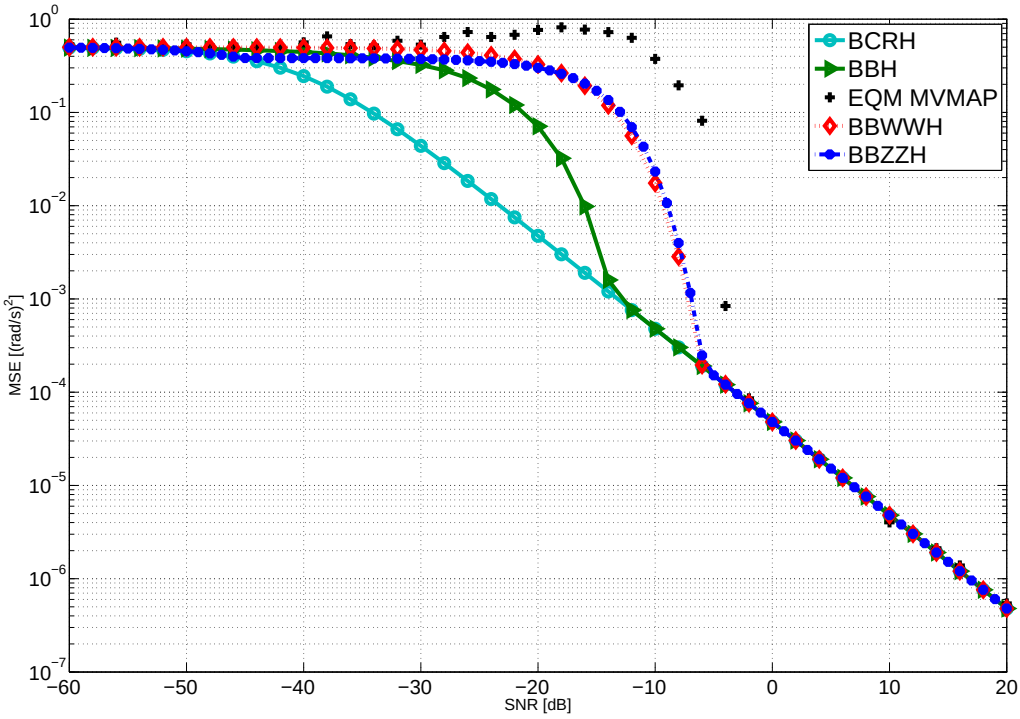

FIGURE 2.2 – Comparaison de la borne de Barankin Ziv-Zakaï hybride aux bornes de la famille de Weiss-Weinstein hybride et à l'EQM du MVMAP

Chapitre 3

Bornes de Cramér-Rao non standards

3.1 Borne de Cramér-Rao hybride sous contraintes

Comme mentionné dans le papier de Gorman et Hero [GO90], la borne de Cramér-Rao déterministe a été établie sous l'hypothèse implicite que l'espace des paramètres déterministes est un sous-ensemble d'ouvert de $\mathbb{R}^D$. Cependant, dans plusieurs applications, ces paramètres peuvent être contraints à appartenir dans un sous-ensemble de $\mathbb{R}^D$. C'est pourquoi, des travaux (par exemple [SN98, MCLB12]) ont été entrepris pour étendre les résultats présentés dans [GO90]. Ces bornes de Cramér-Rao "sous contraintes", qui prennent en compte des contraintes de type égalité entre paramètres, se sont révélées plus pertinentes pour l'étude des problèmes susmentionnés. Le but de cette section est donc de proposer une borne de Cramér-Rao hybride qui prend en compte de possibles contraintes de type égalité entre les paramètres non-aléatoires.

3.1.1 Contexte de l'étude et notations

Nous restons dans le contexte de l'estimation hybride du chapitre précédent. Pour mémoire, nous considérons un vecteur d'observation $\mathbf{x} \in \Omega$. $\boldsymbol{\theta}_d \in \Pi_d \subseteq \mathbb{R}^D$ désigne un vecteur de paramètres inconnus et non-aléatoires et $\boldsymbol{\theta}_r \in \Pi_r \subseteq \mathbb{R}^R$ est un vecteur de paramètres inconnus et aléatoires de densité de probabilité *a priori* $f_{\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}$ connue. Notons $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_d^T \ \boldsymbol{\theta}_r^T\right]^T$ le vecteur de paramètres complets de dimension $D+R$ et $f_{\mathbf{x},\boldsymbol{\theta}} \triangleq f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}$ la densité jointe des observations $\mathbf{x}$ et des paramètres aléatoires $\boldsymbol{\theta}_r$ éventuellement paramétrée par $\boldsymbol{\theta}_d$. La différence entre le contexte d'estimation hybride classique et ce contexte d'étude est que les paramètres non-aléatoires $\boldsymbol{\theta}_d$ sont contraints dans un sous espace $\mathcal{C}$ non vide de Π_d défini par $K < D$ égalités non redondantes :

$$\mathcal{C} = \{\boldsymbol{\theta}_d \in \Pi_d | \mathbf{c}(\boldsymbol{\theta}_d) = \mathbf{0}\}, \tag{3.1}$$

où $\mathbf{c}(\boldsymbol{\theta}_d)$ est un vecteur composé de K fonctions dérivables sur Π_d. Notons $\mathbf{C}(\boldsymbol{\theta}_d)$ une matrice $K \times (D+R)$ définie par

$$\mathbf{C}(\boldsymbol{\theta}_d) = \frac{d\mathbf{c}(\boldsymbol{\theta}_d)}{d\boldsymbol{\theta}^T} = \left[\frac{d\mathbf{c}(\boldsymbol{\theta}_d)}{d\boldsymbol{\theta}_d^T} \ \frac{d\mathbf{c}(\boldsymbol{\theta}_d)}{d\boldsymbol{\theta}_r^T}\right] = [\mathbf{C}_d(\boldsymbol{\theta}_d) \ \ \mathbf{0}], \tag{3.2}$$

où $\mathbf{C}_d(\boldsymbol{\theta}_d)$ est une matrice de dimension $K \times D$. Puisque les contraintes ne sont pas redondantes, le rang de la matrice $\mathbf{C}_d(\boldsymbol{\theta}_d)$ est K pour tout $\boldsymbol{\theta}_d \in \mathcal{C}$. Par conséquent, il existe une matrice $\mathbf{U}_d(\boldsymbol{\theta}_d)$ de dimension $K \times (D-K)$ telle que

$$\mathbf{C}_d(\boldsymbol{\theta}_d)\mathbf{U}_d(\boldsymbol{\theta}_d) = \mathbf{0} \text{ et } \mathbf{U}_d^T(\boldsymbol{\theta}_d)\mathbf{U}_d(\boldsymbol{\theta}_d) = \mathbf{I}_{D-K}, \tag{3.3}$$

où $\mathbf{I}_{D-K}$ est une matrice identité de dimension $D-K$. Par ailleurs, si les équations (3.3) sont vérifiées, alors la matrice étendue $\mathbf{U}(\boldsymbol{\theta}_d) = \begin{bmatrix} \mathbf{U}_d(\boldsymbol{\theta}_d) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_R \end{bmatrix}$ satisfait les relations suivantes

$$\mathbf{C}(\boldsymbol{\theta}_d)\,\mathbf{U}(\boldsymbol{\theta}_d) = \mathbf{0} \text{ et } \mathbf{U}^T(\boldsymbol{\theta}_d)\,\mathbf{U}(\boldsymbol{\theta}_d) = \mathbf{I}_{D+R-K}. \tag{3.4}$$

Notons que les vecteurs colonnes de $\mathbf{U}_d(\boldsymbol{\theta}_d)$ forment une base du noyau de $\mathbf{C}_d(\boldsymbol{\theta}_d)$ et que les vecteurs colonnes de $\mathbf{U}(\boldsymbol{\theta}_d)$ forment une base du noyau de $\mathbf{C}(\boldsymbol{\theta}_d)$. Si nous avions voulu imposer des contraintes sur les paramètres aléatoires, c'est-à-dire $\mathbf{c}(\boldsymbol{\theta}_d, \boldsymbol{\theta}_r) = \mathbf{0}$, alors la matrice $\mathbf{U}$ dépendrait de $\boldsymbol{\theta}_r$ conduisant à une expression qui n'est pas une borne puisqu'elle dépend des estimées de $\boldsymbol{\theta}_r$ (voir section 3.1.2 pour les détails). Il faudrait dans ce cas utiliser des techniques de type Monte-Carlo (coûteuses) pour évaluer la borne.

3.1.2 Borne proposée

Soit $\widehat{\boldsymbol{\theta}}(\mathbf{x})$ un estimateur de $\boldsymbol{\theta}$. Les conditions d'applicabilité de notre borne nécessitent que l'estimateur soit non-biaisé, comme pour la borne de Cramér-Rao classique [RS87,Mes06] c'est-à-dire que :

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right] = \mathbf{0}. \tag{3.5}$$

Rappelons que tout estimateur non biaisé satisfait la relation suivante : $\forall\ 1 \leq i \leq D+R$, nous avons

$$\begin{aligned} & \int_{\mathbb{R}^R}\int_{\Omega}\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)\frac{\partial}{\partial\theta_i} f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)\,d\mathbf{x}d\boldsymbol{\theta}_r \\ = \ & \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_d;\boldsymbol{\theta}_r}\left[\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right] - \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_d;\boldsymbol{\theta}_r}\left[\frac{\partial}{\partial\theta_i}\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)\right] \\ = \ & \mathbf{0} + \mathbf{e}_i, \end{aligned} \tag{3.6}$$

où θ_i est la $i^{\text{ième}}$ composante de $\boldsymbol{\theta}$ et $\mathbf{e}_i$ est un vecteur qui vaut 1 au $i^{\text{ième}}$ élément et 0 ailleurs. l'équation (3.6) est équivalent, sous forme matricielle, à

$$\int_{\mathbb{R}^R}\int_{\Omega}\left(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\right)\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)\,d\mathbf{x}d\boldsymbol{\theta}_r = \mathbf{I}_{D+R}, \tag{3.7}$$

avec $\mathbf{v}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d) = \frac{\partial \ln f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial\boldsymbol{\theta}^T}$.

Dans la suite, pour simplifier les notations, on notera $\widetilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}$, $\mathbf{v} = \mathbf{v}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)$ et $\mathbf{U} = \mathbf{U}(\boldsymbol{\theta}_d)$. Pour toute matrice carrée $\mathbf{M}$:

$$\begin{aligned} & \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\left(\widetilde{\boldsymbol{\theta}} - \mathbf{M}\mathbf{U}\mathbf{U}^T\mathbf{v}\right)\left(\widetilde{\boldsymbol{\theta}} - \mathbf{M}\mathbf{U}\mathbf{U}^T\mathbf{v}\right)^T\right] \\ = \ & \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widetilde{\boldsymbol{\theta}}\widetilde{\boldsymbol{\theta}}^T\right] + \mathbf{M}\mathbf{U}\mathbf{U}^T\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}\mathbf{v}^T\right]\mathbf{U}\mathbf{U}^T\mathbf{M}^T \\ & -\mathbf{M}\mathbf{U}\mathbf{U}^T\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}\widetilde{\boldsymbol{\theta}}^T\right] - \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widetilde{\boldsymbol{\theta}}\mathbf{v}^T\right]\mathbf{U}\mathbf{U}^T\mathbf{M}^T. \end{aligned}$$

Or $\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\left(\widetilde{\boldsymbol{\theta}} - \mathbf{M}\mathbf{U}\mathbf{U}^T\mathbf{v}\right)\left(\widetilde{\boldsymbol{\theta}} - \mathbf{M}\mathbf{U}\mathbf{U}^T\mathbf{v}\right)^T\right]$ est une matrice définie non négative, et comme $\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widetilde{\boldsymbol{\theta}}\mathbf{v}^T\right] = \mathbf{I}_{D+R}$ d'après (3.7), nous obtenons pour toute matrice carrée $\mathbf{M}$:

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widetilde{\boldsymbol{\theta}}\widetilde{\boldsymbol{\theta}}^T\right] \succeq \mathbf{M}\mathbf{U}\mathbf{U}^T + \mathbf{U}\mathbf{U}^T\mathbf{M}^T - \mathbf{M}\mathbf{U}\mathbf{U}^T\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}\mathbf{v}^T\right]\mathbf{U}\mathbf{U}^T\mathbf{M}^T. \tag{3.8}$$

Puisque cette inégalité est vérifiée pour tout $\mathbf{M}$, la meilleure borne est obtenue, sous condition d'existence, en maximisant le membre de droite de (3.8) par rapport à la matrice $\mathbf{M}$. La borne obtenue est appelée la Borne de Cramér-Rao Hybride sous Contraintes notée **BCRHC**. Remarquons que la matrice $\mathbf{U}^T\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}\mathbf{v}^T\right]\mathbf{U}$ est à valeur réelle, symétrique, définie positive. Donc elle est diagonalisable. Il existe une matrice diagonale $\mathbf{D}$ et une matrice orthogonale $\mathbf{Q}$ tel que $\mathbf{U}^T\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}\mathbf{v}^T\right]\mathbf{U} = \mathbf{QDQ}^T$. Par conséquent, la **BCRHC** est égale à :

$$\mathbf{BCRHC} = \sup_{\mathbf{M}}\left\{\mathbf{UQD}^{-1}\mathbf{Q}^T\mathbf{U}^T - \left(\mathbf{UQD}^{-1} - \mathbf{MUQ}\right)\mathbf{D}\left(\mathbf{UQD}^{-1} - \mathbf{MUQ}\right)^T\right\}. \tag{3.9}$$

Nous avons une différence de deux matrices définies non négatives. Puisque le premier terme $\mathbf{UQD}^{-1}\mathbf{Q}^T\mathbf{U}^T$ ne dépend pas de $\mathbf{M}$, le maximum est atteint en minimisant le second terme. Le minimum $\mathbf{M}_{\min}$ est obtenu si :

$$\mathbf{M}_{\min}\mathbf{U} = \mathbf{UQD}^{-1}\mathbf{Q}^T = \mathbf{U}\left(\mathbf{U}^T\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}\mathbf{v}^T\right]\mathbf{U}\right)^{-1}. \tag{3.10}$$

Finalement, en injectant (3.10) dans (3.9), nous obtenons la borne proposée

$$\mathbf{BCRHC} = \mathbf{U}\left(\mathbf{U}^T\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}\mathbf{v}^T\right]\mathbf{U}\right)^{-1}\mathbf{U}^T. \tag{3.11}$$

Remarques :

• en général, il n'est pas nécessaire que la matrice d'information de Fisher $\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}\mathbf{v}^T\right]$ soit inversible mais seulement l'inversibilité de la matrice $\mathbf{U}^T\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}\mathbf{v}^T\right]\mathbf{U}$. Cette condition est également nécessaire pour la borne de Cramér-Rao déterministe sous contraintes [SN98].

• si la matrice $\mathbf{U}$ dépend de $\boldsymbol{\theta}_r$ alors $\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\widetilde{\boldsymbol{\theta}}\mathbf{v}^T\mathbf{U}\right] \neq \mathbf{U}$, il serait toujours possible d'obtenir une inégalité, mais le minorant dépendrait de $\widehat{\boldsymbol{\theta}}$. Or nous n'avons pas une expression explicite de $\widehat{\boldsymbol{\theta}}$ pour des problèmes d'estimation non linéaires donc la borne obtenue ne sera pas simple à évaluer numériquement.

3.1.3 Comparaison avec la borne de Cramér-Rao hybride sans contraintes

La BCRH sans contraintes est donnée par (2.20) :

$$\mathbf{BCRH} = \mathbf{F}_H^{-1}, \text{ où } \mathbf{F}_H = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}\mathbf{v}^T\right] \tag{3.12}$$

est la matrice d'information de Fisher hybride. La BCRH est un cas particulier de la BCRHC (3.11) lorsque $K = 0$ (ce qui implique $\mathbf{U} = \mathbf{I}_{D+R}$). Dans les autres cas, la BCRH et la BCRHC sont évidemment différentes. Cependant, une comparaison entre ces deux bornes est possible lorsque la matrice $\mathbf{F}_H$ est inversible (sinon la BCRH n'existe pas). Comme $\mathbf{F}_H$ est symétrique définie positive, il existe une matrice symétrique inversible $\mathbf{F}_H^{\frac{1}{2}}$ tel que $\mathbf{F}_H = \mathbf{F}_H^{\frac{1}{2}}\mathbf{F}_H^{\frac{1}{2}}$. Donc la BCRHC peut s'écrire sous la forme :

$$\mathbf{BCRHC} = \mathbf{F}_H^{-\frac{1}{2}}\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}\left(\mathbf{U}^T\mathbf{F}_H^{\frac{T}{2}}\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{F}_H^{\frac{T}{2}}\mathbf{F}_H^{-\frac{T}{2}} = \mathbf{F}_H^{-\frac{1}{2}}\mathbf{P}_{\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}}\mathbf{F}_H^{-\frac{T}{2}}$$

où

$$\mathbf{P}_{\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}} = \mathbf{F}_H^{\frac{1}{2}}\mathbf{U}\left(\mathbf{U}^T\mathbf{F}_H^{\frac{T}{2}}\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}\right)^{-1}\left(\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}\right)^T$$

est une matrice de projection orthogonale sur l'espace engendré par les vecteurs colonnes de $\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}$. Si on note $\mathbf{P}^{\perp}_{\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}}$ la matrice de projection sur l'espace orthogonal à $\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}$, alors nous obtenons $\mathbf{P}_{\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}} + \mathbf{P}^{\perp}_{\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}} = \mathbf{I}$ impliquant :

$$\mathbf{BCRHC} = \mathbf{F}_H^{-\frac{1}{2}}\left(\mathbf{I} - \mathbf{P}_{\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}}\right)\mathbf{F}_H^{-\frac{T}{2}} = \mathbf{F}_H^{-1} - \mathbf{F}_H^{-\frac{1}{2}}\mathbf{P}^{\perp}_{\mathbf{F}_H^{\frac{1}{2}}\mathbf{U}}\mathbf{F}_H^{-\frac{T}{2}} \preceq \mathbf{F}_H^{-1}.$$

Par conséquent $\mathbf{BCRHC} \preceq \mathbf{BCRH}$. Ce résultat est logique puisque les contraintes peuvent être interprétées comme des informations supplémentaires permettant de mieux estimer les paramètres d'intérêts. Par ailleurs, il a été montré dans [GO91] que les algorithmes d'estimation qui incluent les contraintes sur les paramètres présentent une EQM en dessous de la borne de Cramér-Rao classique.

3.1.4 Comparaison avec la borne de Cramér-Rao marginale

La borne de Cramér-Rao Marginale (BCRM) s'applique pour les algorithmes qui estiment les paramètres non-aléatoires $\boldsymbol{\theta}_d$ uniquement, les paramètres aléatoires $\boldsymbol{\theta}_r$ sont considérés comme des paramètres de nuisance. La comparaison entre la BCRHC et la BCRM ne concerne donc que les paramètres non-aléatoires $\boldsymbol{\theta}_d$. Pour cela, notons d'abord que, la BCRHC peut se décomposer en quatre blocs :

$$\mathbf{BCRHC} = \begin{bmatrix} \mathbf{BCRHC}_d & \mathbf{BCRHC}_{dr}^T \\ \mathbf{BCRHC}_{dr} & \mathbf{BCRHC}_r \end{bmatrix} \tag{3.13}$$

où les blocs diagonaux $\mathbf{BCRHC}_d$ and $\mathbf{BCRHC}_r$ sont respectivement les bornes inférieures de l'EQM des paramètres non aléatoires $\boldsymbol{\theta}_d$ et des paramètres aléatoires $\boldsymbol{\theta}_r$, c'est-à-dire :

$$\begin{aligned} \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\left(\widehat{\boldsymbol{\theta}}_d(\mathbf{x}) - \boldsymbol{\theta}_d\right)\left(\widehat{\boldsymbol{\theta}}_d(\mathbf{x}) - \boldsymbol{\theta}_d\right)^T\right] &\succeq \mathbf{BCRHC}_d, \\ \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\left(\widehat{\boldsymbol{\theta}}_r(\mathbf{x}) - \boldsymbol{\theta}_r\right)\left(\widehat{\boldsymbol{\theta}}_r(\mathbf{x}) - \boldsymbol{\theta}_r\right)^T\right] &\succeq \mathbf{BCRHC}_r. \end{aligned}$$

Puis, si on note $\mathbf{v}_d = \frac{\partial \ln f(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial \boldsymbol{\theta}_d}$ et $\mathbf{v}_r = \frac{\partial \ln f(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)}{\partial \boldsymbol{\theta}_r}$, alors la matrice d'information de Fisher hybride se décompose également en :

$$\mathbf{F}_H = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\begin{bmatrix} \mathbf{v}_d\mathbf{v}_d^T & \mathbf{v}_d\mathbf{v}_r^T \\ \mathbf{v}_r\mathbf{v}_d^T & \mathbf{v}_r\mathbf{v}_r^T \end{bmatrix}.$$

Par conséquent la borne proposée s'écrit :

$$\mathbf{BCRHC} = \mathbf{U}\mathbb{E}^{-1}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\begin{bmatrix} \mathbf{U}_d^T\mathbf{v}_d\mathbf{v}_d^T\mathbf{U}_d & \mathbf{U}_d^T\mathbf{v}_d\mathbf{v}_r^T \\ \mathbf{v}_r\mathbf{v}_d^T\mathbf{U}_d & \mathbf{v}_r\mathbf{v}_r^T \end{bmatrix}\mathbf{U}^T \tag{3.14}$$

Si nous notons $\mathbf{S}$ le complément de Schur défini par

$$\mathbf{S} = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{U}_d^T\mathbf{v}_d\mathbf{v}_d^T\mathbf{U}_d\right] - \mathbf{R},$$

où

$$\mathbf{R} = \mathbf{U}_d^T\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}_d\mathbf{v}_r^T\right]\mathbb{E}^{-1}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}_r\mathbf{v}_r^T\right]\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}_r\mathbf{v}_d^T\right]\mathbf{U}_d,$$

alors une inversion par bloc de (3.14) permet d'exprimer la BCRHC sous la forme suivante :

$$\begin{aligned}\mathbf{BCRHC} &= \begin{bmatrix} \mathbf{U}_d & \mathbf{0} \\ -\mathbb{E}^{-1}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}_r\mathbf{v}_r^T\right]\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}_r\mathbf{v}_d^T\right]\mathbf{U}_d & \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{S}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbb{E}^{-1}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}_r\mathbf{v}_r^T\right] \end{bmatrix} \\ &\times \begin{bmatrix} \mathbf{U}_d^T & -\mathbf{U}_d^T\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}_d\mathbf{v}_r^T\right]\mathbb{E}^{-1}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}_r\mathbf{v}_r^T\right] \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \end{aligned} \tag{3.15}$$

Par identification entre (3.13) et (3.15), nous établissons l'égalité suivante :

$$\mathbf{BCRHC}_d = \mathbf{U}_d\mathbf{S}^{-1}\mathbf{U}_d^T. \tag{3.16}$$

Puisque $\mathbf{R}$ est une matrice définie non négative, nous avons $\mathbf{S} \succeq \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{U}_d^T\mathbf{v}_d\mathbf{v}_d^T\mathbf{U}_d\right]$, ce qui implique que :

$$\mathbf{BCRHC}_d \preceq \mathbf{U}_d\left(\mathbf{U}_d^T\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{v}_d\mathbf{v}_d^T\right]\mathbf{U}_d\right)^{-1}\mathbf{U}_d^T. \tag{3.17}$$

Nous retrouvons la BCRM dans le membre de droite de l'équation (3.17). Par conséquent, la BCRHC est généralement inférieure à la BCRC marginale. Ce résultat est une extension des relations d'ordre existantes entre les bornes hybrides et les bornes marginales sans contraintes [RM97].

3.1.5 Application à l'estimation de fréquence Doppler

Considérons un système radar actif recevant des échos atténués, retardés et translatés en fréquence d'un signal complexe connu à l'émission $e_T(t)\,e^{j2\pi f_c t}$ où f_c est la fréquence porteuse et $e_T(t)$ est l'enveloppe du signal émis. Les antennes du radar reçoivent un train d'impulsion (burst) composé de L impulsions et l'intervalle de répétition des impulsions est noté T. Chaque impulsion dure T_0, possède une bande passante B et est réfléchie par une cible en mouvement "lent" par rapport à $e_T(t)$, c'est-à-dire que $|2v(L-1)T| << \frac{c}{B}$ (voir [LM04]) et $\frac{2v}{c}T_0 f_c << 1$ (effet Doppler négligeable sur $e_T(t)$), où c est la célérité de la lumière et v est la vitesse de la cible. Sous les hypothèses standards que le bruit est temporellement blanc (bruit thermique) de puissance σ_n^2 et qu'il n'y a pas de fluctuation de la cible durant l'émission du train d'impulsions, un modèle simplifié pour la l^{ième} impulsion, $1 \leq l \leq L$, est donné par [LM04] :

$$x_l(t) = e_T(t-\tau)\,\alpha_l + b_l(t), \quad \alpha_l = \alpha e^{j2\pi f(l-1)}, \tag{3.18}$$

où $f = -2f_c\frac{v}{c}T$, $\frac{-1}{2} \leq f \leq \frac{1}{2}$, est la fréquence Doppler normalisée et α représente l'atténuation complexe du signal. Par souci de simplicité, nous supposons que la distance entre le radar et la cible est connue. Donc le signal en sortie du filtre adapté est donné par

$$y_l = se^{j2\pi f(l-1)} + n_l,\ s = \sqrt{BT_0}\alpha = r + jq. \tag{3.19}$$

Le vecteur de paramètres inconnus à estimer est $\boldsymbol{\theta} = [r\ q\ f]^T$ où le vecteur $[r\ q]^T$ est supposé non aléatoire, f est un paramètre aléatoire de densité de probabilité *a priori* gaussienne $\mathcal{N}\left(f_0, \sigma_f^2\right)$ et le bruit b_l est supposé gaussien complexe circulaire $b_l \sim \mathcal{CN}\left(0, \sigma_n^2\right)$. De plus les variables aléatoires f et b_l sont supposées indépendantes. Ce scénario peut correspondre à un radar multi-fonction qui est entré en mode poursuite après avoir détecté une cible en mode veille (surveillance). Les paramètres $|s|^2$ et f_0, ont été préablement estimés en mode veille. Cependant, la vitesse radiale peut varier légèrement durant le temps de basculement en mode poursuite. C'est

pourquoi nous modélisons la fréquence doppler f par une variable aléatoire avec une densité de probabilité *a priori* centrée en f_0. La question est de savoir si l'accès à la puissance du signal reçu $|s|^2$, permettant d'établir la contrainte $r^2+q^2=|s|^2$, nous aide ou pas à mieux estimer la fréquence Doppler f. Pour cela, il suffit de comparer la BCRHC et la BCRH. En utilisant (3.12), la BCRH classique est donnée par :

$$\begin{pmatrix} \frac{2L}{\sigma_n^2} & 0 & \frac{2\pi qL(1-L)}{\sigma_n^2} \\ 0 & \frac{2L}{\sigma_n^2} & \frac{2\pi rL(L-1)}{\sigma_n^2} \\ \frac{2\pi qL(1-L)}{\sigma_n^2} & \frac{2\pi rL(L-1)}{\sigma_n^2} & \frac{4\pi^2\left(r^2+q^2\right)L(L-1)(2L-1)}{3\sigma_n^2}+\frac{1}{\sigma_f^2T^2} \end{pmatrix}^{-1}. \tag{3.20}$$

La BCRHC est obtenue en utilisant la relation (3.11) et en dérivant la matrice unitaire

$$\mathbf{U}=\begin{pmatrix} \frac{q}{|s|} & \frac{-r}{|s|} & 0 \\ 0 & 0 & 1 \end{pmatrix}^T.$$

Pour valider notre étude, nous comparons la BCRHC et la BCRH à l'EQM du MVMAP définie par :

$$\left(\widehat{r},\widehat{q},\widehat{f}\right)=\underset{(r,q)\in\mathbb{R}^2,\ f\in]-0.5;0.5]}{\arg\max} f_{\mathbf{y},f;r,q}\left(\mathbf{y},f;r,q\right), \tag{3.21}$$

et à l'EQM du MVMAP sous contrainte (MVMAPC) qui consiste à restreindre les paramètres de recherche (r,q) sur le cercle $\mathcal{S}=\left\{(r,q)\,|r^2+q^2=|s|^2\right\}$. Les paramètres de simulation sont les suivants : $r=\frac{1}{\sqrt{2}}$, $|s|^2=0.8$, $f=0.25$, $\sigma_f=0.05$ et $L=32$. Les EQMs empiriques du MVMAP et MVMAPC sont obtenues par 5000 tirages de type Monte-Carlo avec un pas $\delta f=2^{-14}$ pour l'estimation de la fréquence. Dans la figure (3.1), les EQMs empiriques totales du MVMAP et du MVMAPC sont comparées avec la trace de la BCRH et de la BCRHC. On peut noter que l'EQM du MVMAPC est bien en dessous de la BCRH tandis que la BCRHC le prédit parfaitement à fort RSB. Dans la figure (3.2), les performances du MVMAP et du MVMAPC sont comparées aux bornes BCRH et BCRHC pour l'estimation de la fréquence Doppler f. Nous remarquons que la BCRH et la BCRHC coïncident donc l'estimation de f est asymptotiquement indépendante de la contrainte sur la puissance du signal. Ce résultat est également confirmé par les comportements asymptotiques du MVMAP et du MVMAPC qui exhibent exactement les mêmes erreurs avec ou sans la contrainte sur la puissance. C'est une extension de la propriété connue du problème d'estimation de fréquence non-aléatoire [RB74] au cas où la fréquence est aléatoire.

3.2 La borne de Cramér-Rao hybride récursive

Les bornes dites récursives ont été développées pour évaluer les performances optimales de filtrage (bayésien). Le filtrage consiste à estimer de manière récursive (en ligne) l'état d'un système dynamique. Le modèle d'observations d'un tel système est défini dans le cas général par

$$\mathbf{x}_k=\mathbf{h}_k\left(\boldsymbol{\theta}_k,\mathbf{b}_k\right),1\leq k\leq K, \tag{3.22}$$

où $\mathbf{x}_k$ est un vecteur de mesure, $\mathbf{b}_k$ est le bruit de mesure supposé aléatoire, $\mathbf{h}_k(.)$ est une fonction déterministe connue modélisant le processus de mesures, et les paramètres $\{\boldsymbol{\theta}_k\}_{1\leq k\leq K}\in\Theta\subset\mathbb{R}^P$ sont les inconnus à estimer (ils définissent l'état du système). Cependant, les observations $\mathbf{x}_k$

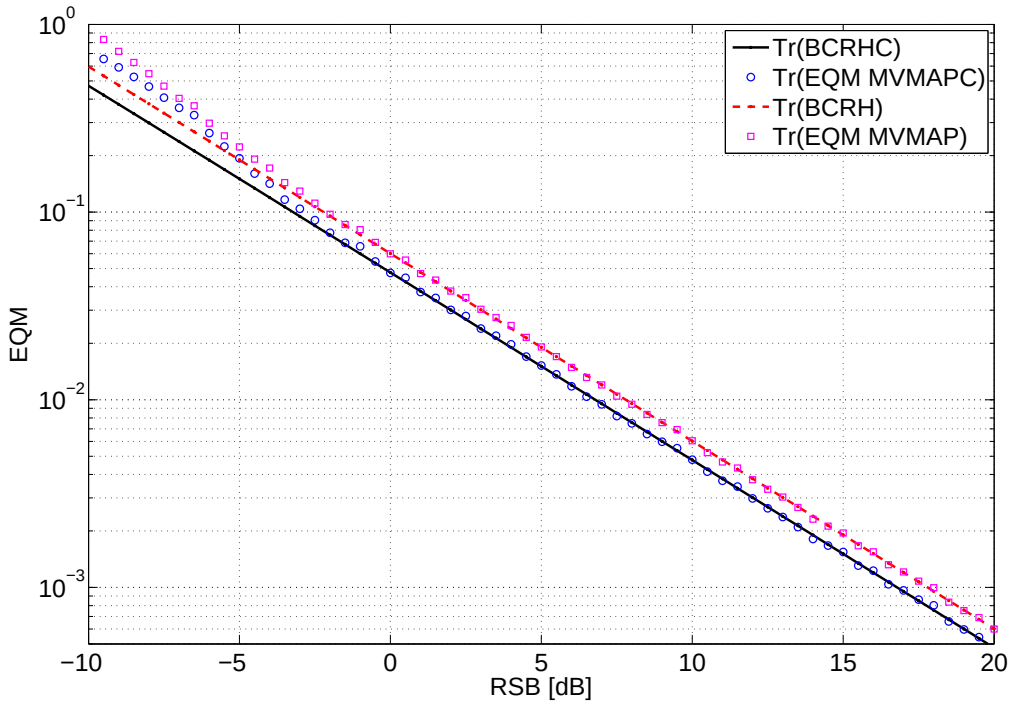

FIGURE 3.1 – Comparaison des EQMs empiriques totales du MVMAP et du MVMAPC aux bornes BCRH et BCRHC en fonction du RSB

ne suffisent pas en général à garantir l'existence et l'unicité des solutions $\left\{\widehat{\boldsymbol{\theta}}_k\right\}_{1\leq k\leq K}$. Une information *a priori* sur les paramètres $\{\boldsymbol{\theta}_k\}_{1\leq k\leq K}$ est nécéssaire : une approche de type

$$\boldsymbol{\theta}_{k+1} = \mathbf{g}_k\left(\boldsymbol{\theta}_k, \mathbf{w}_k\right), \tag{3.23}$$

est souvent utilisée dans la littérature où $\mathbf{w}_k$ est un bruit aléatoire et $\mathbf{g}_k\left(.\right)$ est une fonction déterministe qui décrit la l'évolution *a priori* des états. Il existe de nombreux problèmes pratiques qui relèvent de cette théorie comme les problèmes de localisation, de navigation et de poursuite d'un véhicule. En effet, l'équation (3.22) représente en pratique les données mesurées par les capteurs et l'équation (3.23) décrit le déplacement *a priori* du véhicule. L'estimation de l'état $\boldsymbol{\theta}_k$ au vu des observations $\mathbf{x}_{1:k} = \left[\mathbf{x}_1^T\ \mathbf{x}_2^T\ \cdots\ \mathbf{x}_k^T\right]^T$ dépend donc du modèle *a priori* choisi. Puisque $\boldsymbol{\theta}_k$ est aléatoire, l'estimateur de $\boldsymbol{\theta}_k$, minimisant son EQM, est obtenue par la moyenne *a posteriori* :

$$\widehat{\boldsymbol{\theta}}_k = \mathbb{E}_{\boldsymbol{\theta}_k|\overline{\mathbf{x}}_{1:k}}\left[\boldsymbol{\theta}_k\right]. \tag{3.24}$$

Cependant, il n'existe pas de solution explicite de $\widehat{\boldsymbol{\theta}}_k$ sauf pour certains cas de filtrages linéaires et gaussiens (le filtrage de Kalman) [Sim06]. De même que les estimateurs classiques, les performances de ces filtres sont accessibles uniquement par des simulations (méthode de Monte-Carlo ou approximation particulaire). Ces méthodes sont généralement lourdes en coût de calcul. Une solution alternative est d'utiliser les bornes inférieures pour minorer l'EQM de $\widehat{\boldsymbol{\theta}}_k$. Ces bornes donnent une limite des performances atteignables et elles permettent de déterminer en amont si les performances exigées sont réalistes ou non.

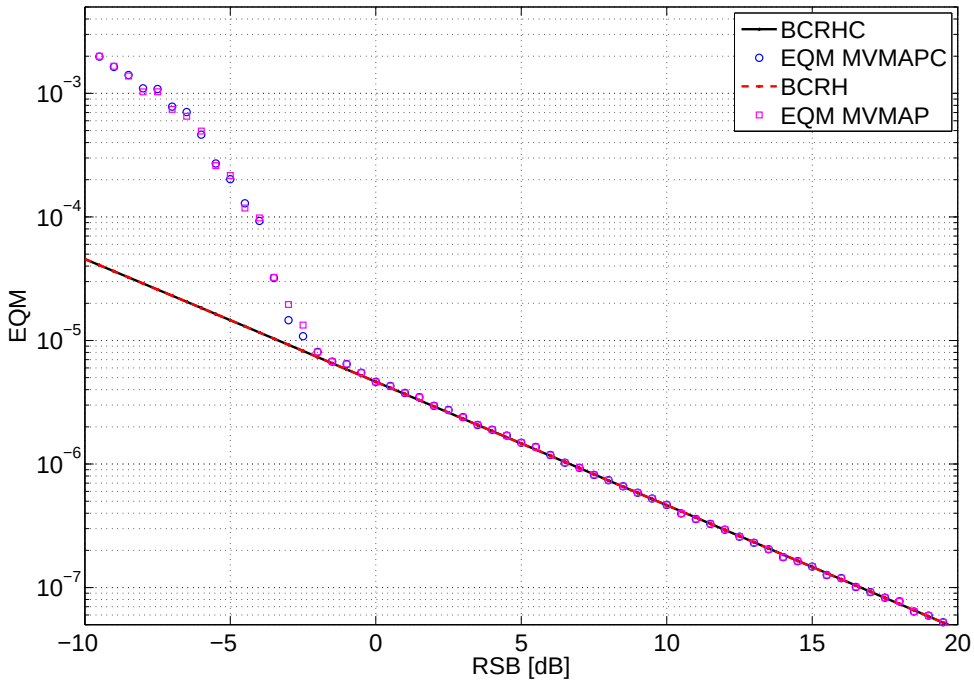

FIGURE 3.2 – Comparaison des EQMs et des bornes pour l'estimation de la fréquence Doppler en fonction du RSB

Plusieurs bornes sont apparues dans la littérature pour les filtrages à temps discret. En effet, Bobrovsky et Zakai [BZ75] ont été les premiers à adapter la borne de Cramér-Rao aux systèmes dynamiques gaussiens à temps discret pour un état scalaire. Puis cette borne a été étendue par Galdos [Gal80] aux cas d'estimation de paramètres multidimensionnels. Enfin, une version plus générale de la borne de Cramér-Rao sous une forme récursive a été proposée par [TMN98] pour des systèmes de filtrage éventuellement non gaussiens. Les extensions aux bornes autres que Cramér-Rao (en particulier Weiss-Weinstein) ont été établies par [RN05, RO07, XGMM13].

3.2.1 Problématique

Considérons le problème de filtrage non-linéaire suivant :

$$\begin{cases} \mathbf{x}_k = \mathbf{h}_k\left(\boldsymbol{\theta}_k, \mathbf{b}_k, \boldsymbol{\lambda}\right) \\ \boldsymbol{\theta}_k = \mathbf{g}_{k-1}\left(\boldsymbol{\theta}_{k-1}, \mathbf{w}_{k-1}, \boldsymbol{\alpha}\right) \end{cases}, 1 \leq k \leq K \tag{3.25}$$

où les fonctions $\mathbf{g}_k(.)$ et $\mathbf{h}_k(.)$ dépendent de paramètres inconnus $\boldsymbol{\lambda}$ et $\boldsymbol{\alpha}$ supposés non aléatoires. Les bruits $\mathbf{w}_k$ et $\mathbf{b}_k$ sont supposés des variables aléatoires mutuellement indépendantes de densités de probabilité conditionnelles respectives $f_{\mathbf{w}_k|\boldsymbol{\beta}}$ et $f_{\mathbf{b}_k|\boldsymbol{\mu}}$ connues pour tout k et les paramètres vectoriels $\boldsymbol{\varphi} \triangleq \left[\boldsymbol{\alpha}^T\ \boldsymbol{\beta}^T\ \boldsymbol{\lambda}^T\ \boldsymbol{\mu}^T\right]^T$ sont supposés inconnus et non aléatoires. Les distributions $f_{\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1},\boldsymbol{\alpha},\boldsymbol{\beta}}$ et $f_{\mathbf{x}_k|\boldsymbol{\theta}_k,\boldsymbol{\lambda},\boldsymbol{\mu}}$ sont déduites de (3.25) et la distribution $f_{\boldsymbol{\theta}_0|\boldsymbol{\alpha}}$ est supposée connue.

Si toutes les distributions $f_{\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1},\boldsymbol{\alpha},\boldsymbol{\beta}}$ et $f_{\mathbf{x}_k|\boldsymbol{\theta}_k,\boldsymbol{\lambda},\boldsymbol{\mu}}$ sont dérivables par rapport à $\boldsymbol{\theta}_k$, $\boldsymbol{\theta}_{k-1}$ et $\boldsymbol{\varphi}$ pour tout $k \geq 1$ alors l'EQM d'estimation sur les paramètres $(\boldsymbol{\theta}_k; \boldsymbol{\varphi}) \triangleq \left[\boldsymbol{\theta}_k^T\ \boldsymbol{\varphi}^T\right]^T$ est minorée

par la borne de Cramér-Rao hybride définie par

$$\mathbf{BCRH}_{\boldsymbol{\theta}_k,\boldsymbol{\varphi}} = \mathbb{E}_{\mathbf{x}_{1:k},\boldsymbol{\theta}_k|\boldsymbol{\varphi}} \left[\frac{\partial \ln f_{\mathbf{x}_{1:k},\boldsymbol{\theta}_k|\boldsymbol{\varphi}} (\mathbf{x}_{1:k}, \boldsymbol{\theta}_k|\boldsymbol{\varphi})}{\partial (\boldsymbol{\theta}_k; \boldsymbol{\varphi})} \frac{\partial \ln f_{\mathbf{x}_{1:k},\boldsymbol{\theta}_k|\boldsymbol{\varphi}} (\mathbf{x}_{1:k}|\boldsymbol{\varphi})}{\partial (\boldsymbol{\theta}_k; \boldsymbol{\varphi})^T} \right]^{-1}. \tag{3.26}$$

Si $f_{\mathbf{x}_{1:k},\boldsymbol{\theta}_k|\boldsymbol{\varphi}} (\mathbf{x}_{1:k}, \boldsymbol{\theta}_k|\boldsymbol{\varphi})$ est deux fois dérivables par rapport à $\boldsymbol{\theta}_k$ et $\boldsymbol{\varphi}$, alors (3.26) équivaut à

$$\mathbf{BCRH}_{\boldsymbol{\theta}_k,\boldsymbol{\varphi}} = -\mathbb{E}_{\mathbf{x}_{1:k},\boldsymbol{\theta}_k|\boldsymbol{\varphi}} \left[\frac{\partial^2 \ln f_{\mathbf{x}_{1:k},\boldsymbol{\theta}_k|\boldsymbol{\varphi}} (\mathbf{x}_{1:k}, \boldsymbol{\theta}_k|\boldsymbol{\varphi})}{\partial (\boldsymbol{\theta}_k; \boldsymbol{\varphi}) \partial (\boldsymbol{\theta}_k; \boldsymbol{\varphi})^T} \right]^{-1}. \tag{3.27}$$

Cependant, la densité de probabilité jointe $f_{\mathbf{x}_{1:k},\boldsymbol{\theta}_k|\boldsymbol{\varphi}} (\mathbf{x}_{1:k}, \boldsymbol{\theta}_k|\boldsymbol{\varphi})$ est difficile à obtenir analytiquement. En effet, une solution est d'exprimer la densité jointe $f_{\mathbf{x}_{1:k},\boldsymbol{\theta}_{0:k}|\boldsymbol{\varphi}} (.)$ via le modèle de récurrence (3.25),

$$\begin{aligned} f_{\mathbf{x}_{1:k},\boldsymbol{\theta}_{0:k}|\boldsymbol{\varphi}} (\mathbf{x}_{1:k}, \boldsymbol{\theta}_{0:k}|\boldsymbol{\varphi}) &= f_{\boldsymbol{\theta}_0|\boldsymbol{\varphi}} (\boldsymbol{\theta}_0|\boldsymbol{\varphi}) \prod_{j=1}^{k} f_{\mathbf{x}_j,\boldsymbol{\theta}_j|\boldsymbol{\varphi}} (\mathbf{x}_j, \boldsymbol{\theta}_j|\boldsymbol{\theta}_{j-1}, \boldsymbol{\varphi}) \\ &= f_{\boldsymbol{\theta}_0|\boldsymbol{\varphi}} (\boldsymbol{\theta}_0|\boldsymbol{\varphi}) \prod_{j=1}^{k} f_{\mathbf{x}_j|\boldsymbol{\theta}_j,\boldsymbol{\varphi}} (\mathbf{x}_j|\boldsymbol{\theta}_j, \boldsymbol{\varphi}) f_{\boldsymbol{\theta}_j|\boldsymbol{\theta}_{j-1},\boldsymbol{\varphi}} (\boldsymbol{\theta}_j|\boldsymbol{\theta}_{j-1}, \boldsymbol{\varphi}), \end{aligned} \tag{3.28}$$

puis de marginaliser la densité jointe (3.28) par rapport aux variables $\boldsymbol{\theta}_{0:k-1}$ pour obtenir la distribution de $f_{\mathbf{x}_{1:k},\boldsymbol{\theta}_k|\boldsymbol{\varphi}} (.)$ et par conséquent la borne de Cramér-Rao. Malheureusement, la marginalisation n'a pas d'expression explicite et elle est en général difficile à évaluer numériquement.

3.2.2 Borne récursive hybride

Comme nous l'avons vu précédemment, la borne de Cramér-Rao (3.27) n'a pas de forme analytique en général. Néanmoins, une approximation de $\mathbf{BCRH}_{\boldsymbol{\theta}_k,\boldsymbol{\varphi}}$ peut être formulée. En effet, nous pouvons étendre les résultats obtenus dans [TMN98] (bornes obtenues dans le contexte d'estimation purement bayésienne, i.e., $\boldsymbol{\varphi}$ connu) à l'estimation hybride ($\boldsymbol{\varphi}$ inconnu). L'idée est de décomposer la matrice d'information de Fisher pour l'estimation des paramètres $(\boldsymbol{\theta}_{0:k}; \boldsymbol{\varphi})$ en plusieurs blocs et d'extraire le sous-bloc correspondant à l'estimation de $(\boldsymbol{\theta}_k; \boldsymbol{\varphi})$. Pour alléger les notations, définissons $\mathbb{E}[.] \widehat{=} \mathbb{E}_{\mathbf{x}_{1:k},\boldsymbol{\theta}_{0:k}|\boldsymbol{\varphi}}[.]$ et la distribution jointe de $\mathbf{x}_{1:k}$ et de $\boldsymbol{\theta}_{0:k}$ par

$$f_k = f_{\mathbf{x}_{1:k},\boldsymbol{\theta}_{0:k}|\boldsymbol{\varphi}} (\mathbf{x}_{1:k}, \boldsymbol{\theta}_{0:k}|\boldsymbol{\varphi}),$$

alors d'après la relation (3.28)

$$f_{k+1} = f_k \times f_{\mathbf{x}_{k+1},\boldsymbol{\theta}_{k+1}|\boldsymbol{\varphi}} (\mathbf{x}_{k+1}|\boldsymbol{\theta}_{k+1}, \boldsymbol{\varphi}) \times f_{\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k,\boldsymbol{\varphi}} (\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k, \boldsymbol{\varphi}). \tag{3.29}$$

Pour tout estimateur $\widehat{\boldsymbol{\theta}}_{0:k}$ à partir des observations $\mathbf{x}_{1:k}$ et du modèle (3.25), nous pouvons borner l'EQM de $\widehat{\boldsymbol{\theta}}_{0:k}$ par la borne de Cramér-Rao hybride donnée par :

$$\mathbf{EQM}_{\widehat{\boldsymbol{\theta}}_{0:k}} \succeq \mathbf{J}_{0:k}^{-1} = \mathbb{E}^{-1} \left[-\frac{\partial^2 \ln f_k}{\partial (\boldsymbol{\theta}_{0:k}; \boldsymbol{\varphi}) \partial (\boldsymbol{\theta}_{0:k}; \boldsymbol{\varphi})^T} \right]. \tag{3.30}$$

Or l'inversion de la matrice d'information de Fisher $\mathbf{J}_{0:k}$ devient problématique, voire impossible, à mesure que k augmente. Fort heureusement, à l'instar de [TMN98], une décomposition de $\mathbf{J}_{0:k}$ en blocs nous permet d'évaluer de manière récursive les performances d'estimation de $(\boldsymbol{\theta}_k;\boldsymbol{\varphi})$. En effet, en se basant sur le modèle d'observation (3.25), l'idée est d'étudier les propriétés de récurrence de la matrice suivante :

$$\mathbf{J}_{0:k} = \mathbb{E}\left[-\frac{\partial^2 \ln f_k}{\partial\left(\boldsymbol{\theta}_{0:k};\boldsymbol{\varphi}\right)\partial\left(\boldsymbol{\theta}_{0:k};\boldsymbol{\varphi}\right)^T}\right] = \begin{bmatrix} \mathbf{A}_k & \mathbf{B}_k \\ \mathbf{B}_k^T & \mathbf{C}_k \end{bmatrix} \tag{3.31}$$

avec

$$\begin{aligned}
\mathbf{A}_k &= \mathbb{E}\left[-\frac{\partial^2 \ln f_k}{\partial\boldsymbol{\theta}_{0:k-1}\partial\boldsymbol{\theta}_{0:k-1}^T}\right], \\
\mathbf{B}_k &= \mathbb{E}_{\mathbf{x}_{1:k},\boldsymbol{\theta}_{0:k}|\boldsymbol{\varphi}}\left[-\frac{\partial^2 \ln f_k}{\partial\boldsymbol{\theta}_{0:k-1}\partial\left(\boldsymbol{\theta}_k;\boldsymbol{\varphi}\right)^T}\right], \\
\mathbf{C}_k &= \mathbb{E}_{\mathbf{x}_{1:k},\boldsymbol{\theta}_{0:k}|\boldsymbol{\varphi}}\left[-\frac{\partial^2 \ln f_k}{\partial\left(\boldsymbol{\theta}_k;\boldsymbol{\varphi}\right)\partial\left(\boldsymbol{\theta}_k;\boldsymbol{\varphi}\right)^T}\right].
\end{aligned}$$

Remarquons d'abord que l'EQM de $\widehat{\boldsymbol{\theta}}_k$ pour l'estimation des paramètres $(\boldsymbol{\theta}_k;\boldsymbol{\varphi})$ est minorée par le bloc inférieur droit de $\mathbf{J}_{0:k}^{-1}$. Donc en utilisant les propriétés d'inversion par blocs des matrices [Seb08, p293], nous avons

$$\mathbf{EQM}_{\widehat{\boldsymbol{\theta}}_k} \succeq \left(\mathbf{C}_k - \mathbf{B}_k^T\mathbf{A}_k^{-1}\mathbf{B}_k\right)^{-1} \widehat{=} \mathbf{J}_k^{-1}. \tag{3.32}$$

La matrice $\mathbf{A}_k$ est bien inversible puisqu'elle est définie positive. Au rang $k+1$, nous avons

$$\mathbf{J}_{0:k+1} = \mathbb{E}\left[-\frac{\partial^2 \ln f_{k+1}}{\partial\left(\boldsymbol{\theta}_{0:k+1};\boldsymbol{\varphi}\right)\partial\left(\boldsymbol{\theta}_{0:k+1};\boldsymbol{\varphi}\right)^T}\right].$$

Or la distribution f_{k+1} satisfait la relation de récurrence (3.29). Nous pouvons donc décomposer $\mathbf{J}_{0:k+1}$ en 9 blocs suivants

$$\mathbf{J}_{0:k+1} = \begin{bmatrix} \mathbf{A}_k & \mathbf{B}_k & \mathbf{0} \\ \mathbf{B}_k^T & \mathbf{C}_k + \mathbf{D}_{k+1} & \mathbf{E}_{k+1} \\ \mathbf{0} & \mathbf{E}_{k+1}^T & \mathbf{F}_{k+1} \end{bmatrix}$$

avec

$$\mathbf{D}_{k+1} = \mathbb{E}\left[-\frac{\partial^2 \ln f_{\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k,\boldsymbol{\varphi}}\left(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k,\boldsymbol{\varphi}\right)}{\partial\boldsymbol{\theta}_k\partial\boldsymbol{\theta}_k^T}\right],$$

$$\mathbf{E}_{k+1} = \mathbb{E}\left[-\frac{\partial^2 \ln f_{\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k,\boldsymbol{\varphi}}\left(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k,\boldsymbol{\varphi}\right)}{\partial\boldsymbol{\theta}_k\partial\left(\boldsymbol{\theta}_{k+1};\boldsymbol{\varphi}\right)^T}\right],$$

et

$$\mathbf{F}_{k+1} = -\mathbb{E}\left[\frac{\partial^2 \ln f_{\mathbf{x}_k,\boldsymbol{\theta}_k|\boldsymbol{\varphi}}\left(\mathbf{x}_{k+1}|\boldsymbol{\theta}_{k+1},\boldsymbol{\varphi}\right)}{\partial\left(\boldsymbol{\theta}_{k+1};\boldsymbol{\varphi}\right)\partial\left(\boldsymbol{\theta}_{k+1};\boldsymbol{\varphi}\right)^T}\right] - \mathbb{E}\left[\frac{\partial^2 \ln f_{\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k,\boldsymbol{\varphi}}\left(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k,\boldsymbol{\varphi}\right)}{\partial\left(\boldsymbol{\theta}_{k+1};\boldsymbol{\varphi}\right)\partial\left(\boldsymbol{\theta}_{k+1};\boldsymbol{\varphi}\right)^T}\right].$$

Les matrices $\mathbf{D}_{k+1}$, $\mathbf{E}_{k+1}$ et $\mathbf{F}_{k+1}$ dépendent uniquement des lois conditionnelles de $\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k,\boldsymbol{\varphi}$ et de $\mathbf{x}_{k+1}|\boldsymbol{\theta}_{k+1},\boldsymbol{\varphi}$, qui peuvent être déduites sans difficulté par le modèle d'observation (3.25). En

suivant le même raisonnement que la borne sur l'EQM de $(\boldsymbol{\theta}_k;\boldsymbol{\varphi})$, la nouvelle borne sur l'EQM de $(\boldsymbol{\theta}_{k+1};\boldsymbol{\varphi})$ est

$$\mathbf{EQM}_{\widehat{\boldsymbol{\theta}}_{k+1}} \succeq \mathbf{J}_{k+1}^{-1},$$

où la matrice d'information de Fisher $\mathbf{J}_{k+1}$ est obtenue par une inversion par bloc de l'élément inférieur droit de $\mathbf{J}_{0:k+1}$

$$\begin{aligned}
\mathbf{J}_{k+1} &= \mathbf{F}_{k+1} - \begin{bmatrix} \mathbf{0} & \mathbf{E}_{k+1}^T \end{bmatrix} \begin{bmatrix} \mathbf{A}_k & \mathbf{B}_k \\ \mathbf{B}_k^T & \mathbf{C}_k + \mathbf{D}_{k+1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{E}_{k+1} \end{bmatrix} \\
&= \mathbf{F}_{k+1} - \mathbf{E}_{k+1}^T \left[\mathbf{C}_k + \mathbf{D}_{k+1} - \mathbf{B}_k^T \mathbf{A}_k \mathbf{B}_k\right]^{-1} \mathbf{E}_{k+1} \\
&= \mathbf{F}_{k+1} - \mathbf{E}_{k+1}^T \left[\mathbf{D}_{k+1} - \mathbf{J}_k\right]^{-1} \mathbf{E}_{k+1}. \qquad (3.33)
\end{aligned}$$

Finalement, nous obtenons une relation de récursion entre $\mathbf{J}_{k+1}$ et $\mathbf{J}_k$, et la matrice de Fisher $\mathbf{J}_0$ est définie par

$$\mathbf{J}_0 = \mathbb{E}\left[-\frac{\partial^2 \ln f_{\boldsymbol{\theta}_0|\boldsymbol{\varphi}}(\boldsymbol{\theta}_0|\boldsymbol{\varphi})}{\partial(\boldsymbol{\theta}_0;\boldsymbol{\varphi})\,\partial(\boldsymbol{\theta}_0;\boldsymbol{\varphi})^T}\right]. \qquad (3.34)$$

Cette relation (3.33) nous permet donc de prédire l'EQM optimale asymptotique pour l'estimation de $(\boldsymbol{\theta}_{k+1};\boldsymbol{\varphi})$. Contrairement à la borne (3.27) qui nécessite la marginalisation de la loi jointe $\mathbf{x}_{1:k}, \boldsymbol{\theta}_{0:k}$ par rapport à $\boldsymbol{\theta}_{0:k-1}$, la borne récursive que nous avons développée dans le contexte d'estimation hybride a besoin uniquement des lois conditionnelles de $\boldsymbol{\theta}_j|\boldsymbol{\theta}_{j-1},\boldsymbol{\varphi}$ et de $\mathbf{x}_j|\boldsymbol{\theta}_j$, $1 \leq j \leq k$. Cette proposition a fait d'objet d'une publication dans *IEEE* Signal Processing Letters (ce document est joint en annexe B)..

3.3 La borne de Cramér-Rao déterministe avec erreurs de modèle : application à la problématique du Radar MIMO

Stimulés par les avancées théoriques en communication MIMO, les radars MIMO ont été beaucoup étudiés pour les problèmes de détection, de localisation pendant les dix dernières années [LS09]. Un radar MIMO est composé de plusieurs antennes émettrices et de plusieurs antennes réceptrices. Chaque émetteur transmet une onde électromagnétique qui, après réflexion sur les cibles, est collectée par les récepteurs. La comparaison entre les ondes émises et les ondes reçues permet de déduire les caractéristiques des cibles (distance, vitesse,...). Contrairement au réseau d'antennes à commande de phase transmettant une forme d'onde unique, les radars MIMO ont la possibilité d'émettre une forme d'onde différente pour chacune des antennes émettrices. Ainsi, l'utilisation des formes d'onde orthogonales limite les interférences et augmente les performances en terme d'estimation et de détection du système [HV10].

Les performances en estimation d'un tel système sont souvent étudiées dans les conditions classiques où la distribution des observations est correctement spécifiée. En effet, les algorithmes optimaux, comme ceux s'appuyant sur la méthode du maximum de vraisemblance, sont généralement basés sur les propriétés statistiques du modèle d'observations établies à partir des connaissances *a priori* du processus de propagation des ondes et du bruit de mesure. Cependant, le modèle d'observation supposé peut être légèrement différent du modèle réel en pratique. Plus particulièrement pour un radar MIMO, différentes erreurs de modèle peuvent intervenir à cause d'un défaut de positionnement des antennes [TRB+11], d'une erreur de modélisation du signal reçu [AN10] et/ou d'une erreur sur la propriété statistique du bruit [GFG14b, GFG14a]. Bien

qu'une phase de calibration puisse être mise en oeuvre pour corriger certains défauts, ce processus est souvent long et peut complexifier le modèle d'observation en ajoutant de nouveaux paramètres à estimer. De plus, il ne garantit pas l'élimination de tous les défauts du système. C'est pourquoi plusieurs solutions ont été proposées dans la littérature pour minimiser les impacts des erreurs sur les performances d'estimation. Les plus connues sont sans doute les méthodes d'estimation robuste [MMY06, ZKCM12] qui séparent généralement les données en deux parties : celles qui correspondent au modèle supposé et le reste (données corrompues).

Une autre méthode est d'étudier l'influence des erreurs de modèle sur l'étude des performances d'estimation. Ces modèles sont dits désadaptés ou mal spécifiés (Mismatched / Misspecified models) si la distribution réelle des données est différente de celle du modèle supposé. Dans ces cas, les propriétés de consistance et de normalité asymptotique de l'estimateur au sens du maximum de vraisemblance ont été établies par Huber et White [Hub67, Whi82]. De plus, comme cet estimateur est asymptotiquement efficace lorsque le modèle d'observation est correctement spécifié [LC03, section 6.3], nous pouvons nous attendre à ce qu'une petite erreur de modèle dégrade légèrement ses performances [GFG14a]. Récemment, des bornes déterministes ont été proposées dans le cas de modèles désadaptés [RH15]. Ces bornes, appelées bornes inférieures (déterministes) désadaptées / mal spécifiées (mismatched / Misspecified lower bounds) sont asymptotiquement atteignables, sous conditions de régularités, par le maximum de vraisemblance désadaptées (mismatched likelihood estimator) qui consiste à estimer les paramètres en se basant sur les données réelles et le modèle d'observation supposé [Hub67].

Le but de cette section est d'étudier, à l'aide de la théorie des bornes désadaptées, l'influence des erreurs de positionnement des antennes émettrices et réceptrices sur les performances d'estimation dans le contexte de radar MIMO. A la différence du travaux proposé dans [TRB+11], dans lequel le vecteur directionnel du réseau d'antennes est perturbé par un vecteur bruit aléatoire de densité de probabilité connue, nous considérons ici un scenario plus réaliste où le modèle d'observation supposé est simplement différent des données mesurées. De telle méthodologie a été récemment utilisée pour l'estimation de la matrice de diffusion en radar [GFG14a] et dans le contexte de traitement d'antenne [RH13].

3.3.1 Rappel sur les bornes de Cramér-Rao désadaptées

Nous définissons d'abord le contexte d'estimation dans lequel les bornes désadaptées peuvent être utiles : soit $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ un vecteur de paramètres inconnus non aléatoires à estimer à partir des observations $\mathbf{x} \in \mathbb{C}^{MN}$ et soit $g_{\mathbf{x};\boldsymbol{\theta}_0}$ la vraie densité de probabilité de $\mathbf{x}$. Cependant, dues aux imperfections du système, le modèle supposé est différent du modèle réel. Ceci signifie, en terme de densité, que la vraie densité de probabilité n'appartient pas à l'ensemble des distributions supposées $f_{\mathbf{y};\boldsymbol{\theta}}$, c'est à dire que $g_{\mathbf{x};\boldsymbol{\theta}_0} \notin \{f_{\mathbf{x};\boldsymbol{\theta}} | \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$. Si les erreurs de modèle sont faibles alors les estimateurs basés sur le modèle supposé restent pertinent. Dans ce contexte, la théorie des bornes désadaptées a été établies récemment dans [RH15]. Cette théorie s'appuie également sur l'inégalité de covariance et la borne de Cramér-Rao désadaptée est obtenue par l'inégalité suivante :

$$\mathbb{E}_{g_{\mathbf{x};\boldsymbol{\theta}_0}}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_A\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_A\right)^T\right] \succeq \mathbf{C}^{-1}\left(\boldsymbol{\theta}_A\right)\mathbf{J}\left(\boldsymbol{\theta}_A\right)\mathbf{C}^{-1}\left(\boldsymbol{\theta}_A\right), \tag{3.35}$$

où $\boldsymbol{\theta}_A$ est le point minimisant la divergence de Kullback-Leibler entre les densités $g_{\mathbf{x};\boldsymbol{\theta}_0}$ et $\{f_{\mathbf{x};\boldsymbol{\theta}} | \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, *i.e.*,

$$\boldsymbol{\theta}_A = \arg\min_{\boldsymbol{\theta}} \int_{\mathbb{C}^{MN}} g_{\mathbf{x};\boldsymbol{\theta}_0}(\mathbf{x}) \log\left(\frac{g_{\mathbf{x};\boldsymbol{\theta}_0}(\mathbf{x})}{f_{\mathbf{x};\boldsymbol{\theta}}(\mathbf{x})}\right) d\mathbf{x}.$$

Les termes du membre de droite [1] de l'équation (3.35) sont donnés par

$$\mathbf{C}\left(\boldsymbol{\theta}_A\right) = \mathbb{E}_{g_{\mathbf{x};\boldsymbol{\theta}_0}}\left[\left.\frac{\partial^2 \ln f_{\mathbf{x};\boldsymbol{\theta}}\left(\mathbf{x}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_A}\right]^{-1}, \tag{3.36}$$

et

$$\mathbf{J}\left(\boldsymbol{\theta}_A\right) = \mathbb{E}_{g_{\mathbf{x};\boldsymbol{\theta}_0}}\left[\left.\frac{\partial \ln f_{\mathbf{x};\boldsymbol{\theta}}\left(\mathbf{x}\right)}{\partial\boldsymbol{\theta}}\frac{\partial \ln f_{\mathbf{x};\boldsymbol{\theta}}\left(\mathbf{x}\right)}{\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_A}\right]. \tag{3.37}$$

Remarquons que le membre de gauche de (3.35) n'est pas l'EQM classique puisque $\boldsymbol{\theta}_A$ est généralement différent de la vraie valeur $\boldsymbol{\theta}_0$. Néanmoins, pour les estimateurs satisfaisant $\mathbb{E}_{g_{\mathbf{x};\boldsymbol{\theta}_0}}\left[\widehat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}_A$, nous avons

$$\mathbb{E}_{g_{\mathbf{x};\boldsymbol{\theta}_0}}\left[\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)^T\right] \succeq \mathbf{BCRD}$$

où la Borne de Cramér-Rao Désadaptée (BCRD) s'écrit

$$\mathbf{BCRD} = \mathbf{C}^{-1}\left(\boldsymbol{\theta}_A\right)\mathbf{J}\left(\boldsymbol{\theta}_A\right)\mathbf{C}^{-1}\left(\boldsymbol{\theta}_A\right) + \left(\boldsymbol{\theta}_A-\boldsymbol{\theta}_0\right)\left(\boldsymbol{\theta}_A-\boldsymbol{\theta}_0\right)^T. \tag{3.38}$$

De plus, des expressions analytiques de type Slepian-Bangs ont été proposées dans le cas d'observations gaussiennes à moyenne paramétrée [RH13]. Plus précisément, lorsque la distribution réelle des données est définie par $g_{\mathbf{x};\boldsymbol{\theta}_0} = \mathbb{C}\mathcal{N}\left(\mathbf{d}\left(\boldsymbol{\theta}_0\right),\mathbf{B}\right)$ et la distribution supposée des données est de la forme $f_{\mathbf{x};\boldsymbol{\theta}} = \mathbb{C}\mathcal{N}\left(\mathbf{m}\left(\boldsymbol{\theta}\right),\mathbf{R}\right)$ où les matrices $\mathbf{B}$ et $\mathbf{R}$ sont indépendantes de $\boldsymbol{\theta}$ et les vecteurs $\mathbf{d}\left(\boldsymbol{\theta}_0\right)$ et $\mathbf{m}\left(\boldsymbol{\theta}\right)$ sont différents pour tout $\boldsymbol{\theta}\in\boldsymbol{\Theta}$. Dans ce contexte, les matrices $\mathbf{C}$ et $\mathbf{J}$ données par l'équation (3.35) ont pour expressions analytiques :

$$\begin{aligned}\mathbf{C}_{ij}\left(\boldsymbol{\theta}\right) &= \frac{\partial^2\mathbf{m}^H\left(\boldsymbol{\theta}\right)}{\partial\theta_i\partial\theta_j}\mathbf{R}^{-1}\left(\mathbf{d}\left(\boldsymbol{\theta}_0\right)-\mathbf{m}\left(\boldsymbol{\theta}\right)\right)+\left(\mathbf{d}\left(\boldsymbol{\theta}_0\right)-\mathbf{m}\left(\boldsymbol{\theta}\right)\right)^H\mathbf{R}^{-1}\frac{\partial^2\mathbf{m}\left(\boldsymbol{\theta}\right)}{\partial\theta_i\partial\theta_j}\\ &\quad -\frac{\partial\mathbf{m}^H\left(\boldsymbol{\theta}\right)}{\partial\theta_i}\mathbf{R}^{-1}\frac{\partial\mathbf{m}\left(\boldsymbol{\theta}\right)}{\partial\theta_j}-\frac{\partial\mathbf{m}^H\left(\boldsymbol{\theta}\right)}{\partial\theta_j}\mathbf{R}^{-1}\frac{\partial\mathbf{m}\left(\boldsymbol{\theta}\right)}{\partial\theta i},\end{aligned} \tag{3.39}$$

et

$$\mathbf{J}_{ij}\left(\boldsymbol{\theta}\right) = \frac{\partial\mathbf{m}^H\left(\boldsymbol{\theta}\right)}{\partial\theta_i}\mathbf{R}^{-1}\mathbf{B}\left(\mathbf{R}^{-1}\right)^*\frac{\partial\mathbf{m}\left(\boldsymbol{\theta}\right)}{\partial\theta_j}+\frac{\partial\mathbf{m}^T\left(\boldsymbol{\theta}\right)}{\partial\theta_i}\mathbf{R}^{-1}\mathbf{B}^*\left(\mathbf{R}^{-1}\right)^*\frac{\partial\mathbf{m}^*\left(\boldsymbol{\theta}\right)}{\partial\theta_j}, \tag{3.40}$$

où * désigne l'opérateur complexe conjugué. Remarquons que si le modèle est correctement spécifié, c'est-à-dire $\mathbf{d}\left(\boldsymbol{\theta}_0\right) = \mathbf{m}\left(\boldsymbol{\theta}\right)$ et $\mathbf{B} = \mathbf{R}$ alors

$$\mathbf{C}\left(\boldsymbol{\theta}\right) = \mathbf{J}\left(\boldsymbol{\theta}\right) = 2\,\mathrm{Re}\left(\frac{\partial\mathbf{m}^H\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\mathbf{R}^{-1}\frac{\partial\mathbf{m}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^T}\right),$$

ce qui conduit à

$$\mathbf{C}\left(\boldsymbol{\theta}\right)\mathbf{J}^{-1}\left(\boldsymbol{\theta}\right)\mathbf{C}\left(\boldsymbol{\theta}\right) = 2\,\mathrm{Re}\left(\frac{\partial\mathbf{m}^H\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\mathbf{R}^{-1}\frac{\partial\mathbf{m}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^T}\right). \tag{3.41}$$

Nous reconnaissons la formule de Slepian-Bangs classique [Sle54, Ban71] pour une distribution gaussienne à moyenne paramétrée sans erreurs de modèle. La borne de Cramér-Rao désadaptée constitue donc une extension de la borne de Cramér-Rao déterministe pour des problèmes d'estimation contenant des erreurs de modèle.

1. Connue aussi sous le nom de "sandwich de Huber"

3.3.2 Application à l'estimation de la direction de départ et d'arrivée pour un système radar MIMO

Nous considérons un système radar MIMO composé de M antennes émettrices et de N antennes réceptrices spatialement séparées (non co-localisées). Les formes d'onde des M antennes d'émission sont supposées orthogonales entre elles et sont collectées par les récepteurs après réflexion sur une seule cible. Nous supposons que la cible se déplace à vitesse lente négligeable durant le temps d'observation (hypothèse de "slow moving target"). Dans ce cas, les données reçues, après un processus de filtrage adapté (distance-Doppler), sont modélisées par (voir [JLL09, équation (2)]) :

$$\mathbf{x} = \sqrt{K}\beta\mathbf{t}\left(\varphi_D\right) \otimes \mathbf{r}\left(\varphi_A\right) + \mathbf{b}, \tag{3.42}$$

où le vecteur $\mathbf{x} \in \mathbb{C}^{MN}$ représente les données traitées après filtrage adapté, le coefficient $K > 0$ est la puissance du signal, le coefficient $\beta \in \mathbb{C}$ correspond aux atténuations à cause des pertes durant le trajet de propagation et à la surface équivalente radar (radar cross section) et l'opérateur $\otimes$ est le produit de Kronecker. La direction d'arrivée de la cible par rapport au centre de phase des antennes réceptrice est noté φ_A et la direction de départ de la cible par rapport au centre de phase des antennes émettrices est noté φ_D (hypothèse de propagation en onde plane). Les vecteurs directionnels sont respectivement donnés par $\mathbf{t}\left(\varphi_D\right) = \left[e^{-j2\pi f_0\tau_1(\varphi_D)} \ \dots \ e^{-j2\pi f_0\tau_M(\varphi_D)}\right]^T$ et $\mathbf{r}\left(\varphi_A\right) = \left[e^{-j2\pi f_0\rho_1(\varphi_A)} \ \dots \ e^{-j2\pi f_0\rho_N(\varphi_A)}\right]^T$ pour le réseau d'émission et pour le réseau de réception. Les retards d'émission $\left\{\tau_j\left(\varphi_D\right)\right\}_{j=1,..,M}$ dépendent de la direction de départ de la cible et de la géométrie des antennes émettrices. Les retards de réception $\left\{\rho_j\left(\varphi_A\right)\right\}_{j=1,..,N}$ varient selon la direction d'arrivée de la cible et la géométrie des antennes réceptrices. La fréquence f_0 est la fréquence porteuse des ondes transmises. Enfin le bruit $\mathbf{b}$ est supposé gaussien complexe circulaire de moyenne nulle et de matrice de covariance $\sigma^2\mathbf{I}_{MN}$. Les paramètres inconnus à estimer sont concaténés dans un vecteur réel $\boldsymbol{\theta} = \left[\mathrm{Re}\left(\beta\right) \ \ \mathrm{Im}\left(\beta\right) \ \ \varphi_D \ \varphi_A\right]^T$. La densité de probabilité des observations du modèle s'écrit alors :

$$f_{\mathbf{x};\boldsymbol{\theta}}\left(\mathbf{x}\right) = \frac{1}{\pi^{MN}\sigma^2} e^{-\frac{1}{\sigma^2}\left\|\mathbf{x}-\sqrt{K}\beta\mathbf{t}(\varphi_D)\otimes\mathbf{r}(\varphi_A)\right\|^2}. \tag{3.43}$$

Basé sur ces hypothèses, l'estimateur au sens du maximum de vraisemblance est obtenu en cherchant la valeur de $\boldsymbol{\theta}$ maximisant la distribution $f_{\mathbf{x};\boldsymbol{\theta}}\left(\mathbf{x}\right)$ c'est-à-dire,

$$\widehat{\boldsymbol{\theta}}_{\mathrm{MV}} = \arg\max_{\boldsymbol{\theta}} f_{\mathbf{x};\boldsymbol{\theta}}\left(\mathbf{x}\right). \tag{3.44}$$

En raison des erreurs de modèles, les données mesurées $\mathbf{x}$ suivent une autre distribution $g_{\mathbf{x};\boldsymbol{\theta}_0} \notin \left\{f_{\mathbf{x};\boldsymbol{\theta}} | \boldsymbol{\theta} \in \boldsymbol{\Theta}\right\}$. De telles erreurs de modèles conduisent à une dégradation des performances de l'estimateur. Par conséquent les bornes classiques (basées sur la distribution supposée $f_{\mathbf{x};\boldsymbol{\theta}}$) ne sont plus asymptotiquement atteignables par l'EQM du maximum de vraisemblance (3.44). Nous allons donc étudier l'influence des ces erreurs sur les performances asymptotiques à l'aide de bornes désadaptées. Plus précisément, on considère le scénario où les positions des antennes sont différentes entre le modèle supposé et la vraie distribution des données. Cette dernière est définie par

$$g_{\mathbf{x};\boldsymbol{\theta}_0}\left(\mathbf{x}\right) = \frac{1}{\pi^{MN}\sigma^2} e^{-\frac{1}{\sigma^2}\left\|\mathbf{x}-\sqrt{K}\beta_0\widetilde{\mathbf{t}}(\varphi_{D_0})\otimes\widetilde{\mathbf{r}}(\varphi_{A_0})\right\|^2}. \tag{3.45}$$

où $\boldsymbol{\theta}_0 = \left[\mathrm{Re}\left(\beta_0\right) \ \ \mathrm{Im}\left(\beta_0\right) \ \ \varphi_{D_0} \ \varphi_{A_0}\right]^T$ contient les vraies valeurs des paramètres, $\widetilde{\mathbf{t}}\left(\varphi_{D_0}\right) = \left[e^{-j2\pi f_0\widetilde{\tau}_1\left(\varphi_{D_0}\right)} \ \dots \ e^{-j2\pi f_0\widetilde{\tau}_M\left(\varphi_{D_0}\right)}\right]^T$ et $\widetilde{\mathbf{r}}\left(\varphi_{A_0}\right) = \left[e^{-j2\pi f_0\widetilde{\rho}_1\left(\varphi_{A_0}\right)} \ \dots \ e^{-j2\pi f_0\widetilde{\rho}_N\left(\varphi_{A_0}\right)}\right]^T$ sont

les vrais vecteurs directionnels. Les retards exacts sont exprimés par $\left\{\widetilde{\tau}_j\left(\varphi_{D_0}\right)\right\}_{j=1,..,M}$ et $\left\{\widetilde{\rho}_j\left(\varphi_{A_0}\right)\right\}_{j=1,..,N}$ et dépendent des positions réelles des antennes.

3.3.2.1 Expression générale de la borne de Cramér-Rao désadaptée pour les erreurs de positions des antennes

Dans cette section, nous dérivons une expression générale de cette borne quelles que soient les erreurs de position des antennes. Rappelons que les paramètres inconnus à estimer sont $\boldsymbol{\theta} = [\mathrm{Re}\,(\beta)\ \ \mathrm{Im}\,(\beta)\ \ \varphi_D\ \varphi_A]$, le modèle supposé est donné par (3.42) et les données sont distribuées suivant la loi $g_{\mathbf{x};\boldsymbol{\theta}_0}(\mathbf{x})$ (équation (3.45)). En appliquant les formules (3.39) et (3.40) pour notre scénario, nous obtenons les expressions suivantes pour les matrices $\mathbf{C}(\boldsymbol{\theta})$ et $\mathbf{J}(\boldsymbol{\theta})$: les éléments diagonaux de $\mathbf{J}(\boldsymbol{\theta})$ sont exprimés par

$$\begin{aligned}
\mathbf{J}_{11}(\boldsymbol{\theta}) &= \mathbf{J}_{22}(\boldsymbol{\theta}) = \frac{2KMN}{\sigma^2}, \\
\mathbf{J}_{33}(\boldsymbol{\theta}) &= \frac{2KN\left\|\beta\right\|^2}{\sigma^2}\left\|\frac{d\mathbf{t}(\varphi_D)}{d\varphi_D}\right\|^2, \\
\mathbf{J}_{44}(\boldsymbol{\theta}) &= \frac{2KM\left\|\beta\right\|^2}{\sigma^2}\left\|\frac{d\mathbf{r}(\varphi_A)}{d\varphi_A}\right\|^2.
\end{aligned}$$

Puisque $\mathbf{J}(\boldsymbol{\theta})$ est une matrice symétrique pour tout $\boldsymbol{\theta}$, les éléments triangulaires inférieurs sont

$$\begin{aligned}
\mathbf{J}_{12}(\boldsymbol{\theta}) &= 0, \\
\mathbf{J}_{13}(\boldsymbol{\theta}) &= \frac{2KN}{\sigma^2}\mathrm{Re}\left(\beta^*\frac{d\mathbf{t}^H(\varphi_D)}{d\varphi_D}\mathbf{a}(\varphi_D)\right), \\
\mathbf{J}_{14}(\boldsymbol{\theta}) &= \frac{2KM}{\sigma^2}\mathrm{Re}\left(\beta^*\frac{d\mathbf{r}^H(\varphi_A)}{d\varphi_A}\mathbf{b}(\varphi_A)\right), \\
\mathbf{J}_{23}(\boldsymbol{\theta}) &= -\frac{2KN}{\sigma^2}\mathrm{Im}\left(\beta^*\frac{d\mathbf{t}^H(\varphi_D)}{d\varphi_D}\mathbf{r}(\varphi_D)\right), \\
\mathbf{J}_{24}(\boldsymbol{\theta}) &= -\frac{2KM}{\sigma^2}\mathrm{Im}\left(\beta^*\frac{d\mathbf{r}^H(\varphi_A)}{d\varphi_A}\mathbf{t}(\varphi_A)\right), \\
\mathbf{J}_{34}(\boldsymbol{\theta}) &= \frac{2K\left\|\beta\right\|^2}{\sigma^2}\mathrm{Re}\left(\frac{d\mathbf{t}^H(\varphi_D)}{d\varphi_D}\mathbf{t}(\varphi_D)\,\mathbf{r}^H(\varphi_A)\frac{d\mathbf{r}(\varphi_A)}{d\varphi_A}\right).
\end{aligned}$$

Avant de déterminer les éléments de la matrice $\mathbf{C}$, remarquons que cette matrice est simplifiable pour $\mathbf{B} = \mathbf{R} = \sigma^2\mathbf{I}_{MN}$:

$$\mathbf{C}_{ij}(\boldsymbol{\theta}) = \frac{\partial^2\mathbf{m}^H(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\mathbf{R}^{-1}\left(\mathbf{d}(\boldsymbol{\theta}_0) - \mathbf{m}(\boldsymbol{\theta})\right) + \left(\mathbf{d}(\boldsymbol{\theta}_0) - \mathbf{m}(\boldsymbol{\theta})\right)^H\mathbf{R}^{-1}\frac{\partial^2\mathbf{m}(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j} - \mathbf{J}_{ij}.$$

Par conséquent, les éléments diagonaux de la matrice $\mathbf{C}$ ont pour expressions analytiques :

$$\begin{aligned}
\mathbf{C}_{11}(\boldsymbol{\theta}) &= \mathbf{C}_{22}(\boldsymbol{\theta}) = -\mathbf{J}_{11}(\boldsymbol{\theta}), \\
\mathbf{C}_{33}(\boldsymbol{\theta}) &= \frac{2K}{\sigma^2}\mathrm{Re}\left(\beta_0\frac{d^2\mathbf{t}^H(\varphi_D)}{d\varphi_D^2}\widetilde{\mathbf{t}}(\varphi_{D_0})\,\mathbf{r}^H(\varphi_A)\,\widetilde{\mathbf{r}}(\varphi_{A_0})\right) - \frac{2KN}{\sigma^2}\mathrm{Re}\left(\beta^*\frac{d^2\mathbf{t}^H(\varphi_D)}{d\varphi_D^2}\mathbf{t}(\varphi_D)\right) - \mathbf{J}_{33}, \\
\mathbf{C}_{44}(\boldsymbol{\theta}) &= \frac{2K}{\sigma^2}\mathrm{Re}\left(\beta_0\mathbf{t}^H(\varphi_D)\,\widetilde{\mathbf{t}}(\varphi_{D_0})\,\frac{d^2\mathbf{r}^H(\varphi_A)}{d\varphi_A^2}\widetilde{\mathbf{r}}(\varphi_{A_0})\right) - \frac{2KM}{\sigma^2}\mathrm{Re}\left(\beta^*\frac{d^2\mathbf{r}^H(\varphi_A)}{d\varphi_A^2}\mathbf{r}(\varphi_A)\right) - \mathbf{J}_{44}.
\end{aligned}$$

Et, puisque $\mathbf{C}$ est également symétrique, ses éléments triangulaires inférieurs sont

$$\begin{aligned}
\mathbf{C}_{12} &= 0, \\
\mathbf{C}_{13} &= \frac{2K}{\sigma^2}\operatorname{Re}\left(\beta_0\frac{d\mathbf{t}^H(\varphi_D)}{d\varphi_D}\widetilde{\mathbf{t}}(\varphi_{D_0})\,\mathbf{r}^H(\varphi_A)\,\widetilde{\mathbf{t}}(\varphi_{A_0})\right) - \frac{4KN}{\sigma^2}\operatorname{Re}(\beta)\operatorname{Re}\left(\frac{d\mathbf{t}^H(\varphi_D)}{d\varphi_D}\mathbf{t}(\varphi_D)\right), \\
\mathbf{C}_{14} &= \frac{2K}{\sigma^2}\operatorname{Re}\left(\beta_0\mathbf{t}^H(\varphi_D)\,\widetilde{\mathbf{t}}(\varphi_{D_0})\,\frac{d\mathbf{r}^H(\varphi_A)}{d\varphi_A}\widetilde{\mathbf{r}}(\varphi_{A_0})\right) - \frac{4KM}{\sigma^2}\operatorname{Re}(\beta)\operatorname{Re}\left(\frac{d\mathbf{r}^H(\varphi_A)}{d\varphi_A}\mathbf{r}(\varphi_A)\right), \\
\mathbf{C}_{23} &= \frac{4KN}{\sigma^2}\operatorname{Im}(\beta)\operatorname{Im}\left(\frac{d\mathbf{t}^H(\varphi_D)}{d\varphi_D}\mathbf{t}(\varphi_D)\right) - \frac{2K}{\sigma^2}\operatorname{Im}\left(\beta_0\frac{d\mathbf{t}^H(\varphi_D)}{d\varphi_D}\widetilde{\mathbf{t}}(\varphi_{D_0})\,\mathbf{r}^H(\varphi_A)\,\widetilde{\mathbf{r}}(\varphi_{A_0})\right), \\
\mathbf{C}_{24} &= \frac{4KM}{\sigma^2}\operatorname{Im}(\beta)\operatorname{Im}\left(\frac{d\mathbf{r}^H(\varphi_A)}{d\varphi_A}\mathbf{r}(\varphi_A)\right) - \frac{2K}{\sigma^2}\operatorname{Im}\left(\beta_0\mathbf{t}^H(\varphi_D)\,\widetilde{\mathbf{t}}(\varphi_{D_0})\,\frac{d\mathbf{r}^H(\varphi_A)}{d\varphi_A}\widetilde{\mathbf{r}}(\varphi_{A_0})\right), \\
\mathbf{C}_{34} &= \frac{2K}{\sigma^2}\operatorname{Re}\left(\beta_0\frac{d\mathbf{t}^H(\varphi_D)}{d\varphi_D}\widetilde{\mathbf{t}}(\varphi_{D_0})\,\frac{d\mathbf{r}^H(\varphi_A)}{d\varphi_A}\widetilde{\mathbf{r}}(\varphi_{A_0})\right) \\
&\quad -\frac{4K}{\sigma^2}\beta^*\operatorname{Re}\left(\frac{d\mathbf{t}^H(\varphi_D)}{d\varphi_D}\mathbf{t}(\varphi_D)\right)\operatorname{Re}\left(\frac{d\mathbf{r}^H(\varphi_A)}{d\varphi_A}\mathbf{r}(\varphi_A)\right).
\end{aligned}$$

3.3.2.2 Simulation

Nous considérons que les réseaux d'antennes émettrices et réceptrices sont supposées linéaires uniformes de distance inter-capteur $\frac{\lambda_0}{2}$ où λ_0 est la longueur d'onde du signal. Par conséquent, le modèle d'observation supposé est défini par l'équation (3.42) où les vecteurs directionnels sont respectivement

$$\mathbf{t}(\varphi_D) = \left[1\ e^{-j\pi\sin(\varphi_D)}\ \cdots\ e^{-j(M-1)\pi\sin(\varphi_D)}\right]$$

et

$$\mathbf{r}(\varphi_A) = \left[1\ e^{-j\pi\sin(f_A)}\ \cdots\ e^{-j(N-1)\pi\sin(f_A)}\right]$$

pour les émetteurs et les récepteurs. Or en réalité, les antennes ne sont pas parfaitement uniforme, les vrais vecteurs directionnels sont donnés par

$$\widetilde{\mathbf{t}}(\varphi_D) = \left[1\ e^{-j\pi(1+e_1)\sin(\varphi_D)}\ \cdots\ e^{-j(M-1+e_{M-1})\pi\sin(\varphi_D)}\right]$$

et

$$\widetilde{\mathbf{r}}(\varphi_A) = \left[1\ e^{-j\pi(1+r_1)\sin(\varphi_A)}\ \cdots\ e^{-j(N-1+r_{N-1})\pi\sin(\varphi_A)}\right]$$

où $\mathbf{e} = [e_1\ e_2\ \cdots\ e_{M-1}]^T$ et $\mathbf{r} = [r_1\ r_2\ \cdots\ r_{N-1}]^T$ désigne les écarts de position entre le modèle supposé et le vrai modèle. Les paramètres de simulation sont les suivants : $M = N = 8$, $K = 1$, $\beta_0 = 0.1\,(1+j)$, $\varphi_{D_0} = 0.3$ rad, $\varphi_{A_0} = 0.2$ rad et les erreurs $\mathbf{e}$ et $\mathbf{r}$ sont générées suivant une loi gaussienne centrée de variance 0.2. Les données $\mathbf{x}$ sont ensuite générées aléatoirement suivant la loi de $g_{\mathbf{x};\boldsymbol{\theta}_0}$ définie par (3.45). Les paramètres inconnues sont estimées à l'aide de l'estimateur au sens du maximum de vraisemblance désadapté défini par

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_{\mathrm{MV}} &= \underset{\beta\in\mathbb{C},\ \varphi_D\in\left[-\frac{\pi}{2};\frac{\pi}{2}\right[,\varphi_A\in\left[-\frac{\pi}{2};\frac{\pi}{2}\right[}{\arg\max}\ f_{\mathbf{x};\boldsymbol{\theta}}(\mathbf{x}) \\
&= \underset{\beta\in\mathbb{C},\ \varphi_D\in\left[-\frac{\pi}{2};\frac{\pi}{2}\right[,\varphi_A\in\left[-\frac{\pi}{2};\frac{\pi}{2}\right[}{\arg\min}\ \left\|\mathbf{x} - \sqrt{K}\beta\mathbf{t}(\varphi_D)\otimes\mathbf{r}(\varphi_A)\right\|^2. \qquad (3.46)
\end{aligned}$$

Nous pouvons remarquer que cet estimateur est séparable. En effet, notons $\mathbf{m}(\varphi_D, \varphi_A) = \sqrt{K}\mathbf{t}(\varphi_D) \otimes \mathbf{r}(\varphi_A)$ et $\Pi_{\mathbf{m}}(\varphi_D, \varphi_A) = \frac{\mathbf{m}(\varphi_D,\varphi_A)\mathbf{m}^H(\varphi_D,\varphi_A)}{\|\mathbf{m}(\varphi_D,\varphi_A)\|^2}$ le projecteur orthogonal sur $\mathbf{m}(\varphi_D, \varphi_A)$. On constate que

$$\begin{aligned}\|\mathbf{x} - \beta\mathbf{m}(\varphi_D, \varphi_A)\|^2 &= \left\|\left(\Pi_{\mathbf{m}}(\varphi_D, \varphi_A) + \Pi_{\mathbf{m}}^{\perp}(\varphi_D, \varphi_A)\right)(\mathbf{x} - \beta\mathbf{m}(\varphi_D, \varphi_A))\right\|^2 \\ &= \|\Pi_{\mathbf{m}}(\varphi_D, \varphi_A)\mathbf{x} - \beta\mathbf{m}(\varphi_D, \varphi_A)\|^2 + \left\|\Pi_{\mathbf{m}}^{\perp}(\varphi_D, \varphi_A)\mathbf{x}\right\|^2,\end{aligned}$$

où $\Pi_{\mathbf{m}}^{\perp}(\varphi_D, \varphi_A) = \mathbf{I} - \Pi_{\mathbf{m}}(\varphi_D, \varphi_A)$. Comme

$$\|\Pi_{\mathbf{m}}(\varphi_D, \varphi_A)\mathbf{x} - \beta\mathbf{m}(\varphi_D, \varphi_A)\|^2 = \|\mathbf{m}(\varphi_D, \varphi_A)\|^2 \left|\frac{\mathbf{m}(\varphi_D, \varphi_A)^H}{\|\mathbf{m}(\varphi_D, \varphi_A)\|^2}\mathbf{x} - \beta\right|^2$$

donc

$$\beta = \frac{\mathbf{m}(\varphi_D, \varphi_A)^H}{\|\mathbf{m}(\varphi_D, \varphi_A)\|^2}\mathbf{x}.$$

Un algorithme simplifié de cet estimateur est donc donné par

$$\widehat{\boldsymbol{\theta}}_{MV} = \begin{cases} [\widehat{\varphi}_D\ \widehat{\varphi}_A]^T = \underset{\varphi_D \in [-\frac{\pi}{2};\frac{\pi}{2}[, \varphi_A \in [-\frac{\pi}{2};\frac{\pi}{2}[}{\arg\min} \left\|\Pi_{\mathbf{m}}^{\perp}(\varphi_D, \varphi_A)\mathbf{x}\right\|^2 \\ \widehat{\beta} = \frac{\mathbf{m}^H(\widehat{\varphi}_D,\widehat{\varphi}_A)}{\|\mathbf{m}(\widehat{\varphi}_D,\widehat{\varphi}_A)\|^2}\mathbf{x} \end{cases} \tag{3.47}$$

De plus,

$$\begin{aligned}\left\|\Pi_{\mathbf{m}}^{\perp}(\varphi_D, \varphi_A)\mathbf{x}\right\|^2 &= \|\mathbf{x}\|^2 - \|\Pi_{\mathbf{m}}(\varphi_D, \varphi_A)\mathbf{x}\|^2 \\ &= \|\mathbf{x}\|^2 - \frac{\left|\mathbf{m}^H(\varphi_D, \varphi_A)\mathbf{x}\right|^2}{MN}.\end{aligned} \tag{3.48}$$

Ce qui donne

$$\widehat{\boldsymbol{\theta}}_{MV} = \begin{cases} [\widehat{\varphi}_D\ \widehat{\varphi}_A]^T = \underset{\varphi_D \in [-\frac{\pi}{2};\frac{\pi}{2}[, \varphi_A \in [-\frac{\pi}{2};\frac{\pi}{2}[}{\arg\max} \left|\mathbf{m}^H(\varphi_D, \varphi_A)\mathbf{x}\right|^2 \\ \widehat{\beta} = \frac{\mathbf{m}^H(\widehat{\varphi}_D,\widehat{\varphi}_A)}{KMN}\mathbf{x} \end{cases}. \tag{3.49}$$

Numériquement, les valeurs estimées de $\widehat{\varphi}_D$ et de $\widehat{\varphi}_A$ sont obtenues en maximisant le critère $\left|\mathbf{m}^H(\varphi_D, \varphi_A)\mathbf{x}\right|^2$ avec un pas de discrétisation de $\frac{\pi}{2^{12}}$.

Sur les figures (3.3) et (3.4), nous comparons l'EQM empirique du maximum de vraisemblance, obtenue à l'aide de 1000 tirages Monte-Carlo, à la borne de Cramér-Rao désadaptée donné par l'expression (3.38), à la vraie borne de Cramér-Rao sans biais définie par $\mathbf{BCR} = \mathbb{E}_{g_{\mathbf{x};\boldsymbol{\theta}_0}}^{-1}\left[\frac{\partial \ln g_{\mathbf{x};\boldsymbol{\theta}}(\mathbf{x})}{\partial\boldsymbol{\theta}}\frac{\partial \ln g_{\mathbf{x};\boldsymbol{\theta}}(\mathbf{x})}{\partial\boldsymbol{\theta}^T}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]$ et à la Borne de Cramér-Rao Biaisée (BCRB) qui prend en compte le biais de l'estimateur. Sous la condition de modèle désadapté, ces deux figures montrent que la seule borne atteignable asymptotiquement par l'estimateur MV désadapté est la borne de BCRD. Notons également que la borne de Cramér-Rao sans biais est la borne la plus basse pour ces deux paramètres, ce qui signifie que l'EQM des estimateurs peut diminuer si le modèle est correctement spécifié. Remarquons finalement que la borne de Cramér-Rao biaisée et la borne de Cramér-Rao désadaptée montrent un phénomène de décrochement [RB74, RFL+08, CGQL08]. En comparant la BCR et la BCRB, l'origine de ce phénomène semble être due au biais de l'estimateur.

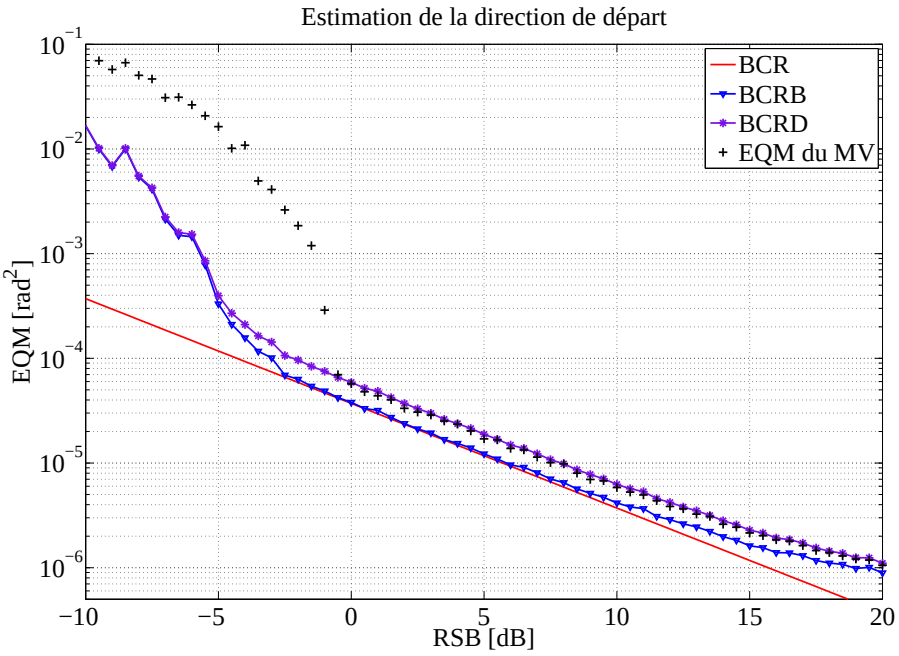

FIGURE 3.3 – Comparaison entre l'EQM du MV et les bornes pour l'estimation de la direction de départ en fonction du RSB

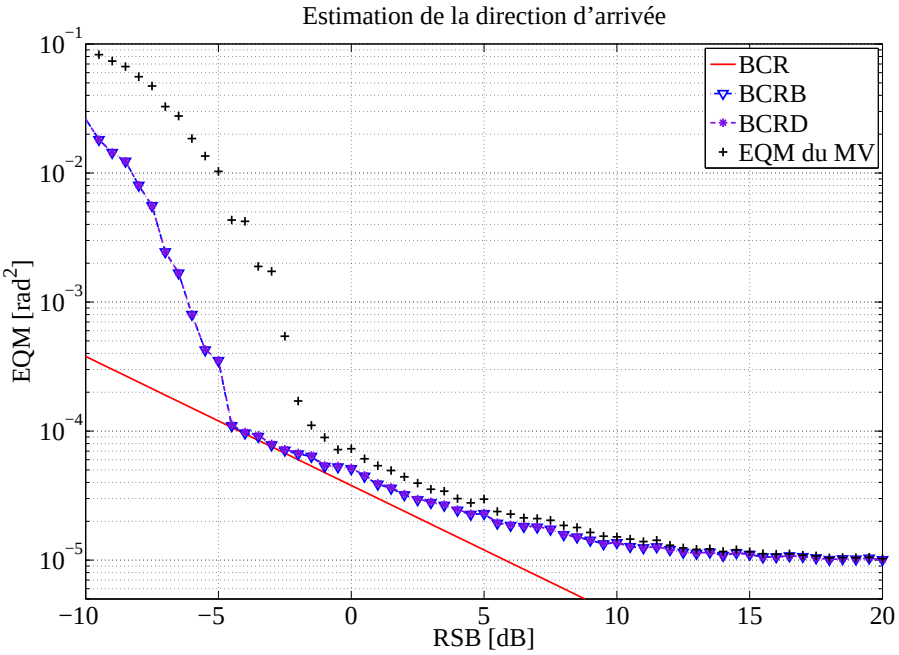

FIGURE 3.4 – Comparaison entre l'EQM du MV et les bornes pour l'estimation de la direction d'arrivée en fonction du RSB

Chapitre 4

Caractérisation des estimateurs en termes de résolution et de précision

4.1 Introduction

La résolution limite caractérise la distance minimale entre les paramètres d'intérêt au-dessus de laquelle la séparation de ces paramètres devient possible. Comme nous l'avons exposé précédemment, les observations sont modélisées, en estimation statistique, par des variables aléatoires dont leurs caractéristiques dépendent des paramètres d'intérêt. Ce comportement aléatoire peut rendre difficile la formulation et la validation d'un critère de résolution limite. En effet, il est possible (avec une probabilité faible) que deux paramètres d'intérêt, bien éloignés entre eux, soient associés à des jeux d'observations qui ne permet pas de les résoudre [1]. De même, il est (peu) probable que deux paramètres d'intérêt, aussi proches que possible, soient associés à des jeux d'observations permettant de les distinguer. Pour ces raisons, il est nécessaire d'introduire la notion de "résolution limite statistique" qui définit de manière probabiliste la résolution des paramètres, en tenant compte des propriétés aléatoires des observations. Dans le contexte déterministe (qui sera aussi notre cas d'éude), les critères de résolution proposés dans la littérature peuvent être regroupés en quatre familles :

1. La première approche de résolution statistique a été proposée par Cox en 1973 [Cox73]. Cette approche est basée sur les critères de minimisation appelés "*null spectrum*" exploitant une fonction de "pseudo-spectre" (voir figure 4.1) notée $C(\mathbf{y}, \theta)$. En effet, si θ_1 et θ_2 sont les paramètres à estimer à partir des observations $\mathbf{y}$, ce critère affirme que θ_1 et θ_2 sont résolvables si les moyennes des valeurs du pseudo-spectre aux points θ_1 et θ_2 sont inférieures à la moyenne de la valeur du pseudo-spectre au point $\frac{\theta_1+\theta_2}{2}$ c'est-à-dire $\mathbb{E}_{\mathbf{y}|\theta_1,\theta_2}\left[C(\mathbf{y}, \theta_1)\right] < \mathbb{E}_{\mathbf{y}|\theta_1,\theta_2}\left[C\left(\mathbf{y}, \frac{\theta_1+\theta_2}{2}\right)\right]$ et $\mathbb{E}_{\mathbf{y}|\theta_1,\theta_2}\left[C(\mathbf{y}, \theta_2)\right] < \mathbb{E}_{\mathbf{y}|\theta_1,f_2}\left[C\left(\mathbf{y}, \frac{\theta_1+\theta_2}{2}\right)\right]$. Ces deux critères, sous certaines conditions, sont équivalents à une étude de la convexité de la fonction $\mathbb{E}_{\mathbf{y}|\theta_1,\theta_2}\left[C(\mathbf{y}, \theta)\right]$ au point $\theta = \frac{\theta_1+\theta_2}{2}$ [SD85]. Cependant, ces critères dépendent du pseudo-spectre associé à l'algorithme utilisé $C(\mathbf{y}, \theta)$.
2. Une deuxième approche basée sur la borne de Cramér-Rao a été proposée dans [Lee93, YB92]. Par la définition de borne, elle exprime les performances minimales en terme d'estimation paramétrique et a l'avantage d'être indépendante de l'algorithme choisi. Par

1. au sens séparer, distinguer

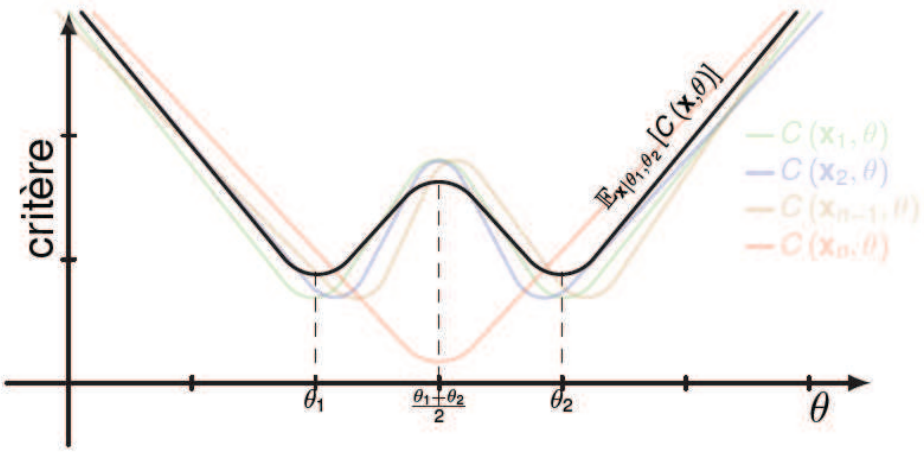

FIGURE 4.1 – Cas de sources résolues selon le critère de Cox.

conséquent, elle est utilisée pour définir une résolution limite. Cette approche définit que deux paramètres θ_1 et θ_2 sont résolus si $|\theta_1 - \theta_2| \geq 2\max\left(\sqrt{BCR(\theta_1)}, \sqrt{BCR(\theta_2)}\right)$. Cependant, elle ne prend pas en compte la dépendance statistique entre les deux paramètres. Ainsi, une amélioration a été proposée par Smith [Smi05], sous la forme suivante $|\theta_1 - \theta_2|^2 \geq BCR(|\theta_1 - \theta_2|)$.

3. Une troisième approche a été proposée par [SM04,AW08] basée sur la théorie de détection. Cette approche reformule le problème de résolvabilité par un test d'hypothèses binaire où l'hypothèse $\mathcal{H}_0$ suppose que les données sont générées à partir d'un modèle contenant un seul paramètre d'intérêt et l'hypothèse $\mathcal{H}_1$ suppose que les données suivent un modèle avec deux paramètres d'intérêt. Cette approche détermine le seuil de résolution en fonction de la probabilité de fausse alarme P_{fa} et de la probabilité de détection P_d en utilisant le test du maximum de vraisemblance généralisé.

4. La dernière approche est basée sur la probabilité de résolution qui permet à la fois de résoudre les paramètres d'intérêt et d'imposer la précision de l'estimation. Elle consiste à déterminer la probabilité P_{res} telle que les estimées appartiennent à des ensembles disjoints. Par conséquent, les paramètres sont résolus avec une probabilité P_{res} et des intervalles de confiance sont établis pour les valeurs estimées. Ce concept a été initialement introduit par Oh et Kashyap [OK91] dans le cadre de paramètres scalaires puis étendu par Clark [Cla95] dans le cas d'estimation de paramètres multidimensionnels, deux à deux décorrélés, suivant une loi gaussienne.

Le but de ce chapitre est d'étendre la dernière approche pour des problèmes d'estimation de paramètres multidimensionnels, non nécessairement gaussiens, ni décorrélés, et de déterminer la probabilité de résolution en utilisant des approximations d'ensembles disjoints.

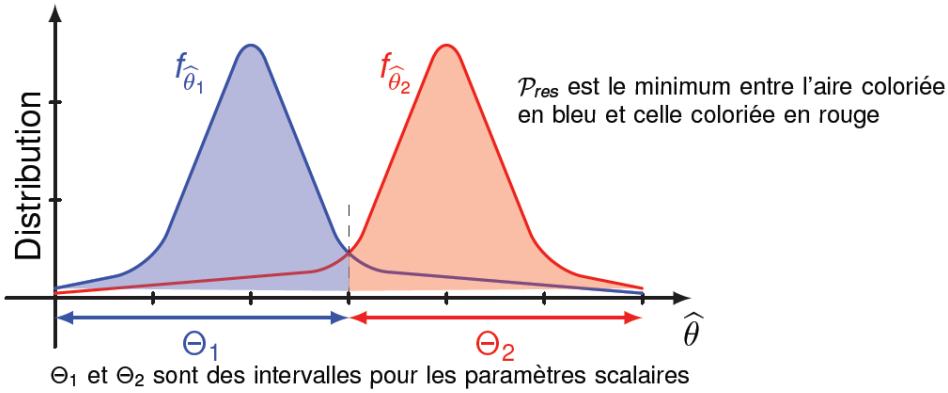

FIGURE 4.2 – Probabilité de résolution pour les paramètres scalaires

4.2 Rappels et définitions

Soit $\theta_1 \in \mathbb{R}$ et $\theta_2 \in \mathbb{R}$, $\theta_1 \neq \theta_2$ deux paramètres déterministes de même nature[2] à estimer à partir des observations $\mathbf{x} \in \Omega$. Soit $\widehat{\boldsymbol{\theta}} = \left[\widehat{\theta}_1\ \widehat{\theta}_2\right]^T$ un estimateur des paramètres θ_1 et θ_2. Dans [OK91], on dit que l'estimateur $\widehat{\boldsymbol{\theta}}$ peut résoudre les paramètres θ_1 et θ_2 avec une probabilité de résolution P_{res} si

$$P_{res} = \min\left(\Pr\left(\widehat{\theta}_1 - \theta_1 \in]-\Delta;\Delta[\right), \Pr\left(\widehat{\theta}_2 - \theta_2 \in]-\Delta;\Delta[\right)\right) \tag{4.1}$$

où $\Delta = \frac{|\theta_1 - \theta_2|}{2}$ est choisi tel que l'intersection entre les intervalles $]\theta_1 - \Delta; \theta_1 + \Delta[$ et $]\theta_2 - \Delta; \theta_2 + \Delta[$ est nulle. Les valeurs estimées $\widehat{\theta}_1$ et $\widehat{\theta}_2$ sont bien distinctes et restreintes à leur intervalle respectif avec une probabilité d'au moins P_{res}.

L'extension aux cas de paramètres vectoriels n'est pas unique et dépend de la forme des domaines délimitant les estimées de chaque paramètres. En effet, il existe déjà une approche proposée par Clark [Cla95] pour des estimateurs distribués suivant une loi gaussienne où les domaines délimitant les estimées sont des hyper-ellipses. Dans la suite, $\left\{\boldsymbol{\theta}_m^0 \in \mathbb{R}^q\right\}_{m=1,\cdots,p}$ désignera un ensemble de vecteurs de paramètres d'intérêt à estimer à partir des observations $\mathbf{x}$ par l'estimateur $\left\{\widehat{\boldsymbol{\theta}}_m \in \mathbb{R}^q\right\}_{m=1,\cdots,p}$. A la différence de [Cla95], nous proposons ici une définition de la probabilité de résolution dans le cas de l'estimation de paramètres multidimensionnels, non nécessairement gaussiens, ni décorrélés, basée sur des domaines hyper-rectangles ; puis nous établissons une relation d'ordre de la probabilité entre les domaines hyper-rectangles et les domaines hyper-ellipses afin d'obtenir une méthode d'évaluation de P_{res}.

2. Par exemples : des angles en localisation de sources, des fréquences en analyse spectrale.

4.3 Une approche pour la résolution multidimensionnelle

4.3.1 Approximation d'ensembles disjoints

Soit $\{\boldsymbol{\varepsilon}_m^+\}_{m=1,\cdots,p}$ et $\{\boldsymbol{\varepsilon}_m^-\}_{m=1,\cdots,p}$ un ensemble de vecteurs réels tel que $\boldsymbol{\varepsilon}_m^+ \in \mathbb{R}^{q+}$ et $\boldsymbol{\varepsilon}_m^- \in \mathbb{R}^{q+}$, $1 \leq m \leq p$. On définit un hyper-rectangle, noté $R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$, $1 \leq m \leq p$, par

$$R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right) = \left\{\boldsymbol{\theta}_m \in \mathbb{R}^q | \forall 1 \leq j \leq q,\ \left\{\boldsymbol{\theta}_m - \boldsymbol{\theta}_m^0\right\}_j \in \left]-\left\{\boldsymbol{\varepsilon}_m^-\right\}_j ; \left\{\boldsymbol{\varepsilon}_m^+\right\}_j\right[\right\}, \tag{4.2}$$

où $\{\boldsymbol{\theta}_m\}_j$ désigne la $j^{ième}$ composante du vecteur $\boldsymbol{\theta}_m$. Donc $R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^+, \boldsymbol{\varepsilon}_m^-\right)$ délimite un espace d'estimation pour le paramètre $\boldsymbol{\theta}_m^0$. Cette définition nous servira pour introduire la précision et la résolvabilité d'un estimateur. L'équation (4.2) peut être réécrite sous la forme :

$$R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right) = \left\{\boldsymbol{\theta}_m \in \mathbb{R}^q | \forall 1 \leq j \leq q,\ \left|\left\{\boldsymbol{\theta}_m - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j\right| \leq \{\overline{\boldsymbol{\varepsilon}}_m\}_j \right\}, \tag{4.3}$$

avec $d\{\boldsymbol{\varepsilon}_m\}_j = \{\boldsymbol{\varepsilon}_m^+ - \boldsymbol{\varepsilon}_m^-\}_j$ et $\{\overline{\boldsymbol{\varepsilon}}_m\}_j = \frac{\{\boldsymbol{\varepsilon}_m^+ + \boldsymbol{\varepsilon}_m^-\}_j}{2}$. On peut alors aisément établir la relation d'ordre suivante

$$E_s^1\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right) \subset R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right) \subset E_s^q\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right), \tag{4.4}$$

avec

$$E_s^r\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right) = \left\{\boldsymbol{\theta}_m \in \mathbb{R}^q | \sum_{j=1}^{q} \left|\frac{\left\{\boldsymbol{\theta}_m - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j}{\{\overline{\boldsymbol{\varepsilon}}_m\}_j}\right|^s \leq r \right\}, \tag{4.5}$$

où $r = 1$ ou q, et $s > 0$. $E_s^r\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$ est une généralisation de l'hyper-ellipse correspondant au cas $s = 2$. La démonstration de (4.4) est basée sur l'idée suivante : $\forall \boldsymbol{\theta} \in R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$, nous avons

$$\forall 1 \leq j \leq q,\ \left|\left\{\boldsymbol{\theta} - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j\right| \leq \{\overline{\boldsymbol{\varepsilon}}_m\}_j \tag{4.6}$$

ce qui est équivalent à

$$\left|\frac{\left\{\boldsymbol{\theta} - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j}{\{\overline{\boldsymbol{\varepsilon}}_m\}_j}\right| \leq 1, \text{ pour tout } \{\overline{\boldsymbol{\varepsilon}}_m\}_j \neq 0. \tag{4.7}$$

Donc, si on élève (4.7) à la puissance s, nous avons

$$\forall 1 \leq j \leq q,\ \left|\frac{\left\{\boldsymbol{\theta} - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j}{\{\overline{\boldsymbol{\varepsilon}}_m\}_j}\right|^s \leq 1, \tag{4.8}$$

conduisant à

$$\sum_{j=1}^{q} \left|\frac{\left\{\boldsymbol{\theta} - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j}{\{\overline{\boldsymbol{\varepsilon}}_m\}_j}\right|^s \leq q.$$

On a finalement $\boldsymbol{\theta} \in E_s^q\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$. D'où $R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right) \subset E_s^q\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$. De même, si $\boldsymbol{\theta} \in E_s^1\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$, alors nous avons

$$\sum_{j=1}^{q} \left|\frac{\left\{\boldsymbol{\theta} - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j}{\{\overline{\boldsymbol{\varepsilon}}_m\}_j}\right|^s \leq 1, \tag{4.9}$$

impliquant la relation suivante

$$\forall 1 \leq j \leq q, \quad \left| \frac{\left\{\boldsymbol{\theta} - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j}{\left\{\overline{\boldsymbol{\varepsilon}}_m\right\}_j} \right| \leq 1. \tag{4.10}$$

Cette dernière est équivalent à

$$\forall 1 \leq j \leq q, \quad \left| \left\{\boldsymbol{\theta} - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j \right| \leq \left\{\overline{\boldsymbol{\varepsilon}}_m\right\}_j, \tag{4.11}$$

donc $\boldsymbol{\theta} \in R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$. On en déduit que $E_s^1\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right) \subset R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$. Par conséquent, nous retrouvons bien l'encadrement (4.4). De plus $\forall \boldsymbol{\theta} \in E_s^q\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$, on a l'équivalence suivante

$$\sum_{j=1}^{q} \left| \frac{\left\{\boldsymbol{\theta} - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j}{\left\{\overline{\boldsymbol{\varepsilon}}_m\right\}_j} \right|^s \leq q \Leftrightarrow \sum_{j=1}^{q} \left| \frac{\left\{\boldsymbol{\theta} - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j}{q^{\frac{1}{s}}\left\{\overline{\boldsymbol{\varepsilon}}_m\right\}_j} \right|^s \leq 1$$

Or $q^{\frac{1}{s}} \underset{s\to\infty}{\to} 1$, nous avons donc

$$\sum_{j=1}^{q} \left| \frac{\left\{\boldsymbol{\theta} - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j}{\left\{\overline{\boldsymbol{\varepsilon}}_m\right\}_j} \right|^s - \sum_{j=1}^{q} \left| \frac{\left\{\boldsymbol{\theta} - \boldsymbol{\theta}_m^0 - \frac{d\boldsymbol{\varepsilon}_m}{2}\right\}_j}{q^{\frac{1}{s}}\left\{\overline{\boldsymbol{\varepsilon}}_m\right\}_j} \right|^s \underset{s\to\infty}{\to} 0.$$

Ce qui implique que

$$R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right) = \lim_{s\to\infty} E_s^q\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right) = \lim_{s\to\infty} E_s^1\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right).$$

D'après ces propriétés, tout hyper-rectangle $R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$ peut être borné par les hypervolumes $E_s^1\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$ et $E_s^q\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$ qui convergent vers une même limite lorsque s tend vers l'infini. Une illustration de ces propriétes est montré dans la figure (4.3) pour différentes valeurs de $s = 1, 2, 4, 16$ en dimension $q = 2$ (dans le plan) avec $\boldsymbol{\varepsilon}_m^- = \boldsymbol{\varepsilon}_m^+ = [1, 4]^T$. Les courbes en trait plein correspondent aux ensembles $E_s^1\left(\mathbf{0}, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$ et les courbes en pointillées correspondent aux ellipses $E_s^2\left(\mathbf{0}, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$. Comme $\boldsymbol{\varepsilon}_m^- = \boldsymbol{\varepsilon}_m^+$, les ensembles $E_s^1\left(\mathbf{0}, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$ et $E_s^2\left(\mathbf{0}, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$ admettent deux axes de symétrie : l'axe des abscisses et l'axe des ordonnées. Donc on illustre ici le quart des volumes $E_s^1\left(\mathbf{0}, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$ et $E_s^2\left(\mathbf{0}, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$. On remarque que $E_s^1\left(\mathbf{0}, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$ croît et $E_s^2\left(\mathbf{0}, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)$ décroît lorsque s augmente et converge vers un rectangle délimité en abscisse par 1 et en ordonné par 4. Cette propriété d'encadrement des hyper-rectangles permettra dans la suite de déterminer des valeurs approchées de la précision d'estimation ou du seuil de résolution limite.

4.3.2 Précision d'estimation

La qualité première d'un estimateur est sa précision. Elle peut être mesurée par la fonction

$$\mathcal{O}_{\boldsymbol{\theta}_m^0}\left(\widehat{\boldsymbol{\theta}}_m, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right) = \Pr\left(\widehat{\boldsymbol{\theta}}_m \in R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)\right) \tag{4.12}$$

pour tout estimateur $\widehat{\boldsymbol{\theta}}_m$ du paramètre $\boldsymbol{\theta}_m^0$. C'est la probabilité que les estimées appartiennent à un hyper-rectangle. A probabilité de résolution P_{res} fixée, plus l'hyper-rectangle est petit, plus l'estimateur $\widehat{\boldsymbol{\theta}}_m$ sera précis. Par la propriété d'encadrement (4.4), nous avons

$$\Pr\left(\widehat{\boldsymbol{\theta}}_m \in E_s^1\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)\right) \leq \mathcal{O}_{\boldsymbol{\theta}_m^0}\left(\widehat{\boldsymbol{\theta}}_m, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right) \leq \Pr\left(\widehat{\boldsymbol{\theta}}_m \in E_s^q\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)\right). \tag{4.13}$$

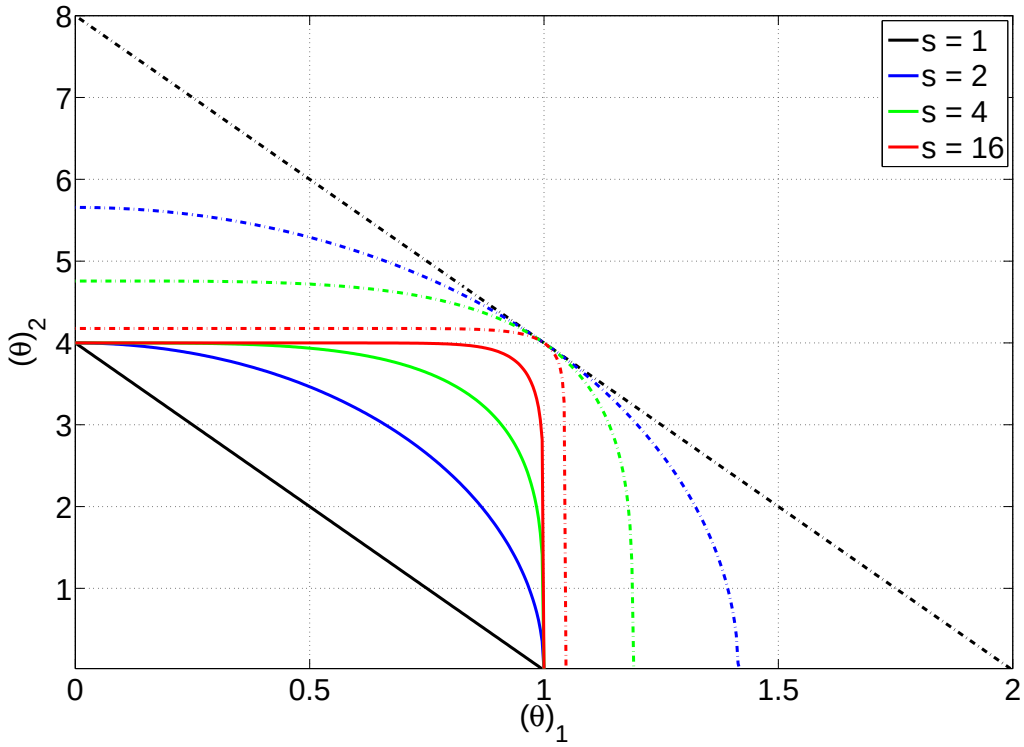

FIGURE 4.3 – Encadrement du rectangle $R\left(\boldsymbol{\theta}^0, \boldsymbol{\varepsilon}^-, \boldsymbol{\varepsilon}^+\right)$ par des ellipses $E_s^1\left(\boldsymbol{\theta}^0, \boldsymbol{\varepsilon}^-, \boldsymbol{\varepsilon}^+\right)$ en trait plein et $E_s^2\left(\boldsymbol{\theta}^0, \boldsymbol{\varepsilon}^-, \boldsymbol{\varepsilon}^+\right)$ en pointillé pour $s \in \{1, 2, 4, 16\}$ et $\boldsymbol{\varepsilon}^- = \boldsymbol{\varepsilon}^+ = [1\ 4]^T$

L'extension de ce critère pour l'estimation d'un ensemble de paramètres $\boldsymbol{\theta}^0 = \left[\left(\boldsymbol{\theta}_1^0\right)^T \cdots \left(\boldsymbol{\theta}_p^0\right)^T\right]^T$ par des estimateurs de type $\widehat{\boldsymbol{\theta}} = \left[\widehat{\boldsymbol{\theta}}_1^T \cdots \widehat{\boldsymbol{\theta}}_p^T\right]^T$ est définie par

$$\mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) = \Pr\left(\bigcap_{m=1}^{p}\left[\widehat{\boldsymbol{\theta}}_m \in R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)\right]\right). \tag{4.14}$$

où $\boldsymbol{\Xi}^{-+} = \left[\left(\boldsymbol{\varepsilon}_1^{-+}\right)^T \cdots \left(\boldsymbol{\varepsilon}_p^{-+}\right)^T\right]^T$. L'équation (4.14) définit la probabilité que l'estimateur $\widehat{\boldsymbol{\theta}}$ appartienne aux domaines définis par les ensembles $\left\{R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)\right\}_{m=1\ldots p}$. Elle équivaut à

$$\mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) = \Pr\left(\widehat{\boldsymbol{\theta}} \in R\left(\boldsymbol{\theta}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)\right), \tag{4.15}$$

où $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^{pq}$ et $R\left(\boldsymbol{\theta}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$ est un hyper-rectangle dans $\mathbb{R}^{pq}$. En appliquant l'encadrement (4.13) dans (4.15), on obtient

$$\Pr\left(\widehat{\boldsymbol{\theta}} \in E_s^1\left(\boldsymbol{\theta}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)\right) \leq \mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) \leq \Pr\left(\widehat{\boldsymbol{\theta}} \in E_s^{pq}\left(\boldsymbol{\theta}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)\right), \tag{4.16}$$

où $E_s^1\left(\boldsymbol{\theta}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$ et $E_s^{pq}\left(\boldsymbol{\theta}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$ sont des hyper-ellipses dans $\mathbb{R}^{pq}$. Cette relation (4.16) sera utile pour évaluer numériqument la précision de l'estimateur $\mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$.

4.3.3 Probabilité de résolution

Puisque les données sont aléatoires, il est nécessaire de donner un critère de résolution probabiliste. L'estimateur $\widehat{\boldsymbol{\theta}}$ peut résoudre les sources avec une probabilité de résolution P_{res} si

$$P_{res} = \mathcal{O}_{\boldsymbol{\theta}^0}\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) = \Pr\left(\bigcap_{m=1}^{p}\left[\widehat{\boldsymbol{\theta}}_m \in R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)\right]\right), \tag{4.17}$$

où $\{\boldsymbol{\varepsilon}_m^-\}_{m=1,\cdots,p}$ et $\{\boldsymbol{\varepsilon}_m^+\}_{m=1,\cdots,p}$ sont soumis aux contraintes

$$\forall i \neq j,\ R\left(\boldsymbol{\theta}_i^0, \boldsymbol{\varepsilon}_i^-, \boldsymbol{\varepsilon}_i^+\right) \cap R\left(\boldsymbol{\theta}_j^0, \boldsymbol{\varepsilon}_j^-, \boldsymbol{\varepsilon}_j^+\right) = \varnothing. \tag{4.18}$$

La résolution est totale pour $\widehat{\boldsymbol{\theta}}$ s'il existe un ensemble d'hyper-rectangles $\left\{R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)\right\}_{m=1,\cdots,p}$ soumis aux contraintes (4.18) tel que $P_{res} = 1$. Cependant, le support de la distribution des estimées est généralement non borné. Par conséquent, la probabilité P_{res} est généralement inférieure à 1. Elle dépend de plusieurs facteurs : le nombre d'observations, le rapport signal à bruit ou encore les volumes considérés $\left\{R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)\right\}_{m=1,\cdots,p}$. La résolution limite consiste à chercher le plus large ensemble $\left\{R\left(\boldsymbol{\theta}_m^0, \boldsymbol{\varepsilon}_m^-, \boldsymbol{\varepsilon}_m^+\right)\right\}_{m=1,\cdots,p}$ disjoint soumis aux contraintes (4.18) pour une probabilité P_{res} fixée.

4.3.4 Cas de distributions gaussiennes

L'évaluation analytique de (4.17) est complexe sauf si l'on dispose de l'expression analytique de la fonction de répartition de l'estimateur $\widehat{\boldsymbol{\theta}}$. Cependant, cette fonction de répartition n'est généralement pas accessible et nous avons au mieux une expression explicite de la densité de probabilité de $\widehat{\boldsymbol{\theta}}$. Une méthode alternative pour évaluer P_{res} est de calculer les encadrements proposés par (4.16). Notamment si l'estimateur est gaussien, c'est-à-dire $\widehat{\boldsymbol{\theta}} \sim \mathcal{N}\left(\boldsymbol{\theta}^0 + \mathbf{m}\left(\boldsymbol{\theta}^0\right), \mathbf{C}\left(\boldsymbol{\theta}^0\right)\right)$ où $\boldsymbol{\theta}^0$ est la vraie valeur des paramètres, $\mathbf{m}\left(\boldsymbol{\theta}^0\right)$ est le biais de l'estimateur $\widehat{\boldsymbol{\theta}}$ et $\mathbf{C}\left(\boldsymbol{\theta}^0\right)$ sa matrice de covariance. En utilisant l'encadrement (4.16) où le domaine hyper-ellipse peut être écrit sous la forme

$$\begin{aligned}
E_2^{\lambda}\left(\boldsymbol{\theta}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) &= \left\{\boldsymbol{\theta} \in \mathbb{R}^{pq} \middle| \left\|diag^{-1}\left(\overline{\boldsymbol{\Xi}}\right)\left(\boldsymbol{\theta} - \boldsymbol{\theta}^0 - \frac{d\boldsymbol{\Xi}}{2}\right)\right\|_2^2 < \lambda\right\} \\
E_s^{\lambda}\left(\boldsymbol{\theta}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) &= \left\{\boldsymbol{\theta} \in \mathbb{R}^{pq} \middle| \left\|diag^{-1}\left(\overline{\boldsymbol{\Xi}}\right)\left(\boldsymbol{\theta} - \boldsymbol{\theta}^0 - \frac{d\boldsymbol{\Xi}}{2}\right)\right\|_2^s < \lambda\right\}
\end{aligned}$$

avec $\overline{\boldsymbol{\Xi}} = \frac{\boldsymbol{\Xi}^- + \boldsymbol{\Xi}^+}{2}$, $d\boldsymbol{\Xi} = \frac{\boldsymbol{\Xi}^+ - \boldsymbol{\Xi}^-}{2}$ et $diag\left(\overline{\boldsymbol{\Xi}}\right)$ qui transforme le vecteur $\overline{\boldsymbol{\Xi}}$ en une matrice diagonale dont les éléments diagonaux sont ceux du vecteur $\overline{\boldsymbol{\Xi}}$, on peut démontrer que

$$\Pr\left(\widehat{\boldsymbol{\theta}} \in E_2^{\lambda}\left(\boldsymbol{\theta}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)\right) = F_{Q_{pq}}\left(\lambda; \left\|diag^{-1}\left(\overline{\boldsymbol{\Xi}}\right)\left(\mathbf{m}\left(\boldsymbol{\theta}^0\right) - \frac{d\boldsymbol{\Xi}}{2}\right)\right\|_2^2, \boldsymbol{\Sigma}\right), \tag{4.19}$$

où $F_{Q_{pq}}\left(x; \mu, \boldsymbol{\nu}\right)$ est une extension de loi de khi deux à pq degrés de liberté de paramètre de non centralité μ [SK61] et de variances $\boldsymbol{\nu} = \left[\sigma_1^2 \cdots \sigma_{pq}^2\right]^T$ où $\left\{\sigma_i^2\right\}_{i=1..pq}$ sont les valeurs propres de la matrice $diag^{-1}\left(\overline{\boldsymbol{\Xi}}\right)\mathbf{C}\left(\boldsymbol{\theta}^0\right)diag^{-1}\left(\overline{\boldsymbol{\Xi}}\right)$. Remarquons que dans le cas où l'estimateur $\widehat{\boldsymbol{\theta}}$ est non

biaisé $\mathbf{m}\left(\boldsymbol{\theta}^0\right) = \mathbf{0}$ et l'intervalle d'erreur admissible est symétrique et centré en $\boldsymbol{\theta}^0$, *i.e.*, $d\boldsymbol{\Xi} = \mathbf{0}$, alors la fonction de répartition s'écrit sous la forme explicite suivante

$$F_{Q_{pq}}\left(\lambda; 0, \boldsymbol{\nu}\right) = \frac{\left(\frac{\lambda}{2}\right)^{\frac{pq}{2}}}{\pi^{\frac{pq}{2}}\sqrt{\prod\limits_{i=1}^{pq}\sigma_i^2}}\sum_{k=0}^{\infty}\frac{\omega_k\left(-\frac{\lambda}{2}\right)^k}{\Gamma\left(\frac{pq}{2}+k+1\right)} \tag{4.20}$$

avec

$$\omega_k = \sum_{i_1+i_2+\cdots+i_{pq}=k}\frac{\Gamma\left(i_1+\frac{1}{2}\right)\cdots\Gamma\left(i_n+\frac{1}{2}\right)}{i_1!\cdots i_n!\sigma_1^{i_1}\cdots\sigma_{pq}^{i_{pq}}}, \tag{4.21}$$

où $\Gamma\left(.\right)$ est la fonction gamma. La forme générale de l'expression de $F_{Q_{pq}}$ est donné dans [SK61]. Une méthode alternative numérique permettant d'évaluer $F_{Q_{pq}}\left(\lambda; \mu, \boldsymbol{\nu}\right)$ est d'utiliser sa densité de probabilité définie par

$$f_{Q_{pq}}\left(\lambda; \mu, \boldsymbol{\nu}\right) = f_\chi\left(\lambda; \mu_1, \sigma_1^2\right) * \cdots * f_\chi\left(\lambda; \mu_{pq}, \sigma_{pq}^2\right) \tag{4.22}$$

où $*$ est le produit de convolution et $f_\chi\left(.\right)$ désigne la densité de probabilité de la variable aléatoire χ donnée par [Mui09]

$$f_\chi\left(\lambda; \mu_i, \sigma_i^2\right) = \frac{e^{-\frac{\lambda+\mu_i}{2\sigma_i^2}}}{2\sigma_i^2}\left(\frac{\lambda\mu_i}{\sigma_i^4}\right)^{-\frac{1}{4}} I_0\left(\sqrt{\frac{\lambda\mu_i}{\sigma_i^4}}\right) \tag{4.23}$$

où $I_0\left(.\right)$ est la fonction de Bessel modifiée de première espèce et d'ordre 0. Notons que l'encadrement établi dans (4.16) devient de plus en plus étroit lorsque s augmente. A notre connaissance, il n'existe pas dans la littérature une méthode d'évaluation de $\Pr\left(\widehat{\boldsymbol{\theta}} \in E_s^\lambda\left(\boldsymbol{\theta}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)\right)$ lorsque $s > 2$. Néanmoins, nous pouvons toujours déterminer directement P_{res} en intégrant numériquement sur le volume hyper-rectangle la distribution des estimées lorsque cette dernière est connue. Pour les distributions gaussiennes multivariées, la probabilité de résolution P_{res} peut être évaluée numériquement à l'aide de l'algorithme proposé par [Gen92].

4.4 Application à la localisation de sources

4.4.1 Performance asymptotique du maximum de vraisemblance conditionnel

La localisation et la séparabilité des sources sont des problèmes courants en traitement d'antennes, motivées au départ par des usages militaires puis par des applications civiles telles que le positionnement par satellites ou la téléphonie mobile. Considérons le problème d'estimation suivant : on souhaite localiser P sources lointaines à l'aide d'un réseau de N antennes. Le modèle d'observation est [Van02]

$$\mathbf{x}\left(t\right) = \mathbf{A}\left(\boldsymbol{\theta}^0\right)\mathbf{s}\left(t\right) + \mathbf{b}\left(t\right), \tag{4.24}$$

où $\mathbf{x}\left(t\right) \in \mathbb{C}^N$ contient les données collectées par le réseau d'antennes à l'instant t, $\boldsymbol{\theta}^0 = \left[\theta_1^0\ \theta_2^0\ \cdots\ \theta_P^0\right]^T \in \mathbb{C}^P$ est le vecteur qui contient les directions d'arrivée des sources, $\mathbf{A}\left(\boldsymbol{\theta}\right) = \left[\mathbf{a}\left(\theta_1^0\right)\ \mathbf{a}\left(\theta_2^0\right)\ \cdots\ \mathbf{a}\left(\theta_P^0\right)\right] \in \mathbb{C}^{N\times P}$ est la matrice directionnelle, $\mathbf{a}\left(.\right)$ est le vecteur directionnel, $\mathbf{s}\left(t\right) \in \mathbb{R}^P$ représente les signaux émis par les sources et $\mathbf{b}\left(t\right) \in \mathbb{C}^N$ est le bruit de mesure. Les paramètres d'intérêt sont les directions d'arrivée des sources $\boldsymbol{\theta}^0$, qui sont supposées inconnues

et non aléatoires. Le bruit $\mathbf{b}(t)$ est aléatoire suivant une loi supposée gaussienne complexe, circulaire de moyenne nulle et de matrice de covariance $\sigma^2\mathbf{I}_N$. Dans le cas du modèle conditionnel (le vecteur $\mathbf{s}(t)$ est inconnu et non aléatoire), il est montré dans [RFCL06] que l'estimateur du maximum de vraisemblance déterministe suit asymptotiquement une loi gaussienne à fort rapport signal sur bruit. Donc, pour une série de T mesures où les instants d'échantillonnage sont définis par le vecteur $\mathbf{t} = [t_1\ t_2\ \cdots\ t_T]^T$, nous avons [RFCL06, eq (26)]

$$\widehat{\boldsymbol{\theta}}_{MV} \sim \mathcal{N}\left(\boldsymbol{\theta}^0, \mathbf{BCR}\right), \tag{4.25}$$

avec

$$\mathbf{BCR} = \frac{\sigma^2}{2T}\left[\mathrm{Re}\left(\mathbf{D}^H\mathbf{P}_{\mathbf{A}^\perp}\mathbf{D} \odot \widehat{\mathbf{S}}^T\right)\right]^{-1},$$

où les expressions intervenant dans la **BCR** sont définies par :

$$\mathbf{D} = \left[\left.\frac{d\mathbf{a}(\theta)}{d\theta}\right|_{\theta=\theta_1^0} \quad \left.\frac{d\mathbf{a}(\theta)}{d\theta}\right|_{\theta=\theta_2^0} \quad \cdots \quad \left.\frac{d\mathbf{a}(\theta)}{d\theta}\right|_{\theta=\theta_P^0}\right], \tag{4.26}$$

$$\mathbf{P}_{\mathbf{A}^\perp} = \mathbf{I}_N - \mathbf{A}\left(\boldsymbol{\theta}^0\right)\left(\mathbf{A}\left(\boldsymbol{\theta}^0\right)\mathbf{A}^H\left(\boldsymbol{\theta}^0\right)\right)^{-1}\mathbf{A}\left(\boldsymbol{\theta}^0\right), \tag{4.27}$$

et

$$\widehat{\mathbf{S}} = \frac{1}{T}\sum_{i=1}^{T}\mathbf{s}(t_i)\,\mathbf{s}^H(t_i). \tag{4.28}$$

4.4.2 Simulations

Le but de cette section est de comparer le critère de résolution proposé avec ceux exhibés dans la littérature. Considérons le cas d'une antenne linéaire uniforme composée de $N = 32$ capteurs. Supposons que les deux sources sont étroitement espacées de $\delta = \theta_1^0 - \theta_2^0 = \frac{1}{kN}$ avec le paramètre $k \in \{2, 4, 8, 16, 32\}$. Le vecteur directionnel s'écrit $\mathbf{a}(\theta) = \left[1\ e^{j2\pi\sin(\theta)}\ \cdots\ e^{j2\pi(N-1)\sin(\theta)}\right]^T$ et l'amplitude $\mathbf{s}(t) = \sqrt{\frac{RSB}{N}}\,[1\ 1]^T$. Nous considérons des hyper-rectangles (dégénérant en des segments dans le cas présent) symétriques de largeur $\boldsymbol{\Xi}^- = \boldsymbol{\Xi}^+ = \left[\frac{\delta}{2}\ \frac{\delta}{2}\right]^T$ pour P_{res} (4.17) tels que les estimées pour chaque paramètre appartiennent à des intervalles disjoints. Il s'agit de trouver le RSB minimal permettant de résoudre les sources avec une probabilité de résolution $P_{res} = 0.95$ pour une séparation angulaire de δ donnée. Comme nous l'avons mentionné à la section (4.3.4), l'évaluation de P_{res} pour une distribution gaussienne est numériquement lourde et il est plus simple de la borner par l'encadrement (4.16). Ces bornes pour $s = 2$ peuvent être obtenues en intégrant la densité de probabilité (4.22) qui s'écrit simplement dans ce cas

$$f_{Q_2}\left(\lambda; 0, \boldsymbol{\sigma}^2\right) = \frac{e^{-\frac{\lambda}{4}\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)}}{2\sigma_1\sigma_2} I_0\left(\frac{\lambda}{4}\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)\right), \tag{4.29}$$

où $I_0(.)$ est la fonction de Bessel modifiée de première espèce et d'ordre 0. Nous montrons dans la figure (4.4), un exemple d'encadrement de P_{res} permettant de résoudre les sources distantes de $\delta = \frac{1}{128}\ rad$ en fonction du RSB. Conformément à l'intuition, plus le RSB augmente, plus la probabilité de résolution P_{res} augmente par encadrement par ces deux bornes. Les deux bornes permettent de déterminer l'intervalle de RSB, $[RSB_{min}, RSB_{max}]$ contenant RSB_{res},

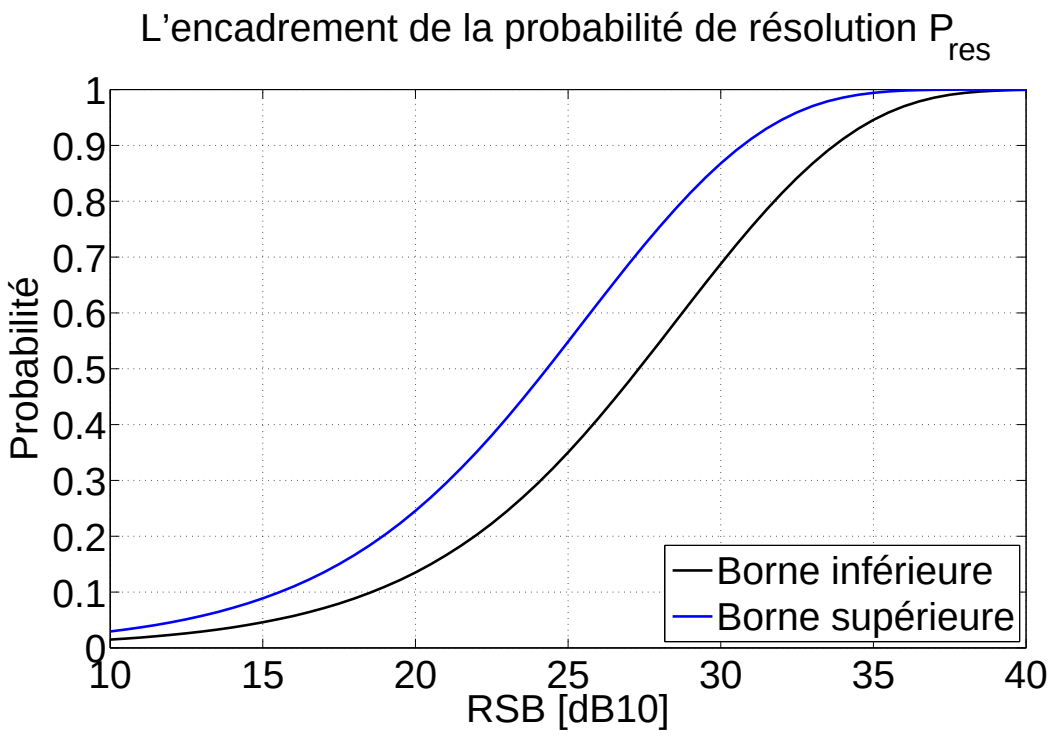

FIGURE 4.4 – Encadrement de la probabilité de résolution en fonction du RSB à l'aide de la relation (4.16) avec $s = 2$.

c'est à dire le RSB garantissant la résolution des 2 sources pour une probabilité P_{res} donnée (Par exemple, $RSB_{res} \in [32, 35]$dB si $P_{res} = 0.95$), le RSB_{min} étant fourni par la borne supérieure. La figure (4.5) compare les différents critères de résolution. Plus précisement, nous donnons les RSB minimaux requis, en fonction de la résolution limite δ, pour le critère de résolution de Lee-Yau-Bressler (noté LYB) [Lee93, YB92], pour le critère de Smith (noté S) [Smi05], pour le critère de Shahram-Milanfar (noté SM) [SM04] obtenu avec une probabilité de détection $P_d = 0.95$ et une probabilité de fausse alarme $P_d = 0.1$, pour le critère de Amar-Weiss (noté AW) [AW08] obtenu avec une probabilité de détection $P_d = 0.95$ et pour les bornes inférieures et supérieures de notre critère de résolution avec une probabilité de résolution de 0.95 (notées respectivement BI $P = 0.95$ et BS $P = 0.95$). Nous remarquons que, parmi tous les critères mentionnés, le critère proposé est le plus exigeant dans le sens où il nécessite un RSB important. Contrairement aux critères de résolution classiques qui se focalisent uniquement sur la résolvabilité du problème, notre critère permet non seulement de résoudre les paramètres mais également d'évaluer la précision de l'estimateur en restreignant les valeurs estimées dans les hypervolumes susmentionnés. Il permet donc de caractériser les performances en termes de résolution et de précision d'un estimateur en présence de plusieurs paramètres. Ceci a fait l'objet d'une publication dans *IEEE* Transactions on Signal Processing. Ce papier est joint en annexe C.

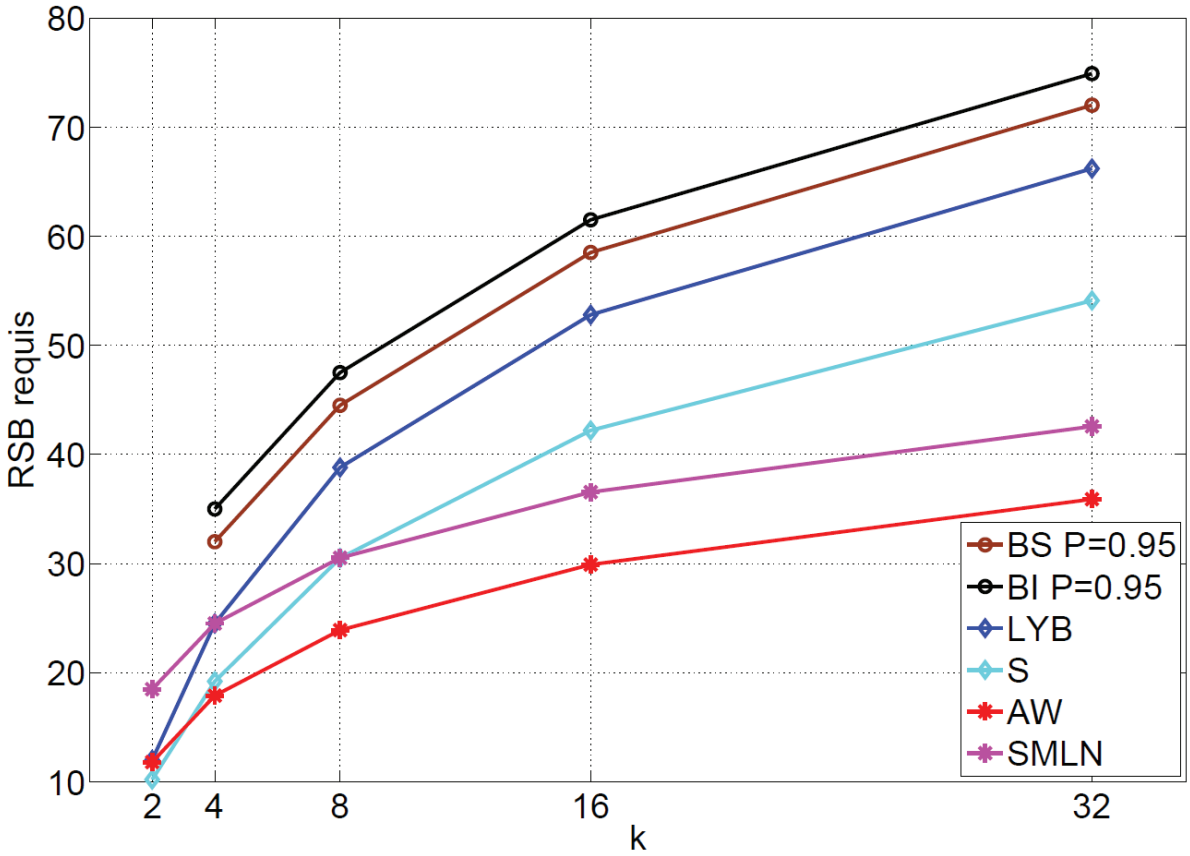

FIGURE 4.5 – Le RSB minimal requis pour la résolution de deux sources proches séparées de $\delta = \frac{1}{kN}$ avec $N = 32$ et $k \in \{2, 4, 8, 16, 32\}$

Chapitre 5

Conclusion et Perspectives

A travers les nombreux développements historiques, force est de constater que la mesure de performance en estimation la plus polyvalente d'un point de vue applicatif est l'EQM pour laquelle des grandes familles de bornes inférieures ont été mises en évidence, quelle que soit la nature des paramètres : déterministes, aléatoires ou hybrides.
La borne inférieure de l'EQM la plus applicative est clairement la borne de Cramér-Rao qui bénéficie non seulement d'une simplicité de formulation (en particulier, au travers de la formule de Slepian-Bang [Sle54, Ban71] et de ses extensions au cas de distributions symétriques elliptiques complexes [BA13, GG13, PR10]) mais également d'extensions particulièrement bien adaptées à l'analyse des performances en estimation : la prise en compte des contraintes d'égalités ou encore ses différentes formes récursives pour les systèmes non linéaires. Bien entendu, la contre-partie est qu'il s'agit d'une borne inférieure qui n'est informative qu'en zone asymptotique. En effet, elle ne permet pas de prédire le seuil de décrochement qui délimite la zone de fonctionnement optimal des estimateurs et qui est en général défini, en le majorant par des valeurs, *a priori*, de bon sens en terme de RSB (par exemple $> 40dB$) et/ou du nombre d'observations indépendantes (par exemple > 100).
Une première étape afin d'améliorer la valeur informative des bornes inférieures de l'EQM est de passer de la borne de Cramér-Rao aux "Large Error Bounds" en utilisant des points-tests dont le principal mérite est la capacité à prendre en compte le seuil de décrochement. Avec le durcissement des cahiers des charges (temps d'acquisitions réduits, environnements électromagnétiques perturbés, etc.), les estimateurs sont désormais amenés à travailler dans une région de plus en plus proche de cette limite, voire en dessous. En conséquence, la prédiction de la valeur du RSB ou du nombre d'observations pour laquelle cette rupture brutale des performances apparaît est fondamentale dans l'étude des performances en estimation. Nos premiers résultats obtenus pour les familles de bornes inférieures hybrides de Barankin-Weiss-Weinstein [RGC$^+$13, RGC$^+$15b] et de Barankin-Ziv-Zakai [RGC$^+$14] tendent à montrer qu'il s'agit dans ce domaine, à moyen terme, principalement d'une recherche académique portant sur la formulation plutôt que sur l'implémentation, en raison d'un coût de calcul encore rédhibitoire pour la plupart des applications pratiques. Dans cette perspective, il serait pertinent :

1. pour une utilisation pratique, de formuler le calcul de la borne Barankin-Weiss-Weinstein hybride [RGC$^+$15b] pour des observations gaussiennes à moyenne et/ou covariance paramétrée dans le cas général (nous ne sommes parvenus à caractériser qu'un sous-ensemble de ces modèles d'observations à savoir le cas gaussien à moyenne paramétrée).
2. d'étendre nos travaux sur les bornes inférieures hybrides de Barankin-Ziv-Zakai [RGC$^+$14]

au cas multiparamètres et multiples points test (ce qui ne devrait pas être trivial au vu de la complexité des calculs dans le cas monoparamètre mono point test). En effet en terme de précision, il est bien connu dans le cas bayésien, que les deux familles de bornes inférieures, Ziv-Zakai et Weiss-Weinstein, se disputent le titre de la borne la plus précise suivant l'application considérée [VRBM14, BSEV97, VB07]. De plus, nous ne sommes parvenus à caractériser cette borne que pour les modèles gaussiens à moyenne paramétrée et il serait intéressant d'étendre ces résultats au cas de modèles gaussiens à covariance paramétrée.

3. d'étendre, au cas de l'estimation de paramètres hybrides, les formes récursives des bornes bayésiennes de Weiss-Weinstein pour les systèmes non linéaires Markoviens établies dans [RO07, XGMM13], ce qui constituerait également une extension de la BCRH récursive.
4. d'étendre la prise en compte des contraintes égalités [GO91,RGC+15a] (tout ou partie) sur les paramètres déterministes aux "Large Error Bounds" dans les contextes déterministes et hybrides.
5. de traiter le cas des moments d'ordres supérieures (> 2) afin de compléter l'analyse des performances des estimateurs. Nous pensons en particulier à utiliser dans un premier temps l'inégalité générale proposée dans [WW88] et à l'étendre au cas de l'estimation hybride.
6. de s'intéresser aux performances des estimateurs hybrides dans un contexte malspécifié. En effet, le premier résultat présenté dans ce document ne traite que de la borne de Cramér-Rao malspécifiée dans un contexte déterministe seulement. Notons que l'obtention d'une borne de Cramér-Rao hybride malspécifiée passera forcément par la mise en place d'un résultat de convergence en probabilité de l'estimateur du maximum de vraisemblance/maximum *a posteriori* similairement aux travaux d'Huber et de White [Hub67, Whi82].

Comme nous l'avons mentionné, l'EQM n'est pas la seule mesure de performance. La seconde étape est donc l'utilisation d'une mesure alternative et potentiellement plus informative, ce qui reste un axe de recherche très prospectif qui a d'ailleurs été relativement peu exploré théoriquement [Bar49, WW88, OK91, Cla95] en traitement statistique du signal et pour lequel nous avons contribué [RkG+14]. Par ailleurs, cet axe bénéficie de peu de résultats pratiques alors qu'il s'agit d'une piste d'un grand intérêt puisqu'elle pourrait permettre à terme de pouvoir caractériser les performances en estimation des systèmes à partir des probabilités et non plus des moments. Nous avons déjà montré que la mesure de précision par la fonction de répartition se simplifie notablement pour peu que la classe d'estimateurs considérés soit gaussienne. C'est pourquoi nous avons proposé un test relativement simple pour établir la "zone de gaussianité" des estimateurs du maximum de vraisemblance pour un modèle d'observation gaussien à moyenne paramétrée. Dans cette perspective, il serait pertinent :

1. de généraliser la procédure de test de la "zone de gaussianité" de l'estimateur du maximum de vraisemblance proposée dans [RkG+14] aux modèles considérés dans [MGCL14]. Ceci permettrait à terme d'unifier ces deux résultats et de présenter un outil général d'aide à la conception des systèmes de mesure basée sur une précision plus informative.
2. de rechercher une procédure de test de la "zone de gaussianité" de l'estimateur du maximum de vraisemblance pour un modèle d'observations gaussiens à covariance paramétrée, sachant que cet estimateur n'est pas gaussien à fort RSB et pour un faible nombre d'observations indépendantes [RFBL07].

Enfin, les travaux traités de ce document concernent des modèles d'observations non-standards, en un certain sens, mais dans lesquels nous avons considéré le bruit gaussien, ce qui reste très standard en traitement du signal. Depuis quelques années, d'autres modèles, plus généraux et incluant le modèle gaussien, ont été étudiés. Nous pensons en particulier aux modèles suivant une loi symétrique elliptique complexe qui permettent de décrire un large ensemble de phénomènes aléatoires et qui, par conséquent, embrassent un grand nombre d'applications en traitement du signal moderne. Les premiers résultats intéressants, obtenus par exemple dans [BA13, GG13, AB15, GMV00, PCO$^+$08, PFOL08], nous suggèrent la perspective d'étendre l'ensemble de nos travaux pour ce type de distribution.

Annexe A

Etude de la matrice C

A.1 La borne de Barankin-Weiss-Weinstein hybride

Pour obtenir la borne de Barankin-Weiss-Weinstein hybride, il faut appliquer l'inégalité de covariance (2.1) avec le choix de $\mathbf{v}(\mathbf{x},\boldsymbol{\theta})$ défini par (2.52) et (2.53). Nous allons montrer que sous les contraintes (R1), (R2) et (R3), la matrice $\mathbf{C}$ est bien indépendante de l'estimateur $\widehat{\boldsymbol{\theta}}$. En effet, si nous décomposons par blocs la matrice $\mathbf{C} \in \mathbb{R}^{(D+R)\times(I+J)}$, c'est-à-dire $\mathbf{C} = \begin{bmatrix} \mathbf{C}_{(D+R)I} & \begin{matrix} \mathbf{C}_{DJ} \\ \mathbf{C}_{RJ} \end{matrix} \end{bmatrix}$ avec $\mathbf{C}_{(D+R)I} \in \mathbb{R}^{(D+R)\times I}$, $\mathbf{C}_{DJ} \in \mathbb{R}^{D\times J}$ et $\mathbf{C}_{RJ} \in \mathbb{R}^{R\times J}$. Notons $\{\mathbf{A}\}_{:,m}$ le m^{ième} vecteur colonne de la matrice $\mathbf{A}$, or d'après (2.45) :

$$\left\{\mathbf{C}_{(D+R)I}\right\}_{:,i} = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r|\boldsymbol{\theta}_d}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\{\mathbf{v}(\mathbf{x},\boldsymbol{\theta})\}_i\right] = \mathbf{h}_i,\ \forall 1 \leq i \leq I \tag{A.1}$$

Avant de calculer les matrices $\mathbf{C}_{DJ} \in \mathbb{R}^{D\times J}$ et $\mathbf{C}_{RJ} \in \mathbb{R}^{R\times J}$, nous donnons un résultat préliminaire : Pour toute fonction réelle mesurable $\mathbf{g}(\mathbf{x},\boldsymbol{\theta}_d)$, indépendante de $\boldsymbol{\theta}_r$, définie sur $\Omega \times \Pi_d$, $0 < m_j < 1$ et pour tout $\mathbf{h} = \left[\mathbf{0}^T\ \mathbf{h}_r^T\right]^T$, nous avons

$$\begin{aligned} &\int_{\Pi_r} \mathbf{g}(\mathbf{x},\boldsymbol{\theta}_d)\left(\frac{f_{\mathbf{x},\boldsymbol{\theta}}^{m_j}(\mathbf{x},\boldsymbol{\theta}+\mathbf{h})}{f_{\mathbf{x},\boldsymbol{\theta}}^{m_j}(\mathbf{x},\boldsymbol{\theta})} - \frac{f_{\mathbf{x},\boldsymbol{\theta}}^{1-m_j}(\mathbf{x},\boldsymbol{\theta}-\mathbf{h})}{f_{\mathbf{x},\boldsymbol{\theta}}^{1-m_j}(\mathbf{x},\boldsymbol{\theta})}\right) f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta})\, d\boldsymbol{\theta}_r \\ = \ & \mathbf{g}(\mathbf{x},\boldsymbol{\theta}_d)\int_{\Pi_r}\left(\begin{array}{c} f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}^{m_j}(\mathbf{x},\boldsymbol{\theta}_r+\mathbf{h}_r;\boldsymbol{\theta}_d)\, f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}^{1-m_j}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d) \\ -f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}^{m_j}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)\, f_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}^{1-m_j}(\mathbf{x},\boldsymbol{\theta}_r-\mathbf{h}_r;\boldsymbol{\theta}_d) \end{array}\right) d\boldsymbol{\theta}_r. \end{aligned}$$

Or Π_r satisfait la relation (R3), donc nous obtenons par le changement de variable $\boldsymbol{\theta}_r' = \boldsymbol{\theta}_r - \mathbf{h}_r$:

$$\int_{\Pi_r} f_{\mathbf{x},\boldsymbol{\theta}}^{m_j}\left(\mathbf{x},\boldsymbol{\theta}_r'+\mathbf{h}_r;\boldsymbol{\theta}_d\right) f_{\mathbf{x},\boldsymbol{\theta}}^{1-m_j}\left(\mathbf{x},\boldsymbol{\theta}_r';\boldsymbol{\theta}_d\right) d\boldsymbol{\theta}_r' = \int_{\Pi_r} f_{\mathbf{x},\boldsymbol{\theta}}^{m_j}(\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)\, f_{\mathbf{x},\boldsymbol{\theta}}^{1-m_j}(\mathbf{x},\boldsymbol{\theta}_r-\mathbf{h}_r;\boldsymbol{\theta}_d)\, d\boldsymbol{\theta}_r,$$

conduisant à

$$\int_{\Pi_r} \mathbf{g}(\mathbf{x},\boldsymbol{\theta}_d)\left(\frac{f_{\mathbf{x},\boldsymbol{\theta}}^{m_j}(\mathbf{x},\boldsymbol{\theta}+\mathbf{h})}{f_{\mathbf{x},\boldsymbol{\theta}}^{m_j}(\mathbf{x},\boldsymbol{\theta})} - \frac{f_{\mathbf{x},\boldsymbol{\theta}}^{1-m_j}(\mathbf{x},\boldsymbol{\theta}-\mathbf{h})}{f_{\mathbf{x},\boldsymbol{\theta}}^{1-m_j}(\mathbf{x},\boldsymbol{\theta})}\right) f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta})\, d\boldsymbol{\theta}_r = \mathbf{0},\ \forall \boldsymbol{\theta}_d \in \Pi_d \tag{A.2}$$

Remarques

— Cette égalité constitue une extension des conditions satisfaites par la famille des bornes de Weiss-Weinsten [WW85, (1)(2)] dans le cas bayésien aux cas de l'estimation hybride en substituant $f(\mathbf{x}, \boldsymbol{\theta}_r; \boldsymbol{\theta}_d)$ par $f(\mathbf{x}, \boldsymbol{\theta}_r)$.

— Notons que si $\mathbf{h} = \left[\mathbf{h}_d^T \ \mathbf{h}_r^T\right]^T$ avec $\mathbf{h}_d \neq 0$, alors la condition d'orthogonalité (A.2) ne sera pas satisfaite, ce qui conduira à une dépendance des matrices $\mathbf{C}_{DJ}$ et $\mathbf{C}_{RJ}$ envers $\widehat{\boldsymbol{\theta}}$.

Par conséquent, $\forall 1 < j < J$, les fonctions $v_{I+j}(\mathbf{x}, \boldsymbol{\theta})$ définies par (2.53) permettent d'obtenir la relation suivante :

$$\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d}\left[\mathbf{g}(\mathbf{x}, \boldsymbol{\theta}_d)\, v_{I+j}(\mathbf{x}, \boldsymbol{\theta})\right] = \mathbf{0}. \tag{A.3}$$

Pour tous les vecteurs colonnes de $\mathbf{C}_{dJ}$ et $\mathbf{C}_{rJ}$ notés respectivement $\{\mathbf{C}_{dJ}\}_{:,j}$ et $\{\mathbf{C}_{rJ}\}_{:,j}$ $1 \leq j \leq J$, nous en déduisons

$$\{\mathbf{C}_{dJ}\}_{:,j} = \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r|\boldsymbol{\theta}_d}\left[\left(\widehat{\boldsymbol{\theta}}_d(\mathbf{x}) - \boldsymbol{\theta}_d\right)\{\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})\}_{I+j}\right] = \mathbf{0}, \tag{A.4}$$

et

$$\begin{aligned}
\{\mathbf{C}_{rJ}\}_{:,j} &= \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r|\boldsymbol{\theta}_d}\left[\left(\widehat{\boldsymbol{\theta}}_r(\mathbf{x}) - \boldsymbol{\theta}_r\right)\{\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})\}_{I+j}\right] \\
&= -\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r|\boldsymbol{\theta}_d}\left[\boldsymbol{\theta}_r\{\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})\}_{I+j}\right] \\
&= \mathbf{h}_{r(I+j)}\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}_r|\boldsymbol{\theta}_d}\left[\frac{f^{1-m_j}(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_{I+j})}{f^{1-m_j}(\mathbf{x}, \boldsymbol{\theta})}\right].
\end{aligned} \tag{A.5}$$

A.2 La borne de Barankin Ziv-Zakaï hybride

Par définition de $\mathbf{V} = \mathbb{E}_{\mathbf{X},\boldsymbol{\theta}}\left[\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})\,\mathbf{v}^T(\mathbf{x}, \boldsymbol{\theta})\right]$, les choix de v_d et v_r donnés par (2.77) et (2.78) conduisent aux expressions (2.81), (2.82) et (2.83) sans difficulté calculatoire majeure. Il reste à montrer que la matrice $\mathbf{C}$ ne dépend pas de $\widehat{\boldsymbol{\theta}}$ pour obtenir une nouvelle borne. Il est évident que les éléments $\{\mathbf{C}\}_{1,1}$ et $\{\mathbf{C}\}_{2,1}$ sont indépendants de $\widehat{\boldsymbol{\theta}}$ en suivant le même raisonnement que pour la borne de Barankin hybride. Pour obtenir une simplification de $\{\mathbf{C}\}_{1,2}$ et $\{\mathbf{C}\}_{2,2}$, nous introduisons d'abord un résultat préliminaire : pour tout fonction réelle mesurable $l(\mathbf{x},\theta_d)$ définie sur $\Omega \times \Pi_d$ et pour tout $\mathbf{h} = (0 \ h_r)^T$ tel que $\theta_r + h_r \in \Pi_r$, nous avons

$$\begin{aligned}
&\int_{\Pi_r} l(\mathbf{x},\theta_d)\left(\min\left(\frac{f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h})}{f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta})}, 1\right) - \min\left(\frac{f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h})}{f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta})}, 1\right)\right) f(\mathbf{x}, \boldsymbol{\theta})\, d\theta_r \qquad \text{(A.6)} \\
= \ & l(\mathbf{x},\theta_d)\int_{\Pi_r}\left(\min\left(f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}), f(\mathbf{x}, \boldsymbol{\theta})\right) - \min\left(f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}), f_{\mathbf{x},\boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta})\right)\right) d\theta_r. \\
= \ & \int_{\Pi_r}\min\left(f_{\mathbf{x},\theta_r;\theta_d}(\mathbf{x},\theta_r + h_r; \theta_d), f_{\mathbf{x},\theta_r;\theta_d}(\mathbf{x},\theta_r; \theta_d)\right) d\theta_r \\
& - \int_{\Pi_r}\min\left(f_{\mathbf{x},\theta_r;\theta_d}(\mathbf{x},\theta_r - h_r; \theta_d), f_{\mathbf{x},\theta_r;\theta_d}(\mathbf{x},\theta_r; \theta_d)\right) d\theta_r \qquad \text{(A.7)}
\end{aligned}$$

Or en effectuant le changement de variable $\theta'_r = \theta_r + h_r$, le domaine d'intégration reste Π_r par l'hypothèse (R3) donc

$$\int_{\Pi_r}\min\left(f(\mathbf{x},\theta_r + h_r; \theta_d), f(\mathbf{x},\theta_r; \theta_d)\right) d\theta_r = \int_{\Pi_r}\min\left(f\left(\mathbf{x},\theta'_r; \theta_d\right), f\left(\mathbf{x},\theta'_r - h_r; \theta_d\right)\right) d\theta'_r. \tag{A.8}$$

En remplaçons (A.8) dans (A.6), nous obtenons :

$$\int_{\Pi_r} l\left(\mathbf{x},\theta_d\right)\left(\min\left(\frac{f\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}\right)}{f\left(\mathbf{x},\boldsymbol{\theta}\right)},1\right)-\min\left(\frac{f\left(\mathbf{x},\boldsymbol{\theta}-\mathbf{h}\right)}{f\left(\mathbf{x},\boldsymbol{\theta}\right)},1\right)\right)f\left(\mathbf{x},\boldsymbol{\theta}\right)d\theta_r=0,\ \forall\theta_d\in\Pi_d. \quad \text{(A.9)}$$

Nous remarquons que :
• ce résultat est similaire à la condition (1) dans [WW88] avec une légère différence près puisque la densité de probabilité jointe dépend de θ_d.
• si nous choisissons $\mathbf{h}=[h_d\ h_r]$ avec $h_d\neq 0$ dans (A.6), alors la propriété (A.9) dépendra de $\mathbf{x}$ et donc les éléments $\{\mathbf{C}\}_{1,2}$ et $\{\mathbf{C}\}_{2,2}$ dépendront de $\widehat{\boldsymbol{\theta}}$.
Maintenant, concernant $\{\mathbf{C}\}_{1,2}$, nous avons

$$\begin{aligned}\{\mathbf{C}\}_{1,2} &= \mathbb{E}_{\mathbf{x},\theta_r;\theta_d}\left[\left(\widehat{\theta}_d-\theta_d\right)v_r\right]\\ &= 0,\end{aligned}$$

en utilisant (A.9) avec $l\left(\mathbf{X},\theta_d\right)=\widehat{\theta}_d-\theta_d$.
Concernant $\{\mathbf{C}\}_{1,2}$, nous avons

$$\begin{aligned}\{\mathbf{C}\}_{2,2} &= \mathbb{E}_{\mathbf{x},\theta_r;\theta_d}\left[\left(\widehat{\theta}_r-\theta_r\right)v_r\right]\\ &= \int_{\Omega}\int_{\Pi_r}\theta_r\left[\begin{array}{c}\min\left(f\left(\mathbf{x},\boldsymbol{\theta}-\mathbf{h}_2\right),f\left(\mathbf{x},\boldsymbol{\theta}\right)\right)\\ -\min\left(f\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_2\right),f\left(\mathbf{x},\boldsymbol{\theta}\right)\right)\end{array}\right]d\theta_r d\mathbf{X},\end{aligned} \quad \text{(A.10)}$$

en utilisant (A.9) avec $l\left(\mathbf{x},\theta_d\right)=\widehat{\theta}_r$. Par un changement de variables $\theta'_r=\theta_r-h_{2r}$, le domaine d'intégration Π_r est inchangé d'après l'hypothèse (R3) et nous obtenons

$$\begin{aligned}&\int_{\Pi_r}\theta_r\min\left(f\left(\mathbf{x},\theta_r-h_{2r};\theta_d\right),f\left(\mathbf{x},\theta_r;\theta_d\right)\right)d\theta_r\\ =&\int_{\Pi_r}\theta'_r\min\left(f\left(\mathbf{x},\theta'_r;\theta_d\right),f\left(\mathbf{x},\theta'_r+h_{2r};\theta_d\right)\right)d\theta'_r\\ &+h_{2r}\int_{\Pi_r}\min\left(f\left(\mathbf{x},\theta'_r;\theta_d\right),f\left(\mathbf{x},\theta'_r+h_{2r};\theta_d\right)\right)d\theta'_r.\end{aligned} \quad \text{(A.11)}$$

Par conséquent, en remplaçant (A.11) dans (A.10), nous avons

$$\{\mathbf{C}\}_{2,2}=h_{2r}\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}}\left[\min\left(\frac{f\left(\mathbf{x},\boldsymbol{\theta}+\mathbf{h}_2\right)}{f\left(\mathbf{x},\boldsymbol{\theta}\right)},1\right)\right]. \quad \text{(A.12)}$$

Nous avons montré finalement que ce choix de $\mathbf{v}\left(\mathbf{x},\boldsymbol{\theta}\right)$ donné par (2.77) et (2.78) aboutit bien à une borne indépendante de $\widehat{\boldsymbol{\theta}}$.

Annexe B

Recursive hybrid Cramér-Rao bound for discrete-time Markovian dynamic systems

IEEE Signal Processing Letters, Volume : 22, Issue : 10, Oct. 2015, pp. 1543-1547

Recursive hybrid Cramér-Rao bound for discrete-time Markovian dynamic systems

Chengfang Ren, Jerome Galy, Eric Chaumette, François Vincent, Pascal Larzabal and Alexandre Renaux

***Abstract*—In statistical signal processing, hybrid parameter estimation refers to the case where the parameters vector to estimate contains both non-random and random parameters. As a contribution to the hybrid estimation framework, we introduce a recursive hybrid Cramér Rao lower bounds for discrete-time Markovian dynamic systems depending on unknown deterministic parameters. Additionnally, the regularity conditions required for its existence and its use are clarified.**

***Index Terms*—Parameter estimation, dynamic Markovian systems, estimation error lower bound**

I. Introduction

Since its introduction in the context of array shape calibration [1], hybrid parameter estimation has given rise to a growing interest as both random and nonrandom parameters occur simultaneously in miscellaneous estimation problems [2]-[9]. However, the hybrid estimation framework is not just the simple concatenation of the Bayesian and non-Bayesian techniques and new estimators have to be derived [10, §1.1]. Similarly, performance analysis methods of such estimators has to be modified accordingly, which is the aim of hybrid lower bounds on the mean square error (MSE).

The first hybrid lower bound, the so-called hybrid Cramér-Rao bound (HCRB), has been introduced in [1] where the random parameters have a prior probability density function (pdf) independent from deterministic parameters. This initial characterization of hybrid estimation has been generalized by Reuven and Messer [2] who introduced the first "large-error" hybrid bound, the so-called hybrid Barankin Bound (HBB), in order to handle the threshold phenomena and of which one limiting form yields the HCRB. This seminal work [2] has been lately extended to new "large-error" hybrid bounds [7][11][12] in order to improve the estimation of the transition region where the threshold phenomena occurs. Unfortunately, the computational cost of hybrid "large-error" bounds is prohibitive in most applications when the number of unknown parameters increases. Concurrently, an extension of the HCRB where the prior pdf of the random parameters depends on deterministic parameters was proposed in [6] and its asymptotic tightness was further analyzed in [13]. All these works have shown that, like the deterministic CRB and Bayesian CRB (BCRB), the HCRB is valid in the asymptotic region only, i.e., when signal to noise ratio is high or the number of observations is large.

In the Bayesian estimation framework, discrete-time Markovian dynamic systems (MDS) arises in various applications such as adaptive control, analysis, and prediction of non-stationary time series [14]. As is well known, the optimal estimator for this problem cannot be built in general, and it is necessary to turn to one of the large number of existing suboptimal filtering techniques [14]. Assessing the achievable performance may be difficult, and we have to resort to simulations and comparing proximity to bayesian lower bounds corresponding to optimum performance [10][15][16]. Actually, most discrete-time MDS incorporate some deterministic parameters which can be either known [10] or unknown [6] according to the experimental conditions. Even when the deterministic parameters are known, some of the true values may originate from a prior calibration process which accuracy impacts on the optimum performance of random parameter estimates. In both cases, there is a need for computationally tractable hybrid lower bounds for discrete-time MDS depending on unknown deterministic parameters.

As a contribution, we introduce the first recursive form of an hybrid lower bound for discrete-time MDS, namely the recursive HCRB, which, provided that one keeps in mind its limitations, is a lower bound of great interest for system analysis and design in the asymptotic region. Additionally we discuss the regularity conditions required for the existence and the use of the recursive HCRB, which are critical to understand not only the applicability limit of the recursive HCRB but also why, in most case, the posterior BCRB cannot be transformed into the recursive HCRB as it is misleadingly suggested in [15][16].

II. Recursive HCRB for Discrete-Time MDS

In hybrid parameter estimation one wishes to estimate an unknown hybrid parameter vector $(\mathbf{x};\boldsymbol{\theta})$[1] from a random observation vector $\mathbf{y} \in \mathbb{R}^{N'}$. Some prior knowledge is available on random parameter $\mathbf{x} \in \mathbb{R}^{P'}$ that is incorporated by an *a priori* pdf $p(\mathbf{x})$ which support is a subset Π_r of $\mathbb{R}^{P'}$. No such knowledge is available on $\boldsymbol{\theta} \in \Pi_d \subset \mathbb{R}^{D'}$ and

Chengfang Ren and Alexandre Renaux are with Université Paris-Sud/LSS 3, Rue Joliot-Curie, 91192 Gif-sur-Yvette, France. Emails: cren@lss.supelec.fr, renaux@lss.supelec.fr

Jérome Galy is with Université de Montpellier 2/LIRMM, 161 rue Ada 34392 Montpellier Cedex 5, France. Email: galy@lirmm.fr

Eric Chaumette and François Vincent are with University of Toulouse-ISAE, DEOS, 10 Avenue Edouard Belin, 31055 Toulouse, France. Emails: eric.chaumette@isae.fr, francois.vincent@isae.fr

Pascal Larzabal is with Université Paris-Sud/SATIE, 61 av. du President Wilson, 94235 Cachan, France. Email: pascal.larzabal@satie.ens-cachan.fr

This work has been partially supported by the European Network of excellence NEWCOM#, by the iCODE institute, research project of the IDEX Paris-Saclay, by the DGA/DGCIS and by the Display-Mastodons project of CNRS.

[1]For L column vectors $\mathbf{a}_l$, $(\mathbf{a}_1;\mathbf{a}_2;\ldots;\mathbf{a}_L) \triangleq (\mathbf{a}_1^T,\mathbf{a}_2^T,\ldots,\mathbf{a}_L^T)^T$ denotes the vertical concatenation

thus it is considered deterministic. In the general case, $p(\mathbf{x})$ may depend on the unknown parameter $\boldsymbol{\theta}$, and it is denoted $p(\mathbf{x}|\boldsymbol{\theta})$. The conditional pdf of $\mathbf{y}$ given $\mathbf{x}$ parameterized by $\boldsymbol{\theta}$ is $p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})$ and their joint pdf parameterized by $\boldsymbol{\theta}$ is given by $p(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})\,p(\mathbf{x}|\boldsymbol{\theta})$. Then, for any estimators $\widehat{\boldsymbol{\theta}}(\mathbf{y})$ of $\boldsymbol{\theta}$ and $\widehat{\mathbf{x}}(\mathbf{y})$ of $\mathbf{x}$, one of the possible lower bounds [10][11][12] deriving from the covariance inequality principle (16) is the HCRB which usual form is given by [10]:

$$E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\mathbf{e}(\mathbf{y})\,\mathbf{e}(\mathbf{y})^T\right] \succeq \mathbf{HCRB}_{\mathbf{x},\boldsymbol{\theta}} = \mathbf{J}_{\mathbf{x},\boldsymbol{\theta}}^{-1}, \tag{1}$$

$$\mathbf{J}_{\mathbf{x},\boldsymbol{\theta}} = E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\frac{\partial \ln p(\mathbf{y},\mathbf{x}|\boldsymbol{\theta})}{\partial(\mathbf{x};\boldsymbol{\theta})}\frac{\partial \ln p(\mathbf{y},\mathbf{x}|\boldsymbol{\theta})}{\partial(\mathbf{x};\boldsymbol{\theta})}^T\right], \tag{2}$$

where $\mathbf{e}(\mathbf{y}) = (\widehat{\mathbf{x}}(\mathbf{y}) - \mathbf{x}; \widehat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta})$, $E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}[\mathbf{g}(\mathbf{y},\mathbf{x})]$ is the statistical expectation of the vector of functions $\mathbf{g}(\,)$ with respect to $\mathbf{y}$ and $\mathbf{x}$ parameterized by $\boldsymbol{\theta}$, and for two matrices, $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semi-definite. The regularity conditions for the hybrid Fisher information matrix (HFIM) $\mathbf{J}_{\mathbf{x},\boldsymbol{\theta}}$ to be of the usual form (2) are (see Section III):
R1): $\Pi_r = \mathbb{R}^P$,
(R2): $E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\frac{\partial \ln p(\mathbf{y},\mathbf{x}|\boldsymbol{\theta})}{\partial x_p}^2\right], E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\frac{\partial \ln p(\mathbf{y},\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_d}^2\right] < \infty$.
Moreover, under its usual form (1), the HCRB is a lower bound for the class of estimates satisfying (R3): $E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\left(\widehat{\mathbf{x}}(\mathbf{y}) - \mathbf{x}; \widehat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}\right)\right] = (\boldsymbol{\mu};\mathbf{0})$.
Our main concern is the derivation of a computationally tractable HFIM (2) for hybrid discrete-time MDS represented with the state and measurement equations:

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1},\mathbf{w}_{k-1},\boldsymbol{\alpha}),\ \mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k,\mathbf{v}_k,\boldsymbol{\lambda}) \tag{3}$$

where $k \geq 1$ is a time index, $\mathbf{x}_k$ is the P-dimensional state vector, $\mathbf{y}_k$ is the N-dimensional measurement vector, $f_k(\ ,\ ,\boldsymbol{\alpha})$ and $\mathbf{h}_k(\ ,\ ,\boldsymbol{\lambda})$ are known parametric vector functions depending on an unknown deterministic parameter vector ($\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ respectively). The process noise sequence $\{\mathbf{w}_k\}$ and the measurement noise sequence $\{\mathbf{v}_k\}$ are mutually independent white sequences described by known pdfs $p(\mathbf{w}_k|\boldsymbol{\beta})$ and $p(\mathbf{v}_k|\boldsymbol{\mu})$, respectively, depending on an unknown deterministic parameter vector ($\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ respectively). The noises are independent of the initial state $\mathbf{x}_0$ described by the known pdf $p(\mathbf{x}_0|\boldsymbol{\alpha})$. Let $\boldsymbol{\theta} = (\boldsymbol{\lambda};\boldsymbol{\mu};\boldsymbol{\alpha};\boldsymbol{\beta})$ be the vector gathering all the unknown deterministic parameters. The state transition and the measurement pdfs depend on unknown deterministic parameters:

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) \triangleq p(\mathbf{x}_k|\mathbf{x}_{k-1},\boldsymbol{\alpha},\boldsymbol{\beta}),\ p(\mathbf{y}_k|\mathbf{x}_k) \triangleq p(\mathbf{y}_k|\mathbf{x}_k,\boldsymbol{\theta}),$$

and we suppose that both $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and $p(\mathbf{y}_k|\mathbf{x}_k)$ are twice differentiable with respect to all their arguments. We adopt the notational convention: $\forall l \leq k, \overline{\boldsymbol{\delta}}_{l:k} = (\boldsymbol{\delta}_l;\ldots;\boldsymbol{\delta}_k)$ where all vectors are of same dimension. Since (3) is a MDS:

$$p(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta}) = p(\mathbf{x}_0|\boldsymbol{\alpha})\prod_{l=1}^{k} p(\mathbf{y}_l|\mathbf{x}_l,\boldsymbol{\theta})\,p(\mathbf{x}_l|\mathbf{x}_{l-1},\boldsymbol{\alpha},\boldsymbol{\beta}) \tag{4}$$

From a theoretical point of view, we are primarily interested in the HFIM (2) on $(\mathbf{x};\boldsymbol{\theta}) \triangleq (\mathbf{x}_k;\boldsymbol{\theta})$ associated to the observation vector $\mathbf{y} \triangleq \overline{\mathbf{y}}_{1:k}$ resulting from the set of k measurements $\mathbf{y}_1$,...,$\mathbf{y}_k$:

$$\mathbf{J}_{\mathbf{x}_k,\boldsymbol{\theta}} = E_{\overline{\mathbf{y}}_{1:k},\mathbf{x}_k|\boldsymbol{\theta}}\left[\frac{\partial \ln p(\overline{\mathbf{y}}_{1:k},\mathbf{x}_k|\boldsymbol{\theta})}{\partial(\mathbf{x}_k;\boldsymbol{\theta})}\frac{\partial \ln p(\overline{\mathbf{y}}_{1:k},\mathbf{x}_k|\boldsymbol{\theta})}{\partial(\mathbf{x}_k;\boldsymbol{\theta})}^T\right]$$

which alternative formula is [1][2][10]:

$$\mathbf{J}_{\mathbf{x}_k,\boldsymbol{\theta}} = E_{\overline{\mathbf{y}}_{1:k},\mathbf{x}_k|\boldsymbol{\theta}}\left[\frac{-\partial^2 \ln p(\overline{\mathbf{y}}_{1:k},\mathbf{x}_k|\boldsymbol{\theta})}{\partial(\mathbf{x}_k;\boldsymbol{\theta})\,\partial(\mathbf{x}_k;\boldsymbol{\theta})^T}\right] \tag{5}$$

Unfortunately the computation of (5) requires the derivation of the marginal pdf $p(\overline{\mathbf{y}}_{1:k},\mathbf{x}_k|\boldsymbol{\theta})$ from (4) which is generally mathematically intractable [1][10][15]. However an upper bound of (5) can be derived from the HFIM (2) on $(\mathbf{x};\boldsymbol{\theta}) \triangleq (\overline{\mathbf{x}}_{0:k};\boldsymbol{\theta})$ associated to the observation vector $\mathbf{y} \triangleq \overline{\mathbf{y}}_{1:k}$:

$$\mathbf{J}_{\overline{\mathbf{x}}_{0:k},\boldsymbol{\theta}} = E_{\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta}}\left[\frac{-\partial^2 \ln p(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta})}{\partial(\overline{\mathbf{x}}_{0:k};\boldsymbol{\theta})\,\partial(\overline{\mathbf{x}}_{0:k};\boldsymbol{\theta})^T}\right]. \tag{6}$$

Indeed, if we decompose $\mathbf{J}_{\overline{\mathbf{x}}_{0:k},\boldsymbol{\theta}}$ as:

$$\mathbf{J}_{\overline{\mathbf{x}}_{0:k},\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{J}_k^{11} & \mathbf{B}_k \\ \mathbf{B}_k^T & \mathbf{C}_k \end{bmatrix},\ \begin{array}{l} \mathbf{J}_k^{11} = E_{\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta}}\left[-\frac{\partial^2 \ln p(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta})}{\partial\overline{\mathbf{x}}_{0:k-1}\partial\overline{\mathbf{x}}_{0:k-1}^T}\right] \\ \mathbf{B}_k = E_{\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta}}\left[-\frac{\partial^2 \ln p(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta})}{\partial\overline{\mathbf{x}}_{0:k-1}\partial(\mathbf{x}_k;\boldsymbol{\theta})^T}\right] \\ \mathbf{C}_k = E_{\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta}}\left[-\frac{\partial^2 \ln p(\overline{\mathbf{y}}_{1:k},\mathbf{x}_k|\boldsymbol{\theta})}{\partial(\mathbf{x}_k;\boldsymbol{\theta})\partial(\mathbf{x}_k;\boldsymbol{\theta})^T}\right] \end{array},$$

then a generalization of Proposition 1 in [17] (which proof follows lines similar to the proof of Theorem 1 in [13]) to hybrid estimation yields:

$$\widetilde{\mathbf{J}}_{\mathbf{x}_k,\boldsymbol{\theta}} \succeq \mathbf{J}_{\mathbf{x}_k,\boldsymbol{\theta}} \ \Leftrightarrow\ \mathbf{HCRB}_{\mathbf{x}_k,\boldsymbol{\theta}} \succeq \widetilde{\mathbf{J}}_{\mathbf{x}_k,\boldsymbol{\theta}}^{-1}, \tag{7}$$

where:

$$\widetilde{\mathbf{J}}_{\mathbf{x}_k,\boldsymbol{\theta}} = \mathbf{C}_k - \mathbf{B}_k^T\left(\mathbf{J}_k^{11}\right)^{-1}\mathbf{B}_k, \tag{8}$$

and leading to:

$$E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\mathbf{e}(\mathbf{y})\,\mathbf{e}(\mathbf{y})^T\right] \succeq \widetilde{\mathbf{J}}_{\mathbf{x}_k,\boldsymbol{\theta}}^{-1}. \tag{9}$$

$\widetilde{\mathbf{J}}_{\mathbf{x}_k,\boldsymbol{\theta}}^{-1}$ defines a looser (in comparison with $\mathbf{HCRB}_{\mathbf{x}_k,\boldsymbol{\theta}}$) but general computable hybrid bound for discrete-time MDS pdf (4). Additionally $\widetilde{\mathbf{J}}_{\mathbf{x}_k,\boldsymbol{\theta}}$ is computationally tractable: it can be assessed without computing the inverse of large matrices such as $(PK \times PK)$ matrix $\mathbf{J}_k^{11}$. Indeed $\widetilde{\mathbf{J}}_{\mathbf{x}_k,\boldsymbol{\theta}}$ (8) can be decomposed into block matrices:

$$\widetilde{\mathbf{J}}_{\mathbf{x}_k,\boldsymbol{\theta}} = \begin{bmatrix} \widetilde{\mathbf{J}}_k^{\mathbf{x}_k,\mathbf{x}_k} & \widetilde{\mathbf{J}}_k^{\mathbf{x}_k,\boldsymbol{\theta}} \\ \widetilde{\mathbf{J}}_k^{\boldsymbol{\theta},\mathbf{x}_k} & \widetilde{\mathbf{J}}_k^{\boldsymbol{\theta},\boldsymbol{\theta}} \end{bmatrix} \tag{10}$$

which obey the recursion (see the Appendix for details):

$$\left|\begin{array}{l} \widetilde{\mathbf{J}}_k^{\mathbf{x}_k,\mathbf{x}_k} = \mathbf{D}_{k-1}^{22} - \left(\mathbf{D}_{k-1}^{12}\right)^T\left(\mathbf{D}_{k-1}^{11} + \widetilde{\mathbf{J}}_{k-1}^{\mathbf{x}_{k-1},\mathbf{x}_{k-1}}\right)^{-1}\mathbf{D}_{k-1}^{12} \\ \widetilde{\mathbf{J}}_k^{\mathbf{x}_k,\boldsymbol{\theta}} = \mathbf{D}_{k-1}^{23} - \left(\mathbf{D}_{k-1}^{12}\right)^T\left(\mathbf{D}_{k-1}^{11} + \widetilde{\mathbf{J}}_{k-1}^{\mathbf{x}_{k-1},\mathbf{x}_{k-1}}\right)^{-1} \\ \qquad\qquad \times\left(\mathbf{D}_{k-1}^{13} + \widetilde{\mathbf{J}}_{k-1}^{\mathbf{x}_{k-1},\boldsymbol{\theta}}\right) \\ \widetilde{\mathbf{J}}_k^{\boldsymbol{\theta},\boldsymbol{\theta}} = \mathbf{D}_{k-1}^{33} + \mathbf{J}_{k-1}^{\boldsymbol{\theta},\boldsymbol{\theta}} - \left(\mathbf{D}_{k-1}^{13} + \widetilde{\mathbf{J}}_{k-1}^{\mathbf{x}_{k-1},\boldsymbol{\theta}}\right)^T \\ \qquad \times\left(\mathbf{D}_{k-1}^{11} + \widetilde{\mathbf{J}}_{k-1}^{\mathbf{x}_{k-1},\mathbf{x}_{k-1}}\right)^{-1}\left(\mathbf{D}_{k-1}^{13} + \widetilde{\mathbf{J}}_{k-1}^{\mathbf{x}_{k-1},\boldsymbol{\theta}}\right) \end{array}\right. \tag{11}$$

where:

$$\left|\begin{array}{l}\mathbf{D}_{k-1}^{11}=E_{\mathbf{x}_k,\mathbf{x}_{k-1}|\boldsymbol{\alpha},\boldsymbol{\beta}}\left[-\frac{\partial^2\ln p(\mathbf{x}_k|\mathbf{x}_{k-1},\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\mathbf{x}_{k-1}\partial\mathbf{x}_{k-1}^T}\right]\\ \mathbf{D}_{k-1}^{12}=E_{\mathbf{x}_k,\mathbf{x}_{k-1}|\boldsymbol{\alpha},\boldsymbol{\beta}}\left[-\frac{\partial^2\ln p((\mathbf{x}_k|\mathbf{x}_{k-1},\boldsymbol{\alpha},\boldsymbol{\beta}))}{\partial\mathbf{x}_{k-1}\partial\mathbf{x}_k^T}\right]\\ \mathbf{D}_{k-1}^{22}=E_{\mathbf{y}_k,\mathbf{x}_k|\boldsymbol{\theta}}\left[-\frac{\partial^2\ln p(\mathbf{y}_k|\mathbf{x}_k,\boldsymbol{\theta})}{\partial\mathbf{x}_k\partial\mathbf{x}_k^T}\right]\\ \qquad+E_{\mathbf{x}_k,\mathbf{x}_{k-1}|\boldsymbol{\alpha},\boldsymbol{\beta}}\left[-\frac{\partial^2\ln p((\mathbf{x}_k|\mathbf{x}_{k-1},\boldsymbol{\alpha},\boldsymbol{\beta}))}{\partial\mathbf{x}_k\partial\mathbf{x}_k^T}\right]\end{array}\right. \tag{12}$$

$$\left|\begin{array}{l}\mathbf{D}_{k-1}^{13}=E_{\mathbf{x}_k,\mathbf{x}_{k-1}|\boldsymbol{\alpha},\boldsymbol{\beta}}\left[-\frac{\partial^2\ln p((\mathbf{x}_k|\mathbf{x}_{k-1},\boldsymbol{\alpha},\boldsymbol{\beta}))}{\partial\mathbf{x}_{k-1}\partial\boldsymbol{\theta}^T}\right]\\ \mathbf{D}_{k-1}^{23}=E_{\mathbf{y}_k,\mathbf{x}_k|\boldsymbol{\theta}}\left[-\frac{\partial^2\ln p(\mathbf{y}_k|\mathbf{x}_k,\boldsymbol{\theta})}{\partial\mathbf{x}_k\partial\boldsymbol{\theta}^T}\right]\\ \qquad+E_{\mathbf{x}_k,\mathbf{x}_{k-1}|\boldsymbol{\alpha},\boldsymbol{\beta}}\left[-\frac{\partial^2\ln p((\mathbf{x}_k|\mathbf{x}_{k-1},\boldsymbol{\alpha},\boldsymbol{\beta}))}{\partial\mathbf{x}_k\partial\boldsymbol{\theta}^T}\right]\\ \mathbf{D}_{k-1}^{33}=E_{\mathbf{y}_k,\mathbf{x}_k|\boldsymbol{\theta}}\left[-\frac{\partial^2\ln p(\mathbf{y}_k|\mathbf{x}_k,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]\\ \qquad+E_{\mathbf{x}_k,\mathbf{x}_{k-1}|\boldsymbol{\alpha},\boldsymbol{\beta}}\left[-\frac{\partial^2\ln p((\mathbf{x}_k|\mathbf{x}_{k-1},\boldsymbol{\alpha},\boldsymbol{\beta}))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]\end{array}\right. \tag{13}$$

$$\widetilde{\mathbf{J}}_{\mathbf{x}_0,\boldsymbol{\theta}}=E_{\mathbf{x}_0|\boldsymbol{\theta}}\left[\frac{-\partial^2\ln p\left(\mathbf{x}_0|\boldsymbol{\alpha}\right)}{\partial\left(\mathbf{x}_0;\boldsymbol{\theta}\right)\partial\left(\mathbf{x}_0^T,\boldsymbol{\theta}^T\right)}\right] \tag{14}$$

Using definition (10) of $\widetilde{\mathbf{J}}_{\mathbf{x}_k,\boldsymbol{\theta}}$ and the associated recursion (11) only involve computations with matrix of dimension $(\max\{D,P\}\times\max\{D,P\})$. Therefore it seems adequate to name $\widetilde{\mathbf{J}}_{\mathbf{x}_k,\boldsymbol{\theta}}$ and $\widetilde{\mathbf{J}}_{\mathbf{x}_k,\boldsymbol{\theta}}^{-1}$ the "recursive" HFIM and HCRB (for discrete-time MDS), respectively. Some special cases of interest can be easily derived by updating the definitions of $\boldsymbol{\theta}$, $p\left(\mathbf{x}_k|\mathbf{x}_{k-1}\right)$, $p\left(\mathbf{y}_k|\mathbf{x}_k\right)$ and $p\left(\mathbf{x}_0|\boldsymbol{\alpha}\right)$ accordingly. For instance, if $\boldsymbol{\theta}=\varnothing$, i.e. there is no unknown deterministic parameter, then (11)(12) reduce to:

$$\left|\begin{array}{l}\widetilde{\mathbf{J}}_k^{\mathbf{x}_k,\mathbf{x}_k}=\mathbf{D}_{k-1}^{22}-\left(\mathbf{D}_{k-1}^{12}\right)^T\left(\mathbf{D}_{k-1}^{11}+\widetilde{\mathbf{J}}_{k-1}^{\mathbf{x}_{k-1},\mathbf{x}_{k-1}}\right)^{-1}\mathbf{D}_{k-1}^{12}\\ \mathbf{D}_{k-1}^{11}=E_{\mathbf{x}_k,\mathbf{x}_{k-1}}\left[\frac{-\partial^2\ln p((\mathbf{x}_k|\mathbf{x}_{k-1}))}{\partial\mathbf{x}_{k-1}\partial\mathbf{x}_{k-1}^T}\right]\\ \mathbf{D}_{k-1}^{12}=E_{\mathbf{x}_k,\mathbf{x}_{k-1}}\left[\frac{-\partial^2\ln p(\mathbf{x}_k|\mathbf{x}_{k-1})}{\partial\mathbf{x}_{k-1}\partial\mathbf{x}_k^T}\right]\\ \mathbf{D}_{k-1}^{22}=E_{\mathbf{y}_k,\mathbf{x}_k}\left[\frac{-\partial^2\ln p(\mathbf{y}_k|\mathbf{x}_k)}{\partial\mathbf{x}_k\partial\mathbf{x}_k^T}\right]\\ \qquad+E_{\mathbf{x}_k,\mathbf{x}_{k-1}}\left[\frac{-\partial^2\ln p(\mathbf{x}_k|\mathbf{x}_{k-1})}{\partial\mathbf{x}_k\partial\mathbf{x}_k^T}\right]\end{array}\right. \tag{15}$$

and $\widetilde{\mathbf{J}}_{\mathbf{x}_0,\boldsymbol{\theta}}\triangleq\mathbf{J}_{\mathbf{x}_0}$, which are (21-25) in [15]. If (3) reduces to:

$$\mathbf{x}_k=\mathbf{f}_{k-1}\left(\mathbf{x}_{k-1},\mathbf{w}_{k-1}\right),\ \mathbf{y}_k=\mathbf{h}_k\left(\mathbf{x}_k,\mathbf{v}_k,\boldsymbol{\lambda}\right),$$

with known pdfs $p\left(\mathbf{w}_k\right)$, $p\left(\mathbf{v}_k\right)$, $p\left(\mathbf{x}_0\right)$, then one only needs to set $\boldsymbol{\theta}=\boldsymbol{\lambda}$, $p\left(\mathbf{x}_k|\mathbf{x}_{k-1}\right)\triangleq p\left(\mathbf{x}_k|\mathbf{x}_{k-1}\right)$ and $\widetilde{\mathbf{J}}_{\mathbf{x}_0,\boldsymbol{\theta}}\triangleq\mathbf{J}_{\mathbf{x}_0}$, yielding: $\mathbf{D}_{k-1}^{13}=\mathbf{0}$, $\mathbf{D}_{k-1}^{23}=E_{\mathbf{y}_k,\mathbf{x}_k|\boldsymbol{\theta}}\left[-\frac{\partial^2\ln p(\mathbf{y}_k|\mathbf{x}_k,\boldsymbol{\theta})}{\partial\mathbf{x}_k\partial\boldsymbol{\theta}^T}\right]$, $\mathbf{D}_{k-1}^{33}=E_{\mathbf{y}_k,\mathbf{x}_k|\boldsymbol{\theta}}\left[-\frac{\partial^2\ln p(\mathbf{y}_k|\mathbf{x}_k,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]$. If (3) reduces to:

$$\mathbf{x}_k=\mathbf{f}_{k-1}\left(\mathbf{x}_{k-1},\mathbf{w}_{k-1},\boldsymbol{\alpha}\right),\ \mathbf{y}_k=\mathbf{h}_k\left(\mathbf{x}_k,\mathbf{v}_k\right),$$

with known pdfs $p\left(\mathbf{w}_k\right)$, $p\left(\mathbf{v}_k\right)$, $p\left(\mathbf{x}_0\right)$, then one only needs to set $\boldsymbol{\theta}=\boldsymbol{\alpha}$, $p\left(\mathbf{x}_k|\mathbf{x}_{k-1}\right)\triangleq p\left(\mathbf{x}_k|\mathbf{x}_{k-1},\boldsymbol{\theta}\right)$ and $\widetilde{\mathbf{J}}_{\mathbf{x}_0,\boldsymbol{\theta}}\triangleq\mathbf{J}_{\mathbf{x}_0}$. Interestingly enough, the application considered in [6] addresses this case with measurement and state equations given by:

$$x_k=x_{k-1}+\alpha+w_{k-1},\ y_k=a_ke^{jx_k}+v_k,$$

where a_k are i.i.d. discrete random variable with $P\left(a_k=\pm1\right)=1/2$. Although authors in [6] were primarily interested in the direct computation of $\mathbf{J}_{\widetilde{\mathbf{x}}_{0:k},\boldsymbol{\theta}}^{-1}$ (6), a careful examination of the derivation reveals that $\left[\mathbf{H}^{-1}\right]_{K,K}$ in [6, (22)] is another possible recursive form for $\left[\widetilde{\mathbf{J}}_{x_K,\alpha}^{-1}\right]_{11}$ (10) allowed by the specific discrete-time MDS considered. Thus [6, (22)] provides an example of the behavior of the recursive HCRB (9) for a practical phase estimation problem.

III. Regularity conditions for the recursive HCRB

We discuss in this section the regularity condition (stated in the previous section) required for the existence and the use of the recursive HCRB (9) deriving from the usual form of the HCRB (1). This discussion clarifies some previous results on HCRB [2][6][13] and is also helpful to understand why, in most case, the posterior BCRB cannot be transformed into the recursive HCRB as it is misleadingly suggested in [15][16].

Let $\Lambda_{\boldsymbol{\theta}}=\left\{\mathbf{h}_d\in\mathbb{R}^D\mid\boldsymbol{\theta}+\mathbf{h}_d\in\Pi_d\right\}$, $\Lambda_{\mathbf{x}}=\left\{\mathbf{h}_r\in\mathbb{R}^P\mid\mathbf{x}+\mathbf{h}_r\in\Pi_r\right\}$ and $\Lambda=\Lambda_{\boldsymbol{\theta}}\times\Lambda_{\mathbf{x}}=\left\{\mathbf{h}\in\mathbb{R}^{D+P}\mid(\mathbf{x};\boldsymbol{\theta})+\mathbf{h}\in\Pi_d\times\Pi_r\right\}$. We assume that for any couple set $(S,T)\subset\Omega\times\Pi_r$ not empty, $\int_S\int_T p\left(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}\right)d\mathbf{y}d\mathbf{x}$ exists. Then, for any L-dimensional real-valued vector $\mathbf{v}\left(\mathbf{y},\mathbf{x},\boldsymbol{\theta}\right)$ with finite second order moment, the covariance inequality principle yields [18]:

$$E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\mathbf{e}\left(\mathbf{y}\right)\mathbf{e}\left(\mathbf{y}\right)^T\right]\succeq\mathbf{C}\mathbf{V}^{-1}\mathbf{C}^T, \tag{16}$$

$$\mathbf{V}=E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\mathbf{v}\left(\mathbf{y},\mathbf{x},\boldsymbol{\theta}\right)\mathbf{v}\left(\mathbf{y},\mathbf{x},\boldsymbol{\theta}\right)^T\right], \tag{17}$$

$$\mathbf{C}=E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\mathbf{e}\left(\mathbf{y}\right)\mathbf{v}\left(\mathbf{y},\mathbf{x},\boldsymbol{\theta}\right)^T\right]. \tag{18}$$

Note that $\mathbf{C}$ (18) depends on the estimation scheme $\left(\widehat{\mathbf{x}}\left(\mathbf{y}\right);\widehat{\boldsymbol{\theta}}\left(\mathbf{y}\right)\right)$ in general; however, some judicious choices of $\mathbf{v}\left(\mathbf{y},\mathbf{x},\boldsymbol{\theta}\right)$ lead to various lower bounds [10][11][12]. Thus, the $\mathbf{v}_{CRB}\left(\mathbf{y},\mathbf{x},\boldsymbol{\theta}\right)$ leading to HCRB is a limiting form of the $\mathbf{v}\left(\mathbf{y},\mathbf{x},\boldsymbol{\theta}\right)$ leading to the HBB which may be derived from the McAulay-Seidman bound (MSB) [19]. The MSB is the usual approximation of the Barankin Bound (BB) [20], the greatest lower bound on the MSE on deterministic parameters $\boldsymbol{\theta}$, resulting from a discretization of the uniform unbiasedness definition[2]:

$$E_{\mathbf{y}|\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\theta}}\left(\mathbf{y}\right)\right]=\int_{\Omega}\widehat{\boldsymbol{\theta}}\left(\mathbf{y}\right)p\left(\mathbf{y}|\boldsymbol{\theta}\right)d\mathbf{y}=\boldsymbol{\theta},\forall\boldsymbol{\theta}\in\Pi_d,$$

expressed for a subset of I test points [19][21][22]:

$$E_{\mathbf{y}|\boldsymbol{\theta}+\mathbf{h}_{di}}\left[\widehat{\boldsymbol{\theta}}\left(\mathbf{y}\right)\right]=\boldsymbol{\theta}+\mathbf{h}_{di},\ \boldsymbol{\theta}+\mathbf{h}_{di}\in\Pi_d,1\leq i\leq I. \tag{19}$$

Unfortunately, in hybrid estimation the MSB can hardly ever be computed to lower bound the MSE of unbiased estimators of $\widehat{\boldsymbol{\theta}}$ since a closed-form expression of $p\left(\mathbf{y}|\boldsymbol{\theta}\right)$ hardly ever exists. Interestingly enough, this major stumbling block can be bypassed for a general class of pdf. Let $1_A\left(\mathbf{y}\right)$ denote the indicator function of subset A of $\mathbb{R}^P$. Then, some easy integral calculuses show that, for any $\mathbf{h}_{ri}\in\mathbb{R}^P$ for which:

$$1_{\Pi_r}\left(\mathbf{x}+\mathbf{h}_{ri}\right)=1_{\Pi_r}\left(\mathbf{x}\right),\ \forall\mathbf{x}\in\mathbb{R}^P, \tag{20}$$

[2]For a straightforward extension to biased estimators see [21][22].

subject to (20), (19) can be recasted as:

$$E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}}\left(\mathbf{y}\right)-\boldsymbol{\theta}\right)v\left(\mathbf{y},\boldsymbol{\theta},\mathbf{h}_{ri},\mathbf{h}_{di}\right)\right]=\mathbf{h}_{di}, \tag{21}$$

$$v\left(\mathbf{y},\boldsymbol{\theta},\mathbf{h}_r,\mathbf{h}_d\right)=\left|\begin{array}{l}\frac{p(\mathbf{y},\mathbf{x}+\mathbf{h}_r|\boldsymbol{\theta}+\mathbf{h}_d)}{p(\mathbf{y},\mathbf{x}|\boldsymbol{\theta})}-1,\ (\mathbf{x},\boldsymbol{\theta})\in\boldsymbol{\Theta},\\ 0,\text{ otherwise.}\end{array}\right.$$

Additionally, if $\mathbf{h}_{ri}$ also satisfies:

$$1_{\Pi_r}\left(\mathbf{x}-\mathbf{h}_{ri}\right)=1_{\Pi_r}\left(\mathbf{x}\right),\ \forall\mathbf{x}\in\mathbb{R}^P, \tag{22}$$

then, some other easy integral calculuses show that, subject to (20) and (22), any estimator $\widehat{\mathbf{x}}$ of $\mathbf{x}$ satisfies:

$$\begin{aligned}E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\left(\widehat{\mathbf{x}}\left(\mathbf{y}\right)-\mathbf{x}\right)v\left(\mathbf{y},\boldsymbol{\theta},\mathbf{h}_{ri},\mathbf{h}_{di}\right)\right]=\mathbf{h}_{ri}+\\ E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}+\mathbf{h}_{di}}\left[\widehat{\mathbf{x}}\left(\mathbf{y}\right)-\mathbf{x}\right]-E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\widehat{\mathbf{x}}\left(\mathbf{y}\right)-\mathbf{x}\right].\end{aligned} \tag{23}$$

Unlike what is stated in [2][6][13], first, (23) proves that for each class of estimates satisfying:

$$E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\left(\widehat{\mathbf{x}}\left(\mathbf{y}\right)-\mathbf{x};\widehat{\boldsymbol{\theta}}\left(\mathbf{y}\right)-\boldsymbol{\theta}\right)\right]=\left(\boldsymbol{\mu}\left(\boldsymbol{\theta}\right);\mathbf{0}\right),$$

a specific HBB (16) (and a specific HCRB) can be computed. The class of wide-sense unbiased estimates [13, (11)] is only the particular case where $\boldsymbol{\mu}\left(\boldsymbol{\theta}\right)=\mathbf{0}$. Second, the regularity condition (20)(22) required for the existence of the HBB only imposes on $1_{\Pi_r}\left(\mathbf{x}\right)$ to be of the following form:

$$1_{\Pi_r}\left(\mathbf{x}\right)=\left|\begin{array}{l}0\text{ if }\sum\limits_{\mathbf{h}_r\in A}\left(\sum\limits_{l\in\mathbb{Z}}1_{\Pi_r^0}\left(\mathbf{x}+l\mathbf{h}_r\right)\right)=0,\\ 1,\text{ otherwise,}\end{array}\right. \tag{24}$$

where A and Π_r^0 are subsets of $\mathbb{R}^P$, what means that Π_r may be a discrete subset of $\mathbb{R}^P$ or a subset of intervals of $\mathbb{R}^P$. Since the HCRB is the limiting case of the HBB obtained for:

$$\begin{aligned}\mathbf{v}\left(\mathbf{y},\mathbf{x},\boldsymbol{\theta}\right)=\left(v\left(\mathbf{y},\boldsymbol{\theta},\mathbf{u}_1h_{r1},\mathbf{0}\right),\ldots,v\left(\mathbf{y},\boldsymbol{\theta},\mathbf{u}_Ph_{rP},\mathbf{0}\right),\right.\\ \left.v\left(\mathbf{y},\boldsymbol{\theta},\mathbf{0},\mathbf{u}_{P+1}h_{d1}\right),\ldots,v\left(\mathbf{y},\boldsymbol{\theta},\mathbf{0},\mathbf{u}_{P+D}h_{dD}\right)\right)\end{aligned}$$

where $\mathbf{u}_i$ is the ith column of the identity matrix, and by letting $\left(h_{r1},\ldots,h_{rP},h_{d1},\ldots,h_{dD}\right)$ be infinitesimally small, therefore (24) reduces to: $\forall\mathbf{x}\in\mathbb{R}^P,1_{\Pi_r}\left(\mathbf{x}\right)=1$, that is $\Pi_r=\mathbb{R}^P$ (R1). Last, the simplest form of $\mathbf{C}$ (18), i.e. $\mathbf{C}=\mathbf{I}$, is obtained for the class of estimates satisfying (R3):

$$E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\left(\widehat{\mathbf{x}}\left(\mathbf{y}\right)-\mathbf{x};\widehat{\boldsymbol{\theta}}\left(\mathbf{y}\right)-\boldsymbol{\theta}\right)\right]=\left(\boldsymbol{\mu};\mathbf{0}\right), \tag{25}$$

leading to the usual form of the HCRB (16):

$$\mathbf{HCRB}_{\mathbf{x},\boldsymbol{\theta}}=E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\frac{\partial\ln p\left(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}\right)}{\partial\left(\mathbf{x};\boldsymbol{\theta}\right)}\frac{\partial\ln p\left(\mathbf{y},\mathbf{x}|\boldsymbol{\theta}\right)}{\partial\left(\mathbf{x};\boldsymbol{\theta}\right)^T}\right]^{-1} \tag{26}$$

which contains elements with finite modulus provided that (R2): $E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\frac{\partial\ln p(\mathbf{y},\mathbf{x}|\boldsymbol{\theta})}{\partial x_p}^2\right],E_{\mathbf{y},\mathbf{x}|\boldsymbol{\theta}}\left[\frac{\partial\ln p(\mathbf{y},\mathbf{x}|\boldsymbol{\theta})}{\partial\theta_d}^2\right]<\infty$. Since the standard form of the HFIM (2)(6) [6][10][13] shares a common analytical form with the posterior BFIM [10]:

$$\mathbf{J}_{\overline{\mathbf{x}}_{0:k}}=E_{\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}}\left[\frac{-\partial^2\ln p\left(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}\right)}{\partial\overline{\mathbf{x}}_{0:k}\partial\overline{\mathbf{x}}_{0:k}^T}\right],$$

one might think that the posterior BFIM becomes naturally a HFIM (6) when some random parameters are transformed into deterministic parameters by discarding their prior information as mentioned in [15] and [16]. For example, if we consider that $p\left(\mathbf{x}_0\right)$ is unknown, then setting $\mathbf{J}_{\mathbf{x}_0}=\mathbf{0}$ in (15) as suggested in [15][16], is equivalent to transform $p\left(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}\right)$ into:

$$p\left(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{1:k}|\mathbf{x}_0\right)=\prod_{l=1}^{k}p\left(\mathbf{y}_l|\mathbf{x}_l\right)p\left(\mathbf{x}_l|\mathbf{x}_{l-1}\right). \tag{27}$$

Even if the recursion (15) still holds when computed with the conditional pdf (27), the block matrices obtained are valid components of the recursive HFIM (10) if and only if (R1) is satisfied. Thus, if any of $p\left(\mathbf{x}_k|\mathbf{x}_{k-1}\right)$, $p\left(\mathbf{y}_k|\mathbf{x}_k\right)$ or $p\left(\mathbf{x}_0\right)$ have a support which is not, $\mathbb{R}^P$, $\mathbb{R}^N$ or $\mathbb{R}^P$, respectively [23], then $\widetilde{\mathbf{J}}_k^{\mathbf{x}_k,\mathbf{x}_k}$ in (15) is not a HFIM. Additionally, even if pdfs support are $\mathbb{R}^P$, $\mathbb{R}^N$ or $\mathbb{R}^P$, respectively, the HFIM obtained is valid only for estimates satisfying (R3) and are no longer valid for any realizable estimate as it is the case for the BFIM [23]. Therefore it is critical to understand that discarding prior information in the computation of the BCRB change thoroughly the estimation problem under consideration and that the correct rationale to address the computation of a HCRB for MDS problem is the one introduced.

IV. Appendix

Let $E\left[\ \right]\triangleq E_{\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta}}\left[\ \right]$. First $\mathbf{J}_{\overline{\mathbf{x}}_{0:k},\boldsymbol{\theta}}$ (6) can be broken down as:

$$\left[\begin{array}{ccc}\mathbf{J}_k^{11}&\mathbf{J}_k^{12}&\mathbf{J}_k^{13}\\ \mathbf{J}_k^{21}&\mathbf{J}_k^{22}&\mathbf{J}_k^{23}\\ \mathbf{J}_k^{31}&\mathbf{J}_k^{32}&\mathbf{J}_k^{33}\end{array}\right],\quad\begin{array}{l}\mathbf{J}_k^{11}=E\left[-\frac{\partial^2\ln p(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta})}{\partial\overline{\mathbf{x}}_{0:k-1}\partial\overline{\mathbf{x}}_{0:k-1}^T}\right],\\ \mathbf{J}_k^{12}=E\left[-\frac{\partial^2\ln p(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta})}{\partial\overline{\mathbf{x}}_{0:k-1}\partial\mathbf{x}_k^T}\right],\end{array}$$

$$\mathbf{J}_k^{22}=E\left[-\frac{\partial^2\ln p(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta})}{\partial\mathbf{x}_k\partial\mathbf{x}_k^T}\right],\mathbf{J}_k^{13}=E\left[-\frac{\partial^2\ln p(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta})}{\partial\overline{\mathbf{x}}_{0:k-1}\partial\boldsymbol{\theta}^T}\right],$$

$$\mathbf{J}_k^{23}=E\left[-\frac{\partial^2\ln p(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta})}{\partial\mathbf{x}_k\partial\boldsymbol{\theta}^T}\right],\mathbf{J}_k^{33}=E\left[-\frac{\partial^2\ln p(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right].$$

Therefore, using block matrix inversion [24, p293]:

$$\begin{array}{l}\widetilde{\mathbf{J}}_k^{\mathbf{x}_k,\mathbf{x}_k}=\mathbf{J}_k^{22}-\mathbf{J}_k^{21}\left(\mathbf{J}_{k-1}^{11}\right)^{-1}\mathbf{J}_k^{12}\\ \widetilde{\mathbf{J}}_k^{\boldsymbol{\theta},\mathbf{x}_k}=\mathbf{J}_k^{23}-\mathbf{J}_k^{21}\left(\mathbf{J}_{k-1}^{11}\right)^{-1}\mathbf{J}_k^{13}\\ \widetilde{\mathbf{J}}_k^{\boldsymbol{\theta},\boldsymbol{\theta}}=\mathbf{J}_k^{33}-\mathbf{J}_k^{31}\left(\mathbf{J}_{k-1}^{11}\right)^{-1}\mathbf{J}_k^{13}\end{array} \tag{28}$$

Moreover, for MDS, (4) leads to:

$$p\left(\overline{\mathbf{y}}_{1:k},\overline{\mathbf{x}}_{0:k}|\boldsymbol{\theta}\right)=p\left(\mathbf{y}_k|\mathbf{x}_k,\boldsymbol{\theta}\right)p\left(\mathbf{x}_k|\mathbf{x}_{k-1},\boldsymbol{\alpha},\boldsymbol{\beta}\right)p\left(\overline{\mathbf{y}}_{1:k-1},\overline{\mathbf{x}}_{0:k-1}|\boldsymbol{\theta}\right)$$

yielding:

$$\mathbf{J}_k^{11}=\left[\begin{array}{cc}\mathbf{J}_{k-1}^{11}&\mathbf{J}_{k-1}^{12}\\ \mathbf{J}_{k-1}^{21}&\mathbf{J}_{k-1}^{22}+\mathbf{D}_{k-1}^{11}\end{array}\right],\ \mathbf{J}_k^{12}=\left[\begin{array}{c}\mathbf{0}\\ \mathbf{D}_{k-1}^{12}\end{array}\right],\ \mathbf{J}_k^{22}=\mathbf{D}_{k-1}^{22},$$

$$\mathbf{J}_k^{13}=\left[\begin{array}{c}\mathbf{J}_{k-1}^{13}\\ \mathbf{J}_{k-1}^{23}+\mathbf{D}_{k-1}^{13}\end{array}\right],\ \mathbf{J}_k^{23}=\mathbf{D}_{k-1}^{23},\ \mathbf{J}_k^{33}=\mathbf{J}_{k-1}^{33}+\mathbf{D}_{k-1}^{33}.$$

where $\mathbf{D}_{k-1}^{ij}$ are given by (12)(13). Second, using once again block matrix inversion:

$$\begin{array}{l}\left(\mathbf{J}_k^{11}\right)^{-1}=\left[\begin{array}{cc}\boldsymbol{\Phi}_{k-1}&-\boldsymbol{\Gamma}_{k-1}\boldsymbol{\Delta}_{k-1}^{-1}\\ -\boldsymbol{\Delta}_{k-1}^{-1}\boldsymbol{\Gamma}_{k-1}^T&\boldsymbol{\Delta}_{k-1}^{-1}\end{array}\right],\\ \boldsymbol{\Delta}_{k-1}=\mathbf{D}_{k-1}^{11}+\mathbf{J}_{k-1}^{22}-\mathbf{J}_{k-1}^{21}\left(\mathbf{J}_{k-1}^{11}\right)^{-1}\mathbf{J}_{k-1}^{12},\\ \boldsymbol{\Gamma}_{k-1}=\left(\mathbf{J}_{k-1}^{11}\right)^{-1}\mathbf{J}_{k-1}^{12},\\ \boldsymbol{\Phi}_{k-1}=\left(\mathbf{J}_{k-1}^{11}\right)^{-1}+\boldsymbol{\Gamma}_{k-1}\boldsymbol{\Delta}_{k-1}^{-1}\boldsymbol{\Gamma}_{k-1}^T\end{array}$$

Finally, by noting that $\boldsymbol{\Delta}_{k-1}=\mathbf{D}_{k-1}^{11}+\widetilde{\mathbf{J}}_{k-1}^{\mathbf{x}_{k-1},\mathbf{x}_{k-1}}$, a few additional lines of calculus allows to show that equivalent forms of $\widetilde{\mathbf{J}}_k^{\mathbf{x}_k,\mathbf{x}_k}$, $\widetilde{\mathbf{J}}_k^{\boldsymbol{\theta},\boldsymbol{\theta}}$and $\widetilde{\mathbf{J}}_k^{\boldsymbol{\theta},\mathbf{x}_k}$ in (28) are given by (11).

REFERENCES

[1] Y. Rockah and P. Schultheiss, "Array shape calibration using sources in unknown locations–part I: Far-field sources," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 35, no. 3, pp. 286-299, Mar. 1987.

[2] I. Reuven and H. Messer, "A Barankin-type lower bound on the estimation error of a hybrid parameter vector," IEEE Transactions on Information Theory, vol. 43, no. 3, pp. 1084-1093, May 1997.

[3] H. L. Van Trees, *Detection, Estimation and Modulation theory: Optimum Array Processing*. New-York, NY, USA: John Wiley & Sons, Mar. 2002, vol. 4.

[4] P. Tichavsk´y and K. Wong, "Quasi-fluid-mecanics based quasi-Bayesian Cramér-Rao bounds for towed-array direction finding," IEEE Transactions on Signal Processing, vol. 52, no. 1, pp. 36-47, Jan. 2004.

[5] S. Buzzi, M. Lops, and S. Sardellitti, "Further result on Cramér-Rao bounds for parameter estimation in long-code DS/CDMA systems," IEEE Transactions on Signal Processing, vol. 53, no. 3, pp. 1216-1221, Mar. 2005.

[6] S. Bay, B. Geller, A. Renaux, J.-P. Barbot, and J.-M. Brossier, "On the hybrid Cramér-Rao bound and its application to dynamical phase estimation," IEEE Signal Processing Letters, vol. 15, pp. 453-456, 2008.

[7] K. Todros and J. Tabrikian, "Hybrid lower bound via compression of the sampled CLR function," in Proc. of IEEE Workshop on Statistical Signal Processing (SSP), Cardiff, Wales, UK, Aug. 2009, pp. 602-605.

[8] J. Yang, B. Geller, and S. Bay, "Bayesian and hybrid Cramér-Rao bounds for the carrier recovery under dynamic phase uncertain channels," IEEE Transactions on Signal Processing, vol. 59, no. 2, pp. 667-680, 2011.

[9] J. Vila-Valls, L. Ros, and J. M. Brossier, "Joint oversampled carrier and time-delay synchronization in digital communications with large excess bandwidth," ELSEVIER Signal Processing, vol. 92, no. 1, pp. 76-88, Jan. 2012.

[10] H. L. Van Trees and K. L. Bell, Eds., Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking. New-York, NY, USA: Wiley/IEEE Press, Sep. 2007.

[11] C. Ren, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux, "Hybrid lower bound on the MSE based on the Barankin and Weiss-Weinstein bounds," in Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vancouver, Canada, May 2013, pp. 5534-5538.

[12] C. Ren, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux, "A Ziv-Zakaï type bound for hybrid parameter estimation," in Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy, May 2014, pp. 4663-4667.

[13] Y. Noam and H. Messer, "Notes on the tightness of the hybrid Cramér-Rao lower bound," IEEE Transactions on Signal Processing, vol. 57, no. 6, pp. 2074–2084, 2009.

[14] D. Simon, *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*, Wiley InterScience, 2006

[15] P. Tichavsky, C. Muravchik and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear Filtering", IEEE Transactions on Signal Processing, vol. 46, no. 5, pp. 1386-1396, 1998

[16] M. Simandl, J. Kralovec and P. Tichavsky, "Filtering, predictive, and smoothing Cramér-Rao bounds for discrete-time nonlinear dynamic systems,", Automatica, vol 37, pp. 1703-1716, 2001

[17] B.Z. Bobrovsky, E. Mayer-Wolf and M. Zakai, "Some Classes of Global Cramer-Rao Bounds", The Annals of Statistics, 15(4): 1421-1438, 1987

[18] E. L. Lehmann and G. Casella, *Theory of Point Estimation (2nd ed.)*. Springer, 1998

[19] R. McAulay, L.P. Seidman, "A useful form of the Barankin Lower Bound and its application to PPM threshold analysis", IEEE Trans. on IT, 15(2): 273-279, 1969

[20] E.W. Barankin, "Locally best unbiased estimates", Ann. Math. Stat., 20(4): 477-501, 1949

[21] K. Todros and J. Tabrikian, "General Classes of Performance Lower Bounds for Parameter Estimation-Part I: Non-Bayesian Bounds for Unbiased Estimators", IEEE Trans. on IT, 56(10): 5064-5082, 2010

[22] E. Chaumette, J. Galy, A. Quinlan, P. Larzabal, "A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum-Likelihood Estimators", IEEE Trans. on SP, 56(11): 5319-5333, 2008

[23] E. Weinstein and A. J. Weiss, "A general class of lower bounds in parameter estimation," IEEE Trans. Inf. Theory, vol. IT-34, no. 2, pp. 338–342, Mar. 1988.

[24] G.A.F. Seber, *Matrix Handbook forr Statisticians*, Wiley Series in Probability and Statistics, 2008

Annexe C

On the Accuracy and Resolvability of Vector Parameter Estimates

IEEE Transactions on Signal Processing, Volume : 62, Issue : 14, July. 2014, pp. 3682-3694

3682 IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 62, NO. 14, JULY 15, 2014

On the Accuracy and Resolvability of Vector Parameter Estimates

Chengfang Ren, Mohammed Nabil El Korso, Jérôme Galy, Eric Chaumette, Pascal Larzabal, *Member, IEEE*, and Alexandre Renaux

***Abstract*—In this paper we address the problem of fundamental limitations on resolution in deterministic parameters estimation. We introduce a definition of resolvability based on probability and incorporating a requirement for accuracy unlike most existing definitions. Indeed in many application the key problem is to obtain distributions of estimates that are not only distinguishable but also accurate and compliant with a required precision. We exemplify the proposed definition with estimators that produce normal estimates, as in the conditional model for which the Gaussianity and efficiency of maximum likelihood estimators (MLEs) in the asymptotic region of operation (in terms of signal-to-noise ratio and/or in large number of snapshots) is well established, even for a single snapshot. In order to measure the convergence in distribution, we derive a simple test allowing to check whether the conditional MLEs operate in the asymptotic region of operation. Last, we discuss the resolution of two complex exponentials with closely spaced frequencies and compare the results obtained with the ones provided by the various statistical resolution limit released in the open literature.**

***Index Terms*—Cramer-Rao bound, minimum probability of error, parameter estimation, performance analysis, resolution, statistical resolution limit.**

I. Introduction

In many applications (as target classification in radar or identification of constellation diagrams in telecommunication) the key problem is to obtain distributions of estimates that are not only distinguishable but also accurate and compliant both with a required precision and a given probability of confidence, these two features being requested to assess the performance of subsequent processing (a classifier for instance). Hence the need of a definition of resolvability based on probability and incorporating a requirement for accuracy unlike most existing definitions. Indeed, strictly speaking, the idea of resolvability in the signal processing open literature does not necessarily take into account how well parameters are estimated, only whether multiple signals can be distinguished [1]–[6].

Interestingly enough, there is little contribution to this approach in papers or monographs [7]–[10]. Oh and Kashyap [11], Section V seem to have introduced the first theoretical definition on the resolution based on probability and incorporating a requirement for accuracy, in the simplified case of two signal sources with a single unknown parameter (frequency of two complex exponentials) in order to characterize their separability (resolvability) by a proposed robust estimation method. Later, this work has been substantially extended by Clark [12] to normally distributed vector parameter estimates with covariance matrix equal to the Cramér-Rao bound (CRB). Since the CRB gives the minimum variance attainable by any unbiased estimates, it can be used to find a lower bound on the actual resolution threshold of any normally distributed estimators. Then, at the expense of neglecting the cross-information (cross-covariance) between estimators (as in [11]), Clark has shown that an extension of the definition on the resolution given by [11] to a vector of parameters for each signal source can be obtained with a metric depending on the distance between ellipsoids of constant probability. When the ellipsoids are disjoint the parameters are deemed resolvable. Based on the analysis of the distance problem, a simple method was developed for finding the noise level at which the ellipsoids become tangent, which is the resolution threshold noise level. Unfortunately, in many applications the hypothesis of uncorrelated estimators is unrealistic and the cross-covariance terms of the associated CRB cannot be neglected (see for instance (29)). To bypass this limitation of Clark's approach, we present a definition of the probability of resolvability based on the cumulative distribution function (c.d.f.) which is the most general way to characterize a random vector, whichever its distribution. Thus, the proposed definition does not depend on the nature of the distributions of estimates (as in [11] and [12]) and allows to take into account their cross-information (unlike in [11] and [12]). As the evaluation (by calculus or by simulation) of a P-dimensional c.d.f. is generally difficult, we propose a general method to generate the lower and upper bounds of confidence intervals on the probability of resolvability for a given required estimation precision. This method is based on the approximation of hyper-rectangles

Manuscript received October 16, 2013; revised February 21, 2014; accepted May 14, 2014. Date of publication June 03, 2014; date of current version June 25, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stefano Marano. This work was supported by the European Network of Excellence NEWCOM, by the iCODE Institute, Research Project of the IDEX Paris-Saclay, by the DGA/DGCIS, and by the Display-Mastodons project of CNRS. This work appeared in part in the *Proceedings of the 21st European Signal Processing Conference*, Marrakech, Morocco, September 9–13, 2013 [13].

C. Ren and A. Renaux are with the Universite Paris-Sud, 91192 Gif-sur-Yvette, France (e-mail: cren@lss.supelec.fr; renaux@lss.supelec.fr).

M. N. E. Korso is with the Laboratoire Energétique Mécanique Electromagnétisme, Université Paris Ouest Nanterre La Défense, IUT de Ville d'Avray, France (e-mail: m.elkorso@u-paris10.fr).

J. Galy is with the Université de Montpellier, 34392 Montpellier, France (e-mail: galy@lirmm.fr).

E. Chaumette is with Department of Electronics, Optronics, and Signal, University of Toulouse, 31055 Toulouse, France (e-mail: eric.chaumette@isae.fr).

P. Larzabal is with the Université Paris-Sud/SATIE, 94235 Cachan, France (e-mail: pascal.larzabal@satie.ens-cachan.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2014.2328322

1053-587X © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

by some ad hoc hypervolumes (generalization of hyper-ellipsoids in non-Euclidean norms) which can be used to substitute the computation of a 1-dimensional c.d.f. for the computation of a P-dimensional c.d.f.

We exemplify the proposed definition (and bounds usefulness) with estimators that produce normal estimates, as in the conditional model for which the Gaussianity and efficiency of MLEs in the asymptotic region of operation (in terms of signal-to-noise ratio (SNR) and/or in large number of snapshots) is well established, even for a single snapshot. Indeed for normal estimates, numerical methods exist (and some closed-form expressions as well in the simplest cases) to compute a confidence interval on the probability of resolvability. Additionally, the conditional model is the privileged model for an active system. In order to measure the convergence in distribution, we have derived a simple test allowing to check whether the conditional maximum likelihood estimators operate in the asymptotic region of operation.

Although the proposed resolvability definition is not limited to a given number of signal sources (see [13] for additional examples), we focus in the present paper on the resolvability of two closely spaced frequencies in order to compare the various definition of statistical resolution limit (SRL) released in the open literature and a SRL deriving from the proposed resolvability definition. The underlying idea is to measure the cost, in terms of scenario requirements (SNR, number of snapshots, number of sensors), of a SRL deriving from a resolvability definition which takes into account not only the ability to distinguish multiple signals (as all definitions do) but also how well parameters are estimated. In other words, we intend to bring to the reader attention that most previously released SRL criteria only take into account the ability to distinguish multiple signals and therefore do not take into account all aspects of performance from an estimation point of view. The significant differences in the region of operation (for a given observation model) requested to fulfill the different SRL criteria should question the reader on which information he looks for: distinguishability only or distinguishability and accuracy, since in order to get both there is a price to pay.

Last, from a practical point of view, the results obtained with the proposed rationale must be regarded as a "lower bound" in terms of actual scenario requirements (SNR, number of snapshots, number of sensors), in the sense that it assumes that the number of signals is known and that all the signals are present as well.

The paper is organized as follows. Section II is tutorial in nature. It presents the proposed resolvability definition and a general scheme to generate the lower and upper bounds of confidence intervals on the probability of resolvability for a given required estimation precision. In Section III we introduce a simple test allowing to check whether the conditional MLEs operate in the asymptotic region of operation in order to exploit its asymptotic Gaussianity and efficiency. Last in Section IV, as an application example, we discuss the resolution of two complex exponentials with closely spaced frequencies in order to compare the various definition of SRL available in the open literature.

II. Probability of Resolvability

In the following, $\mathbf{x}$ denotes the random observation vector of dimension N, Ω denotes the observation space and $L^2(\Omega)$ denotes the complex Hilbert space of square integrable functions over Ω. The probability density function (p.d.f.) of $\mathbf{x}$ is denoted $p(\mathbf{x};\boldsymbol{\Theta})$ and depends on a vector of P real parameters $\boldsymbol{\Theta} = (\theta_1, \ldots, \theta_P) \in \Phi$, where Φ denotes the parameter space. The probability of an event $\mathcal{D} \subset \Omega$ is denoted $\mathcal{P}(\mathcal{D};\boldsymbol{\Theta})$. Let $\boldsymbol{\Theta}^0$ be a selected value of the parameter $\boldsymbol{\Theta}$, and $\widehat{\mathbf{g}\left(\boldsymbol{\Theta}^0\right)}$ an estimator of $\mathbf{g}\left(\boldsymbol{\Theta}^0\right)$ where $\mathbf{g}(\boldsymbol{\Theta}) = (g_1(\boldsymbol{\Theta}), \ldots, g_Q(\boldsymbol{\Theta}))^T$ is a vector of Q real-valued functions of $\boldsymbol{\Theta}$. For any selected value $\boldsymbol{\Theta}^0$, $\widehat{\mathbf{g}\left(\boldsymbol{\Theta}^0\right)} \triangleq \mathbf{g}\left(\widehat{\boldsymbol{\Theta}^0}\right)(\mathbf{x})$ stands for a mapping of the observation space Ω into an estimate of $\mathbf{g}\left(\boldsymbol{\Theta}^0\right)$.

A. Approximation of a Hyper-Rectangle

In the following, $\boldsymbol{\Xi}^- = \left(\varepsilon_1^-, \ldots, \varepsilon_Q^-\right)^T$, $\boldsymbol{\Xi}^+ = \left(\varepsilon_1^+, \ldots, \varepsilon_Q^+\right)^T$, $\mathbf{s} = (s_1, \ldots, s_Q)^T$ are vectors of $\mathbb{R}^Q$ which components verify $\varepsilon_q^- > 0$, $\varepsilon_q^+ > 0$ and $s_q \geq 1$. Then $R\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$, $\mathbf{y}^0 \in \mathbb{R}^Q$, denotes the hyper-rectangle (also called an orthotope or a box) of $\mathbb{R}^Q$ containing $\mathbf{y}^0$ defined by:

$$R\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) = \left\{ y : \bigwedge_{q=1}^{Q} \left(y_q \in \left]y_q^0 - \varepsilon_q^-, y_q^0 + \varepsilon_q^+\right[\right) \right\} \tag{1}$$

or equivalently by:

$$R\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) = \left\{ \mathbf{y} : \bigwedge_{q=1}^{Q} \left(\left| y_q - y_q^0 - \frac{d\varepsilon_q}{2} \right| < \overline{\varepsilon_q} \right) \right\} \tag{2}$$

where:

$$d\varepsilon_q = \varepsilon_q^+ - \varepsilon_q^- = \left(\boldsymbol{\Xi}^+\right)_q - \left(\boldsymbol{\Xi}^-\right)_q = (d\boldsymbol{\Xi})_q,$$
$$\overline{\varepsilon_q} = \frac{\varepsilon_q^+ + \varepsilon_q^-}{2} = \frac{\left(\boldsymbol{\Xi}^+\right)_q + \left(\boldsymbol{\Xi}^-\right)_q}{2} = \left(\overline{\boldsymbol{\Xi}}\right)_q.$$

Let $\Sigma_{\mathbf{s}}^{\lambda}\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$, $\lambda > 0$, be the Q-dimensional hypervolume defined by:

$$\Sigma_{\mathbf{s}}^{\lambda}\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) = \left\{ \mathbf{y} : \sum_{q=1}^{Q} \left| \frac{y_q - y_q^0 - \frac{d\varepsilon_q}{2}}{\overline{\varepsilon_q}} \right|^{s_q} < \lambda \right\} \tag{3}$$

$\Sigma_{\mathbf{s}}^{\lambda}\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$ is the generalization of a hyper-ellipsoid obtained when $s_q = 2$. Let us note that $\mathbf{y}^0$ is the centre of $R\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$ and $\Sigma_{\mathbf{s}}^{\lambda}\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$ if and only if $\boldsymbol{\Xi}^- = \boldsymbol{\Xi}^+ = \boldsymbol{\Xi}$. In this case $R\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$ and $\Sigma_{\mathbf{s}}^{\lambda}\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$ are simply denoted $R\left(\mathbf{y}^0, \boldsymbol{\Xi}\right)$ and $\Sigma_{\mathbf{s}}^{\lambda}\left(\mathbf{y}^0, \boldsymbol{\Xi}\right)$. A first result is the existence of lower and upper bounds on $R\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$ in the inclusion sense:

$$\Sigma_{\mathbf{s}}^{1}\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) \subset R\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) \subset \Sigma_{\mathbf{s}}^{Q}\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) \tag{4}$$

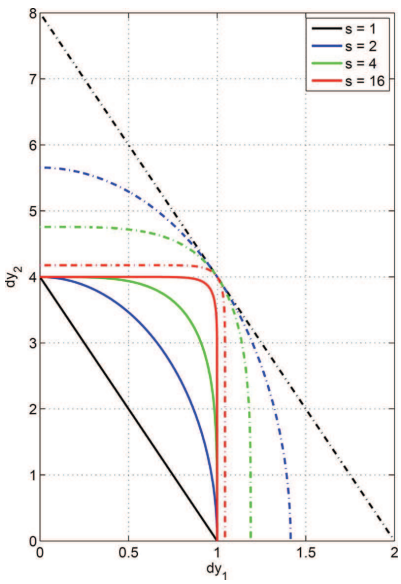

Fig. 1. Approximation of $R(\mathbf{y}^0, \Xi) = \{\mathbf{y} : (|dy_1| < 1) \wedge (|dy_2| < 4)\}$, where $dy_1 = y_1 - y_1^0$ and $dy_2 = y_2 - y_2^0$, by $\Sigma_{\mathbf{s}}^{\lambda}(\mathbf{y}^0, \Xi)$ where $\mathbf{s}^T = (s, s)$ and $\lambda \in \{1, 2\}$, for $s \in \{1, 2, 4, 16\}$, $\Xi = (1, 4)^T$. The solid lines are for $\lambda = 1$ and the dash-dot lines are for $\lambda = 2$.

as shown in Appendix VI-A. Second, since:

$$\sum_{q=1}^{Q} \left| \frac{y_q - y_q^0 - \frac{d\varepsilon_q}{2}}{\overline{\varepsilon_q}} \right|^{s_q} < Q \Leftrightarrow \sum_{q=1}^{Q} \left| \frac{y_q - y_q^0 - \frac{d\varepsilon_q}{2}}{Q^{\frac{1}{s_q}} \overline{\varepsilon_q}} \right|^{s_q} < 1$$

and $\lim_{s \to \infty} Q^{1/s} = 1$, then:

$$\lim_{\min\left\{\{s_q\}_{q=1}^{Q}\right\} \to \infty} \Sigma_{\mathbf{s}}^{Q}\left(\mathbf{y}^0, \Xi^-, \Xi^+\right) - \Sigma_{\mathbf{s}}^{1}\left(\mathbf{y}^0, \Xi^-, \Xi^+\right) = 0$$

and:

$$\lim_{\min\left\{\{s_q\}_{q=1}^{Q}\right\} \to \infty} \Sigma_{\mathbf{s}}^{1}\left(\mathbf{y}^0, \Xi^-, \Xi^+\right) = R\left(\mathbf{y}^0, \Xi^-, \Xi^+\right). \quad (5)$$

Therefore any hyper-rectangle $R\left(\mathbf{y}^0, \Xi^-, \Xi^+\right)$ of $\mathbb{R}^Q$ can be lower and upper bounded by Q-dimensional hypervolumes $\Sigma_{\mathbf{s}}^{1}\left(\mathbf{y}^0, \Xi^-, \Xi^+\right)$ and $\Sigma_{\mathbf{s}}^{Q}\left(\mathbf{y}^0, \Xi^-, \Xi^+\right)$ that converge towards the hyper-rectangle as $\min\left\{\{s_q\}_{q=1}^{Q}\right\}$ increases. This is illustrated by Fig. 1 in the special case of the approximation of a rectangle ($Q = 2$, more precisely, the upper right quarter of the rectangle) by $\Sigma_{\mathbf{s}}^{\lambda}\left(\mathbf{y}^0, \Xi\right)$, $\mathbf{s}^T = (s, s)$, $\lambda \in \{1, 2\}$ (solid line for $\lambda = 1$ and dash-dot line for $\lambda = 2$) showing how the increase of s allows very tight bounds (4).

B. Precision in Probability Sense and Bounds

The quality (i.e., the precision) of an estimator $\mathbf{g}\left(\widehat{\boldsymbol{\Theta}^0}\right)$ can be measured using the following exhaustive canonical risk function:

$$\mathcal{O}_{\boldsymbol{\Theta}^0}\left(\mathbf{g}\left(\widehat{\boldsymbol{\Theta}^0}\right), \Xi^-, \Xi^+\right) = \mathcal{P}\left(\mathbf{g}\left(\widehat{\boldsymbol{\Theta}^0}\right) \in R\left(\mathbf{g}\left(\boldsymbol{\Theta}^0\right), \Xi^-, \Xi^+\right); \boldsymbol{\Theta}^0\right) \quad (6)$$

where Ξ^- and Ξ^+ define the left and right errors on the estimation of $\mathbf{g}\left(\boldsymbol{\Theta}^0\right)$, and $\mathcal{O}_{\boldsymbol{\Theta}^0}\left(\mathbf{g}\left(\widehat{\boldsymbol{\Theta}^0}\right), \Xi^-, \Xi^+\right)$[1] is the probability that errors do not exceed Ξ^- and Ξ^+. Indeed according to (2):

$$\mathcal{O}_{\boldsymbol{\Theta}^0}\left(\mathbf{g}\left(\widehat{\boldsymbol{\Theta}^0}\right), \Xi^-, \Xi^+\right) = F_{\mathbf{y}}\left(\overline{\Xi}\right) \text{ where } (\mathbf{y})_q = \left| g_q\left(\widehat{\boldsymbol{\Theta}^0}\right)(\mathbf{x}) - g_q\left(\boldsymbol{\Theta}^0\right) - \frac{d\varepsilon_q}{2} \right|, \quad (7)$$

i.e., $\mathcal{O}_{\boldsymbol{\Theta}^0}\left(\mathbf{g}\left(\widehat{\boldsymbol{\Theta}^0}\right), \Xi^-, \Xi^+\right)$ is the c.d.f. of a random vector. This risk function is termed "canonical" since it is deduced naturally from the problem under study: the match between the observations of a random vector and a deterministic vector of interest. This risk function is also *exhaustive* , in the sense that it incorporates all the available information on the problem, in other words the probabilities. In the following $\mathcal{O}_{\boldsymbol{\Theta}^0}\left(\mathbf{g}\left(\widehat{\boldsymbol{\Theta}^0}\right), \Xi^-, \Xi^+\right)$ (6) defines the precision of estimator $\mathbf{g}\left(\widehat{\boldsymbol{\Theta}^0}\right)$ in probability sense which is bounded by (4):

$$\begin{aligned} &\mathcal{P}\left(\mathbf{g}\left(\widehat{\boldsymbol{\Theta}^0}\right) \in \Sigma_{\mathbf{s}}^{1}\left(\mathbf{g}\left(\boldsymbol{\Theta}^0\right), \Xi^-, \Xi^+\right); \boldsymbol{\Theta}^0\right) \leq \\ &\qquad \mathcal{O}_{\boldsymbol{\Theta}^0}\left(\mathbf{g}\left(\widehat{\boldsymbol{\Theta}^0}\right), \Xi^-, \Xi^+\right) \\ &\leq \mathcal{P}\left(\mathbf{g}\left(\widehat{\boldsymbol{\Theta}^0}\right) \in \Sigma_{\mathbf{s}}^{Q}\left(\mathbf{g}\left(\boldsymbol{\Theta}^0\right), \Xi^-, \Xi^+\right); \boldsymbol{\Theta}^0\right) \end{aligned} . \quad (8)$$

The underlying idea behind the use of probability bounds (8) is to substitute the computation of the c.d.f. of one random variable for the computation of a Q-dimensional integral over a Q-dimensional hyper-rectangle. In the following, for the sake of legibility:

- without loss of generality, $\mathbf{g}(\boldsymbol{\Theta}) = \boldsymbol{\Theta}$ $(Q = P)$,
- $\boldsymbol{\Theta}^0$ (respectively $\widehat{\boldsymbol{\Theta}^0}$) is denoted $\boldsymbol{\Theta}$ (respectively $\widehat{\boldsymbol{\Theta}}$) wherever it is unambiguous.

C. Probability of Resolvability

We consider a parameter estimation problem where the parameters of interest are the vectors $\{\boldsymbol{\theta}^m\}_{m=1}^{M}$, where $\boldsymbol{\theta} \in \mathbb{R}^{P'}$ and $\boldsymbol{\theta}^m \neq \boldsymbol{\theta}^l$, $\forall l \neq m \in [1, M]$. Among all various possible realizations of this setting, the most studied realization in signal processing is that of separating the components of data formed from a linear superposition of individual signals to noisy data

[1] $\Xi^- \neq \Xi^+$ takes into account that the p.d.f. of $g_q\left(\widehat{\boldsymbol{\theta}^0}\right)(\mathbf{x})$ is not expected to be symmetric at the vicinity of $g_q\left(\boldsymbol{\theta}^0\right)$ in the general case.

(nuisance). Each vector $\boldsymbol{\theta}^m$ parameterizes a single signal of this superposition.

In an ideal estimation problem, the set of estimators $\left\{\widehat{\boldsymbol{\theta}^m}\right\}_{m=1}^M$ would yield the set of unknown parameter vectors $\{\boldsymbol{\theta}^m\}_{m=1}^M$ for all trials : $\widehat{\boldsymbol{\theta}^m} = \boldsymbol{\theta}^m$. However in an actual estimation problem, the set of estimators $\left\{\widehat{\boldsymbol{\theta}^m}\right\}_{m=1}^M$ is a set of random vectors. Therefore given the statistics of the estimates, we desire to know when vector parameters $\{\boldsymbol{\theta}^m\}_{m=1}^M$, which may be closely spaced in $\mathbb{R}^{P'}$, can be resolved by the realizations of $\left\{\widehat{\boldsymbol{\theta}^m}\right\}_{m=1}^M$. So far, most of the definitions of resolvability that have already been proposed (except [12], see Section IV) focuses only on the ability to distinguish the distributions of estimates without any particular requirement on the estimates precision. However in many applications (as target classification in radar or identification of constellation diagram in telecommunication), the key problem is rather to assign the correct vector of estimates to each signal source with a required precision compliant with subsequent processing. This is the problem we want to address by introducing a definition of resolvability that combines the ability to distinguish[2] the distributions of estimates and to measure their precision (see also [12]). The rationale is the following. Let $\boldsymbol{\Theta}^T = \left(\left(\boldsymbol{\theta}^1\right)^T, \ldots, \left(\boldsymbol{\theta}^M\right)^T\right)$ $(P = P'M)$, $\left(\boldsymbol{\Xi}^+\right)^T = \left(\left(\boldsymbol{\varepsilon}^{1+}\right)^T, \ldots, \left(\boldsymbol{\varepsilon}^{M+}\right)^T\right)$ and $\left(\boldsymbol{\Xi}^-\right)^T = \left(\left(\boldsymbol{\varepsilon}^{1-}\right)^T, \ldots, \left(\boldsymbol{\varepsilon}^{M-}\right)^T\right)$ where $\boldsymbol{\varepsilon}^{m+} = \left(\varepsilon_1^{m+}, \ldots, \varepsilon_{P'}^{m+}\right)^T$ and $\boldsymbol{\varepsilon}^{m-} = \left(\varepsilon_1^{m-}, \ldots, \varepsilon_{P'}^{m-}\right)$. Any choice of $\boldsymbol{\Xi}^+$ and $\boldsymbol{\Xi}^-$ in such a manner that the set of M hyper-rectangles $R\left(\boldsymbol{\theta}^m, \boldsymbol{\varepsilon}^{m-}, \boldsymbol{\varepsilon}^{m+}\right)$ are disjoint, i.e.,

$$R\left(\boldsymbol{\theta}^m, \boldsymbol{\varepsilon}^{m-}, \boldsymbol{\varepsilon}^{m+}\right) \cap R\left(\boldsymbol{\theta}^l, \boldsymbol{\varepsilon}^{l-}, \boldsymbol{\varepsilon}^{l+}\right) = \varnothing,\ l \neq m, \quad (9)$$

defines a resolvability criterion, in the sense that any realization $\left\{\widehat{\boldsymbol{\theta}^m}\right\}_{m=1}^M$ that yields an unique estimate per $R\left(\boldsymbol{\theta}^m, \boldsymbol{\varepsilon}^{m-}, \boldsymbol{\varepsilon}^{m+}\right)$ can be regarded as successful. Nevertheless such a resolvability criterion does not guaranty that for all successful trials $\widehat{\boldsymbol{\theta}^m} \in R\left(\boldsymbol{\theta}^m, \boldsymbol{\varepsilon}^{m-}, \boldsymbol{\varepsilon}^{m+}\right)$, $1 \leq m \leq M$, since some estimate switches among hyper-rectangles $R\left(\boldsymbol{\theta}^m, \boldsymbol{\varepsilon}^{m-}, \boldsymbol{\varepsilon}^{m+}\right)$ are statistically possible. And this is exactly what we do not want to allow. Therefore we consider as successful only trials for which:

$$\begin{aligned} &\widehat{\boldsymbol{\theta}^m} \in R\left(\boldsymbol{\theta}^m, \boldsymbol{\varepsilon}^{m-}, \boldsymbol{\varepsilon}^{m+}\right), 1 \leq m \leq M, \text{ s.t.} \\ &R\left(\boldsymbol{\theta}^m, \boldsymbol{\varepsilon}^{m-}, \boldsymbol{\varepsilon}^{m+}\right) \cap R\left(\boldsymbol{\theta}^l, \boldsymbol{\varepsilon}^{l-}, \boldsymbol{\varepsilon}^{l+}\right) = \varnothing, l \neq m \end{aligned}$$

leading to the following probability of resolvability $\mathcal{P}_{\text{res}}$:

$$\begin{aligned} &\mathcal{P}_{\text{res}} = \mathcal{O}_{\boldsymbol{\Theta}}\left(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) \text{ where } \left(\boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) \text{ s.t.} \\ &R\left(\boldsymbol{\theta}^m, \boldsymbol{\varepsilon}^{m-}, \boldsymbol{\varepsilon}^{m+}\right) \cap R\left(\boldsymbol{\theta}^l, \boldsymbol{\varepsilon}^{l-}, \boldsymbol{\varepsilon}^{l+}\right) = \varnothing,\ l \neq m. \quad (10) \end{aligned}$$

[2]A similar idea appears in [14] in connection with the definition of optimally distinguishable distributions in order to obtain a partition in the model space via a partition in the parameter space.

In an ideal estimation problem:

$$\begin{aligned} &\mathcal{P}_{\text{res}} = 1 = \mathcal{O}_{\boldsymbol{\Theta}}\left(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right), \forall\left(\boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) \text{ s.t.} \\ &R\left(\boldsymbol{\theta}^m, \boldsymbol{\varepsilon}^{m-}, \boldsymbol{\varepsilon}^{m+}\right) \cap R\left(\boldsymbol{\theta}^l, \boldsymbol{\varepsilon}^{l-}, \boldsymbol{\varepsilon}^{l+}\right) = \varnothing,\ l \neq m. \end{aligned}$$

In an actual estimation problem, statistics of the estimates depends on the observation model and of its conditions of operation (SNR of sources, number of independent observations, ...). The requirements on precision (choice of $\boldsymbol{\Xi}_{\text{req}}^-$ and $\boldsymbol{\Xi}_{\text{req}}^+$) and on the probability of resolvability $\mathcal{P}_{\text{res}}$ (10) generally come from subsequent processing for which the set $\left\{\widehat{\boldsymbol{\theta}^m}\right\}_{m=1}^M$ is an input: they are application-defined parameters. Therefore for a given application, vectors of parameters $\{\boldsymbol{\theta}^m\}_{m=1}^M$ will be said "resolved" by estimators $\left\{\widehat{\boldsymbol{\theta}^m}\right\}_{m=1}^M$ with probability $\mathcal{P}_{\text{res}}$ if:

$$\begin{aligned} &\mathcal{O}_{\boldsymbol{\Theta}}\left(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Xi}_{\text{req}}^-, \boldsymbol{\Xi}_{\text{req}}^+\right) = \mathcal{P}_{\text{res}} \text{ where } \left(\boldsymbol{\Xi}_{\text{req}}^-, \boldsymbol{\Xi}_{\text{req}}^+\right) \text{ s.t.} \\ &R\left(\boldsymbol{\theta}^m, \boldsymbol{\varepsilon}_{\text{req}}^{m-}, \boldsymbol{\varepsilon}_{\text{req}}^{m+}\right) \cap R\left(\boldsymbol{\theta}^l, \boldsymbol{\varepsilon}_{\text{req}}^{l-}, \boldsymbol{\varepsilon}_{\text{req}}^{l+}\right) = \varnothing,\ l \neq m. \quad (11) \end{aligned}$$

As $\mathcal{P}_{\text{res}} = \mathcal{O}_{\boldsymbol{\Theta}}\left(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Xi}_{\text{req}}^-, \boldsymbol{\Xi}_{\text{req}}^+\right)$ is generally not computable, the use of bounds (8) whenever they can be computed more easily (for example if $\widehat{\boldsymbol{\Theta}}$ is a Gaussian random vector, see hereinafter) may allow to check whether resolvability requirements (11) are fulfilled, that is to check whether:

$$\begin{aligned} &\mathcal{P}\left(\widehat{\boldsymbol{\Theta}} \in \Sigma_{\mathbf{s}}^1\left(\boldsymbol{\Theta}, \boldsymbol{\Xi}_{\text{req}}^-, \boldsymbol{\Xi}_{\text{req}}^+\right); \boldsymbol{\Theta}^0\right) \leq \mathcal{P}_{\text{res}} \\ &\qquad \leq \mathcal{P}\left(\widehat{\boldsymbol{\Theta}} \in \Sigma_{\mathbf{s}}^P\left(\boldsymbol{\Theta}, \boldsymbol{\Xi}_{\text{req}}^-, \boldsymbol{\Xi}_{\text{req}}^+\right); \boldsymbol{\Theta}^0\right). \quad (12) \end{aligned}$$

Conversely, the use of bounds (12), whenever they are computable, allows to obtain a confidence interval on the probability of resolvability $\mathcal{P}_{\text{res}}$ for a given required precision, confidence interval that can be taken into account in the assessment of performance of subsequent processing. If $\mathcal{P}\left(\widehat{\boldsymbol{\Theta}} \in \Sigma_{\mathbf{s}}^\lambda\left(\boldsymbol{\Theta}, \boldsymbol{\Xi}_{\text{req}}^-, \boldsymbol{\Xi}_{\text{req}}^+\right); \boldsymbol{\Theta}^0\right)$ is computable for a sub-set of values of $\mathbf{s}$, the tightness of the confidence interval is improved by increasing $\min\left\{\{s_p\}_{p=1}^P\right\}$. The possible exploitation of independence among estimates can improve the tightness of the confidence interval on $\mathcal{P}_{\text{res}}$ as well. Indeed if $\{\boldsymbol{\theta}^m\}_{m=1}^M$ are jointly independent, then:

$$\mathcal{O}_{\boldsymbol{\Theta}}\left(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Xi}_{\text{req}}^-, \boldsymbol{\Xi}_{\text{req}}^+\right) = \prod_{m=1}^M \mathcal{O}_{\boldsymbol{\theta}^m}\left(\widehat{\boldsymbol{\theta}^m}, \boldsymbol{\varepsilon}_{\text{req}}^{m-}, \boldsymbol{\varepsilon}_{\text{req}}^{m+}\right), \quad (13)$$

leading to:

$$\begin{aligned} &\prod_{m=1}^M \mathcal{P}\left(\widehat{\boldsymbol{\theta}^m} \in \Sigma_{\mathbf{s}}^1\left(\boldsymbol{\theta}^m, \boldsymbol{\varepsilon}_{\text{req}}^{m-}, \boldsymbol{\varepsilon}_{\text{req}}^{m+}\right); \boldsymbol{\theta}^m\right) \leq \mathcal{P}_{\text{res}} \\ &\qquad \leq \prod_{m=1}^M \mathcal{P}\left(\widehat{\boldsymbol{\theta}^m} \in \Sigma_{\mathbf{s}}^{P'}\left(\boldsymbol{\theta}^m, \boldsymbol{\varepsilon}_{\text{req}}^{m-}, \boldsymbol{\varepsilon}_{\text{req}}^{m+}\right); \boldsymbol{\theta}^m\right) \quad (14) \end{aligned}$$

which are tighter bounds than (12). An even more thorough exploitation of independence has been proposed by Clark [12] when $\left\{\widehat{\boldsymbol{\theta}^m}\right\}_{m=1}^M$ are jointly Gaussian with bias vectors $\mathbf{b}^m$ and covariance matrices $\mathbf{C}^m$: $\widehat{\boldsymbol{\theta}^m} \sim \mathcal{N}\left(\boldsymbol{\theta}^m + \mathbf{b}^m, \sigma^2\mathbf{C}^m\right)$. Then, as noticed by Clark, it is worth substituting ellipsoids for

hyper-rectangles in the definition of the precision in probability sense. Indeed if we consider the following ellipsoids of constant probability:

$$E\left(\boldsymbol{\theta}^m, \sigma r\right) = \left\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - (\boldsymbol{\theta}^m + \mathbf{b}^m)\|^2_{2;\mathbf{C}^m} < \sigma^2 r^2\right\}$$

where $\|\mathbf{u}\|^2_{2;\mathbf{A}} = \left(\|\mathbf{u}\|_{2;\mathbf{A}}\right)^2 \triangleq \mathbf{u}^T\mathbf{A}^{-1}\mathbf{u}$ and $\mathbf{A}$ is a symmetric positive definite matrix, then $\mathcal{P}\left(\widehat{\boldsymbol{\theta}^m} \in E\left(\boldsymbol{\theta}^m, \sigma r\right); \boldsymbol{\theta}^m\right) = \mathcal{P}\left(r\right)$ and:

$$\mathcal{P}\left(\bigcap_{m=1}^{M}\left(\widehat{\boldsymbol{\theta}^m} \in E\left(\boldsymbol{\theta}^m, \sigma r\right)\right); \boldsymbol{\Theta}\right) = \mathcal{P}\left(r\right)^M. \quad (15)$$

When the M ellipsoids with a given probability $\mathcal{P}\left(r\right)$ are disjoint, the parameters are deemed resolvable with probability $\mathcal{P}_{\text{res}} = \mathcal{P}\left(r\right)^M$. Therefore when $\{\boldsymbol{\theta}^m\}_{m=1}^M$ are jointly independent, Clark's approach allows to compute the exact $\mathcal{P}_{\text{res}}$ instead of a confidence interval (14), but at the expense of a more complex method to check that $E\left(\boldsymbol{\theta}^m, \sigma r\right)$ are disjoint (in comparison with hyper-rectangle). Based on the analysis of the distance problem between ellipsoids, Clark provides a method that allows to find the noise level σ^2 at which the ellipsoids become tangent with probability $\mathcal{P}_{\text{res}}$, which is the resolution threshold noise level σ^2_{res}. Unfortunately the assumption of vector of estimates decorrelated (independent) from signal to signal is generally not verified; as a consequence it is generally impossible to predict the accuracy of (15) as an approximation of the exact probability of resolvability.

D. Gaussian p.d.f

The (lower and upper) bounds on $\mathcal{O}_{\boldsymbol{\Theta}}\left(\widehat{\boldsymbol{\Theta}}, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$ given by (8) are easily computable when $\widehat{\boldsymbol{\Theta}}\left(\mathbf{x}\right) - \boldsymbol{\Theta} \sim \mathcal{N}\left(\mathbf{b}\left(\boldsymbol{\Theta}\right), \mathbf{C}\left(\boldsymbol{\Theta}\right)\right)$ and $\mathbf{s} = \mathbf{2}_P$, where $\boldsymbol{\alpha}_P$ is a P-dimensional vector with components equal to $\alpha \in \mathbb{R}$, that is $\widehat{\boldsymbol{\Theta}}$ is a Gaussian estimator of $\boldsymbol{\Theta}$ with bias vector $\mathbf{b}\left(\boldsymbol{\Theta}\right)$ and covariance matrix $\mathbf{C}\left(\boldsymbol{\Theta}\right)$ and we consider the approximation of a hyper-rectangle by a hyper-ellipsoid:

$$\Sigma^{\lambda}_{\mathbf{2}_P}\left(\boldsymbol{\Theta}, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) = \left\{\widehat{\boldsymbol{\Theta}}\left(\mathbf{x}\right) : \left\|\mathbf{D}^{-1}_{\overline{\boldsymbol{\Xi}}}\left(\widehat{\boldsymbol{\Theta}}\left(\mathbf{x}\right) - \boldsymbol{\Theta} - d\boldsymbol{\Xi}\right)\right\|^2_2 < \lambda\right\} \quad (16)$$

where $\|\mathbf{y}\|^2_2 = \left(\|\mathbf{y}\|_2\right)^2 = \mathbf{y}^T\mathbf{y}$ and $\mathbf{D}_{\boldsymbol{\alpha}}$ is the diagonal matrix with the components of vector $\boldsymbol{\alpha}$ on the main diagonal: $\left(\mathbf{D}_{\boldsymbol{\alpha}}\right)_{q,p} = \alpha_q \delta^q_p$. As the following unitary transformation of the Gaussian random vector $\widehat{\boldsymbol{\Theta}}$:

$$\widehat{\mathbf{O}}\left(\mathbf{x}\right) = \mathbf{U}^T_{\overline{\boldsymbol{\Xi}}}(\boldsymbol{\Theta})\, D^{-1}_{\overline{\boldsymbol{\Xi}}}\left(\widehat{\boldsymbol{\Theta}}\left(\mathbf{x}\right) - \boldsymbol{\Theta} - d\boldsymbol{\Xi}\right) \mid \mathbf{D}^{-1}_{\overline{\boldsymbol{\Xi}}}\mathbf{C}\left(\boldsymbol{\Theta}\right)\mathbf{D}^{-1}_{\overline{\boldsymbol{\Xi}}} = \mathbf{U}_{\overline{\boldsymbol{\Xi}}}(\boldsymbol{\Theta})\, D_{\boldsymbol{\sigma}^2_{\overline{\boldsymbol{\Xi}}}(\boldsymbol{\Theta})}\mathbf{U}^T_{\overline{\boldsymbol{\Xi}}}(\boldsymbol{\Theta})$$

preserves the norm $\|\ \|_2$, then:

$$\mathcal{P}\left(\widehat{\boldsymbol{\Theta}} \in \Sigma^{\lambda}_{\mathbf{2}_P}\left(\boldsymbol{\Theta}, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right); \boldsymbol{\Theta}\right) = \mathcal{P}\left(\left\|\widehat{\mathbf{O}}\left(\mathbf{x}\right)\right\|^2_2 < \lambda; \boldsymbol{\Theta}\right) = F_{Q_P}\left(\lambda; \left\|\boldsymbol{\delta}\left(\boldsymbol{\Theta}, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)\right\|^2_2, \boldsymbol{\sigma}^2_{\overline{\boldsymbol{\Xi}}}(\boldsymbol{\Theta})\right) \quad (17)$$

where $\boldsymbol{\delta}\left(\boldsymbol{\Theta}, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) = \mathbf{U}^T_{\overline{\boldsymbol{\Xi}}}\left(\boldsymbol{\Theta}\right)\mathbf{D}^{-1}_{\overline{\boldsymbol{\Xi}}}\left(\mathbf{b}\left(\boldsymbol{\Theta}\right) - d\boldsymbol{\Xi}\right)$ and $F_{Q_P}\left(v; \mu, \boldsymbol{\sigma}^2\right)$ is the c.d.f. of a quadratic form for non-central normal variates [15], that is an extension of non-central chi-square with corresponding degrees of freedom in P and positive noncentrality parameters in μ where the power of each component is not constant: $Q_P \sim \sum_{p=1}^P \sigma^2_p \left|z_p + \sqrt{\mu_p}\right|^2$ where $\boldsymbol{\sigma}^2 = \left(\sigma^2_1, \ldots, \sigma^2_P\right)^T$, $\mu = \sum_{p=1}^P \mu_p$ and $\mathbf{z} = \left(z_1, \ldots, z_P\right)^T \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}\right)$. If $\mu = 0$ (unbiased estimates and symmetric errors $\boldsymbol{\Xi}^- = \boldsymbol{\Xi}^+$) then [16]:

$$\begin{aligned} F_{Q_P}\left(v; 0, \boldsymbol{\sigma}^2\right) &= F_P\left(\frac{v}{2}; \boldsymbol{\sigma}^2\right) \\ F_P\left(t; \mathbf{a}\right) &= \frac{1}{\sqrt{|\mathbf{D}_{\mathbf{a}}|}}\frac{t^{\frac{P}{2}}}{\pi^{\frac{P}{2}}}\sum_{k=0}^{\infty}\frac{w^k_P\left(-t\right)^k}{\Gamma\left(\frac{P}{2} + k + 1\right)} \\ w^k_P &= \sum_{i_1 + \cdots + i_P = k}\frac{\Gamma\left(i_1 + \frac{1}{2}\right)\ldots\Gamma\left(i_n + \frac{1}{2}\right)}{i_1!\ldots i_n!\, a_1^{i_1}\ldots a_P^{i_P}} \end{aligned} \quad (18)$$

where $\Gamma\left(t\right)$ is the Gamma function. The general formula for the c.d.f. of Q_P is given in [15]. Unfortunately series expansion of the c.d.f. of a quadratic form for central (18) or non-central normal variates given in [15], [16] are not easy to program. Therefore for small P it is worth considering the alternative numerical method consisting in computing the p.d.f. of Q_P by iterative convolutions since:

$$p_{Q_P}\left(v; \mu, \boldsymbol{\sigma}^2\right) = p_{\chi^2_1}\left(v; \mu_1, \sigma^2_1\right) * \cdots * p_{\chi^2_1}\left(v; \mu_P, \sigma^2_P\right)$$

where each χ^2_1 p.d.f. is given by [18]:

$$p_{\chi^2_1}\left(v; \mu_p, \sigma^2_p\right) = \frac{e^{-\frac{v + \mu_p}{2\sigma^2_p}}}{2\sigma^2_p}\left(\frac{v\mu_p}{\sigma^4_p}\right)^{-\frac{1}{4}} I_0\left(\sqrt{\frac{v\mu_p}{\sigma^4_p}}\right) \quad (19)$$

where $I_0\left(z\right)$ is the modified Bessel function of the first kind with order zero.

It should be noted that the tightness of the confidence interval (12) on the probability of resolvability $\mathcal{P}_{\text{res}}$ decreases, for a given $\mathbf{s}$, as the product $P'M$ increases. Therefore in some applications it may be necessary to resort to values of $\mathbf{s}$ larger than 2 in order to obtain a tight enough confidence interval, as illustrated in Fig. 1. Unfortunately in that case we have not found in the open literature a simple numerical method to compute the c.d.f. of $\Sigma^{\lambda}_{\mathbf{s}}\left(\boldsymbol{\Theta}, \boldsymbol{\Xi}^-_{\text{req}}, \boldsymbol{\Xi}^+_{\text{req}}\right)$ for correlated $\widehat{\boldsymbol{\Theta}}\left(\mathbf{x}\right)$; nevertheless it is still possible (but computationally expensive) to assess directly $\mathcal{P}_{\text{res}}$ by resorting to algorithms proposed by Genz [17] for numerical evaluation of multivariate normal distributions.

III. Asymptotic Performance of Conditional Model

Historically the first MSE lower bound for deterministic parameters to be derived was the CRB, which was introduced to investigate fundamental limits of a parameter estimation problem or to assess the relative performance of a specific estimator (efficiency) [9]. It has since become the most popular lower bound due to its simplicity of calculation and the fact that in many cases it can be achieved asymptotically (in terms of high SNR [RenauxCML] and/or large number of snapshots [9]) by MLEs. Additionally, it is well known that in non-linear estimation problems three distinct regions of operation can be ob-

served [20]–[22]. In the asymptotic region, the MSE of MLEs is small and, in many cases, close to the CRB. In the *a priori* performance region where the number of independent snapshots and/or the SNR are very low, the observations provide little information and the MSE is close to that obtained from the prior knowledge about the problem. Between these two extremes, there is an additional ambiguity region, also called the transition region. In this region, the MSE of MLEs usually deteriorates rapidly with respect to CRB and exhibits a threshold behaviour corresponding to a "performance breakdown". The nature of this phenomenon is specified by a complicated non-smooth behaviour of the likelihood function in the "threshold" area where it tends to generate outliers.

A. Characterization of the Asymptotic Region of Operation for the Conditional Model

The choice of focusing on the (Gaussian) conditional model comes from the asymptotic (in terms of SNR [19] and/or in large number of snapshots [9]) Gaussianity and efficiency of CMLEs in the multiple parameters case (what is not true for the unconditional model [23]). Moreover the conditional model is the privileged model for an active system. A typical example of an active system is a localization system such as a radar or a sonar where a known waveform is transmitted, and the signals scattered from the targets of interest are used to estimate their parameters. In an active system, as the waveform parametric model is known and deterministic (in opposition with a passive system where a probabilistic modelling of the waveform is generally considered), the most accurate statistical prediction for an observation will be obtained when considering the signal amplitudes as deterministic (since it is well known that the complex Gaussian amplitude modelling provide an average unconditional CRB higher than the corresponding conditional CRB [24]). Let us remind that the asymptotic Gaussianity and efficiency of CMLEs in the multiple parameters case has been proved under the assumption that the maximum of the (reduced) log likelihood function belongs to its main lobe [19], or equivalently that the probability of outlier is equal to zero [20], [21]. An approximation of the probability of outlier for the three regions of operation of CMLE has already been derived in the form of the probability of the union of events that one of the sidelobe peaks is higher than the mainlobe peak, probability itself approximated by the union bound [21], Section IV. The main result is that the derived approximation of probability of outlier is quite accurate in modeling the performance of CMLE. Only in the no-information region does the approximation deviate from the simulation results: this is due to the fact that the union bound tightness decreases as the observation model enters the transition region. Nevertheless, in the multi-source case, the computation is rather cumbersome. Indeed, first, one must look for the positions of the sidelobe peaks which are usually not available in closed form and must be calculated by some numerical method. And second, one must compute multiples c.d.f. of noncentral indefinite quadratic form in complex Gaussian random variables.

A far simpler lower bound on the probability of outlier is proposed hereinafter (25). Its advantage is its computational simplicity but at the expense of the information provided: the proposed lower bound is expected to be tight only in the asymptotic region. We focus on the general complex (circular) linear conditional model where the noise correlation matrix is supposed to be known up to a scale factor $\sigma_{\mathbf{n}}^2$ [24]. For the sake of legibility and without loss of generality, we consider the following simplified instance (narrow band signals and spatially white noise):

$$\mathbf{x}_t\left(\boldsymbol{\Theta}^0\right) = \mathbf{A}\left(\boldsymbol{\Theta}^0\right)\mathbf{s}_t + \mathbf{n}_t, \quad t \in [1, T] \tag{20}$$

where T is the number of independent observations, M is the number of signal sources, $\mathbf{s}_t = (s_{t,1}, \ldots, s_{t,M})^T$ is the vector of complex amplitudes of the M sources for the tth observation, $\mathbf{A}(\boldsymbol{\Theta}) = \left[\mathbf{a}\left(\boldsymbol{\theta}^1\right), \ldots, \mathbf{a}\left(\boldsymbol{\theta}^M\right)\right]$ and $\mathbf{a}(\ )$ is a vector of N parametric functions depending on the parameters vector $\boldsymbol{\theta}$, $\mathbf{n}_t$ are Gaussian complex circular independent noises with spatially white covariance matrix $\mathbf{C}_{\mathbf{n}} = \sigma_{\mathbf{n}}^2\mathbf{I}_N$ with unknown noise power $\sigma_{\mathbf{n}}^2$, independent from the M sources. Then the reduced log likelihood function $L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right)$ is given by [24]:

$$L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) = \sum_{t=1}^{T} \frac{\left\|\boldsymbol{\Pi}_{\mathbf{A}(\boldsymbol{\Theta})}\mathbf{x}_t\left(\boldsymbol{\Theta}^0\right)\right\|^2}{TM} \tag{21}$$

and:

$$\mathcal{P}\left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) \leq \lambda\right) = \mathcal{P}\left(\mathcal{CX}_{MT}^2 \leq \lambda; F\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right), \frac{\sigma_{\mathbf{n}}^2}{TM}\right) \tag{22}$$

where $F\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) = \sum_{t=1}^{T} \frac{\left\|\boldsymbol{\Pi}_{\mathbf{A}(\boldsymbol{\Theta})}\mathbf{A}\left(\boldsymbol{\Theta}^0\right)\mathbf{s}_t\right\|^2}{TM}$ is a generalized correlation function (aka generalized matched filter) and $\mathcal{P}\left(\mathcal{CX}_K^2 \leq \lambda; \delta, \sigma^2\right)$ denotes the c.d.f. of a non-central complex (circular) chi-square with corresponding degrees of freedom in K and positive noncentrality parameter in δ. Let $\widehat{\boldsymbol{\Theta}} \triangleq \widehat{\boldsymbol{\Theta}^0}(\mathbf{x})$ denote the CMLE of $\boldsymbol{\Theta}^0$:

$$\widehat{\boldsymbol{\Theta}} = \underset{\boldsymbol{\Theta}}{\arg\max}\left\{L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right)\right\} \tag{23}$$

then:

$$\boldsymbol{\Theta} \in image\left(\widehat{\boldsymbol{\Theta}}\right) \Rightarrow \exists\left\{\mathbf{x}_t\right\}_{t=1}^{T} \mid L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)$$

or conversely:

$$L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) < L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right), \forall\left\{\mathbf{x}_t\right\}_{t=1}^{T} \Rightarrow \boldsymbol{\Theta} \notin image\left(\widehat{\boldsymbol{\Theta}}\right)$$

leading to:

$$\mathcal{P}\left(\boldsymbol{\Theta} \notin image\left(\widehat{\boldsymbol{\Theta}}\right)\right) \geq \mathcal{P}\left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) < L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)\right). \tag{24}$$

Then, some additional probability calculus detailed in Appendix VI-A allows to prove that, $\forall\alpha \in]0, 1[$:

$$\mathcal{P}\left(\boldsymbol{\Theta} \notin image\left(\widehat{\boldsymbol{\Theta}}\right)\right) \geq 1 - \left(\mathcal{P}\left(\frac{L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)}{F\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)} < \alpha\right) + \mathcal{P}\left(\frac{L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right)}{F\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)} > \alpha\right)\right) \tag{25}$$

or conversely:

$$\mathcal{P}\left(\boldsymbol{\Theta} \in image\left(\widehat{\boldsymbol{\Theta}}\right)\right) \leq \mathcal{P}\left(\frac{L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)}{F\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)} < \alpha\right) + \mathcal{P}\left(\frac{L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right)}{F\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)} > \alpha\right). \quad (26)$$

Let $\Upsilon_{\boldsymbol{\Theta}^0}(\alpha)$ denote the main lobe at α $(0 < \alpha < 1)$:

$$\Upsilon_{\boldsymbol{\Theta}^0}(\alpha) = \left\{\boldsymbol{\Theta} : \frac{F\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right)}{F\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)} > \alpha\right\}. \quad (27)$$

Then they are many ways of exploiting (25)(26) according to:

- the choice of the main lobe $\Upsilon_{\boldsymbol{\Theta}^0}(\alpha)$ including $image\left(\widehat{\boldsymbol{\Theta}}\right)$ requested to enforce the asymptotic Gaussianity and efficiency of CMLEs,
- the acceptable probability of an outlier, in the sense of being outside $\Upsilon_{\boldsymbol{\Theta}^0}(\alpha)$: $\mathcal{P}\left(\boldsymbol{\Theta} \in image\left(\widehat{\boldsymbol{\Theta}}\right) \mid \boldsymbol{\Theta} \notin \Upsilon_{\boldsymbol{\Theta}^0}(\alpha)\right)$.

We opt for a simplified didactic approach where we consider $\Upsilon_{\boldsymbol{\Theta}^0}\left(\frac{1}{2}\right)$ (i.e., the usual main lobe at -3 dB) and $\mathcal{P}\left(\boldsymbol{\Theta} \in image\left(\widehat{\boldsymbol{\Theta}}\right) \mid \boldsymbol{\Theta} \notin \Upsilon_{\boldsymbol{\Theta}^0}\left(\frac{1}{2}\right)\right) \simeq 0$.

The quasi-nullity of the probability of an outlier can be demonstrated by computing the p.d.f. of $L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)$ and $L\left(\boldsymbol{\Theta}_{3\text{ dB}};\boldsymbol{\Theta}^0\right)$ where $F\left(\boldsymbol{\Theta}_{3\text{ dB}};\boldsymbol{\Theta}^0\right) = \frac{1}{2}$ and by checking that their supports do not overlap above a certain p.d.f. threshold value, as small as possible (10^{-30} in the present paper). The underlying hypothesis is to consider that below this threshold value, any p.d.f. can be rounded to zero. As the p.d.f. of $\mathcal{CX}^2_{MT}\left(F\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right), \frac{\sigma^2}{TM}\right)$ is an increasing function in $F\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right)$, it is sufficient to check that the p.d.f. of $L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)$ and $L\left(\boldsymbol{\Theta}_{3\text{ dB}};\boldsymbol{\Theta}^0\right)$ do not overlap to ensure that this property is valid for any $\boldsymbol{\Theta} \notin \Upsilon_{\boldsymbol{\Theta}^0}\left(\frac{1}{2}\right)$, what proves that (under the approximations above mentioned) $image\left(\widehat{\boldsymbol{\Theta}}\right) \subset \Upsilon_{\boldsymbol{\Theta}^0}\left(\frac{1}{2}\right)$. We illustrate the soundness of the proposed approach with the application example given in[19] where two $(M = 2)$ fully correlated sources separated by $\frac{3}{10}$ of the beamwidth impinge on a uniform linear array (ULA) of $N = 10$ sensors with half-wavelength spacing. The high SNR Gaussianity of the CMLE was checked via a Lilliefors goodness-of-fit test [25] and was observed for a SNR at output of the single source matched filter higher than approximately 19 dB (that is a SNR of 9 dB at sensor level [19]). Fig. 2 displaying the p.d.f. of the log likelihood functions $L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)$ (solid lines) and $L\left(\boldsymbol{\Theta}_{3\text{ dB}};\boldsymbol{\Theta}^0\right)$ (dash lines), for $SNR \in \{19, 20\}$ dB, shows that Gaussianity is checked for a SNR of 20 dB according to the proposed approach (and p.d.f. threshold value) which is in accordance with the Lilliefors goodness-of-fit test.

B. Asymptotic Performance of Conditional Model

Then, in the asymptotic region of operation:

$$\widehat{\boldsymbol{\Theta}} \sim \mathcal{N}\left(\boldsymbol{\Theta}^0, \mathbf{CRB}\left(\boldsymbol{\Theta}^0\right)\right) \quad (28)$$

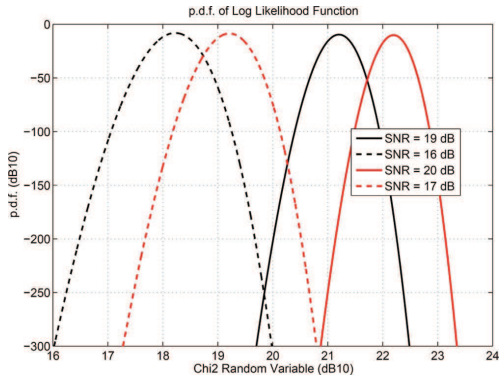

Fig. 2. Loglikelihood p.d.f. (solid lines for $L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)$ and dash lines for $L\left(\boldsymbol{\Theta}_{3\text{ dB}};\boldsymbol{\Theta}^0\right)$) for two correlated signal sources impinging on a ULA of $N = 10$ sensors, separated by $\delta\theta = \frac{3}{10}\theta_{3\text{ dB}}$ and $T = 10$ snapshots.

where, for the general linear observation model (20) [24], [26][3]:

$$\mathbf{CRB}(\boldsymbol{\Theta}) = \frac{\sigma_{\mathbf{n}}^2}{2T}\mathrm{Re}\left\{\mathbf{H}(\boldsymbol{\Theta}) \odot \left(\widehat{\mathbf{R}}_{\mathbf{s}}^T \otimes \mathbf{1}_{P\times P}\right)\right\}^{-1}$$

$$\mathbf{H}(\boldsymbol{\Theta}) = \begin{bmatrix} \mathbf{H}(\boldsymbol{\Theta})_{1,1} & \dots & \mathbf{H}(\boldsymbol{\Theta})_{1,P} \\ \vdots & \ddots & \vdots \\ \mathbf{H}(\boldsymbol{\Theta})_{P,1} & \dots & \mathbf{H}(\boldsymbol{\Theta})_{P,P} \end{bmatrix},$$

$$\mathbf{H}(\boldsymbol{\Theta})_{p,q} = \frac{\partial \mathbf{A}(\boldsymbol{\Theta})}{\partial \theta_p}^H \boldsymbol{\Pi}^{\perp}_{\mathbf{A}(\boldsymbol{\Theta})} \frac{\partial \mathbf{A}(\boldsymbol{\Theta})}{\partial \theta_q}, \quad \widehat{\mathbf{R}}_{\mathbf{s}} = \sum_{t=1}^{T} \frac{\mathbf{s}_t\mathbf{s}_t^H}{T} \quad (29)$$

and $\mathbf{1}_{P\times P}$ is a $P \times P$ matrix of ones. Additionally it has been proved that for each source [26], [30]: the highest (worst) variance is obtained when the sources amplitudes are fully correlated and the lowest (best) variance is obtained when the sources amplitudes are uncorrelated.

IV. Statistical Resolution Limit for Two Closely Spaced Frequencies

A reference application example is the multiple tones estimation problem where in (20):

$$\mathbf{a}(\theta) = \left[1, \dots, e^{j2\pi n\theta}, \dots, e^{j2\pi(N-1)\theta}\right]^T. \quad (30)$$

Although the proposed resolvability criterion (11) is not limited to a given number of signal sources (tones herein) (see [13] for additional examples), we focus in the present paper on the resolvability of two closely spaced frequencies.

A. Background on Statistical Resolution Limit (SRL) Criteria

The resolvability of closely spaced signals, in terms of their parameters of interest, for a given scenario (e.g., for a given number of snapshots, for a given SNR and/or for a given number of sensors) is an old and challenging problem which was recently updated by Smith [4], Shahram and Milanfar [5], Liu and

[3]More general CRB expressions can be found in [27]–[29].

Nehorai [1], Amar and Weiss [2] and El Korso *et al.* [31]. Historically, the statistical resolution limit (SRL) has been empirically defined as the minimum distance between the parameters of interest that allows '*distinguishing*' two closely spaced noisy signals. The concept of SRL does not necessarily take into account how well parameters are estimated, only whether multiple signals can be distinguished. Several criteria were introduced to describe and derive the SRL. We can sum up most of them in four different categories.

1) The first one, and most probably the oldest one, is an algorithm dependent criterion based on the cost function of the considered parametric scheme. It was mainly introduced and applied using the concept of mean null spectrum. Let us considered the problem of distinguishing two signals parameterized by their frequencies, namely f_1 and f_2. In 1974, Cox states that these two signals are resolved, w.r.t. a given parametric algorithm, *if the mean of the null spectrum at the frequencies f_1 and f_2 is lower than the mean of the null spectrum at the frequency midpoint $\frac{f_1+f_2}{2}$* [3]. Two decades after, Sharman and Durrani proposed another commonly used criterion in [6] which states *that two sources are resolved if the second derivative of the mean of the null spectrum at the midpoint $\frac{f_1+f_2}{2}$ is negative*. These intuitive criteria were successfully applied to derive the SRL for a specific algorithm, as MUSIC and Min-Norm scheme, in many different situations (the reader can refer to [3], [6], [32]–[35]).
2) The second approach is based on detection theory. This approach reformulates the problem of distinguishing two signals as a binary hypothesis test. More precisely, the feature of interest of one signal is tested against the feature of interest of the other signal. Let us consider the spectral analysis problem where the main frequency is considered as the feature of interest. In this context, Sharman and Milanfar [5] proposed to use a binary hypothesis test to define/derive the SRL. The hypothesis $\mathcal{H}_1$ represents the case in which two signals are resolvable/distinguishable, i.e., $f_1 - f_2 \neq 0$, whereas, the hypothesis $\mathcal{H}_0$ represents the case where the two signals coalise/combine into one signal, i.e., $f_1 - f_2 = 0$. To solve such a hypothesis test, many strategies can be considered. The commonly used one is the generalized likelihood ratio test (GLRT). This choice is mainly motivated by the known fact that, under certain conditions, the GLRT is asymptotically uniformly most powerful test among all the invariant statistical tests [36], which is one of the strongest statements of optimality that one could expect to obtain [37]. Using this strategy, the main difficulty is to derive the SRL (i.e., the separation $\delta = f_1 - f_2$ which resolves the binary hypothesis test) as a function of the probability of false alarm $\mathcal{P}_{\mathrm{fa}}$ and the probability of detection $\mathcal{P}_{\mathrm{d}}$. This approach was recently applied to several parametric problems (the reader can refer to [1], [5], [31], [38]and references in [31] for more details). Finally note that, in this context, another strategy was applied by Amar and Weiss in [2]. The authors have proposed to determine the SRL of complex sinusoids with nearby frequencies using the Bayesian approach for a given correct decision probability instead of considering $\mathcal{P}_{\mathrm{fa}}$ and $\mathcal{P}_{\mathrm{d}}$ which are relevant generally in a non Bayesian detection strategy.
3) The third approach is based on the estimation accuracy. Unlike the first approach and like the second one, this approach is independent from the estimation algorithm and uses mainly tools as lower bounds (LBs) on the mean square error (MSE). LBs on MSE characterize the ultimate performance of an unbiased estimator in terms of MSE [39], [40]. Due to its simplicity and tightness under certain conditions (see Section III) the CRB is the LB mainly used practically to assess the ultimate estimation accuracy in terms of MSE. In this way, Lee [41], [42], Yau and Bressler [26], and Smith [4] proposed different criteria to describe the SRL. In [26, Section VIII], [41], authors state that *two signals are said to be resolvable w.r.t. the frequencies if the maximum standard deviation of each frequency estimate is less than half the difference between f_1 and f_2*. Under conditions for which the CRB is a tight bound [10], the standard deviations σ_{f_1} and σ_{f_2}, of $\hat{f}_1$ and $\hat{f}_2$ using MLEs, can be well approximated by $\sqrt{\mathrm{CRB}(f_1)}$ and $\sqrt{\mathrm{CRB}(f_2)}$, respectively. Leading to the following separation '*limit*', δ, which represents the SRL: $\delta = 2\max\left\{\sqrt{\mathrm{CRB}(f_1)}, \sqrt{\mathrm{CRB}(f_2)}\right\}$. The second strategy within this estimation accuracy approach was formulated by Smith in [4] (also alluded to in [26, Section VIII-A]) and states that *two signals are resolvable w.r.t. their frequencies if the difference between the frequencies, is greater than the standard deviation of the frequencies difference estimation.* Considering that the mild conditions are met for the CRB, the separation '*limit*' δ representing the SRL in the Smith sense, is then given as the solution of the following equation: $\delta^2 = \mathrm{CRB}(\delta)$ where $\delta = f_1 - f_2$. Which can be written as $\delta^2 = \mathrm{CRB}(f_1) + \mathrm{CRB}(f_2) - 2\mathrm{CRB}(f_1, f_2)$.
4) The fourth approach is based on the definition of a probability of resolvability [11], [12] of parameter vectors by estimators as detailed in Section I.

B. Comparison of SRL Predictions

The aim of this section is to compare the SRL provided by (11) with the existing ones given by Lee-Yaw-Bressler [26], [41], Smith [4], Shahram-Milanfar-Liu-Nehorai [1], [5], [43] and Amar-Weiss [2]. The underlying idea is to measure the cost, in terms of scenario requirements (SNR, number of snapshots, number of sensors), of the proposed SRL definition which takes into account not only the ability to distinguish multiple signals (as all definitions do) but also how well parameters are estimated. To get an idea, let us consider the problem of predicting the SNR required (according to a given criterion) to achieve a given resolution limit δ, $\delta = \theta_2 - \theta_1, \theta_1 < \theta_2$. As a setting, let:

- $N = 32$, $\delta = \frac{1}{N}\frac{1}{k}$ where $k \in \{2, 4, 8, 16, 32\}$,
- the tones amplitude be fully correlated (equal amplitudes but with a phase shift different from zero [43]) $\mathbf{s}_t^T = \sqrt{\frac{SNR}{N}}\left(1, e^{j\frac{\pi}{8}}\right)$ where the SNR is computed at output of the single source matched filter (note that it may not be the worst correlation case [26]).

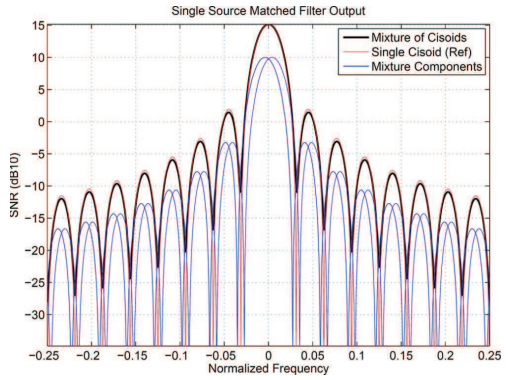

Fig. 3. Single source matched filter output: 2 complex exponentials separated by $\delta = \frac{1}{4N}$, $N = 32$.

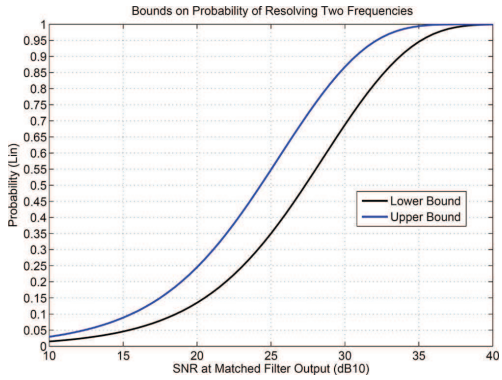

Fig. 4. Bounds (12) on the probability of resolving two complex exponentials separated by $\delta = \frac{1}{4N}$, $N = 32$, $\boldsymbol{\Xi}_{\text{req}}^{T} = \left[\frac{\delta}{2}, \frac{\delta}{2}\right]$, $\mathbf{s}^T = (2, 2)$, $\mathcal{P}_{\text{res}} = 0.95$.

As an instance of the resolvability criterion (11), we consider the combination of symmetric and isotropic required estimation precisions $\boldsymbol{\Xi}_{\text{req}}^{+} = \boldsymbol{\Xi}_{\text{req}}^{-} = \boldsymbol{\Xi}_{\text{req}} = \left[\frac{\delta}{2}, \frac{\delta}{2}\right]^T$ defining disjoint intervals (1-dimensional hyper-rectangles) and $\mathcal{P}_{\text{res}} \in \{0.68, 0.95\}$. For two Gaussian unbiased estimators, the bounds of the confidence interval on the probability of resolvability $\mathcal{P}_{\text{res}}$ (12) can be computed for $\mathbf{s}^T = (2, 2)$ (17) by numerical integration of the p.d.f. of a quadratic form for central normal variates (18) which has then the following closed-form expression (derived in Appendix VI-C for sake of completeness):

$$p_{Q_2}\left(v; 0, \boldsymbol{\sigma}^2\right) = \frac{e^{-\frac{v}{2}\frac{1}{2}\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)}}{2\sigma_1\sigma_2} I_0\left(\frac{v}{2}\frac{1}{2}\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)\right).$$

Then the main features of each high resolution scenario and of the proposed probabilistic approach associated to a given value of δ can be described with the help of 3 figures (exemplified for $k = 4$):

- Fig. 3: the output of the single source matched filter $\left(\frac{1}{N}\left\|\mathbf{a}\left(\boldsymbol{\theta}\right)^H \mathbf{x}_t\left(\boldsymbol{\theta}^0\right)\right\|^2\right)$ which could be the first step in a practical implementation of the CMLE (Clean algorithm, Alternating Projection algorithm) [24].
- Fig. 4: the probability (lower and upper) bounds (PLB and PUB) defined by (12) for $\mathbf{s}^T = (2, 2)$ as a function of the SNR. These bounds allow to determine the SNR interval containing the SNR_{res} for which $\mathcal{P}_{\text{res}}$ is obtained since: $SNR\left(PUB = \mathcal{P}_{\text{res}}\right) \leq SNR_{\text{res}} \leq SNR\left(PLB = \mathcal{P}_{\text{res}}\right)$.
- Fig. 5: the p.d.f. of $L\left(\boldsymbol{\Theta}^0; \boldsymbol{\Theta}^0\right)$ and $L\left(\boldsymbol{\Theta}_{3\text{ dB}}; \boldsymbol{\Theta}^0\right)$ (22) for $SNR\left(PUB = \mathcal{P}_{\text{res}}\right)$ and $SNR\left(PLB = \mathcal{P}_{\text{res}}\right)$ to prove that within $\left[SNR\left(PUB = \mathcal{P}_{\text{res}}\right), SNR\left(PLB = \mathcal{P}_{\text{res}}\right)\right]$ the condition of asymptotic region of operation for CMLE is valid, that is $\mathcal{P}\left(\boldsymbol{\Theta} \in image\left(\widehat{\boldsymbol{\Theta}}\right) \mid \boldsymbol{\Theta} \notin \Upsilon_{\boldsymbol{\Theta}^0}\left(\frac{1}{2}\right)\right) \simeq 0$.

The comparison of the SNR required to achieve a given resolution limit δ is displayed on figure (6), where LYB-SNR is

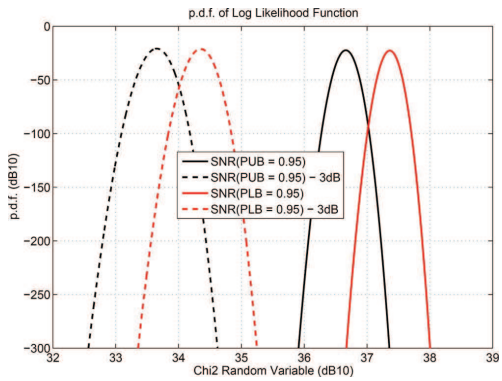

Fig. 5. Loglikelihood p.d.f. at limits of SNR interval allowing the resolvability of two complex exponentials separated by $\delta = \frac{1}{4N}$, $N = 32$, $\mathcal{P}_{\text{res}} = 0.95$.

the one obtained with the Lee-Yaw-Bressler criterion (LYB-criterion) [26], [41], S-SNR is the one obtained with the Smith criterion (S-criterion) [4], SMLN-SNR is the one obtained with the GLRT approach (SMLN-criterion) where $\mathcal{P}_{\text{d}} = 0.95$ and $\mathcal{P}_{\text{fa}} = 0.1$ [1], [5], [43], and AW-SNR is the one obtained with the Bayesian hypothesis test approach (AW-criterion) [2] where $\mathcal{P}d = 0.95$. This figure raises the following comments:

- there is no bound on SNR_{res} for the proposed approach when $k = 2$. The reason is simple: the condition of asymptotic region of operation for CMLEs is not valid within $\left[SNR\left(PUB = \mathcal{P}_{\text{res}}\right), SNR\left(PLB = \mathcal{P}_{\text{res}}\right)\right]$, $\mathcal{P}_{\text{res}} \in \{0.68, 0.95\}$, when $k = 2$. Therefore the probability (lower and upper) bounds defined by (12) (17) are not trustworthy and should not be taken into account.
- in the setting considered, the LYB-criterion takes into account the same estimation errors $\boldsymbol{\Xi}_{\text{req}} = \frac{\delta}{2}\mathbf{1}_M$, except that the LYB-criterion expresses requirements on marginal probability of errors, that is $\mathcal{P}\left(\left|\widehat{\theta}_i - \theta_i\right| \leq \frac{\delta}{2}\right) = 0.68$, $i \in \{1, 2\}$, where $\widehat{\theta}_i$ is the MLE of θ_i in the asymptotic region of operation. Since

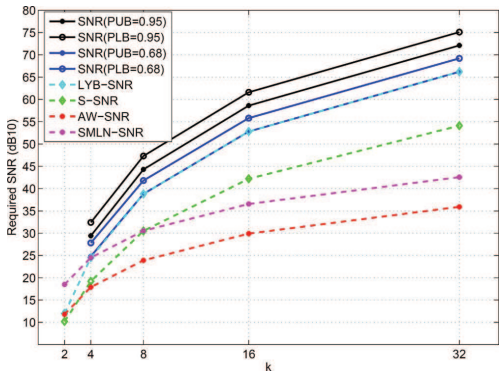

Fig. 6. Required SNR to achieve a given SRL for two complex exponentials separated by $\delta = \frac{1}{kN}$, $N = 32$, $k \in \{2, 4, 8, 16, 32\}$.

$\mathcal{P}(A \cap B) \leq \min\{\mathcal{P}(A), \mathcal{P}(B)\}$, these requirements are less demanding than the requirement on the joint probability of error with same occurrence probability as formulated by (12) (17) when $\mathcal{P}_{\text{res}} = 0.68$; hence the smaller (or equal) required LYB-SNR to achieve a given resolution limit δ. The difference between SNR_{res} and LYB-SNR increases as $\mathcal{P}_{\text{res}}$ increases above 0.68, as shown when $\mathcal{P}_{\text{res}} = 0.95$.

- as already noticed in [26, Section VIII,] the LYB-criterion implies the S-criterion which finally states that the SNR required (S-SNR) is the one for which $\mathcal{P}\left(\left|\left(\widehat{\theta}_2 - \widehat{\theta}_1\right) - \delta\right| \leq \delta\right) = 0.68$, or equivalently the one for which $\mathcal{P}\left(\widehat{\theta}_2 \leq \widehat{\theta}_1\right) = 0.16$ and $\mathcal{P}\left(\widehat{\theta}_2 \geq \left(\widehat{\theta}_1 + \delta\right) + \delta\right) = 0.16$. The second form of the S-criterion clearly highlights that this criterion is mainly focused on parameters distinguishability as it only requires a loose estimation precision for each parameter. Hence the lower S-SNR values in comparison with LYB-SNR and $[SNR(PUB = 0.68), SNR(PLB = 0.68)]$.
- in order to obtain tractable solutions, the AW-SNR and the SMLN-SNR reported here are derived from [2] and [43] in the case of a known central frequency and a known signals power or amplitude. After some proper modifications, one obtains:

$$\text{AW-SNR} = \frac{0.675\pi N}{\delta^2(N-1)(N+1)} \tag{31}$$

$$\text{SMLN-SNR} = \lambda \frac{2}{\delta^2 \varrho^2} \tag{32}$$

where $\varrho = \|\boldsymbol{s}_{t,2} - \boldsymbol{s}_{t,1}\| \, \|\dot{\boldsymbol{a}}\|$, $\dot{\boldsymbol{a}}$ denotes the derivative of $\boldsymbol{a}$ w.r.t. θ and λ is numerically computed as the solution of $\mathcal{Q}^{-1}_{\chi^2_1}(\mathcal{P}_{\text{fa}}; 0, 1) = \mathcal{Q}^{-1}_{\chi^2_1}(\mathcal{P}_{\text{d}}; 0, \lambda)$ for a given $\mathcal{P}_{\text{fa}}$ and $\mathcal{P}_{\text{d}}$, where $\mathcal{Q}_{\chi^2_1}(v; 0, \sigma^2)$ denotes the right tail of the distribution of a central χ^2_1 with mean σ^2.
- Let us remind that both the SMLN-criterion and the AW-criterion exploit the historical definition of resolvability, that is the problem of distinguishing whether the observation consists of a single signal or of two signals, irrespective of any requirement on the precision of parameters estimation. Thus on a scale representing the ability of a criterion to incorporate a requirement on the precision of parameters estimation, we may say that the SMLN-criterion and the AW-criterion are based on an approach of resolvability at the opposite of our proposed approach (and the one of LYB), and that the S-criterion defines an intermediate approach.

The significant differences in the SNR requested to fulfill the different SRL criteria are representative of the position of each approach on this scale: indeed each approach allows to discern the presence of two signals but the estimates of the signal parameters may not be sufficiently accurate according to a prescribed precision (hyper-rectangles). Hence the question on which information one looks for: distinguishability only or distinguishability and accuracy, since in order to get both there is a price to pay (possibly overstressed in our example[4]).

V. Conclusion

We have proposed a resolvability definition which takes into account not only the ability to distinguish multiple signals (as all definitions do) but also how well parameters are estimated. In order to widen the practical use of this definition, we have also introduced a general scheme to generate the lower and upper bounds of confidence intervals on the probability of resolvability for a given required estimation precision. The application of this approach to the conditional model has shown that significant differences in the region of operation (SNR, number of snapshots, number of sensors) requested to fulfill the different SRL criteria are expected according to which information is looked for: distinguishability only or distinguishability and accuracy. Possible application of the proposed approach to the unconditional model will be pursued in future work, since asymptotic efficiency of MLEs does no longer hold for small number of observations [23].

Appendix

Approximation of a Hyper-Rectangle:

As:

$$\left|y_q - y_q^0 - \frac{d\varepsilon_q}{2}\right| < \overline{\varepsilon_q} \Leftrightarrow \left|\frac{y_q - y_q^0 - \frac{d\varepsilon_q}{2}}{\overline{\varepsilon_q}}\right| < 1$$

then, $\forall s_q \geq 1$:

$$\left|y_q - y_q^0 - \frac{d\varepsilon_q}{2}\right| < \overline{\varepsilon_q} \Rightarrow \left|\frac{y_q - y_q^0 - \frac{d\varepsilon_q}{2}}{\overline{\varepsilon_q}}\right|^{s_q} < 1$$

and:

$$\bigwedge_{q=1}^{Q}\left(\left|y_q - y_q^0 - \frac{d\varepsilon_q}{2}\right| < \overline{\varepsilon_q}\right) \Rightarrow \sum_{q=1}^{Q}\left|\frac{y_q - y_q^0 - \frac{d\varepsilon_q}{2}}{\overline{\varepsilon_q}}\right|^{s_q} < Q$$

[4]The hypotheses of a known central frequency and a known signals power or amplitude in ((31), (32)) introduce an *a priori* information that has not been taken into account in the conditional CRB considered (29), therefore higher than the one incorporating this *a priori* information.

that is $R\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) \subset \Sigma_{\mathbf{s}}^Q\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$. Conversely:

$$\sum_{q=1}^{Q}\left|\frac{y_q - y_q^0 - \frac{d\varepsilon_q}{2}}{\overline{\varepsilon_q}}\right|^{s_q} < 1 \Rightarrow \bigwedge_{q=1}^{Q}\left(\left|\frac{y_q - y_q^0 - \frac{d\varepsilon_q}{2}}{\overline{\varepsilon_q}}\right|^{s_q} < 1\right).$$

As, $\forall s_q \geq 1$:

$$\left|\frac{y_q - y_q^0 - \frac{d\varepsilon_q}{2}}{\overline{\varepsilon_q}}\right|^{s_q} < 1 \Leftrightarrow \left|\frac{y_q - y_q^0 - \frac{d\varepsilon_q}{2}}{\overline{\varepsilon_q}}\right| < 1$$

therefore:

$$\sum_{q=1}^{Q}\left|\frac{y_q - y_q^0 - \frac{d\varepsilon_q}{2}}{\overline{\varepsilon_q}}\right|^{s_q} < 1 \Rightarrow \bigwedge_{q=1}^{Q}\left(\left|y_q - y_q^0 - \frac{d\varepsilon_q}{2}\right| < \overline{\varepsilon_q}\right)$$

that is $\Sigma_{\mathbf{s}}^1\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right) \subset R\left(\mathbf{y}^0, \boldsymbol{\Xi}^-, \boldsymbol{\Xi}^+\right)$. QED.

Lower Bound on $\mathcal{P}\left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) < L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)\right)$:

$$\begin{aligned}
&\forall\beta > 0: \\
&\mathcal{P}\left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)\right) \\
&= \mathcal{P}\left(\begin{array}{l}\left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > L\left(L\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right)\right) \wedge \\ \left(\left(\beta \geq L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right)\right) \vee \left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > \beta\right)\right)\end{array}\right) \\
&= \mathcal{P}\left(\left(\beta \geq L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right)\right) \wedge \left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)\right)\right) + \\
&\quad \mathcal{P}\left(\left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)\right) \wedge \left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > \beta\right)\right) \\
&= \mathcal{P}\left(\beta \geq L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)\right) + \\
&\quad \mathcal{P}\left(\left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)\right) \wedge \left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > \beta\right)\right).
\end{aligned}$$

Therefore:

$$\begin{aligned}
&\mathcal{P}\left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)\right) \leq \\
&\qquad \mathcal{P}\left(L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right) < \beta\right) + \mathcal{P}\left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > \beta\right)
\end{aligned}$$

what yields (25):

$$\begin{aligned}
&\mathcal{P}\left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) < L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right)\right) \geq \\
&\qquad 1 - \left(\mathcal{P}\left(L\left(\boldsymbol{\Theta}^0;\boldsymbol{\Theta}^0\right) < \beta\right) + \mathcal{P}\left(L\left(\boldsymbol{\Theta};\boldsymbol{\Theta}^0\right) > \beta\right)\right).
\end{aligned}$$

Derivation of $p_{Q_2}\left(v; 0, \left(\sigma_1^2, \sigma_2^2\right)^T\right)$:

Since $Q_2 = \sigma_1^2 z_1^2 + \sigma_2^2 z_2^2$ where $\mathbf{z} = (z_1, z_2)^T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, then:

$$\begin{aligned}
p_{Q_2}\left(v; 0, \boldsymbol{\sigma}^2\right) &= \int_0^v p_{\chi_1^2}(v - x; 0, \sigma_1^2) p_{\chi_1^2}(x; 0, \sigma_2^2) dx \\
&= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{v}{2\sigma_1^2}} \int_0^v \frac{1}{\sqrt{v-x}\sqrt{x}} e^{-ax} dx
\end{aligned}$$

where $a = \frac{1}{2}\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)$. After a first change of variable $x = \frac{v}{2} - u$ yielding:

$$\int_0^v \frac{e^{-ax}}{\sqrt{(v-x)\,x}} dx = \frac{2}{v} e^{-\frac{a}{2}v} \int_{-\frac{v}{2}}^{\frac{v}{2}} \frac{e^{au}}{\sqrt{1 - \left(2\frac{u}{v}\right)^2}} du,$$

a second change of variable $\sin\theta = 2\frac{u}{v}$ allows to show that:

$$\int_0^v \frac{1}{\sqrt{(v-x)\,x}} e^{-ax} dx = e^{-\frac{a}{2}v} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{\frac{av}{2}\sin\theta} d\theta.$$

Therefore:

$$p_{Q_2}\left(v; 0, \boldsymbol{\sigma}^2\right) = \frac{e^{-\frac{v}{2\sigma_1^2}} e^{-\frac{va}{2}}}{2\sigma_1\sigma_2}\left(\frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{\frac{av}{2}\sin\theta} d\theta\right).$$

Additionally as:

$$\frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{z\sin\theta} d\theta = \frac{1}{\pi}\int_0^{\pi} e^{-z\cos(u)} du = I_0(-z) = I_0(z)$$

one finally has:

$$p_{Q_2}\left(v; 0, \boldsymbol{\sigma}^2\right) = \frac{e^{-\frac{v}{2}\frac{1}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)}}{2\sigma_1\sigma_2} I_0\left(\frac{v}{2}\frac{1}{2}\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)\right).$$

REFERENCES

[1] Z. Liu and A. Nehorai, "Statistical angular resolution limit for point sources," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5521–5527, Nov. 2007.

[2] A. Amar and A. Weiss, "Fundamental limitations on the resolution of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5309–5318, Nov. 2008.

[3] H. Cox, "Resolving power and sensitivity to mismatch of optimum array processors," *J. Acoust. Soc.*, vol. 3, no. 54, pp. 771–785, 1973.

[4] S. T. Smith, "Statistical resolution limits and the complexified Cramér Rao bound," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1597–1609, May 2005.

[5] M. Shahram and P. Milanfar, "On the resolvability of sinusoids with nearby frequencies in the presence of noise," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2579–2585, Jul. 2005.

[6] K. Sharman and T. Durrani, "Resolving power of signal subspace methods for fnite data lengths," presented at the IEEE ICASSP, 1995.

[7] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[8] L. L. Scharf, *Detection, Estimation and Time Series Analysis*. Boston, MA, USA: Adison Wesley, 1991.

[9] H. L. VanTrees, *Detection, Estimation and Modulation Theory*. Hoboken, NJ: Wiley, 1968, vol. 1.

[10] E. L. Lehmann, *Theory of point estimation*. : Wiley, 1983.

[11] S. Oh and R. Kashyap, "A robust approach for high resolution frequency estimation," *IEEE Trans. Signal Process.*, vol. 39, no. 3, pp. 627–643, Mar. 1991.

[12] M. P. Clark, "On the resolvability of normally distributed vector parameter estimates," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2975–2981, Dec. 1995.

[13] C. Ren, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux, "High resolution techniques for radar : myth or reality? ," presented at the Eurasip EUSIPCO, 2013.

[14] J. Rissanen, *Information and complexity in statistical modeling*. Berlin, Germany: Springer-Verlag, 2007.

[15] B. K. Shah and C. G. Khatri, "Distribution of a definite quadratic form for non-central normal variates," *Ann. Math. Statist.*, vol. 32, no. 3, pp. 883–887, 1961.

[16] J. Pachares, "Note on the distribution of a definite quadratic form," *Ann. Math. Statist.*, vol. 26, no. 1, pp. 128–131, 1955.

[17] A. Genz, "Numerical computation of multivariate normal probabilities," *J. Comput. Graph. Statist.*, pp. 1:141–1:149, 1992.

[18] R. Muirhead, *Aspects of multivariate statistical theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2005.

[19] A. Renaux, P. Forster, E. Chaumette, and P. Larzabal, "On the high SNR CML estimator full statistical characterization," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4840–4843, Dec. 2006.

[20] D. C. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 5, pp. 591–598, May 1974.

[21] Athley, "Threshold region performance of maximum likelihood direction of arrival estimators," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1359–1373, Apr. 2005.

[22] C. D. Richmond, "Capon algorithm mean squared error threshold SNR prediction and probability of resolution," *IEEE Trans. Singal Process.*, vol. 53, no. 8, pp. 2748–2764, Aug. 2005.

[23] A. Renaux, P. Forster, E. Boyer, and P. Larzabal, "Unconditional maximum likelihood performance at finite number of samples and high signal-to-Noise ratio," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2358–2364, May 2007.

[24] H. L. Van Trees, *Optimum Array Processing*. New York, NY, USA: Wiley-Interscience, 2002.

[25] H. Lilliefors, "On the kolmogorov-smimov test for normality with mean and variance unknown," *J. Amer. Statist. Assoc.*, vol. 62, no. 318, pp. 399–402, 1967.

[26] S. F. Yau and Y. Bresler, "Worst case Cramér-Rao bounds for parametric estimation of superimposed signals," *IEEE Trans. Signal Process.*, vol. 40, no. 12, pp. 2973–2986, Dec. 1992.

[27] M. C. Wicks, E. L. Mokole, S. D. Blunt, R. S. Schneible, and V. J. Amuso, *Principles of Waveform Diversity and Design*. Hertfordshire, U.K.: SciTech Publishing, 2010.

[28] T. Menni, E. Chaumette, P. Larzabal, and J. P. Barbot, "CRB for Active Radar," presented at the Eurasip EUSIPCO, 2011.

[29] T. Menni, J. Galy, E. Chaumette, and P. Larzabal, "On the versatility of constrained Cramér-Rao bound for estimation performance analysis and design of a system of measurement," *IEEE Trans. Aerospace Electron. Syst.*, Sep. 2013, accepted for publication.

[30] T. Menni, E. Chaumette, P. Larzabal, and J. P. Barbot, "New results on deterministic Cramér-Rao bounds for real and complex parameters," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1032–1049, Mar. 2012.

[31] M. N. El Korso, R. Boyer, A. Renaux, and S. Marcos, "Statistical resolution limit for the multidimensional harmonic retrieval model: hypothesis test and Cramér-Rao bound approaches," *J. Appl. Soc. Psychol.*, Jun. 2011.

[32] M. Kaveh and A. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 2, pp. 331–341, Apr. 1986.

[33] S. Marcos, *Les Méthodes á Haute Résolution: Traitement d'Antenne et Analyse spectrale*. Paris, France: Hermes, 1998.

[34] H. Abeida and J.-P. Delmas, "Statistical performance of MUSIC-like algorithms in resolving noncircular sources," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 4317–4329, Jun. 2008.

[35] A. Ferreol, P. Larzabal, and M. Viberg, "Statistical analysis of the MUSIC algorithm in the presence of modeling errors, taking into account the resolution probability," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 58, no. 8, pp. 4156–4166, Aug. 2010.

[36] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. Berlin, Germany: Springer-Verlag, 2005.

[37] S. M. Kay, *Fundamentals of Statistical Signal Processing : Detection Theory*. Englewood Cliffs, NJ, USA: Prentice Hall, 1998.

[38] M. Shahram and P. Milanfar, "Imaging below the diffraction limit: A statistical analysis," *IEEE Trans. Image Process.*, vol. 13, no. 5, pp. 677–689, May 2004 .

[39] E. Chaumette, J. Galy, A. Quinlan, and P. Larzabal, "A new barankin bound approximation for the prediction of the threshold region performance of maximum-likelihood estimators," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5319–5333, Nov. 2008.

[40] K. Todros and J. Tabrikian, "General classes of performance lower bounds for parameter estimation-Part I: Non-Bayesian bounds for unbiased estimators," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5064–5082, Oct. 2010.

[41] H. B. Lee, "The Cramér-Rao bound on frequency estimates of signals closely spaced in frequency," *IEEE Trans. Signal Process.*, vol. 40, no. 6, pp. 1507–1517, Jun. 1992.

[42] H. B. Lee, "The Cramér-Rao bound on frequency estimates of signals closely spaced in frequency (unconditional case)," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1569–1572, Jun. 1994.

[43] M. N. El Korso, R. Boyer, A. Renaux, and S. Marcos, "On the Asymptotic Resolvability Of Two Point Source in Known Sub-space Interference Using a GLRT-Based Framework," *Elsevier Signal Process.*, vol. 92, no. 10, pp. 2471–2483, 2012.

Chengfang Ren was born in Zhejiang, province of China, in 1987. He came in France since 1997 and obtained the French nationality in 2010. He received the Agrégation in applied Physics, the M.Sc. degrees in electrical engineering from École Normale Supérieure de Cachan, France, in 2011, and 2012, respectively. He is currently working toward the Ph.D. degree at the Laboratory of Signals and Systems (L2S), University of Paris-Sud, France. His research interests include performance bounds and resolution limits for parameter estimation applied to statistical signal processing.

Mohammed Nabil El Korso was born in Oran, Algeria. He received the M.Sc. in Electrical Engineering from the National Polytechnic School, Algeria in 2007. He obtained the Master Research degree in Signal and Image Processing from Paris-Sud XI University/Supélec, France in 2008. In, 2011, he obtained his Ph.D. degree from Paris-Sud XI University. Between 2011 and 2012, he was a research scientist in the Communication Systems Group at Technische Universität Darmstadt, Germany. He was a temporary assistant professor at l'école normale supérieure de Cachan, between 2012 and 2013. Currently, he is assistant professor at Université Paris Ouest Nanterre la Défense and a member of LEME (EA4416) laboratory. His research interests include statistical signal processing, estimation/detection theory with applications to array signal processing.

Jérôme Galy was born in 1970 at Toulouse (France). He studied Electronics and Signal Processing at ENAC high school (Toulouse, France), where he obtained a diploma of engineer in 1993. He received the Ph.D. in Signal Processing from ENSICA high school (Toulouse, France) in 1998. He has been since an Assistant Professor in electronics at University of Montpellier II (France) and a researcher at the LIRMM (Laboratoire d'Informatique, de Robotique et de Micro-Electronique de Montpellier). Main domains of interest are related to signal processing and its applications such as radar or microelectronics (Functional Electrical Stimulation).

Eric Chaumette was born in 1965 at Chartres (France). He studied Electronics and Signal Processing both at ENAC (Toulouse, France), where he obtained a diploma of engineer in 1989, and at Toulouse University where he obtained a M.Sc. degrees in Signal Processing in 1989. From 1990 to 2007, he was with Thales in the radar studies departments and received a Ph.D. degree in 2004 at laboratory SATIE, CNRS, Ecole Normale Supérieure de Cachan, France. From 2007 to 2013, he was with the Electromagnetic and Radar Division of the French Aerospace Lab (ONERA), Palaiseau, France, as a research engineer. He is now with the Department of Electronics, Optronics and Signal of ISAE (Institut Supérieur de l'Aéronautique et de l'Espace). Main domains of interest are related to radar scene modelling, detection and estimation theory and navigation.

Pascal Larzabal (M'93) was born in the Basque country in the south of France in 1962. He entered the Ecole Normale Supérieure of Cachan (France) in 1985 where he received the Agrégation in Electrical Engineering in 1988. He received the Ph.D. in 1992 and the "habilitationà diriger les recherches" in 1998. He is now Professor of Electrical Engineering at University of ParisSud 11, France. He teaches electronic, signal processing, control and mathematics. From 1998 to 2003 he was the director of Electrical Engineering IUP of the University of ParisSud 11. From March 2003 to March 2007 he was at the head of electrical engineering department in the Institut Universitaire de Technologie of Cachan. Since January 2007 he is the director of the laboratory SATIE (UMR CNRS 8029, Ecole Normale Supérieure de Cachan). Since 1993 he is at the head of the Signal Processing Group of the laboratory SATIE. His research concerns estimation in array processing and spectral analysis for wavefront identification, radars, communications, tomography and medical imaging. His recent works concern estimator performances, associated minimal bounds, modelling error and geographical positioning.

Alexandre Renaux received the Ph.D. degrees in electrical engineering from École Normale Supérieure de Cachan, Cachan, France, in 2006. From 2006 to 2007, he was a Postdoctoral Research Associate at the Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO, USA. Since 2007, he has been an Assistant Professor (and currently Associate Professor since 2011) at the Physics Department of University Paris-Sud, Orsay, France. His research interests are in the field of estimation theory and lower bounds on the mean square error applied to statistical signal processing.

Table des figures

1.1 EQM du maximum de vraisemblance pour l'estimation de la fréquence d'une sinusoïde 4

2.1 Comparaison des bornes hybrides et de l'EQM du MVMAP en fonction du RSB. 28
2.2 Comparaison de la borne de Barankin Ziv-Zakaï hybride aux bornes de la famille de Weiss-Weinstein hybride et à l'EQM du MVMAP 34

3.1 Comparaison des EQMs empiriques totales du MVMAP et du MVMAPC aux bornes BCRH et BCRHC en fonction du RSB 41
3.2 Comparaison des EQMs et des bornes pour l'estimation de la fréquence Doppler en fonction du RSB 42
3.3 Comparaison entre l'EQM du MV et les bornes pour l'estimation de la direction de départ en fonction du RSB 52
3.4 Comparaison entre l'EQM du MV et les bornes pour l'estimation de la direction d'arrivée en fonction du RSB 52

4.1 Cas de sources résolues selon le critère de Cox. 54
4.2 Probabilité de résolution pour les paramètres scalaires 55
4.3 Encadrement du rectangle $R\left(\boldsymbol{\theta}^0, \boldsymbol{\varepsilon}^-, \boldsymbol{\varepsilon}^+\right)$ par des ellipses $E_s^1\left(\boldsymbol{\theta}^0, \boldsymbol{\varepsilon}^-, \boldsymbol{\varepsilon}^+\right)$ en trait plein et $E_s^2\left(\boldsymbol{\theta}^0, \boldsymbol{\varepsilon}^-, \boldsymbol{\varepsilon}^+\right)$ en pointillé pour $s \in \{1, 2, 4, 16\}$ et $\boldsymbol{\varepsilon}^- = \boldsymbol{\varepsilon}^+ = [1\ 4]^T$ . . 58
4.4 Encadrement de la probabilité de résolution en fonction du RSB à l'aide de la relation (4.16) avec $s = 2$. 62
4.5 Le RSB minimal requis pour la résolution de deux sources proches séparées de $\delta = \frac{1}{kN}$ avec $N = 32$ et $k \in \{2, 4, 8, 16, 32\}$ 63

Bibliographie

[AB15] Y. I. Abramovich and O. Besson. Fluctuating target detection in fluctuating K-distributed clutter. *IEEE Signal Processing Letters*, 22(10) :1791–1795, October 2015.

[Abe93] J. S. Abel. A bound on mean square estimate error. *IEEE Transactions on Information Theory*, 39(5) :1675–1680, September 1993.

[AN10] M. Akcakaya and A. Nehorai. MIMO Radar Detection and Adaptive Design Under a Phase Synchronization Mismatch. *IEEE Transactions on Signal Processing*, 58(10) :4994–5005, October 2010.

[AW08] A. Amar and A. J. Weiss. Fundamental limitations on the resolution of deterministic signals. *IEEE Transactions on Signal Processing*, 56(11) :5309–5318, November 2008.

[BA13] O. Besson and Y. I. Abramovich. On the Fisher information matrix for multivariate elliptically contoured distributions. *IEEE Signal Processing Letters*, 20(11) :1130–1133, November 2013.

[Ban71] W. J. Bangs. *Array processing with generalized beamformers.* PhD thesis, Yale University, New Haven, CT, 1971.

[Bar49] E. W. Barankin. Locally best unbiased estimates. *The Annals of Mathematical Statistics*, 20(4) :477–501, December 1949.

[Bel95] K. L. Bell. *Performance bounds in parameter estimation with application to bearing estimation.* PhD thesis, George Mason University, Fairfax, VA, USA, 1995.

[BESV] K. L. Bell, Y. Ephraim, Y. Steinberg, and H. L. Van Trees. Improved bellini-tartara lower bound for parameter estimation. In *Proc. of International Symposium on Information Theory, year = 1994, editor =*.

[BEV95] K. L. Bell, Y. Ephraim, and H. L. Van Trees. Ziv-Zakaï lower bounds in bearing estimation. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 1649–1652, Detroit, MI, USA, May 1995.

[BEV96] K. L. Bell, Y. Ephraim, and H. L. Van Trees. Explicit Ziv-Zakaï lower bound for bearing estimation. *IEEE Transactions on Signal Processing*, 44(11) :2810–2824, November 1996.

[BGR+08] S. Bay, B. Geller, A. Renaux, J.-P. Barbot, and J.-M. Brossier. On the hybrid Cramér-Rao bound and its application to dynamical phase estimation. *IEEE Signal Processing Letters*, 15 :453–456, 2008.

[BLS05] S. Buzzi, M. Lops, and S. Sardellitti. Further result on Cramér-Rao bounds for parameter estimation in long-code DS/CDMA systems. *IEEE Transactions on Signal Processing*, 53(3) :1216–1221, March 2005.

[BMWZ87] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakaï. Some classes of global Cramér-Rao bounds. *The Annals of Statistics*, 15(4) :1421–1438, December 1987.

[BSEV97] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees. Extended Ziv-Zakaï lower bound for vector parameter estimation. *IEEE Transactions on Information Theory*, 43(2) :624–637, March 1997.

[BT74] S. Bellini and G. Tartara. Bounds on error in signal parameter estimation. *IEEE Transactions on Communications*, 22(3) :340–342, March 1974.

[BZ75] B. Z. Bobrovsky and M. Zakaï. A lower bound on the estimation error for Markov processes. *IEEE Transactions on Automatic Control*, 20(6) :785–788, December 1975.

[BZ76] B. Z. Bobrovsky and M. Zakaï. A lower bound on the estimation error for certain diffusion processes. *IEEE Transactions on Information Theory*, 22(1) :45–52, January 1976.

[CGQL08] E. Chaumette, J. Galy, A. Quinlan, and P. Larzabal. A new Barankin bound approximation for the prediction of the threshold region performance of maximum likelihood estimators. *IEEE Transactions on Signal Processing*, 56(11) :5319–5333, November 2008.

[Cha04] E. Chaumette. *Contribution à la caractérisation des performances des problèmes conjoints de détection et d'estimation.* PhD thesis, Ecole Normale Supérieure de Cachan, Cachan, FR, December 2004.

[Cla95] M. P. Clark. On the resolvability of normally distributed vector parameter estimates. *IEEE Transactions on Signal Processing*, 43(12) :2975–2981, December 1995.

[Cox73] H. Cox. Resolving power and sensitivity to mismatch of optimum array processors. *The Journal of the acoustical society of America*, 54(3) :771–785, August 1973.

[CR51] D. G. Chapman and H. Robbins. Minimum variance estimation without regularity assumptions. *The Annals of Mathematical Statistics*, 22(4) :581–586, December 1951.

[Cra46] H. Cramér. *Mathematical Methods of Statistics*, volume 9 of *Princeton Mathematics*. Princeton University Press, New-York, September 1946.

[CRL09] E. Chaumette, A. Renaux, and P. Larzabal. Lower bounds on the mean square error derived from mixture of linear and non-linear transformations of the unbiasedness definition. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3045–3048, Taipei, Taiwan, April 2009.

[CRL10] E. Chaumette, A. Renaux, and P. Larzabal. New trends in deterministic lower bounds and SNR threshold estimation : From derivable bounds to conjectural bounds. In *Proc. of IEEE Workshop on Sensor Array and Multi-channel Processing (SAM)*, pages 121–124, Taipei, Taiwan, October 2010.

[CZZ75] D. Chazan, M. Zakaï, and J. Ziv. Improved lower bounds on signal parameter estimation. *IEEE Transactions on Information Theory*, 21(1) :90–93, January 1975.

[EBRM11] M. N. El Korso, R. Boyer, A. Renaux, and S. Marcos. Statistical resolution limit for the multidimensional harmonic retrieval model : Hypothesis test and Cramér-Rao bound approaches. *EURASIP Journal on Advances in Signal Processing*, doi :10.1186/1687-6180-2011-12 :1–14, July 2011.

[EBRM12] M. N. El Korso, R. Boyer, A. Renaux, and S. Marcos. On the asymptotic resolvability of two point sources in known subspace interference using a glrt-based framework. *ELSEVIER Signal Processing*, 92 :2471–2483, October 2012.

[Eld06] Y. C. Eldar. Uniformly improving the Cramér-Rao bound and maximum-likelihood estimation. *IEEE Transactions on Signal Processing*, 54(8) :2943–2956, August 2006.

[FL01] Ph. Forster and P. Larzabal. Sur les bornes minimales pour l'estimation de paramètres déterministes. In *Proc. du colloque GRETSI sur le traitement du signal et des images*, pages 707–710, Toulouse, FR, September 2001.

[FL02] Ph. Forster and P. Larzabal. On lower bounds for deterministic parameter estimation. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 1137–1140, Orlando, FL, USA, May 2002.

[Gal80] J. Galdos. A Cramér-Rao bound for multidimensional discrete-time dynamical systems. *IEEE Transactions on Automatic Control*, 25(1) :117–119, February 1980.

[Gen92] A. Genz. Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1(2) :141–149, 1992.

[GFG14a] M. S. Greco, S. Fortunati, and F. Gini. Maximum likelihood covariance matrix estimation for complex elliptically symmetric distributions under mismatched conditions. *ELSEVIER Signal Processing*, 104(0) :381–386, November 2014.

[GFG14b] M. S. Greco, S. Fortunati, and F. Gini. Naive, robust or fully-adaptive : An estimation problem for CES distributions. In *Proc. of IEEE Workshop on Sensor Array and Multi-channel Processing (SAM)*, pages 457–460, A Coruña, Spain, June 2014.

[GG13] M. S. Greco and F. Gini. Cramér-Rao lower bounds on covariance matrix estimation for complex elliptically symmetric distributions. *IEEE Transactions on Signal Processing*, 61(24) :6401–6409, December 2013.

[Gla72] F. E. Glave. A new look at the Barankin lower bound. *IEEE Transactions on Information Theory*, 18(3) :349–356, May 1972.

[GMV00] F. Gini, M. Montanari, and L. Verrazzani. Maximum likelihood, ESPRIT, and periodogram frequency estimation of radar signals in K-distributed clutter. *ELSEVIER Signal Processing*, 80(6) :1115–1126, January 2000.

[GO90] J. D. Gordman and A. O.Hero. Lower bounds for parametric estimation with constraints. *IEEE Transactions on Information Theory*, 26(6) :1285–1301, November 1990.

[GO91] J. D. Gordman and A. O.Hero. On the application of Cramér-Rao type lower bounds for constrained estimation. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toronto, Canada, 1991.

[Hub67] P. J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 221–233, 1967.

[HV10] A. Hassanien and S.A. Vorobyov. Phased-MIMO Radar : A Tradeoff Between Phased-Array and MIMO Radars. *IEEE Transactions on Signal Processing*, 58(6) :3137–3151, June 2010.

[JLL09] M. Jin, G. Liao, and J. Li. Joint DOD and DOA estimation for bistatic MIMO radar. *ELSEVIER Signal Processing*, 89(2) :244–251, February 2009.

[Kay93] S. M. Kay. *Fundamentals of Statistical Signal Processing : Estimation Theory*, volume 1. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, March 1993.

[Kot59] V. A. Kotel'nikov. *The Theory of Optimum Noise Immunity.* McGraw-Hill, New-York, NY, USA, 1959.

[LC03] E. L. Lehmann and G. Casella. *Theory of Point Estimation.* Springer Texts in Statistics. Springer, New-York, NY, USA, 2 edition, September 2003.

[Lee93] H.B. Lee. The cramer-rao bound on frequency estimates of signals closely spaced in frequency. *IEEE Transactions on Signal Processing*, 40(6) :1507–1517, June 93.

[LM04] N. Levanon and E. Mozeron. *Radar Signals.* Wiley-Interscience, 2004.

[LS09] J. Li and P. Stoica. *MIMO Radar Signal Processing.* John Wiley & Sons, Inc, New Jersey, USA, 2009.

[MCLB12] T. Menni, E. Chaumette, P. Larzabal, and J. P. Barbot. New result on Deterministic Cramér-Rao bounds for real and complex parameters. *IEEE Transactions on Signal Processing*, 60(3) :1032–1049, March 2012.

[Mes06] H. Messer. The hybrid Cramér-Rao lower bound – from practice to theory. In *Proc. of IEEE Workshop on Sensor Array and Multi-channel Processing (SAM)*, pages 304–307, Waltham, MA, USA, July 2006.

[MGCL14] T. Menni, J. Galy, E. Chaumette, and P. Larzabal. Versatility of constrained CRB for system analysis and design. *IEEE Transactions on Aerospace and Electronic Systems*, 50(3) :1841–1863, July 2014.

[MH71] R. J. McAulay and E. M. Hofstetter. Barankin bounds on parameter estimation. *IEEE Transactions on Information Theory*, 17(6) :669–676, November 1971.

[MMY06] R. A. Maronna, D. Martin, and V. Yohai. *Robust statistics.* John Wiley & Sons, Chichester, UK, 2006.

[MS69] R. J. McAulay and L. P. Seidman. A useful form of the Barankin lower bound and its application to PPM threshold analysis. *IEEE Transactions on Information Theory*, 15(2) :273–279, March 1969.

[Mui09] R. J. Muirhead. *Aspects of multivariate statistical theory.* John Wiley & Sons, New Jersey, USA, 2009.

[NM09] Y. Noam and H. Messer. Notes on the tightness of the hybrid Cramér-Rao lower bound. *IEEE Transactions on Signal Processing*, 57(6) :2074–2084, 2009.

[OK91] S. G. Oh and R. L. Kashyap. A robust approach for high resolution frequency estimation. *IEEE Transactions on Signal Processing*, 39(3) :627–643, March 1991.

[Org12] U. Orguner. Calculation of some expected values for parameterized mean model with Gaussian noise. *arXiv :1209.1670v1*, 2012.

[PCO+08] F. Pascal, Y. Chitour, J. Ovarlez, P. Forster, and P. Larzabal. Covariance structure maximum-likelihood estimates in compound gaussian noise : Existence and

algorithm analysis. *IEEE Transactions on Signal Processing*, 56(1) :34–48, January 2008.

[PFOL08] F. Pascal, P. Forster, J. Ovarlez, and P. Larzabal. Performance analysis of covariance matrix estimates in impulsive noise. *IEEE Transactions on Signal Processing*, 56(6) :2206–2217, June 2008.

[PR10] F. Pascal and A. Renaux. Statistical analysis of the covariance matrix MLE in K-distributed clutter. *ELSEVIER Signal Processing*, 90(4) :1165–1175, April 2010.

[PS08] J. G. Proakis and M. Salehi. *Digital Communications (5th Ed.)*. McGraw-Hill, 2008.

[QCL06] A. Quinlan, E. Chaumette, and P. Larzabal. A direct method to generate approximations of the Barankin bound. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 808–811, Toulouse, FR, May 2006.

[Rao45] C. R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37 :81–91, 1945.

[RB74] D. C. Rife and R. R. Boorstyn. Single tone parameter estimation from discrete time observations. *IEEE Transactions on Information Theory*, 20(5) :591–598, September 1974.

[Rendf] A. Renaux. *Contribution à l'analyse des performances d'estimation en traitement statistique du signal.* PhD thesis, Ecole Normale Supérieure de Cachan, Cachan, FR, July 2006. http ://www.bibli.ens-cachan.fr/THESESENS/Renaux/PDF/Renaux2006.pdf.

[RFBL07] A. Renaux, P. Forster, E. Boyer, and P. Larzabal. Unconditional maximum likelihood performance at finite number of samples and high signal-to-noise ratio. *IEEE Transactions on Signal Processing*, 55(5, Part 2) :2358–2364, May 2007.

[RFCL06] 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*, 54(12) :4840–4843, December 2006.

[RFL+08] A. Renaux, P. Forster, P. Larzabal, C. D. Richmond, and A. Nehorai. A fresh look at the Bayesian bounds of the Weiss-Weinstein family. *IEEE Transactions on Signal Processing*, 56(11) :5334–5352, November 2008.

[RGC+13] C. Ren, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux. Hybrid lower bound on the MSE based on the barankin and weiss-weinstein bounds. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 5534–5538, Vancouver, Canada, May 2013.

[RGC+14] C. Ren, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux. A Ziv-Zakaï type bound for hybrid parameter estimation. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 4663–4667, Florence, Italy, May 2014.

[RGC+15a] C. Ren, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux. A constrained hybrid Cramér-Rao bound for parameter estimation. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Australia, May 2015.

[RGC+15b] C. Ren, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux. Hybrid barankin-weiss-weinstein bounds. *IEEE Signal Processing Letters*, accepted with mandatory minor revisions (AQ), June 2015.

[RH13] C. D. Richmond and L. L. Horowitz. Parameter bounds under misspecified models. In *Proc. of Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 176–180, Pacific Grove, CA, USA, November 2013.

[RH15] C. D. Richmond and L. L. Horowitz. Parameter bounds on estimation accuracy under model misspecification. *IEEE Transactions on Signal Processing*, 63(9) :2263–2278, May 2015.

[RkG+14] C. Ren, M. N. El korso, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux. On the accuracy and resolvability of vector parameter estimates. *IEEE Transactions on Signal Processing*, 14(62) :3682–3694, July 2014.

[RM97] I. Reuven and H. Messer. A Barankin-type lower bound on the estimation error of a hybrid parameter vector. *IEEE Transactions on Information Theory*, 43(3) :1084–1093, May 1997.

[RN05] S. Reece and D. Nicholson. Tighter alternatives to the Cramér-Rao lower bound for discrete-time filtering. In *8th International Conference on Information Fusion*, volume 1, pages 1645–1648, July 2005.

[RO07] I. Rapoport and Y. Oshman. Weiss-Weinstein lower bounds for markovian systems. part I : Theory. *IEEE Transactions on Signal Processing*, 55(5) :2016–2030, May 2007.

[RS87] Y. Rockah and P. Schultheiss. Array shape calibration using sources in unknown locations–part I : Far-field sources. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(3) :286–299, March 1987.

[SD85] K. Sharman and T. S. Durrani. Resolving power of signal subspace methods for finite data lengths. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages 1501–1504, April 1985.

[Seb08] G. A. Seber. *A matrix handbook for statisticians*, volume 15. John Wiley & Sons, 2008.

[Sim06] D. Simon. *Optimal State Estimation : Kalman, H-infinity, and Nonlinear Approaches.* Wiley InterScience, 2006.

[SK61] B. K. Shah and C. G. Khatri. Distribution of a definite quadratic form for non-central normal variates. *The Annals of Mathematical Statistics*, pages 883–887, 1961.

[Sle54] D. Slepian. Estimation of signal parameters in the presence of noise. *Trans. IRE Professional Group on Information Theory PG IT-3*, pages 68–89, 1954.

[SM93] L. L. Scharf and L. T. McWhorter. Geometry of the Cramér-Rao bound. *ELSEVIER Signal Processing*, 31(3) :1–11, April 1993.

[SM04] M. Shahram and P. Milanfar. Imaging below the diffraction limit : a statistical analysis. *IEEE Transactions on Image Processing*, 13(5) :677–689, May 2004.

[SM05] P. Stoica and R. Moses. *Spetral analysis of signals.* Prentice-Hall, Inc., Upper Saddle River, New Jersey, 2005.

[Smi05] S. T. Smith. Statistical resolution limits and the complexified Cramér-Rao bound. *IEEE Transactions on Signal Processing*, 53(5) :1597–1609, May 2005.

[SN98] P. Stoica and B. Chong Ng. On the Cramér-Rao bound under parametric constraints. *IEEE Signal Processing Letters*, 5(7) :177–179, 1998.

[TMN98] P. Tichavsky, C. H. Muravchik, and A. Nehorai. Posterior Cramér-Rao bounds for discrete-time nonlinear filtering. *IEEE Transactions on Signal Processing*, 46(5) :1386–1396, May 1998.

[TRB+11] N. D. Tran, A. Renaux, R. Boyer, S. Marcos, and P. Larzabal. MIMO radar in the presence of modeling errors : A Cramér-Rao Bound investigation. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 2780–2783, May 2011.

[TRB+14] N. D. Tran, A. Renaux, R. Boyer, S. Marcos, and P. Larzabal. Performance bounds for the pulse-phase estimation of X-ray pulsars. *IEEE Transactions on Aerospace and Electronic Systems*, 50(1) :786–793, January 2014.

[TT09] K. Todros and J. Tabrikian. Hybrid lower bound via compression of the sampled CLR function. In *Proc. of IEEE Workshop on Statistical Signal Processing (SSP)*, pages 602–605, Cardiff, Wales, UK, August 2009.

[TT10a] K. Todros and J. Tabrikian. General classes of performance lower bounds for parameter estimation - part I : non-Bayesian bounds for unbiased estimators. *IEEE Transactions on Information Theory*, 56(10) :5045–5063, October 2010.

[TT10b] K. Todros and J. Tabrikian. General classes of performance lower bounds for parameter estimation - part II : Bayesian bounds. *IEEE Transactions on Information Theory*, 56(10) :5064–5082, October 2010.

[TW07] P. Tichavský and K. Wong. Quasi-fluid-mecanics based quasi-Bayesian Cramér-Rao bounds for towed-array direction finding. *IEEE Transactions on Signal Processing*, 52(1) :36–47, January 2007.

[Van68] H. L. Van Trees. *Detection, Estimation and Modulation Theory*, volume 1. John Wiley & Sons, New-York, NY, USA, 1968.

[Van02] H. L. Van Trees. *Detection, Estimation and Modulation theory : Optimum Array Processing*, volume 4. John Wiley & Sons, New-York, NY, USA, March 2002.

[VB07] H. L. Van Trees and K. L. Bell, editors. *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Wiley/IEEE Press, New-York, NY, USA, September 2007.

[VRBM14] D. T. Vu, A. Renaux, R. Boyer, and S. Marcos. Some results on the Weiss-Weinstein bound for conditional and unconditional signal models in array processing. *ELSEVIER Signal Processing*, 95 :126–148, February 2014.

[VVRB12] J. Vilà-Valls, L. Ros, and J. M. Brossier. Joint oversampled carrier and time-delay synchronization in digital communications with large excess bandwidth. *ELSEVIER Signal Processing*, 92(1) :76–88, January 2012.

[Whi82] H. White. Maximum likelihood estimation of misspecified models. *Econometrica : Journal of the Econometric Society*, 34 :1–25, 1982.

[WW85] A. J. Weiss and E. Weinstein. A lower bound on the mean square error in random parameter estimation. *IEEE Transactions on Information Theory*, 31(5) :680–682, September 1985.

[WW88] E. Weinstein and A. J. Weiss. A general class of lower bounds in parameter estimation. *IEEE Transactions on Information Theory*, 34(2) :338–342, March 1988.

[XGMM13] F. Xaver, P. Gerstoft, G. Matz, and C. F. Mecklenbrauker. Analytic sequential weiss-weinstein bounds. *IEEE Transactions on Signal Processing*, 61(20) :5049–5062, October 2013.

[YB92] S. F. Yau and Y. Bresler. Worst case cramer-rao bounds for parametric estimation of superimposed signals with applications. *IEEE Transactions on Signal Processing*, 40(12) :2973–2986, December 1992.

[YGB11] J. Yang, B. Geller, and S. Bay. Bayesian and Hybrid Cramér-Rao bounds for the Carrier Recovery Under Dynamic Phase Uncertain Channels. *IEEE Transactions on Signal Processing*, 59(2) :667–680, February 2011.

[ZKCM12] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma. Robust Estimation in Signal Processing : A Tutorial-Style Treatment of Fundamental Concepts. *IEEE Signal Processing Magazine*, 29(4) :61–80, July 2012.

[ZZ69] J. Ziv and M. Zakai. Some lower bounds on signal parameter estimation. *IEEE Transactions on Information Theory*, 15(3) :386–391, May 1969.