



# **Performances et méthodes pour l'échantillonnage comprimé : Robustesse à la méconnaissance du dictionnaire et optimisation du noyau d'échantillonnage.**

Stéphanie Bernhardt

## **► To cite this version:**

Stéphanie Bernhardt. Performances et méthodes pour l'échantillonnage comprimé : Robustesse à la méconnaissance du dictionnaire et optimisation du noyau d'échantillonnage.. Traitement du signal et de l'image [eess.SP]. Université Paris Saclay (COMUE), 2016. Français. ⟨NNT : 2016SACLS443⟩. ⟨tel-01462276⟩

**HAL Id: tel-01462276**

**<https://theses.hal.science/tel-01462276v1>**

Submitted on 8 Feb 2017

**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.



HAL Authorization

NNT : 2016SACLS443

THÈSE DE DOCTORAT  
DE L'UNIVERSITÉ PARIS-SACLAY  
PRÉPARÉE À L'UNIVERSITÉ PARIS-SUD

Ecole doctorale n°580  
Sciences et Technologies de l'Information et de la  
Communication  
Spécialité de doctorat : traitement du signal et des images  
par

**MME STÉPHANIE BERNHARDT**  
Performances et méthodes pour l'échantillonnage  
comprimé: Robustesse à la méconnaissance du  
dictionnaire et optimisation du noyau d'échantillonnage

Thèse présentée et soutenue à Gif-sur-Yvette, le 5 décembre 2016.

Composition du Jury :

|      |                      |                                                         |                          |
|------|----------------------|---------------------------------------------------------|--------------------------|
| M.   | ALI MOHAMMAD-DJAFARI | Professeur des universités<br>Université Paris-Sud      | (Président du jury)      |
| M.   | DAVID BRIE           | Professeur des universités<br>Université de Lorraine    | (Rapporteur)             |
| M.   | ANDRÉ FERRARI        | Professeur des universités<br>Université de Côte d'Azur | (Rapporteur)             |
| M.   | ERIC CHAUMETTE       | Enseignant - Chercheur<br>ISAE-Supaéro                  | (Examineur)              |
| M.   | NICOLAS DOBIGEON     | Professeur des universités<br>Université de Toulouse    | (Examineur)              |
| M.   | RÉMY BOYER           | Maître de conférences<br>Université Paris-Sud           | (Directeur de thèse)     |
| Mme. | SYLVIE MARCOS        | Directeur de recherche<br>CNRS                          | (Co-directrice de thèse) |
| M.   | PASCAL LARZABAL      | Professeur des universités<br>Université Paris-Sud      | (Co-encadrant)           |



# Remerciements

Voici venu le moment d'écrire mes remerciements. C'est ainsi que cette belle page du doctorat s'achève, après trois années riches en apprentissages autant scientifiques, humains, et finalement surtout en apprentissages sur moi-même. Si j'ai pu mener cette entreprise à son terme, malgré tous les bonheurs et aléas de la vie qui ont jalonné ces trois années, c'est grâce à mon entourage aussi bien professionnel que personnel et c'est pour cela que je souhaite tous vous remercier à présent.

Je souhaite commencer par remercier les membres du jury, sans qui je ne pourrais pas écrire ces mots aujourd'hui. Je les remercie de m'avoir fait l'honneur de participer à mon jury de thèse. Merci à David Brie et André Ferrari d'avoir accepté et pris le temps de rapporter ma thèse. Merci de leur relecture et de leurs remarques sur mon manuscrit, qui n'était sûrement pas le plus agréable à aborder. Merci à Ali Mohammad-Djafari de m'avoir fait l'honneur de présider le jury de thèse. Merci à Eric Chaumette pour sa relecture détaillée et pour ses remarques pertinentes, qui m'ont permis de prendre plus de recul sur mon travail et d'améliorer mon manuscrit. Merci à Nicolas Dobigeon pour sa présence, ces remarques et questions lors de la soutenance.

Un grand merci à mon directeur de thèse Remy Boyer pour l'encadrement dont j'ai bénéficié. Merci de m'avoir fait découvrir l'intérêt des bornes, et d'avoir eu la patience de toujours devoir convaincre la critique que je suis. Merci d'avoir accepté mes contraintes personnelles, et de m'avoir permis de bénéficier d'un environnement de travail très souple qui m'a aidé à allier ma vie de famille avec mon travail de thèse. Merci également pour son investissement dans le sujet, pour toutes ses idées et conseils qui ont permis de toujours continuer à avancer. Merci également pour toutes les heures d'enseignements que nous avons partagées, qui m'ont fait découvrir une facette du métier d'enseignant chercheur que j'ai appréciée. Merci à Pascal Larzabal et Sylvie Marcos pour leur présence, leur soutiens, leur gentillesse et surtout pour toutes les corrections, discussions et remarques qui ont enrichi mon travail.

Je remercie également mes collègues doctorants du laboratoire, Lucien, Diane, Eugénie, Mathieu, qui ont par leur présence donnés de l'intérêt à se déplacer jusqu'à Gif-sur-Yvette. Merci à Chengfang pour sa gentillesse, et pour sa présence fidèle au laboratoire. Merci d'avoir pris le temps de répondre à mes questions, surtout en fin de thèse, j'espère que malgré nos orientations différentes nous aurons un jour l'occasion de collaborer dans notre travail. Merci à Olivier et Chengfang pour les moments passés au GRETSI (et pour avoir porté mon sac), en espérant nous y retrouver en 2017.

Je souhaite également remercier les personnes qui ont jalonné ma vie d'étudiante, et par qui mon intérêt pour le traitement du signal a grandi. Je pense en particulier à Pierre Tercinet, qui m'a fait découvrir le plaisir de la lecture bibliographique et du travail de recherche. Merci d'avoir été là au bon moment pour moi, de m'avoir soutenue et motivée, et merci également de sa présence à ma soutenance qui a apporté une saveur encore plus

---

particulière à cette journée. Merci également à Cyrille Enderli, avec qui j'ai eu tout autant de plaisir à travailler en tant que stagiaire, et avec qui j'ai pu approfondir ma volonté de me lancer dans une thèse. Je suis heureuse de pouvoir avoir de nouveau l'occasion aujourd'hui de travailler ensemble.

Je souhaite également penser à toutes les personnes dans mon entourage familial et amical qui m'ont aidé par leur temps, leurs discussions et leur présence à passer trois années heureuses, et mener à bien ce travail. Merci à Eeva et Alain Gourdin qui ont été présents pour nous soutenir lors de la naissance de nos enfants, ce qui m'a permis de continuer malgré tout à poursuivre mes études. Merci d'avoir gardé Elias autant de fois qu'il a fallu, pour que je puisse notamment me déplacer en conférence. Merci à Diandra qui a partagé l'aventure de la thèse avec moi à distance et félicitations pour ton doctorat ! Merci à mes amies mamans chrétiennes, qui par leur présence et leur soutien m'ont permis de retrouver le moral et la motivation dans les périodes plus difficiles. Merci à mes parents premièrement de m'avoir élevée, d'avoir toujours eu une grande confiance en moi, ce qui m'a permis d'arriver où je suis aujourd'hui. Merci également pour l'organisation des aspects pratiques de ma soutenance. Je voulais aussi remercier mes enfants, Elias et Leif, qui m'ont apporté un équilibre important lors d'un travail de thèse. Merci d'avoir été patient pendant la rédaction, et sage pour que maman puisse travailler. Merci à Leif d'avoir fait ses nuits assez tôt et d'avoir été un bébé calme, ce qui fut une aide non négligeable.

En dernier lieu je voulais remercier mon compagnon de vie Ilmo, qui a vécu tous ses événements avec moi. Je ne saurais pas exprimer à quel point sa présence m'a été et m'est précieuse, mais je suis certaine que sans lui je ne serais pas qui je suis et où je suis aujourd'hui. Il a été là dans toutes les étapes du doctorat, du début des réflexions sur une éventuelle poursuite en thèse ou non, à la soutenance. Cette thèse est un peu la nôtre, merci d'avoir pallié à tous mes manquements domestiques lors de la rédaction, quitte à y consacrer tes congés. Merci d'avoir été mon coach quand j'avais envie de baisser les bras. Merci de me rendre heureuse.

Stéphanie

17 janvier 2017

**Titre : Performances et méthodes pour l'échantillonnage comprimé: Robustesse à la méconnaissance du dictionnaire et optimisation du noyau d'échantillonnage**

**Mots Clés :** parcimonie, erreur de modèles, signaux impulsionnels, noyaux, bornes Bayésiennes, échantillonnage

**Résumé :** Dans cette thèse, nous nous intéressons à deux méthodes permettant de reconstruire un signal parcimonieux largement sous-échantillonné : l'échantillonnage de signaux à taux d'innovation fini et l'acquisition comprimée.

Il a été montré récemment qu'en utilisant un noyau de pré-filtrage adapté, les signaux impulsionnels peuvent être parfaitement reconstruits bien qu'ils soient à bande non-limitée. En présence de bruit, la reconstruction est réalisée par une procédure d'estimation de tous les paramètres du signal d'intérêt. Dans cette thèse, nous considérons premièrement l'estimation des amplitudes et retards paramétrisant une somme finie d'impulsions de Dirac filtrée par un noyau quelconque et deuxièmement l'estimation d'une somme d'impulsions de forme quelconque filtrée par un noyau en somme de sinus cardinaux (SoS). Le noyau SoS est intéressant car il est paramétrable par un jeu de paramètres à valeurs complexes et vérifie les conditions nécessaires à la reconstruction. En se basant sur l'information de Fisher Bayésienne relative aux paramètres d'amplitudes et de retards et sur des outils d'optimisation convexe, nous proposons un nouveau noyau d'échantillonnage.

L'acquisition comprimée permet d'échantillonner un signal en-dessous de la fréquence d'échantillonnage de Shannon, si le vecteur à échantillonner peut être approximé comme une combinaison linéaire d'un nombre réduit de vecteurs extraits d'un dictionnaire sur-complet. Malheureusement, dans des conditions réalistes, le dictionnaire (ou base) n'est souvent pas parfaitement connu, et est donc entaché d'une erreur (DB). L'estimation par dictionnaire, se basant sur les mêmes principes, permet d'estimer des paramètres à valeurs continues en les associant selon une grille partitionnant l'espace des paramètres. Généralement, les paramètres ne se trouvent pas sur la grille, ce qui induit une erreur d'estimation même à fort rapport signal sur bruit (RSB). C'est le problème de l'erreur de grille (EG). Dans cette thèse nous étudions les conséquences des modèles d'erreur DB et EG en terme de performances bayésiennes et montrons qu'un biais est introduit même avec une estimation parfaite du support et à fort RSB. La BCRB est dérivée pour les modèles DB et EG non structurés, qui bien qu'ils soient très proches, ne sont pas équivalents en terme de performances. Nous donnons également la borne de Cramér-Rao moyennée (BCRM) dans le cas d'une petite erreur de grille et étudions l'expression analytique de l'erreur quadratique moyenne bayésienne (BEQM) sur l'estimation de l'erreur de grille à fort RSB. Cette dernière est confirmée en pratique dans le contexte de l'estimation de fréquence pour différents algorithmes de reconstruction parcimonieuse.

Nous proposons deux nouveaux estimateurs : le Bias-Correction Estimator (BiCE) et l'Off-Grid Error Correction (OGEC) permettant de corriger l'erreur de modèle induite par les erreurs DB et EG, respectivement. Ces deux estimateurs principalement basés sur une projection oblique des mesures sont conçus comme des post-traitements, destinés à réduire le biais d'estimation suite à une pré-estimation effectuée par n'importe quel algorithme de reconstruction parcimonieuse. Les biais et variances théoriques du BiCE et du OGEC sont dérivés afin de caractériser leurs efficacités statistiques. Nous montrons, dans le contexte difficile de l'échantillonnage des signaux impulsionnels à bande non-limitée que ces deux estimateurs permettent de réduire considérablement l'effet de l'erreur de modèle sur les performances d'estimation. Les estimateurs BiCE et OGEC sont tout deux des schémas (i) génériques, car ils peuvent être associés à tout estimateur parcimonieux de la littérature, (ii) rapides, car leur coût de calcul reste faible comparativement au coût des estimateurs parcimonieux, et (iii) ont de bonnes propriétés statistiques.

**Title : Performance and methods for sparse sampling : robustness to basis mismatch and kernel optimization**

**Keywords :** sparsity, basis mismatch, finite rate of innovation signals, kernel, sampling, Bayesian bounds

**Abstract :** In this thesis, we are interested in two different low rate sampling schemes that challenge Shannon's theory: the sampling of finite rate of innovation signals and compressed sensing.

Recently it has been shown that using appropriate sampling kernel, finite rate of innovation signals can be perfectly sampled even though they are non-bandlimited. In the presence of noise, reconstruction is achieved by a model-based estimation procedure. In this thesis, we consider the estimation of the amplitudes and delays of a finite stream of Dirac pulses using an arbitrary kernel and the estimation of a finite stream of arbitrary pulses using the Sum of Sincs (SoS) kernel. In both scenarios, we derive the Bayesian Cram r-Rao Bound (BCRB) for the parameters of interest. The SoS kernel is an interesting kernel since it is totally configurable by a vector of weights. In the first scenario, based on convex optimization tools, we propose a new kernel minimizing the BCRB on the delays, while in the second scenario we propose a family of kernels which maximizes the Bayesian Fisher Information, i.e., the total amount of information about each of the parameter in the measures. The advantage of the proposed family is that it can be user-adjusted to favor either of the estimated parameters.

Compressed sensing is a promising emerging domain which outperforms the classical limit of the Shannon sampling theory if the measurement vector can be approximated as the linear combination of few basis vectors extracted from a redundant dictionary matrix. Unfortunately, in realistic scenario, the knowledge of this basis or equivalently of the entire dictionary is often uncertain, i.e. corrupted by a Basis Mismatch (BM) error. The related estimation problem is based on the matching of continuous parameters of interest to a discretized parameter set over a regular grid. Generally, the parameters of interest do not lie in this grid and there exists an estimation error even at high Signal to Noise Ratio (SNR). This is the off-grid (OG) problem. The consequence of the BM and the OG mismatch problems is that the estimation accuracy in terms of Bayesian Mean Square Error (BMSE) of popular sparse-based estimators collapses even if the support is perfectly estimated and in the high Signal to Noise Ratio (SNR) regime. This saturation effect considerably limits the effective viability of these estimation schemes.

In this thesis, the BCRB is derived for CS model with unstructured BM and OG. We show that even though both problems share a very close formalism, they lead to different performances. In the biased dictionary based estimation context, we propose and study analytical expressions of the Bayesian Mean Square Error (BMSE) on the estimation of the grid error at high SNR. We also show that this class of estimators is efficient and thus reaches the Bayesian Cram r-Rao Bound (BCRB) at high SNR. The proposed results are illustrated in the context of line spectra analysis for several popular sparse estimator. We also study the Expected Cram r-Rao Bound (ECRB) on the estimation of the amplitude for a small OG error and show that it follows well the behavior of practical estimators in a wide SNR range.

In the context of BM and OG errors, we propose two new estimation schemes called Bias-Correction Estimator (BiCE) and Off-Grid Error Correction (OGEC) respectively and study their statistical properties in terms of theoretical bias and variances. Both estimators are essentially based on an oblique projection of the measurement vector and act as a post-processing estimation layer for any sparse-based estimator and mitigate considerably the BM (OG respectively) degradation. The proposed estimators are generic since they can be associated to any sparse-based estimator, fast, and have good statistical properties. To illustrate our results and propositions, they are applied in the challenging context of the compressive sampling of finite rate of innovation signals.

# Table des matières

|                                                                                                 |            |
|-------------------------------------------------------------------------------------------------|------------|
| <b>Liste des figures</b>                                                                        | <b>iii</b> |
| <b>Liste des tableaux</b>                                                                       | <b>1</b>   |
| <b>1 Échantillonnage comprimé : comment battre Shannon ?</b>                                    | <b>3</b>   |
| 1.1 Rappels sur l'échantillonnage classique . . . . .                                           | 4          |
| 1.2 Échantillonnage de signaux à modèle paramétrique . . . . .                                  | 5          |
| 1.3 Acquisition comprimée . . . . .                                                             | 13         |
| 1.4 Présentation des contributions de la thèse . . . . .                                        | 27         |
| <b>2 Optimisation du noyau d'échantillonnage pour des signaux impulsionnels</b>                 | <b>31</b>  |
| 2.1 Comment étudier les performances de la reconstruction des signaux impulsionnels ? . . . . . | 32         |
| 2.2 Performance de la reconstruction d'un signal à somme de Dirac . . . . .                     | 33         |
| 2.3 Cas du noyau à sommes de sinus cardinaux . . . . .                                          | 57         |
| 2.4 Estimation par dictionnaire pour les signaux impulsionnels . . . . .                        | 76         |
| <b>3 Robustesse à la méconnaissance du dictionnaire</b>                                         | <b>79</b>  |
| 3.1 Deux types d'erreur de modèle . . . . .                                                     | 80         |
| 3.2 Bornes Bayésiennes de Van Trees . . . . .                                                   | 83         |
| 3.3 BCRM en cas de petites erreurs de grille . . . . .                                          | 93         |
| 3.4 Conséquence de la discrétisation sur l'estimation des fréquences . . . . .                  | 104        |
| <b>4 Correction des erreurs de discrétisation</b>                                               | <b>119</b> |
| 4.1 Algorithme OGEC pour corriger l'erreur EG . . . . .                                         | 120        |
| 4.2 Algorithme BiCE pour corriger le DB . . . . .                                               | 132        |
| <b>5 Conclusion et Perspectives</b>                                                             | <b>149</b> |
| 5.1 Échantillonnage de signaux impulsionnels . . . . .                                          | 149        |
| 5.2 Robustesse des méthodes par dictionnaire . . . . .                                          | 150        |
| 5.3 Correction du biais . . . . .                                                               | 151        |
| 5.4 Perspectives générales . . . . .                                                            | 151        |
| <b>A Liste des acronymes</b>                                                                    | <b>I</b>   |



# Liste des figures

|      |                                                                                                                    |    |
|------|--------------------------------------------------------------------------------------------------------------------|----|
| 1.1  | Schéma général d'échantillonnage en trois étapes . . . . .                                                         | 4  |
| 1.2  | Schéma de l'échantillonnage du signal de somme d'impulsions . . . . .                                              | 7  |
| 1.3  | Noyaux classiques . . . . .                                                                                        | 13 |
| 1.4  | Fonction de cohérence . . . . .                                                                                    | 20 |
| 2.1  | Dérivée des noyaux classiques . . . . .                                                                            | 40 |
| 2.2  | Vérification de l'approximation pour le calcul de $\text{BCR}(\tau_l)$ . . . . .                                   | 41 |
| 2.3  | Vérification de l'approximation pour le calcul de $\text{BCR}(a_l)$ . . . . .                                      | 42 |
| 2.4  | Vérification de l'approximation pour le calcul de $\text{BCR}(\tau_l)$ pour des retards hors de la grille. . . . . | 43 |
| 2.5  | Vérification de l'approximation pour le calcul de $\text{BCR}(a_l)$ pour des retards hors de la grille. . . . .    | 43 |
| 2.6  | Visualisation de $f(t)$ . . . . .                                                                                  | 44 |
| 2.7  | Visualisation de $f'(t)$ . . . . .                                                                                 | 45 |
| 2.8  | Bornes pour 3 retards bien espacés sur la grille . . . . .                                                         | 47 |
| 2.9  | Bornes pour 3 retards en dehors de la grille . . . . .                                                             | 48 |
| 2.10 | Bornes pour des impulsions proches . . . . .                                                                       | 49 |
| 2.11 | Bornes pour des impulsions près des bords . . . . .                                                                | 49 |
| 2.12 | Coefficients optimaux pour des impulsions de Dirac . . . . .                                                       | 56 |
| 2.13 | Noyau optimal et de Dirichlet correspondant en temps . . . . .                                                     | 56 |
| 2.14 | Dérivée du noyau optimal et de Dirichlet . . . . .                                                                 | 57 |
| 2.15 | Coefficients optimaux obtenus . . . . .                                                                            | 65 |
| 2.16 | Noyau optimal pour une impulsion de Dirac . . . . .                                                                | 66 |
| 2.17 | Impulsion gaussienne, en fréquence . . . . .                                                                       | 68 |
| 2.18 | Indice du pic . . . . .                                                                                            | 69 |
| 2.19 | BCRB sur les retards du noyau optimisé . . . . .                                                                   | 70 |
| 2.20 | BCRB sur les amplitudes du noyau optimisé . . . . .                                                                | 71 |
| 2.21 | Coefficients optimaux . . . . .                                                                                    | 72 |
| 2.22 | Noyaux optimaux . . . . .                                                                                          | 73 |
| 2.23 | Coefficients optimaux pour une gaussienne . . . . .                                                                | 74 |
| 2.24 | Noyau optimal pour une gaussienne . . . . .                                                                        | 75 |
| 2.25 | BCRB sur les retards du noyau optimisé . . . . .                                                                   | 75 |
| 2.26 | BCRB sur les amplitudes du noyau optimisé . . . . .                                                                | 76 |
| 2.27 | Schéma de l'acquisition comprimée d'un signal impulsionnel . . . . .                                               | 78 |
| 3.1  | Schéma d'échantillonnage comprimé avec dictionnaire bruité . . . . .                                               | 81 |
| 3.2  | Schéma d'échantillonnage comprimé avec erreur de grille . . . . .                                                  | 82 |
| 3.3  | BCRB EG et DB, pour $N = 100$ et $L = 5$ . . . . .                                                                 | 91 |

|      |                                                                                                                                            |     |
|------|--------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 3.4  | Différence entre les bornes, pour $N = 100$ et $L = 5$                                                                                     | 92  |
| 3.5  | Différence entre les bornes, pour $N = 1000$ et $L = 120$                                                                                  | 92  |
| 3.6  | Différence entre les bornes, en fonction de $L/N$                                                                                          | 93  |
| 3.7  | BCR déterministe, pour $N = 500$ , $L = 10$ et $\sigma^2 = 0.01$                                                                           | 95  |
| 3.8  | BCR déterministe en fonction de la variance de l'erreur, pour $N = 1000$ , $L = 50$ et $\sigma^2 = 0.001$                                  | 96  |
| 3.9  | BCRM exacte et approximée                                                                                                                  | 98  |
| 3.10 | Bornes pour l'EG, pour $N = 1000$ , $L = 50$                                                                                               | 100 |
| 3.11 | Bornes pour l'EG, pour $N = 50$ , $L = 3$                                                                                                  | 100 |
| 3.12 | Bornes pour l'EG, pour $N = 50$ , $L = 10$                                                                                                 | 101 |
| 3.13 | Bornes pour l'EG, pour $N = 1000$ , $L = 50$                                                                                               | 101 |
| 3.14 | Borne déterministe pour les problèmes EG et DB, pour $N = 1000$ , $L = 50$ et $\sigma^2 = 0.001$ .                                         | 103 |
| 3.15 | Borne déterministe pour les problèmes EG et DB, pour $N = 1000$ , $L = 50$ et $\sigma^2 = 0.1$ .                                           | 104 |
| 3.16 | Distributions tronquées modélisant une estimation par dictionnaire                                                                         | 106 |
| 3.17 | Exemple de résultat d'estimation de l'algorithme SPGL1                                                                                     | 113 |
| 3.18 | BEQM pour une distribution uniforme. $L = 1, K = 30, N = 10$ .                                                                             | 115 |
| 3.19 | BEQM pour une distribution uniforme. $L = 1, K = 30, N = 20$ .                                                                             | 115 |
| 3.20 | BEQM pour une distribution uniforme. $L = 1, K = 60, N = 30$ .                                                                             | 116 |
| 3.21 | BEQM pour une distribution gaussienne. $L = 1, K = 60, N = 30$ .                                                                           | 116 |
| 3.22 | BEQM pour une distribution uniforme. $L = 1, K = 130, N = 100$ .                                                                           | 117 |
| 3.23 | BEQM pour une distribution gaussienne. $L = 1, K = 130, N = 100$ .                                                                         | 117 |
| 3.24 | BEQM pour une distribution uniforme. $L = 2, K = 130, N = 100$ .                                                                           | 118 |
| 3.25 | BEQM pour une distribution uniforme. $L = 2, K = 100, N = 90$ .                                                                            | 118 |
| 4.1  | Description de l'algorithme OGEC- $\mathcal{A}$ .                                                                                          | 124 |
| 4.2  | Résultat de l'algorithme $\mathcal{A}$ -OGEC sur l'estimation de l'amplitude.                                                              | 131 |
| 4.3  | Résultat de l'algorithme $\mathcal{A}$ -OGEC sur l'estimation de l'erreur.                                                                 | 132 |
| 4.4  | Description de l'algorithme BiCE- $\mathcal{A}$ .                                                                                          | 134 |
| 4.5  | Efficacité statistique du BiCE                                                                                                             | 143 |
| 4.6  | Performance théorique du BiCE                                                                                                              | 144 |
| 4.7  | Performance exacte et approximée comparée avec les bornes BCRM                                                                             | 146 |
| 4.8  | Résultat de l'algorithme BiCE- $\mathcal{A}$ pour différents $\mathcal{A}$ , $N = 200$ , $L = 10$ , $K = 1000$ et $\sigma_e^2 = 10^{-6}$ . | 147 |
| 4.9  | Résultat de l'algorithme BiCE- $\mathcal{A}$ pour différents $\mathcal{A}$ , $N = 500$ , $L = 10$ , $K = 1000$ et $\sigma_e^2 = 10^{-5}$ . | 147 |

# Liste des tableaux

|     |                                                                              |     |
|-----|------------------------------------------------------------------------------|-----|
| 2.1 | MIB et BCRB sur les retards pour les noyaux SoS de Dirichlet et optimal. . . | 55  |
| 2.2 | BCRB obtenues après l'optimisation sur le premier modèle . . . . .           | 67  |
| 2.3 | BCRB obtenues après l'optimisation sur le 2e modèle . . . . .                | 73  |
| 4.1 | Temps de calcul de différents algorithmes d'estimation. . . . .              | 137 |



# Chapitre 1

## Echantillonnage comprimé : comment battre Shannon ?

### Sommaire

---

|                                                                                                |           |
|------------------------------------------------------------------------------------------------|-----------|
| <b>1.1 Rappels sur l'échantillonnage classique</b> . . . . .                                   | <b>4</b>  |
| <b>1.2 Echantillonnage de signaux à modèle paramétrique</b> . . . . .                          | <b>5</b>  |
| 1.2.1 Qu'appelle-t-on signal à modèle paramétrique ? . . . . .                                 | 5         |
| 1.2.2 Méthodes d'échantillonnage des signaux à modèle paramétrique . . . . .                   | 7         |
| 1.2.3 Les noyaux de pré-filtrage . . . . .                                                     | 10        |
| <b>1.3 Acquisition comprimée</b> . . . . .                                                     | <b>13</b> |
| 1.3.1 Modèle d'échantillonnage comprimé . . . . .                                              | 13        |
| 1.3.2 Reconstruction du signal . . . . .                                                       | 15        |
| 1.3.3 Le problème de l'estimation par dictionnaire . . . . .                                   | 17        |
| 1.3.4 Évaluation des performances d'estimation . . . . .                                       | 21        |
| 1.3.5 Le problème des erreurs de modèles . . . . .                                             | 25        |
| <b>1.4 Présentation des contributions de la thèse</b> . . . . .                                | <b>27</b> |
| 1.4.1 Présentation de la thèse . . . . .                                                       | 27        |
| 1.4.2 Contributions sur le sujet des signaux paramétriques . . . . .                           | 27        |
| 1.4.3 Contributions sur le sujet des erreurs de modèles dans l'acquisition comprimée . . . . . | 28        |
| 1.4.4 Publications . . . . .                                                                   | 28        |

---

## 1.1 Rappels sur l'échantillonnage classique

Le principe d'incertitude de Heisenberg nous laisse à penser que pour échantillonner un signal précisément dans le domaine fréquentiel, il faut nécessairement une longue durée d'échantillonnage dans le domaine temporel. Cette propriété est à la base de l'échantillonnage classique, introduit par Shanon [109] qui lie la fréquence d'échantillonnage minimale nécessaires pour reconstruire parfaitement un signal à bande finie à la largeur de la bande en fréquence. Plus le signal est étendu en fréquence, plus l'intervalle d'échantillonnage est réduit et plus le nombre d'échantillons nécessaires sur une durée d'échantillonnage donnée devient grand. La figure 1.1 résume la chaine de l'échantillonnage, sous sa forme la plus générale. Pour échantillonner un signal, on utilise un filtre de pré-filtrage et de post-filtrage qui peuvent être différents. Dans le cadre de la théorie de Shannon la fonction de pré-filtrage et post-filtrage est la même, et correspond au sinus cardinal ( $\phi(t) = g(t) = \text{sinc}(x)$ ). Suite au travail effectué sur les ondelettes, un regain d'attention a été porté sur le problème de l'échantillonnage et des généralisations du théorème de Shannon ont été proposées utilisant d'autres types de fonctions génératrices, moyennant le respect d'un certain nombres de contraintes. UNSER [121] propose un bon résumé de ces travaux jusqu'aux années 2000. Des travaux ont notamment porté sur la généralisation de l'échantillonnage à des fonctions appartenant à d'autres espaces que celui des fonctions à bande limitée. Cela permet, en utilisant des filtres adaptés d'échantillonner des fonctions à bande non limitées. Une revue de ces travaux, dans la continuité de l'article de Unser [121] est proposée dans ELDAR et MICHAELI [60].

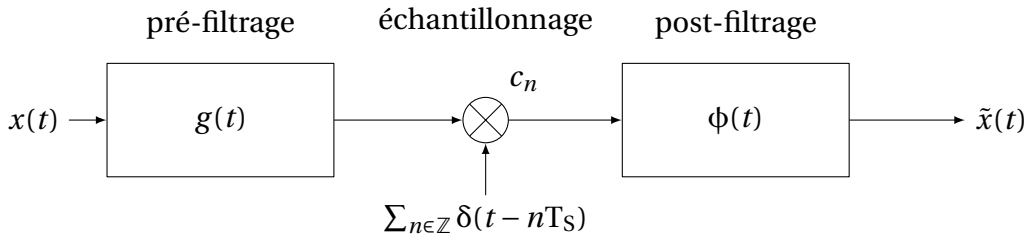


FIGURE 1.1 – Schéma général d'échantillonnage d'un signal  $x(t)$  en trois étapes : filtrage par un noyau  $g(t)$ , échantillonnage avec un intervalle  $T_s$  pour obtenir les échantillons  $c_n$ , puis reconstruction d'un signal analogique  $\tilde{x}(t)$  par filtrage avec un noyau  $\phi(t)$ .

Les limitations sur la précision en fréquence obtenue avec une fenêtre temporelle finie ont également été dépassées dans le domaine de l'estimation spectrale par les méthodes à hautes résolutions type MUSIC ou ESPRIT. Elles permettent de retrouver très précisément les fréquences d'un signal avec un nombre d'échantillons réduit sur une durée d'échantillonnage finie [111]. Ce résultat est possible grâce à une connaissance a priori sur le signal : dans l'exemple des méthodes à haute résolution les éléments composant le signal sont connus. En analyse spectrale notamment on sait que le signal est une combinaison linéaire de sinusoides. Cette approche est appelée paramétrique, car la forme du signal est connue et les seules inconnues sont des paramètres à estimer, nous nous intéressons à ce type de signaux dans la section 1.2. Une autre connaissance a priori sur le signal qui peut être exploitée pour réduire le nombre d'échantillons nécessaires à sa reconstruction est la parcimonie<sup>1</sup>. C'est l'idée derrière l'acquisition comprimée qui a ouvert un grand

1. propriété d'un signal porteur de peu d'information par rapport à sa dimension

nombre de nouvelles perspectives depuis son introduction dans CANDES, ROMBERG et TAO [39]. Nous la présentons de manière générale dans la section 1.3. Ces deux méthodes ont comme point commun de se baser sur un type de parcimonie du signal : les signaux paramétriques sont parcimonieux dans le sens que seul un faible nombre de paramètres porte toute l'information du signal, tandis que dans le cadre de l'acquisition comprimée on considère des signaux déjà discrétisés qui sont parcimonieux dans une base donnée. En ne conservant que les éléments non nuls de la base de parcimonie, les signaux parcimonieux peuvent être réécrits sous la forme de signaux paramétriques. Dans cette thèse nous étudions ces deux méthodes permettant de mettre à profit la parcimonie d'un signal dans le but de réduire le nombre d'échantillons nécessaires à sa reconstruction.

## 1.2 Echantillonnage de signaux à modèle paramétrique

### 1.2.1 Qu'appelle-t-on signal à modèle paramétrique ?

Dans cette thèse, nous appelons "signal à modèle paramétrique" un signal constitué d'une combinaison linéaire d'une forme d'onde connue décalée en temps et pondérée par une amplitude, bien que ce ne soit qu'un modèle restrictif de signal paramétrique. Ce signal sans bruit peut être écrit ainsi :

$$x(t) = \sum_{l \in \mathbb{Z}} a_l h(t - \tau_l) \quad (1.1)$$

On considère que la forme d'onde  $h(t)$  (ou impulsion selon le contexte) est connue et que l'on a un nombre d'impulsions  $L$  fini sur une durée d'échantillonnage finie  $T$ . Ce type de signal est appelé également *signal impulsionnel*, puisqu'il est constitué d'une somme d'impulsions de forme quelconques. On parle en anglais de *stream of pulses*. Le signal impulsionnel reçu est entièrement défini par les paramètres inconnus d'amplitudes  $a_l$  et de retards  $\tau_l$ .

Le taux d'innovation est défini comme le nombre de degrés de liberté du signal  $x(t)$  par unité de temps. On peut le définir formellement en considérant une fonction  $C_x(t_a, t_b)$  qui compte le nombre de paramètres sur l'intervalle de temps  $[t_a, t_b]$  par :

$$\lim_{T \rightarrow +\infty} \frac{1}{T} C_x\left(-\frac{T}{2}, \frac{T}{2}\right) \quad (1.2)$$

Le signal (1.1) a un taux d'innovation de  $2L$ , car on peut reconstruire parfaitement le signal si l'on connaît les  $L$  combinaisons de retard et d'amplitude, la fenêtre étant finie le taux d'innovation est définie sur la durée  $T$ . Il s'agit donc d'un signal à taux d'innovation fini (*finite rate of innovation signal*).

En général dans des applications concrètes, le signal impulsionnel à estimer est bruité, par exemple s'il est reçu suite à une communication entre un émetteur et un récepteur. Dans ce cas le modèle (1.1) devient simplement :

$$x(t) = \sum_{l \in \mathbb{Z}} a_l h(t - \tau_l) + w(t) \quad (1.3)$$

où  $w(t)$  est le bruit additif.

Le modèle paramétrique donné en (1.1) (ou (1.3)) est assez général pour représenter un grand nombre de problèmes connus du traitement du signal. Nous en donnons à présent quelques exemples :

1. *Le problème de l'échantillonnage à bande limitée* : Shannon a montré dans SHANNON [109] que tout signal à bande limitée  $[-B/2, B/2[$  de largeur de bande  $B$  peut s'écrire sous la forme

$$x(t) = \sum_{k \in \mathbb{Z}} x_k \text{sinc}(Bt - k) \quad (1.4)$$

où  $x_k$  sont les échantillons du signal pris avec un intervalle de  $1/B$  :  $x_k = \langle B \text{sinc}(Bt - k), x(t) \rangle = x(k/B)$  où  $\langle f(t), x(t) \rangle = \int_{-\infty}^{+\infty} f^*(t) x(t) dt$  est le produit scalaire de deux fonctions. Les signaux à bande limitée sont représentés comme une fonction linéaire dans la base des sinus cardinaux. Le signal  $x(t)$  a  $B$  degrés de liberté par seconde puisque le signal peut être représenté exactement avec une séquence de nombres réels  $\{x_k\}$  espacés de  $1/B$  secondes. Le taux d'innovation d'un signal de largeur de bande  $B$  est donc égal à  $B$ .

2. *Le radar Doppler à impulsions* : Le principe du radar à impulsions est d'envoyer une impulsion connue  $h(t) = p(t) \cos(2\pi f_0 t + \phi_0)$  pour ensuite en recevoir les échos décalés en temps et en fréquence [81]. Le décalage temporel permet d'estimer la distance à laquelle se trouve l'objet ayant provoqué l'écho (ou cible dans la terminologie radar) et le décalage en fréquence provient de l'effet Doppler et permet d'estimer la vitesse de la cible.  $p(t)$  est une porte permettant d'envoyer plusieurs impulsions courtes,  $f_0$  est la fréquence à laquelle le signal est émis et  $\phi_0$  est sa phase initiale.

L'écho radar est :

$$x(t) = \sum_{l=1}^L A_l h(t - \tau_l) = \sum_{l=1}^L A_l p(t - \tau_l) \cos(2\pi f_0(t - \tau_l) + \phi_0) \quad (1.5)$$

où  $\tau_l = \frac{2d_l}{c} - \frac{2V_l t}{c}$  avec  $c$  la vitesse de la lumière. Les paramètres inconnus sont finalement :  $A_l$  l'amplitude,  $d_l$  la distance de la cible et  $V_l$  la vitesse de la cible. Un signal radar composé de  $L$  impulsions telles que décrites en (1.5) sur une durée d'échantillonnage  $T$  a donc finalement un taux d'innovation fini de  $3L$ .

3. *Le signal à somme de Dirac* : Un cas particulier et limite du modèle (1.1) est le signal composé d'une somme d'impulsions de Dirac. Ce type de signal est intéressant dans le sens qu'il montre une limite de l'échantillonnage classique : une impulsion a une bande infinie qui n'est donc pas échantillonnable au sens de Shannon quel que soit le nombre de points utilisés, puisqu'il n'est pas observable par des échantillons discrets. Ce signal a donc souvent été pris comme référence notamment dans le cadre du développement de l'échantillonnage des signaux à taux d'innovation fini [127, 27], la méthode de reconstruction se basant sur la convolution avec un noyau de pré-filtrage pour "étaier" les impulsions (voir section 1.2.2).

Ce signal est composé d'un nombre fini de  $L$  impulsions de Dirac, pondérées et décalées en temps

$$x(t) = \sum_{l=0}^{L-1} a_l \delta(t - \tau_l). \quad (1.6)$$

Il est intéressant de noter qu'on peut le considérer comme une généralisation du modèle d'une source générant un processus de Poisson qui est un ensemble d'impulsions de Dirac  $\sum_{l \in \mathbb{Z}} \delta(t - \tau_l)$ . Dans le cas d'un processus de Poisson, les différences temporelles entre deux impulsions suivent une distribution exponentielle.

Le taux d'innovation peut être mis en parallèle avec le taux d'information de Shannon, adapté aux signaux paramétriques. Pour ce type de signaux généralement le taux d'innovation est inférieur à  $B$  car on connaît une décomposition plus intéressante que celle proposée dans (1.4). Il est donc intéressant de pouvoir échantillonner ces signaux avec un nombre d'échantillons en rapport avec le taux d'innovation plutôt qu'en rapport avec leur largeur de bande. Les travaux sur ces types de signaux ont montré qu'il est possible en théorie de reconstruire parfaitement des signaux avec un taux d'innovation fini avec un nombre d'échantillons égal au taux d'innovation plus un [122]. Il faut toutefois utiliser plus d'échantillons que cette limite basse dans le cas d'un signal bruité. Le problème d'échantillonnage proprement dit se transforme en un problème d'estimation : il suffit d'estimer correctement les quelques paramètres inconnus pour reconstruire parfaitement le signal de départ.

### 1.2.2 Méthodes d'échantillonnage des signaux à modèle paramétrique

Pour échantillonner des signaux à bande non limitée, il est important de trouver une fonction de pré-filtrage adaptée (rappelons nous le problème de l'échantillonnage présenté dans la figure 1.1). Le processus d'échantillonnage proposé pour les signaux paramétriques (ou à taux d'innovation fini de manière générale) se base sur ce principe. Il est détaillé dans la figure 1.2 et se base sur la convolution du signal  $x(t)$  par un noyau  $g(t)$  avant l'échantillonnage en tant que tel.

Une fois convolué par le noyau  $g(t)$  le signal obtenu est échantillonné de manière classique, avec un intervalle d'échantillonnage  $T_S$  pour finalement obtenir les échantillons :

$$c_n = \langle g(t - nT_S), x(t) \rangle + w_n = \mu_n + w_n \quad (1.7)$$

où  $w_n$  est un bruit blanc gaussien de moyenne nulle et de variance  $\sigma^2$  et  $\langle g(t - nT_S), x(t) \rangle$  est le produit scalaire de deux fonctions continues.

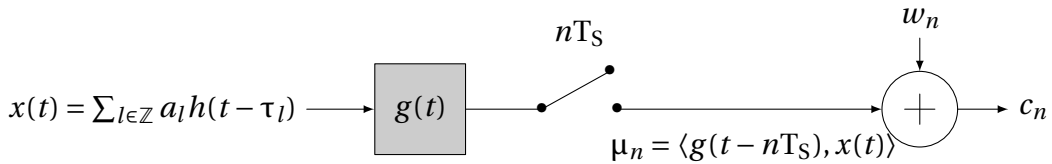


FIGURE 1.2 – Schéma de l'échantillonnage du signal de somme d'impulsions

On utilisera par la suite la notation vectorielle suivante pour les différents paramètres de l'estimation :

$$\mathbf{c} = [c_0 \cdots c_{N-1}]^T \quad (1.8)$$

$$\boldsymbol{\mu} = [\mu_0 \cdots \mu_{N-1}]^T \quad (1.9)$$

$$\mathbf{w} = [w_0 \cdots w_{N-1}]^T. \quad (1.10)$$

où  $\mathbf{x}^T$  est la transposée de  $\mathbf{x}$ .

Le vecteur de mesure peut donc s'écrire  $\mathbf{c} = \boldsymbol{\mu} + \mathbf{w}$  où l'on considère le cas où  $w_n$  est un bruit blanc gaussien réel et discret, de moyenne nulle et de variance  $\sigma^2$ .

VETTERLI, MARZILIANO et BLU [127] ont démontré, en utilisant la méthode de Prony (ou *annihilating filter*), qu'il est possible, connaissant  $\boldsymbol{\mu}$  et le filtre  $g(t)$ , de reconstruire

parfaitement le signal  $x(t)$  à partir de l'estimation de ses paramètres  $\boldsymbol{\theta} = [\tau_1, \dots, \tau_L, a_1, \dots, a_L]^T$ . Les auteurs ont utilisé pour cela un filtre de pré-filtrage en sinus cardinal. Une généralisation de cette méthode est proposée par URIGÜEN et al. [122], qui montrent que l'on peut reconstruire parfaitement les signaux impulsionnels avec la méthode de Prony à condition que le noyau de pré-filtrage suive les propriétés suivantes :

$$G(\omega) \begin{cases} \neq 0 & \omega = \frac{2\pi k}{NT_s}, \quad \{k \in \mathbb{Z} \mid k \in [-K/2, K/2]\} \\ = 0 & \omega = \frac{2\pi k}{NT_s}, \quad \{k \in \mathbb{Z} \mid k \notin [-K/2, K/2]\} \\ \text{arbitraire ailleurs} \end{cases} \quad (1.11)$$

où  $K$  est la largeur de bande du noyau qui doit être choisie de sorte à couvrir toutes les valeurs de  $k$  consécutives pour lesquelles les coefficients de Fourier de l'impulsion  $H\left(\frac{2\pi k}{NT_s}\right)$  sont différents de zéro.  $H(\omega)$  dénote la transformée de Fourier continue de la forme d'impulsion  $h(t)$ . On fait l'hypothèse que pour  $k > K$  les coefficients de Fourier sont tous nuls. L'utilisation de la méthode de Prony pour estimer les paramètres des signaux impulsionnels se base sur une modélisation périodique du signal :

$$\tilde{x}(t) = \sum_{m \in \mathbb{Z}} \sum_{l=0}^{L-1} a_l h(t - \tau_l - mNT_s) \quad (1.12)$$

constitué de la répétition du signal  $x(t)$  de (1.1) échantillonné sur l'intervalle de temps  $\tau = NT_s$ . Une période de  $\tilde{x}$  correspond donc au signal  $x(t)$  que l'on cherche à estimer. Un tel signal est obtenu simplement en faisant une convolution circulaire au moment du pré-filtrage au lieu d'une convolution classique, la convolution circulaire de deux fonctions non périodiques consistant à convoluer de manière classique l'une des fonctions avec la seconde fonction périodisée. Les échantillons peuvent alors être exprimés grâce à la formule sommatoire de Poisson [99] sous la forme d'une transformée de Fourier inverse discrète :

$$\mu_n = \int_{-\infty}^{+\infty} \tilde{x}(t) g(t - nT_s) dt = \sum_{k \in \mathbb{Z}} X[k] e^{\frac{2\pi kn}{N}} G\left(\frac{2\pi k}{NT_s}\right) \quad (1.13)$$

En l'absence de bruit, on peut donc obtenir les coefficients de Fourier  $X[k]$  simplement par une transformée de Fourier de  $\mu_n$ . Si le noyau suit les conditions (1.11), le problème d'estimation peut se résoudre à l'aide de méthodes d'estimation spectrale, par exemple la méthode de Prony proposée par BLU et al. [27], une fois les coefficients de Fourier du signal obtenus à partir des échantillons  $y_n$ .

Il est également possible de combiner les échantillons autrement que par la transformée de Fourier discrète présentée en (1.13), DRAGOTTI, VETTERLI et BLU [56] élargissent les familles de noyaux possibles en changeant cette combinaison, notamment en introduisant une somme d'exponentielles. Les auteurs démontrent que l'échantillonnage est possible pour trois grandes familles de noyaux :

- les noyaux remplissant la condition Strang-Fix [112], autrement dit les noyaux reproduisant des polynômes par une combinaison linéaire du noyau décalé en temps. Un exemple souvent utilisé sont les fonctions B-splines.
- les noyaux qui de la même manière reproduisent des fonctions exponentielles, dont l'exemple le plus courant est le noyau en E-spline.

- les noyaux ayant une transformée de Fourier prenant la forme d'une fonction rationnelle, cas très intéressant en pratique puisqu'il permet de modéliser des méthodes d'acquisition analogique courantes.

Par la suite URIGUEN, BLU et DRAGOTTI [123] ont montré qu'on pouvait utiliser des noyaux arbitraires en cherchant leurs approximations sous la forme du noyau reproduisant des exponentielles. L'estimation est alors possible, bien qu'elle ne soit plus parfaite même en l'absence de bruit.

Les premières méthodes d'estimation des signaux paramétriques se basent sur la représentation des échantillons  $y_n$  comme une combinaison linéaire d'exponentielles, modèle permettant de retrouver les paramètres d'intérêts à l'aide de méthodes dérivées d'estimation spectrale. Une autre approche de résolution est d'exprimer le problème d'estimation sous la forme d'un problème d'optimisation non convexe, et d'utiliser des algorithmes d'optimisation pour trouver le résultat. Ces approches permettent des solutions assez variées, et peuvent prendre en compte la présence du bruit et être ainsi plus robustes. CONDAT, HIRABAYASHI et HIRONAGA [48] par exemple considèrent l'optimisation du maximum de vraisemblance pour un signal composé de sommes de Diracs et le noyau de Dirichlet, et propose un algorithme pour trouver le résultat de ce problème non convexe. MICHAELI et ELDAR [84, 85] proposent d'estimer les paramètres en minimisant l'erreur des moindres carrés entre les échantillons mesurés et les échantillons reconstruits à partir d'un jeu de paramètres donnés. GILLIAM et BLU [64] utilisent un modèle polynomial, adapté au cas particulier des signaux à somme de Dirac échantillonnés avec un noyau sinus cardinal pour faire de l'ajustement de modèle (en anglais *model fitting*), DOĞAN et al. [50] ont élargi la méthode en proposant un modèle général pour les échantillons qu'il utilise pour résoudre le problème à l'aide d'une minimisation de l'erreur de modèle. BENDORY, DEKEL et FEUER [14] et ensuite BENDORY et al. [15] montrent que l'on peut estimer les retards d'une somme d'impulsions sous la forme d'une optimisation convexe, en utilisant des contraintes de types  $l_1$  adaptées au cas continu (a contrario des méthodes basées sur l'acquisition comprimée que nous verrons dans la section 1.3 qui se basent sur des dictionnaires discrets). HUANG et al. [72] proposent d'utiliser directement les méthodes d'estimation par dictionnaire pour estimer les paramètres d'amplitude et de retard d'un signal à somme de Dirac, échantillonné avec un noyau reproducteur d'exponentielles. Cette approche passe par une discrétisation des retards pour générer un dictionnaire, permettant ensuite d'obtenir une solution avec un algorithme tel que l'*Orthogonal Matching Pursuit* (OMP). Cette idée est la même que celle que nous développons dans la section 2.4.1 pour un noyau quelconque.

Des méthodes d'estimation ont été également développées pour des signaux à taux d'innovation finie, mais ne correspondant pas au modèle de la série d'impulsions donnée en (1.1). On peut citer le cas des signaux sinusoïdaux par morceaux traités dans BERENT, DRAGOTTI et BLU [16] où la méthode des filtres annihilants est adaptée aux signaux ayant comme paramètre des fréquences d'oscillation et des phases au lieu des retards que nous considérons. Très récemment MULLETI, SHENOY et SEELAMANTULA [87] ont proposé une méthode pour estimer les signaux composés d'une somme d'exponentielles basés sur de l'annihilation continue. Ces deux modèles de signaux à taux d'innovation fini sont également très utiles en pratique puisqu'ils correspondent au modèle de l'estimation spectrale.

Les méthodes citées jusqu'ici sont toutes algébriques, ou basées sur des algorithmes d'optimisation déterministes. TAN et GOYAL [113] proposent une approche stochastique basée sur une méthode de Gibbs pour estimer un signal de la forme (1.1). ERDOZAIN et

CRESPO [62] étendent cette approche à l'estimation des paramètres d'amplitude et de retards en introduisant des principes d'algorithmes génétiques. Ces approches Bayésiennes permettent de se détacher des conditions habituelles sur les noyaux, l'exemple de TAN et GOYAL [113] le montre puisqu'il utilise un noyau gaussien qui est l'exemple typique d'un noyau ne remplissant pas les conditions Strang-Fix. AMINI, KAMILOV et UNSER [3] proposent également une approche stochastique pour débruiter les échantillons soumis à un bruit généré par un processus de Poisson et ainsi retrouver directement les échantillons, plutôt que de passer par l'estimation des paramètres du signal.

### 1.2.3 Les noyaux de pré-filtrage

Nous avons vu dans 1.2.2 que l'échantillonnage d'un signal à modèle paramétrique est possible avec peu de points grâce à l'utilisation d'un noyau, permettant d'étaler les impulsions avant l'échantillonnage. Dans les premiers travaux sur le sujet deux noyaux classiques ont été mentionnés [127] : le noyau en sinus cardinal qui est à la base de l'échantillonnage de Shannon (1.14) et le noyau gaussien. Ces deux noyaux s'écrivent :

$$g(t) = \text{sinc}(Bt) = \frac{\sin(\pi Bt)}{\pi Bt} \quad (1.14)$$

$$g(t) = e^{-\frac{t^2}{2\sigma^2}} \quad (1.15)$$

où pour le noyau sinus cardinal de (1.14) on utilise la bande passante  $B = 1/T_s$  si l'intervalle entre deux échantillons est  $T_s$ . Ces noyaux ont tous deux l'inconvénient d'être à support infini, ils ne sont donc pas utilisables en pratique sans troncature, ce qui a pour effet de rendre l'estimation des paramètres de délais et d'amplitude instable notamment en présence de bruit. De plus le noyau gaussien ne remplit pas les conditions (1.11) permettant une reconstruction parfaite, ni les conditions Strang-Fix [112] requises par la méthode de DRAGOTTI, VETTERLI et BLU [56].

Eldar et ses collaborateurs ont proposé un noyau paramétrable et à support fini, appelé noyau de sommes de sinus cardinaux, *Sum of Sincs* en anglais (SoS) [122, 120]. Ce noyau est constitué d'une somme de sinus cardinaux dans le domaine fréquentiel :

$$G(\omega) = \frac{T}{\sqrt{2\pi}} \sum_{\substack{k=-K/2 \\ k \in \mathbb{Z}}}^{K/2} b_k \text{sinc}\left(\frac{\omega T}{2\pi} - k\right) \quad (1.16)$$

où les coefficients  $b_k \neq 0$ . Le paramètre  $K$  détermine la largeur de bande du noyau, c'est à dire le nombre de sinus cardinaux le composant.  $T$  est la durée d'échantillonnage, où pour un signal à échantillonner périodique la longueur de la période. Si l'on utilise par exemple  $N$  échantillons espacés d'une durée d'échantillonnage  $T_s$  on a  $T = NT_s$ . Pour que ce noyau soit à valeurs réelles, on ajoute la contrainte que  $b_k = b_{-k}^*$  où  $b^*$  représente le complexe conjugué de  $b$ , si l'on se restreint à des coefficients réels, le noyau SoS doit donc être symétrique. En passant dans le domaine temporel, on obtient

$$g(t) = \text{rect}\left(\frac{t}{NT_s}\right) \sum_{\substack{k=-K/2 \\ k \in \mathbb{Z}}}^{K/2} b_k e^{j2\pi k t NT_s}. \quad (1.17)$$

où la fonction rectangle est définie par :

$$\text{rect}(t) = \begin{cases} 1 & \text{si } -1/2 \leq t \leq 1/2 \\ 0 & \text{sinon} \end{cases} \quad (1.18)$$

Le sinus cardinal aux points  $\omega = \frac{2\pi k'}{NT_S}$  vérifie les propriétés :

$$\text{sinc}\left(\frac{\omega NT_S}{2\pi} - k\right) = \begin{cases} 1 & k' = k \\ 0 & k' \neq k \end{cases} \quad (1.19)$$

ce qui assure que le noyau SoS vérifie par construction la condition (1.11). Ce noyau en théorie n'est pas dérivable, de par la présence de la fenêtre  $\text{rect}(t)$ . Pour contourner ce problème lorsque nous aurons besoin de la dérivée de  $g(t)$ , nous approximerons la fonction porte par une gaussienne généralisée avec un paramètre de forme  $\beta$  grand [88]. En effet on a la propriété que

$$\lim_{\beta \rightarrow \infty} e^{-\left(\frac{2|t|}{NT_S}\right)^\beta} = \text{rect}\left(\frac{t}{NT_S}\right), \quad (1.20)$$

ce qui permet de considérer la fonction gaussienne généralisée comme une approximation continue de la fonction rectangle. Nous utilisons typiquement la valeur de  $\beta = 100$  dans les différentes simulations.

Le noyau SoS est particulièrement intéressant en pratique, car on peut le moduler pour qu'il prenne différentes formes. Le choix des coefficients peut être déterminé soit pour optimiser un objectif, ce que nous nous proposons de faire dans la section 2.3.2 soit de façon à ressembler à un filtre analogique donné. Un cas particulier qui se révèle utile est de choisir tous les  $\{b_k\}$  égaux à 1. On obtient alors le noyau de Dirichlet :

$$g(t) = \text{rect}\left(\frac{t}{NT_S}\right) \sum_{\substack{k=-K/2 \\ k \in \mathbb{Z}}}^{K/2} e^{j2\pi k t NT_S} = \text{rect}\left(\frac{t}{NT_S}\right) D\left(\frac{2\pi t}{NT_S}\right). \quad (1.21)$$

sachant que le noyau de Dirichlet est par définition [99]

$$D(t) = \sum_{\substack{k=-K/2 \\ k \in \mathbb{Z}}}^{K/2} e^{jkt} = \frac{\sin((K/2 + 1/2)t)}{\sin(t/2)}. \quad (1.22)$$

D'autres noyaux ont été utilisés dans la littérature dans le cadre de l'estimation des signaux paramétriques. DRAGOTTI, VETTERLI et BLU [56] proposent différentes familles de noyaux à support fini, que nous avons mentionné dans la section 1.2.2. Dans cette thèse nous allons cependant nous concentrer sur les trois noyaux sinus cardinaux, Gaussiens et SoS. Les deux premiers types de noyaux sont les noyaux les plus couramment rencontrés en pratique, et sont ceux utilisés dans les premiers articles sur les signaux paramétriques [127], tandis que la famille des noyaux SoS est intéressante en tant que famille plus générale permettant de varier les formes de noyaux possibles tout en maintenant la possibilité d'utiliser une méthode d'estimation basée sur l'estimation spectrale par la condition (1.11). Les noyaux reproduisant les exponentielles sont encore plus généraux, leur étude en terme de performances et d'optimisation est faite par URIGUEN, BLU et DRAGOTTI [123], cependant ils demandent de changer la méthode d'estimation utilisée. Pour avoir une meilleure visualisation des noyaux et pouvoir les comparer nous les traçons tous trois sur la figure 1.3. Pour que la comparaison ait un sens, nous avons imposé une paramétrisation de chaque noyau afin qu'ils aient tous trois la même largeur de bande  $B$ , qui est la largeur de bande du premier lobe du noyau en sinus cardinal. Pour le noyau SoS donné en (1.17) cela revient tout simplement à définir  $K = B$ . La largeur de bande du

noyau gaussien se règle grâce au paramètre  $\sigma$ . En effet la transformée de Fourier de (1.15) est

$$G(f) = \sqrt{2\pi\sigma^2} e^{-2\pi^2 f^2 \sigma^2}. \quad (1.23)$$

Pour obtenir une largeur de bande de  $B$ , on choisit d'avoir une amplitude de -3dB à la fréquence  $f = B/2$ , c'est à dire que

$$\frac{G(B/2)}{G(0)} = e^{-\frac{\pi^2 B^2 \sigma^2}{2}} = \frac{1}{\sqrt{2}} \quad (1.24)$$

$$\Leftrightarrow \sigma = \frac{\sqrt{\ln(2)}}{\pi B} \quad (1.25)$$

Sur la figure 1.3, on a choisi une largeur de bande  $B = \frac{1}{T_s}$ , et un intervalle de temps  $[-3s, 3s]$ . Les noyaux ont été normalisés de telle sorte à ce qu'ils aient la même puissance c'est à dire qu'on a

$$\sum_{n=-(N-1)/2}^{(N-1)/2} g(nT_s)^2 = 1. \quad (1.26)$$

Les  $N$  échantillons conservés après échantillonnage sont marqués sur la figure, les échantillons restant n'interviennent pas dans la reconstruction mais sont intéressants à observer afin de différencier les différents noyaux. Nous avons utilisé deux jeux de paramètres pour le noyau SoS : dans le premier cas, tous les coefficients sont égaux pour obtenir le noyau de Dirichlet, dans le deuxième cas on utilise une fenêtre de Hamming donnée en (1.27) [120]) selon :

$$b_k = 0.54 - 0.46 \cos\left(2\pi \frac{k + K/2}{K}\right) \quad (1.27)$$

en considérant que  $K$  est un multiple de 2. On note tout d'abord que les noyaux Sinc, Gaussien et Dirichlet sont confondus une fois échantillonnés, ils ont un pic en zéro et sont nuls aux autres instants considérés, ce qui est la fonction neutre du point de vue de la convolution si l'on regarde uniquement les instants échantillonnés. Cette propriété est le comportement intuitivement attendu d'un tel noyau, qui filtre un signal sans pour autant le distordre temporellement une fois échantillonné. Cependant, les noyaux continus sont bien différents dans leur forme, en dehors des points d'échantillonnage. On verra dans le chapitre 2 que cette différence a son importance quant aux performances d'estimation, celles-ci étant conditionnées par la fonction dérivée du noyau.

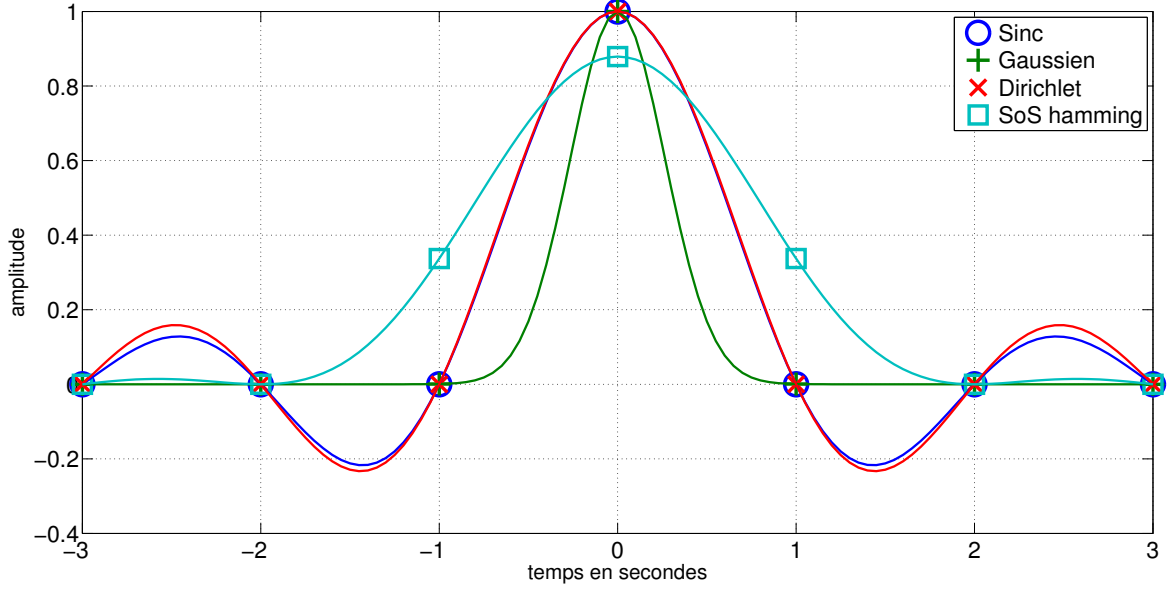


FIGURE 1.3 – Les noyaux utilisés dans le domaine temporel où  $T_S = 10^{-2}$  et  $N = 7$ .

## 1.3 Acquisition comprimée

### 1.3.1 Modèle d'échantillonnage comprimé

Les signaux à modèle paramétrique que nous cherchions à échantillonner dans la section 1.2 sont des signaux à valeurs continues. Nous nous intéressons maintenant à une autre façon d'exploiter la parcimonie d'un signal, cette fois discret, pour en réduire le nombre d'échantillons. Pour ce faire nous allons considérer le signal  $\mathbf{x} \in \mathbb{R}^{K \times 1}$  constitué d'un vecteur de  $K$  échantillons. Ce signal discret peut être représenté comme le produit d'une base orthonormale  $K \times K$   $\Phi = [\phi_1 | \phi_2 | \dots | \phi_K]$  et d'un vecteur parcimonieux  $\mathbf{s}$ , c'est à dire comportant un nombre important de valeurs nulles. On note  $L$  le nombre de valeurs non nulles de  $\mathbf{s}$ . On dit alors que  $\mathbf{x}$  est parcimonieux dans la base  $\Phi$  ce qui formellement s'écrit :

$$\mathbf{x} = \Phi \mathbf{s} = \sum_{k=1}^K s_k \phi_k, \quad (1.28)$$

où  $s_k = \langle \mathbf{x}, \phi_k \rangle$ . Le vecteur  $\mathbf{s}$  est une représentation de  $\mathbf{x}$  obtenue par un simple changement de base.  $\mathbf{x}$  et  $\mathbf{s}$  ont le même degré de parcimonie, qui est défini comme le rapport entre le nombre de valeurs non nulles de  $\mathbf{s}$  et sa dimension :

$$\rho_{\text{par.}} = \frac{L}{K}. \quad (1.29)$$

Ce degré de parcimonie est homogène au taux d'innovation que nous considérons dans la section 1.2. On peut étendre la notion de parcimonie à la notion de compressibilité d'un signal, un signal est dit compressible si le vecteur  $\mathbf{s}$  contient uniquement un faible nombre  $L$  de valeurs significatives pour un grand nombre de valeurs faibles. Un exemple de signal compressible est un signal dont les coefficients, classés en ordre décroissant d'amplitude, sont majorés par une fonction exponentielle. On peut rendre un signal compressible parcimonieux par une simple opération de troncature qui consiste à mettre à zéro les valeurs négligeables. Dans ce cas l'erreur de troncature est bornée par une valeur supérieure [37].

Un grand nombre de signaux rencontrés en traitement du signal peut être exprimé de cette manière. Les images naturelles par exemple sont compressibles dans une représentation de Fourier, ou dans une base d'ondelettes. Cette propriété est mise à profit par les algorithmes de compression d'image tels que JPEG pour une base de Fourier et JPEG2000 utilisant les ondelettes. Ces méthodes classiques de compression échantillonnent les signaux avec un grand nombre d'échantillons, pour tenter de vérifier au mieux le théorème de Shannon puis opèrent une compression dans le but d'en réduire grandement le nombre pour stocker et/ou de transmettre un signal de taille réduite. L'objectif motivant la recherche sur l'acquisition comprimée était initialement de chercher une méthode permettant de n'acquérir que le faible nombre d'échantillons utiles, pour réduire ainsi la complexité et le coût matériel du processus d'échantillonnage [38, 8]. La théorie développée et présentée dans cette section se base sur des signaux discrets, et donc déjà échantillonnés, mais a pour objectif de servir de modèle pour développer des méthodes d'échantillonnage analogiques, comme cela a été fait par exemple dans le projet sur l'appareil photo à un seul pixel par le groupe RICEDSP [102].

La méthode permettant de réduire le nombre d'échantillons de  $\mathbf{x}$  se base sur la multiplication du signal  $\mathbf{x}$  par une matrice de mesure "sur-complète"  $\Psi \in \mathbb{R}^{N \times K}$  où  $N \ll K$  pour obtenir les échantillons  $\mathbf{y}$  :

$$\mathbf{y} = \Psi \mathbf{x} + \mathbf{w} \quad (1.30)$$

Un vecteur  $\mathbf{w}$  contenant le bruit supposé blanc et gaussien a été ajouté car en situation réelle le signal est soumis à du bruit. Cette opération peut être vue comme un processus de mesure faisant le produit scalaire entre le signal  $\mathbf{x}$  et une collection de vecteurs  $\{\Psi_j\}_{j=1}^N$ , ce qui permet de mélanger le grand nombre d'échantillons de  $\mathbf{x}$  dans tous les échantillons de  $\mathbf{y}$ . Ce principe est en quelque sorte similaire à l'opération de pré-filtrage apporté par le noyau  $g(t)$  dans le cas de l'échantillonnage des signaux paramétriques à valeurs continues,  $g(t)$  pouvant être vu comme le pendant analogique de la matrice de mesure. L'échantillonnage classique de Shannon peut être formulé avec le modèle de l'acquisition comprimée en prenant comme matrice de mesure la base canonique de  $\mathbb{R}^K$  :  $\Psi_n(t) = \delta(t - n)$  et une base de Fourier  $\Phi_{k+1}(t) = K^{-1/2} e^{i2\pi kt/K}$ .

Le taux de compression applicable

$$\rho_{\text{mes.}} = \frac{N}{K} \quad (1.31)$$

dépend de propriétés des matrices  $\Psi$  et  $\Phi$  choisies. Dans le cas de l'échantillonnage classique on doit prendre  $N = K$  donné par le théorème de Shannon (voir section 1.3.3 pour un autre exemple). Nous définissons un dernier rapport, permettant de comparer le nombre d'échantillons acquis avec le nombre de valeurs non nulles que l'on cherche à estimer :

$$\rho_{\text{dic.}} = \frac{N}{L} \quad (1.32)$$

ce paramètre entrera en compte dans les performances de reconstruction de  $\mathbf{s}$ .

En combinant les expressions (1.28) et (1.30), on peut exprimer les échantillons en fonction du vecteur parcimonieux :

$$\mathbf{y} = \Psi \Phi \mathbf{s} + \mathbf{w} \triangleq \mathbf{H} \mathbf{s} + \mathbf{w} \quad (1.33)$$

où l'on parlera de  $\mathbf{H} \in \mathbb{R}^{N \times K}$  comme du dictionnaire, que nous considérons réel dans toute cette thèse. Le modèle (1.33) est simplifié, et ne reflète pas toujours parfaitement la réalité. Dans la suite de nos travaux, nous étudions les conséquences des erreurs additives sur le dictionnaire sur les performances de reconstruction.

### 1.3.2 Reconstruction du signal

L'acquisition comprimée a pour objectif de décimer un signal discret en réduisant le nombre d'échantillons nécessaires à sa reconstruction. Cette réduction en nombre d'échantillons est compensée par une étape de reconstruction numérique qui cherche à retrouver le signal d'intérêt à partir des échantillons. Dans la plupart des applications, le vecteur qui nous intéresse en réalité est le vecteur parcimonieux  $\mathbf{s}$  plutôt que  $\mathbf{x}$ . De ce fait le problème de reconstruction une fois un signal échantillonné comme dans (1.33) est de retrouver le vecteur parcimonieux  $\mathbf{s}$  à partir des échantillons  $\mathbf{y}$ . Pour ce faire on considère que la base  $\Phi$  ainsi que la matrice de mesure utilisée  $\Psi$  sont connues. Notons qu'une fois  $\mathbf{s}$  estimé, on peut recouvrer le vecteur  $\mathbf{x}$ . Dans le cas sans bruit, le problème (1.33) est donc simplement un système linéaire à résoudre, sauf que le problème est sous-déterminé puisqu'il y a bien plus d'inconnues que d'équations et n'admet pas une solution unique. Pour choisir une solution, il faut donc ajouter une contrainte, ce qui permet de mettre à profit la propriété de parcimonie de  $\mathbf{s}$ . Finalement l'objectif de la reconstruction est de choisir la solution  $\hat{\mathbf{s}}$  de  $\mathbf{y} = \mathbf{H}\mathbf{s}$  la plus parcimonieuse. Pour ce faire on devrait résoudre le problème d'optimisation suivant [8] :

$$\hat{\mathbf{s}} = \operatorname{argmin} \|\mathbf{s}\|_0 \quad t.q. \quad \mathbf{y} = \mathbf{H}\mathbf{s} \quad (1.34)$$

où la norme  $\ell_0$  est le nombre de valeurs non nulles. Ce problème d'optimisation est numériquement instable et NP-complet, c'est pourquoi la norme  $\ell_0$  est généralement relaxée par la norme  $\ell_1$  qui permet de favoriser également les vecteurs parcimonieux tout en étant convexe. Le problème convexe relaxé avec la norme  $\ell_1$  est équivalent au problème initial considérant la norme  $\ell_0$  si le dictionnaire  $\mathbf{H}$  suit la propriété *Restricted Isometry Property* (RIP) que l'on peut nommer en français "propriété d'isométrie restreinte" [7]. On la définit formellement comme suit.

L'ensemble des vecteurs L-parcimonieux est défini par  $\mathcal{W}_L = \{\mathbf{s} \in \mathbb{R}^K, \|\mathbf{s}\|_0 \leq L\}$ . La matrice  $\mathbf{H}$  vérifie les conditions RIP d'ordre L s'il existe une constante  $\epsilon_L \in (0, 1)$  telle que

$$\|\mathbf{s}\|^2 (1 - \epsilon_L) \leq \|\mathbf{H}\mathbf{s}\|^2 \leq \|\mathbf{s}\|^2 (1 + \epsilon_L) \quad (1.35)$$

pour tout vecteur  $\mathbf{s} \in \mathcal{W}_L$ . Notons que la norme considérée ici est la norme euclidienne  $\ell_2$ . De manière intuitive, la propriété RIP permet d'assurer que le dictionnaire est quasiment orthonormal pour des vecteurs L-parcimonieux. Cette propriété permet d'assurer que les algorithmes d'optimisation utilisant la norme relaxée  $\ell_1$  peuvent retrouver tout vecteur L-parcimonieux à partir d'échantillons bruités. Trouver une matrice  $\mathbf{H}$  suivant la propriété RIP est un problème de type combinatoire, cependant il a été montré que prendre simplement des matrices aléatoires permettait de remplir la condition avec une grande probabilité [7]. Dans le cas général il n'est pas possible de vérifier la propriété pour une matrice quelconque, ce qui a amené à considérer une mesure plus directe : la cohérence. La cohérence est une mesure de l'orthogonalité du dictionnaire, elle est définie formellement par :

$$\mu = \max_{i \neq j} \frac{|\langle \mathbf{H}_i, \mathbf{H}_j \rangle|}{\|\mathbf{H}_i\| \|\mathbf{H}_j\|} \quad (1.36)$$

où  $\mathbf{H}_i$  est la  $i$ -ième colonne du dictionnaire  $\mathbf{H}$ . La cohérence est une mesure intéressante puisqu'elle fournit une borne supérieure sur le nombre de valeurs non nulles de  $\mathbf{s}$  que l'on est capable de retrouver à l'aide d'un dictionnaire donné [57] :

$$L < \frac{1}{2} \left( 1 + \frac{1}{\mu} \right). \quad (1.37)$$

La cohérence est par définition comprise entre 0 et 1, on a donc  $1 \leq L \leq \infty$  suivant les valeurs de  $\mu$ . Toutes les conditions énoncées jusqu'à présent sont des conditions nécessaires mais non suffisantes, un pan de recherche sur le sujet de l'acquisition comprimée est l'étude des conditions de reconstructions dans différents cas de figure, dans le but de réduire encore le nombre d'échantillons nécessaire et d'élargir les matrices utilisables [53]. Une liste de propriétés de parcimonie ainsi que de matrice de mesures pouvant être utilisées se trouve dans le blog de I. CARRON. *Compressive sensing : the Big Picture*. <https://sites.google.com/site/igorcarron2/cs#measurement>.

Les conditions sur les dictionnaires étant énoncées, nous faisons à présent, et pour le reste de cette thèse, l'hypothèse que le dictionnaire  $\mathbf{H}$  respecte les conditions permettant une reconstruction avec la norme  $\ell_1$ . Le problème idéal sans bruit à résoudre est :

$$\hat{\mathbf{s}} = \operatorname{argmin} \|\mathbf{s}\|_1 \quad t.q. \quad \mathbf{y} = \mathbf{H}\mathbf{s}. \quad (1.38)$$

Ce problème, connu sous le nom de LASSO [117] (1.38) peut être résolu par des algorithmes de programmation linéaire [36]. En présence de bruit, le problème consiste à minimiser l'erreur de reconstruction en rajoutant une pénalité sur la norme  $\ell_1$ , on parle alors de Basis Pursuit bruité (BPDN pour *Basis Pursuit Denoising* en anglais) [44] :

$$\hat{\mathbf{s}} = \min_{\mathbf{s}} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \lambda \|\mathbf{s}\|_1 \quad (1.39)$$

De nombreux algorithmes ont depuis été développés pour résoudre le problème de reconstruction  $\ell_1$ . On peut les classer dans quatre grandes catégories :

1. Les algorithmes basés sur des méthodes d'optimisation convexe, on peut notamment utiliser l'interface CVX sur Matlab [68] directement ou des algorithmes spécialisés dans le problème (1.39) tel que le SPLG1 [124] dont on peut trouver une implémentation Matlab [125]. Ces méthodes d'optimisation utilisent directement (1.39). Elles sont donc généralement sensibles à la bonne vérification des conditions RIP.
2. Les algorithmes gloutons (*greedy*) dont l'exemple le plus répandu est l'OMP, présenté par PATI et al. [97] cherchent à chaque itération la colonne de  $\mathbf{H}$  maximisant un critère de corrélation. A chaque itération l'indice correspondant à ce maximum est considéré comme une des valeurs non nulles solution, dont on retranche la contribution au signal avant de passer à la prochaine itération pour trouver le prochain maximum. Ces opérations sont itérées  $L$  fois si l'on connaît l'ordre du modèle ou jusqu'à atteindre un seuil prédéfini. Un algorithme dérivé de l'OMP pour le contexte spécifique de l'acquisition comprimée a été proposé sous le nom de CO-SAMP par NEEDELL et VERSHYNIN [89].
3. Les algorithmes itératifs basés sur un seuillage à chaque itération, un exemple est l'algorithme d'Iterative Soft Thresholding présenté par BECK et TBOULLE [9] qui se base sur un seuillage doux pour obtenir un vecteur parcimonieux en résultat. Les méthodes de passage de messages (*Approximate Message Passing*) comme par exemple DONOHO, MALEKI et MONTANARI [54] entrent également dans cette catégorie.
4. Les algorithmes d'optimisation non convexe : une approche différente consiste à résoudre directement l'équation non convexe avec la norme  $\ell_0$  [110]. Ces méthodes

ont le problème de pouvoir trouver des solutions sous-optimales correspondant à des points stationnaires, mais elles ont le grand avantage de totalement s'affranchir des conditions RIP. Elles peuvent donc être utiles dans des contextes cohérents, tel que c'est le cas pour la plupart des problèmes d'estimation par dictionnaire (voir section 1.3.3).

5. Des algorithmes Bayésiens ont également été proposés, pour résoudre le problème d'estimation parcimonieuse, en se basant sur une connaissance de la distribution du vecteur à estimer ainsi que du bruit [76, 41, 51, 86, 46].

La bibliographie concernant les algorithmes de reconstruction est très vaste, et ne fait pas l'objet de cette thèse, un résumé des méthodes résolvant le problème *Basis Pursuit* (BP) et ses dérivés est fait dans TROPP et WRIGHT [118]. Nous allons principalement utiliser les algorithmes cités que sont SPGL1, OMP et COSAMP.

### 1.3.3 Le problème de l'estimation par dictionnaire

#### Le modèle d'estimation par dictionnaire

Le problème d'estimation paramétrique consiste à retrouver un ensemble de paramètres d'un modèle à partir d'un ensemble d'échantillons bruités  $\{y_n\}_{n=1}^N$ . Pour un modèle d'ordre  $L$ , les échantillons peuvent être exprimés avec le modèle des composantes additives [5] :

$$y_n = \sum_{l=1}^L a_l h(\boldsymbol{\theta}_l, t_n) + w_n. \quad (1.40)$$

La fonction  $h(\boldsymbol{\theta}, t)$  dépend du modèle de signal et donc de l'application qui nous intéresse (radar, estimation de canal...) et est fonction d'un ensemble de paramètres  $\boldsymbol{\theta}$  que l'on cherche à estimer, et du temps  $t$ . Notons par exemple que le modèle de signal à modèle paramétrique donné en (1.1) est un cas particulier du modèle additif (1.40) où le seul paramètre faisant varier la fonction  $h(t)$  est le temps. Comme précédemment, le paramètre  $a_l$  représente l'amplitude et  $w_n$  le bruit additif. Les méthodes traditionnelles d'estimation se basent sur le modèle continu (1.40) pour effectuer l'estimation, elles estiment d'abord l'ordre du modèle  $L$  par des critères d'information du type BIC ou AIC [106] puis se basent par exemple sur l'estimateur du maximum de vraisemblance pour retrouver le reste des paramètres.

Il est également possible de représenter le modèle (1.40) à l'aide d'un dictionnaire, ce qui permet d'utiliser les méthodes d'estimation listées dans la section 1.3.2. Pour ce faire nous allons tout d'abord définir le vecteur de longueur  $N$

$$\mathbf{h}(\boldsymbol{\theta}) = [h(\boldsymbol{\theta}, t_1), \dots, h(\boldsymbol{\theta}, t_N)]^T. \quad (1.41)$$

Définissons

$$\bar{\boldsymbol{\theta}} = \{\bar{\boldsymbol{\theta}}_k\}_{k=1}^K \quad (1.42)$$

un large ensemble comportant  $K$  différents candidats du vecteur de paramètres  $\boldsymbol{\theta}$ , dont le  $k^e$  est noté  $\bar{\boldsymbol{\theta}}_k$ . Le dictionnaire est alors une matrice  $N \times K$  où  $K \gg N$  telle que

$$\mathbf{H}(\bar{\boldsymbol{\theta}}) = [\mathbf{h}(\bar{\boldsymbol{\theta}}_1), \dots, \mathbf{h}(\bar{\boldsymbol{\theta}}_K)], \quad (1.43)$$

et le vecteur d'amplitude parcimonieux est

$$\mathbf{s} = [s_1, \dots, s_K]^T. \quad (1.44)$$

Si l'on fait l'hypothèse que les paramètres composant les échantillons  $\boldsymbol{\theta}_l \in \bar{\boldsymbol{\Theta}}, \quad \forall l$ , on a

$$s_k = \begin{cases} a_l & \text{quand } \bar{\boldsymbol{\theta}}_k = \boldsymbol{\theta}_l \\ 0 & \text{ailleurs} \end{cases}. \quad (1.45)$$

Utilisant cette propriété et ayant défini le dictionnaire en (1.42), on peut exprimer le vecteur des échantillons  $\mathbf{y} = [y_1, \dots, y_N]^T$  par

$$\mathbf{y} = \sum_{l=1}^L a_l \mathbf{h}(\boldsymbol{\theta}_l) + \mathbf{w} = \mathbf{H}(\bar{\boldsymbol{\Theta}}) \mathbf{s} + \mathbf{w} \quad (1.46)$$

Le problème d'estimation revient alors à estimer le vecteur parcimonieux  $\mathbf{s}$  sachant le dictionnaire sur-complet  $\mathbf{H}(\bar{\boldsymbol{\Theta}})$ , ce qui est précisément le problème rencontré dans le cadre de l'acquisition comprimée en (1.33). Les deux problèmes de l'acquisition comprimée et de l'estimation par dictionnaire aboutissent finalement au même problème à résoudre, ce qui explique qu'ils sont souvent confondus dans la littérature et le vocabulaire utilisé est souvent le même (on parle par exemple d'acquisition comprimée appliquée au radar dans le titre de HERMAN et STROHMER [70]). L'estimation par dictionnaire a été proposée pour résoudre une grande variété de problèmes d'estimation classiques tels que le radar [70], l'imagerie *Synthetic Aperture Radar* (SAR) [136], l'estimation de direction d'arrivée BROSSARD et al. [35], BISCH et al. [26] et MALIOUTOV, CETIN et WILLSKY [82], plusieurs de ces applications en traitement d'antennes notamment ont été présentées dans la session spéciale organisée par BOYER et LARZABAL [31]. La parcimonie a également été utilisée pour résoudre les problèmes de détection, par exemple dans PARIS, MARY et FERRARI [95].

Une différence de principe importante existe entre les deux problèmes, c'est la maîtrise du dictionnaire  $\mathbf{H}$ . Dans le cadre de l'acquisition comprimée, l'utilisateur est maître de la chaîne d'acquisition qu'il élabore et peut donc choisir une matrice de mesure adéquate. Il est alors aisé de choisir une matrice aléatoire, ce qui permet d'assurer les conditions RIP et d'éviter ainsi les problèmes de cohérence [7]. D'un autre côté, dans le cadre de l'estimation par dictionnaire,  $\mathbf{H}(\bar{\boldsymbol{\Theta}})$  est connu de l'utilisateur mais il dépend du modèle du problème à estimer. A moins de pouvoir maîtriser le type de signal que l'on envoie comme par exemple dans le contexte du radar [70], le dictionnaire est construit à partir d'une fonction paramétrique dépendant de la physique du problème et étant donc subit par l'utilisateur, et est donc tout sauf aléatoire ce qui explique qu'on a très rapidement une cohérence importante. Concrètement il faudra faire un compromis entre la cohérence du dictionnaire et donc le nombre d'amplitudes non nulles  $L$  que l'on peut estimer et la précision définie par l'espacement dans la grille. Nous donnons un exemple de ce problème dans le cadre de l'estimation spectrale ci-après. L'estimation parcimonieuse se base sur un modèle paramétrique, qui est alors une version simplifiée de la réalité, elle est donc soumise aux erreurs de modèle notamment aux erreurs de grille puisqu'on cherche à estimer des paramètres continus à partir d'un modèle discret. Nous étudions les conséquences des erreurs de modèle pour les deux types de problèmes dans les chapitres 3 et 4.

Les problèmes d'échantillonnages discrets et continus de signaux parcimonieux que nous avons présentés dans les sections 1.2 et 1.3 partagent un même objectif et sont

proches, il est intéressant de noter que les deux champs de recherche de l'échantillonnage de signaux paramétriques et de l'acquisition comprimée initialement distincts ont de nombreux ponts les rapprochant l'un de l'autre. En effet l'acquisition comprimée a initialement été développée dans le cadre des signaux discrets, mais dès le départ un intérêt d'étendre la méthode également aux signaux continus s'est manifesté [59]. De l'autre côté une méthode basée sur la reconstruction parcimonieuse pour estimer les signaux paramétriques a été proposée dans HUANG et al. [72] et par nous-même dans BERNHARDT et al. [19] ou section 2.4.1. De plus OÑATIVIA, LU et DRAGONI [91] a adapté la méthode des filtres annihilants pour l'échantillonnage de signaux discrets.

### Cohérence pour le problème d'estimation spectrale

Dans le problème de l'estimation spectrale on cherche à estimer la fréquence angulaire  $\omega \in \Omega$  à partir de  $N$  échantillons suivant le modèle [111] :

$$y_n = \sum_{l=1}^L a_l e^{j\omega_l n T_S} + w_n \quad (1.47)$$

où  $n \in \{0 \dots N-1\}$  représente les échantillons acquis selon un intervalle d'échantillonnage  $T_S$ ,  $a_l$  et  $\omega_l$  l'amplitude et la fréquence angulaire de la  $l$ e composante fréquentielle et  $w_n$  est le bruit additif qui nous supposons être un bruit blanc gaussien. Pour simplifier les expressions, nous prenons  $T_S = 1$ . Posons maintenant la grille des paramètres  $\bar{\omega} = [\bar{\omega}_1, \dots, \bar{\omega}_K]^T$  avec  $K \gg N$  afin d'exprimer (1.47) sous sa forme parcimonieuse :

$$\mathbf{y} = \mathbf{H}(\bar{\omega})\mathbf{s} + \mathbf{w} \quad (1.48)$$

où  $\mathbf{H}(\bar{\omega}) = [\mathbf{h}(\bar{\omega}_1), \dots, \mathbf{h}(\bar{\omega}_K)]$  avec

$$\mathbf{h}(\bar{\omega}_k) = [1, \dots, e^{j\bar{\omega}_k n}, \dots, e^{j\bar{\omega}_k (N-1)}]^T, \quad (1.49)$$

et  $\mathbf{s} = [s_1, \dots, s_K]^T$  où

$$s_k = \begin{cases} a_l & \text{quand } \bar{\omega}_k = \omega_l \\ 0 & \text{ailleurs.} \end{cases} \quad (1.50)$$

Nous faisons pour le moment l'hypothèse que les paramètres à estimer  $\omega_l$  sont sur la grille  $\bar{\omega}$  et donc qu'il existe pour tout  $l$  un  $k$  tel que  $\bar{\omega}_k = \omega_l$ . Dans le cas contraire (1.48) n'est qu'une approximation. Nous verrons dans la section 3.4 que l'erreur d'estimation induite par cette discrétisation augmente quand l'espacement entre deux éléments de la grille augmente.

Nous prenons à présent la grille la plus courante, et également la plus simple qui consiste à choisir des paramètres espacés régulièrement selon

$$r = \bar{\omega}_{k+1} - \bar{\omega}_k = \frac{\bar{\omega}_K - \bar{\omega}_1}{K-1}. \quad (1.51)$$

La cohérence du dictionnaire  $\mathbf{H}(\bar{\omega})$  donnée en (1.36) peut alors être calculée :

$$\mu(\mathbf{H}(\bar{\omega})) = \max_{1 \leq k_1 \neq k_2 \leq K} \frac{|\langle \mathbf{h}(\bar{\omega}_{k_1})^H, \mathbf{h}(\bar{\omega}_{k_2}) \rangle|}{\|\mathbf{h}(\bar{\omega}_{k_1})\|_2 \|\mathbf{h}(\bar{\omega}_{k_2})\|_2} \quad (1.52)$$

on a  $\|\mathbf{h}(\omega)\|_2 = \sqrt{\sum_{n=0}^{N-1} |e^{j2\omega n}|} = \sqrt{N}$  et étant donné que la grille est uniforme  $\bar{\omega}_{k_2} - \bar{\omega}_{k_1} = mr$  où  $m$  est un nombre entier positif. La cohérence devient :

$$\mu(\mathbf{H}(\bar{\omega})) = \max_{m \in \mathbb{Z} \neq 0} \frac{|\sum_{t=0}^{T-1} e^{jmr t}|}{N} = \max_{m \in \mathbb{Z} \neq 0} \left| \frac{\sin(\frac{Nmr}{2})}{N \sin(\frac{mr}{2})} \right| \quad (1.53)$$

Définissons la fonction  $f(x) = \left| \frac{\sin(\frac{Nx}{2})}{N \sin(\frac{x}{2})} \right|$ . Cette fonction est tracée sur la figure 1.4 pour  $N = 20$  et sur un intervalle où  $x \in [-\pi, \pi]$ . On observe intuitivement que le maximum en dehors de 0 est atteint pour une valeur  $mr$  la plus proche de zéro possible, c'est à dire  $m = 1$ , à condition que cette valeur appartienne au premier lobe de la fonction. Notons que la valeur de cohérence obtenue est la même si l'on considère l'intervalle  $[0, \pi]$  étant donné que la fonction  $f(x)$  est périodique de période  $2\pi$ . La cohérence du dictionnaire associé au problème de l'estimation spectrale est donc égale à  $f(r)$ .

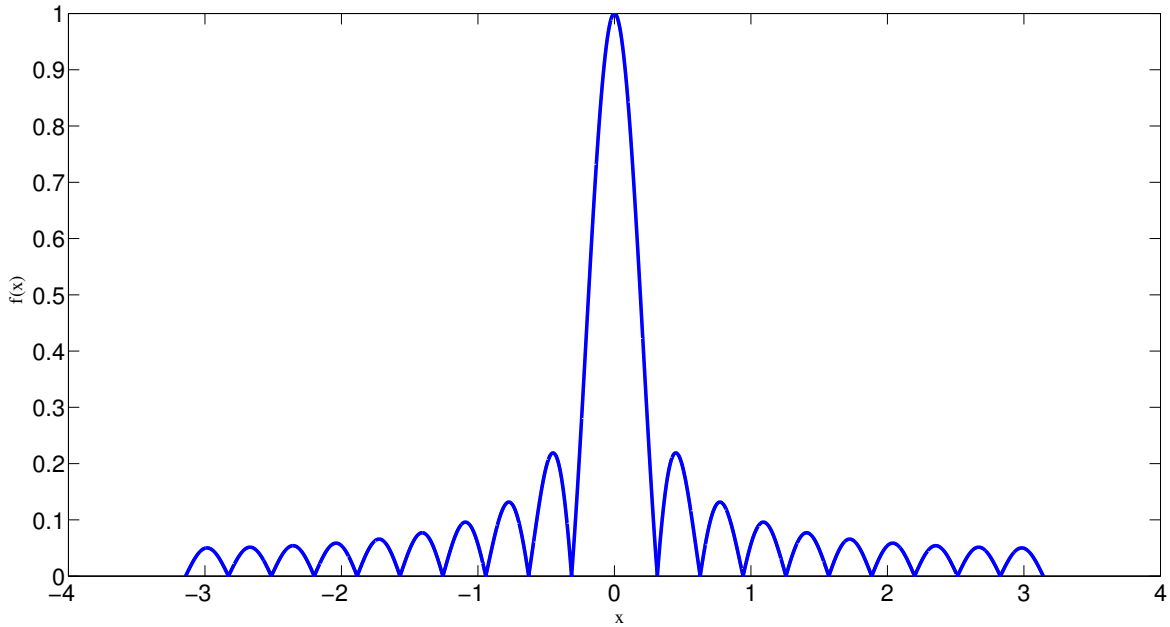


FIGURE 1.4 – Fonction représentant la cohérence  $f(x)$  pour  $N = 20$ .

**Résultat 1.** La cohérence du dictionnaire  $\mathbf{H}(\bar{\omega})$  associée au modèle de l'estimation spectrale, pour  $N \leq K - 1$  est :

$$\mu(\mathbf{H}(\bar{\omega})) = \left| \frac{\sin(\frac{Nr}{2})}{N \sin(\frac{r}{2})} \right| \quad (1.54)$$

ce qui pour une grille allant de 0 à  $\pi$  est égal à :

$$\mu(\mathbf{H}(\bar{\omega})) = \left| \frac{\sin(\frac{N\pi}{K-1})}{N \sin(\frac{\pi}{K-1})} \right| \quad (1.55)$$

*Démonstration.* Pour prouver que le point  $r$  correspond bien au maximum de  $f(x)$  pour  $x = mr, m \in \mathbb{Z}^*$  il suffit de s'assurer que  $f(r) \geq f(\frac{3\pi}{N})$ , le point  $\frac{3\pi}{N}$  correspondant à l'abscisse du maximum du second lobe de  $f(x)$ . On montre tout d'abord que la condition est

vérifiée pour  $r \leq \pi/N$ . On a

$$f\left(\frac{3\pi}{N}\right) = \frac{|\sin(\frac{3\pi}{2})|}{N|\sin(\frac{3\pi}{2N})|} = \frac{1}{N|\sin(\frac{3\pi}{2N})|}$$

tandis que

$$f\left(\frac{\pi}{N}\right) = \frac{|\sin(\frac{\pi}{2})|}{N|\sin(\frac{\pi}{2N})|} = \frac{1}{N|\sin(\frac{\pi}{2N})|}$$

Pour les cas où  $N = 1$  et  $N = 2$  on a  $f(\frac{\pi}{N}) = f(\frac{3\pi}{N})$ . Quand  $N \geq 3$  on utilise le fait que  $|\sin(x)|$  est une fonction strictement croissante sur l'intervalle  $[-\pi/2, \pi/2]$  auquel appartiennent aussi bien  $\frac{\pi}{N}$  que  $\frac{3\pi}{N}$  pour obtenir que  $|\sin(\frac{\pi}{2N})| < |\sin(\frac{3\pi}{2N})|$  et donc par inversion que  $f(\frac{\pi}{N}) > f(\frac{3\pi}{N})$ .  $f(x)$  étant strictement décroissante entre  $x = 0$  et  $x = \frac{\pi}{N}$ , on a démontré que pour  $r \leq \frac{\pi}{N}$ ,  $f(r) \geq f(\frac{\pi}{N}) \geq f(\frac{3\pi}{N})$ .

De par la définition de  $r$  (1.51) on a  $r \leq \frac{\pi}{K-1}$  puisque le paramètre que l'on cherche à estimer est une fréquence angulaire, donc comprise entre  $\bar{\omega}_1 = 0$  et  $\bar{\omega}_K = \pi$ . On a donc que  $r \leq \pi/N$  pour tout  $N \leq K-1$ , cette condition étant par définition vérifiée dans les problèmes d'estimation parcimonieuse, puisqu'elle est la définition d'un dictionnaire sur-complet ayant plus de colonnes que de lignes.  $\square$

On peut à présent utiliser la condition (1.37) pour déterminer le nombre de composantes fréquentielles  $L$  que l'on peut déterminer en l'absence de bruit, étant donné une grille de  $K$  paramètres :

$$\mu(\mathbf{H}(\bar{\boldsymbol{\omega}})) < \frac{1}{2L-1}. \quad (1.56)$$

Pour illustrer le problème, considérons par exemple le cas simple où nous cherchons à retrouver  $L = 2$  fréquences à partir de  $N = 100$  échantillons. Dans ce cas on doit choisir  $K \leq 138$ , ce qui revient à avoir un taux de compression de seulement  $\rho_{\text{mes.}} = 100/139$ . En pratique le nombre d'échantillons  $N$  peut être assez important, ce qui permet de réduire le problème de la cohérence de telle sorte à obtenir de bons résultats pour une bonne précision en fréquence. Nous verrons cependant dans la section 3.4 que la cohérence nous a posé problème dans les simulations, où nous utilisons des  $N$  assez faibles.

### 1.3.4 Évaluation des performances d'estimation

#### Métriques de performances en traitement statistique du signal

Dans le contexte des problèmes d'estimations, des métriques ont été développées pour mesurer les performances théoriques atteignables dans un contexte bruité. Une de ces métriques est couramment utilisée, de par sa simplicité : la borne de Cramer-Rao. Elle permet d'évaluer les performances accessibles par tout un algorithme d'estimation en présence de bruit pour un problème d'estimation donné en donnant une borne inférieure à l'erreur quadratique moyenne de tout estimateur. Un algorithme atteignant la borne est dit efficace, il ne sera alors plus possible d'améliorer ses performances en terme d'erreur quadratique moyenne ( ce qui n'empêche pas de chercher à améliorer d'autres critères comme la complexité ).

Considérons les mesures réelles  $\mathbf{y} = \boldsymbol{\mu} + \mathbf{w}$  où le vecteur  $\boldsymbol{\mu}$  est une fonction dépendant de paramètres  $\boldsymbol{\theta}$  et  $\mathbf{w}$  est le vecteur de bruit additif. La Borne de Cramer Rao (en anglais *Cramer Rao Bound*) (BCR) uniforme est une borne inférieure sur l'Erreur Quadratique

Moyenne (en anglais MSE pour *Mean Square Error*) (EQM) d'un estimateur non biaisé  $\hat{\boldsymbol{\theta}}(\mathbf{y})$  estimant les paramètres inconnus  $\boldsymbol{\theta}$  tels que

$$\text{MSE} = \mathbb{E} \left\{ \left\| \hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta} \right\|^2 \right\} \geq \text{CRB} = \text{Tr}(\mathbf{I}(\boldsymbol{\theta})^{-1}), \quad (1.57)$$

où  $\mathbf{I}(\boldsymbol{\theta})$  est la matrice d'information de Fisher des paramètres  $\boldsymbol{\theta}$  [78] et Tr est la trace d'une matrice. La BCR uniforme est la somme des BCR sur chacun des paramètres à estimer, qui correspondent aux éléments diagonaux de  $\mathbf{I}(\boldsymbol{\theta})^{-1}$ .

La matrice d'information de Fisher déterministe est définie par :

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}} \left\{ -\frac{\partial^2 \log p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right\} \quad (1.58)$$

où  $p(\mathbf{y}|\boldsymbol{\theta})$  est la distribution des mesures conditionnées à un ensemble de paramètres définis.

Dans le cas fréquemment rencontré en pratique d'une distribution gaussienne des observations

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta})), \quad (1.59)$$

on peut utiliser la formule de Slepian-Bang pour des observations réelles par laquelle la matrice d'information de Fisher devient :

$$\mathbf{I}_{ij}(\boldsymbol{\theta}) = \left[ \frac{\partial \boldsymbol{\mu}}{\partial \theta_i} \right]^T \mathbf{C}^{-1}(\boldsymbol{\theta}) \left[ \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \right] + \frac{1}{2} \text{Tr} \left[ \mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_j} \right], \quad (1.60)$$

où

$$\left[ \frac{\partial \boldsymbol{\mu}}{\partial \theta_i} \right] = \begin{bmatrix} \frac{\partial [\boldsymbol{\mu}]_0}{\partial \theta_i} \\ \vdots \\ \frac{\partial [\boldsymbol{\mu}]_{(N-1)}}{\partial \theta_i} \end{bmatrix}.$$

Un cas particulier très utilisé en pratique est le cas d'un bruit additif gaussien de moyenne nulle et de variance  $\sigma^2$ . Dans ce contexte on a  $p(\mathbf{y}|\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2)$  où l'on obtient :

$$\mathbf{I}_{ij} = \frac{1}{\sigma^2} \left[ \frac{\partial \boldsymbol{\mu}}{\partial \theta_i} \right]^T \left[ \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \right]. \quad (1.61)$$

La borne de Cramer Rao donnée ci-dessus concerne uniquement les estimateurs non biaisés, pour un estimateur biaisé dont le biais est  $\mathbf{b}(\boldsymbol{\theta})$  une généralisation de la borne est BEN-HAIM et ELDAR [11] :

$$\text{CRB}(\boldsymbol{\theta}, \mathbf{b}(\boldsymbol{\theta})) \triangleq \text{Tr}[(\mathbf{I} + \mathbf{D}(\boldsymbol{\theta})) \mathbf{I}(\boldsymbol{\theta})^{-1} (\mathbf{I} + \mathbf{D}(\boldsymbol{\theta}))^T] + \|\mathbf{b}(\boldsymbol{\theta})\|^2, \quad (1.62)$$

où  $\mathbf{D}(\boldsymbol{\theta}) = \frac{\partial \mathbf{b}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$  est la matrice dérivée du biais, et  $\mathbf{I}(\boldsymbol{\theta})$  est la matrice d'information de Fisher donnée en (1.58), la matrice  $\mathbf{I}$  étant l'identité.

Dans le contexte Bayésien, nous n'avons plus un vecteur de paramètre déterminé pour lequel on calcule la borne, mais on considère qu'on connaît uniquement la distribution des paramètres à estimer. L'extension de la borne déterministe au cas bayésien peut être faite de deux façons : soit par la borne Bayésienne de Van Trees, notée Borne de Cramer Rao Bayésienne (BCRB) [126] soit en se basant sur la connexion entre le contexte déterministe et bayésien, ce qui permet d'obtenir la borne Borne de Cramer Rao Moyennée, en anglais ECRB pour *Expected Cramer Rao Bound* (BCRM) [11].

Dans [126], Van Trees propose une généralisation de la BCR dans le contexte Bayésien. Notons  $\hat{\boldsymbol{\theta}}(\mathbf{y})$  un estimateur Bayésien. L'erreur d'estimation peut être bornée par :

$$\mathbb{E} \left\{ \left\| \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{y}) \right\|^2 \right\} \geq \text{Tr}(\mathbf{J}^{-1}), \quad (1.63)$$

où  $\mathbf{J}$  est la Matrice d'Information de Fisher Bayésienne (MIB) définie par :

$$\mathbf{J} = \mathbb{E}_{\mathbf{y}, \boldsymbol{\theta}} \left\{ -\frac{\partial^2 \log p(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right\} \quad (1.64)$$

$$= \mathbb{E}_{\mathbf{y}, \boldsymbol{\theta}} \left\{ -\frac{\partial^2 \log p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right\} + \mathbb{E}_{\boldsymbol{\theta}} \left\{ -\frac{\partial^2 \log p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right\} \quad (1.65)$$

$$= \mathbb{E}_{\boldsymbol{\theta}} \left\{ \mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}} \left\{ -\frac{\partial^2 \log p(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right\} \right\} + \mathbb{E}_{\boldsymbol{\theta}} \left\{ -\frac{\partial^2 \log p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right\} \quad (1.66)$$

$$= \mathbb{E}_{\boldsymbol{\theta}} \{\mathbf{I}\} + \mathbb{E}_{\boldsymbol{\theta}} \left\{ -\frac{\partial^2 \log p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right\}. \quad (1.67)$$

On remarque que la MIB peut être décomposée comme la somme de la moyenne de la Matrice d'Information de Fisher déterministe (FIM en anglais) (MIF) sur la distribution des paramètres et une composante dépendant uniquement de la distribution des paramètres aléatoires à estimer.

Une autre borne Bayésienne minimisant la BEQM est la BCRM qui est simplement la CRB déterministe moyennée sur le paramètre aléatoire  $\boldsymbol{\theta}$  [11]

$$\mathbb{E}_{\mathbf{y}, \boldsymbol{\theta}} \left\{ \left( \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{y}) \right)^2 \right\} \geq \mathbb{E}_{\boldsymbol{\theta}} \left\{ \text{Tr}(\mathbf{I}(\boldsymbol{\theta})^{-1}) \right\}. \quad (1.68)$$

On a la propriété que  $\text{BCRB} \leq \text{BCRM} \leq \text{BEQM}$  c'est à dire que la BCRM est une borne plus haute, et donc plus serrée (voir par exemple [67] qui montre cette inégalité). Les deux bornes sont égales si la MIF  $\mathbf{I}(\boldsymbol{\theta})$  n'est pas une fonction de  $\boldsymbol{\theta}$  ce qui est le cas par exemple pour un système linéaire avec un dictionnaire déterministe. Dans le cas contraire,  $\text{BCRB} < \text{BCRM}$  ce qui signifie que la BCRM est plus proche du comportement des estimateurs, la BCRB étant trop optimiste.

D'autres bornes minimisant la Erreur Quadratique Moyenne Bayésienne (en anglais BMSE pour *Bayesian Mean Square Error*) (BEQM) ont également été développées, permettant d'apporter des informations complémentaires à la borne de Cramer-Rao.

### Etude des performances dans le contexte de l'acquisition comprimée

Les expressions de ces bornes sont intéressantes pour tout problème d'estimation, afin de pouvoir comparer les différents algorithmes proposés dans la littérature. L'acquisition comprimée (ou l'estimation parcimonieuse) a comme nous l'avons vu, généré un intérêt important pour le développement d'estimateurs parcimonieux. Les différentes bornes minimales de la BEQM dans ce contexte particulier doivent donc également être étudiées.

La première borne à avoir été étudiée est la borne oracle, c'est à dire connaissant le support du vecteur parcimonieux [6]. Connaissant le support, le problème d'estimation parcimonieuse devient un problème classique de résolution d'une équation linéaire, et la borne sur l'estimation des amplitudes des éléments non nuls peut être dérivée. BABADI,

KALOUPSIDIS et TAROKH [6] dérivent la borne oracle, et montrent que pour un dictionnaire construit à partir d'échantillons tirés aléatoirement d'une distribution gaussienne on peut trouver un estimateur atteignant cette borne même sans connaître le support quand le nombre de mesures  $N$  tend vers l'infini. BOYER et al. [33] apportent une correction à cet article en introduisant le contexte doublement asymptotique où à la fois  $N$  et la taille du support  $L$  tendent vers l'infini. NIAZADEH, BABAIE-ZADEH et JUTTEN [90], quant à eux, généralisent ce résultat à des matrices déterministes et non gaussiennes.

Une étude de la borne parcimonieuse non oracle est faite par BEN-HAIM et ELDAR [12] qui utilisent une borne de Cramer Rao Contrainte (CCRB) en introduisant la contrainte sur la parcimonie du signal dans le calcul de la borne. Les auteurs montrent que quand on connaît le degré de parcimonie  $L$  du vecteur à estimer la borne contrainte est égale à la borne oracle. Cette étude justifie donc l'utilisation de l'hypothèse de support connu pour dériver les bornes sur le problème de l'estimation parcimonieuse, puisque ces bornes sont plus simples et équivalentes à la borne CCRB pour un  $L$  donné. Dans notre travail nous nous concentrons sur le contexte asymptotique, pour lequel l'utilisation de la borne oracle est pertinente, nous ferons donc l'hypothèse que le support est connu. Une utilisation intéressante de la borne de Cramer-Rao est proposée dans HUANG et al. [73], qui proposent d'optimiser itérativement la matrice de mesure de telle sorte à minimiser la borne de Cramer-Rao sur le support obtenu, et ainsi améliorer les performances d'estimation de l'amplitude. Nous utiliserons de la même manière les bornes comme objectifs pour optimiser les noyaux dans le contexte de l'échantillonnage des signaux parcimonieux, dans le chapitre 2.

Plutôt que de dériver directement la borne avec le modèle de mesures parcimonieux (1.33), PAKROOH et al. [94] étudient l'influence de la compression des mesures sur la MIF classique pour le problème d'estimation par dictionnaire. Dans le modèle que les auteurs considèrent la compression opère après l'ajout du bruit additif, ce qui a pour conséquence que le bruit est également multiplié par la matrice de compression  $\mathbf{H}$ . SHAGHAGHI et VOROBYOV [107] étendent ces résultats à l'estimation de paramètres complexes suivant un modèle à rang faible, et pouvant utiliser de multiples vues plutôt qu'un seul vecteur de mesure  $\mathbf{y}$  pris à un temps  $t$ .

Les bornes les plus couramment utilisées, de par leur forme simple sont les bornes de Cramer Rao et leurs dérivées bayésiennes ou biaisées. Ces bornes prédisent le comportement des estimateurs dans le cas asymptotique, c'est à dire pour un rapport signal sur bruit élevé. Quand le bruit augmente, les estimateurs décrochent de la borne à partir d'un certain niveau de bruit ou seuil, qu'il est intéressant de pouvoir prévoir. D'autres bornes sont plus précises dans les régions non asymptotiques, et peuvent être utilisées pour prédire ce seuil, c'est notamment le cas des bornes de Barankin [43], ou des bornes hybrides proposées par REN et al. [101]. JUNG et al. [77] étudient la borne de Barankin pour le modèle simplifié de reconstruction d'un vecteur parcimonieux bruité, sans multiplication par une matrice de mesure. Les auteurs donnent une borne supérieure et inférieure de la borne de Barankin qui ne peut pas être formulée analytiquement. L'effet de seuil et le décrochage est également étudié par PAKROOH, SCHARF et PEZESHKI [93] qui analysent le décrochage en tant que passage des mesures bruitées du sous-espace signal au sous-espace bruit quand le bruit augmente.

La théorie des matrices aléatoires traite des grandes matrices aléatoires, c'est à dire dont les éléments sont tirés d'une distribution aléatoire. Ces matrices ont été étudiées mathématiquement depuis des décennies [133, 132, 83], mais avec l'augmentation des

dimensions des problèmes concrets actuellement, cette théorie a trouvé ces dernières années de nombreuses utilisations pratiques en communication sans fil et traitement du signal [49, 119] notamment pour le calcul des bornes de Cramer Rao [92]. Cette théorie est particulièrement bien adaptée au contexte de l'acquisition comprimée, puisque les matrices aléatoires sont souvent utilisées comme dictionnaire, pour garantir les conditions RIP. Son utilisation permet de dériver des formes analytiques pour les bornes Bayésiennes du problème de l'acquisition comprimée [34]. Les bornes Bayésiennes se placent dans le contexte où le vecteur parcimonieux à estimer est aléatoire. BOYER et al. [34] et BOYER, LARZABAL et FLEURY [32] s'intéressent aux bornes Bayésiennes oracles minimisant la BEQM, et montrent que des estimateurs connus tel que l'OMP ou le *Linear Minimum Mean Square Estimator* ou en français Estimateur de l'erreur quadratique moyenne minimum linéaire (LMMSE) atteignent cette borne à haut Rapport Signal sur Bruit (RSB). EL KORSO et al. [58] étudient les performances pour un modèle Bayésien hiérarchique.

### 1.3.5 Le problème des erreurs de modèles

Dans les applications pratiques, notamment en estimation par dictionnaire, le modèle parcimonieux (1.46) n'est qu'une approximation de la réalité, puisque les paramètres continus n'appartiennent pas dans le cas général à la grille de discrétisation choisie. Le problème de l'acquisition comprimée également peut contenir des erreurs de modèles, dans le cas où la matrice  $\mathbf{H}$  n'est pas parfaitement connue. Nous présentons dans la section 3.1 plus en détail les différents types d'erreurs qui peuvent apparaître et fausser l'estimation du vecteur parcimonieux  $\mathbf{s}$ .

Les conséquences des erreurs de dictionnaires ont été mises en évidence dès 2010 par HERMAN et STROHMER [71], CHI et al. [45] et CHAE, SADEGHI et KENNEDY [42]. CHI et al. [45] se concentrent sur le problème de l'erreur de grille pour le cas particulier d'une base de Fourier, utile dans les applications d'imagerie, et montrent que dans cette situation les performances sont considérablement dégradées. CHAE, SADEGHI et KENNEDY [42] illustrent également les conséquences d'une telle erreur dans le contexte de la reconstruction de signaux sinusoïdaux. HERMAN et STROHMER [71] étudient les conséquences d'une perturbation additive sur le dictionnaire du type

$$\tilde{\mathbf{H}} = \mathbf{H} + \mathbf{E}. \quad (1.69)$$

Les auteurs montrent l'influence de cette erreur sur les conditions RIP et sur les performances d'algorithmes de type BP et montrent notamment que la reconstruction est d'une certaine manière robuste aux perturbations puisque l'erreur observée sur la reconstruction augmente linéairement avec le niveau de perturbation. D'autres types de méconnaissance du dictionnaire ont également été étudiées : EMIYA et al. [61] étudient les perturbations dues à une méconnaissance du processus de détection dans le contexte de l'imagerie, notamment quand on ne connaît pas l'ordre des capteurs, qui peut être modélisé comme une permutation inconnue de la matrice de mesure. ROSENBAUM et TSYBAKOV [103] considèrent le cas où l'estimateur parcimonieux utilise une version bruitée du dictionnaire. Les auteurs montrent que la perturbation causée par ce type d'erreur sur le LASSO est importante, et proposent une version robuste du critère à minimiser. Le cas le plus général de méconnaissance du dictionnaire est le contexte de l'apprentissage de dictionnaire, qui considère qu'on ne connaît pas du tout la matrice de mesure. On parle alors d'acquisition comprimée aveugle [1, 65].

La première méthode envisagée pour réduire le problème de l'erreur de grille dans le cadre de l'estimation par dictionnaire est de recentrer la grille de manière itérative sur les résultats (*grid refinement*) [105, 104, 82], ces méthodes se basant sur une première estimation du signal ne prenant la présence de l'erreur de modèle en compte. A contrario, de nouveaux algorithmes d'estimation parcimonieuse ont été proposés prenant en compte les perturbations sur le dictionnaire, pour en atténuer les effets. Ces algorithmes peuvent être dérivés en faisant des hypothèses sur l'erreur  $\mathbf{E}$ . Deux hypothèses sont fréquemment rencontrées en pratique :

1. Le modèle *Error in Variables* ou en français erreur dans les variables (EIV) où  $\mathbf{E}$  est une matrice aléatoire dont les éléments sont tirés d'une distribution gaussienne d'une variance donnée  $\sigma_e^2$ , par exemple dans ZHU, LEUS et GIANNAKIS [135], que nous étudions dans le chapitre 3.
2. le modèle structuré présenté dans la section 4.1.1 qui est le plus souvent rencontré en pratique puisqu'il modélise les erreurs dues à la discrétisation dans les applications d'estimation par dictionnaire [114, 134] et que nous étudions dans le chapitre 4.

La plupart des méthodes d'estimation prenant en compte l'erreur cherche à améliorer un algorithme de reconstruction existant pour le rendre robuste aux perturbations. Citons par exemple TEKE, GURBUZ et ARIKAN [116] qui proposent une amélioration de l'OMP ou PARKER, CEVHER et SCHNITER [96] qui se concentrent sur les méthodes à seuillage doux. De notre côté, pour les algorithmes proposés dans le chapitre 4 nous nous plaçons dans le contexte où nous avons un processus d'acquisition comprimée classique, que nous ne voulons pas modifier et à partir de l'estimation parcimonieuse obtenue nous corrigeons l'erreur due à l'erreur de modèle en nous basant sur des projections oblique. Les algorithmes obtenus ne procèdent donc qu'à une seule reconstruction parcimonieuse, suivi d'une correction, contrairement aux méthodes de *grid refinement*.

Tout comme pour le problème d'acquisition comprimée idéal, le modèle parcimonieux perturbé permet de dériver des bornes minimales pour les estimateurs parcimonieux soumis à une erreur de dictionnaire. KRZAKALA, MÉZARD et ZDEBOROVÁ [80] étudient une borne inférieure de l'erreur quadratique moyenne de l'estimateur Bayésien optimal différente des bornes de Cramer-Rao, basée sur une méthode de répliques (*replica method*). WIESEL, ELDAR et YEREDOR [131] étudient la borne de Cramer Rao déterministe et les performances de l'estimateur du maximum de vraisemblance pour deux modèles de perturbations différents, ces deux modèles pouvant être reliés aux modèles que nous étudions dans le chapitre 3 et que nous étendons au contexte Bayésien. TANG, CHEN et GU [115] étudient la borne de Cramer Rao Contrainte pour le modèle EIV ainsi que la borne de Hammerlsey qui est plus haute dans le cas particulier d'une matrice de mesure égale à la matrice identité. RAMASAMY, VENKATESWARAN et MADHOW [100] utilisent les bornes de Cramer-Rao déterministe et bayésiennes et de Ziv-Zakai pour étudier le problème de l'acquisition comprimée appliquée à l'estimation de paramètres continus. Ils utilisent ces bornes pour définir le nombre de mesures nécessaire à l'estimation d'un paramètre continu par des méthodes d'acquisition comprimées avec un dictionnaire vérifiant des propriétés d'isométrie bien précises. ALDROUBI, CHEN et POWELL [2] proposent également de nouvelles conditions sur la matrice de mesures, pour rendre le problème de l'acquisition comprimée plus robuste au bruit multiplicatif. Peu d'auteurs ont étudié le problème de l'erreur dans le contexte Bayésien, JAGANNATH et HARI [74] s'intéressent à

la borne Bayésienne mais uniquement dans le contexte bien précis de l'estimation parcimonieuse des directions d'arrivées avec une seule vue.

## 1.4 Présentation des contributions de la thèse

### 1.4.1 Présentation de la thèse

Dans ce premier chapitre, nous avons introduit le contexte de l'échantillonnage des signaux parcimonieux se basant sur deux façons différentes de représenter la parcimonie. Dans le chapitre 2 nous nous intéresserons plus particulièrement aux signaux impulsionnels (ou à modèle paramétrique). Nous dérivons les performances dans différents contextes dans le but d'en tirer le choix d'un noyau optimal. Dans le chapitre 3 nous nous focalisons sur l'étude théorique de l'influence de l'erreur de modèle dans le contexte de l'estimation par dictionnaire et de l'acquisition comprimée, pour établir la robustesse de l'estimation à ces erreurs à haut rapport signal sur bruit. Dans le chapitre 4, nous étudions le cas particulier de l'erreur structurée et proposons des algorithmes permettant de corriger l'écart de performance induit par les erreurs de modèles. Enfin le chapitre 5 résume nos conclusions et apporte des perspectives pour la suite de notre travail.

### 1.4.2 Contributions sur le sujet des signaux paramétriques

Nous présentons ici les différentes contributions que nous apportons dans cette thèse, pour en donner une vision synthétique.

- Section 2.2.1 : nous donnons l'expression analytique de la borne de Cramer Rao déterministe, pour un signal à somme de Dirac et des noyaux quelconques. Cette borne prend en compte le cas à plusieurs impulsions. Nous donnons une approximation très simple de cette borne, valable pour les noyaux les plus courants, permettant de comparer facilement les résultats des noyaux en terme de performance.
- Section 2.2.2 : nous nous plaçons dans le contexte Bayésien et donnons la borne Bayésienne de Van Trees pour un signal à somme de Dirac et des noyaux quelconques et nous utilisons cette borne pour établir le noyau SoS optimal sur le paramètre de retard.
- Section 2.3.1 : nous dérivons la borne Bayésienne de Van Trees pour un signal quelconque et l'utilisation des noyaux SoS.
- Section 2.3.2 : nous trouvons le noyau optimal optimisant la borne Bayésienne à la fois pour les retards et les amplitudes, pour différentes contraintes de normalisation.
- Section 2.4 : nous proposons un modèle permettant d'utiliser l'acquisition comprimée pour estimer les signaux paramétriques composés d'une somme d'impulsions de Dirac et pouvant être généralisé à tout type de signaux paramétriques.

### 1.4.3 Contributions sur le sujet des erreurs de modèles dans l'acquisition comprimée

- Section 3.1 : nous présentons deux modèles d'erreur différents, apparaissant l'un ou l'autre à des étapes différentes du processus d'échantillonnage comprimé.
- Section 3.2 : nous donnons la borne Bayésienne de Van Trees pour ces deux types d'erreur sur l'estimation des amplitudes.
- Section 3.3 : nous donnons la Borne de Cramer Rao Moyennée pour le problème spécifique de l'erreur de grille, dans le cas où l'erreur est faible et comparons cette borne à la borne de Van Trees.
- Section 3.4 : nous donnons dans le cas à haut rapport signal sur bruit où le support du vecteur parcimonieux est correctement estimé l'erreur quadratique moyenne dans le contexte Bayésien sur l'estimation des retards de l'estimation par dictionnaire et nous montrons par simulation que cette borne est atteinte par des algorithmes d'estimation.
- Section 3.4.4 et 1.3.3 : nous montrons les conséquences de l'erreur de grille dans le cas particulier de l'estimation de fréquence. Dans ce contexte applicatif nous donnons aussi la cohérence du dictionnaire et la contrainte que celle-ci représente sur l'estimation.
- Section 4.1.1 : nous nous intéressons également au problème de l'erreur de grille structurée et donnons les bornes de Van Trees sur l'estimation des retards et des amplitudes dans ce contexte.
- Section 4.1 : nous proposons l'algorithme OGEC (Off-Grid Error Correction) comme étape post-traitement pour estimer l'erreur de support et réestimer les amplitudes en présence d'erreur de grille. Nous étudions les performances théoriques de cet algorithme sur l'estimation des retards, et montrons ses résultats par simulation dans le contexte de l'estimation par dictionnaire des signaux paramétriques.
- Section 4.2 : nous proposons également l'algorithme BiCE (Bias Correction Estimator) pour corriger l'erreur sur l'estimation de l'amplitude en présence d'une erreur de bruit additif sur le dictionnaire. Cette correction se base uniquement sur l'estimation du support et peut donc être utilisée en post-traitement de n'importe quel algorithme d'estimation parcimonieuse sans que cela n'influe les performances. Nous dérivons les performances théoriques de l'algorithme, étudions son efficacité et montrons les résultats pratiques dans le cadre de l'acquisition comprimée des signaux paramétriques.

### 1.4.4 Publications

#### Articles dans des revues à comité de lecture

- S. BERNHARDT, R. BOYER, S. MARCOS et P. LARZABAL. « Compressed Sensing with Basis Mismatch : Performance Bounds and Sparse-Based Estimator ». Dans : *IEEE Transactions on Signal Processing* 64.13 (juil. 2016), p. 3483–3494

### Actes de colloques internationaux

- S. BERNHARDT, R. BOYER, S. MARCOS et P. LARZABAL. « Sparse-based estimators improvement in case of Basis mismatch ». Dans : *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*. Déc. 2015, p. 417–420
- S. BERNHARDT, R. BOYER, S. MARCOS et P. LARZABAL. « Compressed Sensing with uncertainty - the Bayesian estimation perspective ». Dans : *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*. Déc. 2015, p. 421–424
- S. BERNHARDT, R. BOYER, S. MARCOS, Y. C. ELDAR et P. LARZABAL. « Sampling FRI signals with the SOS kernel : Bounds and optimal kernel ». Dans : *Signal Processing Conference (EUSIPCO), 2015 23rd European*. Août 2015, p. 2172–2176
- S. BERNHARDT, R. BOYER, S. MARCOS, Y. C. ELDAR et P. LARZABAL. « Cramer-Rao Bound for finite streams of pulses ». Dans : *2014 22nd European Signal Processing Conference (EUSIPCO)*. Sept. 2014, p. 984–988
- S. BERNHARDT, R. BOYER, B. ZHANG, S. MARCOS et P. LARZABAL. « Performance analysis for sparse based biased estimator : Application to line spectra analysis ». Dans : *2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. Invited paper. Juin 2014, p. 365–368

### Actes de colloques nationaux

- S. BERNHARDT, R. BOYER, S. MARCOS, Y. ELDAR et P. LARZABAL. « Nouveaux noyaux pour l'échantillonnage de signaux impulsionnels à bande non-limitée ». Dans : *25ème édition du colloque GRETSI*. 2015

### Communications scientifiques dans des colloques et journées d'étude

- S. BERNHARDT et R. BOYER. « Caractérisation théorique du biais d'estimation du à l'erreur de grille dans le cadre de l'acquisition comprimée ». Dans : *Journée GDR" Représentations parcimonieuses, échantillonnage compressé et imagerie médicale"*. 2014
- Estimation parcimonieuse en présence d'erreur de modèle, Journée des doctorants du laboratoire des signaux et Systèmes (LSS)



# Chapitre 2

## Optimisation du noyau d'échantillonnage pour des signaux impulsionnels

« habababapapapaa »

Leif Gourdin

### Sommaire

---

|            |                                                                                         |           |
|------------|-----------------------------------------------------------------------------------------|-----------|
| <b>2.1</b> | <b>Comment étudier les performances de la reconstruction des signaux impulsionnels?</b> | <b>32</b> |
| <b>2.2</b> | <b>Performance de la reconstruction d'un signal à somme de Dirac</b>                    | <b>33</b> |
| 2.2.1      | Borne déterministe                                                                      | 34        |
| 2.2.2      | Estimation Bayésienne du signal à somme de Diracs                                       | 50        |
| <b>2.3</b> | <b>Cas du noyau à sommes de sinus cardinaux</b>                                         | <b>57</b> |
| 2.3.1      | Borne bayésienne pour un signal à série d'impulsions de forme quelconque                | 57        |
| 2.3.2      | Choix du noyau optimal au sens de l'information                                         | 63        |
| <b>2.4</b> | <b>Estimation par dictionnaire pour les signaux impulsionnels</b>                       | <b>76</b> |
| 2.4.1      | Estimation des signaux à somme de Dirac en discrétisant un dictionnaire                 | 76        |
| 2.4.2      | Acquisition comprimée pour la reconstruction des signaux impulsionnels                  | 77        |

---

## 2.1 Comment étudier les performances de la reconstruction des signaux impulsionnels ?

Nous avons vu dans la section 1.2 qu'il est possible d'échantillonner les signaux impulsionnels en utilisant un nombre très faible d'échantillons, largement inférieur à ce que préconise le théorème de Shannon. Pour ce faire l'étape d'échantillonnage doit être suivie d'une estimation des paramètres de modèle du signal. Une fois ces paramètres estimés, le signal est parfaitement connu et peut donc être reconstruit. Dans ce schéma, le problème d'échantillonnage initial est transformé en un problème d'estimation de paramètres continus. En présence de bruit, l'estimation ne peut être parfaite et une erreur est introduite. Développer des algorithmes d'estimation robuste au bruit nécessite donc de pouvoir évaluer les performances d'estimation dans le cas non idéal, puisque développer des algorithmes sans se préoccuper de l'effet du bruit peut aboutir à des algorithmes non robustes au bruit, tel que l'algorithme idéal des filtres annihilants de BLU et al. [27]. Une métrique utilisée pour évaluer les performances des algorithmes est la BCR (voir section 1.3.4).

Dès les premiers articles concernant l'échantillonnage de signaux impulsionnels, la BCR a été utilisée pour comparer les résultats des algorithmes d'estimation proposés. WEI, BLU et DRAGOTTI [129] et DOĞAN et al. [50] comparent leurs résultats avec la borne, qu'ils calculent numériquement sans en donner une expression analytique. BLU et al. [27] donnent la formule générale de la borne pour le modèle (1.1) avec un bruit discret  $w_n$  blanc et gaussien. Cependant une expression analytique de la borne n'est donnée que dans le cas d'une unique impulsion de Dirac échantillonnée avec un noyau en sinus cardinal. DRAGOTTI et HOMANN [55] donnent une expression analytique de cette borne également pour les noyaux E-spline et B-spline introduits dans DRAGOTTI, VETTERLI et BLU [56]. [122] résumant ces contributions et ajoutent une expression analytique pour le noyau SoS. Toutes ces bornes sont pour des paramètres déterministes et ne considèrent qu'une impulsion par période. Dans cette thèse, nous étendons cette borne au cas de plusieurs impulsions et pour une forme de noyau quelconque dans la section 2.2.1 puis nous nous intéresserons au cas Bayésien dans la section 2.2.2. En parallèle de ce travail sur la borne déterministe généralisée à plusieurs impulsions [21], ANGIERSKI et KUEHN [4] ont également élargi ces bornes dans le cas de plusieurs impulsions, en se basant sur la représentation spectrale des échantillons donnée en (1.13) pour établir les dérivées, cette méthode limite la borne aux noyaux suivant la condition de l'équation (1.11) notamment le noyau SoS. De notre côté nous avons dérivé la borne à partir de la représentation temporelle des échantillons.

Tous ces auteurs se sont placés dans le contexte présenté dans la section 1.2 où le bruit est introduit au moment ou après l'échantillonnage (voir figure 1.2). Ce bruit numérique correspond au bruit de quantification et aux bruits électroniques de la chaîne de traitement. D'autres auteurs se sont intéressés quant à eux au bruit analogique apparaissant avant l'échantillonnage sous la forme  $y(t) = x(t) + w(t)$  où  $w(t)$  est le bruit blanc gaussien analogique. BEN-HAIM, MICHAELI et ELDAR [13] développent la borne continue correspondante sur la reconstruction du signal  $x(t)$  à partir des échantillons. Cette borne permet d'avoir une limite sur les performances atteignables quelle que soit la méthode d'échantillonnage utilisée, et donc quel que soit le modèle choisi pour  $x(t)$  et quel que soit le noyau. Cette borne est une borne minimale intéressante, mais elle peut largement sur-estimer les performances atteignables par un estimateur dans un contexte plus

précis. COMSA et HAIMOVICH [47] proposent une borne pour une seule impulsion, considérant un bruit analogique et une famille de noyaux de pré-filtrage donnée. BEN-HAIM, MICHAELI et ELDAR [13] développent également une expression générale de la borne discrète prenant en compte le bruit analogique et le bruit numérique mais considèrent le problème de l'estimation du signal  $x(t)$ , pas celui de l'estimation des paramètres  $\theta$ . Cette borne permet de montrer qu'en présence uniquement de bruit numérique, augmenter le nombre d'échantillons au-delà de ce qui est nécessaire en théorie ( $N \leq 2L + 1$  pour le signal (1.1)) permet d'améliorer les performances d'estimation. Dans cette thèse nous nous plaçons dans le contexte où seul le bruit numérique est présent, ce qui explique que pour obtenir une méthode d'estimation donnant des performances satisfaisantes nous devons utiliser un nombre d'échantillons  $N \gg 2L + 1$  dans nos simulations.

Les algorithmes d'estimation proposés dans la littérature suivent la borne de Cramer Rao dans le cas asymptotique, à fort RSB. A faible RSB, la borne de Cramer Rao prédit mal le comportement des estimateurs qui décrochent de la borne minimale, pour savoir dans quelle situation on peut espérer développer un estimateur efficace il est intéressant de prédire à quel point ce décrochage a lieu. WEI et DRAGOTTI [130] proposent une méthode pour estimer ce point dans le cas d'une somme de Dirac et des noyaux à reproducteurs d'exponentielles. La borne de Cramer-Rao est également utilisée dans les articles utilisant d'autres modèles de signaux que celui donné en (1.1). On peut citer notamment MULLETI, SHENOY et SEELAMANTULA [87] qui la donnent pour les signaux composés d'une somme d'exponentielles et BERENT, DRAGOTTI et BLU [16] qui eux considèrent des signaux sinusoïdaux.

Après avoir dérivé les bornes, il est intéressant de les utiliser pour essayer de trouver la méthode d'échantillonnage promettant les meilleures performances. La borne de Cramer-Rao est un outil particulièrement adapté pour choisir le noyau de pré-filtrage, étant donné que les performances dérivées à partir des échantillons discrets  $c_n$  dépendent du choix du noyau. Cela permet ensuite de donner une référence pour évaluer les performances des méthodes de reconstruction post-échantillonnage, basées sur le meilleur noyau possible. Une telle démarche a été proposée par BEN-HAIM, MICHAELI et ELDAR [13], qui obtiennent que la famille de noyau minimisant la borne continue en l'absence de bruit numérique est formé d'un ensemble d'exponentielles. La borne discrète analytique est ensuite donnée pour l'utilisation de ce noyau, en prenant en compte uniquement le bruit analogique. La méthode d'estimation proposée par DRAGOTTI, VETTERLI et BLU [56] se base sur les moments pour obtenir un nouveau modèle du problème, utilisé pour effectuer ensuite l'estimation. URIGUEN, BLU et DRAGOTTI [123] se base sur le modèle des moments pour donner la borne de Cramer Rao adaptée à cette technique, et la minimise pour trouver le noyau reproducteur d'exponentielles optimal. Nous allons également utiliser une telle démarche d'optimisation dans les sections 2.2.2 et 2.3.2, basée sur les bornes bayésiennes, qui permettent de trouver le noyau SoS optimal sachant la distribution des paramètres.

## 2.2 Performance de la reconstruction d'un signal à somme de Dirac

Dans cette section nous nous intéressons au signal à somme de Dirac, présenté dans l'équation (1.6) du chapitre 1.2. Nous allons tout d'abord nous concentrer sur la borne dé-

terministe, et comparer les performances des trois noyaux sinus cardinal, gaussien et SoS dans ce cadre. Pour ce faire nous donnons d'abord une expression générale de la borne, puis nous faisons des approximations pour des noyaux quasiment orthogonaux. En second lieu nous nous intéressons au contexte Bayésien où les paramètres d'amplitude et de retards sont des variables aléatoires suivant une distribution donnée. Nous donnons la borne Bayésienne pour estimer les paramètres d'un signal à somme de Dirac, que nous utilisons ensuite pour déterminer un noyau optimisant la borne pour l'estimation des retards.

### 2.2.1 Borne déterministe

Les résultats concernant la borne déterministe pour l'échantillonnage des signaux à somme de Dirac est tirée de l'article BERNHARDT et al. [21].

#### Expression analytique de la borne

L'impulsion  $h(t)$  étant une impulsion de Dirac, la convolution entre le signal  $x(t)$  et le noyau  $g(t)$  s'écrit simplement par

$$\mu_n = \langle g(t - nT_S), x(t) \rangle = \sum_{l=0}^{L-1} a_l g(\tau_l - nT_S) \quad (2.1)$$

pour une durée d'analyse de  $N$  échantillons, espacés par un intervalle d'échantillonnage de  $T_S$ . Finalement les échantillons bruités de (1.7) s'écrivent (voir l'illustration figure 1.2)

$$c_n = \mu_n + w_n = \sum_{l=0}^{L-1} a_l g(\tau_l - nT_S) + w_n. \quad (2.2)$$

Par définition (voir 1.3.4) la BCR déterministe est donnée selon

$$\text{BCR}(\boldsymbol{\theta}) = \text{Tr}(\mathbf{I}(\boldsymbol{\theta})^{-1}). \quad (2.3)$$

Pour simplifier le calcul, nous définissons la matrice  $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_L]$  contenant le noyau  $g(t)$  échantillonné à l'intervalle  $T_S$  pour chaque retard  $\tau_l$  :

$$\mathbf{g}_l = [g(\tau_l), g(\tau_l - T_S), \dots, g(\tau_l - (N-1)T_S)]^T. \quad (2.4)$$

L'expression en détail de la matrice est

$$\mathbf{G} = \begin{bmatrix} g(\tau_1) & \cdots & g(\tau_L) \\ \vdots & \ddots & \vdots \\ g(\tau_1 - (N-1)T_S) & \cdots & g(\tau_L - (N-1)T_S) \end{bmatrix}. \quad (2.5)$$

On introduit également  $\dot{\mathbf{G}}$  qui est une matrice similaire à  $\mathbf{G}$  mais contenant les dérivées du noyau  $g(t)$  aux différents instants d'échantillonnage considérés :

$$\dot{\mathbf{G}} = \begin{bmatrix} \dot{g}(\tau_1) & \cdots & \dot{g}(\tau_L) \\ \vdots & \ddots & \vdots \\ \dot{g}(\tau_1 - (N-1)T_S) & \cdots & \dot{g}(\tau_L - (N-1)T_S) \end{bmatrix} = [\dot{\mathbf{g}}_1, \dots, \dot{\mathbf{g}}_L], \quad (2.6)$$

où

$$\dot{\mathbf{g}}_l = [\dot{g}(\tau_l), \dot{g}(\tau_l - T_S), \dots, \dot{g}(\tau_l - (N-1)T_S)]^T. \quad (2.7)$$

avec  $\dot{g}(t) = \frac{\partial g(t)}{\partial t}$ .

On peut réécrire le vecteur  $\boldsymbol{\mu}$  à l'aide de  $\mathbf{G}$  selon

$$\boldsymbol{\mu} = \mathbf{G}\mathbf{a}, \quad (2.8)$$

pour finalement obtenir les dérivées de  $\boldsymbol{\mu}$  par rapport aux différents paramètres inconnus

$$\frac{\partial \boldsymbol{\mu}}{\partial a_l} = \mathbf{g}_l, \quad (2.9)$$

$$\frac{\partial \boldsymbol{\mu}}{\partial \tau_l} = a_l \dot{\mathbf{g}}_l. \quad (2.10)$$

Etant donné que le bruit est gaussien, on peut écrire la matrice de Fisher comme dans l'équation (1.61) qui devient, une fois les dérivées introduites

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \mathbf{B}^T \mathbf{B} = \frac{1}{\sigma^2} \begin{bmatrix} \mathbf{D}\dot{\mathbf{G}}^T \dot{\mathbf{G}} \mathbf{D} & \mathbf{D}\dot{\mathbf{G}}^T \mathbf{G} \\ \mathbf{G}^T \dot{\mathbf{G}} \mathbf{D} & \mathbf{G}^T \mathbf{G} \end{bmatrix}, \quad (2.11)$$

avec  $\mathbf{B} = [\dot{\mathbf{G}} \mathbf{D} \quad \mathbf{G}]$  où  $\mathbf{D}$  est une matrice diagonale contenant les amplitudes  $\mathbf{a}$ .

Les valeurs de BCR nous intéressant sont celles correspondant à l'erreur sur un paramètre donné, c'est à dire les valeurs de la diagonale de l'inverse de la matrice d'information de Fisher. Nous allons nous concentrer sur chaque élément de la diagonale séparément, c'est à dire la BCR pour chacun des paramètres inconnus. Procédons tout d'abord au calcul de la BCR concernant le retard  $\tau_l$ . Pour ce faire nous nous intéressons à la matrice d'information correspondant au vecteur de paramètres permuté

$$\boldsymbol{\theta}^{(\tau_l)} = [\tau_l, \tau_1, \dots, \tau_{l-1}, \tau_{l+1}, \dots, \tau_L, a_1, \dots, a_L]^T = \begin{bmatrix} \mathbf{P}_l & 0 \\ 0 & \mathbf{I}_L \end{bmatrix} \boldsymbol{\theta} \quad (2.12)$$

où  $\mathbf{P}_l$  est une matrice de permutation telle que  $\mathbf{P}_l = [\mathbf{e}_k, \mathbf{e}_1 \dots \mathbf{e}_{l-1} \quad \mathbf{e}_{l+1} \dots \mathbf{e}_L]$ , où  $\mathbf{e}_k$  est le vecteur unité, ayant un 1 à la  $k^e$  entrée, et des 0 ailleurs. La matrice d'information de Fisher dans ce cas est

$$\mathbf{I}(\boldsymbol{\theta}^{(\tau_l)}) = \frac{1}{\sigma^2} \begin{bmatrix} \mathbf{P}_l^T \mathbf{D}\dot{\mathbf{G}}^T \dot{\mathbf{G}} \mathbf{D} \mathbf{P}_l & \mathbf{P}_l^T \mathbf{D}\dot{\mathbf{G}}^T \mathbf{G} \\ \mathbf{G}^T \dot{\mathbf{G}} \mathbf{D} \mathbf{P}_l & \mathbf{G}^T \mathbf{G} \end{bmatrix}. \quad (2.13)$$

Considérant une matrice par bloc  $\mathbf{A}$  inversible. L'inverse d'une matrice par bloc s'écrit :

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C}_1^{-1} & -\mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{C}_2^{-1} \\ -\mathbf{C}_2^{-1} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} & \mathbf{C}_2^{-1} \end{bmatrix}, \quad (2.14)$$

où on note :

$$\mathbf{C}_1 = \mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21}, \quad (2.15)$$

$$\mathbf{C}_2 = \mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}. \quad (2.16)$$

En se basant sur la formule d'inversion des matrices par blocs (2.14) tirée de PETERSEN et PEDERSEN [98], on obtient en se concentrant sur le bloc contenant uniquement les retards que la BCR sur le paramètre  $\tau_l$  est

$$\text{BCR}(\tau_l) = \frac{\sigma^2}{a_l^2} \left[ (\mathbf{P}_l^T \mathbf{G}^T \mathbf{P}_l \mathbf{G} \mathbf{P}_l)^{-1} \right]_{11} \quad (2.17)$$

où l'on a introduit le projecteur orthogonal  $\mathbf{P}_G^\perp = \mathbf{I} - \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T$  de (2.19) sur le sous-espace vectoriel engendré par les colonnes de  $\langle \mathbf{G} \rangle^\perp$ . L'utilisation des projecteurs orthogonaux et obliques permettent de donner une interprétation géométrique des bornes de Cramer Rao [30, 28]. BEHRENS et SCHARF [10] attestent de l'importance de ces projecteurs dans différentes applications du traitement du signal. Pour différencier les différentes projections on notera  $\mathbf{P}_H$  une projection orthogonale sur l'espace des colonnes de  $\mathbf{H}$ , tandis qu'on utilise  $\mathbf{E}_{(\mathbf{H} \mathbf{S})}$  pour une projection oblique sur l'espace des colonnes de  $\mathbf{H}$  et de noyau  $\mathbf{S}$ . De manière générale une matrice de projection est une matrice imdépotente c'est à dire que  $\mathbf{E}^2 = \mathbf{E}$ . Nous donnons une liste non exhaustive des propriétés des projecteurs, notamment celles que nous utiliserons dans cette thèse :

$$\mathbf{P}_H^H = \mathbf{P}_H \quad (2.18)$$

$$\mathbf{P}_H = \mathbf{H} (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \quad (2.19)$$

$$\mathbf{P}_H^\perp = \mathbf{I} - \mathbf{P}_H \quad (2.20)$$

$$\mathbf{E}_{(\mathbf{H} \mathbf{S})} \mathbf{H} = \mathbf{H} \quad (2.21)$$

$$\mathbf{E}_{(\mathbf{H} \mathbf{S})} \mathbf{S} = \mathbf{0} \quad (2.22)$$

$$\mathbf{E}_{(\mathbf{H} \mathbf{S})} = \mathbf{H} (\mathbf{H}^H \mathbf{P}_S^\perp \mathbf{H})^{-1} \mathbf{H}^T \mathbf{P}_S^\perp \quad (2.23)$$

$$\mathbf{E}_{(\mathbf{H} \mathbf{S})} = \mathbf{P}_H \left( \mathbf{I} - \mathbf{S} (\mathbf{S}^H \mathbf{P}_H^\perp \mathbf{S})^{-1} \mathbf{S}^H \mathbf{P}_H^\perp \right) \quad (2.24)$$

$$\mathbf{E}_{(\mathbf{H} \mathbf{S})} = \mathbf{P}_H (\mathbf{I} - \mathbf{E}_{(\mathbf{S} \mathbf{H})}) \quad (2.25)$$

$$\mathbf{H}^\dagger \mathbf{E}_{(\mathbf{H} \mathbf{S})} = (\mathbf{P}_S^\perp \mathbf{H})^\dagger = (\mathbf{H}^H \mathbf{P}_S^\perp \mathbf{H})^{-1} \mathbf{H}^T \mathbf{P}_S^\perp \quad (2.26)$$

où  $\cdot^H$  est la transposée hermitienne et  $\mathbf{0}$  est la matrice nulle. Notons la matrice  $[\mathbf{H} \mathbf{S}]$  la matrice concaténant les matrices  $\mathbf{H}$  et  $\mathbf{S}$ , on a le lien suivant entre les projections orthogonales et obliques :

$$\mathbf{P}_{[\mathbf{H} \mathbf{S}]} = \mathbf{E}_{(\mathbf{H} \mathbf{S})} + \mathbf{E}_{(\mathbf{S} \mathbf{H})} \quad (2.27)$$

Nous allons à présent chercher une expression analytique de la borne (2.17). Pour ce faire nous détaillons tout d'abord la matrice :

$$\mathbf{P}_l^T \dot{\mathbf{G}}^T \mathbf{P}_G^\perp \dot{\mathbf{G}} \mathbf{P}_l = \begin{bmatrix} \dot{\mathbf{g}}_l^T \mathbf{P}_G^\perp \dot{\mathbf{g}}_l & \dot{\mathbf{g}}_l^T \mathbf{P}_G^\perp \dot{\mathbf{G}}^{(l)} \\ \dot{\mathbf{G}}^{(l)T} \mathbf{P}_G^\perp \dot{\mathbf{g}}_l & \dot{\mathbf{G}}^{(l)T} \mathbf{P}_G^\perp \dot{\mathbf{G}}^{(l)} \end{bmatrix}. \quad (2.28)$$

où l'on note  $\dot{\mathbf{G}}^{(l)}$  la matrice extraite de  $\dot{\mathbf{G}}$  en retirant la  $l$ -ième colonne, c'est à dire le vecteur  $\dot{\mathbf{g}}_l$ .

Après inversion de cette matrice par bloc, on obtient

$$\left[ \left( \mathbf{P}_l^T \dot{\mathbf{G}}^T \mathbf{P}_G^\perp \dot{\mathbf{G}} \mathbf{P}_l \right)^{-1} \right]_{11} = \dot{\mathbf{g}}_l^T \mathbf{P}_G^\perp \dot{\mathbf{g}}_l - \dot{\mathbf{g}}_l^T \mathbf{P}_G^\perp \dot{\mathbf{G}}^{(l)} \left( \dot{\mathbf{G}}^{(l)T} \mathbf{P}_G^\perp \dot{\mathbf{G}}^{(l)} \right)^{-1} \dot{\mathbf{G}}^{(l)T} \mathbf{P}_G^\perp \dot{\mathbf{g}}_l \quad (2.29)$$

$$\text{avec (2.23) on a} \quad = \dot{\mathbf{g}}_l^T \mathbf{P}_G^\perp \dot{\mathbf{g}}_l - \dot{\mathbf{g}}_l^T \mathbf{P}_G^\perp \mathbf{E}_{(\dot{\mathbf{G}}^{(l)} \mathbf{G})} \dot{\mathbf{g}}_l \quad (2.30)$$

$$= \dot{\mathbf{g}}_l^T \mathbf{P}_G^\perp \left( \mathbf{I} - \mathbf{E}_{(\dot{\mathbf{G}}^{(l)} \mathbf{G})} \right) \dot{\mathbf{g}}_l \quad (2.31)$$

$$= \dot{\mathbf{g}}_l^T \left[ \mathbf{I} - \mathbf{E}_{(\dot{\mathbf{G}}^{(l)} \mathbf{G})} - \mathbf{P}_G \left( \mathbf{I} - \mathbf{E}_{(\dot{\mathbf{G}}^{(l)} \mathbf{G})} \right) \right] \dot{\mathbf{g}}_l \quad (2.32)$$

$$\text{avec (2.25) on a} \quad = \dot{\mathbf{g}}_l^T \left[ \mathbf{I} - \mathbf{E}_{(\dot{\mathbf{G}}^{(l)} \mathbf{G})} - \mathbf{E}_{(\mathbf{G} \dot{\mathbf{G}}^{(l)})} \right] \dot{\mathbf{g}}_l \quad (2.33)$$

$$\text{avec (2.27) on a} \quad = \dot{\mathbf{g}}_l^T \left[ \mathbf{I} - \mathbf{P}_{[\dot{\mathbf{G}}^{(l)} \mathbf{G}]} \right] \dot{\mathbf{g}}_l \quad (2.34)$$

$$= \dot{\mathbf{g}}_l^T \mathbf{P}_{[\dot{\mathbf{G}}^{(l)} \mathbf{G}]}^\perp \dot{\mathbf{g}}_l \quad (2.35)$$

$$= \left\| \mathbf{P}_{[\dot{\mathbf{G}}^{(l)} \mathbf{G}]}^\perp \dot{\mathbf{g}}_l \right\|^2. \quad (2.36)$$

On a introduit dans le calcul de la borne le projecteur oblique  $\mathbf{E}_{(\dot{\mathbf{G}}^{(l)} \mathbf{G})}$  sur l'espace des colonnes de  $\langle \dot{\mathbf{G}}^{(l)} \rangle$  et de noyau  $\langle \mathbf{G} \rangle$  ainsi que le projecteur orthogonal  $\mathbf{P}_{[\dot{\mathbf{G}}^{(l)} \mathbf{G}]}^\perp$  sur l'espace des colonnes de la matrice concaténée  $\langle [\dot{\mathbf{G}}^{(l)} \mathbf{G}] \rangle^\perp$ .

On utilise la définition du RSB suivante :  $\text{RSB}_l = a_l^2 / \sigma^2$ , où  $\sigma^2$  est la variance du bruit numérique, dépendant du nombre de bits de la quantification et de la chaîne de traitement. La BCR sur le  $l^e$  retard est

$$\text{BCR}(\tau_l) = \frac{1}{\text{RSB}_l} \frac{1}{\left\| \mathbf{P}_{[\dot{\mathbf{G}}^{(l)} \mathbf{G}]}^\perp \dot{\mathbf{g}}_l \right\|^2}. \quad (2.37)$$

Cette dernière expression de la borne est intéressante car on peut en donner une interprétation géométrique. En effet, on considère le sous-espace généré par  $\langle [\dot{\mathbf{G}}^{(l)} \mathbf{G}] \rangle$  dont on retire le vecteur  $\dot{\mathbf{g}}_l$  quand on s'intéresse à la  $l^e$  impulsion. On obtient alors le sous-espace réduit  $\langle [\dot{\mathbf{G}}^{(l)} \mathbf{G}] \rangle$  de dimension  $2L - 1$ . La BCR est alors proportionnelle à la norme de la projection du vecteur  $\dot{\mathbf{g}}_l$  sur le complément orthogonal de l'espace issu de la dérivée du noyau et du noyau duquel on a retiré ce même vecteur.

On procède maintenant exactement de la même manière pour calculer la borne correspondant à l'amplitude de la  $l^e$  impulsion. Pour ce faire nous allons nous intéresser au premier élément de la matrice de BCR associé au vecteur permuté

$$\boldsymbol{\theta}^{(a_l)} = [\tau_1, \dots, \tau_L, a_l, a_1, \dots, a_{l-1}, a_{l+1}, \dots, a_L]^T = \begin{bmatrix} \mathbf{I}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_l \end{bmatrix} \boldsymbol{\theta}. \quad (2.38)$$

On se retrouve avec la matrice d'information de Fisher suivante :

$$\mathbf{I}(\boldsymbol{\theta}^{(a_l)}) = \frac{1}{\sigma^2} \begin{bmatrix} \mathbf{D}\dot{\mathbf{G}}^T \dot{\mathbf{G}} \mathbf{D} & \mathbf{D}\dot{\mathbf{G}}^T \mathbf{G} \mathbf{P}_l \\ \mathbf{P}_l^T \mathbf{G}^T \dot{\mathbf{G}} \mathbf{D} & \mathbf{P}_l^T \mathbf{G}^T \mathbf{G} \mathbf{P}_l \end{bmatrix}. \quad (2.39)$$

Après inversion de cette matrice on a

$$\text{BCR}(a_l) = \sigma^2 \left[ \left( \mathbf{P}_l^T \mathbf{G}^T \mathbf{P}_G^\perp \mathbf{G} \mathbf{P}_l \right)^{-1} \right]_{11} = \sigma^2 \mathbf{C}_2^{-1}. \quad (2.40)$$

Où  $\mathbf{C}_2$  est donné en (2.16). Notons  $\mathbf{G}^{(l)}$  la matrice extraite de  $\mathbf{G}$  en retirant la  $l^e$  colonne, c'est à dire le vecteur  $\mathbf{g}_l$ . On a

$$\mathbf{C}_2 = \mathbf{g}_l^T \mathbf{P}_{\dot{\mathbf{G}}}^\perp \mathbf{g}_l - \mathbf{g}_l^T \mathbf{P}_{\dot{\mathbf{G}}}^\perp \mathbf{G}^{(l)} \left( \mathbf{G}^{(l)T} \mathbf{P}_{\dot{\mathbf{G}}}^\perp \mathbf{G}^{(l)} \right)^{-1} \mathbf{G}^{(l)T} \mathbf{P}_{\dot{\mathbf{G}}}^\perp \mathbf{g}_l \quad (2.41)$$

$$= \left\| \mathbf{P}_{[\mathbf{G}^{(l)} \dot{\mathbf{G}}]}^\perp \mathbf{g}_l \right\|^2 \quad (2.42)$$

où  $\mathbf{P}_{[\mathbf{G}^{(l)} \dot{\mathbf{G}}]}^\perp$  est le projecteur orthogonal sur l'espace des colonnes de la matrice concaténée  $\langle [\mathbf{G}^{(l)} \dot{\mathbf{G}}] \rangle^\perp$ . La BCR sur la  $l^e$  amplitude est alors

$$\text{BCR}(a_l) = \frac{\sigma^2}{\left\| \mathbf{P}_{[\mathbf{G}^{(l)} \dot{\mathbf{G}}]}^\perp \mathbf{g}_l \right\|^2}. \quad (2.43)$$

Finalement la BCR totale est la somme des différentes bornes sur chaque paramètre, données dans les équations (2.37) et (2.43).

**Résultat 2.** La BCR pour l'estimation des signaux à somme de Dirac est :

$$\begin{aligned} \text{BCR} &= \text{Tr}(\mathbf{C}(\boldsymbol{\theta})) = \sum_{l=1}^L \text{BCR}(\tau_l) + \text{BCR}(a_l) \\ &= \sum_{l=1}^L \frac{1}{\text{RSB}_l} \frac{1}{\left\| \mathbf{P}_{[\dot{\mathbf{G}}^{(l)} \dot{\mathbf{G}}]}^\perp \dot{\mathbf{g}}_l \right\|^2} + \frac{\sigma^2}{\left\| \mathbf{P}_{[\mathbf{G}^{(l)} \dot{\mathbf{G}}]}^\perp \mathbf{g}_l \right\|^2}. \end{aligned} \quad (2.44)$$

Nous obtenons une expression analytique de la borne généralisant au cas à plusieurs impulsions la borne donnée par BLU et al. [27] qui ne considérait qu'une unique impulsion. De plus la borne obtenue est assez générale pour pouvoir comparer toutes les familles de noyaux, même ceux ne suivant pas la condition (1.11) ou la condition Strang-Fix. Ce n'est pas le cas de la borne dérivée par ANGIERSKI et KUEHN [4]. Nous allons par exemple dans la partie simulation, comparer le noyau gaussien avec d'autres noyaux classiques bien qu'il n'entre pas dans les familles de noyaux valables présentés dans la section 1.2.2.

### Approximations grâce aux propriétés d'orthogonalité des noyaux

Dans cette sous-section, nous cherchons à obtenir une expression simplifiée de la borne grâce à différentes approximations sur le comportement des noyaux. Nous allons ensuite montrer par la simulation que ces approximations sont validées pour certains noyaux couramment utilisés, bien qu'elles ne le soient pas nécessairement dans le cas général. Nous nous concentrons ici sur la comparaison de quelques noyaux pour lesquels nous allons vérifier empiriquement les différentes hypothèses. Il s'agit des noyaux en sinus cardinal (1.14), gaussien (1.15), et enfin le noyau SoS (1.17) paramétré pour former le noyau de Dirichlet (1.21) et avec la fenêtre de hamming (1.27). Tous ces noyaux sont introduits dans la section 1.2.3.

Nous considérons que les noyaux et leurs dérivées ont les propriétés d'orthogonalité

suivantes :

$$\mathbf{G}^T \mathbf{G} \approx \begin{bmatrix} \|\mathbf{g}_1\|^2 & 0 & \cdots & 0 \\ 0 & \|\mathbf{g}_2\|^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \|\mathbf{g}_L\|^2 \end{bmatrix} = \mathbf{F} \quad (2.45)$$

$$\dot{\mathbf{G}}^T \dot{\mathbf{G}} \approx \begin{bmatrix} \|\dot{\mathbf{g}}_1\|^2 & 0 & \cdots & 0 \\ 0 & \|\dot{\mathbf{g}}_2\|^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \|\dot{\mathbf{g}}_L\|^2 \end{bmatrix} = \mathbf{E} \quad (2.46)$$

$$\dot{\mathbf{G}}^T \mathbf{G} = \mathbf{G}^T \dot{\mathbf{G}} \approx \mathbf{0}_L \quad (2.47)$$

où  $\mathbf{0}_L$  est la matrice  $L \times L$  ne contenant que des 0. La propriété d'orthogonalité des noyaux (2.45) est valide notamment quand les supports temporels discrétisés des noyaux sont disjoints. Cette propriété est vérifiée de manière assez intuitive étant donné que les noyaux d'échantillonnage sont construits de manière à ce qu'il y ait un pic au temps zéro, qui une fois convolué donnera un pic à chaque retard d'impulsion  $\tau_l$  et ensuite descend rapidement à zéro. Ce comportement peut s'observer sur la figure 1.3. pour que la propriété soit vérifiée il faut donc que les impulsions soient plus espacées que la largeur du support du noyau. Les équations (2.47) et (2.46) sont vérifiées de la même façon quand le support échantillonné de la dérivée du noyau  $\dot{g}(t)$  est étroit et donc que pour deux retards suffisamment éloignés les supports de la dérivée du noyau et du noyau à deux impulsions différentes (respectivement la dérivée du noyau à deux impulsions différentes) sont disjoints. Cette propriété est donc moins naturelle, étant donné que la dérivée du noyau doit également avoir un support étroit, ce qui n'est pas utile directement pour l'échantillonnage. On peut observer sur la figure 2.1 que les supports des dérivés ne sont pas étroits la plupart du temps, comme on pourrait le craindre, cependant les valeurs oscillants autour de zéro permettent une annulation de l'effet de ce support quand on considère la norme. Nous vérifierons par simulation pour quels retards la BCR approximée obtenue à partir de (2.45), (2.46) et (2.47) est valide.

En considérant les propriétés d'orthogonalité validées, on peut développer les expressions des projecteurs :

$$\mathbf{P}_{[\dot{\mathbf{G}}^{(l)} \mathbf{G}]}^\perp = \mathbf{I}_N - [\dot{\mathbf{G}}^{(l)} \quad \mathbf{G}] \begin{bmatrix} \dot{\mathbf{G}}^{(l)T} \dot{\mathbf{G}}^{(l)} & \dot{\mathbf{G}}^{(l)T} \mathbf{G} \\ \mathbf{G}^T \dot{\mathbf{G}}^{(l)} & \mathbf{G}^T \mathbf{G} \end{bmatrix}^{-1} \begin{bmatrix} \dot{\mathbf{G}}^{(l)T} \\ \mathbf{G}^T \end{bmatrix} \quad (2.48)$$

$$\approx \mathbf{I}_N - [\dot{\mathbf{G}}^{(l)} \quad \mathbf{G}] \begin{bmatrix} \mathbf{E}^{(l)} & \mathbf{0}_{(L-1, L)} \\ \mathbf{0}_{(L, L-1)} & \mathbf{F} \end{bmatrix}^{-1} \begin{bmatrix} \dot{\mathbf{G}}^{(l)T} \\ \mathbf{G}^T \end{bmatrix} \quad (2.49)$$

$$\approx \mathbf{I}_N - \dot{\mathbf{G}}^{(l)} \mathbf{E}^{(l)-1} \dot{\mathbf{G}}^{(l)T} - \mathbf{G} \mathbf{F}^{-1} \mathbf{G}^T, \quad (2.50)$$

$$\mathbf{P}_{[\mathbf{G}^{(l)} \dot{\mathbf{G}}]}^\perp = \mathbf{I}_N - [\mathbf{G}^{(l)} \quad \dot{\mathbf{G}}] \begin{bmatrix} \mathbf{G}^{(l)T} \mathbf{G}^{(l)} & \mathbf{G}^{(l)T} \dot{\mathbf{G}} \\ \dot{\mathbf{G}}^T \mathbf{G}^{(l)} & \dot{\mathbf{G}}^T \dot{\mathbf{G}} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{G}^{(l)T} \\ \dot{\mathbf{G}}^T \end{bmatrix} \quad (2.51)$$

$$\approx \mathbf{I}_N - [\mathbf{G}^{(l)} \quad \dot{\mathbf{G}}] \begin{bmatrix} \mathbf{F}^{(l)} & \mathbf{0}_{(L-1, L)} \\ \mathbf{0}_{(L, L-1)} & \mathbf{E} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{G}^{(l)T} \\ \dot{\mathbf{G}}^T \end{bmatrix} \quad (2.52)$$

$$\approx \mathbf{I}_N - \mathbf{G}^{(l)} \mathbf{F}^{(l)-1} \mathbf{G}^{(l)T} - \dot{\mathbf{G}} \mathbf{E}^{-1} \dot{\mathbf{G}}^T \quad (2.53)$$

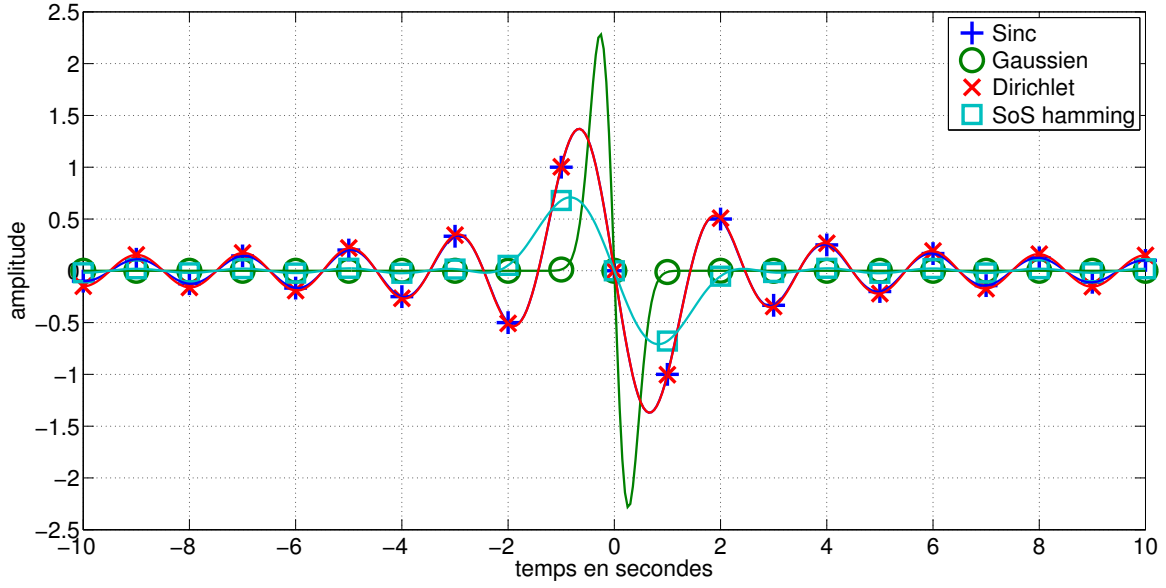


FIGURE 2.1 – La dérivée des noyaux utilisés  $\dot{g}(t)$  où  $T_S = 10^{-2}$  et  $N = 7$ .

où  $\mathbf{E}^{(l)}$  et  $\mathbf{F}^{(l)}$  sont les matrices de dimensions  $(L-1) \times (L-1)$  extraites de  $\mathbf{E}$  et  $\mathbf{F}$  respectivement en enlevant la  $l$ -ième colonne et ligne. Il découle de ces expressions que

$$\mathbf{P}_{[\dot{\mathbf{G}}^{(l)} \mathbf{G}]}^\perp \dot{\mathbf{g}}_l \approx \dot{\mathbf{g}}_l - \underbrace{\dot{\mathbf{G}}^{(l)} \mathbf{E}^{(l)-1} \dot{\mathbf{G}}^{(l)T} \dot{\mathbf{g}}_l}_{=0} - \underbrace{\mathbf{G} \mathbf{F}^{(l)-1} \mathbf{G}^T \dot{\mathbf{g}}_l}_{=0} \quad (2.54)$$

$$\approx \dot{\mathbf{g}}_l, \quad (2.55)$$

$$\mathbf{P}_{[\mathbf{G}^{(l)} \dot{\mathbf{G}}]}^\perp \mathbf{g}_l \approx \mathbf{g}_l - \underbrace{\mathbf{G}^{(l)} \mathbf{F}^{(l)-1} \mathbf{G}^{(l)T} \mathbf{g}_l}_{=0} - \underbrace{\mathbf{G} \mathbf{E}^{-1} \mathbf{G}^T \mathbf{g}_l}_{=0} \quad (2.56)$$

$$\approx \mathbf{g}_l. \quad (2.57)$$

On peut donc à présent donner une formule de la BCR approximée en introduisant (2.55) et (2.57) dans (2.44) pour obtenir le résultat suivant.

**Résultat 3.** La BCR pour l'échantillonnage des signaux à modèles paramétriques avec un noyau  $g(t)$  remplissant les conditions (2.45), (2.46) et (2.47) est :

$$\text{BCR} = \sum_{l=1}^L \frac{1}{\text{RSB}_l} \frac{1}{\|\dot{\mathbf{g}}_l\|^2} + \frac{\sigma^2}{\|\mathbf{g}_l\|^2}. \quad (2.58)$$

Les conditions d'orthogonalité citées ne sont que des approximations pour la majorité des noyaux d'échantillonnage utilisés. De ce fait (2.58) est seulement une BCR approximée. Nous allons à présent valider cette approximation pour les différents noyaux présentés dans la section 1.2.3, qui sont les noyaux les plus couramment utilisés. Il nous importe de valider les expressions (2.55) et (2.57), pour cela nous traçons les rapports entre la norme de  $\dot{\mathbf{g}}$  projeté sur la norme de  $\dot{\mathbf{g}}$ , et la norme de  $\mathbf{g}$  projeté sur la norme de  $\mathbf{g}$ , soient

$$\frac{\|\mathbf{P}_{[\dot{\mathbf{G}}^{(l)} \mathbf{G}]}^\perp \dot{\mathbf{g}}_l\|^2}{\|\dot{\mathbf{g}}_l\|^2} \text{ et } \frac{\|\mathbf{P}_{[\mathbf{G}^{(l)} \dot{\mathbf{G}}]}^\perp \mathbf{g}_l\|^2}{\|\mathbf{g}_l\|^2}. \quad (2.59)$$

Ces deux rapports sont tracés en fonction du retard de l'une des sources sur les figures 2.2 et 2.3 respectivement. Nous utilisons les mêmes paramètres des noyaux que dans les figures 1.3 et 2.1,  $N = 21$  échantillons et toujours un intervalle d'échantillonnage de  $T_S = 1$  seconde. Nous avons considéré deux retards, le premier égal à 10 est la valeur de référence et le deuxième retard varie sur la longueur de la fenêtre d'analyse et est donné en abscisse.

Quand le ratio est égal ou proche de 1 cela signifie que l'approximation est bonne. La première chose qui saute aux yeux est que ce n'est pas le cas pour des valeurs autour de  $\tau_1 = 10$ . Cela correspond à l'intuition qui demande à ce que les supports soient disjoints aussi bien pour les noyaux (rapports que l'on observe surtout dans le cas de la figure 2.3) que pour leur dérivée (qui est dans la figure 2.2). On peut noter que les comportements des différents noyaux correspondent à l'intuition que l'on peut avoir à partir des figures 1.3 et 2.1. Le noyau gaussien vérifie très bien toutes les approximations car il a un support étroit aussi bien directement, que dans sa dérivée. Le noyau SoS avec une fenêtre de Hamming est un peu plus étalé, les retards doivent donc être plus éloignés les uns des autres pour que les hypothèses d'orthogonalité soient respectées, notamment celle de la figure 2.2 qui se base principalement sur la dérivée. De manière générale on constate que pour les quelques noyaux testés, l'approximation permet d'obtenir une bonne intuition de la réalité dès que les retards sont espacés d'au moins  $2T_S$ , ce qui est réaliste dans des applications où le nombre d'impulsions est faible sur une durée d'échantillonnage longue. On peut supposer que de manière générale on peut trouver un intervalle à respecter entre deux retards pour que les approximations soient validées pour tout type de noyau.

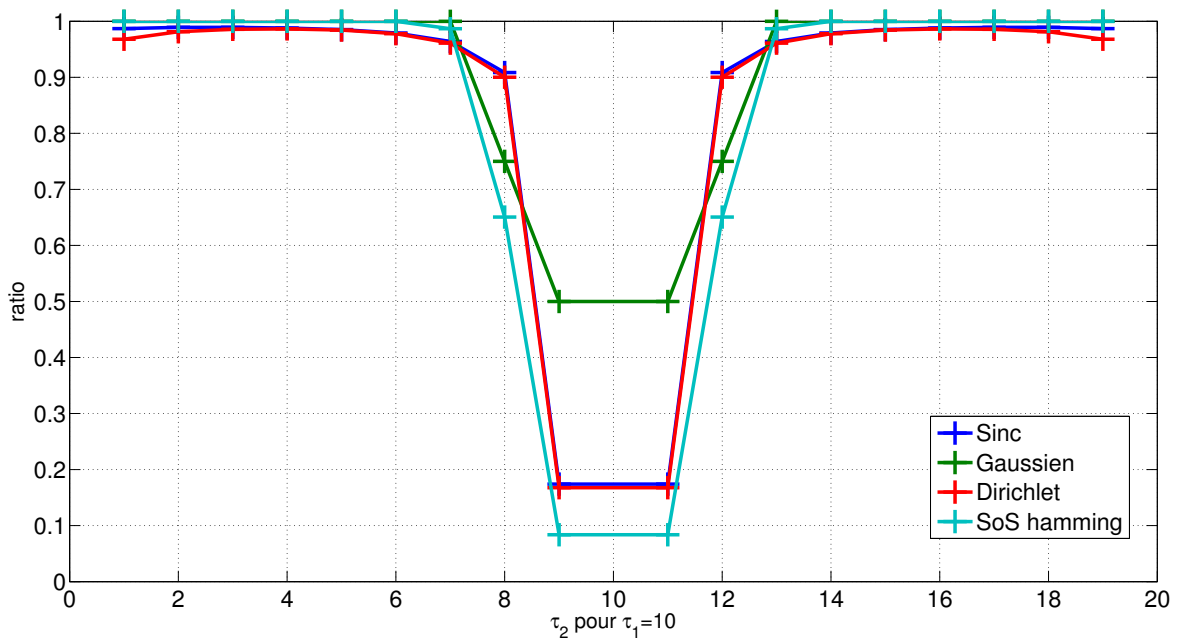


FIGURE 2.2 – Vérification de l'approximation pour le calcul de  $\text{BCR}(\tau_l)$ .

Pour les courbes précédentes, nous avons uniquement utilisé les retards  $\tau = kT_S$  étant un multiple de l'intervalle d'échantillonnage. Sur les figures 2.4 et 2.5 nous utilisons exactement les mêmes paramètres, mais l'abscisse est plus précise : nous considérons des  $\tau_2$  espacés de 0.1 c'est à dire 10 fois plus précisément que l'échantillonnage choisi. Dans ce cas nous observons que les noyaux Sinc et SoS continuent de plutôt bien vérifier les hypothèses d'orthogonalité alors que le noyau gaussien en est très loin. Le noyau gaussien

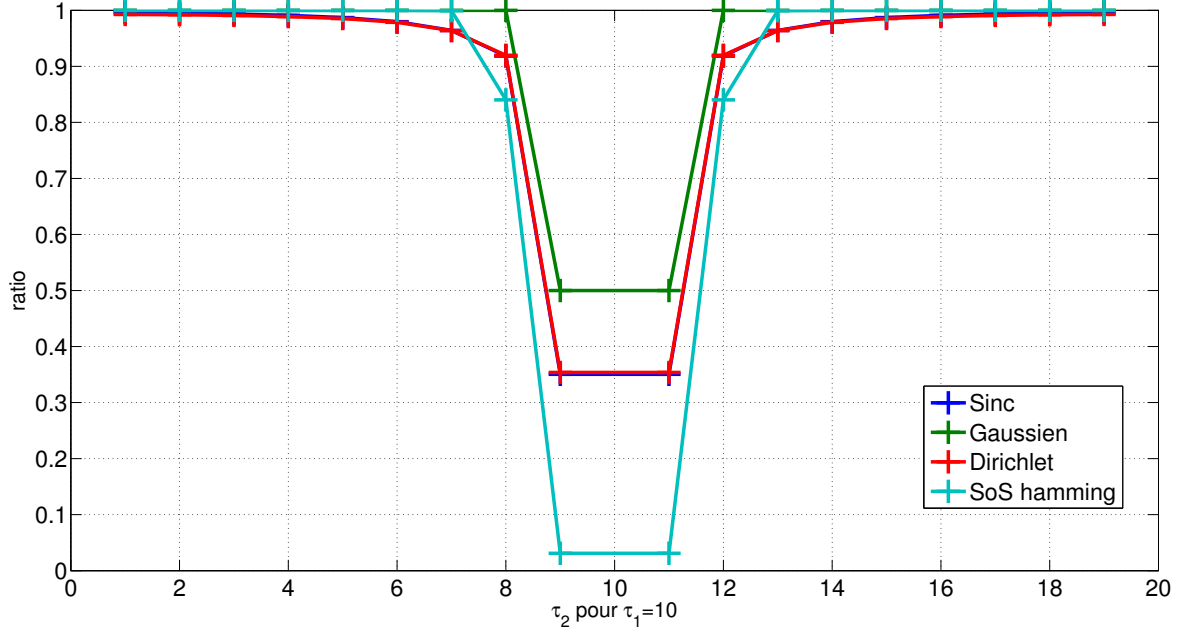


FIGURE 2.3 – Vérification de l'approximation pour le calcul de  $\text{BCR}(a_l)$ .

est donc orthogonal uniquement si les retards sont des multiples de  $T_S$ , l'approximation risque d'être très mauvaise dans le cas concret de retards continus, ce que nous confirmons dans la figure 2.9.

### Indépendance de la borne par rapport aux retards

Nous avons observé par simulation que la borne semble ne pas dépendre des retards  $\tau_l$ . Dans cette partie nous allons vérifier cette propriété qui permet encore de grandement simplifier le calcul de la borne. De plus une borne indépendante des paramètres à estimer est utile pour que la comparaison des différents noyaux par rapport à cette borne soit judicieuse, en effet cela permet de comparer les noyaux entre eux, sans que cette comparaison ne vaille que pour un jeu de retards donnés. Pour montrer que la borne ne dépend pas des retards, on va chercher à prouver que la dérivée de  $\gamma(\tau) = \|\mathbf{g}_l\|^2$  et de  $\tilde{\gamma}(\tau) = \|\dot{\mathbf{g}}_l\|^2$  par rapport à  $\tau$  est nulle.

Considérons tout d'abord le cas de  $\gamma(\tau)$  :

$$\gamma(\tau) = \sum_{n=0}^{N-1} g(\tau - nT_S)^2 \quad (2.60)$$

d'où l'on obtient la dérivée

$$\frac{\partial \gamma(\tau)}{\partial \tau} = \sum_{n=0}^{N-1} 2g(\tau - nT_S) \dot{g}(\tau - nT_S) = \sum_{n=0}^{N-1} f(\tau - nT_S) \quad (2.61)$$

en définissant  $f(t) = 2g(t)\dot{g}(t)$ .

Quand l'on trace la fonction  $f(t)$  pour les différents noyaux (voir figure 2.6) on observe que cette fonction suit les deux propriétés suivantes :

1.  $f(t)$  est une fonction impaire, c'est à dire qu'on a  $f(t) = -f(-t)$  (et en particulier  $f(0) = 0$ ),

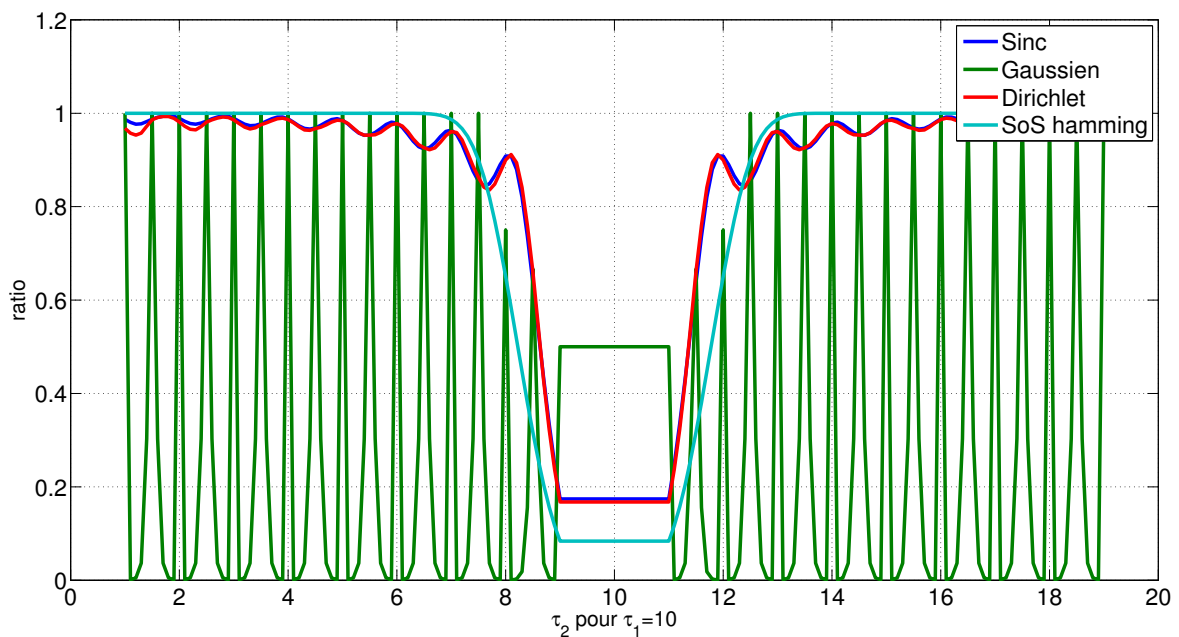


FIGURE 2.4 – Vérification de l'approximation pour le calcul de  $\text{BCR}(\tau_l)$  pour des retards hors de la grille.

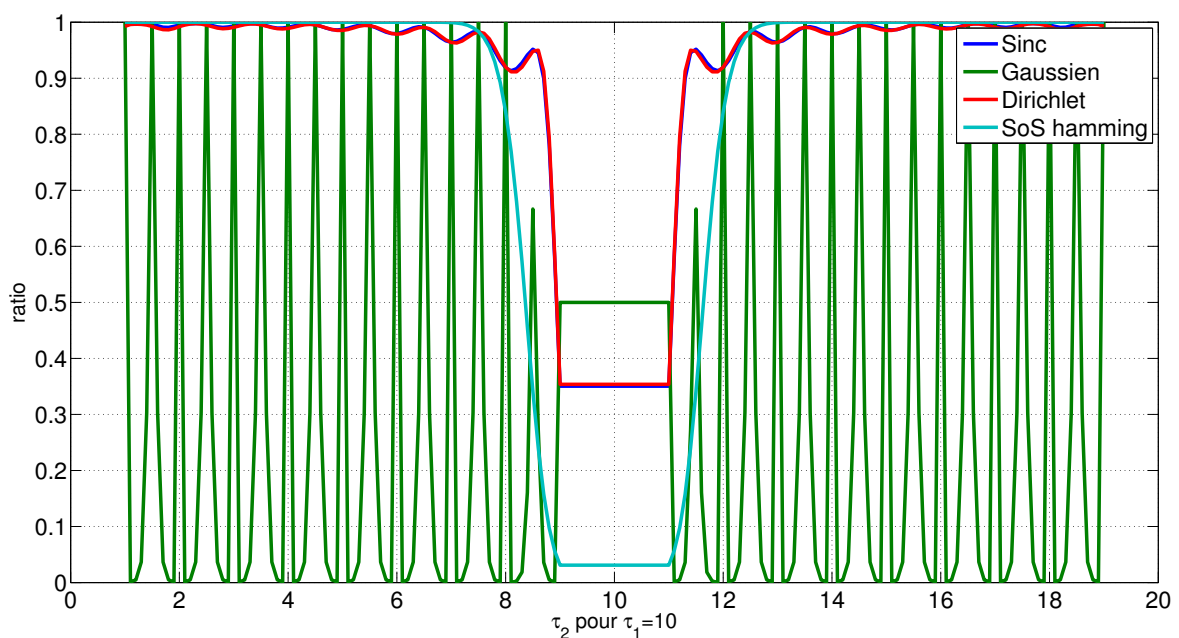


FIGURE 2.5 – Vérification de l'approximation pour le calcul de  $\text{BCR}(a_l)$  pour des retards hors de la grille.

2.  $f(t)$  a un support fini de longueur  $(2N_S + 1)T_S$  où l'on a nommé  $N_S$  la largeur du support. On a donc la propriété que  $f(t) = 0$  pour tout  $|t| \leq N_S T_S$

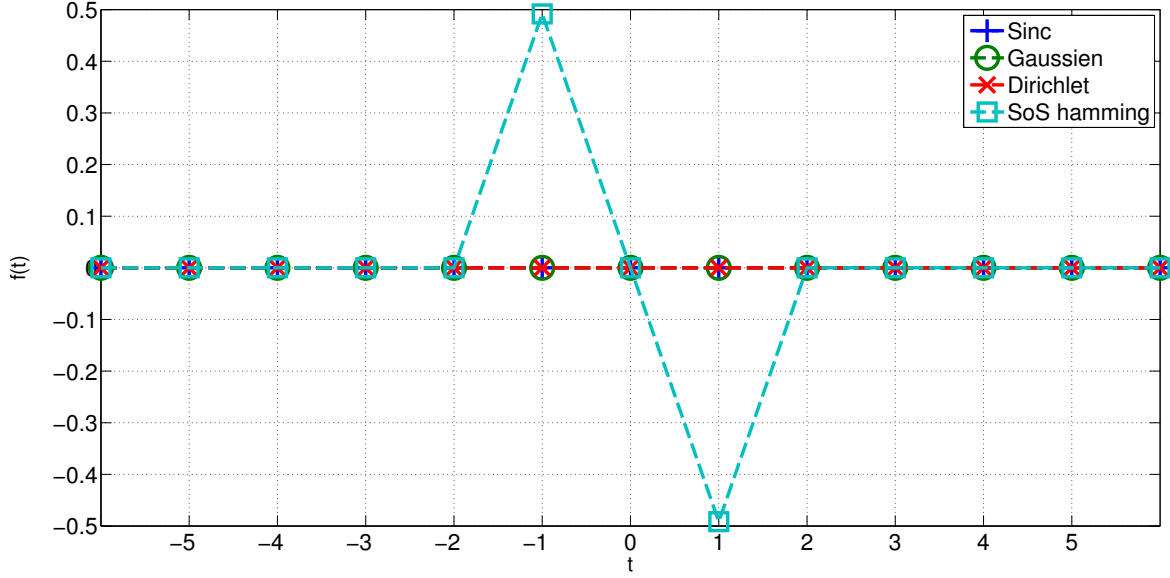


FIGURE 2.6 – Visualisation de  $f(t)$ , utilisée pour montrer que la BCR sur l'amplitude ne dépend pas du retard.

On fait de plus l'hypothèse que les retards des impulsions tombent sur la grille d'échantillonnage, c'est à dire qu'on peut noter  $\tau_l = n_l T_S$ . Dès lors on a :

$$\begin{aligned} \left. \frac{\partial \gamma(\tau)}{\partial \tau} \right|_{\tau=n_l T_S} &= \sum_{n=0}^{N-1} f((n_l - n)T_S) \\ &= f(0) + \sum_{n=0}^{n_l-1} f((n_l - n)T_S) + \sum_{n_l+1}^{N-1} f((n_l - n)T_S). \end{aligned}$$

En introduisant la propriété 2) et en faisant l'hypothèse que  $n_l$  appartient à  $[N_S, N - N_S - 1]$  :

$$\begin{aligned} \left. \frac{\partial \gamma(\tau)}{\partial \tau} \right|_{\tau=n_l T_S} &= f(0) + \sum_{n=n_l-N_S}^{n_l-1} f((n_l - n)T_S) + \sum_{n=n_l+1}^{n_l+N_S} f((n_l - n)T_S) \\ &= f(0) + \sum_{n=1}^{N_S} \underbrace{f(nT_S) + f(-nT_S)}_{=0 \text{ propriété 1)}} = f(0) = 0. \end{aligned}$$

En conclusion on peut considérer que  $\gamma(\tau)$  ne dépend pas du retard  $\tau_l$  à partir du moment où les retards ne sont pas près des bords de la fenêtre d'analyse de l'échantillonnage. En d'autres termes les retards doivent être assez éloignés de 0 et à l'opposé de N. Cela vient du fait qu'on a besoin que tout le support de  $f(t)$  soit considéré dans le calcul de la norme pour tirer parti de la propriété de symétrie.

De la même façon, on considère la dérivée de

$$\tilde{\gamma}(\tau) = \sum_{n=0}^{N-1} \dot{g}(\tau - nT_S)^2 : \quad (2.62)$$

$$\frac{\partial \tilde{\gamma}(\tau)}{\partial \tau} = \sum_{n=0}^{N-1} 2\dot{g}(\tau - nT_s) \ddot{g}(\tau - nT_s) = \sum_{n=0}^{N-1} f'(\tau - nT_s) \quad (2.63)$$

en définissant  $f'(t) = 2\dot{g}(t)\ddot{g}(t)$  où naturellement la fonction  $\ddot{g}(t)$  est la dérivée de  $\dot{g}(t)$  par rapport au temps.

On observe les mêmes propriétés pour  $f'(t)$  que pour  $f(t)$  sur la figure 2.7, de ce fait on peut faire exactement la même démonstration pour montrer que  $\tilde{\gamma}(\tau)$  ne dépend pas de  $\tau_l$ . Cependant notons que sur la figure  $f'(t)$  a un support plus large que  $f(t)$  pour certains noyaux, notamment Dirichlet et Sinc qui sont quasiment confondus. Dans ces cas, la propriété 2) n'est qu'une approximation étant donné qu'un sinus cardinal oscille autour de 0 jusqu'à l'infini mais devient infiniment proche de zéro au fur et à mesure que l'on s'éloigne du centre de symétrie. De ce fait, on peut considérer que le résiduel  $\sum_{n=0}^{n_k-N_s-1} f((n_k-n)T_s)^2 + \sum_{n=N_s+1}^{N-1} f((n_k-n)T_s)^2$  est négligeable, et ce d'autant plus que l'on choisit un  $N_s$  grand, c'est à dire concrètement que les retards sont plus espacés les uns des autres.

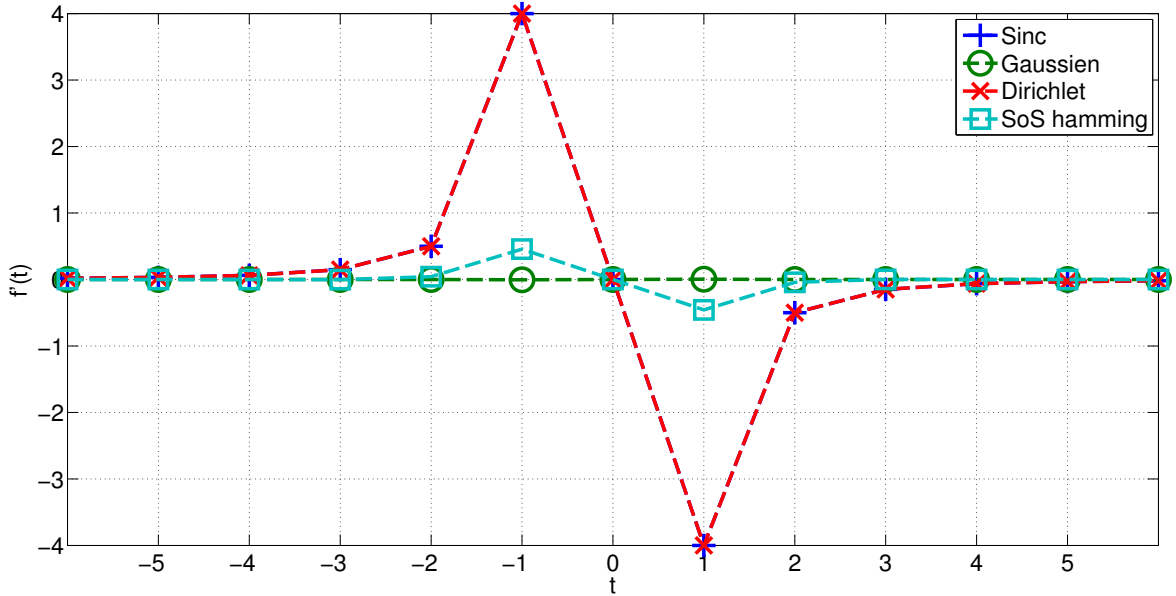


FIGURE 2.7 – Visualisation de  $f'(t)$ , utilisée pour montrer que la BCR sur les retards ne dépend pas du retard.

On utilise la notation  $\gamma(\tau) = \|\mathbf{g}_l\|^2 = \gamma_{\text{noyau}}$  et de  $\tilde{\gamma}(\tau) = \|\dot{\mathbf{g}}_l\|^2 = \tilde{\gamma}_{\text{noyau}}$  où l'on remplacera noyau par le nom du noyau que l'on considère. En effet, quand les retards sont assez espacés et éloignés des bords, les bornes ne dépendent que du choix du noyau  $g(t)$ . Cette hypothèse revient en quelque sorte à considérer l'estimation des paramètres pour chaque impulsion comme totalement indépendante des autres impulsions. On se retrouve donc dans une situation similaire au cas avec une seule impulsion.

**Résultat 4.** Pour des noyaux suivant approximativement les propriétés (2.45), (2.46) et (2.47) et des retards assez espacés et éloignés du bord, les bornes (2.43), (2.37) et (2.44) de-

viennent :

$$\text{BCR}(a_l) \approx \frac{\sigma^2}{\gamma_{\text{noyau}}} \quad (2.64)$$

$$\text{BCR}(\tau_l) \approx \frac{1}{\text{RSB}_l} \frac{1}{\tilde{\gamma}_{\text{noyau}}}. \quad (2.65)$$

pour donner finalement :

$$\text{BCR} \approx \frac{L\sigma^2}{\gamma_{\text{noyau}}} + \frac{1}{\tilde{\gamma}_{\text{noyau}}} \sum_{l=0}^{L-1} \frac{1}{\text{RSB}_l}. \quad (2.66)$$

On remarque que pour des noyaux normalisés suivant le critère de normalisation (1.26) on a  $\|\mathbf{g}\|^2 = 1$  et donc constant quel que soit le noyau. De ce fait, la BCR sur les amplitudes ne dépend pas du choix du noyau, mais uniquement de la normalisation et est simplement proportionnelle à la variance du bruit  $\sigma^2$  :  $\text{BCR}(a_l) = L\sigma^2$ . De ce fait pour optimiser la BCR totale le noyau le plus intéressant est celui minimisant la borne sur les retards, c'est à dire le noyau avec la plus grande norme sur sa dérivée. Il est intéressant de noter que le paramètre déterminant dans le choix du noyau, au regard de la BCR déterministe est la fonction dérivée du noyau  $g(t)$ , alors même que cette dérivée n'est pas intuitivement utile et n'apparaît pas directement dans la méthode d'échantillonnage, figure 1.2. Cette influence de la dérivée a également été pressentie par BENDORY, DEKEL et FEUER [14], puisqu'ils montrent que la distance minimale nécessaire entre deux impulsions dépend de la dérivée en zéro du noyau.

### Comparaison de quelques noyaux standards

Dans cette sous-partie nous allons tracer la BCR pour les différents noyaux que nous avons utilisés jusqu'à présent. Les bornes aux différentes étapes de simplification sont toutes tracées, afin de valider une nouvelle fois les expressions finales obtenues. La première BCR tracée est obtenue de manière numérique, par inversion de la matrice d'information de Fisher (2.3), ensuite nous allons tracer la BCR exacte analytique (2.44) et enfin la BCR obtenue pour des noyaux orthogonaux (2.58). Pour ces simulations nous utilisons  $N = 100$  échantillons, toujours  $T_S = 1$ . Les bornes sont tracées en fonction du RSB moyen qui est défini par  $\text{RSB} = \frac{\|\mathbf{a}\|^2}{\sigma^2}$ . Nous nous sommes placés dans des contextes où les délais sont espacés les uns des autres et éloignées de la bordure, de telle sorte que les expressions (2.66) et (2.58) sont équivalentes. Les figures 2.8, 2.9, 2.10 et 2.11 mettent en évidence les différentes expressions de la BCR pour le noyau Gaussien (en bleu), sinus cardinal (en rouge), Dirichlet (en noir) et enfin SoS avec une fenêtre de Hamming (en rouge). Pour chaque noyau on a tracé en ligne continue la BCR numérique (2.3), l'expression analytique (2.44) correspond aux étoiles tandis que l'expression approximée (2.58) est marquée par un carré.

Pour la figure 2.8 le signal  $x(t)$  est composé de  $L = 3$  impulsions situées aux temps  $\boldsymbol{\tau} = [20T_S, 60T_S, 85T_S]^T$  avec les amplitudes  $\mathbf{a} = [1, 2, 1]^T$ . C'est le cas favorable étant donné que les retards sont à la fois bien espacés, loin des bords et sur la grille. On peut donc observer dans ce cas que les trois expressions de la borne sont confondues ce qui valide l'approximation alors même que  $N$  est relativement petit par rapport aux valeurs qu'on pourrait avoir en pratique. Le noyau avec les meilleures performances pour des impulsions de Dirac est le noyau Sinus cardinal, suivi de près par le Dirichlet qui en est une approximation avec une fenêtre temporelle finie. Le sinus cardinal est le noyau optimal pour

ce type de signal, c'est celui qui a été utilisé pour montrer que les signaux à sommes d'impulsions sont échantillonnables parfaitement [27]. Le noyau gaussien quand à lui est bien en-dessous des autres, en effet la dérivée de la gaussienne est la seule des dérivée avec un support étroit (voir figure 2.1) donc elle a une norme plus faible sur tout le support. Le noyau SoS est très intéressant puisqu'il permet de rejoindre les performances du sinus cardinal tout en ayant un support fini, et peut également être modifié pour éventuellement s'adapter à d'autres formes d'impulsions, nous nous pencherons sur son utilisation dans le cas général dans la section 2.3.

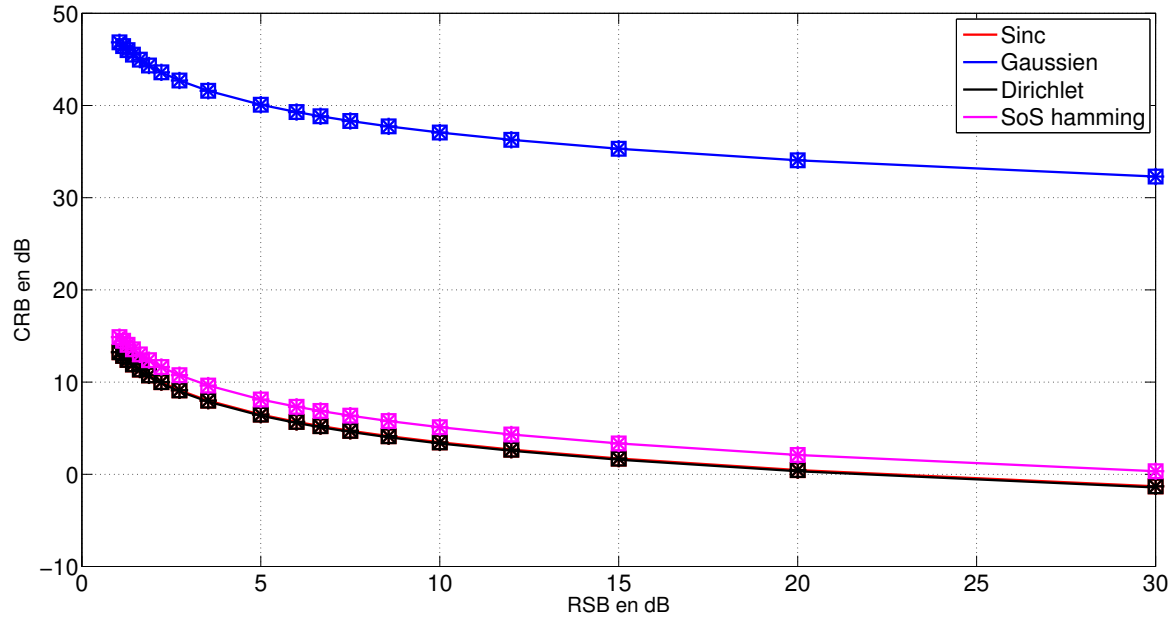


FIGURE 2.8 – Bornes pour 3 retards bien espacés sur la grille :  $\tau = [20, 60, 85]^T$  et  $\mathbf{a} = [1, 2, 1]^T$

La figure 2.9 utilise des retards qui ne tombent pas sur un multiple entier de  $T_s$  :  $\tau = [20.2, 60.6, 85.5]^T$ . Des problèmes dus à l'erreur de grille apparaissent donc au niveau du calcul de la borne. Cette erreur est importante pour le noyau gaussien, qui est très sensible aux erreurs de précision du calcul numérique, ce qui explique que même la borne analytique sans approximation ne correspond pas à la borne numérique. Cela laisse à penser qu'il ne vaut mieux pas utiliser le noyau gaussien pour l'échantillonnage des signaux impulsionnels car sa dérivée a un comportement instable dès que l'on s'éloigne de la grille. De plus ce noyau n'est pas du tout orthogonal dans ces conditions (voir figures 2.4 et 2.5), l'expression approximée est donc très éloignée de la vraie valeur de la BCR qui est fortement sous-estimée. Le noyau gaussien est un bon exemple de noyau ne vérifiant pas les conditions Strang-Fix, qui permettent d'assurer la reconstruction sans erreur des signaux à modèles paramétrique à fort RSB [113], il est intéressant de noter que nos résultats sont moins bons pour des noyaux ne vérifiant pas ces conditions. Les noyaux SoS et sinus cardinaux, qui contrairement au gaussien remplissent les conditions (voir section 1.2) ont toujours les trois bornes qui sont confondues malgré la présence de retards hors de la grille. Cette vérification est importante étant donné que dans la pratique les retards peuvent être n'importe où étant donné que ce sont des paramètres à valeurs continues.

On observe sur la figure 2.10 le résultat pour le cas particulier des retards très proches : on a considéré un signal composé de seulement deux retards  $\tau = [20, 22]^T$  avec des amplitudes différentes  $\mathbf{a} = [1, 2]^T$ . Comme nous l'avons vu dans les figures 2.2 et 2.3 les condi-

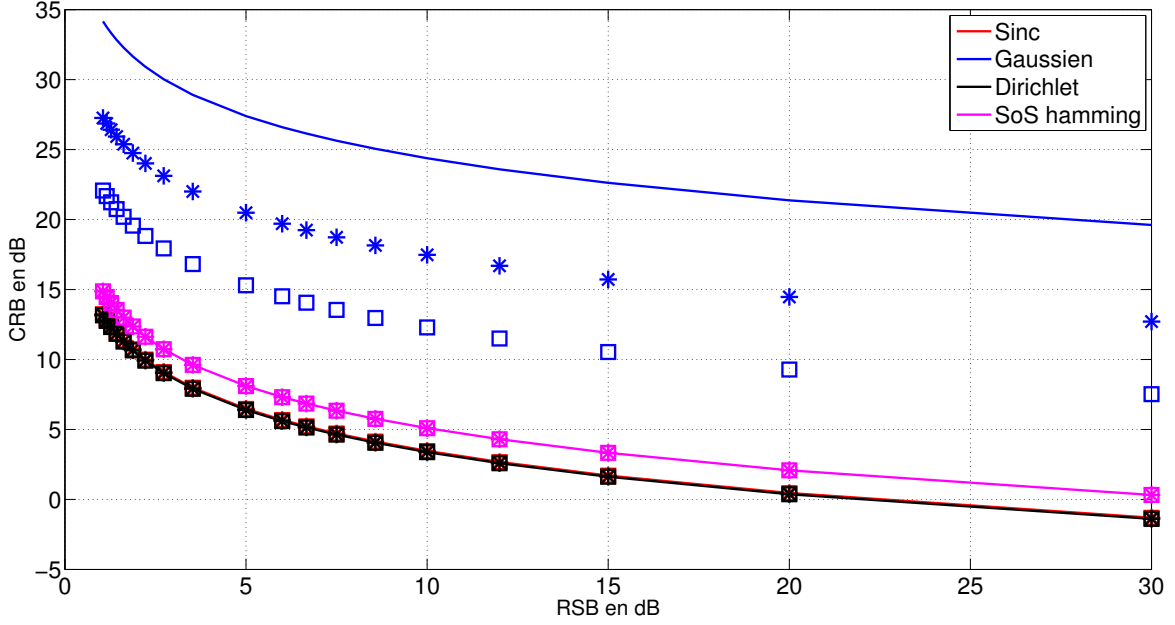


FIGURE 2.9 – Bornes pour 3 retards en dehors de la grille :  $\tau = [20.2, 60.6, 85.5]^T$  et  $\mathbf{a} = [1, 2, 1]^T$

tions d'orthogonalités ne peuvent être respectées que pour des retards assez éloignés. Les simulations confirment ce que nous attendions : les BCR numériques (2.3) et analytiques (2.44) sont bien confondues, tandis que la borne approximée (2.58) s'en éloigne légèrement, ce qui est logique puisque les conditions d'orthogonalité ne sont pas vérifiées dans ce cas précis.

La figure 2.11 considère un dernier cas limite, qui est peu rencontré en pratique notamment pour  $N$  grand : le cas des retards proches des bords. On prend donc des retards  $\tau = [1, 60, 99]^T$  en gardant les amplitudes  $\mathbf{a} = [1, 2, 1]^T$ . Le noyau gaussien ne figure pas sur la courbe, pour qu'on puisse voir de manière plus précise le comportement des autres noyaux. Le résultat observé est assez étonnant : en effet la borne analytique et son approximation sont confondues, ce qui montre que les propriétés d'orthogonalités restent bien vérifiées puisque la norme du noyau projeté est égale à la norme du noyau. Cependant la borne numérique obtenue en inversant directement la matrice est différente, il se peut que pour ces cas limites, les matrices utilisées pour le calcul soient mal conditionnées, ce qui provoque des erreurs dans le calcul numérique de l'inversion des matrices. Étant donné qu'on inverse une matrice à la fois pour le calcul de la BCR numérique, et pour les différentes projections nécessaires pour le calcul analytique on est soumis aux problèmes de l'inversion numérique des matrices ce qui peut fausser le résultat dans des cas limites.

En conclusion des simulations, on observe que l'expression analytique de la BCR est validée dans le cas général pour des noyaux respectant bien les conditions Strang-Fix, ce qui n'est pas le cas du noyau gaussien. Cela laisse à penser qu'il pourrait y avoir un lien entre les conditions d'orthogonalités et les conditions Strang-Fix. Ce lien serait intéressant à étudier, puisque les conditions d'orthogonalité sont simples à utiliser et à démontrer. L'expression analytique est validée également pour des retards continus mais restant assez éloignés des bords de l'échantillonnage pour que les matrices restent bien conditionnées, et ce dans le contexte difficile de la présence de plusieurs impulsions. De plus l'approximation donnée en (2.58) est très proche dans ces conditions de la borne numé-

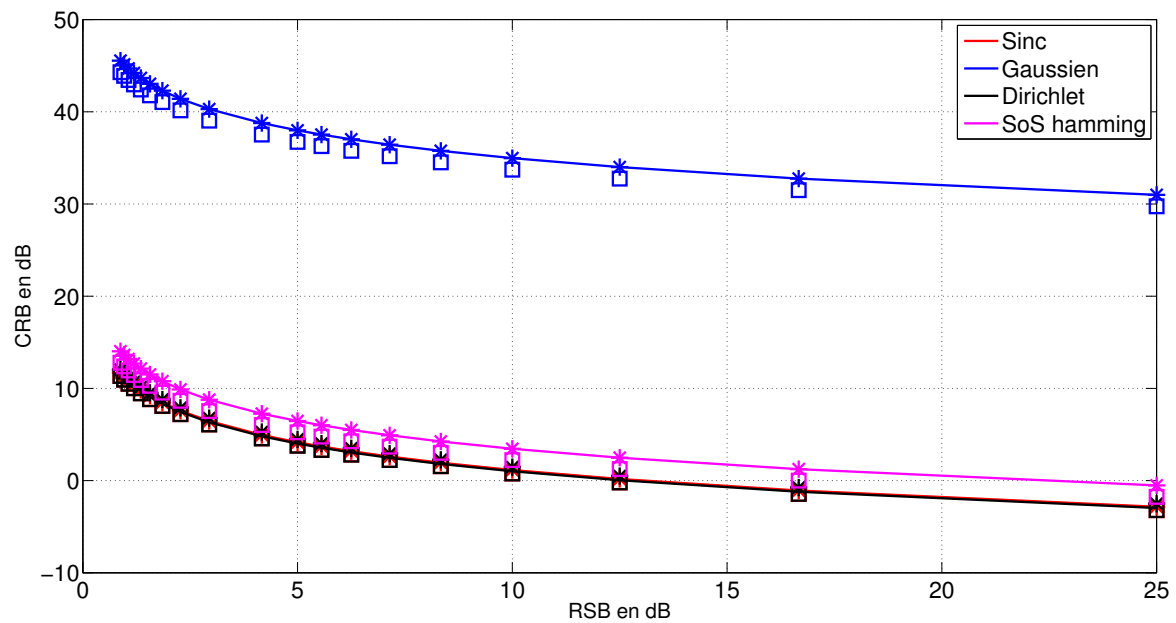


FIGURE 2.10 – Bornes pour des impulsions proches :  $\tau = [20, 22]^T$  et  $\mathbf{a} = [1, 2]^T$

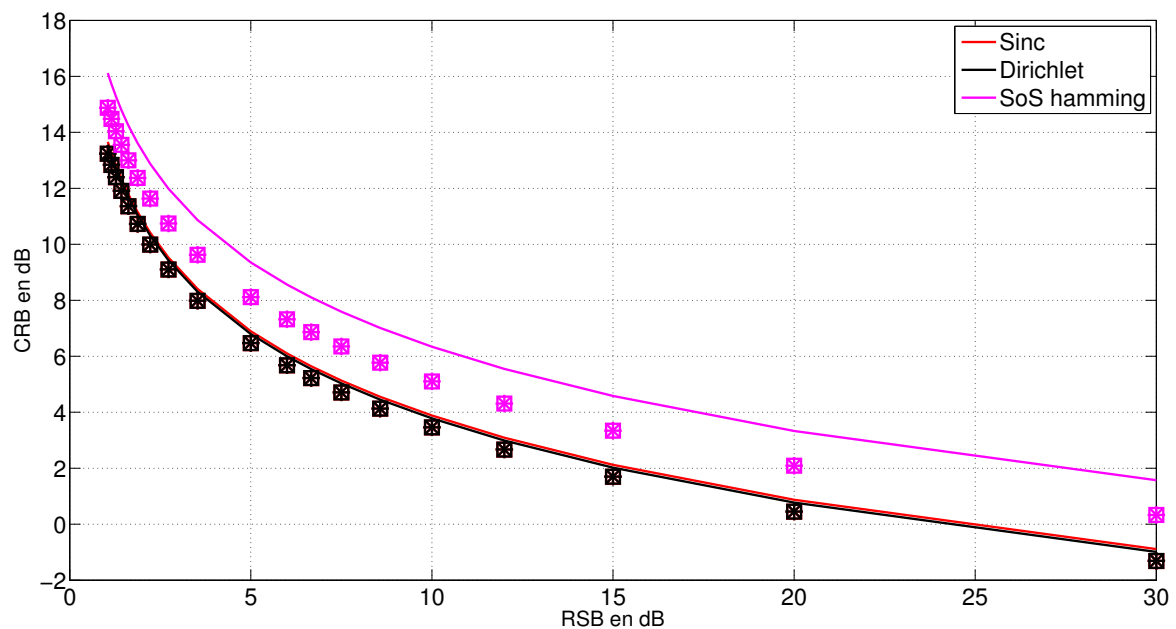


FIGURE 2.11 – Bornes pour des impulsions près des bords :  $\tau = [1, 60, 99]^T$  et  $\mathbf{a} = [1, 2, 1]^T$

rique, bien que le calcul en soit très simple. Par conséquent on peut avoir une première idée de la qualité d'un noyau au regard de la borne assez rapidement, simplement en observant la norme de la dérivée qui doit être la plus importante possible si l'on souhaite minimiser la BCR des retards. Cela minimise également la borne générale si l'on choisit de normaliser les noyaux, ce qui a pour conséquence que la BCR sur les amplitudes est la même quelque soit le noyau.

## 2.2.2 Estimation Bayésienne du signal à somme de Diracs

### Calcul de la borne de Van Trees

Dans la section 2.2.1 nous avons déterminé une expression déterministe de la BCR. La borne donnée en (2.44) est calculable pour une combinaison d'amplitudes et de retards donnés, et ne permet de comparer des noyaux que pour ce jeu de paramètre précis. De plus l'expression analytique obtenue est complexe et ne peut donc pas être utilisée par exemple pour optimiser les noyaux d'échantillonnages. Pour simplifier l'expression afin de la rendre plus facilement exploitable, nous avons dû faire un certain nombre d'hypothèses aussi bien sur la forme des noyaux que sur l'appartenance des retards à la grille pour que la BCR approximée ne soit plus dépendante des retards, et donc utilisable comme outil de comparaison général des noyaux. L'hypothèse de retards multiples de  $T_S$  est peu réaliste en pratique, en effet l'objectif de la méthode d'échantillonnage des signaux impulsionnels par convolution avec un noyau est de réduire le nombre d'échantillons, et donc d'utiliser un  $T_S$  grand ce qui réduit d'autant la probabilité que les retards soient proches des éléments de la grille. Pour pouvoir comparer et donc choisir un noyau d'échantillonnage de manière pertinente, nous voulons faire un choix indépendant de la connaissance a priori des paramètres à estimer. Nous nous intéressons donc à la borne Bayésienne qui est une moyenne des performances sur tous les tirages de paramètres possibles, sachant la distribution des paramètres.

Dans le contexte Bayésien, on considère que les paramètres à estimer suivent une distribution donnée. On considère que les  $a_l$  et les  $\tau_l$  sont indépendants et identiquement distribués (i.i.d.) et sont de plus indépendants entre eux. Cette hypothèse est courante, bien que réductrice étant donné qu'on considère que chaque impulsion est indépendante des autres. De cette indépendance, on tire la propriété :

$$p(\boldsymbol{\theta}|\mathbf{L}) = \prod_{l=1}^{2L} p(\theta_l|\mathbf{L}) = \prod_{l=1}^L p(a_l|\mathbf{L}) \prod_{l=1}^L p(\tau_l|\mathbf{L}). \quad (2.67)$$

On va s'intéresser au calcul de la borne  $C_L(\boldsymbol{\theta})$  qui est la BCRB connaissant le nombre d'impulsions  $L$ . On a par définition  $C_L(\boldsymbol{\theta}) = \text{Tr}(\mathbf{J}(\boldsymbol{\theta}|\mathbf{L})^{-1})$  (1.63). Elle s'écrit donc d'après (1.67) :

$$\mathbf{J}(\boldsymbol{\theta}|\mathbf{L}) = \mathbb{E}_{\boldsymbol{\theta}|\mathbf{L}} \{\mathbf{I}(\boldsymbol{\theta})\} + \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}|\mathbf{L}} \left\{ -\frac{\partial^2 \log p(\boldsymbol{\theta}|\mathbf{L})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} \right\}. \quad (2.68)$$

où la MIF  $\mathbf{I}(\boldsymbol{\theta})$  est celle dérivée dans la section 2.2.1, équation (2.11).

Le premier élément de l'addition est donc :

$$\mathbb{E}_{\boldsymbol{\theta}|\mathbf{L}} \{\mathbf{I}(\boldsymbol{\theta})\} = \frac{1}{\sigma^2} \begin{bmatrix} \mathbb{E}_{\boldsymbol{\theta}|\mathbf{L}} \{\mathbf{D}\dot{\mathbf{G}}^T \dot{\mathbf{G}}\mathbf{D}\} & \mathbb{E}_{\boldsymbol{\theta}|\mathbf{L}} \{\mathbf{D}\dot{\mathbf{G}}^T \mathbf{G}\} \\ \mathbb{E}_{\boldsymbol{\theta}|\mathbf{L}} \{\mathbf{G}^T \dot{\mathbf{G}}\mathbf{D}\} & \mathbb{E}_{\boldsymbol{\theta}|\mathbf{L}} \{\mathbf{G}^T \mathbf{G}\} \end{bmatrix} \quad (2.69)$$

La contribution des paramètres aléatoires quant à elle est :

$$\mathbf{J}_\theta = \mathbb{E}_{\theta|L} \left\{ -\frac{\partial^2 \log p(\theta|L)}{\partial \theta^T \partial \theta} \right\} = \int_{S_\theta} -\sum_{l=1}^{2L} \frac{\partial^2 \log p(\theta_l|L)}{\partial \theta^T \partial \theta} \prod_{l=1}^{2L} p(\theta_l|L) d\theta. \quad (2.70)$$

où l'on note  $S_A$  le support du paramètre  $A$  et où l'on a introduit la propriété d'indépendance des paramètres (2.67). La fonction à dériver ne dépend que d'un paramètre à la fois, c'est pourquoi tous les éléments hors de la diagonale sont nuls c'est à dire que  $\forall l \neq i, [\mathbf{J}_\theta]_{li} = 0$  tandis que sur la diagonale on a

$$\begin{aligned} [\mathbf{J}_\theta]_{ll} &= \int_{S_\theta} -\frac{\partial^2 \log p(\theta_l|L)}{\partial^2 \theta_l} \prod_{l=1}^{2L} p(\theta_l|L) d\theta \\ &= \int_{S_{\theta_l}} -\frac{\partial^2 \log p(\theta_l|L)}{\partial^2 \theta_l} p(\theta_l|L) d\theta_l \underbrace{\int_{S_{\theta^{-l}}} \prod_{i=1 \neq l}^{2L} p(\theta_i|L) d\theta^{-l}}_{=1} \\ &= \mathbb{E}_{\theta_l|L} \left\{ -\frac{\partial^2 \log p(\theta_l|L)}{\partial^2 \theta_l} \right\} \end{aligned} \quad (2.71)$$

La MIB dans le cas général est donc

$$\begin{aligned} \mathbf{J}(\theta|L) &= \frac{1}{\sigma^2} \begin{bmatrix} \mathbb{E}_{\theta|L} \{\mathbf{D} \dot{\mathbf{G}}^T \dot{\mathbf{G}} \mathbf{D}\} & \mathbb{E}_{\theta|L} \{\mathbf{D} \dot{\mathbf{G}}^T \mathbf{G}\} \\ \mathbb{E}_{\theta|L} \{\mathbf{G}^T \dot{\mathbf{G}} \mathbf{D}\} & \mathbb{E}_{\theta|L} \{\mathbf{G}^T \mathbf{G}\} \end{bmatrix} \\ &+ \text{diag} \left( \mathbb{E}_{\theta_1|L} \left\{ -\frac{\partial^2 \log p(\theta_1|L)}{\partial^2 \theta_1} \right\}, \dots, \mathbb{E}_{\theta_{2L}|L} \left\{ -\frac{\partial^2 \log p(\theta_{2L}|L)}{\partial^2 \theta_{2L}} \right\} \right) \end{aligned} \quad (2.72)$$

où l'on définit  $\text{diag}(\cdot)$  comme la matrice diagonale ayant le vecteur en argument comme diagonale. Nous noterons par la suite également  $\text{diag}(\mathbf{A})$  la diagonale de la matrice  $\mathbf{A}$ . Pour aller plus loin dans le calcul, on considère que l'on connaît les moments d'ordre un et deux de la distribution  $p(a_l)$  et que toutes les amplitudes suivent cette même distribution. La moyenne de  $a_l$  est notée  $\mu_a$  et la variance  $\sigma_a^2$ . Etant donnée l'indépendance des paramètres  $a_l$  et  $\tau_l$  on peut séparer les contributions des différents paramètres dans le calcul de la moyenne de la MIF sur les paramètres croisés :

$$\mathbb{E}_{\theta|L} \{\mathbf{D} \dot{\mathbf{G}}^T \mathbf{G}\} = \mathbb{E}_{\mathbf{a}|L} \{\mathbf{D}\} \mathbb{E}_{\tau|L} \{\dot{\mathbf{G}}^T \mathbf{G}\} = \mu_a \mathbf{I}_L \cdot \mathbb{E}_{\tau|L} \{\dot{\mathbf{G}}^T \mathbf{G}\} \quad (2.73)$$

$$\mathbb{E}_{\theta|L} \{\mathbf{G}^T \dot{\mathbf{G}} \mathbf{D}\} = \mathbb{E}_{\tau|L} \{\mathbf{G}^T \dot{\mathbf{G}}\} \mathbb{E}_{\mathbf{a}|L} \{\mathbf{D}\} = \mu_a \mathbf{I}_L \cdot \mathbb{E}_{\tau|L} \{\mathbf{G}^T \dot{\mathbf{G}}\} \quad (2.74)$$

On peut également détailler la moyenne de la MIF sur les retards en séparant le cas sur la diagonale du cas hors de la diagonale :

$$[\mathbb{E}_{\theta|L} \{\mathbf{D} \dot{\mathbf{G}}^T \dot{\mathbf{G}} \mathbf{D}\}]_{li} = \mu_a^2 \mathbb{E}_{\tau|L} \{\dot{\mathbf{g}}_l^T \dot{\mathbf{g}}_i\} \quad (2.75)$$

$$[\mathbb{E}_{\theta|L} \{\mathbf{D} \dot{\mathbf{G}}^T \dot{\mathbf{G}} \mathbf{D}\}]_{ll} = \sigma_a^2 \mathbb{E}_{\tau|L} \{\|\dot{\mathbf{g}}_l\|^2\} \quad (2.76)$$

**Résultat 5.** *Considérant des amplitudes distribuées suivant la même distribution  $p(a)$  de moyenne  $\mu_a$  et de variance  $\sigma_a^2$ , et des paramètres indépendants les uns des autres, la MIB est :*

$$\mathbf{J}(\theta|L) = \begin{bmatrix} \text{RSB}_1 \mathbb{E}_{\tau|L} \{\|\dot{\mathbf{g}}_1\|^2\} & \cdots & \frac{\mu_a^2}{\sigma_a^2} \mathbb{E}_{\tau|L} \{\dot{\mathbf{g}}_1^T \dot{\mathbf{g}}_L\} \\ \vdots & \ddots & \vdots \\ \frac{\mu_a^2}{\sigma_a^2} \mathbb{E}_{\tau|L} \{\dot{\mathbf{g}}_L^T \dot{\mathbf{g}}_1\} & \cdots & \text{RSB}_1 \mathbb{E}_{\tau|L} \{\|\dot{\mathbf{g}}_L\|^2\} \end{bmatrix} + \begin{bmatrix} \mathbf{D}_\tau & \frac{\mu_a}{\sigma_a^2} \mathbf{I}_L \cdot \mathbb{E}_{\tau|L} \{\dot{\mathbf{G}}^T \mathbf{G}\} \\ \frac{\mu_a}{\sigma_a^2} \mathbf{I}_L \cdot \mathbb{E}_{\tau|L} \{\mathbf{G}^T \dot{\mathbf{G}}\} & \frac{1}{\sigma_a^2} \mathbb{E}_{\theta|L} \{\mathbf{G}^T \mathbf{G}\} + \mathbf{D}_a \end{bmatrix} \quad (2.77)$$

où  $\mathbf{D}_a$  est la matrice diagonale ayant pour  $l$ -ième élément  $\mathbb{E}_{a_l|L} \left\{ -\frac{\partial^2 \log p(a_l|L)}{\partial^2 a_l} \right\}$  et  $\mathbf{D}_\tau$  est la matrice diagonale ayant pour  $l$ -ième élément  $\mathbb{E}_{\tau_l|L} \left\{ -\frac{\partial^2 \log p(\tau_l|L)}{\partial^2 \tau_l} \right\}$ .

Dans le cas fort RSB, on peut considérer que  $\sigma^2$  est très petit, et donc que  $\frac{1}{\sigma^2}$  est grand. Cela rend les composantes ne dépendant pas du RSB  $\mathbf{D}_a$  et  $\mathbf{D}_\tau$  négligeables [126]. On peut également simplifier la borne en considérant que la moyenne de la distribution des amplitudes  $\mu_a = 0$ .

**Résultat 6.** Pour un niveau de bruit faible, et une distribution des amplitudes de moyenne nulle, la MIB est

$$\mathbf{J}(\boldsymbol{\theta}|L) = \begin{bmatrix} \left[ \begin{array}{ccc} \text{RSB}_1 \mathbb{E}_{\tau|L} \left\{ \|\dot{\mathbf{g}}_1\|^2 \right\} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \text{RSB}_L \mathbb{E}_{\tau|L} \left\{ \|\dot{\mathbf{g}}_L\|^2 \right\} \end{array} \right] & \mathbf{0}_L \\ \mathbf{0}_L & \frac{1}{\sigma^2} \mathbb{E}_{\boldsymbol{\theta}|L} \{ \mathbf{G}^T \mathbf{G} \} \end{bmatrix}, \quad (2.78)$$

La matrice de BCR est simplement obtenue en inversant cette matrice bloc-diagonale : qui une fois inversée donne la borne pour un nombre d'impulsions  $l$  :

$$\mathbf{C}_l = \begin{bmatrix} \left[ \begin{array}{ccc} \frac{1}{\text{RSB}_1 \mathbb{E}_{\tau|L} \left\{ \|\dot{\mathbf{g}}_1\|^2 \right\}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\text{RSB}_L \mathbb{E}_{\tau|L} \left\{ \|\dot{\mathbf{g}}_L\|^2 \right\}} \end{array} \right] & \mathbf{0}_L \\ \mathbf{0}_L & \frac{\sigma^2}{\mathbb{E}_{\boldsymbol{\theta}|L} \{ \mathbf{G}^T \mathbf{G} \}} \end{bmatrix}. \quad (2.79)$$

La borne sur les retards est directement donnée par l'inversion de la MIB, étant donné que le bloc de la MIB concernant les retards est inversible. On a :

$$\text{BCRB}_{\tau_l} = \frac{1}{\text{RSB}_l \cdot \mathbb{E}_{\tau|L} \left\{ \|\dot{\mathbf{g}}_l\|^2 \right\}} \quad (2.80)$$

Si l'on considère comme pour les amplitudes que tout les retards sont identiquement distribués, l'espérance de la norme de la dérivée du noyau est la même quel que soit le retard considéré. En effet quelle que soit l'impulsion considérée, on se retrouve à calculer la même intégrale.

On cherche à présent à calculer la BCRB sur les amplitudes, pour laquelle on a besoin uniquement des éléments diagonaux de  $\mathbb{E}_{\boldsymbol{\theta}|L} \{ \mathbf{G}^T \mathbf{G} \}^{-1}$ . Pour ce faire nous allons à nouveau introduire la permutation  $\mathbf{P}_l$  permettant d'isoler chaque amplitude  $l$ . On s'intéresse donc à la BCR permutée, obtenue par l'inversion de la matrice

$$\mathbb{E}_{\boldsymbol{\theta}|L} \{ \mathbf{P}_l^T \mathbf{G}^T \mathbf{G} \mathbf{P}_l \} = \mathbb{E}_{\boldsymbol{\theta}|L} \left\{ \left[ \begin{array}{cc} \|\mathbf{g}_l\|^2 & \mathbf{g}_l^T \mathbf{G}^{(l)} \\ \mathbf{G}^{(l)T} \mathbf{g}_l & \mathbf{G}^{(l)T} \mathbf{G}^{(l)} \end{array} \right] \right\} \quad (2.81)$$

$$= \begin{bmatrix} \mathbb{E}_{\tau_l|L} \left\{ \|\mathbf{g}_l\|^2 \right\} & \mathbb{E}_{\tau|L} \{ \mathbf{g}_l^T \mathbf{G}^{(l)} \} \\ \mathbb{E}_{\tau|L} \{ \mathbf{G}^{(l)T} \mathbf{g}_l \} & \mathbb{E}_{\tau|L} \{ \mathbf{G}^{(l)T} \mathbf{G}^{(l)} \} \end{bmatrix} \quad (2.82)$$

En inversant le premier élément on obtient la borne sur l'amplitude  $a_i$  :

$$\left[ \mathbb{E}_{\boldsymbol{\theta}|L} \{ \mathbf{P}_l^T \mathbf{G}^T \mathbf{G} \mathbf{P}_l \}^{-1} \right]_{11} = \mathbb{E}_{\tau_l|L} \left\{ \|\mathbf{g}_l\|^2 \right\} - \mathbb{E}_{\tau|L} \left\{ \mathbf{g}_l^T \mathbf{G}^{(l)} \right\} \mathbb{E}_{\tau|L} \left\{ \mathbf{G}^{(l)T} \mathbf{G}^{(l)} \right\}^{-1} \mathbb{E}_{\tau|L} \left\{ \mathbf{G}^{(l)T} \mathbf{g}_l \right\} \quad (2.83)$$

**Résultat 7.** La BCRB sur les retards et les amplitudes est :

$$\text{BCRB}_{\tau} = \frac{1}{\text{RSB}_l \cdot \mathbb{E}_{\tau} \left\{ \|\dot{\mathbf{g}}\|^2 \right\}} \quad (2.84)$$

$$\text{BCRB}_{a_l} = \sigma^2 \left[ \mathbb{E}_{\tau|L} \left\{ \|\mathbf{g}_l\|^2 \right\} - \mathbb{E}_{\tau|L} \left\{ \mathbf{g}_l^T \mathbf{G}^{(l)} \right\} \mathbb{E}_{\tau|L} \left\{ \mathbf{G}^{(l)T} \mathbf{G}^{(l)} \right\}^{-1} \mathbb{E}_{\tau|L} \left\{ \mathbf{G}^{(l)T} \mathbf{g}_l \right\} \right] \quad (2.85)$$

La borne totale sur tout les  $L$  impulsions est donc :

$$C_L(\boldsymbol{\theta}) = \text{LBCRB}_{\tau} + \sum_{l=1}^L \text{BCRB}_{a_l} \quad (2.86)$$

Contrairement à la borne sur les retards qui prend une forme assez simple, la borne sur les amplitudes garde une forme difficile à exploiter. Les espérances sur le paramètre  $\tau$  demandent le calcul d'une intégrale pour chaque retard considéré, quand le nombre de retard augmente ce calcul est très vite très complexe à effectuer numériquement. Pour cette raison nous nous sommes concentrés sur l'optimisation uniquement de la borne sur les retards, cela montre un exemple de la façon dont on peut choisir les paramètres du noyau SoS.

### Application au noyau SoS

On s'intéresse à présent à la borne sur les retards (2.84) appliquée au noyau SoS (1.17). Notre objectif est d'isoler l'influence des paramètres  $b_k$  sur la borne afin de choisir les coefficients permettant de la minimiser. Minimiser la borne permet de minimiser l'erreur quadratique moyenne optimale. La démarche de minimiser la BCR pour choisir le noyau dans le cadre de la reconstruction des signaux impulsionnels est également proposée par URIGUEN, BLU et DRAGOTTI [123]. Dans cet article les auteurs se sont intéressés aux noyaux reproducteurs d'exponentielles tandis que nous nous focalisons sur le noyau SoS.

Nous allons tout d'abord exprimer (1.17) sous forme matricielle, en faisant l'hypothèse qu'on a un noyau réel c'est à dire que  $b_k = b_{-k}^*$  :

$$g(t) = \text{rect}\left(\frac{t}{NT_S}\right) \sum_{k=-K/2}^{K/2} b_k e^{\frac{j2\pi kt}{NT_S}} \quad (2.87)$$

$$\begin{aligned} &= \text{rect}\left(\frac{t}{NT_S}\right) \left( b_0 + \sum_{k=1}^{K/2} b_k \left( e^{\frac{-j2\pi kt}{NT_S}} + e^{\frac{j2\pi kt}{NT_S}} \right) \right) \\ &= \text{rect}\left(\frac{t}{NT_S}\right) \left( b_0 + \sum_{k=1}^{K/2} 2b_k \cos\left(\frac{2\pi kt}{NT_S}\right) \right) \end{aligned} \quad (2.88)$$

$$= \mathbf{b}^T \mathbf{v}(t) \quad (2.89)$$

où tous les coefficients à déterminer sont dans le vecteur  $\mathbf{b} = [b_0, \dots, b_{K/2}]^T$  et  $\mathbf{v}(t) = [v_0(t), \dots, v_{K/2}(t)]^T$  avec

$$\begin{cases} v_k(t) = 2\text{rect}\left(\frac{t}{NT_S}\right) \cos\left(\frac{2\pi kt}{NT_S}\right) \text{ for } k \neq 0 \\ v_0(t) = \text{rect}\left(\frac{t}{NT_S}\right). \end{cases} \quad (2.90)$$

De la même façon on peut exprimer la fonction dérivée sous forme matricielle :

$$\dot{\mathbf{g}}(t) = \frac{\partial \mathbf{g}(t)}{\partial t} = \mathbf{b}^T \frac{\partial \mathbf{v}(t)}{\partial t} \quad (2.91)$$

Etant donné que la dérivée de la fonction rect n'est pas définie, on l'approxime avec une gaussienne généralisée :

$$v_l(t) = 2e^{-\left(\frac{2|t|}{NT_s}\right)^\beta} \cos\left(\frac{2\pi l t}{NT_s}\right) \quad (2.92)$$

En utilisant l'expression de la valeur absolue  $|t| = (t^2)^{1/2}$ , on obtient la dérivée

$$\frac{\partial v_k(t)}{\partial t} = -2\text{rect}\left(\frac{t}{NT_s}\right) \left[ \left(\frac{2}{NT_s}\right)^\beta \beta t |t|^{\beta-2} \cos\left(\frac{2\pi k t}{NT_s}\right) + \frac{2\pi k}{NT_s} \sin\left(\frac{2\pi k t}{NT_s}\right) \right] \quad (2.93)$$

et pour  $k = 0$  :

$$\frac{\partial v_0(t)}{\partial t} = -\left(\frac{2}{NT_s}\right)^\beta \beta t |t|^{\beta-2} \text{rect}\left(\frac{t}{NT_s}\right) \quad (2.94)$$

On peut à présent exprimer la borne (2.84) pour ce noyau :

$$\begin{aligned} \mathbb{E}_\tau \left\{ \|\dot{\mathbf{g}}\|^2 \right\} &= \mathbb{E}_\tau \left\{ \dot{\mathbf{g}}^T \dot{\mathbf{g}} \right\} \\ &= \mathbf{b}^T \underbrace{\mathbb{E}_\tau \left\{ \sum_{n=0}^{N-1} \frac{\partial \mathbf{v}(\tau - nT_s)}{\partial (\tau - nT_s)} \cdot \frac{\partial \mathbf{v}^T(\tau - nT_s)}{\partial (\tau - nT_s)} \right\}}_{\dot{\Delta}} \mathbf{b} \end{aligned} \quad (2.95)$$

d'où

$$\text{BCRB}_{\tau_i} = \frac{1}{\text{RSB}_i \cdot \mathbf{b}^T \dot{\Delta} \mathbf{b}} \quad (2.96)$$

où  $\dot{\Delta}$  est une matrice carré et symétrique de dimension  $K/2 + 1$ .

Pour minimiser la BCRB, on souhaite maximiser la fonction quadratique  $\mathbf{b}^T \dot{\Delta} \mathbf{b}$  qui ne dépend pas de  $\tau$ . De plus on souhaite obtenir un noyau normalisé (1.26) c'est à dire que

$$\sum_{n=-(N-1)/2}^{(N-1)/2} g(nT_s)^2 = \mathbf{b}^T \underbrace{\sum_{n=-(N-1)/2}^{(N-1)/2} \mathbf{v}(nT_s) \cdot \mathbf{v}(nT_s)^T}_{\Delta} \mathbf{b} = 1 \quad (2.97)$$

où similairement à  $\dot{\Delta}$ ,  $\Delta$  est une matrice symétrique de dimension  $K/2 + 1$ . Le problème d'optimisation devient finalement

$$\max_{\mathbf{b}} \mathbf{b}^T \dot{\Delta} \mathbf{b} \text{ t.q. } \mathbf{b}^T \Delta \mathbf{b} = 1. \quad (2.98)$$

Une optimisation pour un problème tout à fait similaire est proposée en SHAHBAZPANAHI et al. [108]. La fonction de Lagrange obtenue est

$$L_\lambda(\mathbf{b}, \lambda) = \mathbf{b}^T \dot{\Delta} \mathbf{b} + \lambda (1 - \mathbf{b}^T \Delta \mathbf{b}) \quad (2.99)$$

où  $\lambda$  est un multiplicateur de Lagrange. En prenant le gradient de (2.99) et en l'annulant on trouve que la solution de (2.98) est donnée par le problème de valeur propre généralisée :

$$\dot{\Delta} \mathbf{b} = \lambda \Delta \mathbf{b}. \quad (2.100)$$

On a la propriété que les matrices  $\hat{\Delta}$  et  $\Delta$  sont symétriques par construction, de plus  $\Delta = \mathbf{V}\mathbf{V}^T$  où  $\mathbf{V}$  est la matrice  $(K/2 + 1) \times N$  ayant les vecteurs  $\mathbf{v}_n$  pour colonnes. Etant donné que  $\mathbf{V}$  est de rang  $K/2 + 1$ , la matrice  $\Delta$  est définie positive c'est à dire que toutes ses valeurs propres sont supérieures à zéro (propriété (A.5.1) de [111]). Ces deux propriétés permettent à (2.100) d'entrer dans le cadre des problèmes de valeurs propres généralisées qui a des méthodes de résolutions standard [63, 66]. Une de ces méthodes propose de se baser sur une décomposition de Cholesky, c'est celle utilisée par défaut par Matlab quand on utilise la fonction `eig` avec deux matrices symétriques. Nous allons donc utiliser cette fonction pour trouver les vecteurs propres qui sont les différents vecteurs de coefficients candidats. Etant donné que nous cherchons à maximiser une fonction quadratique, la solution de (2.98) est le vecteur propre généralisé associé à la plus grande valeur propre généralisée [108].

Le vecteur  $\mathbf{b}_{\text{opt}}$  obtenu pour  $N = 101$  et  $T_S = 1$  est présenté dans la figure 2.12. Les calculs ont été faits sur Matlab de manière numérique. Pour calculer l'espérance sur  $\tau$  nécessaire pour obtenir la matrice  $\hat{\Delta}$  nous avons utilisé la fonction `trapz` avec un pas de 0.01 secondes et des  $\tau \in [0, NT_S]$ . La gaussienne généralisée a été paramétrée avec  $\beta = 100$  pour approximer la fonction `rect`, et le noyau a une largeur en fréquence de  $K = N$ . Le noyau  $g_{\text{opt}}(t)$  correspondant ainsi que sa dérivée sont présentés sur les figures 2.13 et 2.14 respectivement. Les courbes sont obtenues exactement de la même manière que dans les figures 1.3 et 2.1 : les croix correspondent aux points réellement utilisés une fois le noyau échantillonné, tandis que le trait continu montre le noyau de manière continue. Le noyau de Dirichlet est présent afin de permettre la comparaison.

Les coefficients du noyau sont symétriques et ont un pic dans les hautes fréquences. Ce pic induit une oscillation une fois le noyau temporel considéré, la fréquence de l'oscillation dépendant de l'indice du pic. Le noyau optimal est beaucoup moins compact dans le domaine temporel que le standard noyau de Dirichlet. Il faut noter qu'il est optimal seulement considérant les performances sur le retard moyenné. Il est assez intuitif en observant la dérivée que ce noyau a une dérivée moyenne bien plus importante que le Dirichlet ce qui induit une borne plus petite : en effet l'oscillation importante de la dérivée induit une norme grande. Les valeurs de la MIB et de la BCRB sur les retards pour les deux noyaux présentés sont comparés dans le tableau 2.1 où l'on considère un RSB de 1 :  $\text{MIB}_\tau = \mathbf{b}^T \hat{\Delta} \mathbf{b}$  et  $\text{BCRB}_\tau = 1/\text{MIB}_\tau$ . La MIB optimisée est bien plus grande que la MIB pour le noyau de Dirichlet, on a bien obtenu un noyau pour lequel ce paramètre est maximal. Il faut noter cependant que nous n'avons pas du tout considéré l'estimation des amplitudes, étant donné que la borne sur les amplitudes a une forme trop compliquée pour être exploitée. Cependant ce noyau peut être intéressant pour les applications où les temps d'arrivée des impulsions sont les paramètres les plus importants. C'est le cas par exemple en radar, où à partir du retard à l'arrivée d'un signal connu envoyé à un instant précis, on estime la distance d'une cible par rapport au radar [81].

TABLEAU 2.1 – MIB et BCRB sur les retards pour les noyaux SoS de Dirichlet et optimal.

|           | $\text{MIB}_\tau$ | $\text{BCRB}_\tau$   |
|-----------|-------------------|----------------------|
| Dirichlet | 326.5             | $3.1 \times 10^{-3}$ |
| Optimal   | 857.7             | $1.2 \times 10^{-3}$ |

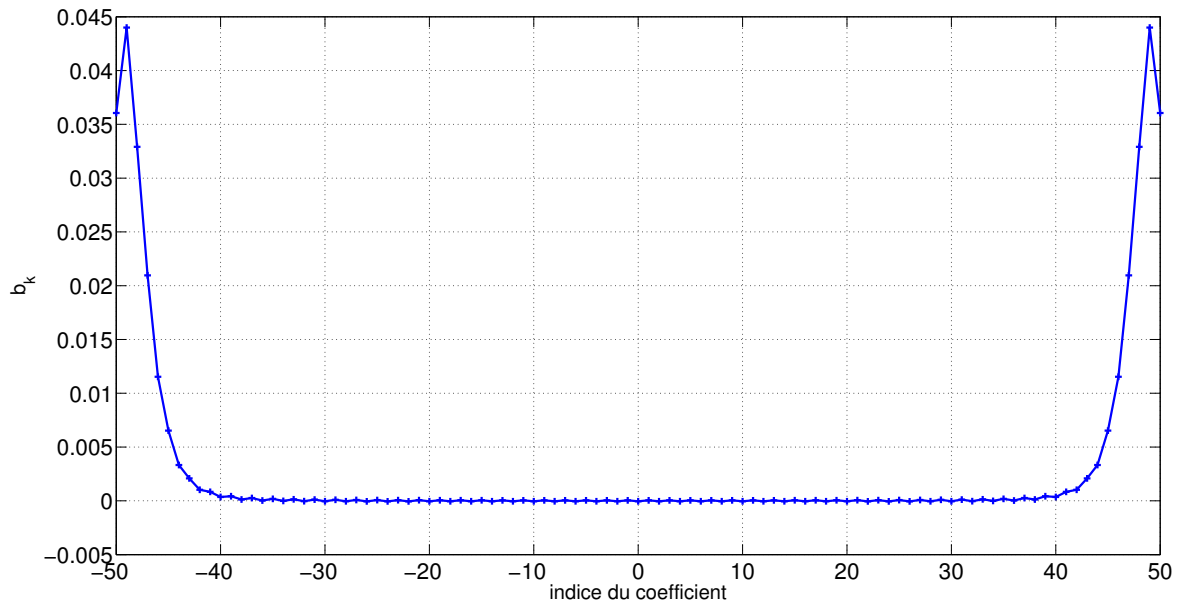


FIGURE 2.12 – Coefficients obtenus après l'optimisation de (2.98), pour  $N = 101$  et  $T_S = 1$

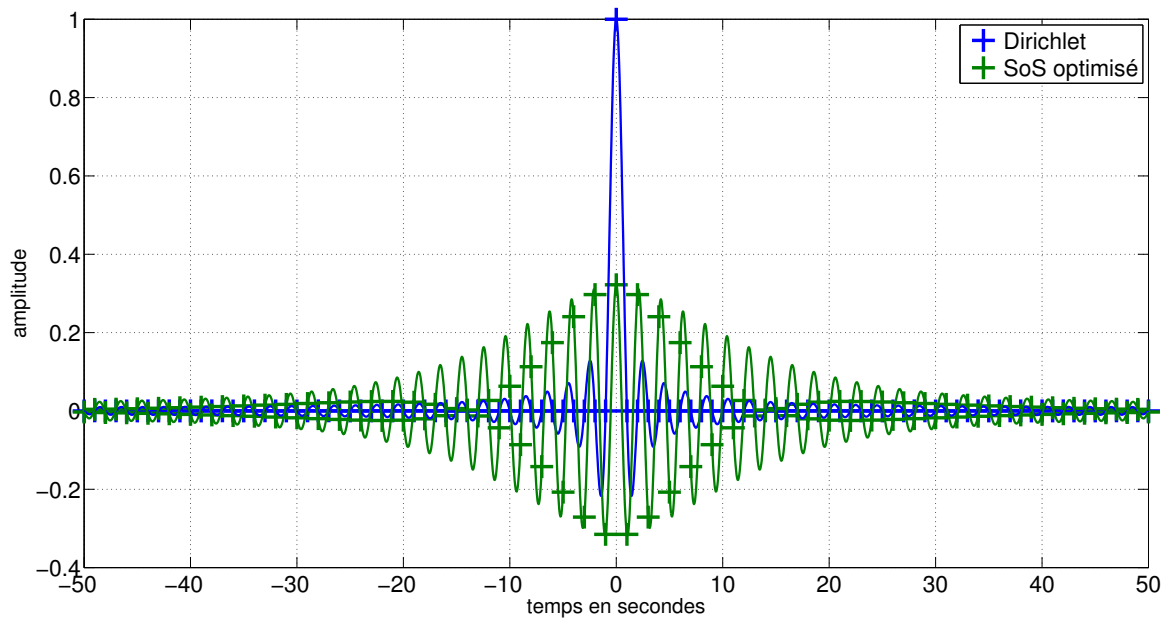


FIGURE 2.13 – Noyau optimal et de Dirichlet correspondant en temps

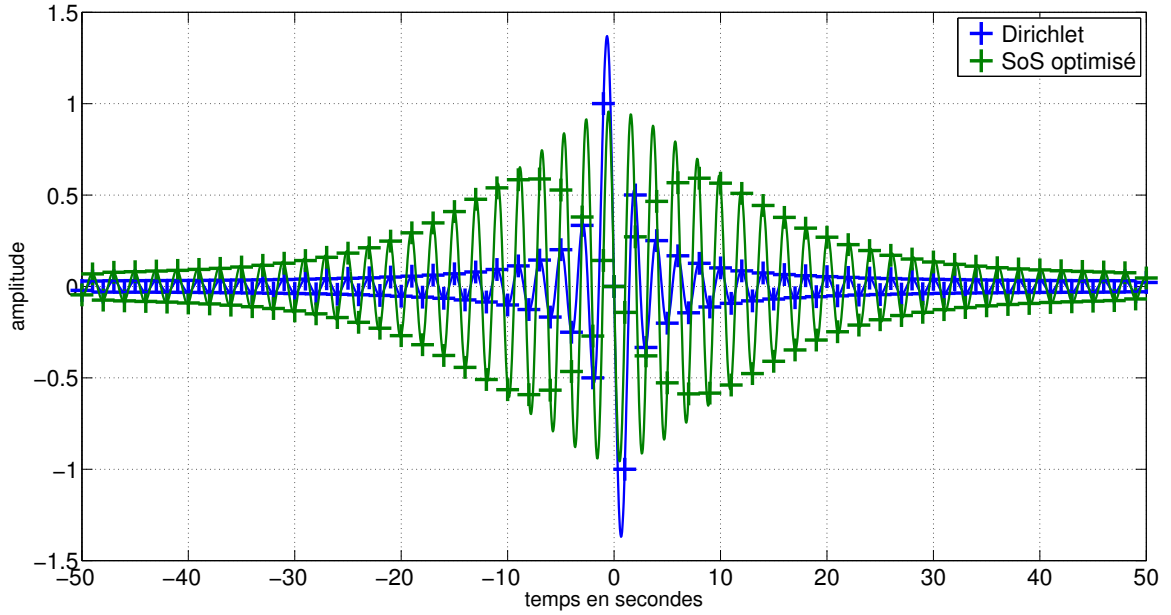


FIGURE 2.14 – Dérivée du noyau optimal et de Dirichlet

## 2.3 Cas du noyau à sommes de sinus cardinaux

Nous nous intéressons à présent à la borne pour tout type de forme d'impulsion  $h(t)$  en nous restreignant au noyau SoS. Cela permet d'introduire directement la forme du noyau (1.17) dans l'expression des échantillons de mesures (1.7) et de trouver une forme analytique pour la borne Bayésienne. Le noyau SoS est intéressant à utiliser dans ce but car il permet en choisissant les paramètres avec soin d'approximer des noyaux analogiques connus tels que le noyau en sinus cardinal tout en maintenant un support en temps fini. De ce fait l'utilisation du noyau SoS reste assez générale malgré la restriction à un modèle assez précis. Une fois la BCRB dérivée, nous nous intéressons à sa minimisation pour trouver le noyau optimal, suivant des contraintes différentes. Nous montrons qu'un noyau sinusoïdal est optimal au sens de la borne bayésienne en l'absence de bruit, ce qui est contre-intuitif au regard de la forme des noyaux habituellement utilisés. Une grande partie de cette section est tirée de l'article BERNHARDT et al. [24], mais nous rajoutons ici d'autres contraintes d'optimisation qui aboutissent à des noyaux optimaux différents.

### 2.3.1 Borne bayésienne pour un signal à série d'impulsions de forme quelconque

Le problème d'estimation introduit dans la section 1.2.2 peut être totalement reformulé sous une forme matricielle, en isolant le vecteur de paramètres  $\mathbf{b}$  tel que présenté dans TUR, ELDAR et FRIEDMAN [120].

Les coefficients de la transformée de Fourier du signal  $x(t)$  sont :

$$X[k] = \frac{1}{NT_s} H\left(\frac{2\pi k}{NT_s}\right) \sum_{l=0}^{L-1} a_l e^{\frac{-2\pi k \tau_l}{NT_s}}. \quad (2.101)$$

Le noyau SoS que l'on utilise a une largeur de bande en fréquence de  $K+1$  par sa définition

(1.16). De ce fait après échantillonnage seuls les  $K + 1$  éléments de la transformée de Fourier de  $\tilde{x}(t)$  correspondants sont utiles. On fait l'hypothèse que la largeur de bande choisie correspond à la largeur de bande de l'impulsion  $h(t)$  c'est à dire qu'on a  $H\left(\frac{2\pi kt}{NT_s}\right) \neq 0$  pour  $k$  allant de  $-K/2$  à  $K/2$ . On peut à présent réécrire (2.101) sous forme matricielle par le vecteur  $\mathbf{x}$  de longueur  $K + 1$  dont le  $k$ -ième élément est  $X[k]$  qui peut s'écrire :

$$\mathbf{x} = \mathbf{H}\mathbf{V}(\boldsymbol{\tau})\mathbf{a}, \quad (2.102)$$

où  $\mathbf{V}(\boldsymbol{\tau})$  est la matrice  $(K + 1) \times L$  dont le  $(k, l)$ -ième élément est  $e^{\frac{-j2\pi k\tau_l}{NT_s}}$ ,  $\mathbf{H}$  est la matrice diagonale de dimensions  $(K + 1) \times (K + 1)$  dont le  $k$ -ième élément est  $h_k = \frac{1}{NT_s} H\left(\frac{2\pi k}{NT_s}\right)$ . Tout comme dans (1.8) nous mettons le vecteur des échantillons sous forme vectorielle  $\mathbf{c} = [c_0, \dots, c_{N-1}]^T$  et  $\boldsymbol{\mu} = [\mu_0, \dots, \mu_{N-1}]^T$ .

De plus le noyau SoS a une bande limitée de par la présence de la fonction rect, la somme infinie (1.13) devient donc la somme finie :

$$\mu_n = \sum_{\substack{k=-K/2 \\ k \in \mathbb{Z}}}^{K/2} X[k] e^{\frac{2\pi kn}{N}} G\left(\frac{2\pi k}{NT_s}\right) \quad (2.103)$$

$$= \frac{NT_s}{\sqrt{2\pi}} \sum_{\substack{k=-K/2 \\ k \in \mathbb{Z}}}^{K/2} b_k X[k] e^{\frac{2\pi kn}{N}} \quad (2.104)$$

Cette expression peut être mise sous forme matricielle en introduisant les matrices  $\mathbf{V}(-\mathbf{t}_s)$  qui est une matrice  $N \times (K + 1)$  dont le  $(n, k)$ -ième élément est  $e^{j2\pi kn/N}$ , et  $\mathbf{B}$  qui est une matrice diagonale de dimensions  $K + 1 \times K + 1$  ayant  $\{b_k\}$  comme éléments diagonaux :

$$\boldsymbol{\mu} = \frac{NT_s}{\sqrt{2\pi}} \mathbf{V}(-\mathbf{t}_s) \mathbf{B} \mathbf{x} \quad (2.105)$$

$$= \frac{NT_s}{\sqrt{2\pi}} \mathbf{V}(-\mathbf{t}_s) \mathbf{B} \mathbf{H} \mathbf{V}(\boldsymbol{\tau}) \mathbf{a} \text{ en introduisant (2.102)} \quad (2.106)$$

$$\triangleq \boldsymbol{\Delta} \mathbf{V}(\boldsymbol{\tau}) \mathbf{a} \quad (2.107)$$

où  $\mathbf{t}_s = \{nT_s : 0 \leq n \leq N - 1, n \in \mathbb{N}\}$  est le vecteur des temps d'échantillonnage. On a défini la matrice  $\boldsymbol{\Delta}$  comme la matrice comprenant tous les éléments qui sont connus, il est donc possible de la construire à partir du moment où on a choisi le vecteur  $\mathbf{b}$ . Cette expression permet d'isoler les paramètres connus des paramètres  $\boldsymbol{\tau}$  et  $\mathbf{a}$  à estimer.

Il est à présent possible de dériver la borne Bayésienne pour le modèle (2.106). Nous étudions le problème de l'estimation du vecteur de paramètres inconnus  $\boldsymbol{\theta} = [\boldsymbol{\tau}^T \mathbf{a}^T]^T$  basé sur un nombre fini d'échantillons  $N$ , obtenus uniformément avec un intervalle d'échantillonnage  $T_s$ . Nous nous intéressons au cas Bayésien où les paramètres d'amplitudes  $a_l$  sont des variables aléatoires indépendantes et identiquement distribuées suivant une distribution centrée  $p(a_l)$  de variance  $\sigma_a^2$ , tandis que les paramètres de retard  $\tau_l$  sont indépendamment et identiquement distribués suivant une distribution de gaussienne généralisée (notée GN) de paramètre de position nul, de paramètre d'échelle  $\alpha > 0$  et de paramètre de forme  $\beta > 0$  notée  $\tau_l \sim \mathcal{GN}(0, \alpha, \beta)$ , on note  $\sigma_\tau^2$  la variance de  $p(\tau_l)$ . La distribution GN est une distribution intéressante car elle est générale, et englobe les distributions courantes laplaciennes ( $\beta = 1$ ), gaussiennes ( $\beta = 2$ ) et uniformes ( $\beta \rightarrow \infty$ ) [88]. De plus elle permet d'exprimer la distribution uniforme par une fonction continue, ce qui la rend dérivable et permet d'en tirer des bornes de Cramer-Rao.

On se place dans le cas fort RSB pour lequel la MIB est simplement l'espérance sur les paramètres à estimer de la MIF. La restriction au cas fort RSB est pertinente étant donné que nous souhaitons dériver la borne dans le but de choisir un noyau optimal. Le second terme de la MIB (1.67) ne dépend pas de la forme du noyau, de ce fait il n'influencera pas l'optimisation, c'est pourquoi il n'est pas nécessaire de le développer. De plus l'influence du noyau est visible surtout à fort RSB, car pour un bruit faible les erreurs d'estimations deviennent importantes et l'influence du noyau sur ces erreurs devient négligeable. Étant donné que le bruit additif  $w_n$  est gaussien, les échantillons conditionnés à un tirage donné suivent une distribution gaussienne  $\mathbf{c}|\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_N)$ . On peut donc utiliser la formule de Slepian-Bang pour le calcul de la MIF :

$$\mathbf{J}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \mathbb{E}_{\boldsymbol{\theta}} \left\{ \left[ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} \right] \right\} \quad (2.108)$$

La dérivée partielle relative aux amplitudes est :

$$\frac{\partial \boldsymbol{\mu}}{\partial a_l} = \Delta \mathbf{V}(\tau_l) \quad (2.109)$$

où  $\mathbf{V}(\tau_l)$  est un vecteur colonne de longueur  $K+1$  correspondant à l'extraction de la  $l$ -ième colonne de  $\mathbf{V}$  :

$$\mathbf{V}(\tau_l) = e^{\frac{-j2\pi \mathbf{k} \tau_l}{NT_S}} \quad \text{avec } \mathbf{k} = \left[ -\frac{K}{2}, \dots, \frac{K}{2} \right]^T, k \in \mathbb{Z}, \quad (2.110)$$

tandis que la dérivée partielle relative aux retards est :

$$\frac{\partial \boldsymbol{\mu}}{\partial \tau_l} = \Delta \frac{\partial \mathbf{V}(\tau)}{\partial \tau_l} \mathbf{a} = \Delta \dot{\mathbf{V}}(\tau_l) a_l \quad (2.111)$$

où la dérivée

$$\dot{\mathbf{V}}(\tau_l) = \frac{-j2\pi \mathbf{k}}{NT_S} \cdot e^{\frac{-j2\pi \mathbf{k} \tau_l}{NT_S}}. \quad (2.112)$$

où  $\cdot$  représente le produit de Hadamard.

La MIF peut s'écrire sous la forme d'une matrice bloc :

$$\left[ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} \right] = \begin{bmatrix} \left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau} \right] & \left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{a}} \right] \\ \left[ \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{a}} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau} \right] & \left[ \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{a}} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{a}} \right] \end{bmatrix} \quad (2.113)$$

En détail les différents éléments de cette matrice sont :

$$\left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau_l} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau_{l'}} \right] = a_l \dot{\mathbf{V}}(\tau_l)^H \Delta^H \Delta \dot{\mathbf{V}}(\tau_{l'}) a_{l'} \quad (2.114)$$

$$\left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau_l} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial a_{l'}} \right] = a_{l'} \dot{\mathbf{V}}(\tau_l)^H \Delta^H \Delta \mathbf{V}(\tau_{l'}) \quad (2.115)$$

$$\left[ \frac{\partial \boldsymbol{\mu}}{\partial a_l} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau_{l'}} \right] = \mathbf{V}(\tau_l)^H \Delta^H \Delta \dot{\mathbf{V}}(\tau_{l'}) a_{l'} \quad (2.116)$$

$$\left[ \frac{\partial \boldsymbol{\mu}}{\partial a_l} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial a_{l'}} \right] = \mathbf{V}(\tau_l)^H \Delta^H \Delta \mathbf{V}(\tau_{l'}) \quad (2.117)$$

La matrice  $\Delta^H \Delta$  est intéressante car elle apparait dans chaque élément de la MIF. On a

$$\Delta^H \Delta = \frac{N^2 T_S^2}{2\pi} \mathbf{H} \mathbf{B} \mathbf{V}(-\mathbf{t}_S)^H \mathbf{V}(-\mathbf{t}_S) \mathbf{B} \mathbf{H}. \quad (2.118)$$

Or le produit des matrices  $\mathbf{V}(-\mathbf{t}_S)^H \mathbf{V}(-\mathbf{t}_S)$  se réduit simplement :

$$[\mathbf{V}(-\mathbf{t}_S)^H \mathbf{V}(-\mathbf{t}_S)]_{i_1 i_2} = \sum_{n=0}^{N-1} e^{-\frac{j2\pi i_1 n}{N}} e^{\frac{j2\pi i_2 n}{N}} = \sum_{n=0}^{N-1} e^{\frac{j2\pi(i_2-i_1)n}{N}} \quad (2.119)$$

$$= \frac{1 - e^{j2\pi(i_2-i_1)}}{1 - e^{\frac{j2\pi(i_2-i_1)}{N}}} = \frac{e^{j\pi(i_2-i_1)} \sin(\pi(i_2-i_1))}{e^{j\frac{\pi(i_2-i_1)}{N}} \sin\left(\frac{\pi(i_2-i_1)}{N}\right)} = 0 \text{ pour } i_1 \neq i_2 \quad (2.120)$$

On peut donc en conclure que  $\mathbf{V}(-\mathbf{t}_S)^H \mathbf{V}(-\mathbf{t}_S)$  est une matrice diagonale. De plus on peut approximer pour  $i_1 \approx i_2$ ,  $\sin\left(\frac{\pi(i_2-i_1)}{N}\right) \approx \frac{\pi(i_2-i_1)}{N}$ , et on a donc sur la diagonale :

$$[\mathbf{V}(-\mathbf{t}_S)^H \mathbf{V}(-\mathbf{t}_S)]_{i_1 i_1} \approx \frac{N \sin(\pi(i_2-i_1))}{\pi(i_2-i_1)} = N \text{sinc}(\pi(i_2-i_1)) = N \text{ pour } i_1 = i_2. \quad (2.121)$$

On peut donc en déduire que  $\Delta^H \Delta$  est une matrice diagonale, étant donné qu'elle est le produit de matrices diagonales. Le  $k$ -ième élément est

$$[\Delta^H \Delta]_{kk} = \frac{N^3 T_S^2}{2\pi} b_k^2 h_k^2 = \frac{N}{2\pi} b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2. \quad (2.122)$$

Les éléments de la MIF sont donc :

$$\left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau_l} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau_{l'}} \right] = a_l a_{l'} \sum_{k=-K/2}^{K/2} \frac{2\pi k^2}{NT_S^2} b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2 e^{\frac{j2\pi k(\tau_l - \tau_{l'})}{NT_S}} \quad (2.123)$$

$$\left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau_l} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial a_{l'}} \right] = a_l \sum_{k=-K/2}^{K/2} \frac{-jk}{T_S} b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2 e^{\frac{j2\pi k(\tau_l - \tau_{l'})}{NT_S}} \quad (2.124)$$

$$\left[ \frac{\partial \boldsymbol{\mu}}{\partial a_l} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau_{l'}} \right] = a_{l'} \sum_{k=-K/2}^{K/2} \frac{-jk}{T_S} b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2 e^{\frac{j2\pi k(\tau_l - \tau_{l'})}{NT_S}} \quad (2.125)$$

$$\left[ \frac{\partial \boldsymbol{\mu}}{\partial a_l} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial a_{l'}} \right] = \sum_{k=-K/2}^{K/2} \frac{N}{2\pi} b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2 e^{\frac{j2\pi k(\tau_l - \tau_{l'})}{NT_S}} \quad (2.126)$$

Pour passer en Bayésien on calcule l'espérance sur les paramètres de (2.123), (2.124), (2.125) et (2.126). Etant donné que les paramètres sont indépendants on peut séparer les espérances sur les amplitudes et les retards. La distribution sur les amplitudes étant centrée (c'est à dire de moyenne nulle) on a (2.124) et (2.125) qui sont nulles, ainsi que (2.123) pour  $l \neq l'$ . Les termes non nuls de la MIB sont donc les termes sur les amplitudes, ainsi que la diagonale de la MIB sur les retards, qui est :

$$\mathbb{E}_{\boldsymbol{\tau}, \mathbf{a}} \left\{ \left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau_l} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial \tau_l} \right] \right\} = \sigma_a^2 \sum_{k=-K/2}^{K/2} \frac{2\pi k^2}{NT_S^2} b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2 \left( \int_{S_\tau} e^0 p(\tau) d\tau \right)^2 \quad (2.127)$$

$$= \frac{2\pi}{NT_S^2} \sigma_a^2 \sum_{k=-K/2}^{K/2} k^2 b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2 \quad (2.128)$$

La MIB concernant les amplitudes est :

$$\mathbb{E}_{\tau, \mathbf{a}} \left\{ \left[ \frac{\partial \boldsymbol{\mu}}{\partial a_l} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial a_{l'}} \right] \right\} = \sum_{k=-K/2}^{K/2} \frac{N}{2\pi} b_k^2 H \left( \frac{2\pi k}{NT_S} \right)^2 \mathbb{E}_{\tau_l, \tau_{l'}} \left\{ e^{\frac{j2\pi k(\tau_l - \tau_{l'})}{NT_S}} \right\} \quad (2.129)$$

$$= \begin{cases} \frac{N}{2\pi} \sum_{k=-K/2}^{K/2} b_k^2 H \left( \frac{2\pi k}{NT_S} \right)^2 & \text{pour } l = l' \\ \frac{N}{2\pi} \sum_{k=-K/2}^{K/2} b_k^2 H \left( \frac{2\pi k}{NT_S} \right)^2 \mathbb{E}_{\tau} \left\{ e^{\frac{j2\pi k\tau}{NT_S}} \right\} \cdot \mathbb{E}_{\tau} \left\{ e^{\frac{-j2\pi k\tau}{NT_S}} \right\} & \text{pour } l \neq l' \end{cases} \quad (2.130)$$

Prenons par exemple une distribution uniforme des  $\tau$ , qui pour être rendue dérivable est considérée comme une gaussienne généralisée avec un  $\beta \rightarrow \infty$ ,  $p(\tau) = \frac{1}{NT_S}$  sur le support de l'échantillonnage c'est à dire pour  $\tau \in [0, NT_S]$ . Dans ce cas on peut calculer le terme hors diagonale de (2.130) :

$$\mathbb{E}_{\tau} \left\{ e^{\frac{j2\pi k\tau}{NT_S}} \right\} \cdot \mathbb{E}_{\tau} \left\{ e^{\frac{-j2\pi k\tau}{NT_S}} \right\} = \frac{1}{N^2 T_S^2} \left[ \frac{-j}{2\pi k} e^{\frac{j2\pi k\tau}{NT_S}} \right]_0^{NT_S} \left[ \frac{j}{2\pi k} e^{\frac{-j2\pi k\tau}{NT_S}} \right]_0^{NT_S} \quad (2.131)$$

$$\text{pour } k \neq 0 : = 0 \quad (2.132)$$

Pour  $k = 0$  on utilise l'approximation  $e^x \approx 1 + x$  pour  $x \approx 0$  afin d'éviter la division par zéro. On a donc :

$$\begin{aligned} \mathbb{E}_{\tau} \left\{ e^{\frac{j2\pi k\tau}{NT_S}} \right\} \cdot \mathbb{E}_{\tau} \left\{ e^{\frac{-j2\pi k\tau}{NT_S}} \right\} \Big|_{k \approx 0} &= \frac{1}{N^2 T_S^2} \left[ \frac{-j}{2\pi k} \left( 1 + \frac{j2\pi k\tau}{NT_S} \right) \right]_0^{NT_S} \left[ \frac{j}{2\pi k} \left( 1 + \frac{-j2\pi k\tau}{NT_S} \right) \right]_0^{NT_S} \\ &= \frac{1}{N^2 T_S^2} \end{aligned} \quad (2.133)$$

Finalement on obtient que

$$\mathbb{E}_{\tau, \mathbf{a}} \left\{ \left[ \frac{\partial \boldsymbol{\mu}}{\partial a_l} \right]^H \left[ \frac{\partial \boldsymbol{\mu}}{\partial a_{l'}} \right] \right\} = \frac{1}{2\pi NT_S^2} b_0^2 H(0)^2 \text{ pour } l \neq l' \quad (2.134)$$

**Résultat 8.** La MIB pour le problème d'estimation donné en (2.106), avec une distribution  $p(\tau)$  tendant vers une distribution uniforme de support  $[0, NT_S]$  est :

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{J}_{\tau} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{\mathbf{a}} \end{bmatrix} \quad (2.135)$$

avec

$$\mathbf{J}_{\tau} = \left( \frac{2\pi\sigma_a^2}{\sigma^2 NT_S^2} \sum_{k=-K/2}^{K/2} k^2 b_k^2 H \left( \frac{2\pi k}{NT_S} \right)^2 \right) \mathbf{I}_L \quad (2.136)$$

et

$$\mathbf{J}_{\mathbf{a}} = \frac{1}{\sigma^2 2\pi} \left( \left( N \sum_{k=-K/2}^{K/2} b_k^2 H \left( \frac{2\pi k}{NT_S} \right)^2 \right) \mathbf{I}_L + \frac{1}{NT_S^2} b_0^2 H(0)^2 (\mathbf{1}_L \mathbf{1}_L^T - \mathbf{I}_L) \right) \quad (2.137)$$

où  $\mathbf{1}_L$  est un vecteur colonne contenant  $L$  1s.

On cherche à présent à inverser  $\mathbf{J}(\boldsymbol{\theta})$  pour obtenir la borne Bayésienne de Van Trees. Etant donné que la matrice est bloc-diagonale, les BCRB sur les retards et sur les amplitudes sont indépendantes et égales simplement à l'inverse de la MIB sur les retards et les

amplitudes respectivement.  $\mathbf{J}_\tau$  étant une matrice diagonale on obtient la BCRB sur les retards directement. Pour calculer la BCRB sur les amplitudes on remarque que (2.137) peut s'écrire de cette façon :

$$\mathbf{J}_a = \frac{1}{\sigma^2 2\pi} \left[ \left( N \sum_{k=-K/2}^{K/2} b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2 - \frac{1}{NT_S^2} b_0^2 H(0)^2 \right) \mathbf{I}_L + \frac{1}{NT_S^2} b_0^2 H(0)^2 \mathbf{1}_L \mathbf{1}_L^T \right]. \quad (2.138)$$

On peut utiliser la formule de Sherman-Morrisson [98] suivante

$$(\mathbf{A} + \mathbf{b}\mathbf{c}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{b}\mathbf{c}^T\mathbf{A}^{-1}}{1 + \mathbf{c}^T\mathbf{A}^{-1}\mathbf{b}} \quad (2.139)$$

pour inverser la MIB :

$$[\mathbf{J}_a]^{-1} = \frac{\sigma^2 2\pi}{N \sum_{k=-K/2}^{K/2} b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2 - \frac{1}{NT_S^2} b_0^2 H(0)^2} \mathbf{I}_L - \frac{\frac{\sigma^2 2\pi}{NT_S^2} b_0^2 H(0)^2 \mathbf{1}_L \mathbf{1}_L^T}{\left( N \sum_{k=-K/2}^{K/2} b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2 - \frac{1}{NT_S^2} b_0^2 H(0)^2 \right)^2} \quad (2.140)$$

**Résultat 9.** Les BCRB sur les retards et les amplitudes pour une distribution  $p(\tau)$  tendant vers une distribution uniforme de support  $[0, NT_S]$  et pour un RSB élevé sont :

$$\text{BCRB}(\tau_l) \approx \frac{\sigma^2 NT_S^2}{2\pi\sigma_a^2} \frac{1}{\sum_{k=-K/2}^{K/2} k^2 b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2} \quad (2.141)$$

$$\text{BCRB}(a_l) \approx \frac{\sigma^2 2\pi}{N \sum_{k=-K/2}^{K/2} b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2 - \frac{1}{NT_S^2} b_0^2 H(0)^2} \left( \frac{1 + (L-1) \frac{\frac{1}{NT_S^2} b_0^2 H(0)^2}{N \sum_{k=-K/2}^{K/2} b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2 - \frac{1}{NT_S^2} b_0^2 H(0)^2}}{1 + L \frac{\frac{1}{NT_S^2} b_0^2 H(0)^2}{N \sum_{k=-K/2}^{K/2} b_k^2 H\left(\frac{2\pi k}{NT_S}\right)^2 - \frac{1}{NT_S^2} b_0^2 H(0)^2}} \right) \quad (2.142)$$

Pour les deux paramètres la borne obtenue est proportionnelle à la variance du bruit. Dans le cas de l'impulsion de Dirac, la transformée de Fourier de l'impulsion  $H(\omega)$  est constante. La borne sur les retards avec le premier modèle dépendait principalement de la dérivée du noyau, alors qu'en se basant sur l'estimation directement avec le noyau SoS on observe que le paramètre déterminant les performances sur le retard est la norme du vecteur  $\|\mathbf{k}\mathbf{b}\|^2$ . Ces deux expressions ne sont pas égales, la différence venant du fait qu'en introduisant la forme du noyau SoS dans le problème initial d'estimation, et en développant ce modèle on aboutit finalement à un problème d'estimation différent de celui présenté dans la première section. Cependant on peut noter que l'expression (2.141) reflète également l'importance de la dérivée du noyau de manière intuitive. En effet, si l'on omet la fonction rectangle qui permet d'avoir un noyau de support fini, la dérivée du noyau SoS est :

$$\frac{\partial g_{\text{SoS}}(t)}{\partial t} = \sum_{k=-K/2}^{K/2} j2\pi NT_S k b_k e^{j2\pi NT_S k} \quad (2.143)$$

le facteur exponentiel n'ayant pas de poids dans la norme de la dérivée, on peut observer que maximiser la dérivée passe également par maximiser la norme de  $k b_k$ . Nous allons voir plus en détail le comportement de ces bornes, en visualisant le noyau les maximisant pour différentes formes d'impulsion.

### 2.3.2 Choix du noyau optimal au sens de l'information

#### Choix de la fonction objectif

Nous allons à présent chercher un noyau optimal minimisant l'erreur quadratique moyenne dans un contexte Bayésien. Pour ce faire nous allons utiliser une démarche similaire à celle de la section 2.2.2 où nous avons cherché le noyau minimisant la borne sur les retards. La borne sur les amplitudes avec le modèle précédent était trop complexe pour être utilisée pour une optimisation, nous n'avons donc pas pu trouver de noyau optimal global. Dans cette section nous allons en dériver un pour le modèle étudié en 2.3.1 qui non seulement est plus général car il permet tout type de forme d'impulsions, mais en plus prend une forme analytique que nous pouvons utiliser. La BCRB totale que nous devrions optimiser est la somme des BCRB (2.141) et (2.142). Malheureusement cette somme n'est pas une fonction convexe, et est donc difficile à utiliser. Cependant ce problème peut être résolu en cherchant non pas à minimiser l'erreur quadratique à travers la borne, mais en maximisant l'information de Fisher qui est l'information disponible relative à chaque paramètre. Cette information est mesurée par  $\text{tr}(\mathbf{J}_\tau)$  (2.136) et  $\text{tr}(\mathbf{J}_a)$  (2.137) respectivement.

Tout comme dans la section 2.2.2, nous allons nous intéresser uniquement à la recherche d'un noyau réel, c'est à dire répondant à la contrainte  $b_k = b_{-k}$ . Nous souhaitons donc exprimer les éléments diagonaux de la matrice de Fisher sous forme matricielle, de manière à faire apparaître  $\mathbf{b} = [b_0, \dots, b_{K/2}]^T$  qui sont les paramètres inconnus à optimiser. On fait l'hypothèse que la forme d'impulsion  $h(t)$  est une fonction réelle et paire, ce qui assure que  $H(\omega) = H(-\omega)$ , c'est à dire que la transformée de Fourier de  $h(t)$  est également paire. On obtient donc que :

$$[\mathbf{J}_\tau]_l = \frac{2\pi\sigma_a^2}{\sigma^2 N T_s^2} \sum_{k=1}^{K/2} k^2 b_k^2 H\left(\frac{2\pi k}{N T_s}\right)^2 \quad (2.144)$$

$$= \mathbf{b}^T \underbrace{\frac{2\pi\sigma_a^2}{\sigma^2 N T_s^2} \text{diag}\left(0, 2H\left(\frac{2\pi}{N T_s}\right)^2, \dots, 2\frac{K^2}{2} H\left(\frac{2\pi \frac{K}{2}}{N T_s}\right)^2\right)}_{\mathbf{D}_\tau} \mathbf{b} \quad (2.145)$$

$$[\mathbf{J}_a]_l = \frac{N}{\sigma^2 2\pi} \left( \sum_{k=1}^{K/2} 2b_k^2 H\left(\frac{2\pi k}{N T_s}\right)^2 + b_0^2 H(0)^2 \right) \quad (2.146)$$

$$= \mathbf{b}^T \underbrace{\frac{N}{\sigma^2 2\pi} \text{diag}\left(H(0)^2, 2H\left(\frac{2\pi}{N T_s}\right)^2, \dots, 2H\left(\frac{2\pi \frac{K}{2}}{N T_s}\right)^2\right)}_{\mathbf{D}_a} \mathbf{b} \quad (2.147)$$

Notons que les valeurs sur la diagonale de  $\mathbf{J}_\tau$  et  $\mathbf{J}_a$  sont constantes, de ce fait la trace de ces matrices est simplement  $L$  fois la valeur sur leur diagonale.

Nous cherchons à trouver un noyau optimal global, c'est à dire qui soit un bon compromis entre une optimisation des performances d'estimation sur les retards et sur les amplitudes. Pour ce faire il est important que les deux MIB soient comparables quelle que soit la distribution des paramètres choisie. Dans ce but nous allons optimiser la somme des deux MIB normalisées. La BEQM normalisée est définie par [126] :

$$\text{BEQM}_{\text{norm}}(\theta_i) = \frac{\text{BEQM}(\theta_i)}{\sigma_{\theta_i}^2}, \quad (2.148)$$

le lien entre la BEQM et la BCRB étant  $\text{BEQM}(\theta_i) \geq \text{BCRB}(\theta_i)$ , la BCRB normalisée est

$$\text{BEQM}_{\text{norm}}(\theta_i) \geq \frac{\text{BCRB}(\theta_i)}{\sigma_{\theta_i}^2} = \text{BCRB}_{\text{norm}}(\theta_i) \quad (2.149)$$

Pour normaliser la BCRB on divise la BCRB sur chaque paramètre par la variance de chaque paramètre, supposant que les variances  $\sigma_{\tau}^2$  et  $\sigma_a^2$  sont indépendants de l'impulsion, normaliser la MIB revient à multiplier  $\mathbf{J}_{\tau}$  par  $\sigma_{\tau}^2$  et  $\mathbf{J}_a$  par  $\sigma_a^2$ . De plus nous introduisons un paramètre  $\lambda \in [0, 1]$  afin de régler l'importance relative attribuée à chaque type de paramètres à estimer. Pour  $\lambda = 0.5$  chaque MIB normalisée a la même importance, c'est à dire que les retards et les amplitudes auront le même poids sur le choix du noyau optimisé. La fonction à optimiser est donc finalement la somme des MIB sur les retards et les amplitudes normalisées, pondérées par le paramètre  $\lambda$  :

$$f_{\lambda}(\mathbf{b}) = \lambda \sigma_{\tau}^2 \text{tr}(\mathbf{J}_{\tau}) + (1 - \lambda) \sigma_a^2 \text{tr}(\mathbf{J}_a) \quad (2.150)$$

$$= \lambda \sigma_{\tau}^2 \mathbf{L} \mathbf{b}^T \frac{2\pi \sigma_a^2}{\sigma^2 N T_S^2} \mathbf{D}_{\tau} \mathbf{b} + (1 - \lambda) \sigma_a^2 \mathbf{L} \mathbf{b}^T \frac{N}{\sigma^2 2\pi} \mathbf{D}_a \mathbf{b} \quad (2.151)$$

$$= \mathbf{b}^T \mathbf{D}_{\lambda} \mathbf{b} \quad (2.152)$$

où

$$\mathbf{D}_{\lambda} \triangleq \text{LRSB} \left( \lambda \frac{2\pi \sigma_{\tau}^2}{N T_S^2} \mathbf{D}_{\tau} + (1 - \lambda) \frac{N}{2\pi} \mathbf{D}_a \right) \quad (2.153)$$

où le RSB dans un contexte Bayésien est  $\sigma_a^2/\sigma^2$ . On peut dès à présent remarquer que la fonction objectif (2.152) prend la forme intéressante d'une fonction quadratique par rapport à  $\mathbf{b}$ . De plus  $\mathbf{D}_{\lambda}$  est une matrice diagonale dont les valeurs propres sont donc les valeurs sur la diagonale, c'est à dire positive. On peut en conclure que  $\mathbf{D}_{\lambda}$  est également une matrice définie positive, ce qui implique que la fonction objectif est convexe.

### Choix des contraintes et noyau optimal

Une contrainte dans la construction d'un noyau SoS est que tous les  $b_k$  soient différents de zéro. Dans la fonction objectif, seul le carré des coefficients est utilisé, de ce fait le signe des  $b_k$  n'influe pas sur la borne. On peut donc simplifier le problème en ne considérant que les solutions positives, ce qui permet de transformer la contrainte  $b_k \neq 0$  en  $b_k \geq \epsilon > 0$  où le paramètre  $\epsilon$  est réglé par l'utilisateur. En théorie on peut choisir un  $\epsilon$  infiniment proche de zéro, mais on verra qu'en présence de bruit le choix du palier est important. Il faut également contraindre la puissance disponible, car augmenter la valeur des coefficients à l'infini permet d'augmenter  $f_{\lambda}(\tilde{\mathbf{b}})$  à l'infini. Une première méthode pour ce faire est d'introduire la contrainte  $\|\mathbf{b}\|^2 = \mathbf{b}^T \mathbf{b} = K + 1$ . Dans ce cas le problème d'optimisation que nous cherchons à résoudre est

$$\max_{\mathbf{b}} \quad \mathbf{b}^T \mathbf{D}_{\lambda} \mathbf{b} \quad \text{s.t.} \quad \begin{cases} \mathbf{b}^T \mathbf{b} = K + 1 \\ \forall k \in [-K/2 \dots K/2] : b_k \geq \epsilon \end{cases} \quad (2.154)$$

Pour avoir une idée générale du type de solution que l'on obtient, nous considérons d'abord le problème simplifié pour  $\epsilon = 0$ . Cette solution correspond au cas idéal sans bruit, où l'on fait tendre  $\epsilon$  vers 0. La fonction de Lagrange obtenue dans ce cas là est :

$$\forall \lambda \in [0, 1], \mathcal{L}_{\lambda}(\mathbf{b}, \alpha) = \mathbf{b}^T \tilde{\mathbf{D}}_{\lambda} \mathbf{b} - \alpha(\mathbf{b}^T \mathbf{b} - K - 1). \quad (2.155)$$

Le vecteur  $\mathbf{b}^*$  optimal est obtenu en dérivant  $\mathcal{L}_\lambda(\mathbf{b}, \alpha)$  par rapport à  $\mathbf{b}$ , puis en cherchant pour quelle valeur cette dérivée s'annule. Les vecteurs  $\mathbf{b}$  candidats sont donc les vecteurs qui satisfont

$$\mathbf{D}_\lambda \mathbf{b} = \alpha \mathbf{b} \quad (2.156)$$

c'est à dire les vecteurs propres de la matrice  $\mathbf{D}_\lambda$ . La solution amenant un maximum est le vecteur propre associé à la plus grande valeur propre de cette matrice. La matrice  $\mathbf{D}_\lambda$  étant diagonale, les vecteurs propres sont des vecteurs avec une seule valeur à 1, et le reste à zéro ; nous notons  $\mathbf{e}_m$  ce type de vecteur, où l'indice  $m$  correspond à l'indice de la valeur à 1. On obtient donc  $\mathbf{b}_\lambda^* = \mathbf{e}_{m_\lambda^*}$ . Une fois la symétrie appliquée sur le vecteur  $\mathbf{b}$  on obtient un vecteur contenant deux pics symétriques par rapport à zéro. Les coefficients correspondant à une vision fréquentielle du noyau  $g(t)$ , le noyau temporel optimal pour  $\epsilon = 0$  est une sinusoïde parfaite donc la fréquence est déterminée par la localisation de la valeur propre la plus grande de  $\mathbf{D}_\lambda$ , qui, elle dépend principalement de la forme  $H(\omega)^2$ .

Ce noyau n'est pas utilisable en pratique, car beaucoup de valeurs  $b_k$  sont à zéro, c'est pourquoi on ajoute maintenant la contrainte de palier  $\epsilon$ , avec  $0 < \epsilon \leq 1$ . Le problème d'optimisation convexe (2.154) avec  $\epsilon > 0$  ne peut pas être résolu analytiquement, on utilise donc le package CVX sur Matlab pour le résoudre [68, 69]. Concrètement on recherche les  $\beta_k = b_k^2$  solutions, et on choisira la solution positive correspondante. Le résultat des simulations montre que le noyau obtenu est similaire à celui obtenu pour  $\epsilon = 0$ , mais avec un palier de niveau  $\epsilon$  (voir Figure 2.15). Le pic obtenu prend une valeur  $u$  de telle sorte que  $\sum_0^{K/2} \beta_k = K + 1$ .

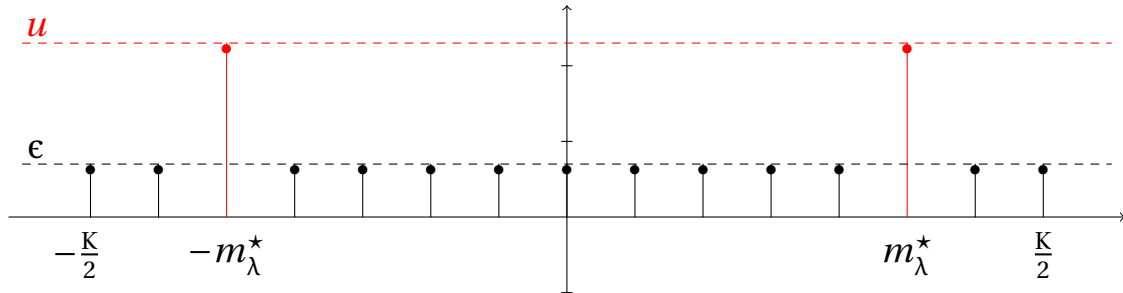


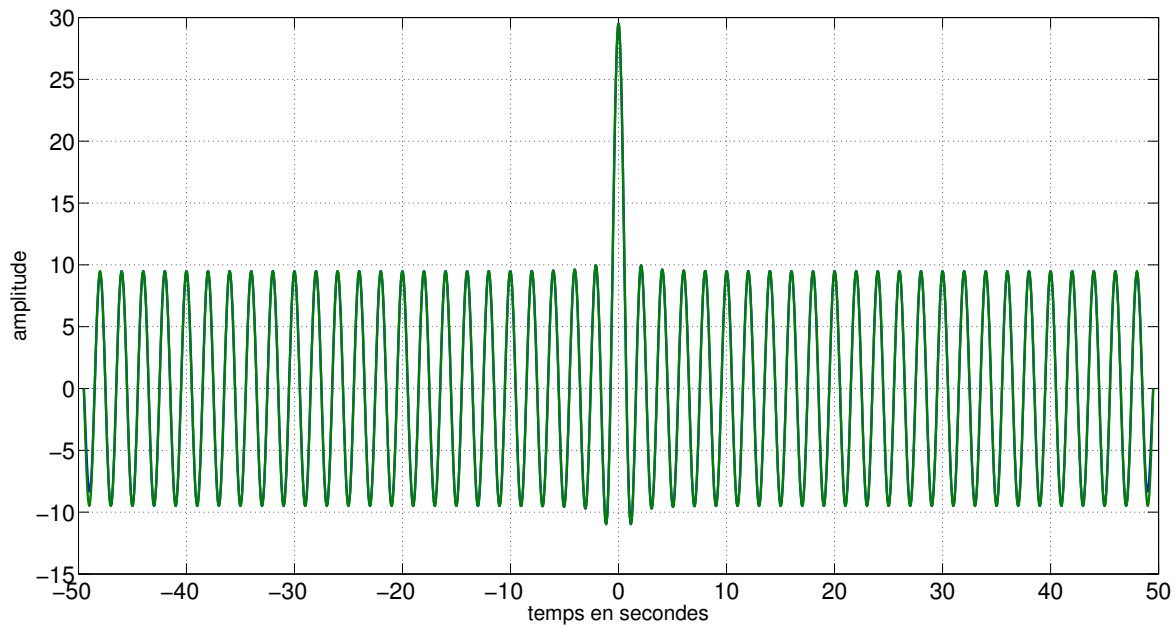
FIGURE 2.15 – Coefficients optimaux obtenus

Le noyau obtenu dans le domaine temporel (voir un exemple Figure 2.16) est :

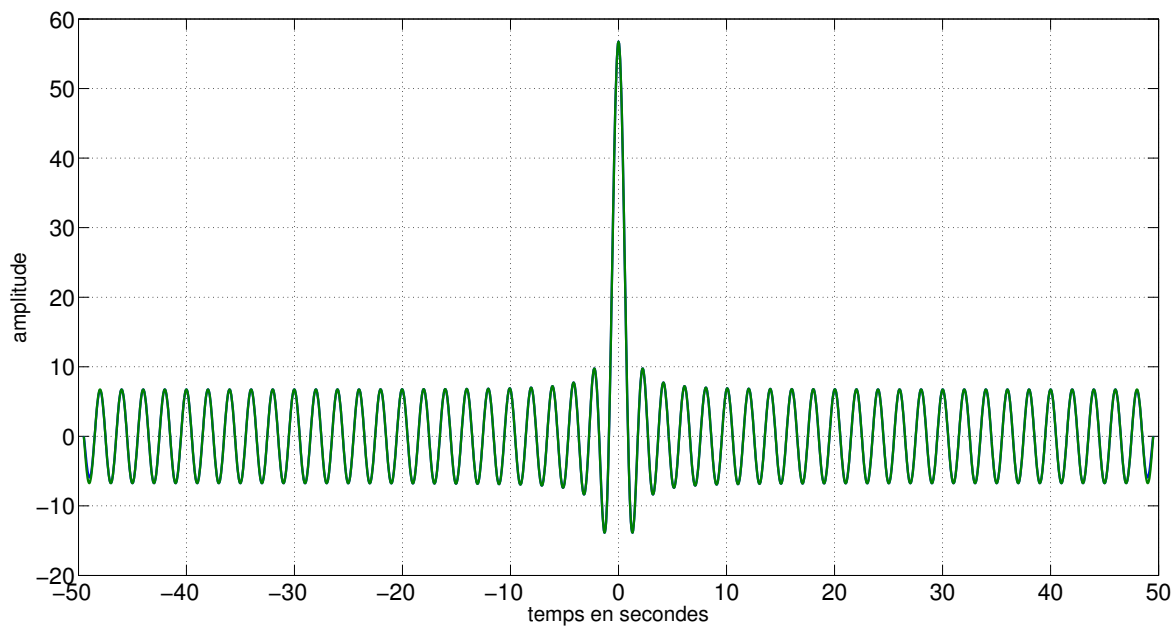
$$g^*(t) = 2(u - \epsilon) \cos\left(\frac{2\pi m^* t}{NT_S}\right) + \epsilon \sum_{k=-K/2}^{K/2} e^{\frac{j2\pi kt}{NT_S}} \quad (2.157)$$

Quand  $\epsilon = 0$  on a une simple sinusoïde tandis que pour  $\epsilon = 1$  le noyau optimal devient un simple noyau de Dirichlet. En optimisant à partir de la BCRB, on minimise l'erreur théorique dans le régime asymptotique, le noyau optimal dans ce cas est obtenu pour  $\epsilon \rightarrow 0$ . Cependant, augmenter  $\epsilon$  permet d'obtenir un pic au niveau du temps  $t = 0$  qui permet de distinguer une source parmi d'autre, et d'estimer son retard en présence de bruit. Avec une simple sinusoïde, la source serait indétectable en présence de bruit. Le choix d' $\epsilon$  va donc être naturellement dépendant du niveau de bruit considéré, plus  $\epsilon$  est choisi grand, plus le noyau sera robuste. En effet, il faut garder en mémoire que nous nous intéressons ici à l'optimisation de performances théoriques moyennés sur tous les retards, et qui concernent de plus le régime asymptotique. Les noyaux obtenus par cette méthode

sont intéressants d'un point de vue théorique et permettent d'avoir une idée de la forme de noyaux à retenir, cependant d'un point de vue pratique il y a toujours d'autres paramètres à prendre en compte, notamment le bruit, et la faisabilité d'une méthode d'estimation correspondant au noyau.



(a) pour  $\epsilon = 0.2$



(b) pour  $\epsilon = 0.5$

FIGURE 2.16 – Noyau optimal pour une impulsion de Dirac où  $N = K = 100$ ,  $\lambda = 0.5$ .

### Performances avec le noyau optimal

Dans cette sous-section, nous allons chercher à visualiser quels sont les résultats de l'optimisation dans un cas concret par la simulation. Nous allons nous intéresser à la

BCRB sur les retards et les amplitudes du noyau optimisé pour différents  $\epsilon$  en fonction de  $\lambda$ , ainsi qu'au choix de l'indice  $m^*$  correspondant. Nous prenons à présent  $N = 400$ ,  $K = 300$  et toujours  $T_S = 1$ . Le rapport signal sur bruit est choisi à 30dB, et nous prenons un  $\sigma_\tau^2 = 1$ . Cette valeur est petite par rapport à une valeur qu'on utiliserait en pratique, mais nous avons fait ce choix pour pouvoir observer l'évolution des bornes en fonction de  $\lambda$ . Si l'on prend  $\sigma_\tau^2$  grand, cela apporte un poids important à la MIF sur les retards qui sera alors prépondérante dès que  $\lambda \neq 0$ .

Les résultats pour une impulsion de Dirac sont donnés dans le tableau 2.2. Quand  $h(t)$  est une impulsion de Dirac, les coefficients fréquentiels  $h_k$  sont tous égaux, ce qui a pour conséquence que  $\mathbf{J}_a$  ne dépend pas de  $\mathbf{b}$ . En effet, sur les trois noyaux donnés en exemples dans le tableau, la BCRB sur les amplitudes est constante. C'est pourquoi quand seules les amplitudes sont considérées ( $\lambda = 0$ ) tous les noyaux possibles sont des solutions, dont en particulier le noyau de Dirichlet (où tous les  $b_k$  sont égaux). Dans ce cas précis, l'algorithme d'optimisation s'arrête sur le premier noyau solution obtenu qui dans notre simulation donne un résultat moins bon que le noyau de Dirichlet. Le noyau optimal global pour  $\lambda > 0$  dépend donc des paramètres de retards et ne varie pas avec  $\lambda$ . Dans ce contexte, la valeur propre la plus élevée est celle correspondant au  $k^2$  maximum, c'est à dire la dernière. Le noyau correspondant en temps pour différentes valeurs de  $\epsilon$  est tracé figure 2.16 : on observe une sinusoïde dont l'amplitude varie selon le  $\epsilon$  choisi, ainsi qu'un pic autour du temps  $t = 0$ . La BCRB correspondante est effectivement plus petite que celle du noyau Dirichlet.

Il est intéressant de comparer ce résultat à celui obtenu dans 2.12. Dans la section 2.2.2 le problème traité était très similaire : nous cherchions le noyau SoS permettant de minimiser la BCRB pour une impulsion de Dirac. Les deux noyaux obtenus ne se ressemblent pas dans le domaine temporel, mais sont assez proches au niveau du choix des coefficients optimaux. En effet le noyau de la figure 2.12 a également un pic dans les hautes fréquences, à la différence que ce pic s'étale sur plusieurs fréquences en largeur. La borne obtenue dans les deux cas n'est pas tout à fait comparable car nous nous sommes intéressés soit à la minimisation de la BCRB sur les retards, soit à la maximisation de l'information de Fisher. Dans ce 2e cas, le noyau final obtenu prend une forme très simple puisque la MIF est une matrice diagonale, ce qui implique la restriction des noyaux à une forme très simple également contenant un seul pic en fréquence. De plus la normalisation utilisée n'est pas la même, ce qui rend la comparaison des performances difficile, le noyau avec la plus grande norme est celui ayant la BCRB la plus basse, car il permet d'estimer les échantillons avec plus d'énergie. Nous nous intéresserons à ce problème dans la section 2.3.2.

|                             | Amplitude              | Delays                 |
|-----------------------------|------------------------|------------------------|
| Noyau de Dirichlet          | $2.609 \times 10^{-6}$ | $9.338 \times 10^{-6}$ |
| Noyau optimal $\lambda = 0$ | $2.609 \times 10^{-6}$ | $9.966 \times 10^{-6}$ |
| Noyau optimal $\lambda > 0$ | $2.609 \times 10^{-6}$ | $3.154 \times 10^{-6}$ |

TABEAU 2.2 – BCRB obtenues par l'optimisation, pour une impulsion de Dirac

Pour mieux visualiser l'évolution du résultat de l'optimisation en fonction de  $\lambda$  nous nous intéressons à présent à une impulsion gaussienne, puisque nous venons de le voir dans le cas d'une impulsion de Dirac le noyau optimal ne dépend pas vraiment de  $\lambda$ . Le noyau utilisé dans le domaine fréquentiel est présenté 2.17. Le noyau gaussien a son pic

en zéro, et donc apporte un poids assez important aux coefficients proches de  $k = 0$ , ce qui dans l'optimisation viendra contrebalancer l'effet de  $k^2$  sur les performances liées aux retards.

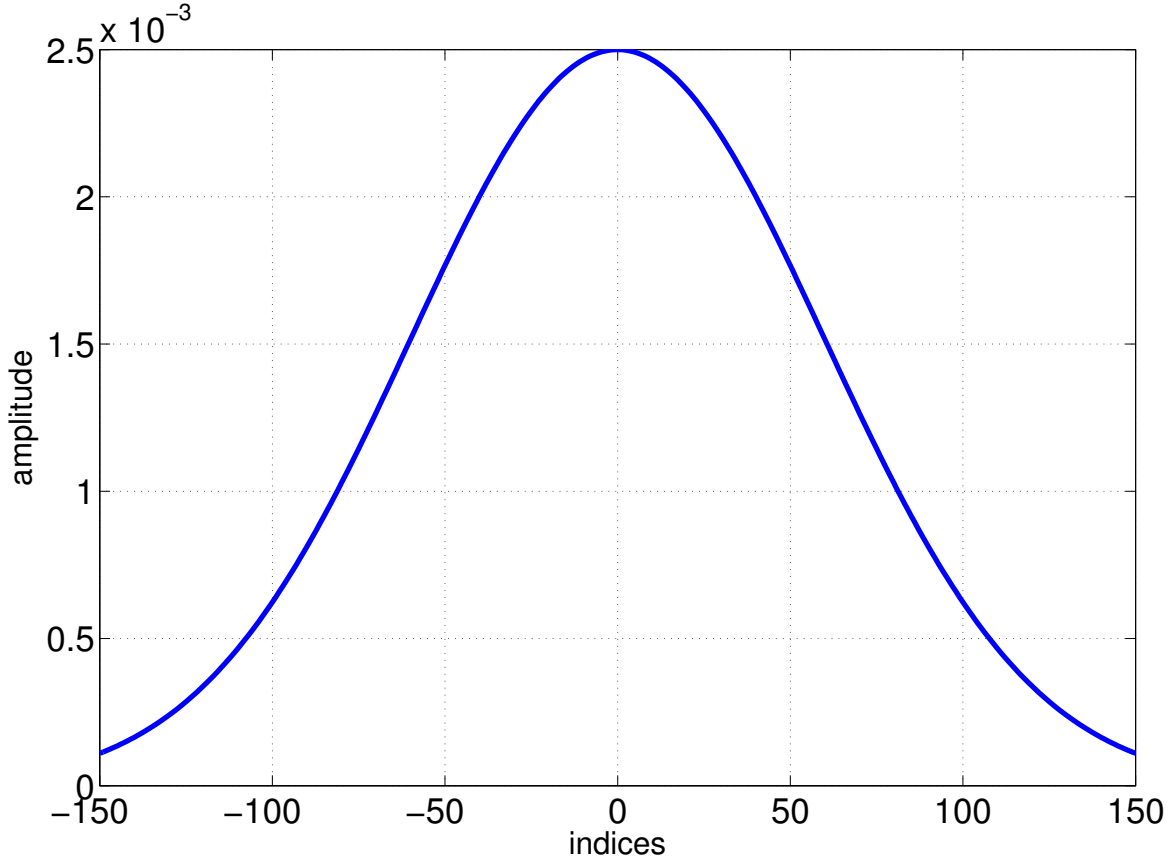


FIGURE 2.17 – Impulsion gaussienne, en fréquence

L'évolution du choix de l'indice  $m^*$  en fonction de  $\lambda$  est visible sur la figure 2.18 : quand  $\lambda = 0$  seule la borne sur l'amplitude est considérée, c'est à dire qu'on cherche à maximiser  $H\left(\frac{2\pi k}{NT_s}\right)$ , on obtient un indice le plus petit possible c'est à dire  $m^* = 0$ , puisque une fonction gaussienne est maximale en zéro. Notons que sur la courbe cela correspond à un indice 1, étant donné que sur Matlab les vecteurs sont indexés à partir de 1. Quand  $\lambda$  augmente on doit trouver un compromis entre maximiser  $H\left(\frac{2\pi k}{NT_s}\right)$  et  $k^2 H\left(\frac{2\pi k}{NT_s}\right)$ , jusque  $\lambda = 0.5$  le maximum de la gaussienne reste prépondérant et on garde donc  $m^* = 0$ . Ensuite la borne sur l'amplitude commence à prendre un poids plus important que celle sur les retards, ce qui décale le pic vers les hautes fréquences progressivement à mesure que l'importance du paramètre d'amplitude augmente. Cela se traduit finalement par un indice  $m^* = 60$  quand on ne considère que les amplitudes, ce qui correspond donc au  $k^2 H\left(\frac{2\pi k}{NT_s}\right)$  maximum.

Sur les figures 2.19 et 2.20 on peut voir les performances de ces noyaux sur les retards et les amplitudes respectivement. On y a tracé les BCRB sur ces deux paramètres dans une échelle semi-logarithmique pour les noyaux optimisés dans le cas de  $\epsilon = 0.5$  et  $\epsilon = 0.1$ , pour permettre la comparaison nous avons également tracé en rouge les bornes pour le noyau de Dirichlet c'est à dire pour  $b_k = 1$ . La première chose à noter est que les bornes sur les retards et les amplitudes n'ont pas du tout le même ordre de grandeur : la borne sur

les amplitudes a un ordre de grandeur au moins 10 fois plus petit que la borne sur les retards, ou de manière équivalente l'information sur les amplitudes est 10 fois plus grande en ordre de grandeur que l'information sur les retards. Cela illustre immédiatement l'importance de la normalisation, sinon dès que  $\lambda \leq 1$  seule l'amplitude serait utilisée pour choisir l'indice  $m^*$  optimal. La BCRB totale est simplement la somme des deux bornes, la BCRB sur les retards étant négligeable devant la BCRB sur les amplitudes j'ai choisi de ne pas l'afficher, c'est exactement la même courbe visuellement que dans 2.20. Pour  $0 \leq \lambda \leq 0.5$  la MIF sur l'amplitude est prépondérante de ce fait la BCRB sur les retards obtenus par le noyau optimisé est bien moins bonne que celle du noyau de Dirichlet. A l'opposé, on a une BCRB sur les amplitudes bien meilleure. Au fur et à mesure que  $\lambda$  augmente, la BCRB sur les amplitudes augmente également pour rejoindre les performances du noyau de Dirichlet : quel que soit  $\lambda$  le noyau optimal est meilleur au regard des amplitudes. Une fois que  $\lambda > 0.5$  la BCRB sur les retards devient de plus en plus importante dans le choix du noyau, ce qui se traduit par une nette amélioration des performances du noyau optimal : elles sont meilleures que le noyau de Dirichlet dès  $\lambda = 0.6$ . Du point de vue de la BCRB, le  $\epsilon = 0$  est optimal, cette intuition est confirmée par la simulation : quand on augmente  $\epsilon$  les performances diminuent, jusqu'à rejoindre celles du noyau de Dirichlet, qui peut également être vu comme le résultat de l'optimisation pour  $\epsilon = 1$ .

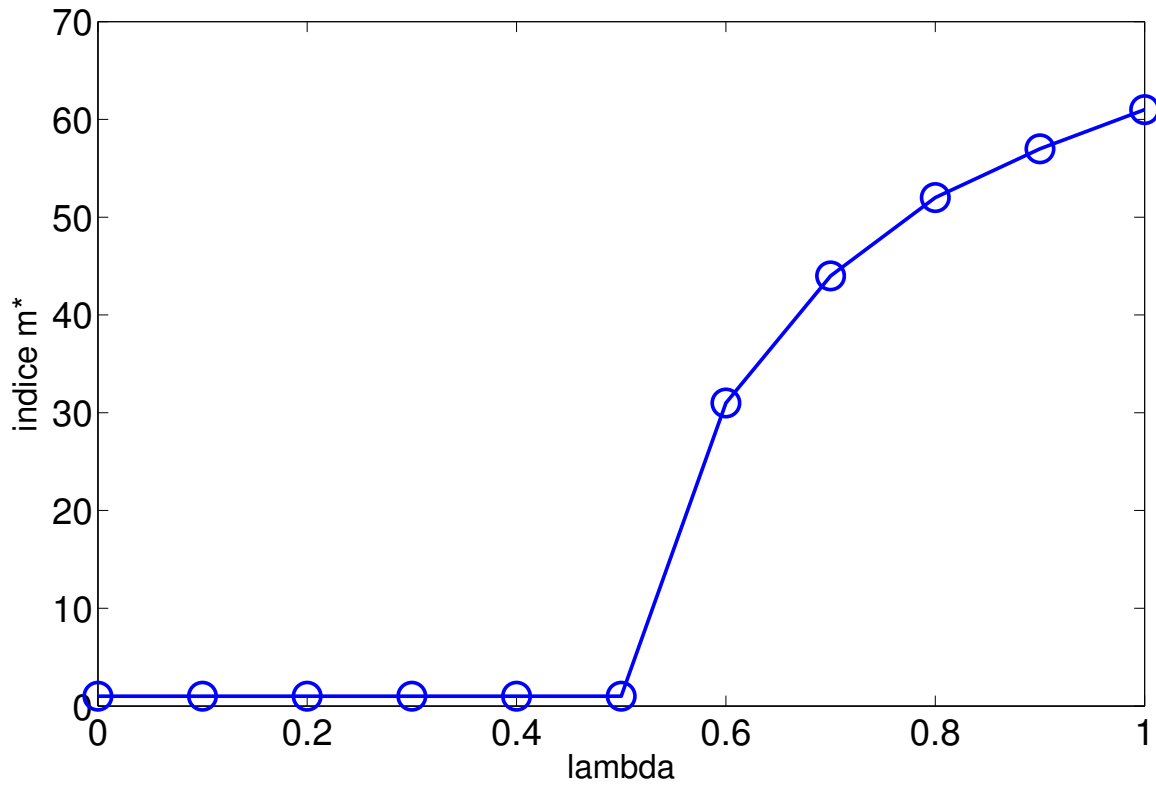


FIGURE 2.18 – Indice du pic, en fonction de  $\lambda$

### Résultat pour une autre contrainte de normalisation

La normalisation utilisée dans (2.154) est celle que nous avons proposée dans l'article [24], suite à son utilisation également par TUR, ELDAR et FRIEDMAN [120]. Il s'agit de

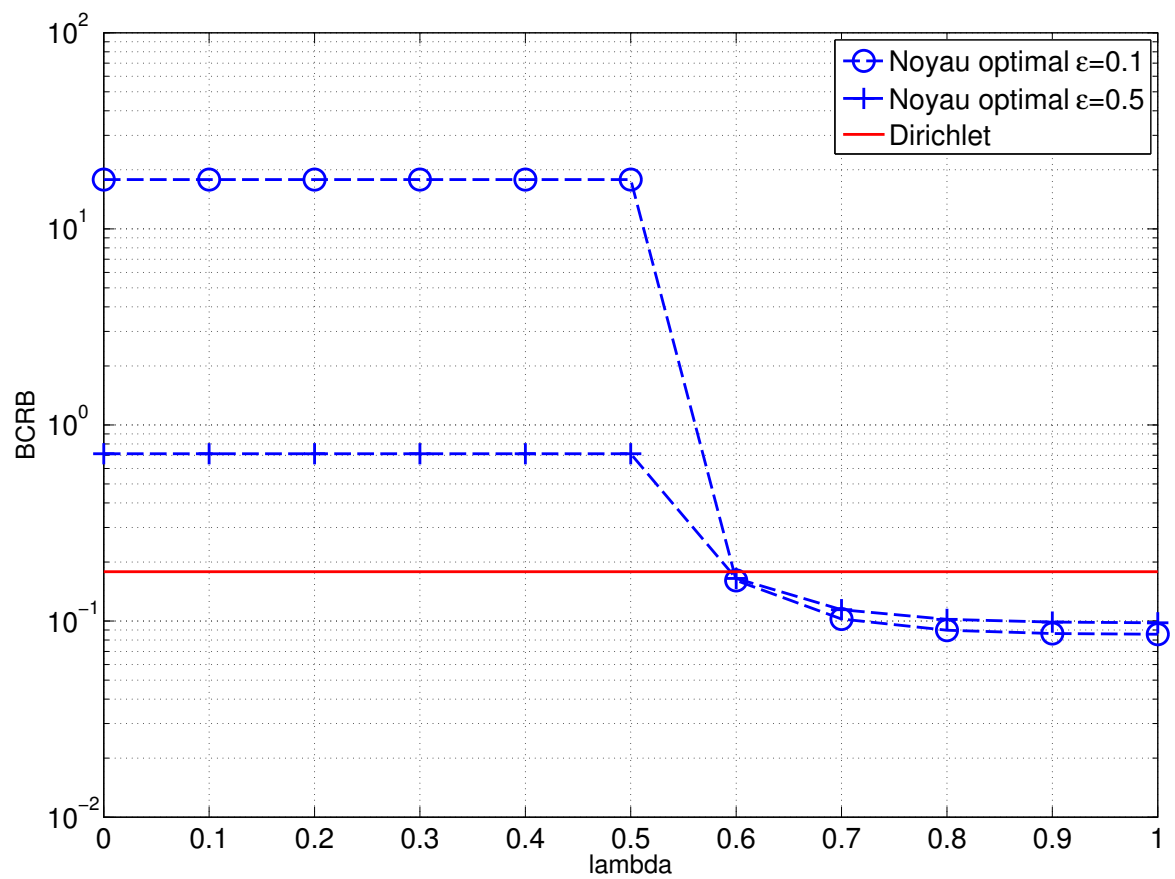


FIGURE 2.19 – BCRB sur les retards, pour le noyau de Dirichlet et le noyau optimisé pour différent  $\epsilon$

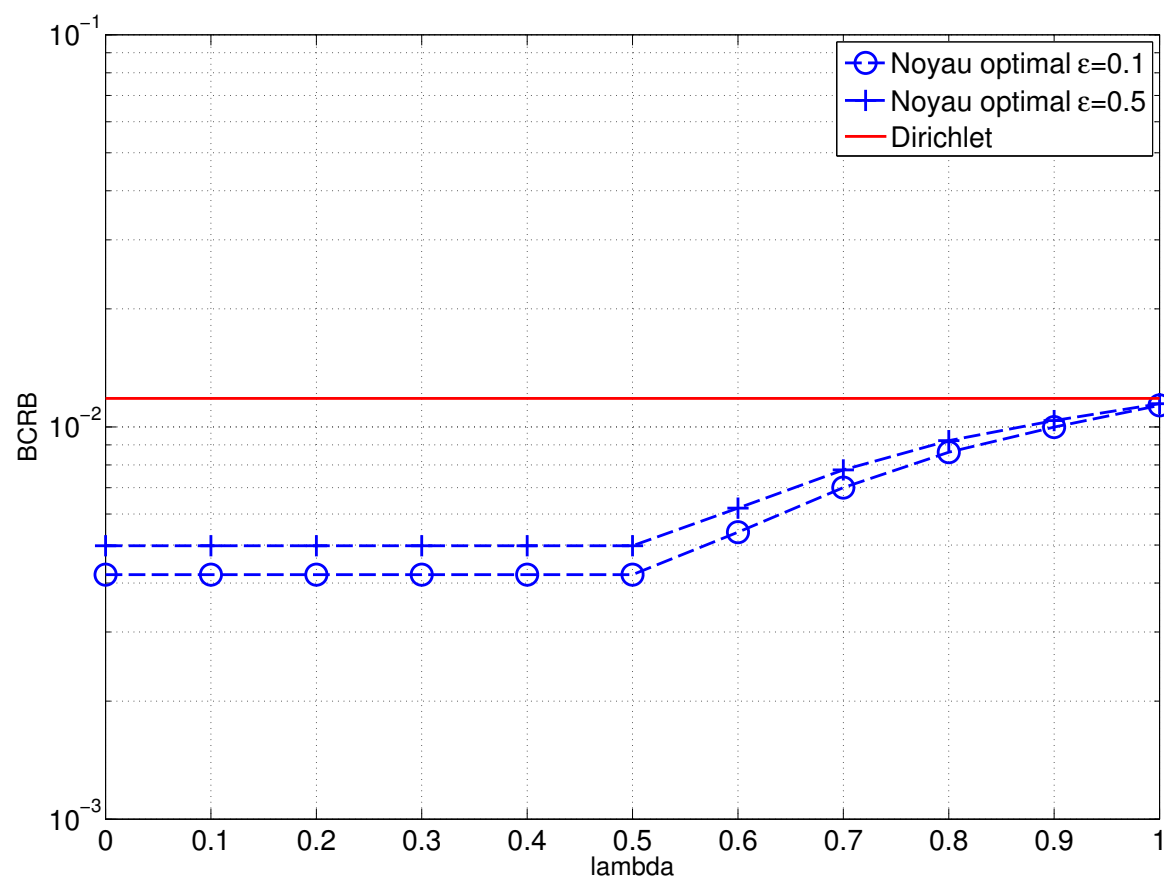


FIGURE 2.20 – BCRB sur les amplitudes, pour le noyau de Dirichlet et le noyau optimisé pour différent  $\epsilon$

la fonction la plus simple permettant d'empêcher la puissance d'être choisie infinie. Cependant une autre normalisation ayant plus de sens en terme de noyaux est de conserver celle proposée dans (2.98) qui ne dépend non pas de  $\mathbf{b}$  mais directement du noyau  $g(t)$ . Cela a l'avantage de rendre tous les types de noyaux comparables, alors que la normalisation précédente ne peut fonctionner que pour des noyaux SoS. On prendra donc que  $\mathbf{g}^T \mathbf{g} = \mathbf{b}^T \Delta \mathbf{b} = 1$ . Le nouveau problème d'optimisation est alors :

$$\max_{\mathbf{b}} \quad \mathbf{b}^T \mathbf{D}_\lambda \mathbf{b} \text{ t.q. } \begin{cases} \mathbf{b}^T \Delta \mathbf{b} = 1 \\ \forall k \in [0 \cdots K/2] : b_k \geq \epsilon \end{cases} \quad (2.158)$$

Pour  $\epsilon = 0$  le problème se résout exactement de la même manière que (2.98) : le noyau optimal correspond au vecteur propre associé à la valeur propre généralisée maximale solution du système :

$$\mathbf{D}_\lambda \mathbf{b} = \lambda \Delta \mathbf{b} \quad (2.159)$$

Pour une impulsion de Dirac, le noyau obtenu est montré dans la figure 2.21. Pour rendre la comparaison avec la solution de (2.98) valable, nous avons choisi d'optimiser uniquement la borne sur les retards, c'est à dire de prendre  $\lambda = 1$ . Nous avons utilisé les mêmes paramètres que dans la section précédente c'est à dire  $K = 400$ ,  $N = 300$ ,  $T_s = 1$ , un RSB de 30 et  $\sigma_\tau^2 = 1$ . Les deux noyaux représentés sont les solutions des problèmes (2.98) et (2.158), tous deux ont un pic assez haut sur les hautes fréquences. Le noyau obtenu ici a un pic plus étroit, mais des valeurs plus importantes en dehors du pic. On peut noter que l'on obtient un noyau pour lequel la contrainte  $b_k \neq 0$  est donc respectée. Les noyaux correspondant dans le domaine temporel sont représentés sur la figure 2.22 : le noyau obtenu pour cette dernière optimisation (en bleu) prend une forme intéressante, il a un pic important en zéro, ce qui est intéressant pour distinguer les impulsions entre elles, tandis qu'autour du pic on a les oscillations caractéristiques obtenues en solution de (2.154). Intuitivement ce noyau semble être un bon compromis dans le cas pratique entre un noyau de Dirichlet et le sinus parfait que nous avons obtenus dans les autres problèmes.

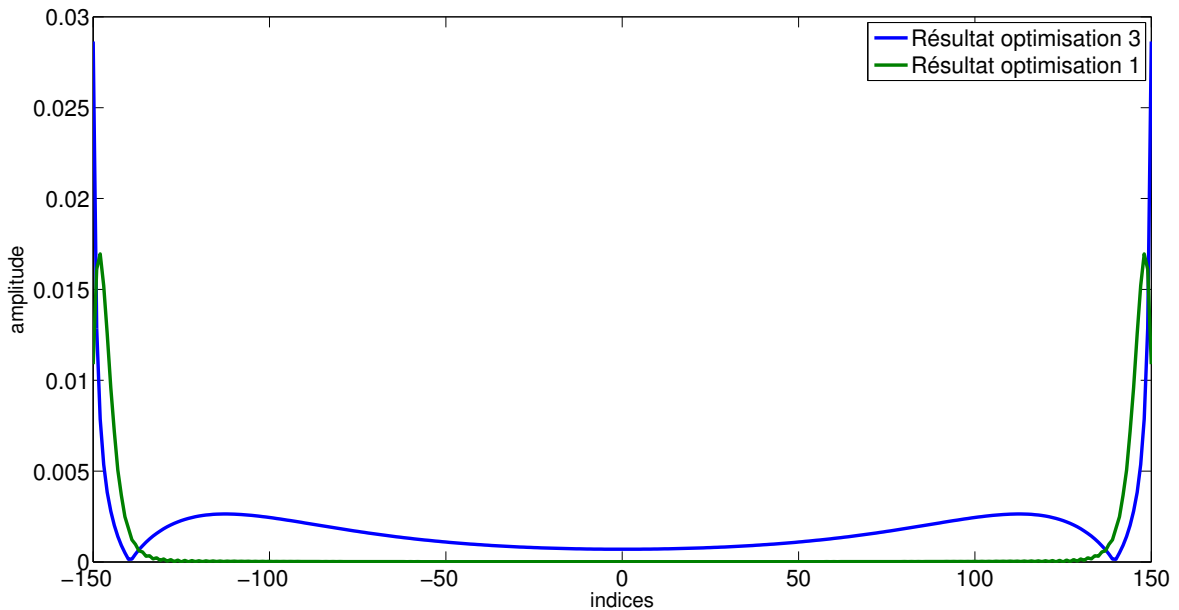


FIGURE 2.21 – Coefficients optimaux résolvant (2.98) en vert et (2.158) en bleu

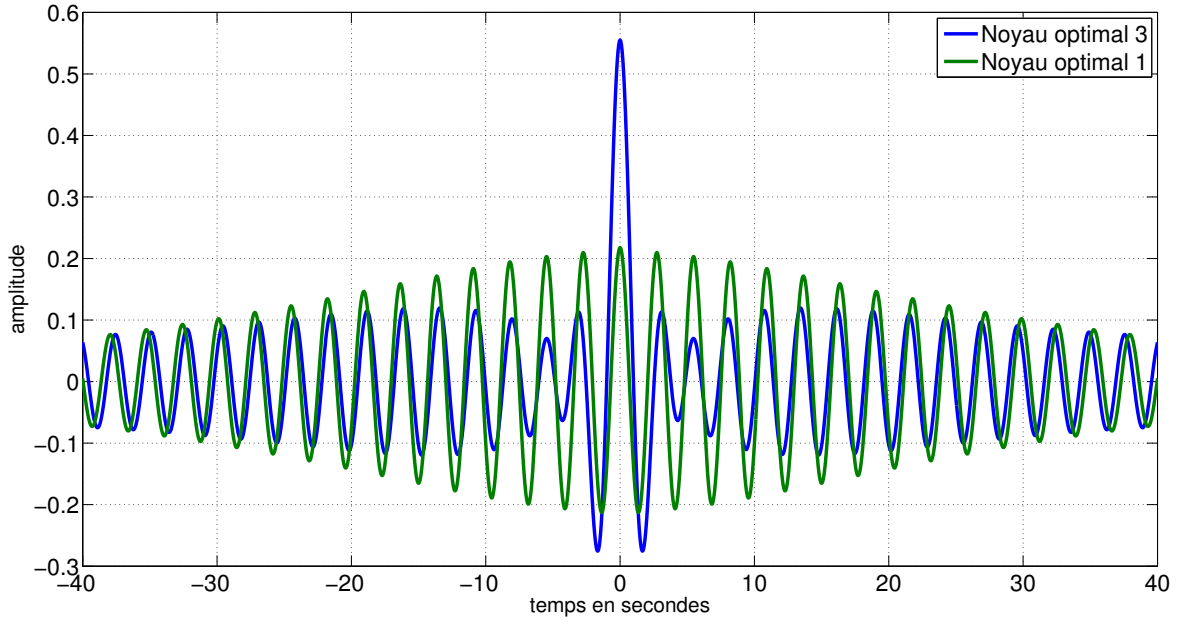


FIGURE 2.22 – Noyaux optimaux solution de (2.98) en vert et (2.158) en bleu

Nous nous intéressons à présent au résultat obtenu sur les bornes de ces deux noyaux, ainsi que le noyau de Dirichlet normalisé par  $\sqrt{\mathbf{b}^T \Delta \mathbf{b}}$  qu'il est toujours intéressant de comparer. Dans le tableau 2.3 figurent la BCRB sur les retards avec le modèle de la section 2.2.2 de l'équation 2.96 et la BCRB sur les retards du modèle appliqué au noyau SoS (2.141). Dans les différentes figures et tableau je considère (2.98) comme le problème d'optimisation 1 et (2.158) comme le problème d'optimisation 3 (le 2 correspondant à (2.154)). On observe que les deux noyaux optimaux ont des performances très proches pour les deux bornes, qui sont bien meilleures que celles du noyau de Dirichlet. Le noyau 1 étant le noyau minimisant 2.96, il obtient de meilleurs résultats sur le calcul de cette borne tandis que le noyau 3 minimise 2.141. Notons que ces valeurs sont inférieures à celles obtenues dans le tableau 2.2 car la normalisation n'est pas la même, la norme ici est inférieure ce qui explique que les bornes sont également inférieures.

|                    | BCRB 2.96              | BCRB 2.141             |
|--------------------|------------------------|------------------------|
| Noyau de Dirichlet | $4.505 \times 10^{-5}$ | $1.124 \times 10^{-2}$ |
| Noyau optimal 1    | $1.629 \times 10^{-5}$ | $3.909 \times 10^{-3}$ |
| Noyau optimal 3    | $1.698 \times 10^{-5}$ | $3.727 \times 10^{-3}$ |

TABLEAU 2.3 – BCRB obtenues par l'optimisation, pour une impulsion de Dirac

Une remarque à faire est qu'avec cette normalisation on ne peut pas choisir  $K = N$ , en effet dans ce cas on a

$$v_{K/2}(t) = \cos\left(\frac{2\pi \frac{K}{2} t}{NT_s}\right) = \cos\left(\frac{\pi t}{NT_s}\right) \quad (2.160)$$

et étant donné qu'on choisit les  $t$  comme des multiples de  $T_s$  pour le calcul de la norme (1.26), celle-ci est une somme de  $\cos(\pi n)$  et est donc théoriquement nulle. Cela donne des valeurs très petites et aléatoires dans la simulation. C'est pourquoi nous avons utilisé un  $K < N$

Il n'est pas possible de résoudre numériquement (2.158) pour  $\epsilon \neq 0$ , car CVX n'accepte pas de résoudre directement le problème avec comme variable inconnue les  $b_k$ . Pour le problème (2.154) nous avons rendu le problème convexe en substituant  $\beta_k = b_k^2$ , ce qui donnerait la contrainte de normalisation  $\sqrt{\mathbf{b}}^T \Delta \sqrt{\mathbf{b}} = 1$  qui n'est pas acceptée par CVX. Cependant le noyau obtenu sans cette contrainte vérifie  $b_k \neq 0$ , ce qui rend la contrainte  $b_k > \epsilon$  peu utile.

Nous nous intéressons maintenant à la forme d'impulsion gaussienne de la figure 2.17. Le noyau obtenu dans ce contexte, pour  $\lambda = 1$  c'est à dire optimisant les retards est montré figure 2.24, les coefficients optimaux correspondant étant représentés figure 2.23. Le noyau obtenu est comparé au noyau de Dirichlet normalisé, on peut observer que le noyau optimal obtenu est similaire au noyau de Dirichlet, avec moins d'oscillations. En fréquence on observe un pic à une fréquence plus basse que pour l'impulsion de Dirac, résultat expliqué par le compromis à trouver entre  $k^2$  et  $H\left(\frac{2\pi k}{NT_s}\right)$ . L'effet de l'impulsion gaussienne sur le résultat de l'optimisation est donc le même qu'avec la normalisation précédente, le noyau obtenu ayant simplement un pic plus bas et plus large que celui résolvant (2.154). Les bornes sur les retards et les amplitudes résultant de cette optimisation sont représentées figure 2.25 et 2.26 respectivement. Comme précédemment nous les comparons à la borne du noyau de Dirichlet qui nous sert de référence. Quand  $\lambda$  augmente on observe que la borne sur les amplitudes diminue tandis que la borne sur les retards augmente, ce qui reflète parfaitement leur importance relative dans l'optimisation.

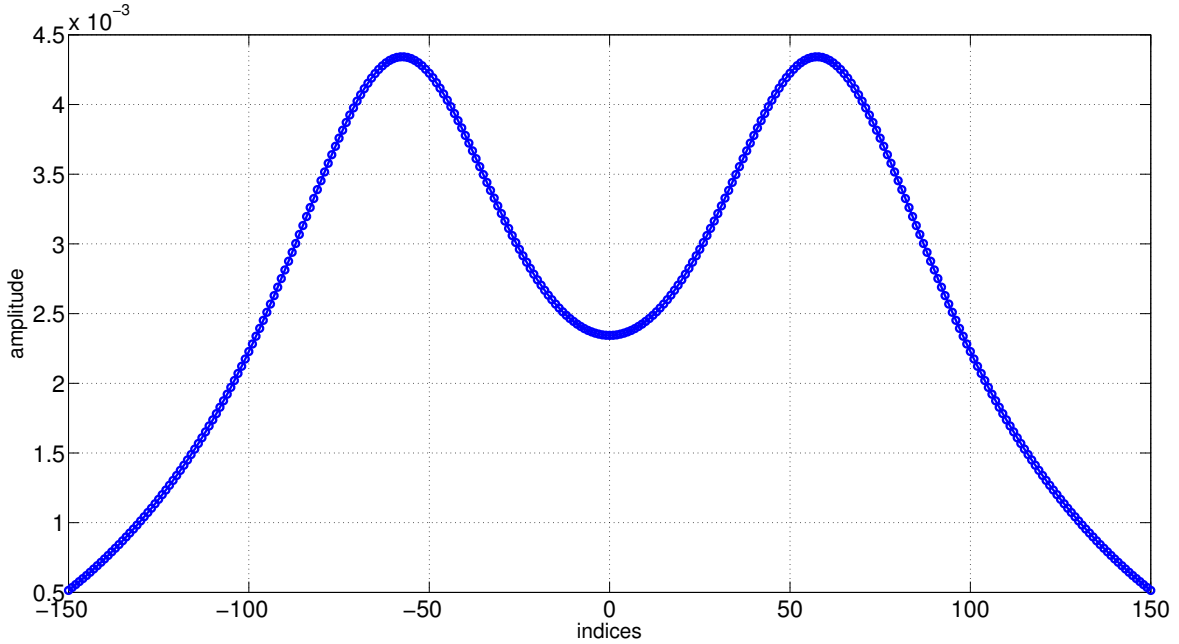


FIGURE 2.23 – Coefficients optimaux pour une impulsion gaussienne et  $\lambda = 1$

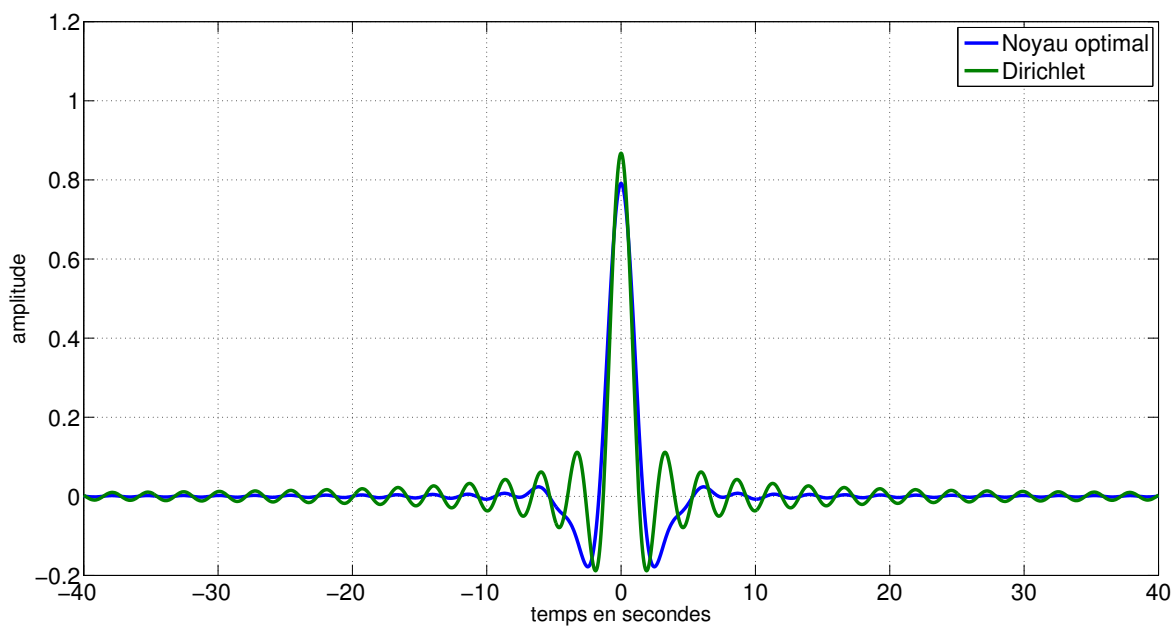


FIGURE 2.24 – Noyau temporel optimal pour une impulsion gaussienne et  $\lambda = 1$

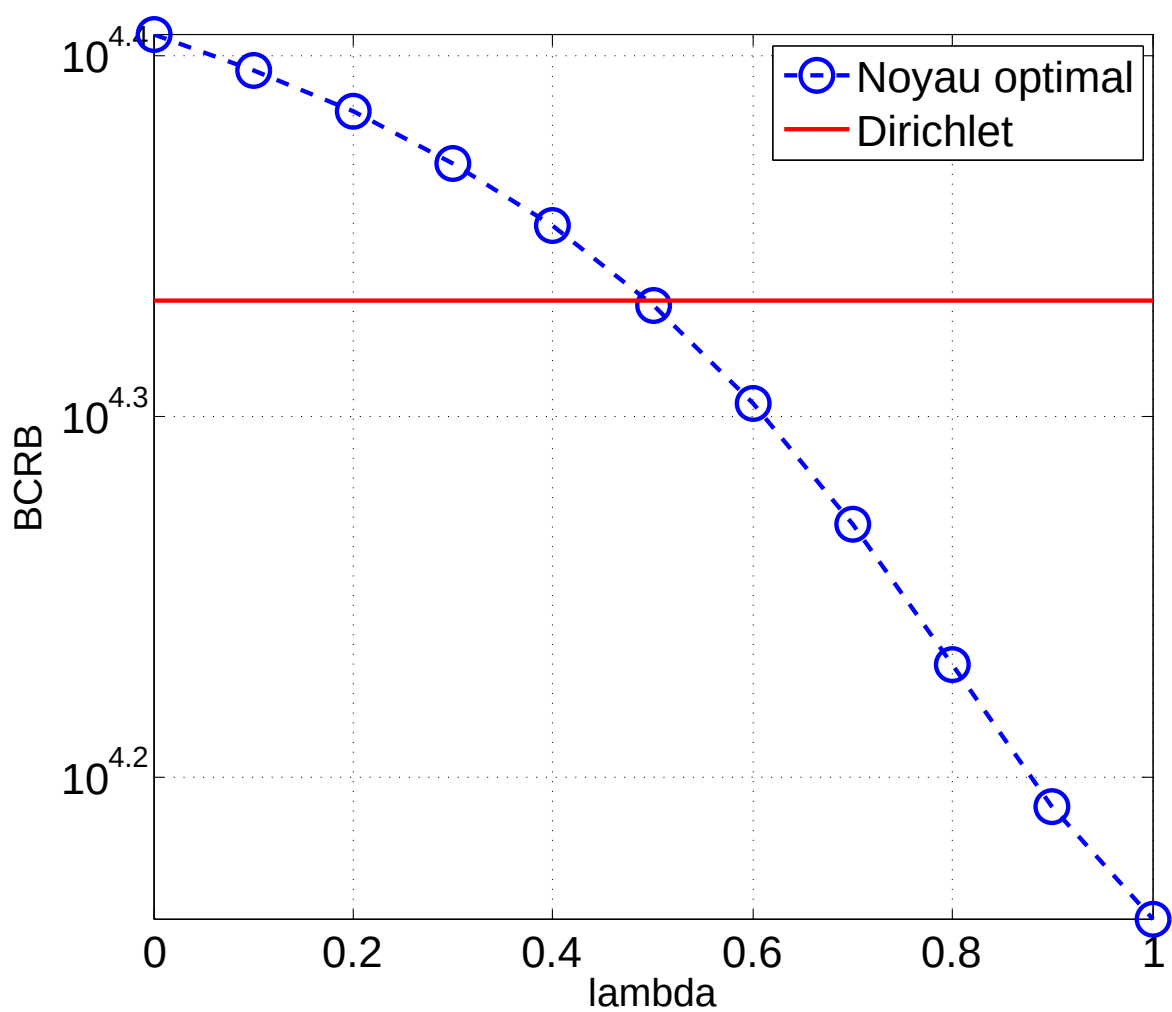


FIGURE 2.25 – BCRB sur les retards, en fonction de  $\lambda$

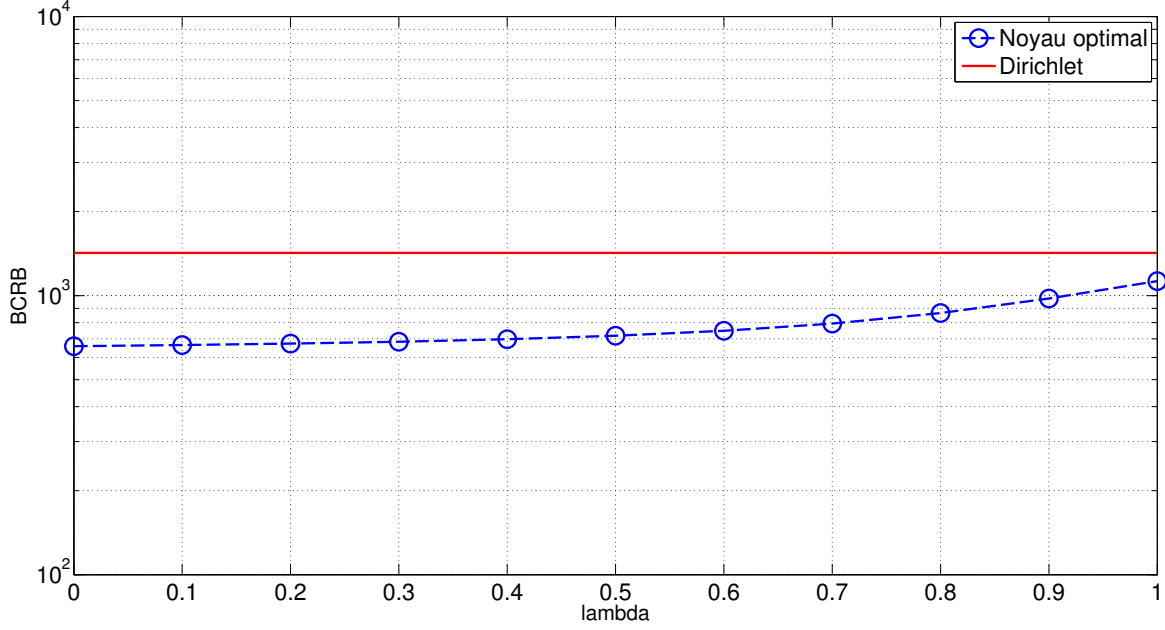


FIGURE 2.26 – BCRB sur les amplitudes, en fonction de  $\lambda$

## 2.4 Estimation par dictionnaire pour les signaux impulsionnels

### 2.4.1 Estimation des signaux à somme de Dirac en discrétisant un dictionnaire

Les signaux à série d'impulsions sont parcimonieux par nature : en effet il y a peu d'impulsions par rapport à toute la durée d'échantillonnage. De ce fait on peut imaginer utiliser l'échantillonnage par dictionnaire présenté dans la section 1.3.3 pour estimer leurs paramètres. En parallèle de nos travaux, cette idée a également été émise par les auteurs de HUANG et al. [72] qui ont développé une approche similaire à la notre ici dans le cas de la reconstruction avec des noyaux reproducteurs d'exponentielles. Les paramètres des signaux impulsionnels sont continus de nature, de ce fait l'estimation par dictionnaire passe par une discrétisation de l'espace des paramètres, ce qui peut réduire les performances d'estimation. Nous nous penchons néanmoins sur cette méthode d'estimation, car nous utiliserons justement ce problème pour illustrer nos résultats sur le problème de la grille dans le cadre de l'acquisition comprimée dans la suite de cette thèse. Nous considérons par simplicité le cas d'une impulsion de Dirac, sachant qu'il est possible d'étendre la méthode à tout type d'impulsions, ce qui n'a pour effet que de modifier le dictionnaire. Le signal que nous cherchons à reconstruire est celui donné dans l'équation (1.6). Les échantillons bruités, obtenus après convolution avec le noyau  $g(t)$  sont exprimés par (2.2) :

$$c_n = \sum_{l=0}^{L-1} a_l g(\tau_l - nT_S) + w_n. \quad (2.161)$$

La durée d'échantillonnage étant de  $NT_S$ , on peut former une grille discrète des temps  $t$  allant de 0 à  $NT_S$  espacés d'un temps  $T'_S$  choisi par l'utilisateur. Plus on choisit  $T'_S$  petit, plus l'estimation sera précise en théorie, cependant réduire le pas de la grille augmente

la cohérence et réduit donc le nombre d'impulsions que l'on peut distinguer tout en augmentant le temps de calcul. Nous notons  $M = NT'_S/T'_S$  la taille de la grille. A partir de cette grille, on construit le dictionnaire  $\Phi$  de dimension  $N \times M$  tel que

$$[\Phi]_{nm} = g(t_m - nT'_S) \quad (2.162)$$

où  $t_m = mT'_S$ . On peut écrire les échantillons sous forme vectorielle et introduire le dictionnaire  $\Phi$  :

$$\mathbf{c} = \Phi \bar{\boldsymbol{\theta}} + \mathbf{w} \quad (2.163)$$

où  $\bar{\boldsymbol{\theta}}$  est un vecteur de longueur  $M$  qui est nul à tous les indices  $m$  pour lesquels  $\forall l, t_m \neq \tau_l$ , et qui contient  $a_l$  aux indices  $m_l$  pour lesquels  $t_{m_l} = \tau_l$ . Dans cette section nous allons utiliser uniquement des retards appartenant à la grille, hypothèse qui est fausse dans le cas de retards continus, mais nous traiterons ce problème par la suite de la thèse. L'équation (2.163) correspond au modèle d'estimation parcimonieux (1.46) :  $\Phi$  et  $\mathbf{c}$  sont connus, et l'on cherche à estimer le vecteur parcimonieux  $\bar{\boldsymbol{\theta}}$ . L'estimation des retards correspond à l'indice des éléments non nuls de  $\bar{\boldsymbol{\theta}}$  tandis que les amplitudes associées à chaque retard seront les valeurs non nulles. On peut retrouver le vecteur  $\bar{\boldsymbol{\theta}}$  à partir de la connaissance de  $\mathbf{c}$  et du dictionnaire  $\Phi$  avec les algorithmes de reconstruction utilisés dans le cadre de l'estimation parcimonieuse tels que l'OMP (voir plus de détails sur cette méthode dans la section 1.3.2).

### 2.4.2 Acquisition comprimée pour la reconstruction des signaux impulsionnels

Nous avons vu dans la section 1.2.2 que le schéma d'acquisition utilisé pour acquérir tout d'abord puis estimer les signaux impulsionnels se basait sur le choix d'un noyau de pré-filtrage par l'utilisateur. Cela implique que l'on se place dans des applications où l'on peut modifier le schéma d'acquisition, ce qui permet également de rajouter une matrice de mesure aléatoire dans le processus. Dans la section 2.4.1 nous avons proposé l'utilisation d'une méthode de type discrète et parcimonieuse pour résoudre le problème d'estimation de ces signaux. Néanmoins ce type d'approche par dictionnaire est soumis aux contraintes fortes sur le modèle, et donc sur la base  $\Phi$  qui pose d'important problèmes de cohérence. La précision d'estimation nécessite une grille relativement dense, qui discrétisée à partir d'un modèle continu donné a une cohérence trop importante pour que les algorithmes de reconstruction basés sur la norme  $\ell_1$  donnent de bon résultats. De manière à réduire la cohérence du dictionnaire, nous proposons de procéder à une étape de compression aléatoire post filtrage qui consiste simplement à multiplier les échantillons non bruités obtenus  $\boldsymbol{\mu}$  par une matrice de mesure  $\Psi$  aléatoire. Cette étape est rajoutée dans le schéma d'acquisition figure 2.27 où l'on note  $\Psi_n$  la  $n$ -ième ligne de la matrice de mesure  $\Psi$ , correspondant au  $n$ -ième échantillon. Dans cette configuration la matrice de mesure  $\Psi$  est de dimension  $N \times K$ , et opère donc la compression attendue. Pour cette raison on peut alors utiliser une base de mesure carrée, c'est à dire que  $\Phi$  est de dimension  $K \times K$ , ce qui permet d'être beaucoup plus précis au niveau de la discrétisation de la grille. On construit dans ce cas la matrice carrée  $\Phi$  en ayant chacune de ses dimensions à une précision d'échantillonnage de  $T'_S$  avec :

$$[\Phi]_{nm} = g((m - n)T'_S). \quad (2.164)$$

Les échantillons ainsi obtenus sont :

$$\mathbf{c} = \mathbf{\Psi}\mathbf{\Phi}\bar{\mathbf{\theta}} + \mathbf{w} = \mathbf{H}\bar{\mathbf{\theta}} + \mathbf{w} \quad (2.165)$$

où l'on peut définir le dictionnaire  $\mathbf{H} \triangleq \mathbf{\Psi}\mathbf{\Phi}$ . Le modèle ainsi obtenu correspond au modèle de l'acquisition comprimée présenté en (1.30) où l'on peut choisir la matrice de mesure, c'est pourquoi on parle d'acquisition comprimée appliquée à la reconstruction des signaux impulsionnels, en opposition avec l'estimation par dictionnaire qui n'introduit pas de matrice aléatoire.

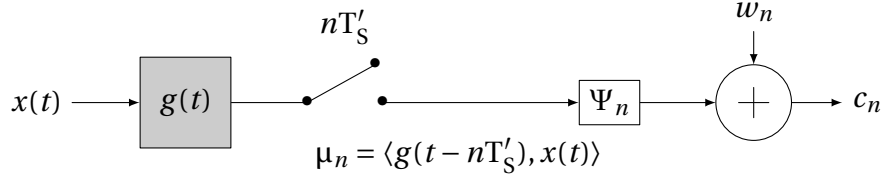


FIGURE 2.27 – Schéma de l'acquisition comprimée d'un signal impulsionnel

Nous utiliserons ce contexte applicatif pour étudier les algorithmes de correction de l'erreur de grille dans le chapitre 4. Nous démontrons ainsi que les amplitudes et les retards peuvent être retrouvés par cette méthode, puisque dans nos simulations nous observons que les erreurs d'estimation à fort RSB proviennent uniquement de l'erreur de modèle, étudiée dans le chapitre 3.

# Chapitre 3

## Robustesse à la méconnaissance du dictionnaire

« Maman, tu finis ta thèse à quelle heure ? »

Elias Gourdin

### Sommaire

|                                                                                        |            |
|----------------------------------------------------------------------------------------|------------|
| <b>3.1 Deux types d'erreur de modèle</b>                                               | <b>80</b>  |
| 3.1.1 Le problème du dictionnaire bruité (DB)                                          | 80         |
| 3.1.2 Le problème de l'erreur de grille (EG)                                           | 81         |
| 3.1.3 Lien entre ces deux modèles                                                      | 82         |
| <b>3.2 Bornes Bayésiennes de Van Trees</b>                                             | <b>83</b>  |
| 3.2.1 BCRB générale pour les deux modèles d'erreur                                     | 83         |
| 3.2.2 Expression analytique des bornes                                                 | 86         |
| 3.2.3 Comparaison des deux modèles                                                     | 88         |
| <b>3.3 BCRM en cas de petites erreurs de grille</b>                                    | <b>93</b>  |
| 3.3.1 Approximation de la BCR déterministe pour une petite erreur                      | 93         |
| 3.3.2 BCRM pour une petite erreur de grille                                            | 95         |
| 3.3.3 Problème de dictionnaire bruité                                                  | 99         |
| <b>3.4 Conséquence de la discrétisation sur l'estimation des fréquences</b>            | <b>104</b> |
| 3.4.1 Présentation du modèle bayésien d'erreur sur le support                          | 104        |
| 3.4.2 Performances à fort RSB des méthodes d'estimation par dictionnaire               | 107        |
| 3.4.3 Exemple de distributions                                                         | 108        |
| 3.4.4 Illustration de l'erreur de grille dans le contexte de l'estimation de fréquence | 112        |

### 3.1 Deux types d'erreur de modèle

Le modèle d'échantillonnage des signaux discrets parcimonieux introduit dans la section 1.3 a été à la base d'un travail important permettant d'établir des propriétés de reconstruction correctes à partir d'un nombre réduit d'échantillons. Cependant nous avons vu dans la section 1.3.5 qu'en pratique le modèle ne représente pas parfaitement la réalité, ce qui introduit des erreurs de modèle et perturbe la reconstruction notamment à fort RSB. Nous souhaitons dans ce chapitre étudier plus en détail les performances asymptotiques en présence d'erreurs. Pour cela nous utilisons deux façons différentes de considérer l'erreur, selon l'étape de l'échantillonnage où l'erreur est introduite. Celles-ci peuvent sembler similaires au premier abord mais nous montrons qu'elles n'aboutissent pas exactement à la même borne dans le contexte Bayésien. Les deux modèles d'erreur permettent de modéliser deux causes différentes de méconnaissance du dictionnaire, se présentant l'une ou l'autre suivant que l'on se place plutôt dans le contexte de l'acquisition comprimée ou de l'estimation parcimonieuse. Le problème de l'erreur de grille apparaissant en pratique dans toutes les applications d'estimation par dictionnaire, est celui le plus souvent traité dans la littérature notamment sous sa forme structurée. Le problème de dictionnaire bruité apparaît plutôt dans le contexte de l'acquisition comprimée. Ces deux erreurs sont la plupart du temps confondues, puisque pour un seul jeu de mesure dans le contexte déterministe elles sont équivalentes, et le terme de méconnaissance de dictionnaire (ou *basis mismatch* en anglais) est alors utilisé pour parler des erreurs de modèle de manière générale. Nous avons mis en évidence l'existence de ces deux modèles dans l'article [20], que nous reprenons et complétons ici.

#### 3.1.1 Le problème du dictionnaire bruité (DB)

Nous présentons tout d'abord le problème de dictionnaire bruité, que nous avons appelé *basis mismatch* dans notre article introduisant les erreurs suivant les deux modèles [20]. Ce modèle est similaire à celui présenté dans ROSENBAUM et TSYBAKOV [103]. On se base sur le modèle (1.33) où l'on considère que les  $N$  échantillons comprimés sont obtenus à partir d'un dictionnaire sur-complet  $\mathbf{H}$  de dimension  $N \times K$  :

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w}. \quad (3.1)$$

Cependant l'algorithme de reconstruction n'a pas une connaissance parfaite du dictionnaire, car celui-ci est bruité, on effectue donc la reconstruction à partir d'un dictionnaire erroné  $\tilde{\mathbf{H}} = \mathbf{H} + \mathbf{E}$ . Ce modèle d'erreur est décrit sous forme de schéma dans la figure 3.1.

Ce type d'erreur apparaît surtout dans le contexte de l'acquisition comprimée. En effet en acquisition comprimée l'utilisateur choisit le dictionnaire, souvent aléatoire, qu'il utilise pour obtenir les échantillons à partir du signal à échantillonner. De ce fait au moment de la compression, le dictionnaire est connu et fixé. Cependant dans certains contextes des erreurs peuvent être introduites au moment de la reconstruction, en effet ce n'est pas toujours le même utilisateur qui opère la compression et la reconstruction et des erreurs peuvent donc apparaître entre ces deux étapes. De même, dans les applications où la matrice de mesure est un modèle servant à représenter une chaîne d'acquisition matérielle, celle-ci peut être soumise à un bruit de traitement. Ce bruit, même faible, a des conséquences sur les performances de reconstruction, comme nous le verrons par la suite dans ce chapitre.

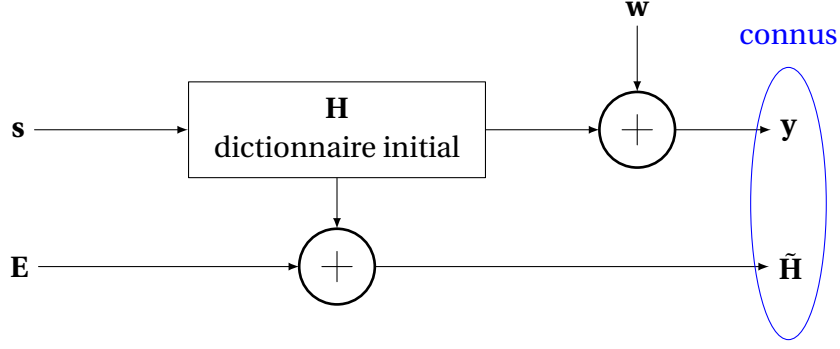


FIGURE 3.1 – Schéma d'échantillonnage comprimé avec dictionnaire bruité

Le signal  $\mathbf{s}$  est aléatoire, pour étudier les performances de reconstruction de manière générale on s'intéresse donc au contexte bayésien, et pas à un seul jeu de mesures  $\mathbf{y}$ . Dans ce contexte, l'erreur sur le dictionnaire est aléatoire, on a donc  $\tilde{\mathbf{H}}$  qui est aléatoire, tandis que la matrice  $\mathbf{H}$  est déterministe puisqu'on garde le même dictionnaire pour toutes les mesures.  $\mathbf{H}$  est dans de nombreuses applications tiré aléatoirement à partir d'une distribution donnée (typiquement gaussienne), ce qui assure que le dictionnaire vérifie les conditions RIP. Cependant cette matrice est tout de même déterministe par rapport aux observations vu qu'elle est choisie une seule fois et que la même matrice est utilisée pour toutes les mesures  $\mathbf{y}$ . Le problème de la reconstruction parcimonieuse en présence d'un dictionnaire bruité consiste à retrouver le vecteur parcimonieux  $\mathbf{s}$  à partir des observations  $\mathbf{y}$  en utilisant le dictionnaire corrompu  $\tilde{\mathbf{H}}$ . On note donc la sortie de l'estimateur  $\hat{\mathbf{s}}(\mathbf{y}, \tilde{\mathbf{H}})$ , le dictionnaire utilisé est considéré comme une observation au même titre que le vecteur de mesure  $\mathbf{y}$ . Dans ce contexte Dictionnaire bruité (DB) la BEQM est donc :

$$\text{BEQM}_{\text{DB}} = \mathbb{E}_{\mathbf{y}, \tilde{\mathbf{H}}, \mathbf{s}} \left[ \|\mathbf{s} - \hat{\mathbf{s}}(\mathbf{y}, \tilde{\mathbf{H}})\|^2 \right]. \quad (3.2)$$

### 3.1.2 Le problème de l'erreur de grille (EG)

Le second type d'erreur qui peut apparaître est une erreur dans la modélisation utilisée pour exprimer les mesures  $\mathbf{y}$  sous la forme (1.33). Dans cette situation, c'est le modèle choisi pour exprimer les échantillons  $\mathbf{y}$  qui ne colle pas avec la réalité, ce qui peut se résumer par le schéma de la figure 3.2. Le cas le plus fréquemment rencontré de ce type d'erreur est le problème de l'erreur de grille, qui apparaît dans toutes les applications utilisant l'estimation parcimonieuse cherchant à estimer un paramètre continu en cherchant des solutions dans une grille discrète. C'est pour cette raison que nous utilisons le terme d'erreur de grille pour parler du problème présenté dans 3.2, bien qu'en réalité ce modèle est plus général puisqu'il couvre toutes les situations où le modèle utilisé pour exprimer  $\mathbf{H}$  ne correspond pas exactement à la réalité des mesures  $\mathbf{y}$ . Pour modéliser ce problème, nous notons les échantillons  $\tilde{\mathbf{y}}$  puisqu'ils sont obtenus à partir du dictionnaire erroné comportant un bruit additif  $\mathbf{E}$  aléatoire. De ce fait, dans le même contexte bayésien où  $\mathbf{s}$  est aléatoire et contrairement au problème de dictionnaire bruité, chaque vecteur d'échantillons reçu par l'algorithme de reconstruction est en réalité obtenu à partir d'un dictionnaire différent, tandis que l'algorithme lui utilise toujours le même dictionnaire déterministe  $\mathbf{H}$  pour opérer la reconstruction. On note le résultat de l'estimateur en présence d'erreur de grille  $\hat{\mathbf{s}}(\tilde{\mathbf{y}})$ .

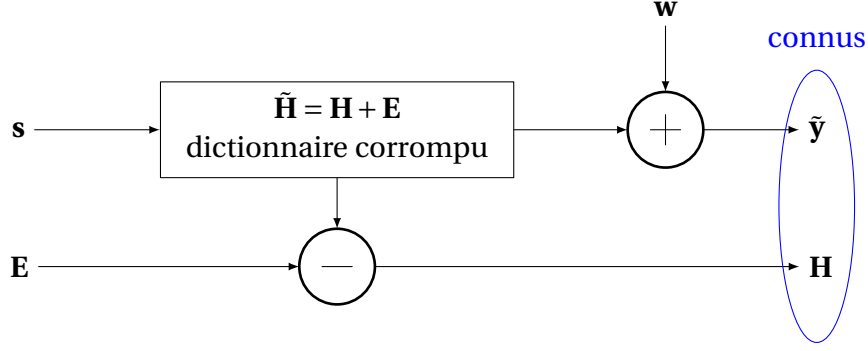


FIGURE 3.2 – Schéma d'échantillonnage comprimé avec erreur de grille

Dans ce contexte Erreur de grille (EG) la BEQM est :

$$\text{BEQM}_{\text{EG}} = \mathbb{E}_{\tilde{\mathbf{y}}, \mathbf{s} | \mathbf{H}} \left[ \|\mathbf{s} - \hat{\mathbf{s}}(\tilde{\mathbf{y}})\|^2 \right]. \quad (3.3)$$

Le dictionnaire utilisé par l'estimateur ne change pas, il est donc une connaissance a priori.

### 3.1.3 Lien entre ces deux modèles

Dans les deux cas on peut écrire les échantillons reçus par l'étape de reconstruction sous une forme très similaire en utilisant le modèle additif d'erreur de dictionnaire  $\tilde{\mathbf{H}} = \mathbf{H} + \mathbf{E}$  :

$$\text{problème DB : } \mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w} = \tilde{\mathbf{H}}\mathbf{s} - \mathbf{E}\mathbf{s} + \mathbf{w}, \quad (3.4)$$

$$\text{problème EG : } \tilde{\mathbf{y}} = \tilde{\mathbf{H}}\mathbf{s} + \mathbf{w} = \mathbf{H}\mathbf{s} + \mathbf{E}\mathbf{s} + \mathbf{w}. \quad (3.5)$$

Les deux problèmes sont à première vue très similaires au niveau des mesures  $\mathbf{y}$  ; le bruit additif est dans les deux cas  $\pm \mathbf{E}\mathbf{s} + \mathbf{w}$  qui est un terme dépendant des données, le signe n'ayant pas d'influence dans un contexte aléatoire où  $\mathbf{E}$  suit une distribution centrée. Dans le contexte déterministe où  $\mathbf{y}$  est un vecteur de mesure donné, ces deux problèmes se rejoignent. Nous allons cependant montrer qu'une différence de performance existe dans un contexte Bayésien et qu'on n'est donc pas dans des situations équivalentes du point de vue de l'estimation, parce qu'elle ne possède pas les mêmes informations dans les deux cas. Dans tout ce chapitre, nous considérerons que  $\mathbf{w}$  est un bruit blanc gaussien de moyenne nulle et de variance  $\sigma^2$ .

Les deux modèles d'erreurs présentés peuvent être rapprochés des problèmes étudiés par WIESEL, ELDAR et YEREDOR [131] qui étudient le modèle à variable aléatoire où  $\mathbf{y}$  est estimé à l'aide d'une matrice aléatoire de moyenne déterministe et connue et de variance connue. Ce modèle correspond à ce que nous avons appelé "erreur de grille", où le dictionnaire connu et discrétisé est la moyenne tandis que la partie aléatoire est modélisée par la matrice  $\mathbf{E}$ . Ce modèle est comparé au modèle EIV où la matrice avec laquelle sont construits les échantillons est déterministe tandis qu'on observe une version bruitée de cette matrice, ce modèle correspond donc à ce que nous nommons le "dictionnaire bruité". Cependant dans WIESEL, ELDAR et YEREDOR [131], l'estimation avec le modèle EIV cherche à estimer la valeur de l'erreur en plus du vecteur de paramètre, tandis que dans le modèle que nous considérons l'erreur est un bruit et non un paramètre à estimer.

Le calcul de la borne que nous effectuons se base sur le même modèle que dans WIESEL, ELDAR et YEREDOR [131], et est donc similaire. Dans cet article les auteurs ont dérivé les bornes déterministes uniquement, nous l'étendons ici au contexte Bayésien et nous nous concentrons sur le problème de l'acquisition comprimée, et donc sur le cas d'un vecteur à estimer parcimonieux contenant  $L$  valeurs non nulles.

## 3.2 Bornes Bayésiennes de Van Trees

Dans cette section, nous étudions la borne Bayésienne de Van Trees [126] sur l'estimation des amplitudes en présence de chacun des deux types d'erreur. Nous étudions le contexte du modèle EIV où l'erreur de dictionnaire additive aléatoire  $\mathbf{E}$  est une matrice dont les valeurs suivent une distribution gaussienne de variance  $\sigma_e^2$  connue. Notre objectif est de comparer les deux modèles, et d'étudier leur différence et leur impact du point de vue des performances. Nous nous placerons dans tous nos travaux dans le contexte asymptotique où les erreurs dues aux perturbations sur le dictionnaire sont prédominantes par rapport aux erreurs dues au bruit. Les bornes oracles c'est à dire où l'on suppose que l'on connaît le support du vecteur parcimonieux correspondent aux performances d'estimateurs existant, et sont donc équivalentes aux bornes contraintes non oracles dans le cas sans erreur de modèles [12, 6, 90]. Nous faisons la même hypothèse pour le contexte des estimations erronées, et nous montrerons par nos simulations que les bornes ainsi développées sont également atteintes en présence d'erreur de modèles par un estimateur ne connaissant pas a priori le support, pour un choix judicieux des paramètres. Pour cette raison nous ferons à présent l'hypothèse d'un support connu, afin de pouvoir donner des expressions analytiques des bornes. Nous donnons tout d'abord l'expression générale de la BCRB pour les deux modèles, puis nous donnons une expression analytique pour une distribution des amplitudes gaussiennes centrées. Enfin nous comparons les deux erreurs en mesurant la différence entre les performances dans les deux situations, différence que nous illustrons par simulation. Nous montrons que les performances dans le contexte de dictionnaire bruité sont meilleures que celles en présence d'erreur de grille et que la différence entre les deux est augmentée avec la perturbation sur la matrice. Les résultats de cette section sont tirés en grande partie de notre article BERNHARDT et al. [20].

### 3.2.1 BCRB générale pour les deux modèles d'erreur

#### BCRB en cas de dictionnaire bruité

La borne bayésienne minimisant la BEQM 3.2 est donnée par la trace de l'inverse de la matrice bayésienne d'information de Fisher, définie par (1.67) :

$$\mathbf{J}_{DB} = \mathbb{E}_{\mathbf{y}, \tilde{\mathbf{H}}, \mathbf{s}} \left\{ - \frac{\partial^2 \log p(\mathbf{y}, \tilde{\mathbf{H}}, \mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right\}. \quad (3.6)$$

Étant donné que  $\tilde{\mathbf{H}}$  et  $\mathbf{s}$  sont indépendants on peut décomposer la distribution jointe :

$$p(\mathbf{y}, \tilde{\mathbf{H}}, \mathbf{s}) = p(\mathbf{y} | \tilde{\mathbf{H}}, \mathbf{s}) p(\tilde{\mathbf{H}}) p(\mathbf{s}). \quad (3.7)$$

La MIB peut donc également être décomposée :

$$\mathbf{J}_{DB} = \mathbb{E}_{\mathbf{y}, \tilde{\mathbf{H}}, \mathbf{s}} \left\{ - \frac{\partial^2 \log p(\mathbf{y} | \tilde{\mathbf{H}}, \mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right\} + \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{s}} \left\{ - \frac{\partial^2 \log p(\tilde{\mathbf{H}})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right\} + \mathbb{E}_{\mathbf{s}} \left\{ - \frac{\partial^2 \log p(\mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right\}, \quad (3.8)$$

où  $p(\tilde{\mathbf{H}})$  ne dépend pas de  $\mathbf{s}$  du coup le deuxième terme s'annule.

On utilise le modèle des échantillons

$$\mathbf{y} = \tilde{\mathbf{H}}\mathbf{s} - \mathbf{E}\mathbf{s} + \mathbf{w}, \quad (3.9)$$

où le bruit  $\mathbf{w}$  est gaussien avec une variance de  $\sigma^2$ . Chaque élément de  $\mathbf{E}$  étant tirés indépendamment les uns des autres d'une distribution gaussienne d'une variance de  $\sigma_e^2$ , on obtient le même modèle des observations que celui étudié dans WIESEL, ELDAR et YEREDOR [131]. De ce fait, les échantillons conditionnés à la connaissance du dictionnaire erroné et des paramètres suivent une distribution gaussienne vectorielle :

$$\mathbf{y}|\tilde{\mathbf{H}}, \mathbf{s} \sim \mathcal{N}(\tilde{\mathbf{H}}\mathbf{s}, (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) \mathbf{I}). \quad (3.10)$$

On peut en déduire la MIF conditionnelle en utilisant la formule de Slepian-Bang (1.60) :

$$\mathbf{F}_{\text{DB}} = \mathbb{E}_{\mathbf{y}|\tilde{\mathbf{H}}, \mathbf{s}} \left\{ -\frac{\partial^2 \log p(\mathbf{y}|\tilde{\mathbf{H}}, \mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right\} = \frac{\tilde{\mathbf{H}}^T \tilde{\mathbf{H}}}{\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2} + \frac{2\sigma_e^4 \mathbf{N} \mathbf{s} \mathbf{s}^T}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2}. \quad (3.11)$$

L'espérance de la MIF est

$$\mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{s}} \{\mathbf{F}_{\text{DB}}\} = \mathbb{E}_{\tilde{\mathbf{H}}} (\tilde{\mathbf{H}}^T \tilde{\mathbf{H}}) \mathbb{E}_{\mathbf{s}} \left( \frac{1}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2} \right) + 2\sigma_e^4 \mathbf{N} \mathbb{E}_{\mathbf{s}} \left( \frac{\mathbf{s} \mathbf{s}^T}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2} \right). \quad (3.12)$$

On peut développer le terme sur la matrice aléatoire :

$$\mathbb{E}_{\tilde{\mathbf{H}}} (\tilde{\mathbf{H}}^T \tilde{\mathbf{H}}) = \mathbb{E}_{\tilde{\mathbf{H}}} (\mathbf{H}^T \mathbf{H} + \mathbf{H}^T \mathbf{E} + \mathbf{E}^T \mathbf{H} + \mathbf{E}^T \mathbf{E}) \quad (3.13)$$

$$= \mathbf{H}^T \mathbf{H} + \mathbb{E}_{\tilde{\mathbf{H}}} (\mathbf{E}^T \mathbf{E}) \quad (3.14)$$

$$= \mathbf{H}^T \mathbf{H} + \mathbf{N} \sigma_e^2 \mathbf{I}, \quad (3.15)$$

où la deuxième égalité découle du fait que  $\mathbf{E}$  a une moyenne nulle, et que  $\mathbf{H}$  est fixée tandis que dans la troisième on introduit le résultat de l'espérance de  $\mathbf{E}^T \mathbf{E}$ . En introduisant (3.15) dans la borne (3.8) on obtient le résultat qui suit.

**Résultat 10.** La BCRB pour le modèle de dictionnaire bruité, avec une erreur additive suivant une distribution gaussienne de variance  $\sigma_e^2$  est :

$$\text{BCRB}_{\text{DB}} = \text{Tr} [\mathbf{J}_{\text{DB}}^{-1}], \quad (3.16)$$

où

$$\mathbf{J}_{\text{DB}} = (\mathbf{H}^T \mathbf{H} + \mathbf{N} \sigma_e^2 \mathbf{I}) \mathbb{E}_{\mathbf{s}} \left( \frac{1}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2} \right) + 2\sigma_e^4 \mathbf{N} \mathbb{E}_{\mathbf{s}} \left( \frac{\mathbf{s} \mathbf{s}^T}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2} \right) + \mathbb{E}_{\mathbf{s}} \left( -\frac{\partial^2 \log p(\mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right). \quad (3.17)$$

### BCRB dans le cas de l'erreur de grille

Pour permettre l'étude des bornes Bayésiennes et la comparaison avec le problème du dictionnaire bruité, nous utilisons le même modèle gaussien de l'erreur  $\mathbf{E}$  pour étudier le problème de l'erreur de grille, bien que cette approche modélise mal le problème concret

de l'erreur de discrétisation. La distribution naturelle pour modéliser l'erreur de discrétisation est la distribution uniforme plutôt que gaussienne, cependant dans le cas haute résolution où la grille est dense, la somme des gaussiennes centrés sur chaque élément de la grille est proche d'une distribution uniforme. Nous verrons dans la section 4.1.1 les bornes pour le modèle d'erreur structuré, qui est plus proche de ce que l'on peut observer comme erreur en pratique lorsque l'on discrétise un dictionnaire généré à partir d'un modèle paramétrique. Dans le contexte de l'erreur de grille, le dictionnaire utilisé par la reconstruction est connu et déterministe, de ce fait on dérive la borne conditionnelle à ce dictionnaire pour minimiser la BEQM (3.3). La MIB correspondante est définie par :

$$\mathbf{J}_{\text{EG}} = \mathbb{E}_{\tilde{\mathbf{y}}, \mathbf{s}} \left\{ -\frac{\partial^2 \log p(\tilde{\mathbf{y}}, \mathbf{s} | \mathbf{H})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right\}. \quad (3.18)$$

La distribution peut être décomposée en

$$p(\tilde{\mathbf{y}}, \mathbf{s} | \mathbf{H}) = p(\tilde{\mathbf{y}} | \mathbf{s}, \mathbf{H}) p(\mathbf{s} | \mathbf{H}) = p(\tilde{\mathbf{y}} | \mathbf{s}, \mathbf{H}) p(\mathbf{s}), \quad (3.19)$$

puisque le dictionnaire et le vecteur parcimonieux à estimer sont indépendants. On en déduit la décomposition de la MIB :

$$\mathbf{J}_{\text{EG}} = \mathbb{E}_{\tilde{\mathbf{y}}, \mathbf{s}} \left\{ -\frac{\partial^2 \log p(\tilde{\mathbf{y}} | \mathbf{s}, \mathbf{H})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right\} + \mathbb{E}_{\mathbf{s}} \left\{ -\frac{\partial^2 \log p(\mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right\}. \quad (3.20)$$

Les échantillons suivant le modèle  $\tilde{\mathbf{y}} = \mathbf{H}\mathbf{s} + \mathbf{E}\mathbf{s} + \mathbf{w}$  suivent une distribution gaussienne de la même façon que pour le problème de dictionnaire bruité :

$$\tilde{\mathbf{y}} | \mathbf{H}, \mathbf{s} \sim \mathcal{N}(\mathbf{H}\mathbf{s}, (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) \mathbf{I}). \quad (3.21)$$

On peut en dériver la MIF à l'aide de la formule de Slepian-Bang :

$$\mathbf{F}_{\text{EG}} = \mathbb{E}_{\mathbf{y} | \mathbf{s}} \left\{ -\frac{\partial^2 \log p(\mathbf{y} | \mathbf{H}, \mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right\} = \frac{\mathbf{H}^T \mathbf{H}}{\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2} + \frac{2\sigma_e^4 \mathbf{N} \mathbf{s} \mathbf{s}^T}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2}. \quad (3.22)$$

L'espérance de la MIF est

$$\mathbb{E}_{\mathbf{s}} \{\mathbf{F}_{\text{EG}}\} = \mathbf{H}^T \mathbf{H} \mathbb{E}_{\mathbf{s}} \left( \frac{1}{\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2} \right) + 2\sigma_e^4 \mathbf{N} \mathbb{E}_{\mathbf{s}} \left( \frac{\mathbf{s} \mathbf{s}^T}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2} \right). \quad (3.23)$$

**Résultat 11.** La BCRB pour le modèle de l'erreur de grille, avec une erreur additive suivant une distribution gaussienne de variance  $\sigma_e^2$  est :

$$\text{BCRB}_{\text{EG}} = \text{Tr} [\mathbf{J}_{\text{EG}}^{-1}], \quad (3.24)$$

où

$$\mathbf{J}_{\text{EG}} = \mathbf{H}^T \mathbf{H} \mathbb{E}_{\mathbf{s}} \left( \frac{1}{\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2} \right) + 2\sigma_e^4 \mathbf{N} \mathbb{E}_{\mathbf{s}} \left( \frac{\mathbf{s} \mathbf{s}^T}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2} \right) + \mathbb{E}_{\mathbf{s}} \left( -\frac{\partial^2 \log p(\mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right). \quad (3.25)$$

On peut remarquer que les bornes pour les deux modèles sont très similaires, dans le cas de l'erreur de dictionnaire bruité  $\mathbf{N}\sigma_e^2 \mathbf{I}$  est ajouté au premier terme de la MIF. On en déduit que

$$\text{BCRB}_{\text{DB}} < \text{BCRB}_{\text{EG}} \text{ pour } \sigma_e \neq 0 \quad (3.26)$$

puisque toutes les matrices impliquées dans le calcul des bornes sont définies positives. Quand  $\sigma_e^2 = 0$  les deux bornes sont égales, ce qui est cohérent puisqu'on trouve avec la borne bayésienne classique pour un dictionnaire connu et sans erreur :

$$\text{BCRB}_{\text{DB}} = \text{BCRB}_{\text{EG}} = \text{Tr} \left[ \frac{\mathbf{H}^T \mathbf{H}}{\sigma^2} + \mathbb{E}_{\mathbf{s}} \left( -\frac{\partial^2 \log p(\mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right) \right]^{-1}. \quad (3.27)$$

### 3.2.2 Expression analytique des bornes

Nous allons à présent détailler ces bornes pour un contexte particulier, permettant d'obtenir une expression analytique. Pour ce faire nous faisons les deux hypothèses suivantes :

1. Nous prenons tout d'abord l'hypothèse d'un faible niveau de bruit, c'est à dire que  $\sigma^2 \rightarrow 0$ . Cette hypothèse est justifiée par le fait que les erreurs dues à la grille ou aux erreurs de modèle ne sont prédominantes que quand le bruit est faible. En effet si  $\sigma^2 \gg \sigma_e^2$ , les erreurs de reconstruction seront importantes, et l'influence de l'erreur de modèle est alors négligeable et ne peut être mesurée. Nous nous plaçons donc dans le contexte où la variance du bruit est dominée par la variance de l'erreur de modèle. Nous nous sommes focalisés uniquement sur cette dernière source d'erreur pendant tout mon travail de thèse.
2. En deuxième lieu, nous choisissons un modèle pour la variable aléatoire  $\mathbf{s}$  afin de pouvoir donner un exemple de forme analytique pour les bornes  $\text{BCRB}_{\text{DB}}$  et  $\text{BCRB}_{\text{EG}}$ .  $\mathbf{s}$  est un vecteur parcimonieux, contenant  $L$  éléments non nuls dont les amplitudes suivent une distribution gaussienne centrée ayant une variance de  $\sigma_s^2$ . Les  $K - L$  autres valeurs de  $\mathbf{s}$  sont nulles. Pour un support connu, et ne variant pas, le terme lié à cette distribution dans la borne est alors :

$$\mathbb{E}_{\mathbf{s}} \left( -\frac{\partial^2 \log p(\mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right) = \mathbb{E}_{\mathbf{s}} \left( -\frac{\partial^2 \log \left( \frac{1}{\sigma_s^2 \sqrt{2\pi}} \right) - \frac{\mathbf{s}^2}{2\sigma_s^2}}{\partial \mathbf{s}^T \partial \mathbf{s}} \right) = \frac{1}{\sigma_s^2} \mathbf{S}, \quad (3.28)$$

où on note  $\mathbf{S}$  la matrice diagonale  $K \times K$  contenant le support de  $\mathbf{s}$  sur sa diagonale c'est à dire un 1 aux  $L$  indices de  $\mathbf{s}$  contenant la variable aléatoire gaussienne, et un 0 ailleurs.

Pour dériver le reste des termes de la borne on cherche à exprimer

$$\mathbf{P}_1(\sigma_e^2, \sigma^2) = \mathbb{E}_{\mathbf{s}} \left( \frac{1}{\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2} \right) \quad (3.29)$$

$$\mathbf{P}_2(\sigma_e^2, \sigma^2) = \mathbb{E}_{\mathbf{s}} \left( \frac{\mathbf{s} \mathbf{s}^T}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2} \right). \quad (3.30)$$

On note tout d'abord que  $\|\mathbf{s}\|^2 = \sum_{l=1}^L s_l^2$  où l'on note  $s_l$  la  $l$ -ième amplitude non nulle du vecteur parcimonieux. De ce fait seules les composantes non nulles entrent en compte dans la norme, et son espérance. Du fait que les  $x_l$  suivent une distribution normale,

$$\frac{\sigma_s^2}{\|\mathbf{s}\|^2} \sim \text{Inv} - \chi_L^2, \quad (3.31)$$

où  $\text{Inv} - \chi_L^2$  est la loi du khi-2 inverse avec  $L$  degré de liberté [17]. On connaît son espérance :

$$\mathbb{E}_{\mathbf{s}} \left( \frac{\sigma_s^2}{\|\mathbf{s}\|^2} \right) = \frac{1}{L-2} \text{ pour } L > 2. \quad (3.32)$$

Suivant l'hypothèse 1. où  $\sigma^2$  est négligeable devant  $\sigma_e^2$  on peut en déduire que :

$$\lim_{\sigma^2 \rightarrow 0} P_1(\sigma_e^2, \sigma^2) = \mathbb{E}_{\mathbf{s}} \left( \frac{1}{\sigma_e^2 \|\mathbf{s}\|^2} \right) = \frac{1}{\sigma_s^2 \sigma_e^2 (L-2)}. \quad (3.33)$$

Quand  $\sigma^2 \rightarrow 0$  on peut approximer  $\mathbf{P}_2(\sigma_e^2, \sigma^2)$  par :

$$\lim_{\sigma^2 \rightarrow 0} \mathbf{P}_2(\sigma_e^2, \sigma^2) = \mathbb{E}_{\mathbf{s}} \left( \frac{\mathbf{ss}^T}{\sigma_e^4 \|\mathbf{s}\|^4} \right). \quad (3.34)$$

Étant donné que  $\|\mathbf{s}\|^4$  est un scalaire et que  $\mathbb{E}_{\mathbf{s}}(\mathbf{ss}^T) = \sigma_s^2 \mathbf{S}$  où  $\mathbf{S}$  est la matrice support on obtient finalement que  $\mathbb{E}_{\mathbf{s}} \left( \frac{\mathbf{ss}^T}{\sigma_e^4 \|\mathbf{s}\|^4} \right) = t\mathbf{S}$  est forcément proportionnel à  $\mathbf{S}$ . On a :

$$\begin{aligned} \text{Tr} \left[ \mathbb{E}_{\mathbf{s}} \left( \frac{\mathbf{ss}^T}{\sigma_e^4 \|\mathbf{s}\|^4} \right) \right] &= \mathbb{E}_{\mathbf{s}} \left( \text{Tr} \left[ \frac{\mathbf{ss}^T}{\sigma_e^4 \|\mathbf{s}\|^4} \right] \right) = \mathbb{E}_{\mathbf{s}} \left( \frac{\text{Tr}[\mathbf{ss}^T]}{\sigma_e^4 \|\mathbf{s}\|^4} \right) \\ &= \mathbb{E}_{\mathbf{s}} \left( \frac{\|\mathbf{s}\|^2}{\sigma_e^4 \|\mathbf{s}\|^4} \right) = \mathbb{E}_{\mathbf{s}} \left( \frac{1}{\sigma_e^4 \|\mathbf{s}\|^2} \right) = \frac{1}{\sigma_s^2 \sigma_e^4 (L-2)}. \end{aligned} \quad (3.35)$$

En utilisant la propriété de proportionnalité on peut en déduire d'une autre façon que

$$\text{Tr} \left[ \mathbb{E}_{\mathbf{s}} \left( \frac{\mathbf{ss}^T}{\sigma_e^4 \|\mathbf{s}\|^4} \right) \right] = Lt, \quad (3.36)$$

d'où

$$\lim_{\sigma^2 \rightarrow 0} \mathbf{P}_2(\sigma_e^2, \sigma^2) = \frac{\mathbf{S}}{\sigma_s^2 \sigma_e^4 L (L-2)}. \quad (3.37)$$

On peut à présent déduire l'expression des deux MIB en introduisant (3.28), (3.33) et (3.37) dans (3.17) et (3.25).

$$\mathbf{J}_{\text{DB}} = \frac{\mathbf{H}^T \mathbf{H}}{\sigma_s^2 \sigma_e^2 (L-2)} + \frac{\mathbf{N} \mathbf{I}}{\sigma_s^2 (L-2)} + \frac{2\mathbf{N}}{\sigma_s^2 L (L-2)} \mathbf{S} + \frac{1}{\sigma_s^2} \mathbf{S} \quad (3.38)$$

$$\mathbf{J}_{\text{EG}} = \frac{\mathbf{H}^T \mathbf{H}}{\sigma_s^2 \sigma_e^2 (L-2)} + \frac{2\mathbf{N}}{\sigma_s^2 L (L-2)} \mathbf{S} + \frac{1}{\sigma_s^2} \mathbf{S}. \quad (3.39)$$

Ce qui permet en inversant d'obtenir le résultat suivant.

**Résultat 12.** Les bornes de Van Trees pour les modèles de dictionnaire bruité et d'erreur de grille, pour un vecteur d'amplitude gaussienne de moyenne nulle et de variance  $\sigma_s^2$  de support connu  $\mathbf{S}$ , une erreur additive  $\mathbf{E}$  aléatoire suivant également une gaussienne de moyenne nulle et de variance  $\sigma_e^2$  et un niveau de bruit négligeable sont :

$$\text{BCRB}_{\text{DB}} = \frac{\sigma_s^2 (L-2)}{N} \text{Tr} \left[ \left( \frac{1}{N \sigma_e^2} \mathbf{H}^T \mathbf{H} + \mathbf{I} + \left( \frac{2}{L} + \frac{L-2}{N} \right) \mathbf{S} \right)^{-1} \right], \quad (3.40)$$

$$\text{BCRB}_{\text{EG}} = \frac{\sigma_s^2 (L-2)}{N} \text{Tr} \left[ \left( \frac{1}{N \sigma_e^2} \mathbf{H}^T \mathbf{H} + \left( \frac{2}{L} + \frac{L-2}{N} \right) \mathbf{S} \right)^{-1} \right]. \quad (3.41)$$

Le support étant connu, les seuls paramètres à estimer sont les  $L$  amplitudes non nulles du vecteur  $\mathbf{s}$ . De ce fait, nous nous intéressons aux bornes réduites sur ce support, que l'on appelle également bornes oracles [Boyer2016, 32]. Pour ce faire nous définissons  $\mathbf{H}_S$  comme étant la matrice  $N \times K$  extraite de  $\mathbf{H}$  contenant les  $L$  colonnes correspondant au support de  $\mathbf{s}$ . Les bornes réduites au support sont :

$$\text{BCRB}_{\text{DB}_S} = \frac{\sigma_s^2(L-2)}{N} \text{Tr} \left[ \left( \frac{1}{N\sigma_e^2} \mathbf{H}_S^T \mathbf{H}_S + \left( 1 + \frac{2}{L} + \frac{L-2}{N} \right) \mathbf{I}_L \right)^{-1} \right], \quad (3.42)$$

$$\text{BCRB}_{\text{EG}_S} = \frac{\sigma_s^2(L-2)}{N} \text{Tr} \left[ \left( \frac{1}{N\sigma_e^2} \mathbf{H}_S^T \mathbf{H}_S + \left( \frac{2}{L} + \frac{L-2}{N} \right) \mathbf{I}_L \right)^{-1} \right], \quad (3.43)$$

où les MIF correspondantes sont :

$$\mathbf{J}_{\text{DB}_S} = \frac{\mathbf{H}_S^T \mathbf{H}_S}{\sigma_s^2 \sigma_e^2 (L-2)} + \frac{N}{\sigma_s^2 (L-2)} \mathbf{I}_L + \frac{2N}{\sigma_s^2 L (L-2)} \mathbf{I}_L + \frac{1}{\sigma_s^2} \mathbf{I}_L \quad (3.44)$$

$$\mathbf{J}_{\text{EG}_S} = \frac{\mathbf{H}_S^T \mathbf{H}_S}{\sigma_s^2 \sigma_e^2 (L-2)} + \frac{2N}{\sigma_s^2 L (L-2)} \mathbf{I}_L + \frac{1}{\sigma_s^2} \mathbf{I}_L. \quad (3.45)$$

### 3.2.3 Comparaison des deux modèles

Pour comparer les erreurs des deux modèles, nous nous concentrons sur la matrice oracle de la BCRB de dimension  $L \times L$ , qui contient sur sa diagonale la borne sur chaque élément amplitude non nulle du vecteur inconnu  $\mathbf{s}$ . Cette matrice est l'inverse de la matrice d'information de Fisher Bayésienne, et sa trace est égale à la BCRB. On la note  $\mathbf{B}_{\text{DB}} = \mathbf{J}_{\text{DB}_S}^{-1}$  et  $\mathbf{B}_{\text{EG}} = \mathbf{J}_{\text{EG}_S}^{-1}$ , respectivement. Pour comparer ces deux matrices nous définissons le critère permettant de mesurer leur écart :

$$\eta \triangleq \|\mathbf{B}_{\text{DB}}^{-1} (\mathbf{B}_{\text{DB}} - \mathbf{B}_{\text{EG}})\|_2. \quad (3.46)$$

La norme de la différence entre  $\mathbf{B}_{\text{DB}}$  et  $\mathbf{B}_{\text{EG}}$  est un critère naturel pour comparer ces deux matrices. On ajoute le terme de normalisation  $\mathbf{B}_{\text{DB}}^{-1}$  car les deux matrices peuvent prendre de très petites valeurs numériques ce qui rend le calcul numérique de la norme de leur différence instable et peu utilisable en pratique. Le terme de normalisation permet de contrebalancer l'effet des petites valeurs sur la différence, et permet de rendre le critère  $\eta$  robuste.

On a la relation suivante entre les deux MIB :

$$\mathbf{J}_{\text{DB}_S} = \mathbf{J}_{\text{EG}_S} + \frac{N}{\sigma_s^2 (L-2)} \mathbf{I}, \quad (3.47)$$

ce qui permet de simplifier (3.46) :

$$\begin{aligned} \eta &= \|\mathbf{I} - \mathbf{B}_{\text{EG}} \mathbf{J}_{\text{DB}_S}\|_2 = \left\| \mathbf{I} - \mathbf{B}_{\text{EG}} \left( \mathbf{J}_{\text{EG}_S} + \frac{N}{\sigma_s^2 (L-2)} \mathbf{I} \right) \right\|_2 \\ &= \frac{N}{\sigma_s^2 (L-2)} \|\mathbf{B}_{\text{EG}}\|_2. \end{aligned} \quad (3.48)$$

où  $\mathbf{J}_{\text{EG}_S}$  est une combinaison linéaire de  $\mathbf{H}_S^T \mathbf{H}_S$  qui par construction est une matrice symétrique et définie positive et de la matrice  $\mathbf{I}$  également définie positive. Par conséquent

$\mathbf{J}_{\text{EG}_S}$  est symétrique et définie positive et son inverse  $\mathbf{B}_{\text{EG}}$  l'est également. De ce fait sa norme est définie par [98] :

$$\|\mathbf{B}_{\text{EG}}\|_2 = \lambda_{\max}(\mathbf{B}_{\text{EG}}), \quad (3.49)$$

où  $\lambda_{\max}(\mathbf{A})$  est la plus grande valeur propre de la matrice  $\mathbf{A}$ . On a

$$\begin{aligned} \eta &= \lambda_{\max} \left( \left( \frac{1}{N\sigma_e^2} \mathbf{H}_S^T \mathbf{H}_S + \left( \frac{2}{L} + \frac{L-2}{N} \right) \mathbf{I} \right)^{-1} \right) = \frac{1}{\lambda_{\min} \left( \frac{1}{N\sigma_e^2} \mathbf{H}_S^T \mathbf{H}_S + \left( \frac{2}{L} + \frac{L-2}{N} \right) \mathbf{I} \right)} \\ &= \frac{N\sigma_e^2}{\lambda_{\min}(\mathbf{H}_S^T \mathbf{H}_S + N\sigma_e^2 \left( \frac{2}{L} + \frac{L-2}{N} \right) \mathbf{I})} = \frac{N\sigma_e^2}{\lambda_{\min}(\mathbf{H}_S^T \mathbf{H}_S) + \sigma_e^2 \left( \frac{2N}{L} + L - 2 \right)}. \end{aligned} \quad (3.50)$$

**Résultat 13.** La différence entre les performances de l'estimation en présence d'une erreur de dictionnaire bruité ou d'erreur de grille, mesurée suivant le critère (3.46) est :

$$\eta = \frac{1}{\frac{\lambda_{\min}(\mathbf{H}_S^T \mathbf{H}_S)}{N\sigma_e^2} + \frac{2}{L} + \frac{L-2}{N}}. \quad (3.51)$$

De manière logique, quand l'erreur est nulle ( $\sigma_e^2 = 0$ ), la différence entre les deux bornes est nulle également, cela se voit directement de l'expression (3.50). Comme  $\mathbf{H}_S^T \mathbf{H}_S$  est définie positive,  $\lambda_{\min}(\mathbf{H}_S^T \mathbf{H}_S) > 0$ , à partir de la formulation donnée en (3.51) on observe par conséquent que  $\eta$  est une fonction croissante par rapport à la variance de l'erreur  $\sigma_e^2$  pour des  $N$  et  $L$  finis. Cela montre que pour une erreur suffisamment importante, les deux types d'erreurs que nous avons présentés en section 3.1 ne sont pas équivalents.

Nous considérons à présent le cas particulier où les éléments de  $\mathbf{H}_S$  sont tirés aléatoirement suivant une distribution gaussienne de variance  $1/L$ , pour ce type de dictionnaire la théorie des matrices aléatoires, permet de donner une expression approximative simple de cette différence, dans le cas asymptotique où  $N, L$  sont très grand avec  $\frac{N}{L} \rightarrow \rho_{\text{dic}}$ .

Un des résultats centraux de la théorie des matrices aléatoires est la loi de Marčcenko-Pastur. Dans un contexte asymptotique, où  $N, L \rightarrow \infty$  avec  $\frac{L}{N} \rightarrow c$  c'est à dire que les dimensions de la matrices sont infiniment grandes, tout en conservant un rapport constant et pour  $\mathbf{H}$  de dimension  $N \times L$  dont les éléments sont tirés d'une distribution gaussienne de moyenne nulle et de variance  $\frac{1}{L}$ , la distribution empirique des valeurs propres de  $\mathbf{H}^T \mathbf{H}$  converge de façon presque sûre vers une distribution précise appelée de Marčcenko-Pastur [119]. De cette propriété on peut déduire plusieurs relations utiles :

$$\lim_{L \rightarrow \infty} \frac{1}{L} \text{Tr} \left[ (\mathbf{H}^T \mathbf{H})^{-1} \right] = \frac{1}{c-1}, \quad (3.52)$$

$$\lim_{L \rightarrow \infty} \frac{1}{L} \text{Tr} \left[ (\mathbf{H}^T \mathbf{H})^{-2} \right] = \frac{c}{(c-1)^3}. \quad (3.53)$$

On connaît aussi les propriétés asymptotiques des valeurs propres de  $\mathbf{H}^T \mathbf{H}$ , dans les mêmes conditions :

$$\lim_{L \rightarrow \infty} \lambda_{\min}(\mathbf{H}^T \mathbf{H}) = (1 - \sqrt{c})^2 \quad (3.54)$$

$$\lim_{L \rightarrow \infty} \lambda_{\max}(\mathbf{H}^T \mathbf{H}) = (1 + \sqrt{c})^2. \quad (3.55)$$

Pour  $N$  et  $L$  assez grand, on peut introduire le résultat (3.54) dans l'expression de la différence entre les performances DB et EG (3.50) :

$$\eta \approx \frac{N\sigma_e^2}{(1 - \sqrt{\frac{1}{\rho_{\text{dic.}}}})^2 + \sigma_e^2(2\rho_{\text{dic.}} + L - 2)}, \quad (3.56)$$

qui est une valeur finie, puisque le rapport  $N/L$  reste une valeur finie même quand  $N$  et  $L$  tendent vers l'infini. La différence  $\eta$  augmente toujours proportionnellement avec la variance de l'erreur.

Pour mieux visualiser le comportement des bornes et leurs différences, nous avons tracé leur comportement pour des exemples de jeux de paramètres, pour un dictionnaire tiré aléatoirement à partir d'une distribution gaussienne. Nous prenons pour cela une distribution gaussienne de  $\mathbf{s}$  de variance  $\sigma_s^2 = 1$ . Le choix de cette variance n'influence pas le comportement des bornes étant donné qu'elles sont simplement proportionnelles à  $\sigma_s^2$ . Nous avons également choisi un niveau de bruit  $\sigma^2 = 0.1$  relativement faible, pour respecter les conditions de fort RSB. La figure 3.3 compare les bornes (3.42) et (3.43) pour un dictionnaire  $\mathbf{H}_S$  de dimensions  $100 \times 5$  tracées en fonction du logarithme de la variance de l'erreur. On peut tout d'abord confirmer sur cette courbe que les deux bornes sont confondues quand l'erreur tend vers zéro et que la différence entre les deux bornes augmente avec la variance de l'erreur. Les performances que l'on peut obtenir avec un dictionnaire bruité au moment de la réception sont bien meilleures que celles que l'on peut espérer en présence d'une erreur de grille, la comparaison des bornes donne à penser que le problème de dictionnaire bruité est plus simple que le problème de l'erreur de grille du point de vue de l'estimation. On peut également observer une saturation des bornes pour un  $\sigma_e^2$  assez grand, l'erreur dégrade les performances mais celles-ci atteignent un palier haut. Dans les figures 3.4, 3.5 et 3.6 nous nous intéressons exclusivement à l'écart  $\eta$  entre les deux bornes. Pour cela sont tracés  $\eta$  suivant la formule (3.46) calculée numériquement en bleu, (3.50) qui est la formule analytique générale en vert et enfin la formule asymptotique (3.56) pour des dictionnaire grands en rouge. Naturellement ces courbes confirment que la différence augmente avec la variance de l'erreur. De plus les deux courbes numérique et analytique sont confondues pour toutes les tailles de dictionnaire, ce qui valide le calcul effectué. La courbe de la différence asymptotique quant à elle est légèrement différente des autres quand on prend un dictionnaire de taille  $100 \times 20$  suivant les tirages du dictionnaire choisi. C'est un exemple où la valeur propre calculée numériquement n'est pas égale à la valeur propre asymptotique car le dictionnaire est trop petit pour entrer dans les conditions où la relation (3.54) est vérifiée. La formule très simple asymptotique est par contre confondue avec les autres formulations pour la deuxième taille de dictionnaire  $1000 \times 120$  testée. Nous avons vu sur la figure 3.3 que les deux bornes atteignaient un palier pour une erreur grande, il est donc logique que  $\eta$  atteigne ce palier également. On peut noter que pour le deuxième dictionnaire, l'erreur a un ordre de grandeur plus important que pour le premier. La taille du dictionnaire, et surtout le rapport  $\rho_{\text{dic.}}$  joue un rôle important sur la différence entre les bornes, on peut le remarquer notamment via l'expression asymptotique (3.56). Pour illustrer cette relation nous traçons également  $\eta$  (suivant ses trois formulations) en fonction du rapport  $L/N = 1/\rho_{\text{dic.}}$ . Pour ce faire nous avons fixé  $\sigma_e^2 = 1$  et  $N = 1000$  puis nous avons fait varier  $L$  entre 1 et  $N$  puisque par définition nous devons maintenir  $L < N$ , le rapport est donc compris entre 0 et 1. Le comportement de cette courbe n'est pas intuitif, en effet la différence entre les deux bornes est

croissante au départ puis atteint un maximum pour un rapport de  $L/N = 0.05$  c'est à dire  $\rho_{\text{dic.}} = 20$  puis diminue lentement jusqu'à  $\rho_{\text{dic.}} = 1$  où l'on a  $\eta = \rho_{\text{dic.}} = 1$

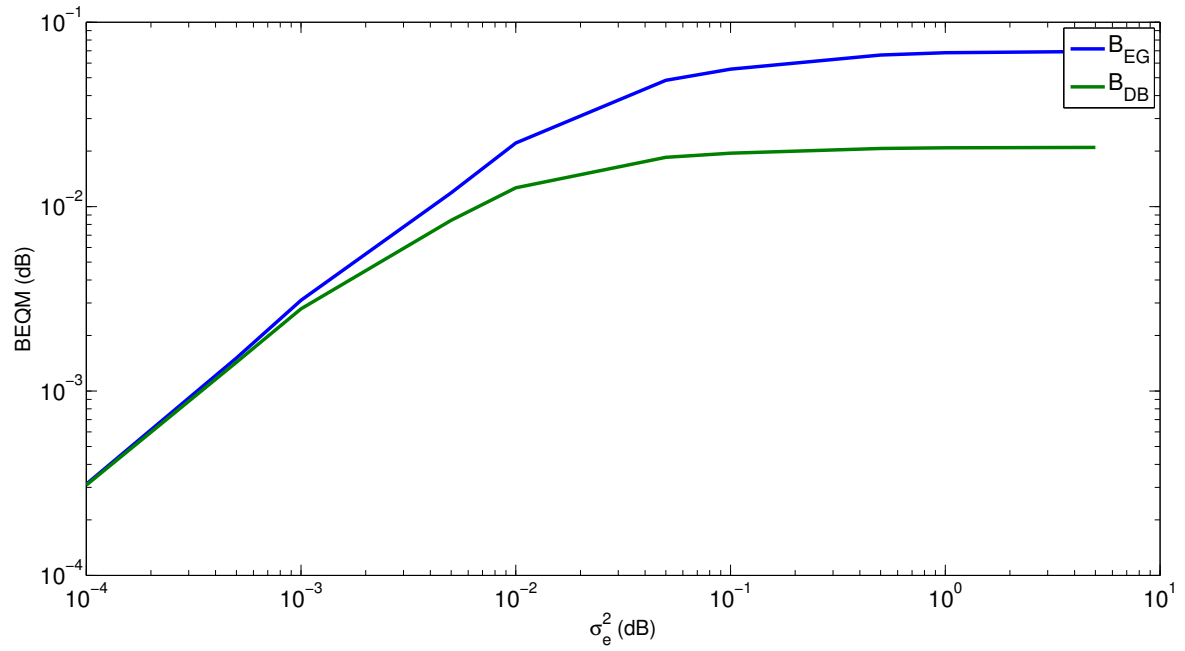


FIGURE 3.3 – BCRB EG et DB en fonction de la variance de l'erreur, pour  $N = 100$  et  $L = 5$

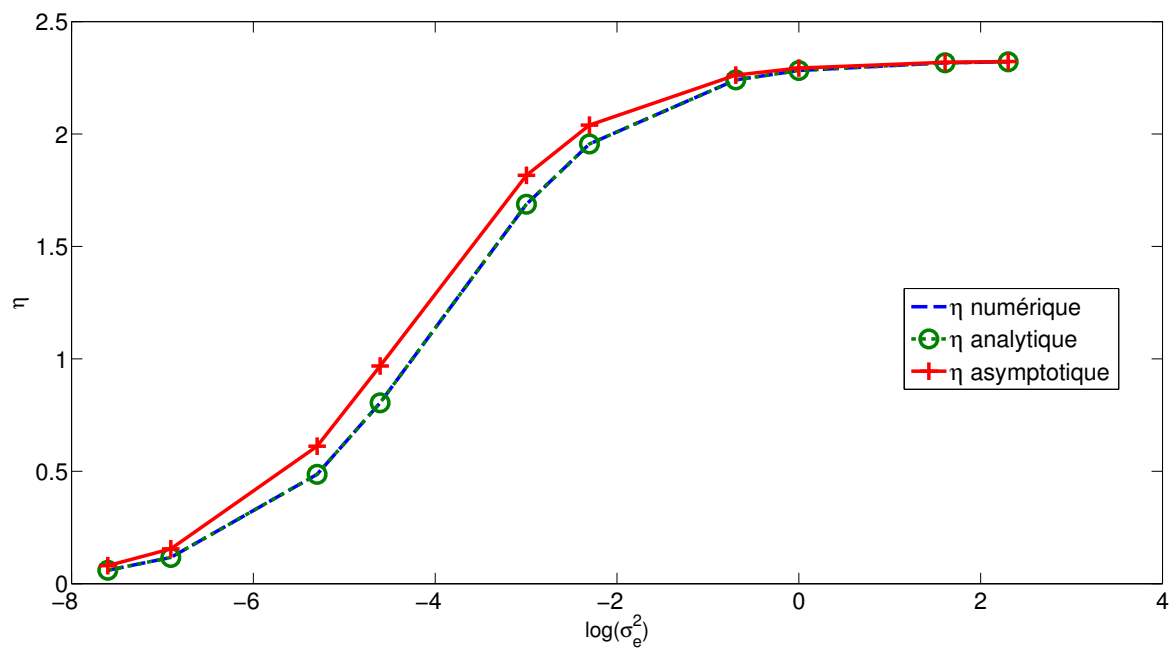


FIGURE 3.4 – Critère de différence entre les bornes  $\eta$  en fonction de la variance de l'erreur, pour  $N = 100$  et  $L = 5$

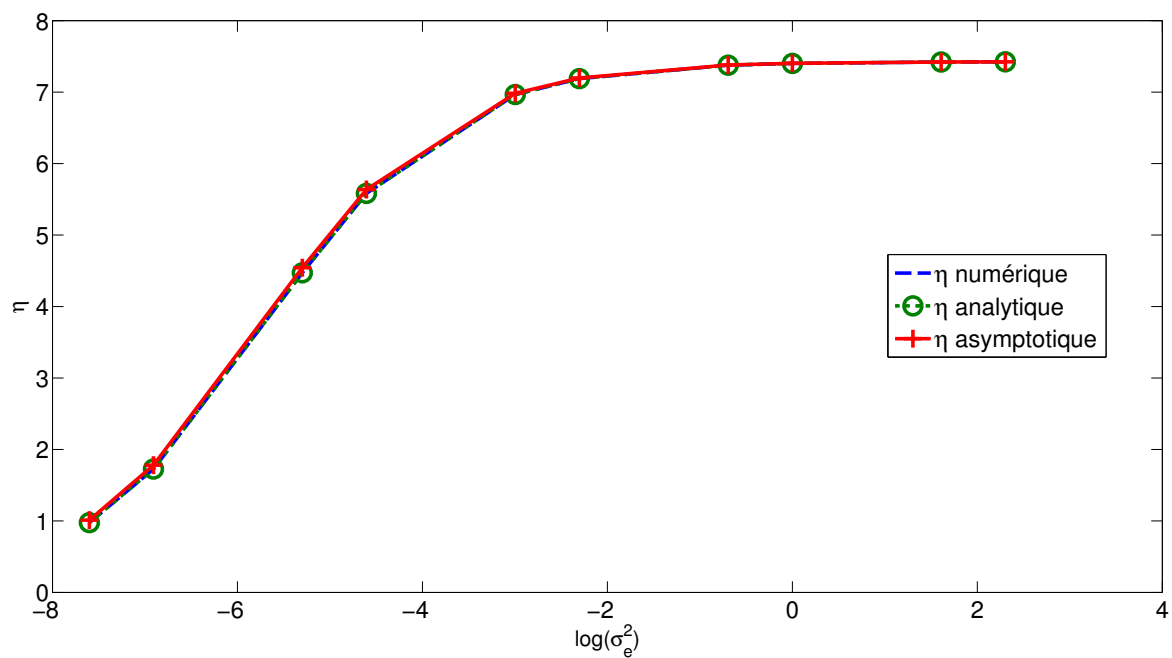
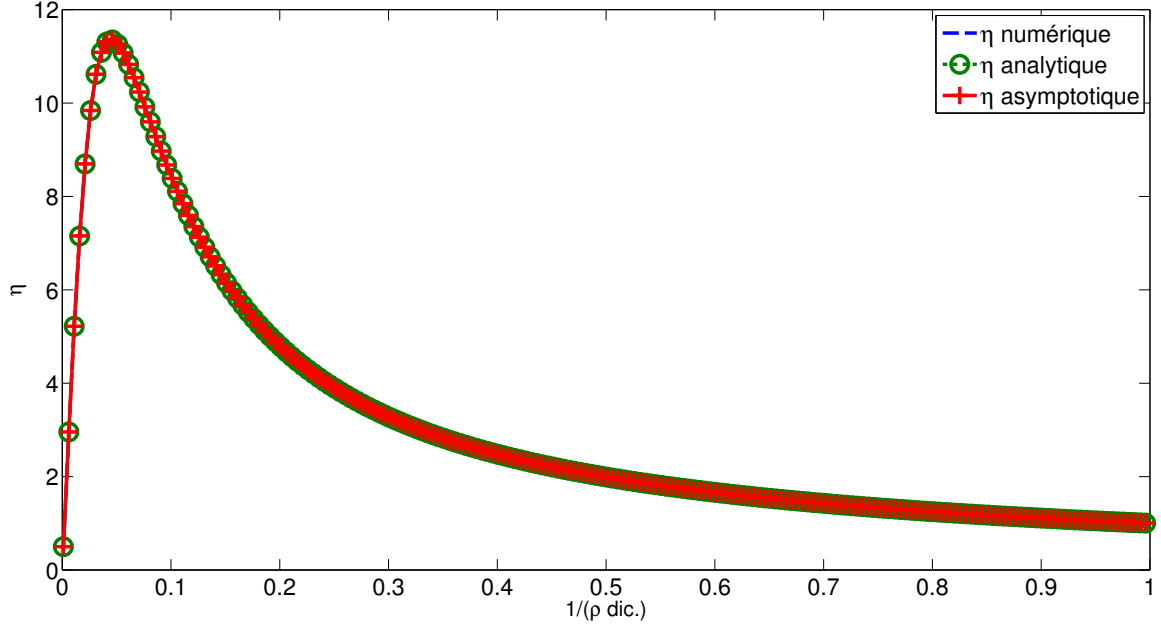


FIGURE 3.5 – Critère de différence entre les bornes  $\eta$  en fonction de la variance de l'erreur, pour  $N = 1000$  et  $L = 120$


 FIGURE 3.6 – Critère de différence entre les bornes  $\eta$  en fonction du rapport  $1/\rho_{\text{dic.}}$ , pour  $\sigma_e^2 = 1$ 

### 3.3 BCRM en cas de petites erreurs de grille

Nous nous intéressons à présent à une autre borne Bayésienne, qui est la borne BCRM (1.68) et qui est définie comme la moyenne de la borne BCR déterministe. L'étude de cette borne permet d'obtenir des résultats plus proches des résultats des estimateurs que la BCRB de Van Trees, tout en minimisant la BEQM. En l'absence d'erreur de grille les deux bornes bayésiennes sont égales puisque dans ce cas la MIF ne dépend pas du paramètre  $\mathbf{s}$  [67]. Quand  $\sigma_e^2 \neq 0$ , l'erreur de grille s'ajoute au terme de bruit qui devient dépendant à la fois de l'erreur et de  $\mathbf{s}$ , ce qui modifie les statistiques du bruit et conduit à une différence entre la BCRM et la BCRB. Il n'est pas possible de donner une expression analytique de la BCRM dans le cas général, mais nous dérivons une approximation de la borne dans le cas d'une petite erreur de grille dans le contexte EG. La borne déterministe générale est donnée dans WIESEL, ELDAR et YEREDOR [131], nous en faisons une approximation et la moyennons pour obtenir la borne BCRM pour une petite erreur de grille. Pour ce faire nous faisons la même hypothèse que WIESEL, ELDAR et YEREDOR [131] de l'indépendance entre le bruit additif  $\mathbf{w}$  et la matrice d'erreur de modèle  $\mathbf{E}$ , qui est toujours gaussienne. Nous comparons ensuite le comportement de la BCRB et de la BCRM par simulation, et nous montrons que la BCRM étant une borne plus serrée, on peut avoir une bonne prédiction du comportement des estimateurs pour toutes les plages de  $\sigma_e^2$ , notamment dans le cas de  $\sigma_e^2$  très petit où la BCRB sur-estime largement les performances.

#### 3.3.1 Approximation de la BCR déterministe pour une petite erreur

Pour donner la BCRM, il faut d'abord déterminer la BCR déterministe pour un vecteur  $\mathbf{s}$  donné. Nous nous intéressons au problème de l'erreur de grille où

$$\tilde{\mathbf{y}}|\mathbf{H}, \mathbf{s} \sim \mathcal{N}(\mathbf{H}\mathbf{s}, (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) \mathbf{I}). \quad (3.57)$$

Nous rappelons que :

$$\text{BCRM}_{\text{EG}} = \mathbb{E}_{\mathbf{s}} \text{Tr} [\mathbf{F}_{\text{EG}}^{-1}] \quad (3.58)$$

où  $\mathbf{F}_{\text{EG}}$  est la MIF donnée grâce à la formule de Slepian-Bang [131] :

$$\mathbf{F}_{\text{EG}} = \frac{\mathbf{H}^T \mathbf{H}}{\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2} + \frac{2\sigma_e^4 \mathbf{N} \mathbf{s} \mathbf{s}^T}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2}, \quad (3.59)$$

qui après inversion donne :

$$\mathbf{F}_{\text{EG}}^{-1} = (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) (\mathbf{H}^T \mathbf{H})^{-1} - \Delta, \quad (3.60)$$

avec

$$\Delta = \frac{2\sigma_e^4 \mathbf{N} (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{s} \mathbf{s}^T (\mathbf{H}^T \mathbf{H})^{-1}}{1 + 2\sigma_e^4 \mathbf{N} \mathbf{s}^T (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{s}}. \quad (3.61)$$

Cette borne a une expression difficile à exploiter, nous allons en donner une expression approximative. Pour ce faire nous considérons le cas où  $\sigma_e^4 \rightarrow 0$ . On cherche l'approximation de

$$h(\sigma_e^4) = \frac{2\sigma_e^4 \mathbf{N}}{1 + 2\sigma_e^4 \mathbf{N} \mathbf{s}^T (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{s}}, \quad (3.62)$$

à l'aide du développement de Taylor

$$h(\sigma_e^4) = \left. \frac{\partial h(\sigma_e^4)}{\partial \sigma_e^4} \right|_{\sigma_e^4=0} \sigma_e^4 + 0(\sigma_e^8). \quad (3.63)$$

On a la dérivée

$$\frac{\partial h(\sigma_e^4)}{\partial \sigma_e^4} = \frac{2\mathbf{N}}{(1 + 2\sigma_e^4 \mathbf{N} \mathbf{s}^T (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{s})^2}, \quad (3.64)$$

qui permet d'obtenir le développement  $h(\sigma_e^4) = 2\mathbf{N}\sigma_e^4 + 0(\sigma_e^8)$  et :

$$\Delta = 2\mathbf{N}\sigma_e^4 (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{s} \mathbf{s}^T (\mathbf{H}^T \mathbf{H})^{-1} + 0(\sigma_e^8). \quad (3.65)$$

L'inverse de la MIF déterministe peut donc être approximée par :

$$\mathbf{F}_{\text{EG}}^{-1} = ((\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) (\mathbf{H}^T \mathbf{H})^{-1}) (\mathbf{I} - 2\mathbf{N}\sigma_e^4 \mathbf{s} \mathbf{s}^T (\mathbf{H}^T \mathbf{H})^{-1}) + 0(\sigma_e^8). \quad (3.66)$$

**Résultat 14.** La BCR déterministe pour une petite erreur de grille peut être approximée par :

$$\text{BCR}_{\text{EG}} \approx \text{Tr} [((\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) (\mathbf{H}^T \mathbf{H})^{-1}) (\mathbf{I} - 2\mathbf{N}\sigma_e^4 \mathbf{s} \mathbf{s}^T (\mathbf{H}^T \mathbf{H})^{-1})]. \quad (3.67)$$

Pour donner une idée de l'ordre de grandeur de  $\sigma_e^2$  pour laquelle l'approximation faite à partir du développement de Taylor est valable, nous traçons la BCR déterministe exacte en calculant numériquement la trace de la MIF (3.60) que nous comparons à la BCR approximée donnée en (3.67). Nous nous intéressons dans la simulation à la BCR sur les paramètres non nuls du vecteur  $\mathbf{s}$ , nous utilisons donc la matrice  $\mathbf{H}_S$  de dimension  $N \times L$  dont les éléments sont tirés aléatoirement d'une distribution normale de moyenne 1 et de variance  $1/L$ . Les amplitudes à estimer sont tirées également à partir d'une distribution normalement de moyenne nulle et de variance  $1/L$ . Les bornes ainsi obtenues sont tracées en fonction de la variance de l'erreur pour deux jeux de paramètres différents sur

les figures 3.7 et 3.8, la borne exacte est tracée en bleu tandis que la borne approximative est tracée en rouge. Il est premièrement intéressant de noter que la borne approximée décroche de la borne exacte pour le même niveau de bruit, quelque soit les paramètres de  $N$ ,  $L$  et  $\sigma_e^2$  choisis. Nous avons tracé ici deux exemples pour l'illustrer, mais ce comportement a été observé sur toutes les simulations effectuées. La borne approximée colle parfaitement avec la borne exacte pour  $\sigma_e^2 \leq 5 \cdot 10^{-2}$ . Au delà de  $\sigma_e^2 = 0.1$  la borne approximative ne peut plus être calculée car le terme en  $2N\sigma_e^2$  devient supérieur au premier terme, la borne devient alors négative ce qui n'a pas de sens. L'approximation faite est donc assez contraignante, puisqu'elle n'est valable que pour des variances d'erreur très faibles.

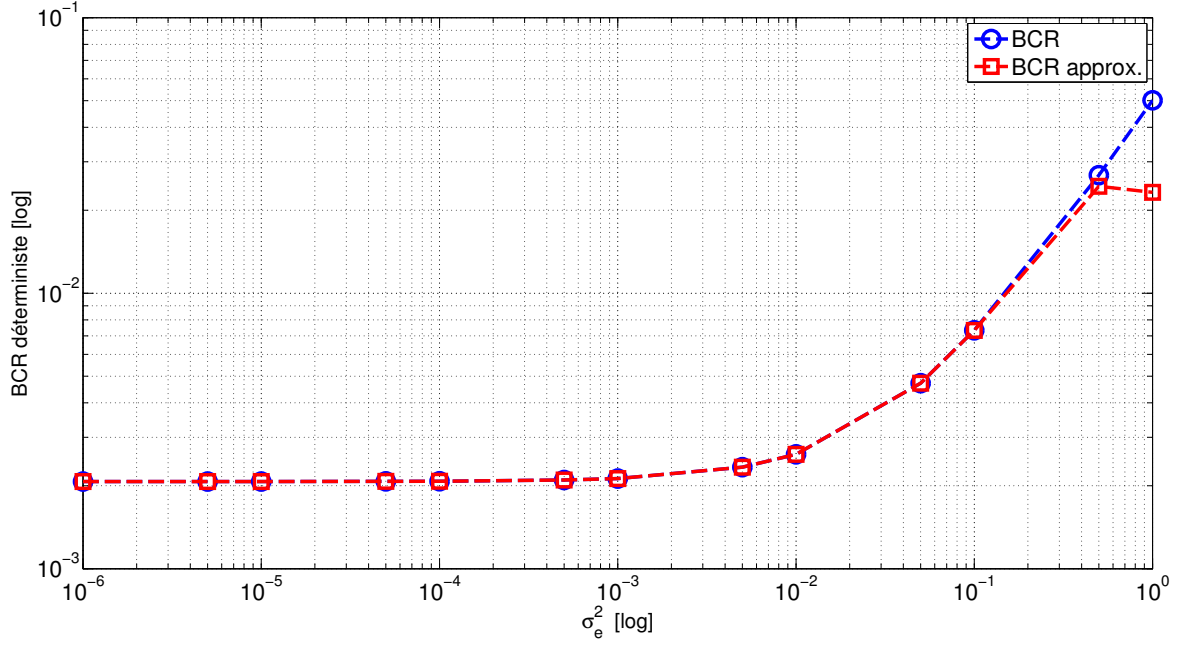


FIGURE 3.7 – BCR déterministe en fonction de la variance de l'erreur, pour  $N = 500$ ,  $L = 10$  et  $\sigma^2 = 0.01$

### 3.3.2 BCRM pour une petite erreur de grille

Il est à présent possible de calculer la BCRM en moyennant la BCR déterministe approximée pour les différentes valeurs de  $\mathbf{s}$ . Pour ce faire on fait l'hypothèse que celui-ci est un vecteur parcimonieux, dont les  $L$  valeurs non nulles  $s_l$  sont des variables aléatoires de moyenne nulle, de variance  $\sigma_s^2/L$  et l'on définit leur moment d'ordre 4 par :  $\mathbb{E}_{s_l}\{s_l^4\} = m_s^4/L^2$ . Contrairement à la borne de Van Trees, on n'a pas besoin de connaître en détail la distribution de  $s_l$ , mais seulement sa moyenne et ses moments d'ordre 2 et 4. De plus on choisit une variance de  $\sigma_s^2/L$  de telle sorte que  $\mathbb{E}\{\|\mathbf{s}\|^2\} = \sigma_s^2$  soit une quantité finie, même quand on choisit des  $L$  infinis, cette propriété est importante puisque nous étudierons par la suite les résultats asymptotiques, pour lesquels  $N$  et  $L$  sont très grands.

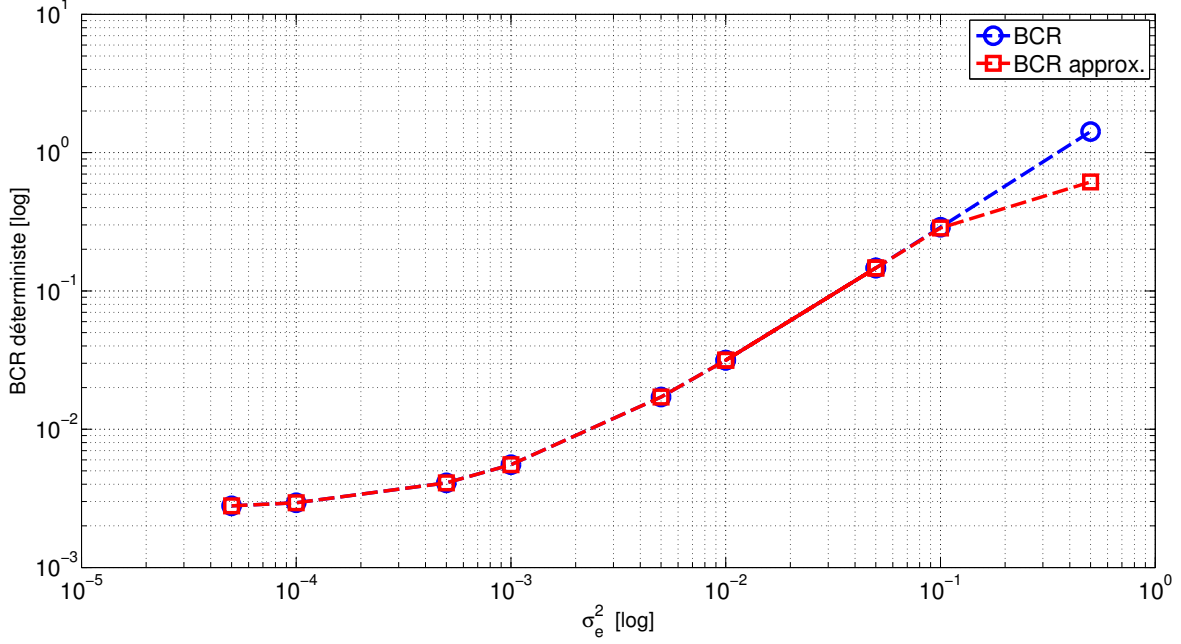


FIGURE 3.8 – BCR déterministe en fonction de la variance de l’erreur, pour  $N = 1000$ ,  $L = 50$  et  $\sigma^2 = 0.001$

Dans ces conditions la BCRM est :

$$\begin{aligned}
 \text{BCRM}_{\text{EG}} &= \mathbb{E}_{\mathbf{s}} \text{Tr} [\mathbf{F}_{\text{EG}}^{-1}] = \text{Tr} [\mathbb{E}_{\mathbf{s}} (\mathbf{F}_{\text{EG}}^{-1})] \\
 &\approx \text{Tr} [\mathbb{E}_{\mathbf{s}} \{ (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) (\mathbf{H}^T \mathbf{H})^{-1} \} - \mathbb{E}_{\mathbf{s}} \{ (2N\sigma_e^4 (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{s} \mathbf{s}^T (\mathbf{H}^T \mathbf{H})^{-1}) \}] \\
 &= (\sigma_e^2 \sigma_s^2 + \sigma^2) \text{Tr} [(\mathbf{H}^T \mathbf{H})^{-1}] - 2N\sigma_e^4 \text{Tr} [(\mathbf{H}^T \mathbf{H})^{-1} \mathbb{E}_{\mathbf{s}} \{ (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) \mathbf{s} \mathbf{s}^T \} (\mathbf{H}^T \mathbf{H})^{-1}].
 \end{aligned}$$

On considère maintenant que le support de  $\mathbf{s}$  est connu, et ne change pas sur les différents tirages aléatoires. La deuxième espérance peut alors être développée :

$$\begin{aligned}
 \mathbb{E}_{\mathbf{s}} \{ (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) \mathbf{s} \mathbf{s}^T \} &= \sigma_e^2 \mathbb{E}_{\mathbf{s}} \{ \|\mathbf{s}\|^2 \mathbf{s} \mathbf{s}^T \} + \sigma^2 \mathbb{E}_{\mathbf{s}} \{ \mathbf{s} \mathbf{s}^T \} = \sigma_e^2 \mathbb{E}_{\mathbf{s}} \left\{ \sum_{l=1}^L s_l^2 \mathbf{s} \mathbf{s}^T \right\} + \sigma^2 \sigma_s^2 \mathbf{S} \\
 &= \left( \frac{\sigma_e^2}{L^2} (m_s^4 + (L-1)\sigma_s^4) + \frac{\sigma_s^2}{L} \right) \mathbf{S},
 \end{aligned} \tag{3.68}$$

où  $\mathbf{S}$  est la matrice diagonale représentant le support de  $\mathbf{s}$  comme précédemment. En introduisant (3.68) dans le calcul de la BCRB on a :

$$\text{BCRM}_{\text{EG}} \approx (\sigma_e^2 \sigma_s^2 + \sigma^2) \text{Tr} [(\mathbf{H}^T \mathbf{H})^{-1}] - 2N\sigma_e^4 \left( \frac{\sigma_e^2}{L^2} (m_s^4 + (L-1)\sigma_s^4) + \frac{\sigma_s^2}{L} \right) \text{Tr} [(\mathbf{H}^T \mathbf{H})^{-2} \mathbf{S}]. \tag{3.69}$$

**Résultat 15.** Pour  $\sigma_e^4$  petit, et  $\mathbf{s}$  un vecteur parcimonieux dont les amplitudes non nulles sont des variables aléatoires de moyenne nulle, de variance  $\sigma_s^2/L$  de moment d’ordre 4  $m_s^4/L^2$  la BCRM est :

$$\text{BCRM}_{\text{EG}} \approx (\sigma_e^2 \sigma_s^2 + \sigma^2) \text{Tr} [(\mathbf{H}^T \mathbf{H})^{-1}] - \left( 2 \frac{N}{L^2} \sigma_e^6 (m_s^4 + (L-1)\sigma_s^4) + 2 \frac{N}{L} \sigma_e^4 \sigma_s^2 \right) \text{Tr} [(\mathbf{H}_S^T \mathbf{H}_S)^{-2}], \tag{3.70}$$

tandis que la BCRM si l'on s'intéresse seulement aux amplitudes non nulle est :

$$\text{BCRM}_{\text{EG}} \approx (\sigma_e^2 \sigma_s^2 + \sigma^2) \text{Tr}[(\mathbf{H}_S^T \mathbf{H}_S)^{-1}] - \left( 2 \frac{N}{L^2} \sigma_e^6 (m_s^4 + (L-1)\sigma_s^4) + 2 \frac{N}{L} \sigma_e^4 \sigma_s^2 \right) \text{Tr}[(\mathbf{H}_S^T \mathbf{H}_S)^{-2}]. \quad (3.71)$$

On se place maintenant dans le contexte asymptotique où  $N, L \rightarrow \infty$  avec  $\frac{N}{L} \rightarrow \rho_{\text{dic.}}$ . Les dimensions de la matrice aléatoire  $\mathbf{H}_S$  tendent alors vers l'infini tandis que le rapport entre ses dimensions reste fini. Nous faisons également l'hypothèse que  $\mathbf{H}_S$  est une matrice dont les éléments sont tirés aléatoirement d'une distribution gaussienne de moyenne nulle et d'une variance  $1/L$ . On définit la BCRM asymptotique par :

$$\eta^\infty = \lim_{N, L \rightarrow \infty} \frac{1}{L} \text{BCRM}. \quad (3.72)$$

Sous ces conditions on a :

$$\begin{aligned} \eta_{\text{EG}}^\infty &\approx \lim_{L \rightarrow \infty} (\sigma_e^2 \sigma_s^2 + \sigma^2) \frac{\text{Tr}[(\mathbf{H}_S^T \mathbf{H}_S)^{-1}]}{L} \\ &\quad - \left( 2 \frac{\rho_{\text{dic.}}}{L} \sigma_e^6 (m_s^4 - \sigma_s^4) + 2 \sigma_e^6 \rho_{\text{dic.}} \sigma_s^4 + 2 \rho_{\text{dic.}} \sigma_e^4 \sigma_s^2 \right) \frac{\text{Tr}[(\mathbf{H}_S^T \mathbf{H}_S)^{-2}]}{L} \end{aligned} \quad (3.73)$$

$$\begin{aligned} &= \lim_{L \rightarrow \infty} (\sigma_e^2 \sigma_s^2 + \sigma^2) \frac{\text{Tr}[(\mathbf{H}_S^T \mathbf{H}_S)^{-1}]}{L} \\ &\quad - (2 \sigma_e^6 \rho_{\text{dic.}} \sigma_s^4 + 2 \rho_{\text{dic.}} \sigma_e^4 \sigma_s^2) \frac{\text{Tr}[(\mathbf{H}_S^T \mathbf{H}_S)^{-2}]}{L}, \end{aligned} \quad (3.74)$$

étant donné que  $L$  tend vers l'infini le terme en  $1/L$  s'annule. En introduisant dans cette expression les résultats (3.52) et (3.53) de la théorie des matrices aléatoires on trouve le résultat qui suit.

**Résultat 16.** La BCRM asymptotique pour le problème de l'erreur de grille, pour une variance d'erreur faible et un dictionnaire aléatoire où  $L$  et  $N$  tendent vers l'infini avec un rapport fini est :

$$\eta_{\text{EG}}^\infty \approx \frac{(\sigma_e^2 \sigma_s^2 + \sigma^2)}{\rho_{\text{dic.}} - 1} - \frac{2 \sigma_e^4 \sigma_s^2 (\sigma_e^2 \sigma_s^2 + 1) \rho_{\text{dic.}}^2}{(\rho_{\text{dic.}} - 1)^3}. \quad (3.75)$$

Dans la figure 3.9 sont tracées la borne BCRM approximée (3.71) en bleu et la borne  $\eta_{\text{EG}}^\infty$  correspondante (3.75) en rouge. Pour que les deux bornes soient comparables, la courbe bleue trace la fonction  $\frac{\text{BCRM}_{\text{EG}}}{L}$ . Les bornes sont tracées en fonction de  $\frac{L}{N} = \frac{1}{\rho_{\text{dic.}}}$ , pour ce faire on fixe  $N = 1000$  et on fait varier  $L$  entre 1 et  $N$ . La variance du bruit est choisie assez faible ( $\sigma_e^2 = 10^{-3}$ ) pour que l'approximation de la BCR pour des petites erreurs soit correcte et égale à la BCRM. Pour une erreur trop importante on ne peut plus faire l'approximation et le calcul de la BCRM donné en (3.71) ne correspond plus à une borne. La variance du bruit a également été choisie petite de telle sorte qu'on conserve une variance d'erreur supérieure à la variance du bruit, pour se concentrer sur les effets de cette erreur, même si elle est faible. Pour  $N = 1000$ , le dictionnaire est assez large pour que l'hypothèse asymptotique soit vérifiée, les courbes bleu et rouge sont confondues ce qui valide l'expression (3.75). Nous verrons dans les autres figures que cette expression est valide même pour des dictionnaires relativement petits. Quand le nombre de paramètres  $L$  augmente, la difficulté de l'estimation augmente, la borne augmente donc assez naturellement quand  $L$  augmente.

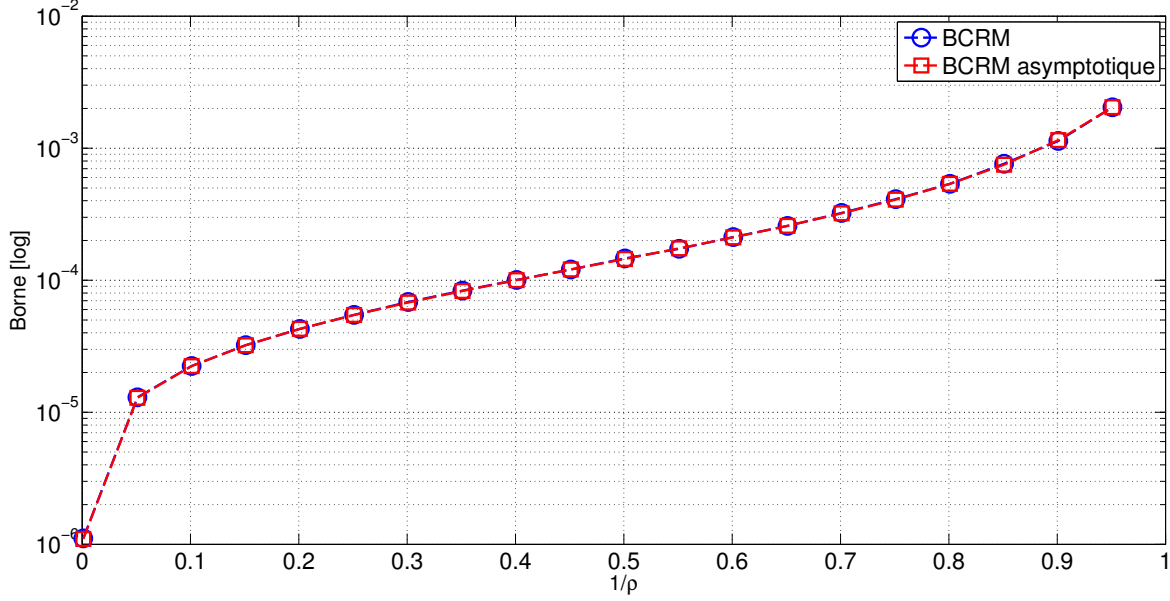


FIGURE 3.9 – BCRM exacte et approximée en fonction de  $1/\rho_{\text{dic.}}$ , pour  $N = 1000$ ,  $\sigma_e^2 = 10^{-3}$  et  $\sigma^2 = 10^{-4}$

Nous allons à présent étudier le comportement des différentes bornes obtenues pour le problème de l'erreur de grille, que nous comparons aux bornes sans erreur. Toutes les bornes sont normalisées par  $L$  pour rendre la comparaison pertinente. Sur les figures 3.10 à 3.13 sont tracées :

- en magenta, la borne BCRM sans ajout d'erreur :

$$\frac{\text{BCRB}_{\sigma_e^2=0}}{L} = \frac{\sigma^2}{L} \text{Tr} \left[ (\mathbf{H}_S^T \mathbf{H}_S)^{-1} \right]. \quad (3.76)$$

La borne de Van Trees sans erreur est égale à la BCRM sans erreur, à laquelle on ajoute le terme lié à la distribution des paramètres. Dans le cas à fort RSB où nous nous plaçons, ces deux bornes sont égales puisque l'on peut négliger le second terme de la BCRB de Van Trees.

- en noir, la borne sans bruit asymptotique obtenue pour un dictionnaire  $\mathbf{H}_S$  aléatoire et gaussien :

$$\eta_{\sigma_e^2=0}^\infty = \frac{\sigma^2}{\rho_{\text{dic.}} - 1}. \quad (3.77)$$

- en bleu, la BCRM normalisée pour une petite erreur générale à partir de l'expression (3.71) :  $\text{BCRM}/L$ .
- en rouge, la BCRM asymptotique donnée en (3.75).
- en vert la BCRB de Van Trees normalisée associée au problème de l'erreur de grille donnée en (3.43) :  $\text{BCRB}_{\text{EG}_S}/L$ .

Les trois premières figures tracent les bornes en fonction de la variance de l'erreur  $\sigma_e^2$  pour un bruit fixé à  $\sigma^2 = 0.001$ . On fait varier la variance de l'erreur sur la plage pour laquelle

l'expression de la BCR approximée est égale à la BCR numérique comme nous l'avons observé sur la figure 3.7. La figure 3.10 montre le cas idéal pour lequel le dictionnaire est assez grand ( $1000 \times 50$ ), les courbes asymptotiques y sont toutes deux confondues avec la borne non asymptotique correspondante. La première remarque que l'on peut faire est que la borne BCRM est bien tout le temps supérieure à la BCRB de Van Trees. Quand la variance de l'erreur décroît, la borne BCRM rejoint la borne en l'absence d'erreur, ce qui se vérifie par leur expression. De son côté la borne de Van Trees continue de décroître à mesure que la variance de l'erreur décroît jusqu'à atteindre une valeur infiniment faible quand  $\sigma_e^2 \rightarrow 0$  alors que la borne sans erreur atteint un palier. Ce comportement montre que la borne de Van Trees est certes une borne minimale, mais elle se montre trop optimiste dans certaines situations, notamment pour un bruit et une erreur faible où elle peut atteindre une erreur moyenne nulle alors que le calcul de la BCRM montre qu'en réalité il n'est pas possible d'atteindre de telles performances par un estimateur, puisque la BCRM est également une borne inférieure de la BEQM. Quand la variance de l'erreur augmente, de manière intuitive les deux bornes Bayésiennes montrent que les performances se dégradent fortement en comparaison du cas idéal sans erreur. Dans la figure 3.11 nous avons tracé les mêmes courbes, pour une taille de dictionnaire très petite :  $50 \times 3$ . On peut observer sur cette courbe que l'on a atteint un cas limite de la théorie des matrices aléatoires, le dictionnaire est trop petit pour que la formulation très simple de la BCRM asymptotique soit vérifiée. Cependant on peut voir sur la figure 3.12 que même pour un dictionnaire relativement petit de taille  $50 \times 10$ , l'approximation apportée par la théorie des matrices aléatoires peut être utilisée, pourtant ces valeurs sont loin d'être asymptotiques. Ce résultat est intéressant, on peut observer que dans des plages de valeurs utilisées en pratique, même loin d'être asymptotique, le cas asymptotique reflète très bien le comportement des bornes. Cela permet d'obtenir des expressions vraiment simples à calculer, ne dépendant pas d'une réalisation donnée de la matrice  $\mathbf{H}$ . On peut en conclure que les performances d'estimation pour une matrice aléatoire gaussienne ne dépendent pas d'une réalisation précise du dictionnaire, mais uniquement de ses dimensions.

Les mêmes courbes sont tracées également sur la figure 3.13 avec une perspective différente : l'axe des abscisses est cette fois-ci le rapport entre l'interférence apportée par l'erreur de grille et le bruit. On définit le rapport Interférence sur Bruit par  $\text{INR} = \sigma_e^2 / \sigma^2$ . On a choisi sur cette courbe les mêmes paramètres que la figure 3.10, et l'on a fixé  $\sigma_e^2 = 10^{-3}$  et fait varier le bruit  $\sigma^2$ . Rappelons que la borne de Van Trees donnée en (3.43) est obtenue pour un bruit faible où  $\sigma_e^2$  domine sur  $\sigma^2$ , son expression pour un INR faible n'est donc pas représentative de la borne Bayésienne exacte et numérique. Cette borne approximative ne dépend pas de  $\sigma^2$  et est donc constante sur la figure 3.13. Elle représente le palier qu'atteignent les performances en présence d'une erreur de grille, même quand le bruit diminue. La BCRM quant à elle suit la borne sans erreur quand le bruit est important, en effet pour une variance d'erreur donnée, quand le bruit devient prédominant les bornes avec ou sans erreur se confondent. Au fur et à mesure que le bruit diminue, la BCRM rejoint le palier donné par la borne de Van Trees. Cette borne suit le comportement que l'on peut attendre des estimateurs.

### 3.3.3 Problème de dictionnaire bruité

Nous nous intéressons à présent au calcul de la BCR déterministe en présence d'une erreur de dictionnaire bruité. Dans ce contexte les observations sont  $\mathbf{y}$  et  $\tilde{\mathbf{H}}$  et le paramètre

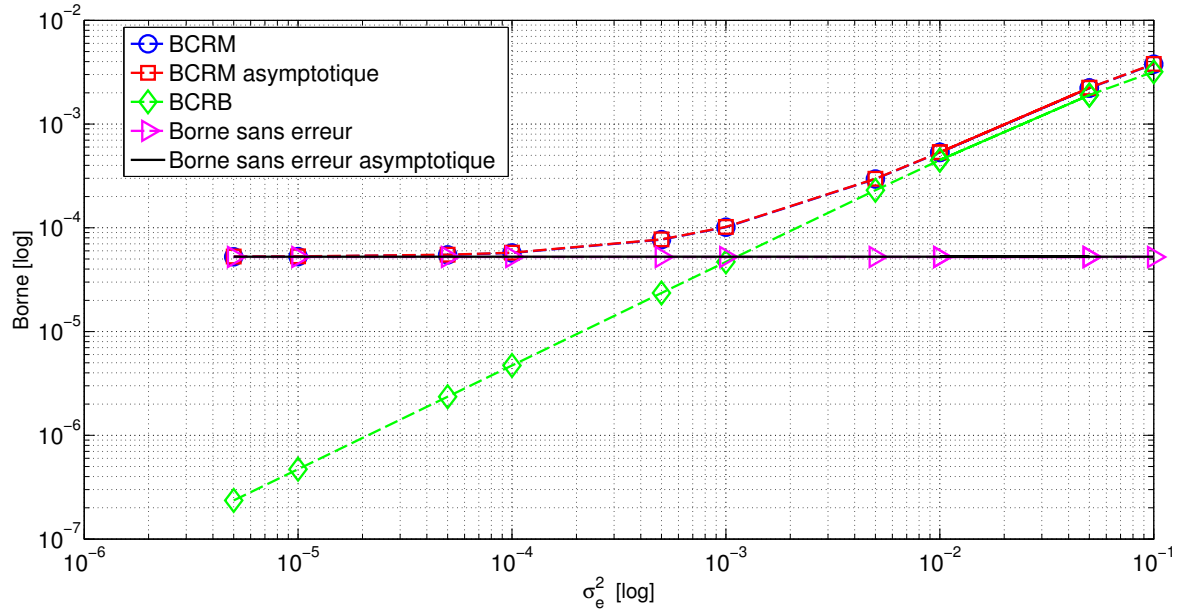


FIGURE 3.10 – Bornes pour le problème de l’erreur de grille en fonction de la variance de l’erreur, pour  $N = 1000$ ,  $L = 50$

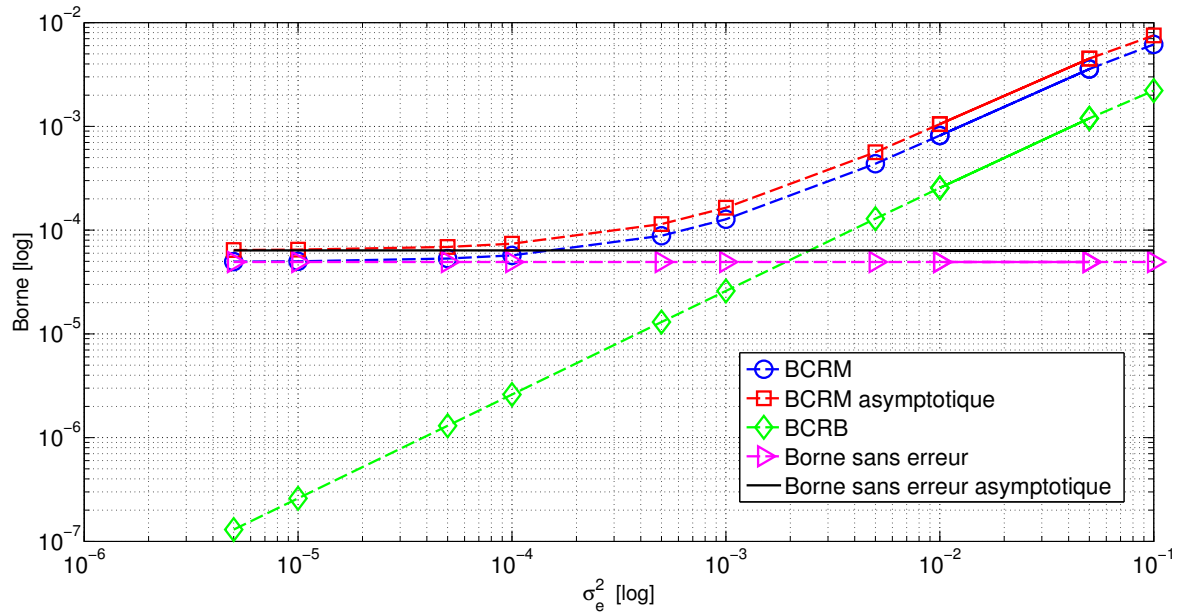


FIGURE 3.11 – Bornes pour le problème de l’erreur de grille en fonction de la variance de l’erreur, pour  $N = 50$ ,  $L = 3$

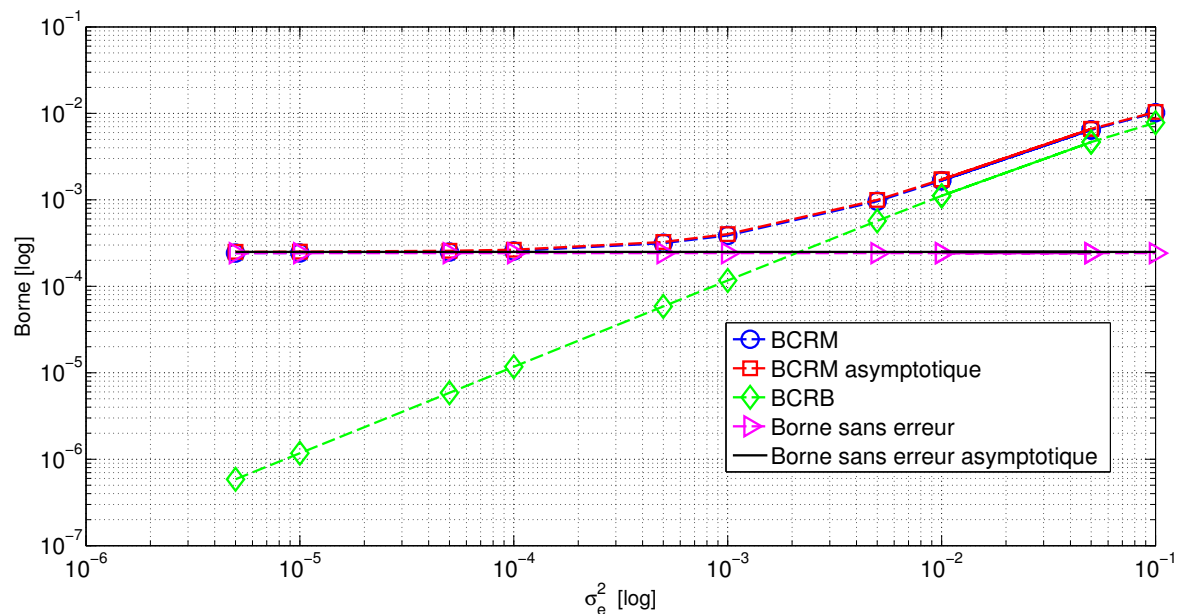


FIGURE 3.12 – Borne pour le problème de l’erreur de grille en fonction de la variance de l’erreur, pour  $N = 50$ ,  $L = 10$

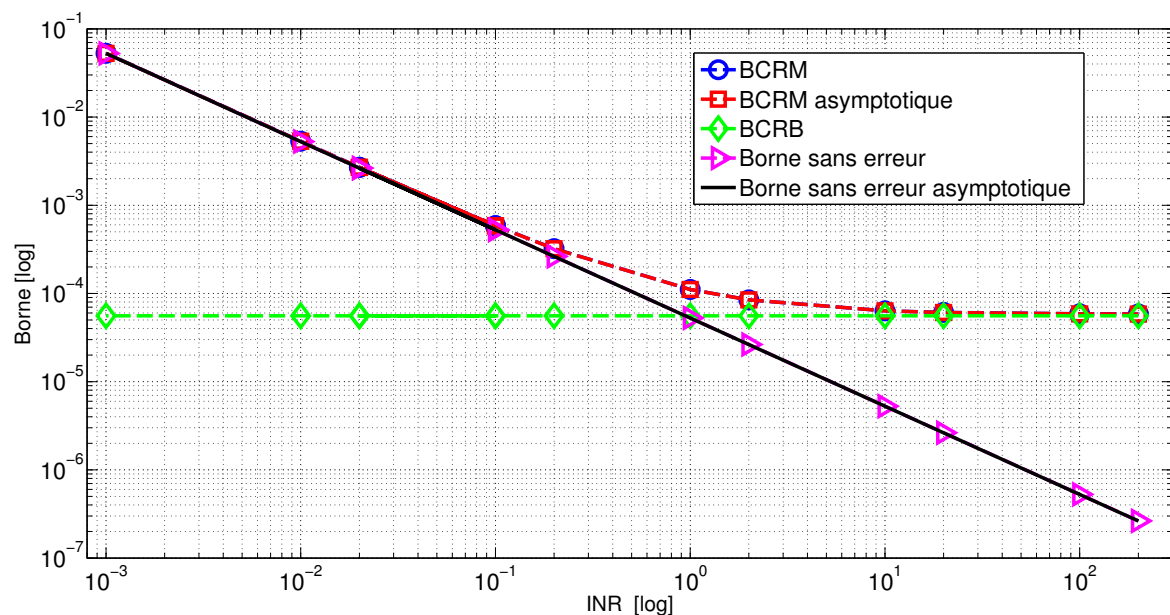


FIGURE 3.13 – Borne pour le problème de l’erreur de grille en fonction du rapport  $\sigma_e^2/\sigma^2$ , pour  $N = 1000$ ,  $L = 50$  et  $\sigma_e^2 = 10^{-3}$

à estimer est toujours  $\mathbf{s}$ . Contrairement au problème EIV tel que le considèrent WIESEL, ELDAR et YEREDOR [131], nous ne cherchons pas à estimer  $\mathbf{H}$  dans ce chapitre mais nous plaçons dans un contexte où l'algorithme d'estimation utilise un dictionnaire erroné de manière habituelle, sans savoir que ce dictionnaire est bruité. La matrice d'information de Fisher à inverser pour obtenir la BCR est dans ce contexte :

$$\mathbf{F}_{\text{DB}} = \mathbb{E}_{\mathbf{y}, \tilde{\mathbf{H}}|\mathbf{s}} \left\{ -\frac{\partial^2 \log p(\mathbf{y}, \tilde{\mathbf{H}}|\mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right\} \quad (3.78)$$

$$= \mathbb{E}_{\mathbf{y}, \tilde{\mathbf{H}}|\mathbf{s}} \left\{ -\frac{\partial^2 \log p(\mathbf{y}|\tilde{\mathbf{H}}, \mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right\} \quad (3.79)$$

$$= \mathbb{E}_{\tilde{\mathbf{H}}} \left\{ \mathbb{E}_{\mathbf{y}|\tilde{\mathbf{H}}, \mathbf{s}} \left\{ -\frac{\partial^2 \log p(\mathbf{y}|\tilde{\mathbf{H}}, \mathbf{s})}{\partial \mathbf{s}^T \partial \mathbf{s}} \right\} \right\}. \quad (3.80)$$

On obtient la deuxième égalité en utilisant le fait que  $p(\tilde{\mathbf{H}})$  ne dépend pas de  $\mathbf{s}$  et la troisième vient de la formule de Bayes des espérances.

En se basant sur le modèle gaussien  $\mathbf{y}|\tilde{\mathbf{H}}, \mathbf{s} \sim \mathcal{N}(\tilde{\mathbf{H}}\mathbf{s}, (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)\mathbf{I})$ , on peut déduire la MIF conditionnelle en utilisant la formule de Slepian-Bang (1.60) :

$$\mathbf{F}_{\text{DB}} = \mathbb{E}_{\tilde{\mathbf{H}}} \left\{ \frac{\tilde{\mathbf{H}}^T \tilde{\mathbf{H}}}{\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2} + \frac{2\sigma_e^4 \mathbf{N} \mathbf{s} \mathbf{s}^T}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2} \right\} \quad (3.81)$$

$$= \frac{\mathbb{E}_{\tilde{\mathbf{H}}} \{\tilde{\mathbf{H}}^T \tilde{\mathbf{H}}\}}{\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2} + \frac{2\sigma_e^4 \mathbf{N} \mathbf{s} \mathbf{s}^T}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2} \quad (3.82)$$

$$= \frac{\mathbf{H}^T \mathbf{H} + \mathbf{N} \sigma_e^2 \mathbf{I}}{\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2} + \frac{2\sigma_e^4 \mathbf{N} \mathbf{s} \mathbf{s}^T}{(\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2)^2}, \quad (3.83)$$

où l'on a introduit (3.15) pour obtenir la dernière égalité.

On peut à présent procéder à l'inversion de la MIF à l'aide de la formule d'inversion par bloc (2.14) :

$$\mathbf{F}_{\text{DB}}^{-1} = (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) (\mathbf{H}^T \mathbf{H} + \mathbf{N} \sigma_e^2 \mathbf{I})^{-1} - \Delta, \quad (3.84)$$

avec

$$\Delta = \frac{2\sigma_e^4 \mathbf{N} (\sigma_e^2 \|\mathbf{s}\|^2 + \sigma^2) (\mathbf{H}^T \mathbf{H} + \mathbf{N} \sigma_e^2 \mathbf{I})^{-1} \mathbf{s} \mathbf{s}^T (\mathbf{H}^T \mathbf{H} + \mathbf{N} \sigma_e^2 \mathbf{I})^{-1}}{1 + 2\sigma_e^4 \mathbf{N} \mathbf{s}^T (\mathbf{H}^T \mathbf{H} + \mathbf{N} \sigma_e^2 \mathbf{I})^{-1} \mathbf{s}}. \quad (3.85)$$

La borne de Cramer Rao correspondante est égale à la trace de  $\mathbf{F}_{\text{DB}}^{-1}$ . La présence de l'erreur dans le dictionnaire reçu complique encore l'expression, ce qui ne permet pas d'en obtenir une expression simple permettant d'obtenir une expression analytique de la BCRM. Il n'est pas possible d'obtenir une expression plus simple pour cette borne, même pour des  $\sigma_e^2$  petit.

Dans les figures 3.14 et 3.15 les bornes de Cramer Rao déterministes exactes pour le problème d'erreur de grille et le problème de dictionnaire bruité sont tracées en bleu et en rouge respectivement, en fonction de la variance de l'erreur  $\sigma_e^2$ . Sur les deux figures on peut confirmer le résultat obtenu par la BCRB de Van Trees : le problème du bruit sur le dictionnaire permet de bien meilleures performances que celui de l'erreur de grille. Quand l'erreur est nulle, les deux BCR sont égales. Quand l'erreur de dictionnaire augmente, la  $\text{BCR}_{\text{DB}}$  atteint un palier haut, tandis que la  $\text{BCR}_{\text{EG}}$  continue de se dégrader. De plus une propriété très intéressante du problème de dictionnaire bruité est que quand le bruit est important, et que donc l'erreur sur le dictionnaire est plus faible que le bruit,

cette erreur permet d'améliorer les performances de l'estimation. En effet pour  $\sigma^2 = 0.1$ , la  $\text{BCR}_{\text{DB}}$  diminue avec l'augmentation de l'incertitude sur le dictionnaire. Plus  $\sigma^2$  augmente plus cette amélioration est importante. Ce comportement peut être expliqué par le fait qu'en présence d'un bruit important, l'ajout d'un terme aléatoire important sur le dictionnaire permet de mieux refléter le comportement des échantillons ; ceux-ci devenant eux même aléatoires à mesure que le rapport signal sur bruit diminue.

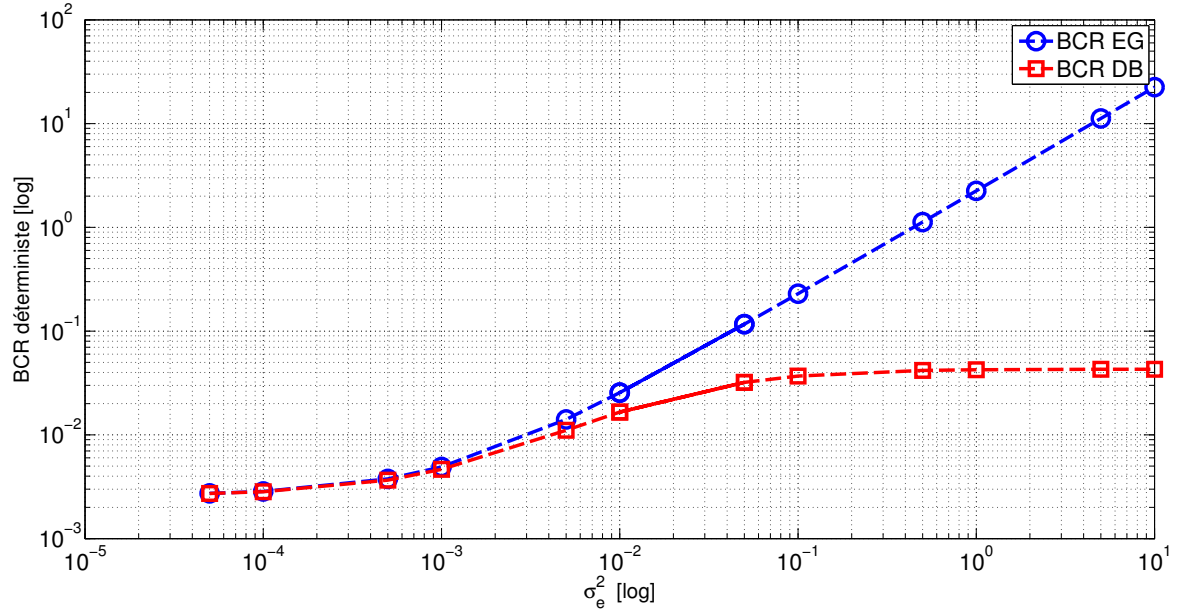
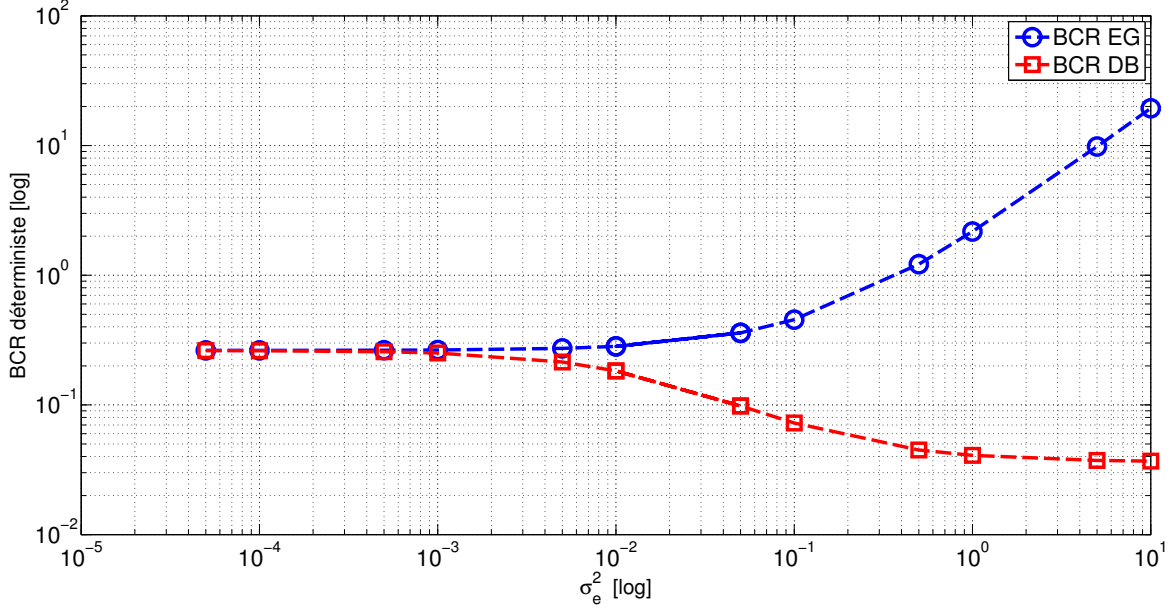


FIGURE 3.14 – Borne déterministe pour les problèmes EG et DB, pour  $N = 1000$ ,  $L = 50$  et  $\sigma^2 = 0.001$ .


 FIGURE 3.15 – Borne déterministe pour les problèmes EG et DB, pour  $N = 1000$ ,  $L = 50$  et  $\sigma^2 = 0.1$ .

### 3.4 Conséquence de la discrétisation sur l'estimation des fréquences

Nous nous sommes concentrés pour le moment sur les performances d'estimation des amplitudes en présence d'erreur, pour un support connu. Dans cette section, nous nous intéressons spécifiquement au problème de l'estimation par dictionnaire présenté dans l'équation (1.46), un problème pour lequel l'indice des valeurs non nulles du support est le paramètre principal à estimer, les amplitudes n'étant que secondaires. Nous nous intéressons donc ici à l'estimation du paramètre de support, plutôt qu'à l'amplitude. Nous avons vu dans la section 1.3.3 que pour ce type d'estimation, on fait l'hypothèse intrinsèque que les paramètres à estimer appartiennent à une grille donnée, et connue. Cette hypothèse ne peut pas être vérifiée, quelle que soit la grille choisie étant donné que nous cherchons à estimer des paramètres continus, qui ne peuvent donc pas être tous représentés dans une grille ayant un nombre fini de points. Ce problème est un cas particulier du problème EG, spécifique à l'estimation par dictionnaire. Pour étudier les conséquences de l'erreur de discrétisation sur les performances d'estimation par dictionnaire, nous établissons un modèle Bayésien du problème, puis à partir de ce modèle nous dérivons l'erreur quadratique moyenne Bayésienne, supposant toujours que l'on se place dans un contexte asymptotique où le support est correctement estimé, mis à part l'erreur de discrétisation. Nous donnons ensuite les expressions analytiques de la BEQM pour des distributions gaussiennes et uniformes avant d'illustrer ces résultats dans le contexte de l'estimation de fréquence. Les résultats présentés dans cette section ont été présentés dans l'article BERNHARDT et al. [23].

#### 3.4.1 Présentation du modèle bayésien d'erreur sur le support

Pour simplifier les notations en comparaison avec le modèle donné en (1.3.3) nous allons considérer le cas à une dimension où il n'y a qu'un seul type de paramètre à estimer

en plus de l'amplitude. Le modèle d'estimation par dictionnaire, cherchant à retrouver  $L$  paramètres  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_L]^T$  est alors :

$$\mathbf{y} \approx \mathbf{H}(\bar{\boldsymbol{\theta}})\mathbf{s} + \mathbf{w} \quad (3.86)$$

Notons  $\Omega$  l'espace des paramètres, c'est à dire l'espace des valeurs que peuvent prendre les  $\theta$ . L'espace des paramètres  $\Omega$  est discrétisé dans la grille  $\bar{\boldsymbol{\theta}} = [\bar{\theta}_1, \dots, \bar{\theta}_K]^T$ . Nous allons plus spécifiquement étudier le cas de la grille uniforme, avec un espacement de  $r = \bar{\theta}_{k+1} - \bar{\theta}_k$ . On peut diviser l'espace des paramètres en intervalles de décision centrés sur chaque élément du dictionnaire. Nous définissons formellement ces intervalles par :

$$\begin{aligned} \Omega_1 &= \left[ \bar{\theta}_1, \bar{\theta}_1 + \frac{r}{2} \right] \\ \Omega_k &= \left[ \bar{\theta}_k - \frac{r}{2}, \bar{\theta}_k + \frac{r}{2} \right] \\ \Omega_K &= \left[ \bar{\theta}_K - \frac{r}{2}, \bar{\theta}_K \right]. \end{aligned} \quad (3.87)$$

La grille couvrant tout l'espace des paramètres, nous avons :

$$\Omega = [\bar{\theta}_1, \bar{\theta}_K] = \Omega_1 \cup_{k=2}^{K-1} \Omega_k \cup \Omega_K. \quad (3.88)$$

Dans un contexte Bayésien, les paramètres à estimer sont aléatoires, tirés chacun d'une distribution indépendante  $p(\theta_l; l)$ . Nous utilisons une telle notation pour exprimer le fait que toutes les variables suivent la même famille de distribution  $p(\theta)$  (du type uniforme, gaussienne...) mais paramétrée différemment : elles peuvent par exemple être des gaussiennes avec des variances  $\sigma_l^2$  différentes. La figure 3.16 permet de visualiser les différentes distributions que nous allons définir par la suite,  $p(\theta_l; l)$  est représentée par la courbe pointillée noire, il s'agit d'une distribution allant de  $-\infty$  à  $+\infty$  et représentant donc les paramètres non réduit à l'espace  $\Omega$ . Le vecteur contenant les  $L$  paramètres inconnus  $\boldsymbol{\theta}$  est un vecteur de variables aléatoires qui suit une distribution connue  $p(\boldsymbol{\theta}) = \prod_{l=1}^L p(\theta_l; l)$  et appartient à l'ensemble  $\Omega^L = \Omega \times \dots \times \Omega$ . Nous nous intéressons à la distribution conditionnelle  $p(\boldsymbol{\theta} | \boldsymbol{\theta} \in \Omega^L)$  qui est la distribution des paramètres sachant qu'ils appartiennent à l'espace de recherche qui nous intéresse, celle-ci peut être écrite comme une distribution tronquée :

$$p(\boldsymbol{\theta} | \boldsymbol{\theta} \in \Omega^L) = \frac{p(\boldsymbol{\theta}) \cdot 1_{\Omega^L}(\boldsymbol{\theta})}{\Pr(\boldsymbol{\theta} \in \Omega^L)}, \quad (3.89)$$

où  $1_{\Omega^L}(\boldsymbol{\theta})$  est la fonction indicatrice de l'ensemble  $\Omega$  qui est égale à 1 si  $\boldsymbol{\theta} \in \Omega^L$  et 0 autrement, et

$$\Pr(\boldsymbol{\theta} \in \Omega^L) = \int_{\Omega^L} p(\boldsymbol{\theta}) d\boldsymbol{\theta} = \prod_{l=1}^L P_{\Omega}^{(l)}. \quad (3.90)$$

étant donné que les paramètres  $\theta$  sont supposés indépendants. On a défini :

$$P_{\Omega}^{(l)} \triangleq \int_{\Omega} p(\theta_l; l) d\theta_l. \quad (3.91)$$

En combinant ces expressions et en utilisant le fait que la fonction caractéristique sur l'ensemble  $\Omega^L$  est le produit des fonctions caractéristiques  $1_{\Omega}(\theta_l)$  on a :

$$p(\boldsymbol{\theta} | \boldsymbol{\theta} \in \Omega^L) = \prod_{l=1}^L p(\theta_l | \theta_l \in \Omega), \quad (3.92)$$

où la distribution de chaque paramètre sachant dans quel ensemble il se trouve est la version tronquée sur l'ensemble des paramètres de  $p(\theta_l; l)$  :

$$p(\theta_l | \theta_l \in \Omega) = \frac{p(\theta_l; l) \cdot 1_{\Omega}(\theta_l)}{P_{\Omega}^{(l)}}. \quad (3.93)$$

Cette distribution tronquée et son facteur de normalisation  $P_{\Omega}^{(l)}$  sont tracés sur la figure 3.16 avec une ligne pointillée bleue et une aire hachurée bleue respectivement. On se place dans la situation où  $\theta_l \in \Omega$ , les paramètres suivent donc la distribution (3.93), qui est la distribution  $p(\theta_l; l)$  réduite sur l'espace  $\Omega$ .

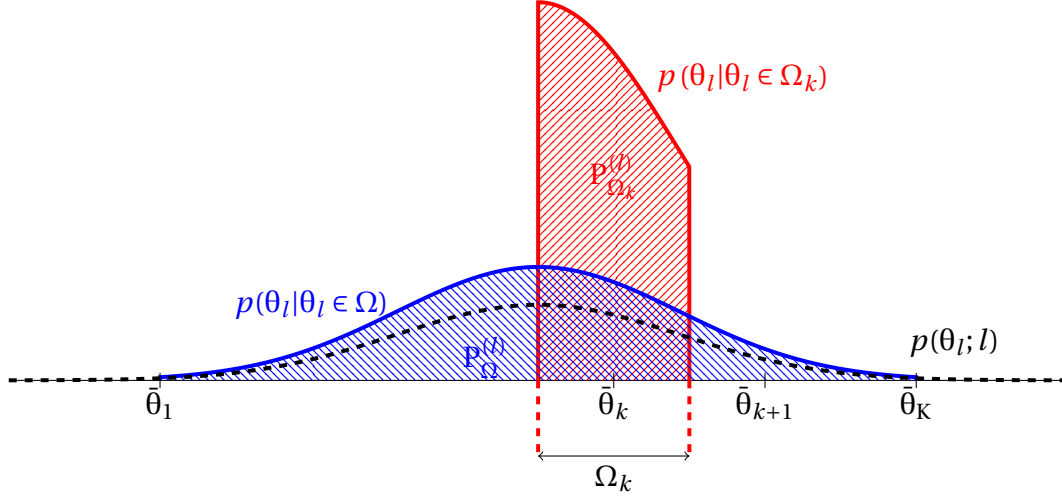


FIGURE 3.16 – Distributions tronquées modélisant une estimation par dictionnaire

Nous pouvons à présent représenter la distribution des paramètres comme la somme des distributions sur chaque intervalle de décision défini en (3.87) grâce à la loi des probabilités totales :

$$p(\theta_l | \theta_l \in \Omega) = \sum_{k=1}^K P_{\Omega_k}^{(l)} p(\theta_l | \theta_l \in \Omega_k), \quad (3.94)$$

où  $P_{\Omega_k}^{(l)} = \int_{\Omega_k} p(\theta_l | \theta_l \in \Omega) d\theta_l$  est la probabilité que le paramètre  $\theta_l$  soit dans l'intervalle de décision  $\Omega_k$ . Cette distribution et le facteur de normalisation  $P_{\Omega_k}^{(l)}$  associé sont tracés en rouge sur la figure 3.16.

Notre objectif est d'étudier les conséquences de l'erreur de dictionnaire, non du bruit, c'est pourquoi jusqu'à présent nous avons fait l'hypothèse d'un support connu pour le calcul des bornes sur le paramètre d'amplitude. De façon similaire, nous souhaitons isoler dans cette section uniquement les problèmes liés à la grille, et donc considérer que tout le reste se déroule de façon optimale : les conditions RIP ou de cohérence sont respectées et on considère que les algorithmes sont efficaces et donc estiment au plus juste les paramètres en l'absence de bruit. Plus formellement, nous nous plaçons dans le régime que nous appelons "à fort RSB", qui est le régime dans lequel le niveau de bruit est assez faible pour que :

$$\forall l, \exists k_l, \forall \theta_l \in \Omega_{k_l}, \quad \hat{\theta}_l(\mathbf{y}) = \bar{\theta}_{k_l}, \quad (3.95)$$

ce qui explique le fait que l'estimateur  $\hat{\theta}_l(\mathbf{y})$  choisit comme estimée la valeur du dictionnaire la plus proche de la vraie valeur  $\theta_l$ . Nous indexons l'intervalle de décision dans lequel se trouve  $\theta_l$  par  $k_l$ , la valeur la plus proche étant alors  $\bar{\theta}_{k_l}$ . Une conséquence directe

de cette propriété est que le résultat de l'estimation devient déterministe, malgré le fait qu'en entrée on considère des paramètres aléatoires, parce que le résultat est entièrement déterminé par l'intervalle de décision.

### 3.4.2 Performances à fort RSB des méthodes d'estimation par dictionnaire

Le biais sur le paramètre vectoriel  $\boldsymbol{\theta}$  est  $\mathbf{b}(\boldsymbol{\theta}) = [b(\theta_1), \dots, b(\theta_L)]^T$  où le biais sur chaque paramètre  $\theta_l$  est défini par

$$b(\theta_l) = \mathbb{E}_{\mathbf{y}}(\hat{\theta}_l(\mathbf{y})) - \theta_l = \bar{\theta}_{k_l} - \theta_l, \quad \forall \theta_l \in \Omega_{k_l}. \quad (3.96)$$

Cette définition du biais découle directement de l'hypothèse que nous sommes dans un régime fort RSB (3.95). La BEQM est conditionnée à l'ensemble  $\Omega^L$  et est définie par :

$$\text{BEQM}_{\Omega^L} = \int_{\mathbb{R}^L} \underbrace{\mathbb{E}_{\mathbf{y}}(\|\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}\|^2)}_{\text{EQM}} p(\boldsymbol{\theta} | \boldsymbol{\theta} \in \Omega^L) d\boldsymbol{\theta}, \quad (3.97)$$

où  $\mathbb{R}^L = \mathbb{R} \times \dots \times \mathbb{R}$  et  $p(\boldsymbol{\theta} | \boldsymbol{\theta} \in \Omega^L)$  est la distribution définie en (3.89). L'erreur quadratique moyenne d'un paramètre vectoriel peut être décomposé en :

$$\mathbb{E}_{\mathbf{y}}(\|\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}\|^2) = \mathbb{E}_{\mathbf{y}}\left(\sum_{l=1}^L (\hat{\theta}_l(\mathbf{y}) - \theta_l)^2\right) = \sum_{l=1}^L \mathbb{E}_{\mathbf{y}}\left((\hat{\theta}_l(\mathbf{y}) - \theta_l)^2\right). \quad (3.98)$$

On utilise cette propriété et l'expression (3.92) pour obtenir que :

$$\text{BEQM}_{\Omega^L} = \sum_{l=1}^L \int_{\mathbb{R}^L} \mathbb{E}_{\mathbf{y}}\left((\hat{\theta}_l(\mathbf{y}) - \theta_l)^2\right) p(\boldsymbol{\theta} | \boldsymbol{\theta} \in \Omega^L) d\boldsymbol{\theta} \quad (3.99)$$

$$= \sum_{l=1}^L \int_{\theta_l} \mathbb{E}_{\mathbf{y}}\left((\hat{\theta}_l(\mathbf{y}) - \theta_l)^2\right) p(\theta_l | \theta_l \in \Omega) d\theta_l, \quad (3.100)$$

où l'on note  $\boldsymbol{\theta}^{-l}$  le vecteur de longueur  $L - 1$  correspondant à  $\boldsymbol{\theta}$  sans la  $l^e$  colonne.

On peut exprimer la BEQM de deux manières en utilisant les deux façons d'exprimer  $p(\theta_l | \theta_l \in \Omega)$ . Premièrement, si l'on se place du point de vue des intervalles de décision, on peut exprimer la BEQM à l'aide de l'expression (3.94) :

$$\text{BEQM}_{\Omega^L} = \sum_{l=1}^L \sum_{k=1}^K P_{\Omega_k}^{(l)} \int_{\Omega_k} \mathbb{E}_{\mathbf{y}}\left((\hat{\theta}_l(\mathbf{y}) - \theta_l)^2\right) p(\theta_l | \theta_l \in \Omega_k) d\theta_l \quad (3.101)$$

$$\xrightarrow{\text{fort RSB}} \sum_{l=1}^L \sum_{k=1}^K P_{\Omega_k}^{(l)} \underbrace{\int_{\Omega_k} b(\theta_l)^2 p(\theta_l | \theta_l \in \Omega_k) d\theta_l}_{\mathbb{E}_{\theta_l | \theta_l \in \Omega_k}(b(\theta_l)^2)}. \quad (3.102)$$

On a utilisé le fait que dans le régime à fort RSB  $\hat{\theta}_l$  devient déterministe, et ne dépend plus que de l'intervalle  $\Omega_k$  ce qui permet de retirer l'espérance sur le bruit qui n'affecte pas le résultat de l'estimation. On obtient ainsi le résultat sur la BEQM.

**Résultat 17.** Dans le régime à fort RSB et pour des paramètres indépendants aléatoires de distribution connue  $p(\theta_l; l)$ , la BEQM conditionnée à l'ensemble  $\Omega^L$  est donnée par

$$\text{BEQM}_{\Omega^L, \text{fort}} = \sum_{l=1}^L \sum_{k=1}^K P_{\Omega_k}^{(l)} \mathbb{E}_{\theta_l | \theta_l \in \Omega_k} (b(\theta_l)^2). \quad (3.103)$$

Une expression équivalente est obtenue en utilisant plutôt l'expression (3.93) dans (3.100) et le fait que  $P_{\Omega}^{(l)}$  ne dépend pas non plus de  $\theta_l$  :

$$\begin{aligned} \text{BEQM}_{\Omega^L} &= \sum_{l=1}^L \frac{1}{P_{\Omega}^{(l)}} \int_{\Omega} \mathbb{E}_{\mathbf{y}} ((\hat{\theta}_l(\mathbf{y}) - \theta_l)^2) p(\theta_l; l) d\theta_l \\ &\xrightarrow{\text{fort RSB}} \sum_{l=1}^L \frac{1}{P_{\Omega}^{(l)}} \sum_{k=1}^K \underbrace{\int_{\Omega_k} b(\theta_l)^2 p(\theta_l; l) d\theta_l}_{\mathbb{E}_{\theta_l} (b(\theta_l)^2 1_{\Omega_k}(\theta_l))}. \end{aligned}$$

Pour obtenir la dernière expression, nous avons décomposé l'intégrale sur tout l'espace  $\Omega$  en la somme des intégrales sur chaque espace  $\Omega_k$  consécutif, intervalles sur lesquels le biais est déterministe et ne dépend plus de la réalisation  $\mathbf{y}$ . On obtient le résultat suivant :

**Résultat 18.** Dans le régime à fort RSB et pour des paramètres indépendants aléatoires de distribution connue  $p(\theta_l; l)$ , la BEQM conditionnée à l'ensemble  $\Omega^L$  est également donnée par

$$\text{BEQM}_{\Omega^L, \text{fort}} = \sum_{l=1}^L \frac{1}{P_{\Omega}^{(l)}} \sum_{k=1}^K \mathbb{E}_{\theta_l} (b(\theta_l)^2 1_{\Omega_k}(\theta_l)). \quad (3.104)$$

La BEQM augmente proportionnellement avec le nombre de paramètres à estimer, puisque l'on voit que chacun de ses paramètres contribue individuellement à l'erreur totale de manière additive. Le second paramètre essentiel, bien qu'il n'apparaisse pas directement dans les expressions est la taille de la grille  $r$  qui détermine la taille de l'espace  $\Omega_k$  sur lequel on intègre. Nous allons montrer pour des distributions particulières permettant de faire apparaître explicitement l'influence de  $r$  que la BEQM diminue quand on réduit  $r$ . Nous allons prendre pour exemple les distributions uniformes et gaussiennes, et pour chacune d'elle nous allons utiliser une formule différente pour calculer la BEQM. Ces deux exemples montrent l'intérêt d'exprimer l'erreur de deux façons différentes, qui peuvent chacune être plus appropriée selon les situations.

### 3.4.3 Exemple de distributions

Prenons tout d'abord le cas d'une distribution uniforme sur l'intervalle  $\Omega = [\bar{\theta}_1, \bar{\theta}_K]$  qui est la distribution la plus commune étant donné qu'elle est utilisée quand on n'a pas d'a priori sur des valeurs plus probables que d'autres. Selon l'intervalle où nous nous situons, les distributions conditionnées à un intervalle donné  $\theta_k$  sont :

1. pour  $2 \leq k \leq K-1$  : la distribution uniforme sur un intervalle de longueur  $r$  est

$$p(\theta_l | \theta_l \in \Omega_k) = \begin{cases} \frac{1}{r} & \text{dans } \Omega_k = [-\frac{r}{2} + \bar{\theta}_k, \bar{\theta}_k + \frac{r}{2}] \\ 0 & \text{ailleurs.} \end{cases} \quad (3.105)$$

2. pour  $k = 1$  : les intervalles de bords  $\Omega_1$  et  $\Omega_K$  ont une longueur de  $r/2$ , ce qui donne une distribution tronquée :

$$p(\theta_l | \theta_l \in \Omega_k) = \begin{cases} \frac{2}{r} & \text{dans } \Omega_1 = [\bar{\theta}_1, \bar{\theta}_1 + \frac{r}{2}] \\ 0 & \text{ailleurs.} \end{cases} \quad (3.106)$$

3. et pour  $k = K$  :

$$p(\theta_l | \theta_l \in \Omega_K) = \begin{cases} \frac{2}{r} & \text{dans } \Omega_K = [-\frac{r}{2} + \bar{\theta}_K, \bar{\theta}_K] \\ 0 & \text{ailleurs.} \end{cases} \quad (3.107)$$

Pour calculer la BEQM à partir de l'expression (3.103) on dérive  $\mathbb{E}_{\theta_l | \theta_l \in \Omega_k} (|\bar{\theta}_k - \theta_l|^2)$  dans chacun de ces cas :

$$\mathbb{E}_{\theta_l | \theta_l \in \Omega_k} [|\bar{\theta}_k - \theta_l|^2] = \int_{\theta_l} |\bar{\theta}_k - \theta_l|^2 p(\theta_l | \theta_l \in \Omega_k) d\theta_l = \frac{1}{3r} \left[ |\theta_l - \bar{\theta}_k|^3 \right]_{-\frac{r}{2} + \bar{\theta}_k}^{\bar{\theta}_k + \frac{r}{2}} = \frac{r^2}{12}$$

$$\mathbb{E}_{\theta_l | \theta_l \in \Omega_1} [|\bar{\theta}_1 - \theta_l|^2] = \int_{\theta_l} |\bar{\theta}_1 - \theta_l|^2 p(\theta_l | \theta_l \in \Omega_1) d\theta_l = \frac{2}{3r} \left[ |\theta_l - \bar{\theta}_1|^3 \right]_{\bar{\theta}_1}^{\bar{\theta}_1 + \frac{r}{2}} = \frac{r^2}{12}$$

$$\mathbb{E}_{\theta_l | \theta_l \in \Omega_K} [|\bar{\theta}_K - \theta_l|^2] = \int_{\theta_l} |\bar{\theta}_K - \theta_l|^2 p(\theta_l | \theta_l \in \Omega_K) d\theta_l = \frac{2}{3r} \left[ |\theta_l - \bar{\theta}_K|^3 \right]_{-\frac{r}{2} + \bar{\theta}_K}^{\bar{\theta}_K} = \frac{r^2}{12}.$$

On en conclut que la BEQM est :

$$\text{BEQM}_{\Omega, \text{fort}}(\text{uniforme}) = \frac{r^2}{12} \underbrace{\sum_{l=1}^L \sum_{k=1}^K \text{Pr}(\theta_l \in \Omega_k)}_{=1},$$

ce qui amène le résultat suivant.

**Résultat 19.** *Pour une distribution uniforme, à fort RSB la BEQM est :*

$$\text{BEQM}_{\Omega, \text{fort}}(\text{uniforme}) = \frac{Lr^2}{12}. \quad (3.108)$$

Il est intéressant de noter que ce résultat est exactement égal à l'erreur de quantification. En effet l'opération de discrétisation d'un paramètre continu est très similaire à une opération de quantification qui cherche la valeur appartenant à une grille uniforme la plus proche pour chaque entrée continue. Le résultat est le même quand on considère l'hypothèse fort RSB en isolant l'erreur de discrétisation. On peut noter que l'erreur est proportionnelle au nombre de paramètres  $L$  comme nous l'avons déjà remarqué précédemment, mais on remarque aussi qu'elle est proportionnelle au carré de  $r$  qui est donc un paramètre essentiel.

La distribution gaussienne est également courante, elle permet de modéliser les situations où on a une connaissance a priori sur  $\theta_l$ . On considère une distribution gaussienne de moyenne  $\alpha_l$  et de variance  $\sigma_l^2$  :

$$p(\theta_l; l) = \frac{1}{\sqrt{2\pi}\sigma_l} e^{-\frac{1}{2}\left(\frac{\theta_l - \alpha_l}{\sigma_l}\right)^2}. \quad (3.109)$$

Pour dériver l'expression de la BEQM pour une distribution gaussienne, nous partons cette fois-ci de l'expression (3.104). Nous allons tout d'abord calculer le terme

$$A_k \triangleq \mathbb{E}_{\theta_l} \left( (\bar{\theta}_k - \theta_l)^2 1_{\Omega_k}(\theta_l) \right). \quad (3.110)$$

De la même manière que pour la distribution uniforme, nous prenons tout d'abord les intervalles  $2 \leq k \leq K-1$ . Dans ces cas :

$$A_k = \frac{1}{\sqrt{2\pi}\sigma_l} \int_{\bar{\theta}_k - \frac{r}{2}}^{\bar{\theta}_k + \frac{r}{2}} (\bar{\theta}_k - \theta_l)^2 e^{-\frac{1}{2} \left( \frac{\theta_l - \alpha_l}{\sigma_l} \right)^2} d\theta_l$$

on utilise le changement de variable  $b(\theta_l) = \bar{\theta}_k - \theta_l$

$$\begin{aligned} &= -\frac{1}{\sqrt{2\pi}\sigma_l} \int_{\frac{r}{2}}^{-\frac{r}{2}} b(\theta_l)^2 e^{-\frac{1}{2} \left( \frac{\bar{\theta}_k - b(\theta_l) - \alpha_l}{\sigma_l} \right)^2} db(\theta_l) \\ &= \frac{1}{\sqrt{2\pi}\sigma_l} \int_{-\frac{r}{2}}^{\frac{r}{2}} b(\theta_l)^2 e^{-\frac{1}{2} \left( \frac{b(\theta_l) + \alpha_l - \bar{\theta}_k}{\sigma_l} \right)^2} db(\theta_l). \end{aligned}$$

On utilise la formule de l'intégrale de la gaussienne :

$$\int x^2 \phi(a + bx) dx \triangleq b^{-3} [(a^2 + 1) \Phi(a + bx) + (a - bx) \phi(a + bx)] + C, \quad (3.111)$$

où l'on pose

$$\phi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \text{ est la gaussienne centrée de variance 1 et} \quad (3.112)$$

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt = \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{x}{\sqrt{2}} \right) \right) \text{ est sa fonction de répartition.} \quad (3.113)$$

$\operatorname{erf}$  est la fonction d'erreur  $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ . En utilisant (3.111) on peut en déduire l'intégrale :

$$\begin{aligned} A_k &= \left[ \left( (\alpha_l - \bar{\theta}_k)^2 + \sigma_l^2 \right) \Phi \left( \frac{b(\theta_l) + \alpha_l - \bar{\theta}_k}{\sigma_l} \right) + \sigma_l (\alpha_l - \bar{\theta}_k - b(\theta_l)) \phi \left( \frac{b(\theta_l) + \alpha_l - \bar{\theta}_k}{\sigma_l} \right) \right]_{-\frac{r}{2}}^{\frac{r}{2}} \\ &= \left[ \frac{1}{2} \left( (\alpha_l - \bar{\theta}_k)^2 + \sigma_l^2 \right) \left( 1 + \operatorname{erf} \left( \frac{b(\theta_l) + \alpha_l - \bar{\theta}_k}{\sqrt{2}\sigma_l} \right) \right) + \frac{\sigma_l}{\sqrt{2\pi}} (\alpha_l - \bar{\theta}_k - b(\theta_l)) e^{-\frac{1}{2} \left( \frac{b(\theta_l) + \alpha_l - \bar{\theta}_k}{\sigma_l} \right)^2} \right]_{-\frac{r}{2}}^{\frac{r}{2}} \\ &= \frac{1}{2} \left( (\alpha_l - \bar{\theta}_k)^2 + \sigma_l^2 \right) \left( \operatorname{erf} \left( \frac{r + 2(\alpha_l - \bar{\theta}_k)}{2\sqrt{2}\sigma_l} \right) - \operatorname{erf} \left( \frac{-r + 2(\alpha_l - \bar{\theta}_k)}{2\sqrt{2}\sigma_l} \right) \right) \\ &\quad + \frac{\sigma_l}{2\sqrt{2\pi}} \left( (2(\alpha_l - \bar{\theta}_k) - r) e^{-\frac{1}{2} \left( \frac{r + 2(\alpha_l - \bar{\theta}_k)}{2\sigma_l} \right)^2} - (2(\alpha_l - \bar{\theta}_k) + r) e^{-\frac{1}{2} \left( \frac{-r + 2(\alpha_l - \bar{\theta}_k)}{2\sigma_l} \right)^2} \right). \quad (3.114) \end{aligned}$$

Pour les cas  $k = 1$  on a de la même façon :

$$A_1 = \frac{1}{\sqrt{2\pi}\sigma_l} \int_{-\frac{r}{2}}^0 b(\theta_l)^2 e^{-\frac{1}{2} \left( \frac{b(\theta_l) + \alpha_l - \bar{\theta}_1}{\sigma_l} \right)^2} db(\theta_l),$$

ce qui en utilisant (3.111) devient :

$$\begin{aligned}
 A_1 = & \frac{1}{2} \left( (\alpha_l - \bar{\theta}_1)^2 + \sigma_l^2 \right) \left( \operatorname{erf} \left( \frac{\alpha_l - \bar{\theta}_1}{\sqrt{2}\sigma_l} \right) - \operatorname{erf} \left( \frac{-r + 2(\alpha_l - \bar{\theta}_1)}{2\sqrt{2}\sigma_l} \right) \right) \\
 & + \frac{\sigma_l}{2\sqrt{2\pi}} \left( 2(\alpha_l - \bar{\theta}_1) e^{-\frac{1}{2} \left( \frac{\alpha_l - \bar{\theta}_1}{2\sigma_l} \right)^2} - (2(\alpha_l - \bar{\theta}_1) + r) e^{-\frac{1}{2} \left( \frac{-r + 2(\alpha_l - \bar{\theta}_1)}{2\sigma_l} \right)^2} \right), \quad (3.115)
 \end{aligned}$$

et enfin pour  $k = K$  de la même façon on obtient :

$$\begin{aligned}
 A_K = & \frac{1}{\sqrt{2\pi}\sigma_l} \int_0^{\frac{r}{2}} b(\theta_l)^2 e^{-\frac{1}{2} \left( \frac{b(\theta_l) + \alpha_l - \bar{\theta}_K}{\sigma_l} \right)^2} db(\theta_l) \\
 = & \frac{1}{2} \left( (\alpha_l - \bar{\theta}_K)^2 + \sigma_l^2 \right) \left( \operatorname{erf} \left( \frac{r + 2(\alpha_l - \bar{\theta}_K)}{2\sqrt{2}\sigma_l} \right) - \operatorname{erf} \left( \frac{\alpha_l - \bar{\theta}_K}{\sqrt{2}\sigma_l} \right) \right) \\
 & + \frac{\sigma_l}{2\sqrt{2\pi}} \left( (2(\alpha_l - \bar{\theta}_K) - r) e^{-\frac{1}{2} \left( \frac{r + 2(\alpha_l - \bar{\theta}_K)}{2\sigma_l} \right)^2} - 2(\alpha_l - \bar{\theta}_K) e^{-\frac{1}{2} \left( \frac{\alpha_l - \bar{\theta}_K}{2\sigma_l} \right)^2} \right). \quad (3.116)
 \end{aligned}$$

On a également besoin de calculer  $P_\Omega^{(l)}$  pour connaître tous les termes de (3.104) :

$$\begin{aligned}
 P_\Omega^{(l)} = & \frac{1}{\sqrt{2\pi}\sigma_l} \int_{\bar{\theta}_1}^{\bar{\theta}_K} e^{-\frac{1}{2} \left( \frac{\theta_l - \alpha_l}{\sigma_l} \right)^2} d\theta_l = \frac{1}{\sqrt{2\pi}} \int_{\frac{\bar{\theta}_1 - \alpha_l}{\sigma_l}}^{\frac{\bar{\theta}_K - \alpha_l}{\sigma_l}} e^{-\frac{x^2}{2}} dx \\
 = & \left[ \frac{1}{2} \left( 1 + \operatorname{erf} \left( \frac{x}{\sqrt{2}} \right) \right) \right]_{\frac{\bar{\theta}_1 - \alpha_l}{\sigma_l}}^{\frac{\bar{\theta}_K - \alpha_l}{\sigma_l}} = \frac{1}{2} \left[ \operatorname{erf} \left( \frac{\bar{\theta}_K - \alpha_l}{\sqrt{2}\sigma_l} \right) - \operatorname{erf} \left( \frac{\bar{\theta}_1 - \alpha_l}{\sqrt{2}\sigma_l} \right) \right]. \quad (3.117)
 \end{aligned}$$

En combinant les expressions (3.114), (3.115), (3.116) et (3.117) dans (3.104) on obtient le résultat qui suit.

**Résultat 20.** La  $\text{BEQM}_{\Omega, \text{fort}}$  pour la distribution  $\theta_l \sim \mathcal{N}(\alpha_l, \sigma_l^2)$  est donnée par :

$$\text{BEQM}_{\Omega, \text{fort}} = \sum_{l=1}^L \frac{1}{\operatorname{erf} \left( \frac{\bar{\theta}_K - \alpha_l}{\sqrt{2}\sigma_l} \right) - \operatorname{erf} \left( \frac{\bar{\theta}_1 - \alpha_l}{\sqrt{2}\sigma_l} \right)} \sum_{k=1}^K A_k, \quad (3.118)$$

où

$$A_k = \left( (\alpha_l - \bar{\theta}_k)^2 + \sigma_l^2 \right) \left( \operatorname{erf} \left( \frac{a_k + 2(\alpha_l - \bar{\theta}_k)}{2\sqrt{2}\sigma_l} \right) - \operatorname{erf} \left( \frac{b_k + 2(\alpha_l - \bar{\theta}_k)}{2\sqrt{2}\sigma_l} \right) \right) \quad (3.119)$$

$$+ \frac{\sigma_l}{\sqrt{2\pi}} \left( (2(\alpha_l - \bar{\theta}_k) - a_k) e^{-\frac{1}{2} \left( \frac{a_k + 2(\alpha_l - \bar{\theta}_k)}{2\sigma_l} \right)^2} - (2(\alpha_l - \bar{\theta}_k) - b_k) e^{-\frac{1}{2} \left( \frac{b_k + 2(\alpha_l - \bar{\theta}_k)}{2\sigma_l} \right)^2} \right), \quad (3.120)$$

où les paramètres  $a_k$  et  $b_k$  prennent les valeurs suivantes selon  $k$  :

$$a_k = \begin{cases} r & \text{pour } 2 \leq k \leq N-1 \\ 0 & \text{pour } k = 1 \\ r & \text{pour } k = K \end{cases}$$

$$b_k = \begin{cases} -r & \text{pour } 2 \leq k \leq N-1 \\ -r & \text{pour } k = 1 \\ 0 & \text{pour } k = K. \end{cases}$$

#### 3.4.4 Illustration de l'erreur de grille dans le contexte de l'estimation de fréquence

Dans cette section, nous allons illustrer les différents résultats obtenus en nous plaçant dans le contexte de l'estimation de fréquence que nous avons présenté dans la section 1.3.3. Pour ce faire nous allons générer des échantillons suivant le modèle (1.47) en utilisant un paramètre  $\omega_l$ , que nous allons estimer en utilisant un dictionnaire tel que présenté en (1.49) avec des  $\tilde{\omega}$  allant de 0 à  $\pi$ . Nous utiliserons pour la reconstruction des algorithmes utilisés dans le cadre de l'acquisition comprimée et de l'estimation parcimonieuse : l'Orthogonal Matching Pursuit (OMP) [97], sa variante parcimonieuse Compressive Sensing Approximate Message Passing (CoSaMP) [89] et enfin le SPGL1 [124] qui est une implémentation du Basis Pursuit (BP) [44]. Dans toutes les simulations nous faisons l'hypothèse que nous connaissons le nombre de fréquences à estimer  $L$ . L'OMP et le COSAMP étant des algorithmes gloutons, cette information est utilisée pour définir le nombre d'itération à effectuer. Il n'est pas possible au contraire de spécifier l'ordre du modèle à l'algorithme SPGL1 qui utilise un critère d'arrêt basé sur l'erreur entre les échantillons reconstruits à partir du modèle et les échantillons mesurés. De ce fait l'ordre est la plupart du temps surestimé, pour que cette erreur sur l'ordre n'entre pas en compte dans la BEQM mesurée, nous mettons un seuil et concrètement ne considérons que les  $L$  plus grandes valeurs. La figure 3.17 montre un résultat typique en sortie de l'algorithme SPGL1, quand il y a deux fréquences à estimer. On a pris ici des paramètres permettant d'obtenir une cohérence de  $\mu \approx 0.266$  qui vérifie donc les conditions de reconstruction (1.37) (voir section 1.3.spectre). La ligne rouge correspond aux fréquences à estimer, chaque pic correspond à une fréquence que l'on retrouve en abscisse, l'amplitude considérée étant de 1. Les points marqués en bleu permettent de visualiser le vecteur parcimonieux en sortie de l'algorithme SPGL1 qui donne une amplitude estimée pour chaque élément du dictionnaire. Quelques comportements peuvent être observés sur cet exemple, que l'on retrouve dans toutes nos simulations quels que soient les paramètres choisis :

1. L'ordre estimé est plus important que l'ordre réel. Cela est dû en grande partie à l'erreur de grille : au lieu de choisir la fréquence la plus proche comme nous l'avons supposé en hypothèse, cet algorithme ressort les deux fréquences autour de la vraie valeur.
2. Etant donné que l'ordre estimé est plus important, l'algorithme répartit la puissance sur plus de composantes ce qui a pour effet direct de sous-estimer l'amplitude. Nous nous intéressons ici uniquement à la fréquence estimée, l'erreur sur l'amplitude n'aura donc pas d'influence par la suite sur la BEQM.

3. en dehors des deux fréquences obtenues pour chaque paramètre dans le cas où les paramètres ne sont pas sur la grille, il y a également quelques autres petites valeurs qui ressortent : le vecteur obtenu n'est pas tout à fait parcimonieux, car les valeurs restent non nulles mêmes en dehors des pics. Celles-ci sont négligeables quand les paramètres sont sur la grille, mais pas tout à fait nulles. Pour le calcul de la BEQM nous ne gardons que les  $L$  plus grandes valeurs, puisque  $L$  est connu, ce qui permet d'éviter le problème de choix d'un seuil.
4. Si les paramètres choisis ne permettent pas de vérifier la condition (1.37), l'erreur de modèle existe même pour des paramètres sur la grille. On observe que l'algorithme basé sur le Basis Pursuit est plus sensible aux problèmes de cohérence que l'OMP.

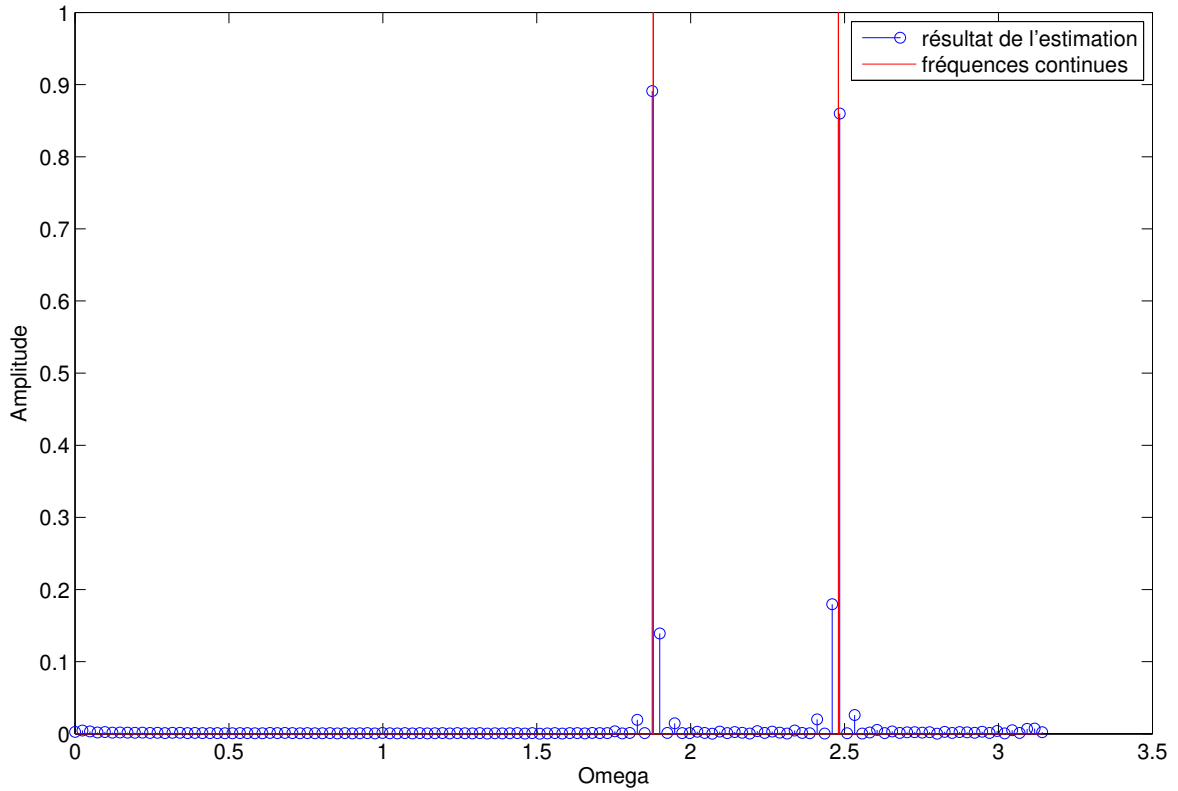


FIGURE 3.17 – Exemple de résultat d'estimation de l'algorithme SPGL1.  $L = 2, K = 130, N = 100, \text{RSB} = 50$ .

Nous traçons à présent la BEQM obtenue avec un moyennage de l'erreur d'estimation sur 1500 réalisations des algorithmes OMP, COSAMP et SPGL1 pour des fréquences angulaires appartenant à l'intervalle  $\Omega = [0, \pi]$ . Celles-ci sont tirées d'une distribution uniforme dans les figures 3.18, 3.19 et 3.21 et d'une distribution gaussienne centrée sur  $\pi/3$  et de variance  $r^2$  sur la figure 3.21. Les BEQM obtenus par simulation sont comparés dans chaque figure à la BEQM théorique donnée en (3.108) et (3.118) pour les distributions uniformes et gaussiennes respectivement, tracée par une ligne continue noire. Pour ces quatre courbes nous ne prenons qu'un seul paramètre à estimer, et nous avons fait varier les dimensions du dictionnaire  $K$  et  $N$ . On remarque tout d'abord que tous les algorithmes rejoignent la BEQM théorique à fort RSB, sachant que le point auquel le régime fort RSB est atteint dépend de l'algorithme et de la taille du dictionnaire choisie. Ces simulations

permettent donc de valider la pertinence de l'erreur obtenue et de la borne. Pour une seule source et de petits dictionnaires comme ceux utilisés dans ces simulations le régime à fort RSB est atteint globalement pour des RSB entre 20 et 30dB. En dehors de la zone à fort RSB où tous les algorithmes ont les mêmes performances, on observe que l'algorithme OMP est celui donnant les meilleurs résultats, car il a l'erreur la plus faible. A l'opposé l'algorithme SPGL1 basé sur le Basis Pursuit commet plus d'erreur, mais cela s'explique par le fait qu'il est plus sensible à l'erreur de grille. Comme on peut s'y attendre, au fur et à mesure que le bruit diminue l'erreur diminue également, mais ne peut pas atteindre zéro étant donné la présence de l'erreur de modèle. Au mieux les algorithmes peuvent atteindre le palier donné par la BEQM donné en (3.103) ou (3.104). Quand le bruit est important, quels que soit les paramètres choisis on obtient une erreur importante, du même ordre de grandeur. Ce point correspond à l'état où l'algorithme n'arrive plus à reconnaître les points correspondant à des fréquences à estimer par rapport à des pics dus au bruit. Quand  $K$  augmente, la taille de la grille  $r$  diminue ce qui a pour effet direct de baisser la BEQM théorique et le palier. C'est le cas par exemple sur la courbe 3.22 qui donne un exemple avec des valeurs plus importantes, tout en gardant une cohérence de 0.2665. L'intervalle est divisé en 130 ce qui permet d'obtenir une bonne précision, au coût d'un taux de compression assez faible. Avec cette cohérence, l'OMP et le COSAMP donnent des résultats très similaires et permettent d'atteindre le régime à fort RSB dès 0dB, qui peut être une valeur de RSB rencontrée en pratique. Au contraire augmenter  $N$  n'a pas d'effet direct sur la BEQM théorique, mais cela réduit le taux de compression du dictionnaire, pour le rendre plus orthogonal. Par exemple les paramètres de la courbe 3.18 donnent une cohérence  $\mu = 0.8172$  tandis que ceux sur la courbe 3.19 ont une cohérence de  $\mu = 0.3828$ . Notez que dans les deux cas il n'est pas possible d'estimer plus qu'un paramètre avec ces cohérences importantes. Il est intéressant de remarquer que sur l'exemple donné par ces deux courbes, le SPGL1 qui est un algorithme spécifiquement conçu dans le cadre de l'acquisition comprimée donne de meilleurs résultats pour le dictionnaire ayant une forme plus conforme à l'acquisition comprimée ( $N \ll K$ ), bien que la cohérence soit moins bonne dans ces conditions. Le paramètre  $N$  n'influe pas sur le régime à fort RSB, par contre de ce qu'on observe dans les simulations, il joue un rôle sur le rapport signal sur bruit à atteindre pour entrer dans le régime fort RSB. Une perspective intéressante pour la suite de nos travaux serait de chercher à étudier ce point à l'aide d'autres bornes bayésiennes. Les courbes 3.20 et 3.21 utilisent les mêmes paramètres pour une distribution uniforme et gaussienne respectivement, on peut observer sur cet exemple que le type de distribution choisi n'a pas vraiment de conséquence sur le comportement des différents algorithmes. Cette observation est vérifiée aussi pour des paramètres plus grands, comme on peut le voir en comparant les courbes 3.22 et 3.23. L'erreur est simplement plus faible pour la distribution gaussienne étant donné qu'on a utilisé une gaussienne de variance faible, les angles à estimer sont assez rapprochés. Pour un dictionnaire de taille  $K = 130$  allant entre 0 et  $\pi$ , les éléments sont trop rapprochés et les algorithmes n'arrivent pas à obtenir une estimation précise pour  $L = 2$ , l'erreur reste donc importante sur la courbe 3.24 quel que soit le niveau de bruit. Pour obtenir de bons résultats avec deux sources, on élargit le dictionnaire pour aller de  $\Omega = [-\pi, \pi]$ , tout en utilisant un dictionnaire peu cohérent de dimensions  $90 \times 100$ . Ces paramètres permettent d'obtenir une cohérence faible de  $\mu = 0.0987$ , qui permet d'atteindre le régime à fort RSB même avec deux sources (courbe 3.25).

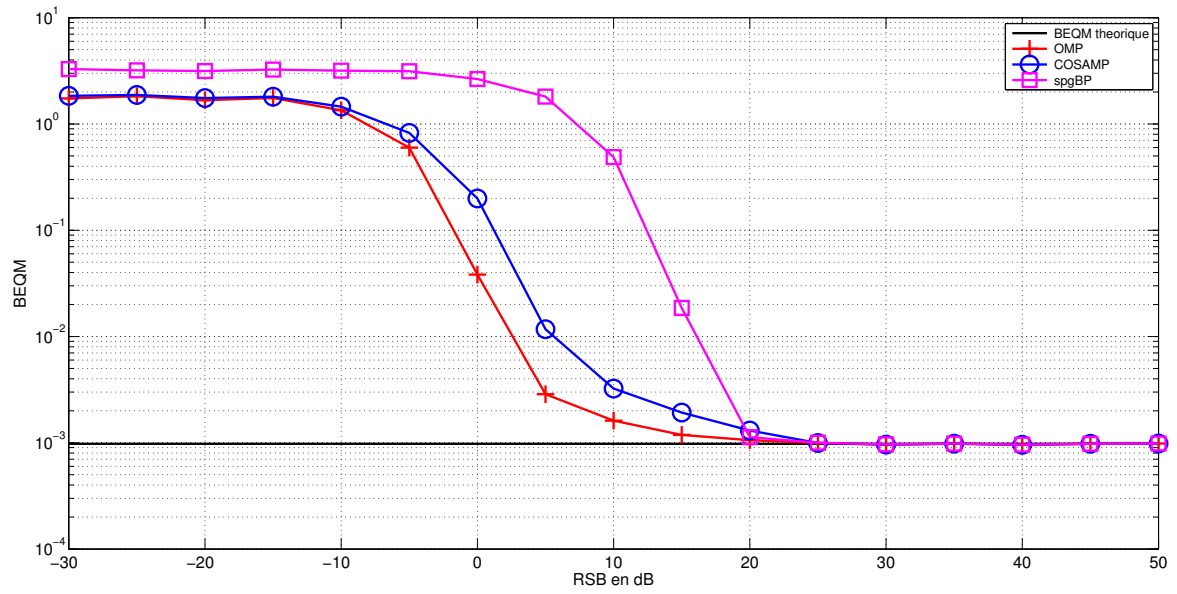


FIGURE 3.18 – BEQM pour une distribution uniforme.  $L = 1, K = 30, N = 10$ .

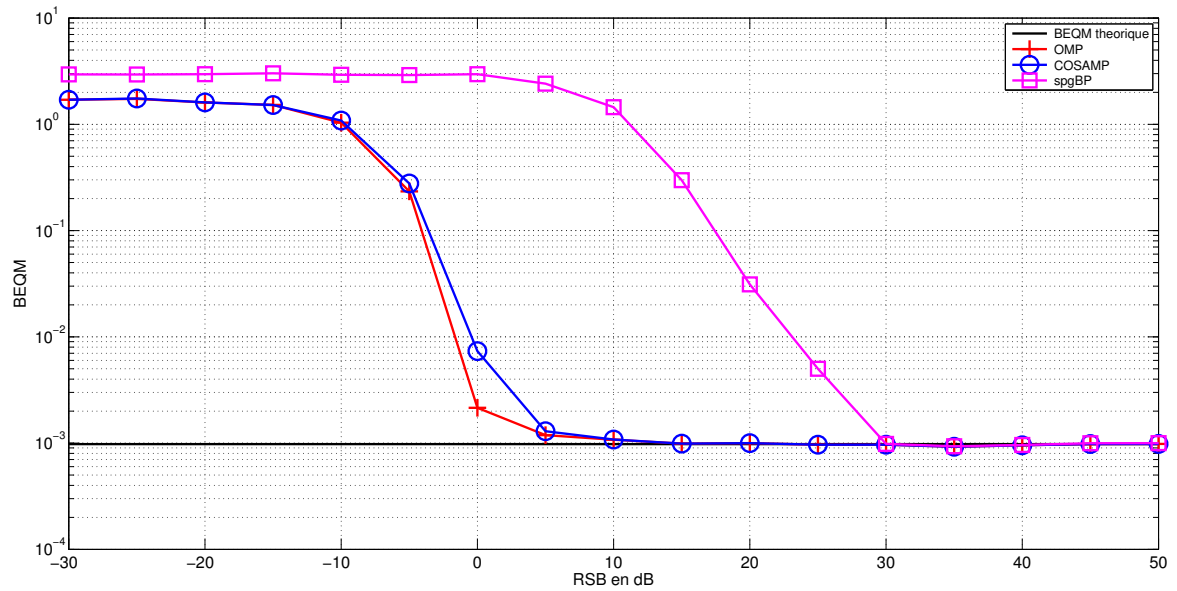


FIGURE 3.19 – BEQM pour une distribution uniforme.  $L = 1, K = 30, N = 20$ .

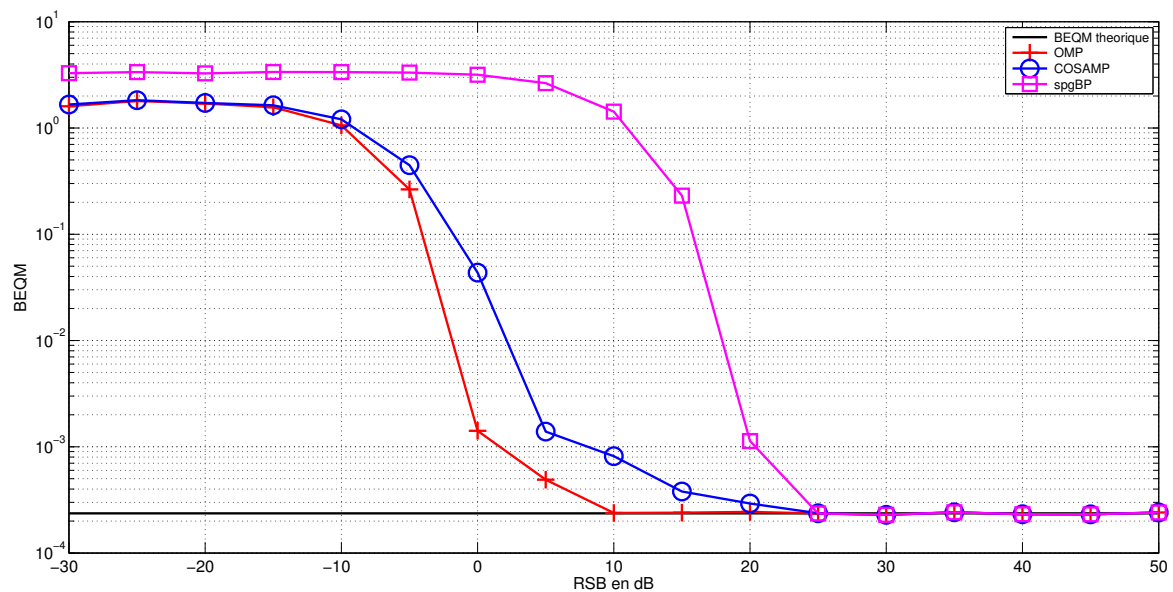


FIGURE 3.20 – BEQM pour une distribution uniforme.  $L = 1, K = 60, N = 30$ .

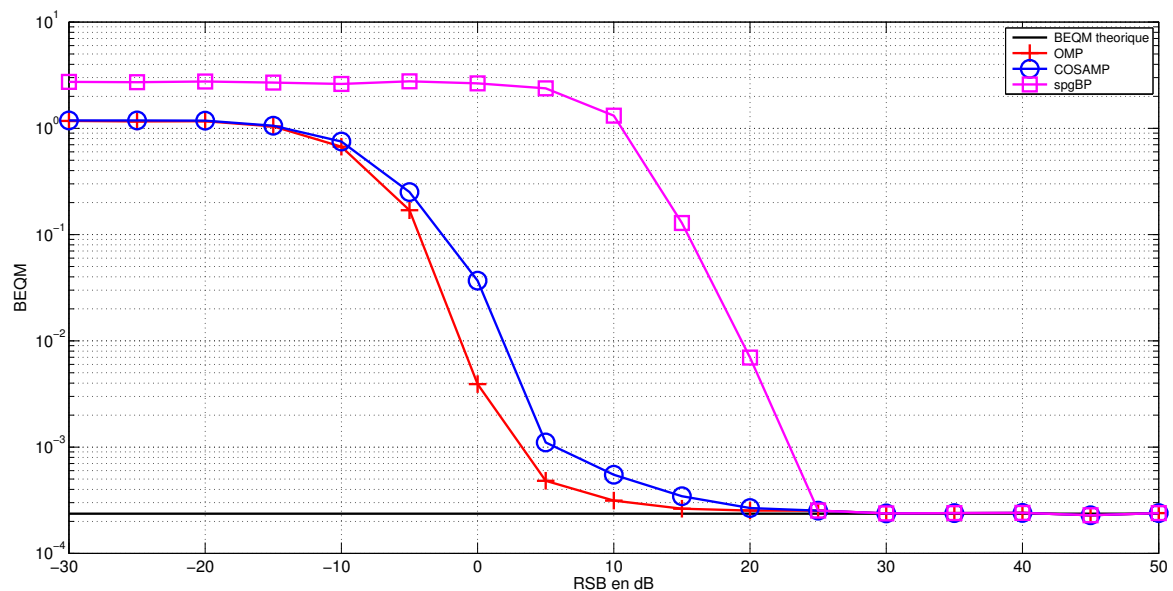


FIGURE 3.21 – BEQM pour une distribution gaussienne.  $L = 1, K = 60, N = 30$ .

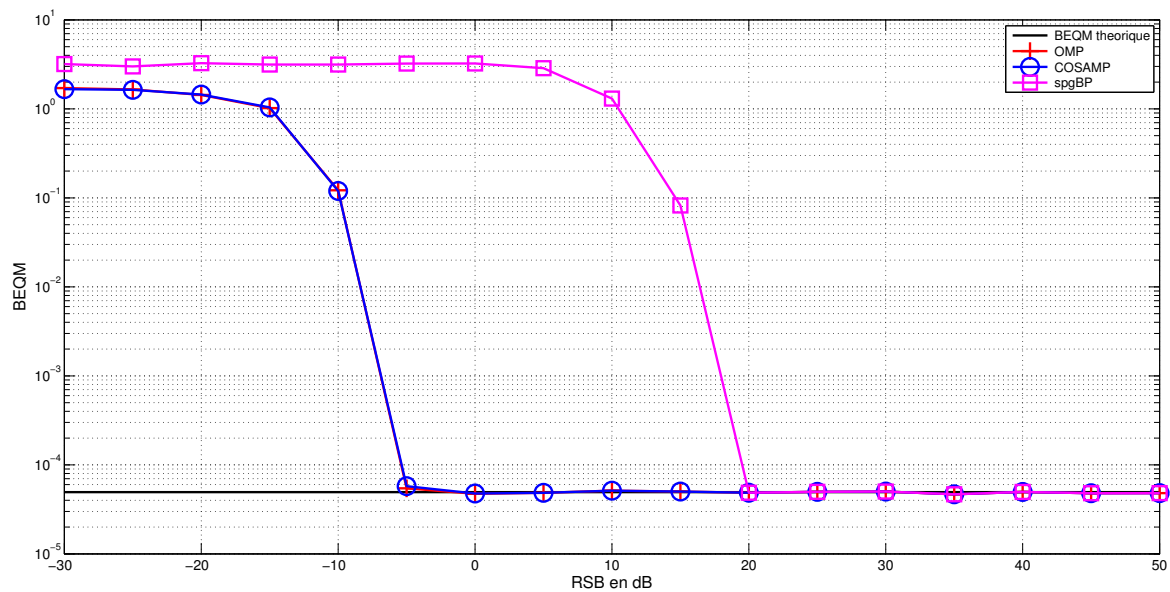


FIGURE 3.22 – BEQM pour une distribution uniforme.  $L = 1, K = 130, N = 100$ .

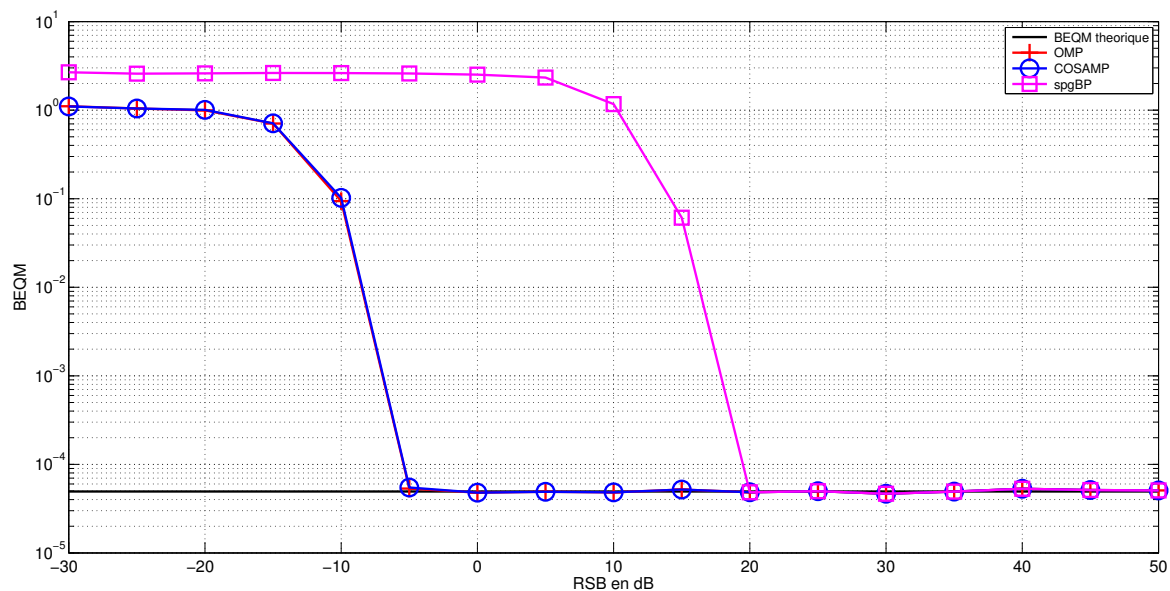


FIGURE 3.23 – BEQM pour une distribution gaussienne.  $L = 1, K = 130, N = 100$ .

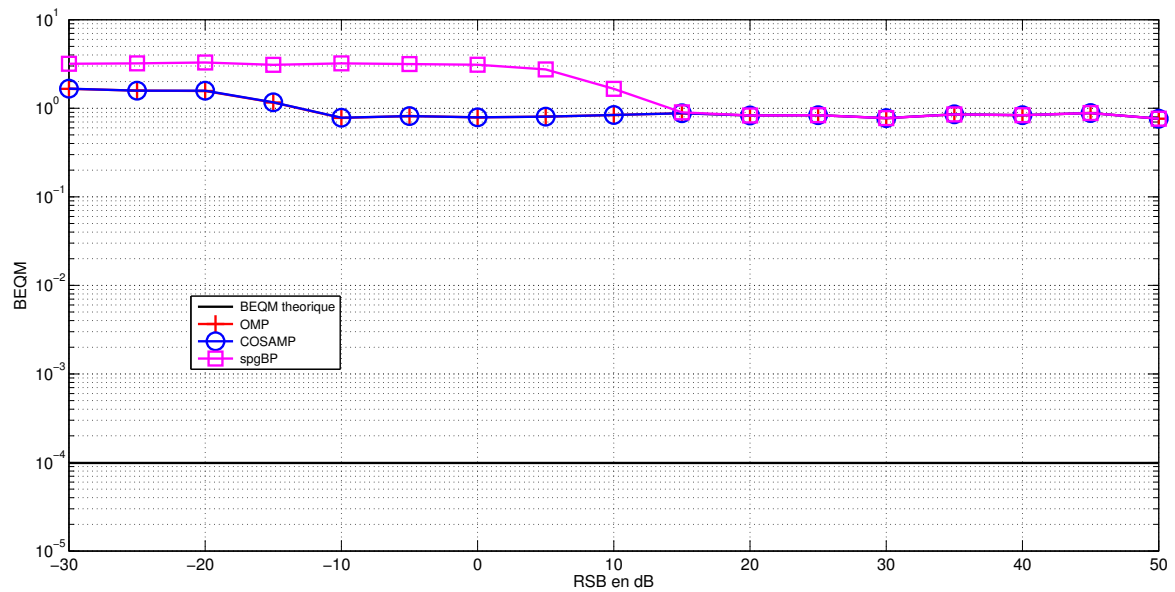


FIGURE 3.24 – BEQM pour une distribution uniforme.  $L = 2, K = 130, N = 100$ .

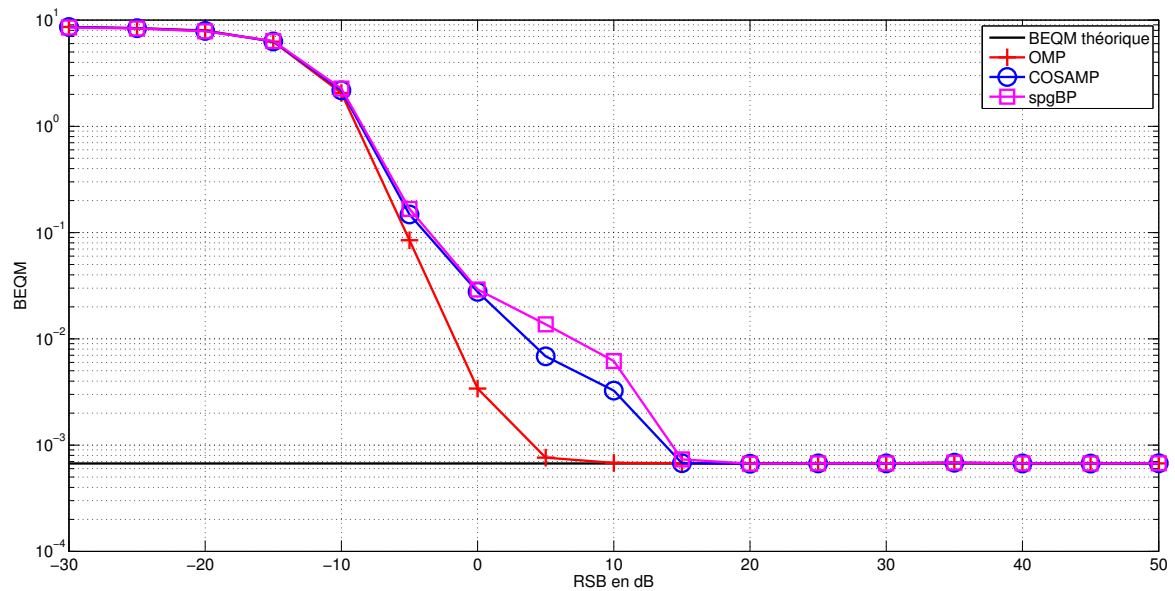


FIGURE 3.25 – BEQM pour une distribution uniforme.  $L = 2, K = 100, N = 90$ .

# Chapitre 4

## Correction des erreurs de discrétisation

### Sommaire

---

|                                                                                          |            |
|------------------------------------------------------------------------------------------|------------|
| <b>4.1 Algorithme OGEC pour corriger l'erreur EG</b>                                     | <b>120</b> |
| 4.1.1 Performances d'estimation dans le contexte structuré                               | 120        |
| 4.1.2 Présentation de l'algorithme OGEC                                                  | 123        |
| 4.1.3 Performances théoriques de l'algorithme                                            | 126        |
| 4.1.4 Résultat des simulations                                                           | 129        |
| <b>4.2 Algorithme BiCE pour corriger le DB</b>                                           | <b>132</b> |
| 4.2.1 Modèle structuré dans le contexte de dictionnaire bruité                           | 132        |
| 4.2.2 Présentation de l'algorithme BiCE                                                  | 133        |
| 4.2.3 Performances théoriques de l'algorithme                                            | 137        |
| 4.2.4 Résultat de l'algorithme pour l'estimation parcimonieuse des signaux impulsionnels | 145        |

---

Dans le chapitre 3 nous avons étudié les conséquences induites par la connaissance incertaine du modèle sur les performances que l'on peut attendre des algorithmes de reconstruction parcimonieuse. Nous avons vu que même en l'absence de bruit additif et avec des estimateurs donnant le meilleur résultat possible, la présence d'une erreur de modèle du type dictionnaire bruité ou de discrétisation introduit un palier dans les performances de l'estimation parcimonieuse. Nous nous intéressons à présent dans ce chapitre à la correction de ce palier. Pour cela nous nous plaçons dans le contexte d'une erreur de modèle structurée, la structure de l'erreur permettant d'isoler la contribution due aux erreurs de modèles pour l'annuler. Nous nous intéressons aux deux contextes premièrement de l'erreur de grille, puis de dictionnaire bruité. Nous proposons pour corriger ces erreurs deux algorithmes de post-traitement : l'*Off Grid Error Correction* (OGEC) et le *Bias Correction Estimator* (BiCE) respectivement. Nous nous plaçons toujours dans le contexte asymptotique à haut RSB où les estimateurs classiques (type OMP, BP...) proposent une bonne solution en l'absence d'erreur de modèle, les deux algorithmes introduits ayant pour objectif de rectifier uniquement la dégradation des performances due à l'erreur de modèle. Pour les RSB inférieurs, la correction des erreurs de modèle n'est pas pertinente puisque ces erreurs sont masquées par les erreurs dues au bruit.

Pour chacun des contextes d'erreur de grille et de dictionnaire bruité, nous présentons le modèle structuré correspondant. Puis nous proposons deux algorithmes de post-traitement permettant dans les deux contextes de corriger l'estimation de l'amplitude, et également d'estimer l'erreur de grille dans le contexte EG. Nous étudions ensuite les performances théoriques des algorithmes et leurs efficacités statistiques, puis nous démontrons leur utilisation par simulation dans le contexte présenté en 2.4.2. Le modèle de l'erreur de grille structurée est celui utilisé pour étudier l'erreur de discrétisation de l'estimation par dictionnaire, dans le cas à une dimension. Nous l'étudions dans la section 4.1.1 et proposons la borne Bayésienne oracle correspondante sur l'estimation des amplitudes et de l'erreur de grille.

## 4.1 Correction de l'erreur de grille par l'algorithme OGEC

### 4.1.1 Performances d'estimation dans le contexte structuré

#### Erreur additive structurée

Dans le chapitre 3, nous avons étudié les bornes sur l'estimation en présence d'une erreur de modèle de type additive, suivant le modèle  $\tilde{\mathbf{H}} = \mathbf{H} + \mathbf{E}$ . Nous avons étudié le modèle naïf d'une erreur additive gaussienne permettant d'avoir une idée des conséquences de l'erreur en terme de performances Bayésiennes. Nous nous intéressons maintenant à une sous-famille du modèle de l'erreur de grille représentant de manière plus réaliste l'erreur de discrétisation dans l'estimation par dictionnaire où le dictionnaire est obtenu à partir de la discrétisation d'une fonction ou d'un noyau donné : il s'agit du modèle d'erreur structurée. Ce modèle est généralement utilisé dans la littérature quand on cherche à estimer l'erreur de grille, puisque la connaissance de la structure du dictionnaire rend cette estimation possible [114, 134]. Ce type de dictionnaire est notamment obtenu dans le contexte de l'estimation par dictionnaire qui cherche à modéliser un problème de la forme donnée en (1.40) (voir section 1.3.3 pour une présentation de ce modèle) en discrétisant une fonction, ou noyau sur l'espace des paramètres à estimer. Cette discrétisation

est la cause des erreurs de grille qui peuvent donc être approximées au premier ordre par un modèle structuré, comme nous allons le présenter dans cette section.

Le modèle d'échantillonnage de l'estimation par dictionnaire, ne prenant pas en compte la présence de l'erreur de grille, est donné dans (3.86) où l'on choisit pour simplifier les notations de ne prendre en compte que des paramètres à une dimension :

$$\mathbf{y} \approx \mathbf{H}(\bar{\bar{\boldsymbol{\theta}}})\mathbf{s} + \mathbf{w}, \quad (4.1)$$

où le dictionnaire de dimension  $K$  est  $\bar{\bar{\boldsymbol{\theta}}} = [\bar{\bar{\theta}}_1, \dots, \bar{\bar{\theta}}_K]$ . Pour que les notations soient comparables à celles utilisées dans le chapitre 3 nous allons simplement écrire que  $\mathbf{H}(\bar{\bar{\boldsymbol{\theta}}}) = \mathbf{H}$  où les colonnes de  $\mathbf{H}$  notées  $\mathbf{H}(\bar{\theta}_k)$  sont construites à partir de l'échantillonnage de la fonction  $h(t, \theta)$  pour une valeur de paramètre  $\bar{\theta}_k$  :

$$\mathbf{H}(\bar{\theta}_k) = h(\mathbf{t}, \bar{\theta}_k), \quad (4.2)$$

où  $\mathbf{t}$  est le vecteur des temps échantillonnés de longueur  $N$ . Si l'échantillonnage est uniforme avec un intervalle d'échantillonnage de durée  $T_S$  on a par exemple  $[\mathbf{t}]_n = nT_S$ . Pour que l'égalité soit vraie dans l'expression (3.86), il a fallu faire l'hypothèse que les paramètres à estimer  $\theta_l \in \bar{\bar{\boldsymbol{\theta}}}$ . En pratique les  $\theta_l$  sont continus, ce qui génère une erreur de grille. Notons  $\bar{\theta}_{k_l}$  l'élément de la grille le plus proche de  $\theta_l$ , on a  $\theta_l = \bar{\theta}_{k_l} + e_l$  et :

$$\mathbf{y} = \sum_{l=1}^L h(\mathbf{t}, \theta_l) s_l + \mathbf{w} = \sum_{l=1}^L h(\mathbf{t}, \bar{\theta}_{k_l} + e_l) s_l + \mathbf{w}. \quad (4.3)$$

Le modèle d'erreur structurée est obtenu en faisant une interpolation de Taylor de  $h(\mathbf{t}, \bar{\theta}_{k_l} + e_l)$  en considérant une erreur  $e_l$  proche de zéro :

$$h(\mathbf{t}, \bar{\theta}_{k_l} + e_l) = h(\mathbf{t}, \bar{\theta}_{k_l}) + \left. \frac{\partial h(\mathbf{t}, \theta)}{\partial \theta} \right|_{\theta=\bar{\theta}_{k_l}} e_l + O(e_l^2). \quad (4.4)$$

En utilisant ce développement, on a, pour une erreur petite :

$$\mathbf{y} \approx \sum_{l=1}^L s_l \left( h(\mathbf{t}, \bar{\theta}_{k_l}) + \left. \frac{\partial h(\mathbf{t}, \theta)}{\partial \theta} \right|_{\theta=\bar{\theta}_{k_l}} e_l \right) + \mathbf{w} \quad (4.5)$$

$$= \tilde{\mathbf{H}}_S \mathbf{x} + \mathbf{w}, \quad (4.6)$$

où le dictionnaire corrompu est :

$$\tilde{\mathbf{H}}_S \approx \mathbf{H}_S + \dot{\mathbf{H}}_S \text{diag}(\mathbf{e}), \quad (4.7)$$

où  $\mathbf{e} = [e_1, \dots, e_L]^T$  et

$$\dot{\mathbf{H}}_S = \left[ \left. \frac{\partial h(\mathbf{t}, \theta)}{\partial \theta} \right|_{\theta=\bar{\theta}_{k_1}}, \dots, \left. \frac{\partial h(\mathbf{t}, \theta)}{\partial \theta} \right|_{\theta=\bar{\theta}_{k_L}} \right]. \quad (4.8)$$

Dans ce modèle, on n'a conservé que les amplitudes non nulles, c'est à dire que  $\mathbf{H}_S$  est le dictionnaire réduit conservant uniquement les  $L$  colonnes du dictionnaire  $\mathbf{H}$  correspondant aux  $\bar{\theta}_{k_l}$ , tandis que le vecteur  $\mathbf{x}$  est le vecteur contenant uniquement les amplitudes non nulles de  $\mathbf{s}$ . Le modèle d'erreur donné en (4.7) est utilisé pour développer des algorithmes d'estimation permettant d'estimer à la fois  $\mathbf{s}$  et l'erreur associée  $\mathbf{e}$  [114, 74].

### Borne bayésienne pour une erreur de grille structurée

Pour étudier le problème spécifique de l'erreur de grille, nous nous plaçons toujours dans le cas à haut RSB pour lequel le support de  $\mathbf{s}$  est connu. En effet nous avons vu dans le chapitre 3 que même si le support est connu et le RSB grand, une erreur d'estimation persiste. De ce fait les paramètres restant à estimer sont  $\mu = [\mathbf{x}, \mathbf{e}]$ . Nous dérivons la borne oracle dans le contexte d'une erreur de grille structurée, présenté en (4.7). Nous nous plaçons dans le contexte bayésien, où  $\mathbf{x}$  est un vecteur aléatoire suivant une distribution centrée ayant comme matrice de covariance  $\sigma_x^2 \mathbf{I}$  et  $\mathbf{e}$  est un vecteur aléatoire suivant une distribution centrée ayant comme matrice de covariance  $\sigma_e^2 \mathbf{I}$ . Les observations sont les échantillons  $\mathbf{y}$  et les matrices  $\mathbf{H}_S$  et  $\dot{\mathbf{H}}_S$  qui sont connues.

La matrice d'information de Fisher bayésienne est dans ce contexte :

$$\mathbf{F}_S = \mathbb{E}_{\mathbf{y}, \mu | \mathbf{H}_S} \left\{ - \frac{\partial^2 \log p(\mathbf{y}, \mu | \mathbf{H}_S)}{\partial \mu^T \partial \mu} \right\}. \quad (4.9)$$

Le logarithme de la distribution jointe peut être décomposé par :

$$\log p(\mathbf{y}, \mu | \mathbf{H}_S) = \log p(\mathbf{y} | \mu, \mathbf{H}_S) + \log p(\mu | \mathbf{H}_S). \quad (4.10)$$

On considère que les variables  $\mathbf{x}$  et  $\mathbf{e}$  sont indépendantes, on a donc la propriété que

$$\frac{\partial^2 p(\mu | \mathbf{H}_S)}{\partial \mathbf{x}^T \partial \mathbf{x}} = \frac{\partial^2 p(\mathbf{x} | \mathbf{H}_S)}{\partial \mathbf{x}^T \partial \mathbf{x}}, \quad (4.11)$$

$$\frac{\partial^2 p(\mu | \mathbf{H}_S)}{\partial \mathbf{e}^T \partial \mathbf{e}} = \frac{\partial^2 p(\mathbf{e} | \mathbf{H}_S)}{\partial \mathbf{e}^T \partial \mathbf{e}}, \quad (4.12)$$

$$\frac{\partial^2 p(\mu | \mathbf{H}_S)}{\partial^2 \mathbf{x}^T \partial \mathbf{e}} = \frac{\partial p(\mu | \mathbf{H}_S)}{\partial \mathbf{e}^T \partial \mathbf{x}} = 0. \quad (4.13)$$

La MIB est finalement :

$$\mathbf{F}_S = \mathbb{E}_{\mu} \left( \mathbb{E}_{\mathbf{y} | \mu, \mathbf{H}_S} \left\{ - \frac{\partial^2 \log p(\mathbf{y} | \mu, \mathbf{H}_S)}{\partial \mu^T \partial \mu} \right\} \right) + \begin{bmatrix} \mathbf{P}_x & 0 \\ 0 & \mathbf{P}_e \end{bmatrix}, \quad (4.14)$$

où la matrice associée à la distribution de chaque paramètre à estimer est

$$\mathbf{P}_x = \mathbb{E}_x \left\{ - \frac{\partial^2 \log p(\mathbf{x} | \mathbf{H}_S)}{\partial \mathbf{x}^T \partial \mathbf{x}} \right\}, \quad (4.15)$$

$$\mathbf{P}_e = \mathbb{E}_e \left\{ - \frac{\partial^2 \log p(\mathbf{e} | \mathbf{H}_S)}{\partial \mathbf{e}^T \partial \mathbf{e}} \right\}. \quad (4.16)$$

Le bruit additif  $\mathbf{w}$  est un bruit blanc gaussien de variance  $\sigma^2$ , de ce fait la distribution conditionnelle des échantillons suit une distribution gaussienne :  $\mathbf{y} | \mu \sim \mathcal{N}(\tilde{\mathbf{H}}_S \mathbf{x}, \sigma^2 \mathbf{I})$ . Les dérivées de la moyenne de cette distribution en fonction des différents paramètres à estimer sont :

$$\frac{\partial \tilde{\mathbf{H}}_S \mathbf{x}}{\partial \mathbf{x}} = \tilde{\mathbf{H}}_S, \quad (4.17)$$

$$\frac{\partial \tilde{\mathbf{H}}_S \mathbf{x}}{\partial \mathbf{e}} = \frac{\partial (\mathbf{H}_S \mathbf{x} + \dot{\mathbf{H}}_S \text{diag}(\mathbf{x}) \mathbf{e})}{\partial \mathbf{e}} = \dot{\mathbf{H}}_S \text{diag}(\mathbf{x}). \quad (4.18)$$

$$(4.19)$$

A partir de la formule de Slepian-Bang (1.60) et des dérivées ci-dessus, on peut exprimer la MIB conditionnelle :

$$\mathbb{E}_{\mathbf{y}|\mu, \mathbf{H}_S} \left\{ -\frac{\partial^2 \log p(\mathbf{y}|\mu, \mathbf{H}_S)}{\partial \mu^T \partial \mu} \right\} = \frac{1}{\sigma^2} \begin{bmatrix} \tilde{\mathbf{H}}_S^T \tilde{\mathbf{H}}_S & \dot{\mathbf{H}}_S^T \tilde{\mathbf{H}}_S \text{diag}(\mathbf{x}) \\ \tilde{\mathbf{H}}_S^T \dot{\mathbf{H}}_S \text{diag}(\mathbf{x}) & \text{diag}(\mathbf{x}) \dot{\mathbf{H}}_S^T \dot{\mathbf{H}}_S \text{diag}(\mathbf{x}) \end{bmatrix}. \quad (4.20)$$

Le vecteur  $\mathbf{x}$  suit une distribution centrée, on obtient que la moyenne de la MIB conditionnelle est une matrice bloc-diagonale, ce qui permet de conclure que la BCRB sur chaque paramètre est découpée. On a

$$\begin{aligned} \mathbb{E}_{\mu} \{ \tilde{\mathbf{H}}_S^T \tilde{\mathbf{H}}_S \} &= \mathbf{H}_S^T \mathbf{H}_S + (\mathbf{H}_S^T \dot{\mathbf{H}}_S + \dot{\mathbf{H}}_S^T \mathbf{H}_S) \mathbb{E}_{\mathbf{e}} \{ \text{diag}(\mathbf{e}) \} + \dot{\mathbf{H}}_S^T \dot{\mathbf{H}}_S \{ \text{diag}(\mathbf{e})^2 \} \\ &= \mathbf{H}_S^T \mathbf{H}_S + \sigma_e^2 \dot{\mathbf{H}}_S^T \dot{\mathbf{H}}_S, \end{aligned} \quad (4.21)$$

$$\mathbb{E}_{\mu} \{ \text{diag}(\mathbf{x}) \dot{\mathbf{H}}_S^T \dot{\mathbf{H}}_S \text{diag}(\mathbf{x}) \} = \sigma_x^2 \text{diag}(\dot{\mathbf{H}}_S^T \dot{\mathbf{H}}_S). \quad (4.22)$$

On peut en déduire le résultat suivant sur la BCRB.

**Résultat 21.** *La BCRB pour le modèle d'erreur structuré, et des amplitudes i.i.d. de moyenne nulle et de variance  $\sigma_x^2$  et pour des erreurs i.i.d. de moyenne nulle et de variance  $\sigma_e^2$  est découpée pour chaque paramètre et est donnée par :*

$$\text{BCRB}_{\mathbf{x}} = \text{Tr} \left[ \left( \frac{\mathbf{H}_S^T \mathbf{H}_S + \sigma_e^2 \dot{\mathbf{H}}_S^T \dot{\mathbf{H}}_S}{\sigma^2} + \mathbf{P}_{\mathbf{x}} \right)^{-1} \right], \quad (4.23)$$

$$\text{BCRB}_{\mathbf{e}} = \text{Tr} \left[ \left( \frac{\sigma_x^2 \dot{\mathbf{H}}_S^T \dot{\mathbf{H}}_S}{\sigma^2} + \mathbf{P}_{\mathbf{e}} \right)^{-1} \right]. \quad (4.24)$$

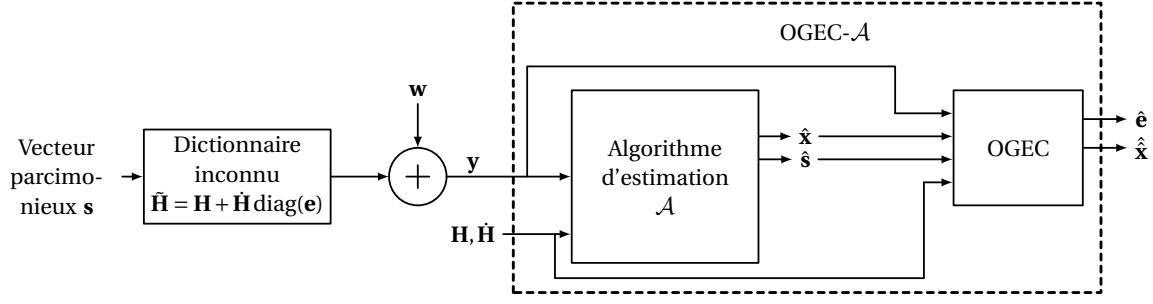
Pour une BCRB totale de  $\text{BCRB}_{\text{struct.}} = \text{BCRB}_{\mathbf{x}} + \text{BCRB}_{\mathbf{e}}$ .

## 4.1.2 Présentation de l'algorithme OGEC

### Présentation de l'algorithme

L'algorithme OGEC est un algorithme à la fois d'estimation de l'erreur de grille et de correction de l'erreur d'amplitude à haut RSB, dans le but d'améliorer la précision de l'estimation par dictionnaire. Le contexte dans lequel l'algorithme se place est celui présenté dans la section 3.4.1 qui correspond au problème de l'estimation par dictionnaire avec un paramètre à une seule dimension. On se base sur la première estimée du vecteur parcimonieux donnée par l'algorithme d'estimation parcimonieuse classique, pour en tirer le support qu'on considère correct et la première estimation entachée d'une erreur de modèle des amplitudes. Notre objectif est d'estimer tout d'abord  $\mathbf{e}$  pour corriger l'erreur de grille et avoir une estimation corrigée du paramètre  $\theta$ . On opère ensuite une ré-estimation de l'amplitude  $\mathbf{x}$  à l'aide du dictionnaire corrigé prenant en compte l'erreur de grille. Le principe de l'estimateur OGEC- $\mathcal{A}$  est résumé par le schéma 4.1. On observe que l'ajout OGEC utilise une première estimation à la fois des amplitudes non nulles  $\hat{\mathbf{x}}$  et du support correspondant obtenu par un estimateur parcimonieux quelconque noté  $\mathcal{A}$ . L'algorithme BiCE que nous présenterons dans la section 4.2.2 pour le problème de dictionnaire bruité n'utilise quant à lui que l'estimée du support, ce qui le rend moins sensible à la qualité du premier estimateur.

On se place dans un contexte à haut RSB de telle sorte qu'on peut faire l'hypothèse que le support du vecteur parcimonieux est correctement estimé. Dans le contexte d'erreur


 FIGURE 4.1 – Description de l'algorithme OGEC- $\mathcal{A}$ .

de grille structurée, on peut modéliser les mesures avec le modèle structuré réduit au support  $\hat{\mathcal{S}}$ , pour simplifier les notations on considère que le support est correctement estimé, et donc que  $\mathcal{S} = \hat{\mathcal{S}}$ . On a

$$\mathbf{y} = \tilde{\mathbf{H}}_{\mathcal{S}} \mathbf{x} + \mathbf{w} = \mathbf{H}_{\mathcal{S}} \mathbf{x} + \dot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\mathbf{e}) \mathbf{x} + \mathbf{w}. \quad (4.25)$$

Pour estimer l'erreur  $\mathbf{e}$  on cherche à projeter ces mesures de telle sorte à conserver uniquement le terme contenant les erreurs de dictionnaire en annulant le premier terme de (4.25). Pour cela on utilise le projecteur oblique  $\mathbf{E}_{\dot{\mathbf{H}}_{\mathcal{S}} \mathbf{H}_{\mathcal{S}}}$  qui permet d'obtenir les mesures projetées :

$$\boldsymbol{\Upsilon} = \mathbf{E}_{\dot{\mathbf{H}}_{\mathcal{S}} \mathbf{H}_{\mathcal{S}}} \mathbf{y} = \mathbf{E}_{\dot{\mathbf{H}}_{\mathcal{S}} \mathbf{H}_{\mathcal{S}}} \mathbf{H}_{\mathcal{S}} \mathbf{x} + \mathbf{E}_{\dot{\mathbf{H}}_{\mathcal{S}} \mathbf{H}_{\mathcal{S}}} \dot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\mathbf{e}) \mathbf{x} + \mathbf{E}_{\dot{\mathbf{H}}_{\mathcal{S}} \mathbf{H}_{\mathcal{S}}} \mathbf{w} \quad (4.26)$$

$$= \dot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\mathbf{x}) \mathbf{e} + \mathbf{E}_{\dot{\mathbf{H}}_{\mathcal{S}} \mathbf{H}_{\mathcal{S}}} \mathbf{w}. \quad (4.27)$$

Cette formulation permet d'obtenir des mesures dépendantes de l'erreur, multipliées par une matrice  $\dot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\mathbf{x})$  et soumises à un bruit additif  $\mathbf{E}_{\dot{\mathbf{H}}_{\mathcal{S}} \mathbf{H}_{\mathcal{S}}} \mathbf{w}$  interprété comme un bruit gaussien projeté de manière oblique. Dans un environnement idéal, connaissant parfaitement  $\mathbf{x}$  et en l'absence de bruit, on peut retrouver le vecteur d'erreur par l'estimation :

$$\mathbf{e} = \text{diag}(\mathbf{x})^{-1} \dot{\mathbf{H}}_{\mathcal{S}}^{\dagger} \boldsymbol{\Upsilon}. \quad (4.28)$$

L'algorithme effectue cette minimisation, en utilisant le vecteur d'amplitude  $\hat{\mathbf{x}}$  précédemment estimé par  $\mathcal{A}$ . Une fois l'erreur estimée, on construit un nouveau dictionnaire corrigé pour ré-estimer les amplitudes. Cette ré-estimation peut exploiter n'importe quel algorithme résolvant le problème linéaire :

$$\mathbf{y} = \tilde{\mathbf{H}}_{\mathcal{S}} \mathbf{x} + \mathbf{w}. \quad (4.29)$$

Dans nos simulations et dans l'algorithme que nous proposons nous choisissons d'utiliser l'estimateur LMMSE qui est l'estimateur linéaire permettant d'obtenir la BEQM minimum, puisqu'il est la version linéaire de l'estimateur optimal du *Minimum Mean Square Estimator* ou en français Estimateur de l'erreur quadratique moyenne minimum (MMSE). Quand la distribution de la variable aléatoire est gaussienne ces deux estimateurs sont égaux [78]. L'estimateur LMMSE pour une distribution  $\mathbf{x}$  de moyenne nulle et de covariance  $\mathbf{R}_x$  et un bruit centré de covariance  $\sigma^2 \mathbf{I}$  est donné en (4.34). L'algorithme complet proposé est décrit dans l'algorithme 1 en détail.

**Données :** les mesures  $\mathbf{y}$ , le dictionnaire  $\mathbf{H}$ , et sa dérivée  $\dot{\mathbf{H}}$

**Résultat :**  $\hat{\mathbf{s}}$

1. Exécuter l'algorithme d'estimation parcimonieuse  $\mathcal{A}$  pour obtenir une première estimation  $\hat{\mathbf{s}}$ ,

$$\mathcal{A}(\mathbf{y}, \mathbf{H}) \rightarrow \hat{\mathbf{s}}. \quad (4.30)$$

2. A partir de l'estimée  $\hat{\mathbf{s}}$  extraire le support estimé  $\hat{\mathcal{S}}$ .
3. Extraire les matrices  $\mathbf{H}_{\hat{\mathcal{S}}}$  et  $\dot{\mathbf{H}}_{\hat{\mathcal{S}}}$  de  $\mathbf{H}$  et  $\dot{\mathbf{H}}$  respectivement.
4. Calculer le projecteur oblique  $\mathbf{E}_{\dot{\mathbf{H}}_{\hat{\mathcal{S}}}\mathbf{H}_{\hat{\mathcal{S}}}}$  suivant l'expression (2.23) :

$$\mathbf{E}_{\dot{\mathbf{H}}_{\hat{\mathcal{S}}}\mathbf{H}_{\hat{\mathcal{S}}}} = \dot{\mathbf{H}}_{\hat{\mathcal{S}}} \left( \dot{\mathbf{H}}_{\hat{\mathcal{S}}}^T \mathbf{P}_{\mathbf{H}_{\hat{\mathcal{S}}}}^{\perp} \dot{\mathbf{H}}_{\hat{\mathcal{S}}} \right)^{-1} \dot{\mathbf{H}}_{\hat{\mathcal{S}}}^T \mathbf{P}_{\mathbf{H}_{\hat{\mathcal{S}}}}^{\perp}. \quad (4.31)$$

5. Extraire  $\hat{\mathbf{x}}$  en gardant les amplitudes non nulles de  $\hat{\mathbf{s}}$ .
6. Utiliser  $\hat{\mathbf{x}}$  pour effectuer l'estimation des erreurs de dictionnaire suivant la formule :

$$\hat{\mathbf{e}} = \text{diag}(\hat{\mathbf{x}})^{-1} \dot{\mathbf{H}}_{\hat{\mathcal{S}}}^{\dagger} \mathbf{E}_{\dot{\mathbf{H}}_{\hat{\mathcal{S}}}\mathbf{H}_{\hat{\mathcal{S}}}} \mathbf{y} \quad (4.32)$$

7. Corriger le dictionnaire :

$$\hat{\hat{\mathbf{H}}}_{\hat{\mathcal{S}}} = \mathbf{H}_{\hat{\mathcal{S}}} + \dot{\mathbf{H}}_{\hat{\mathcal{S}}} \text{diag}(\hat{\mathbf{e}}) \quad (4.33)$$

8. Réévaluer les amplitude  $\hat{\hat{\mathbf{x}}}_{\hat{\mathcal{S}}}$ , en utilisant par exemple un estimateur LMMSE :

$$\hat{\hat{\mathbf{x}}} = \text{hat} \tilde{\mathbf{H}}_{\hat{\mathcal{S}}}^T \left( \hat{\hat{\mathbf{H}}}_{\hat{\mathcal{S}}} \hat{\hat{\mathbf{H}}}_{\hat{\mathcal{S}}}^T + \sigma^2 \mathbf{R}_x^{-1} \right)^{-1} \mathbf{y} \quad (4.34)$$

9. Remplacer les amplitudes non nulles de  $\hat{\mathbf{s}}$  par les estimées  $\hat{\hat{\mathbf{x}}}_{\hat{\mathcal{S}}}$  pour obtenir le vecteur parcimonieux estimé  $\hat{\hat{\mathbf{s}}}$

**Algorithme 1 :** Algorithme OGEC- $\mathcal{A}$

### 4.1.3 Performances théoriques de l'algorithme

#### Erreur quadratique moyenne et biais

Nous avons donné dans la section 4.1.1 la borne sur l'estimation des amplitudes et des erreurs de grille minimisant l'erreur des algorithmes cherchant à estimer ces quantités dans un contexte structuré tel que l'OGEC. Dans cette section nous allons donner les performances théoriques de l'algorithme sur l'estimation de l'erreur de grille. L'erreur sur l'estimation des amplitudes dépend du type d'estimateur choisi pour faire la seconde estimation des amplitudes, les performances de l'algorithme correspondant aux performances du LMMSE dans notre cas, mais celles-ci sont trop optimistes puisqu'en réalité on fait le LMMSE avec un dictionnaire erroné. Dans les applications favorisant l'apparition d'un bruit sur le dictionnaire telle que l'acquisition comprimée, l'estimation de l'amplitude est primordiale. A contrario, les problèmes d'estimation par dictionnaire ont pour objectif de retrouver l'indice ayant servi à générer le dictionnaire et se focalisent pour cette raison sur l'estimation de l'erreur de grille. Pour ces deux raisons nous nous concentrons ici seulement sur l'estimation de l'erreur de grille, qui est au coeur de l'algorithme que nous proposons. Nous nous concentrons sur les performances de l'algorithme OGEC uniquement, c'est pourquoi nous considérons que l'on connaît une première estimée du support  $\hat{\mathcal{S}}$  et des amplitudes non nulles correspondantes  $\hat{\mathbf{x}}$ . Les performances ainsi obtenues sont donc conditionnées à des amplitudes  $\hat{\mathbf{x}}$  données. Il n'est pas possible de donner une formulation générale de la BEQM, étant donné que celle-ci dépend de l'estimation  $\hat{\mathbf{x}}$  et donc des performances de l'algorithme  $\mathcal{A}$  choisi. Si l'on connaît les statistiques concernant  $\hat{\mathbf{x}}$ , l'expression générale de la BEQM peut être donnée par :

$$\mathbb{E}_{\hat{\mathbf{x}}} [\text{BEQM}(\hat{\mathbf{e}}|\hat{\mathbf{x}})], \quad (4.35)$$

où la BEQM conditionnelle sur l'estimation de l'erreur est :

$$\text{BEQM}(\hat{\mathbf{e}}|\hat{\mathbf{x}}) = \mathbb{E}_{\mathbf{y}, \mathbf{x}, \mathbf{e}|\hat{\mathbf{x}}} \left[ \|\hat{\mathbf{e}}(\mathbf{y}, \hat{\mathbf{x}}) - \mathbf{e}\|^2 \right]. \quad (4.36)$$

L'expression générale ne peut pas être donnée de manière analytique, puisqu'elle dépend du choix de l'estimateur  $\mathcal{A}$ , cependant on peut l'obtenir par des simulations de Monte-Carlo.

On suppose une estimation exacte du support, et l'on pose donc que  $\hat{\mathcal{S}} = \mathcal{S}$ . L'algorithme OGEC estime l'erreur de grille  $\mathbf{e}$  par l'expression donnée en (4.32). Les mesures  $\mathbf{y}$  pouvant s'écrire sous la forme (4.25) on peut introduire cette expression dans le calcul de l'estimée (4.32) pour obtenir :

$$\hat{\mathbf{e}}(\mathbf{y}, \hat{\mathbf{x}}) - \mathbf{e} = \text{diag}(\hat{\mathbf{x}})^{-1} \dot{\mathbf{H}}_{\mathcal{S}}^{\dagger} \mathbf{E}_{\dot{\mathbf{H}}_{\mathcal{S}} \mathbf{H}_{\mathcal{S}}} (\mathbf{H}_{\mathcal{S}} \mathbf{x} + \dot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\mathbf{x}) \mathbf{e} + \mathbf{w}) - \mathbf{e} \quad (4.37)$$

$$= (\text{diag}(\hat{\mathbf{x}})^{-1} \text{diag}(\mathbf{x}) - \mathbf{I}) \mathbf{e} + \text{diag}(\hat{\mathbf{x}})^{-1} \dot{\mathbf{H}}_{\mathcal{S}}^{\dagger} \mathbf{E}_{\dot{\mathbf{H}}_{\mathcal{S}} \mathbf{H}_{\mathcal{S}}} \mathbf{w} \quad (4.38)$$

Le biais conditionnel de l'estimateur est donc :

$$\mathbf{b}(\hat{\mathbf{e}}|\hat{\mathbf{x}}) = \mathbb{E}_{\mathbf{y}|\mathbf{x}, \hat{\mathbf{x}}, \mathbf{e}} [\hat{\mathbf{e}}(\mathbf{y}, \hat{\mathbf{x}}) - \mathbf{e}] = (\text{diag}(\hat{\mathbf{x}})^{-1} \text{diag}(\mathbf{x}) - \mathbf{I}) \mathbf{e} \quad (4.39)$$

pour un bruit centré. Notons que si l'estimation du premier algorithme est parfaite, on a  $\mathbf{x} = \hat{\mathbf{x}}$  et le biais de l'étape de l'algorithme estimant les erreurs de grille est nul. OGEC permet de corriger les erreurs, mais ce n'est qu'une étape de correction qui est dépendante de la qualité de l'estimation faite par l'algorithme  $\mathcal{A}$ . On verra dans la prochaine section

que l'algorithme BiCE quant à lui n'utilise pas du tout la première estimation de l'amplitude, ce qui permet de choisir d'obtenir le même résultat quel que soit l'algorithme de reconstruction parcimonieuse choisi.

La BEQM quant à elle est :

$$\begin{aligned} \text{BEQM}_{\text{OGEC}}(\hat{\mathbf{e}}|\hat{\mathbf{x}}) &= \mathbb{E} \left[ \text{Tr} \left[ \left( \text{diag}(\hat{\mathbf{x}})^{-1} \text{diag}(\mathbf{x}) - \mathbf{I} \right)^2 \mathbf{e} \mathbf{e}^T \right] \right] \\ &\quad + \mathbb{E} \left[ \text{Tr} \left[ \mathbf{w} \mathbf{w}^T \text{diag}(\hat{\mathbf{x}})^{-1} \dot{\mathbf{H}}_{\mathcal{S}}^T \mathbf{E}_{\dot{\mathbf{H}}_{\mathcal{S}} \mathbf{H}_{\mathcal{S}}} \text{diag}(\hat{\mathbf{x}})^{-1} \mathbf{E}_{\dot{\mathbf{H}}_{\mathcal{S}} \mathbf{H}_{\mathcal{S}}} \dot{\mathbf{H}}_{\mathcal{S}}^{\dagger T} \right] \right] \\ &= \text{Tr} \left[ \mathbb{E}_{\mathbf{x}} \left[ \left( \text{diag}(\hat{\mathbf{x}})^{-1} \text{diag}(\mathbf{x}) - \mathbf{I} \right)^2 \right] \boldsymbol{\Sigma}_e \right] + L \sigma_e^2 \text{Tr} \left[ \text{diag}(\hat{\mathbf{x}})^{-2} \left( \dot{\mathbf{H}}_{\mathcal{S}}^T \mathbf{P}_{\dot{\mathbf{H}}_{\mathcal{S}}}^{\perp} \dot{\mathbf{H}}_{\mathcal{S}} \right)^{-1} \right] \end{aligned}$$

où l'on a fait l'hypothèse que le bruit et l'erreur de grille sont indépendants et  $\boldsymbol{\Sigma}_e = \mathbb{E}(\mathbf{e} \mathbf{e}^T)$  est la matrice de covariance de l'erreur. On a que

$$\left[ \left( \text{diag}(\hat{\mathbf{x}})^{-1} \text{diag}(\mathbf{x}) - \mathbf{I} \right)^2 \right]_{ll} = \frac{(s_l - \hat{s}_l)^2}{\hat{s}_l^2} \quad (4.40)$$

De plus, pour une erreur de distribution gaussienne centrée de covariance  $\boldsymbol{\Sigma}_e = \frac{\sigma_e^2}{L} \mathbf{I}$  on peut donner le résultat qui suit.

**Résultat 22.** *Pour une erreur de matrice de covariance  $\frac{\sigma_e^2}{L} \mathbf{I}$ , un bruit centré de matrice de covariance  $\sigma_e^2 \mathbf{I}$  indépendant, la BEQM de l'algorithme OGEC, conditionnée au résultat de l'algorithme  $\mathcal{A}$  est, quand le support est parfaitement estimé :*

$$\text{BEQM}_{\text{OGEC}}(\hat{\mathbf{e}}|\hat{\mathbf{x}}) = \sigma_e^2 \sum_{l=1}^L \frac{\text{BSE}_{\mathcal{A}}(\hat{s}_l)}{\hat{s}_l^2} + \sigma_e^2 \text{Tr} \left[ \text{diag}(\hat{\mathbf{x}})^{-2} \left( \dot{\mathbf{H}}_{\mathcal{S}}^T \mathbf{P}_{\dot{\mathbf{H}}_{\mathcal{S}}}^{\perp} \dot{\mathbf{H}}_{\mathcal{S}} \right)^{-1} \right] \quad (4.41)$$

où  $L$  est le nombre d'amplitudes non nulles, et  $\text{BSE}_{\mathcal{A}}$  est l'erreur quadratique moyenne de l'estimateur  $\mathcal{A}$  sur la  $l^e$  amplitude définie par :

$$\text{BSE}_{\mathcal{A}}(\hat{s}_l) = \mathbb{E}_{\mathbf{x}} \left[ (s_l - \hat{s}_l)^2 \right]. \quad (4.42)$$

Les performances de l'algorithme OGEC, même sur l'estimation de l'erreur de grille dépendent de la qualité d'estimation de l'algorithme que l'on cherche à corriger. En plus de l'estimation correcte du support, la qualité d'estimation de l'amplitude joue également un rôle important. La borne ainsi obtenue ne peut donc pas être comparée de manière analytique à la borne donnée en (4.24). Elle peut néanmoins être calculée numériquement pour un exemple d'estimateur donné, comme nous le verrons dans la partie simulation.

L'erreur quadratique moyenne du LMMSE est bien connue [78] :

$$\text{BEQM}(\hat{\mathbf{x}}|\hat{\mathbf{H}}) = \text{Tr} \left[ \mathbf{R}_x - \mathbf{R}_x \hat{\mathbf{H}}_{\mathcal{S}}^T \left( \hat{\mathbf{H}}_{\mathcal{S}} \mathbf{R}_x \hat{\mathbf{H}}_{\mathcal{S}}^T + \sigma^2 \mathbf{I} \right)^{-1} \hat{\mathbf{H}}_{\mathcal{S}} \mathbf{R}_x \right]. \quad (4.43)$$

La BEQM sur l'estimation de l'amplitude dépend donc de la performance d'estimation de l'erreur, qui dépend elle-même de la performance d'estimation de l'algorithme  $\mathcal{A}$ . La BEQM conditionnée à l'estimation de l'erreur est :

$$\begin{aligned} \text{BEQM}_{\text{OGEC}}(\hat{\mathbf{x}}|\hat{\mathbf{e}}) &= \text{Tr} \left[ \mathbf{R}_x - \mathbf{R}_x (\mathbf{H}_{\mathcal{S}} + \dot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\hat{\mathbf{e}}))^T \right. \\ &\quad \cdot \left. \left( (\mathbf{H}_{\mathcal{S}} + \dot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\hat{\mathbf{e}})) \mathbf{R}_x (\mathbf{H}_{\mathcal{S}} + \dot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\hat{\mathbf{e}}))^T + \sigma^2 \mathbf{I} \right)^{-1} (\mathbf{H}_{\mathcal{S}} + \dot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\hat{\mathbf{e}})) \mathbf{R}_x \right]. \end{aligned} \quad (4.44)$$

Pour obtenir la BEQM non conditionnelle, il faut moyenner suivant  $\hat{\mathbf{e}}$  :

$$\text{BEQM}_{\text{OGEC}}(\hat{\mathbf{x}}) = \mathbb{E}_{\hat{\mathbf{e}}} [\text{BEQM}_{\text{OGEC}}(\hat{\mathbf{x}}|\hat{\mathbf{e}})]. \quad (4.45)$$

Il n'est pas possible de donner une expression analytique, cependant on peut en obtenir une approximation en moyennant la BEQM conditionnelle par simulation de Monte-Carlo.

### Efficacité statistique

L'estimation de l'erreur de grille est une estimation biaisée, c'est pourquoi on s'intéresse à la borne BCRM biaisée, les bornes de Cramer Rao classique n'étant valables que pour des estimateurs non biaisés. La borne déterministe pour un estimateur biaisé est donné en (1.62), la BCRM biaisée consistant à faire la moyenne de la borne déterministe. L'algorithme OGEC effectue une projection des mesures avant d'effectuer l'estimation de l'erreur, c'est pourquoi on s'intéresse au modèle des mesures projetées sur l'espace orthogonal de  $\mathbf{H}_S$  :  $\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}$  [29]. La borne moyennée sur les mesures projetées est définie par :

$$\mathcal{C}_{S|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}} = \text{Tr} \left[ \mathbb{E}_{\mathbf{e}} \left( \left( \mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{e})}{\partial \mathbf{e}} \right) \mathbf{F}(\mathbf{e}|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y})^{-1} \left( \mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{e})}{\partial \mathbf{e}} \right)^T \right) \right] + \mathbb{E}_{\mathbf{e}} (\|\mathbf{b}(\mathbf{e})\|^2), \quad (4.46)$$

où l'on a simplement fait la moyenne de la borne déterministe projetée définie dans l'équation (1.62). Le biais  $\mathbf{b}(\mathbf{e})$  est le biais de l'estimateur, son expression conditionnée à l'estimation de  $\hat{\mathbf{x}}$  est donnée en (4.39). Les mesures projetées sont définies par :

$$\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y} = \mathbf{P}_{\mathbf{H}_S}^\perp \dot{\mathbf{H}}_S \text{diag}(\mathbf{x}) \mathbf{e} + \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{w}. \quad (4.47)$$

Ces mesures conditionnées à une réalisation de  $\mathbf{e}$  suivent donc la distribution suivante

$$\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}|\mathbf{e} \sim \mathcal{N} \left( \mathbf{P}_{\mathbf{H}_S}^\perp \dot{\mathbf{H}}_S \text{diag}(\mathbf{x}) \mathbf{e}, \sigma^2 \mathbf{P}_{\mathbf{H}_S}^\perp \right), \quad (4.48)$$

ce qui permet en utilisant la formule de Slepian-Bang (1.60) d'obtenir la MIF :

$$\mathbf{F}(\mathbf{e}|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}) = \frac{1}{\sigma^2} \text{diag}(\mathbf{x})^2 \dot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \dot{\mathbf{H}}_S. \quad (4.49)$$

De l'expression (4.39) donnant l'erreur sur l'erreur de grille, on peut donner l'espérance du biais au carré :

$$\mathbb{E}_{\mathbf{e}, \mathbf{x}|\hat{\mathbf{x}}} (\|\mathbf{b}(\mathbf{e})\|^2) = \mathbb{E}_{\mathbf{e}, \mathbf{x}|\hat{\mathbf{x}}} \left( \text{Tr} \left[ (\text{diag}(\hat{\mathbf{x}})^{-1} \text{diag}(\mathbf{x}) - \mathbf{I})^2 \mathbf{e} \mathbf{e}^T \right] \right) \quad (4.50)$$

$$= \sigma_e^2 \sum_{l=1}^L \frac{\text{BSE}_{\mathcal{A}}(\hat{s}_l)}{\hat{s}_l^2}, \quad (4.51)$$

et la dérivée du biais :

$$\frac{\partial \mathbf{b}(\mathbf{e})}{\partial \mathbf{e}} = \text{diag}(\hat{\mathbf{x}})^{-1} \text{diag}(\mathbf{x}) - \mathbf{I}. \quad (4.52)$$

En introduisant (4.49), (4.51) et (4.52) dans (4.46), on trouve :

$$\mathcal{C}_{S|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}} = \sigma^2 \text{Tr} \left[ \mathbb{E}_{\mathbf{e}} \left( \text{diag}(\hat{\mathbf{x}})^{-2} \text{diag}(\mathbf{x})^2 \text{diag}(\mathbf{x})^{-2} \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \dot{\mathbf{H}}_S \right)^{-1} \right) \right] + \sigma_e^2 \sum_{l=1}^L \frac{\text{BSE}_{\mathcal{A}}(\hat{s}_l)}{\hat{s}_l^2} \quad (4.53)$$

$$= \sigma^2 \text{Tr} \left[ \text{diag}(\hat{\mathbf{x}})^{-2} \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \dot{\mathbf{H}}_S \right)^{-1} \right] + \sigma_e^2 \sum_{l=1}^L \frac{\text{BSE}_{\mathcal{A}}(\hat{s}_l)}{\hat{s}_l^2}. \quad (4.54)$$

**Résultat 23.** *Le post traitement OGEC est statistiquement efficace au sens de la BCRM biaisée sur les mesures projetées  $\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}$  pour l'estimation des erreurs de grille. En effet on a*

$$\text{BEQM}_{\text{OGEC}}(\hat{\mathbf{e}}|\hat{\mathbf{x}}) = \mathcal{C}_{S|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}} \quad (4.55)$$

*conditionnellement à l'estimation correcte du support et la connaissance de  $\hat{\mathbf{x}}$*

La matrice d'information de Fisher correspondant au modèle non projeté dépend des deux paramètres à estimer  $\mathbf{x}$  et  $\mathbf{e}$ . En effet on a :

$$\mathbf{y}|\mathbf{x}, \mathbf{e} \sim \mathcal{N}(\mathbf{H}_S \mathbf{x} + \dot{\mathbf{H}}_S \text{diag}(\mathbf{e})\mathbf{x}, \sigma^2 \mathbf{I}). \quad (4.56)$$

Les dérivées associées à ce modèle ont été données dans (4.11) et (4.12), et permettent d'obtenir la MIF :

$$\mathbf{F}([\mathbf{x}, \mathbf{e}]|\mathbf{y}) = \frac{1}{\sigma^2} \begin{bmatrix} \tilde{\mathbf{H}}_S^T \tilde{\mathbf{H}}_S & \dot{\mathbf{H}}_S^T \tilde{\mathbf{H}}_S \text{diag}(\mathbf{x}) \\ \tilde{\mathbf{H}}_S^T \dot{\mathbf{H}}_S \text{diag}(\mathbf{x}) & \text{diag}(\mathbf{x}) \dot{\mathbf{H}}_S^T \dot{\mathbf{H}}_S \text{diag}(\mathbf{x}) \end{bmatrix}, \quad (4.57)$$

où  $\tilde{\mathbf{H}} = \mathbf{H} + \dot{\mathbf{H}} \text{diag}(\mathbf{e})$ . Si l'on s'intéresse uniquement à la borne sur l'erreur, seul le bloc concernant le vecteur  $\mathbf{e}$  nous intéresse. D'après la formule de l'inverse d'une matrice par bloc (2.14) on a :

$$\mathbf{F}(\mathbf{e}|\mathbf{y})^{-1} = \sigma^2 \left( \text{diag}(\mathbf{x})^{-2} (\dot{\mathbf{H}}_S^T \dot{\mathbf{H}}_S)^{-1} - \text{diag}(\mathbf{x})^2 \tilde{\mathbf{H}}_S^T \dot{\mathbf{H}}_S (\tilde{\mathbf{H}}_S^T \tilde{\mathbf{H}}_S)^{-1} \dot{\mathbf{H}}_S^T \tilde{\mathbf{H}}_S \right). \quad (4.58)$$

La BCRM sur les mesures non projetée est donc :

$$\mathcal{C}_{S|\mathbf{y}} = \text{Tr} \left[ \mathbb{E}_{\mathbf{e}} \left( \text{diag}(\hat{\mathbf{x}})^{-2} \text{diag}(\mathbf{x})^2 \mathbf{F}(\mathbf{e}|\mathbf{y})^{-1} \right) \right] + \sigma_e^2 \sum_{l=1}^L \frac{\text{BSE}_{\mathcal{A}}(\hat{s}_l)}{\hat{s}_l^2} \quad (4.59)$$

$$\begin{aligned} &= \sigma^2 \text{Tr} \left[ \text{diag}(\hat{\mathbf{x}})^{-2} (\dot{\mathbf{H}}_S^T \dot{\mathbf{H}}_S)^{-1} - \text{diag}(\hat{\mathbf{x}})^{-2} \text{diag}(\mathbf{x})^4 \mathbb{E}_{\mathbf{e}} \left( \tilde{\mathbf{H}}_S^T \dot{\mathbf{H}}_S (\tilde{\mathbf{H}}_S^T \tilde{\mathbf{H}}_S)^{-1} \dot{\mathbf{H}}_S^T \tilde{\mathbf{H}}_S \right) \right] \\ &\quad + \sigma_e^2 \sum_{l=1}^L \frac{\text{BSE}_{\mathcal{A}}(\hat{s}_l)}{\hat{s}_l^2}. \end{aligned} \quad (4.60)$$

On ne peut pas obtenir une expression analytique de l'espérance puisque la matrice  $\tilde{\mathbf{H}}_S$  dépend du paramètre à estimer  $\mathbf{e}$ . On doit donc passer par la simulation pour étudier son comportement et le comparer à celui de la borne sur les mesures projetées.

#### 4.1.4 Résultat des simulations

##### Contexte applicatif

Pour illustrer le comportement des deux algorithmes OGEC et BiCE que nous présenterons dans la section suivante, nous nous plaçons dans le contexte de l'acquisition comprimée des signaux à modèles paramétriques, présenté dans la section 2.4.2. Nous considérons le cas d'école du signal à somme de Diracs, avec pour noyau d'échantillonnage un noyau en sinus cardinal tel que  $g(t) = \frac{1}{T_S} \text{sinc}(t)$  où  $T_S$  est l'intervalle d'échantillonnage utilisé. Connaissant ce noyau et la forme de l'impulsion, on utilise l'expression

(2.164) pour obtenir la base utilisée pour reconstruire le signal  $\Phi$ , ainsi que sa dérivée et sa dérivée seconde :

$$[\Phi]_{kk'} = \frac{1}{T_S} \text{sinc}((k' - k)T'_S), \quad (4.61)$$

$$[\dot{\Phi}]_{kk'} = \begin{cases} 0 & \text{pour } k = k' \\ \frac{1}{T_S((k' - k)T'_S)} (\cos(\pi((k' - k)T'_S)) - \text{sinc}((k' - k)T'_S)) & \text{ailleurs,} \end{cases} \quad (4.62)$$

$$[\ddot{\Phi}]_{kk'} = \begin{cases} 0 & \text{pour } k = k' \\ \left( \frac{2}{T_S((k' - k)T'_S)^2} + \frac{\pi^2}{T_S} \right) \text{sinc}((k' - k)T'_S) - \frac{2}{T_S} \frac{\cos(\pi((k' - k)T'_S))}{((k' - k)T'_S)^2} & \text{ailleurs.} \end{cases} \quad (4.63)$$

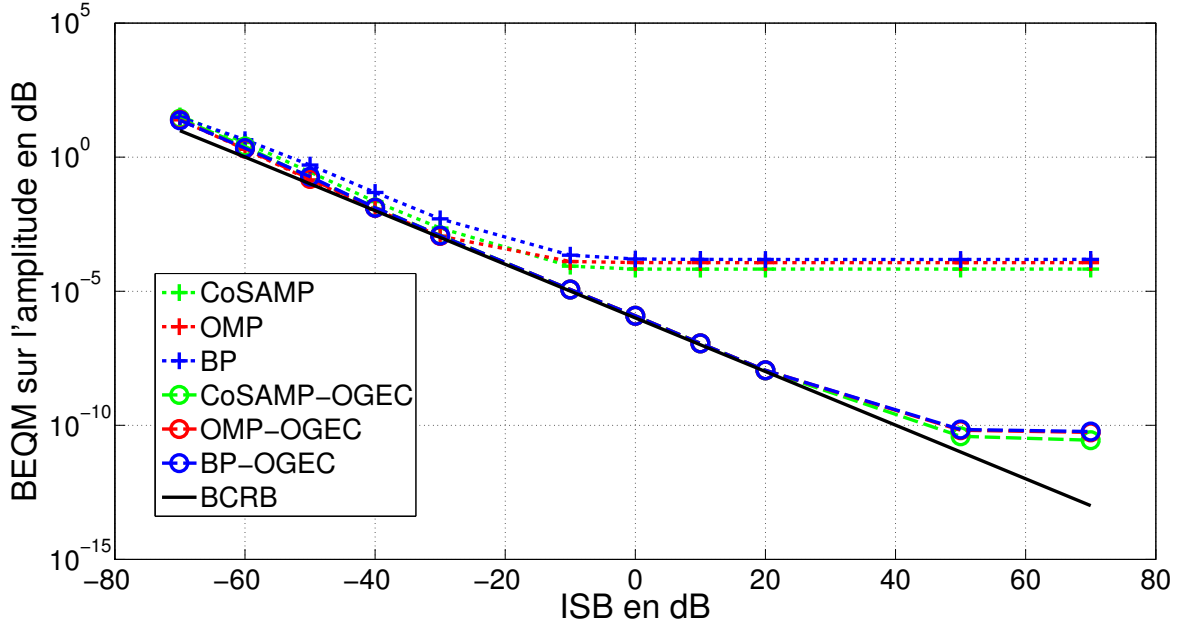
où  $T'_S$  est l'intervalle d'échantillonnage de la grille et  $k$  et  $k'$  sont deux entiers allant de 1 à  $K$  étant donné que la base est carrée. Pour simplifier les calculs on prendra  $T'_S = 1$  dans les simulations. La matrice de mesure  $\Psi$  quant à elle est choisie aléatoire, chaque élément de la matrice est tiré d'une distribution gaussienne centrée de variance  $1/N$ . Choisir une telle distribution permet de se prémunir contre les problèmes de cohérence, car cette matrice suit les conditions RIP [7].

## Résultats

Pour obtenir les résultats de l'algorithme OGEC, nous utilisons trois algorithmes différents d'estimation parcimonieuse, les mêmes que dans la section 3.4.4 (OMP, *Compressive Sensing Approximate Message Passing* (COSAMP) et SPGL1). Nous considérons le cas où le nombre d'inconnu  $L$  est fixé. De la même façon que dans la section 3.4.4, pour que l'algorithme SPGL1 soit dans les mêmes conditions que l'OMP et le COSAMP, nous gardons les  $L$  plus grandes amplitudes estimées par l'algorithme. Le résultat de ces algorithmes est ensuite utilisé pour faire une estimation de l'erreur sur le support, puis de l'amplitude avec l'algorithme LMMSE. Nous utilisons les paramètres  $N = 100$ ,  $L = 10$  et  $K = 300$  et faisons 100 réalisations pour calculer les espérances. La contrainte principale dans le choix des paramètres est de garantir l'estimation correcte du support à haut RSB. Nous prenons pour cela une variance de l'erreur  $\sigma_e^2 = 10^{-5}$ . Cette erreur peut sembler faible, mais elle est nécessaire pour rester dans les conditions où le support est correctement estimé, sans quoi les résultats ne sont pas pertinents puisqu'ils ne correspondent plus seulement à l'erreur de modèle. Avec une erreur trop importante, le support est mal estimé et l'algorithme OGEC n'opère pas de correction. Les courbes sont tracées en fonction du Rapport Interférence sur Bruit (ISB), que nous définissons par :  $\text{ISB} = \sigma_e^2 / \sigma^2$ .

La figure 4.2 montre le résultat de l'algorithme pour l'estimation des amplitudes. Le résultat des différents algorithmes sans correction OGEC est indiqué par une croix, tandis que la BEQM sur l'amplitude après correction LMMSE est indiqué par des ronds. On observe que quand le bruit diminue, les estimateurs atteignent assez rapidement un palier sur l'amplitude (pour un ISB de -10dB) tandis que la correction effectuée permet d'améliorer nettement les performances, qui atteignent cependant un palier vers 50dB d'ISB. La borne sur l'amplitude (4.23), tracée en noir, diminue avec le bruit jusqu'à atteindre une valeur infiniment petite. On peut observer que tant que le palier de la correction OGEC n'est pas atteinte, l'estimation corrigée des amplitudes permet d'atteindre la borne, et permet donc de corriger efficacement l'erreur, sur la plage de bruit d'intérêt.

Nous étudions à présent le problème de l'estimation de l'erreur de grille, dont les résultats sont visualisés sur la figure 4.3. Sont tracés à la fois le résultat de l'estimation

FIGURE 4.2 – Résultat de l'algorithme  $\mathcal{A}$ -OGEC sur l'estimation de l'amplitude.

sans aucune correction d'erreur marquée par une croix, l'erreur moyenne sur l'estimation des retards une fois la correction OGEC effectuée marquée par un rond et enfin la borne BEQM (4.41) pour chaque algorithme en pointillé et marquée par une étoile. La borne ne peut pas être directement tracée puisqu'elle n'a pas une forme analytique et dépend de l'erreur d'estimation moyenne de l'algorithme  $\mathcal{A}$ . Pour obtenir son expression on procède par une méthode de Monte-Carlo, c'est à dire qu'on fait une moyenne de cette erreur, et du résultat  $\hat{x}$  sur 100 réalisations, de cette façon on obtient une approximation de (4.41). En dernier lieu la borne BCRB structurée sur l'erreur donnée en (4.23) est tracée en ligne continue noire. Notons que les courbes de la BEQM correspondent également à la borne BCRM biaisée projetée puisque nous avons vu que ces deux mesures sont égales. Tout comme pour les amplitudes, l'algorithme OGEC permet de grandement améliorer les performances d'estimation sur les retards. En effet, sans correction, l'erreur obtenue correspond à l'erreur de discrétisation de la grille à haut RSB qui produit un palier constant dès un ISB de -10dB. Ce palier étant le fruit de la discrétisation de la grille, il est exactement le même quel que soit l'algorithme utilisé pour l'estimation. Quand le bruit est assez faible pour que l'erreur atteigne le palier de discrétisation, l'utilisation de l'algorithme OGEC améliore les performances et permet d'estimer convenablement l'erreur. Tout comme pour l'amplitude, cette estimation a ses limites à très haut RSB puisque l'erreur d'estimation finit par atteindre également un palier à partir d'un ISB de 50dB. On remarque cependant, que bien que l'estimation des retards soit grandement améliorée, les performances de l'algorithme OGEC n'atteignent pas la borne (4.24), ce qui confirme nos résultats concernant l'efficacité de l'algorithme, qui n'a été démontrée que pour le modèle d'erreur projetée. Les résultats pratiques de l'algorithme suivent approximativement la BEQM théorique, dans la région de bruit d'intérêt. En effet, pour des bruits plus importants, le calcul de la BEQM est trop optimiste puisqu'il part du principe que les supports sont correctement estimés ce qui n'est pas le cas dans cette région de bruit. Par la suite, les résultats de l'estimation OGEC suivent la BEQM, puis à haut ISB la BEQM est un peu plus optimiste que les résultats réels de l'algorithme. Les différences de résultats

peuvent être expliquées par la façon de calculer la BEQM, qui ne correspond pas tout à fait au calcul d'espérance exact, ces différences ayant plus de poids pour des erreurs très faibles.

En conclusion, on peut noter que l'algorithme OGEC atteint son objectif d'améliorer l'estimation à la fois de l'erreur et de l'amplitude sous la condition que le support est correctement estimé. Cependant l'algorithme n'atteint pas la BCRB, ce qui montre qu'une amélioration est encore possible. De plus quand le support est mal estimé, le post traitement OGEC obtient le même résultat que les estimateurs non corrigés, ce qui permet de l'utiliser dans toutes les situations puisque dans un cas il ne dégrade pas les performances, alors qu'à haut RSB il les améliore.

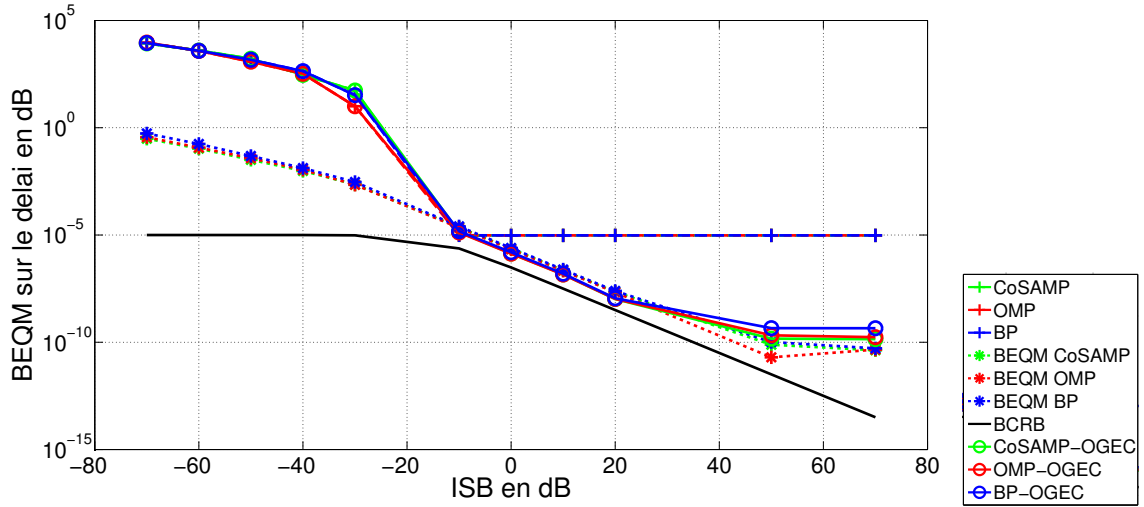


FIGURE 4.3 – Résultat de l'algorithme  $\mathcal{A}$ -OGEC sur l'estimation de l'erreur.

## 4.2 Correction de l'erreur de dictionnaire bruité par l'algorithme BiCE

Nous présentons à présent l'algorithme BiCE qui se concentre sur le problème de l'erreur de dictionnaire bruité. Cet algorithme a été présenté dans les articles [19] et [25].

### 4.2.1 Modèle structuré dans le contexte de dictionnaire bruité

L'erreur additive structurée apparaît immédiatement lorsque l'on considère le problème de l'estimation parcimonieuse, où le dictionnaire est toujours construit à partir de la discrétisation d'une fonction continue. Cependant dans le cadre de l'acquisition comprimée également, il est courant que la base  $\Phi$  introduite dans l'équation (1.28) et représentant la base dans laquelle le vecteur est parcimonieux, soit une discrétisation d'un noyau continu connu  $\phi(t, \theta)$ . Prenons le cas où les échantillons  $\mathbf{y}$  sont obtenus à partir de la discrétisation  $\tilde{\boldsymbol{\theta}} = \{\tilde{\theta}_1, \dots, \tilde{\theta}_K\}$ , et que l'algorithme d'estimation à la réception connaît la fonction  $\phi(t, \theta)$  mais qu'une erreur s'est ajoutée au moment de la transmission de l'information sur le vecteur  $\tilde{\boldsymbol{\theta}}$ . L'algorithme utilise alors une version erronée  $\tilde{\tilde{\boldsymbol{\theta}}} = \{\tilde{\tilde{\theta}}_1, \dots, \tilde{\tilde{\theta}}_K\}$  pour

discrétiser  $\phi(t, \theta)$  et donc une base  $\tilde{\Phi}$  erronée. Si l'on pose que  $\tilde{\theta}_k = \bar{\theta}_k + e_k$ , la base erronée peut alors être écrite sous la forme d'un développement de Taylor pour une erreur  $e_k$  exactement de la même façon que dans le problème de l'estimation par dictionnaire. Le vecteur composant la  $k$ -ième colonne de  $\tilde{\Phi}$  peut s'écrire :

$$[\tilde{\Phi}]_k = \Phi(\tilde{\theta}_k) = \Phi(\bar{\theta}_k) + \left. \frac{\partial \Phi(\theta)}{\partial \theta} \right|_{\theta=\bar{\theta}_k} e_k + O(e_k^2). \quad (4.64)$$

Dans ce contexte, pour une erreur suffisamment faible, l'algorithme d'estimation dispose d'un modèle de la base  $\tilde{\Phi}$  qui peut être approximé par :

$$\tilde{\Phi} \approx \Phi + \dot{\Phi} \text{diag}(\mathbf{e}), \quad (4.65)$$

où

$$[\dot{\Phi}]_k = \left. \frac{\partial \Phi(\theta)}{\partial \theta} \right|_{\theta=\bar{\theta}_k}, \quad (4.66)$$

et  $\mathbf{e} = [e_1, \dots, e_K]^T$ . Le dictionnaire est obtenu en multipliant la base par la matrice de mesure que l'on suppose connue, le dictionnaire utilisé alors par l'algorithme d'estimation est :

$$\tilde{\mathbf{H}} = \Psi \tilde{\Phi} \approx \Psi \Phi + \Psi \dot{\Phi} \text{diag}(\mathbf{e}) = \mathbf{H} + \dot{\mathbf{H}} \text{diag}(\mathbf{e}). \quad (4.67)$$

Pour développer l'algorithme BiCE, on considère que l'on connaît  $\tilde{\theta}$ , la fonction  $\phi(t, \theta)$  ainsi que la matrice de mesure  $\Psi$ . A partir de ces paramètres, on peut générer aussi bien  $\tilde{\mathbf{H}}$  que  $\dot{\tilde{\mathbf{H}}}$  qui sont donc connus également. La matrice  $\dot{\tilde{\mathbf{H}}}$  est générée de la même façon que  $\dot{\mathbf{H}}$  mais avec le vecteur de paramètres  $\tilde{\theta}$ .

## 4.2.2 Présentation de l'algorithme BiCE

L'algorithme BiCE a pour objectif d'améliorer les résultats de n'importe quel estimateur parcimonieux utilisant un dictionnaire erroné pour l'estimation d'un vecteur parcimonieux  $\mathbf{s}$ . Il est conçu comme une étape simple de post-traitement, se basant sur un estimateur  $\mathcal{A}$  au choix que l'on ne souhaite pas modifier. Cette approche a donc un objectif différent des algorithmes dans la littérature cherchant à proposer des estimateurs parcimonieux directement robustes à l'erreur de dictionnaire, au coût d'une complexification des algorithmes pouvant se répercuter sur le temps de calcul. L'algorithme BiCE de son côté est simple et peu coûteux en temps de calcul, et a pour objectif de corriger les estimateurs parcimonieux en présence d'un bruit structuré sur le dictionnaire.

### Présentation de l'algorithme

La procédure complète depuis l'acquisition jusqu'à la correction de l'erreur est schématisée sur la figure 4.4. Le vecteur parcimonieux que l'on cherche à estimer est multiplié par le dictionnaire  $\mathbf{H}$  qui permet d'obtenir les échantillons  $\mathbf{y}$ . L'algorithme d'estimation  $\mathcal{A}$  connaît ces échantillons  $\mathbf{y}$ , ainsi que le dictionnaire erroné  $\tilde{\mathbf{H}}$  et sa dérivée  $\dot{\tilde{\mathbf{H}}}$ . Selon la méthode d'estimation de l'algorithme  $\mathcal{A}$ , la connaissance de l'ordre du modèle  $L$  peut également être supposée (comme par exemple pour l'algorithme OMP). Le vecteur  $\hat{\mathbf{s}}$  estimé est ensuite utilisé par l'algorithme BiCE qui ré-estime les amplitudes correspondant aux valeurs non nulles de  $\hat{\mathbf{s}}$  afin d'en corriger l'influence de l'erreur sur le dictionnaire.

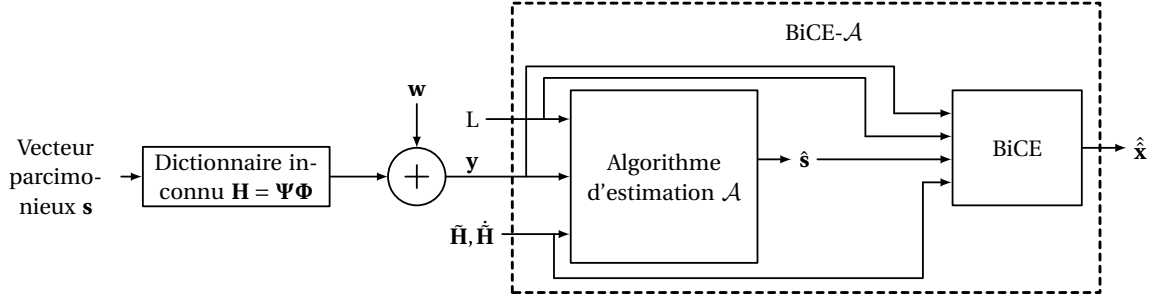


FIGURE 4.4 – Description de l'algorithme BiCE-A.

Nous nommons l'algorithme constitué de la succession de l'étape d'estimation  $\mathcal{A}$  suivi de la correction BiCE "BiCE-A".

L'algorithme BiCE se concentre sur la correction uniquement des amplitudes non nulles de  $\hat{\mathbf{s}}$ , rassemblées dans le vecteur  $\mathbf{x}$ . Pour ce faire on utilise le dictionnaire réduit au support  $\mathcal{S}$  noté  $\tilde{\mathbf{H}}_{\mathcal{S}}$  et sa dérivée  $\dot{\tilde{\mathbf{H}}}_{\mathcal{S}}$ . En introduisant (4.67) dans la formulation des mesures on a :

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w} = \mathbf{H}_{\mathcal{S}}\mathbf{x} + \mathbf{w} = \tilde{\mathbf{H}}_{\mathcal{S}}\mathbf{s} - \dot{\tilde{\mathbf{H}}}_{\mathcal{S}} \text{diag}(\mathbf{e})\mathbf{x} + \mathbf{w}. \quad (4.68)$$

De la même façon qu'en (4.67), on peut également procéder au développement de Taylor concernant le dictionnaire dérivé erroné :

$$\dot{\tilde{\mathbf{H}}}_{\mathcal{S}} \approx \tilde{\mathbf{H}}_{\mathcal{S}} + \ddot{\tilde{\mathbf{H}}}_{\mathcal{S}} \text{diag}(\mathbf{e}), \quad (4.69)$$

où la dérivé seconde du dictionnaire est  $\ddot{\tilde{\mathbf{H}}}_{\mathcal{S}} \approx \Psi\ddot{\Phi}$  où les colonnes de  $\ddot{\Phi}$  sont définies par :

$$[\ddot{\Phi}]_k = \left. \frac{\partial^2 \Phi(\theta)}{\partial^2 \theta} \right|_{\theta=\bar{\theta}_k}. \quad (4.70)$$

De ce fait on peut en conclure que les échantillons sont approximativement modélisés selon :

$$\mathbf{y} \approx (\tilde{\mathbf{H}}_{\mathcal{S}} - \dot{\tilde{\mathbf{H}}}_{\mathcal{S}} \text{diag}(\mathbf{e}))\mathbf{x} + \ddot{\tilde{\mathbf{H}}}_{\mathcal{S}} \text{diag}(\mathbf{e})^2 \mathbf{x} + \mathbf{w}, \quad (4.71)$$

où de la même façon que dans la section 4.1.1,  $\mathbf{x}$  est le vecteur de longueur  $L$  des amplitudes non nulles de  $\mathbf{s}$ .

L'expression (4.71) nous permet de remarquer que l'influence de l'erreur sur les échantillons est apportée par la matrice  $\dot{\tilde{\mathbf{H}}}$  principalement. Le reste de l'influence de l'erreur étant en  $O(\text{diag}(\mathbf{e})^2)$  est donc négligeable, puisque dans le développement de Taylor nous avons fait l'hypothèse que le terme en  $O(\text{diag}(\mathbf{e})^2)$  était négligeable. L'objectif de l'algorithme BiCE est de réduire cette influence. Pour ce faire, il se base sur la projection oblique des échantillons mesurés sur l'espace généré par le dictionnaire  $\langle \tilde{\mathbf{H}} \rangle$  tout en annulant la composante liée à l'erreur, qui est sur l'espace généré par le dictionnaire dérivé  $\langle \dot{\tilde{\mathbf{H}}} \rangle$ . On utilisera pour cela le projecteur oblique  $\mathbf{E}_{\tilde{\mathbf{H}}_{\mathcal{S}}\dot{\tilde{\mathbf{H}}}_{\mathcal{S}}}$  dont l'espace image est  $\langle \tilde{\mathbf{H}} \rangle$  et le noyau  $\langle \dot{\tilde{\mathbf{H}}} \rangle$  dont l'expression est donnée en (2.23). On utilisera ensuite les nouveaux échantillons projetés pour estimer le nouveau vecteur d'amplitude  $\hat{\mathbf{x}}$ . Pour cela on fait simplement une minimisation des moindres carrés basée sur le modèle  $\mathbf{y} = \tilde{\mathbf{H}}_{\mathcal{S}}\mathbf{x}$  qui est le terme mis en avant par la projection effectuée. L'algorithme BiCE procède en deux étapes :

1. On procède à la projection oblique des échantillons  $\mathbf{y}$  :

$$\mathbf{y} = \mathbf{E}_{\tilde{\mathbf{H}}_{\mathcal{S}}\dot{\tilde{\mathbf{H}}}_{\mathcal{S}}} \mathbf{y} \quad (4.72)$$

2. On estime à présent les amplitudes  $\hat{\mathbf{x}}$  avec ces nouvelles mesures projetées, par une minimisation classique au moindre carré de la fonction

$$\Gamma(\mathbf{x}) = \|\mathbf{Y} - \tilde{\mathbf{H}}_{\mathcal{S}}\mathbf{x}\|^2 \quad (4.73)$$

dont le résultat est

$$\hat{\mathbf{x}} = \tilde{\mathbf{H}}_{\mathcal{S}}^{\dagger} \mathbf{Y} \quad (4.74)$$

où l'on note  $\dagger$  la pseudo-inverse d'une matrice non carrée.

Notons que, contrairement à l'algorithme OGEC, le seul résultat de l'algorithme d'estimation  $\mathcal{A}$  que l'on utilise dans le BiCE est l'estimation du support, puisque l'on procède à une réestimation ensuite des amplitudes. De ce fait nous verrons que n'importe quel algorithme  $\mathcal{A}$  estimant le support de  $\mathbf{s}$  peut être utilisé de manière interchangeable, sans que les performances d'estimation sur les amplitudes ne varient. L'algorithme combiné BiCE- $\mathcal{A}$  est résumé dans l'algorithme 2.

**Données :**  $\mathbf{y}, \tilde{\boldsymbol{\theta}}, \phi(t, \theta), \Psi$

**Résultat :**  $\hat{\mathbf{s}}$

1. Générer le dictionnaire  $\tilde{\mathbf{H}} = \Psi \tilde{\boldsymbol{\Phi}}$  ainsi que sa dérivée  $\dot{\tilde{\mathbf{H}}} = \Psi \dot{\tilde{\boldsymbol{\Phi}}}$  à l'aide de la discrétisation de la fonction  $\phi(t, \theta)$  avec la grille  $\tilde{\boldsymbol{\theta}}$ .
2. Exécuter l'algorithme d'estimation parcimonieuse  $\mathcal{A}$  pour obtenir une estimation  $\hat{\mathbf{s}}$ ,

$$\mathcal{A}(\mathbf{y}, \tilde{\mathbf{H}}) \rightarrow \hat{\mathbf{s}}. \quad (4.75)$$

3. A partir de l'estimée  $\hat{\mathbf{s}}$  extraire le support estimé  $\hat{\mathcal{S}}$ .
4. Utiliser  $\hat{\mathcal{S}}$  pour extraire les dictionnaires réduits au supports  $\tilde{\mathbf{H}}_{\hat{\mathcal{S}}}$  et  $\dot{\tilde{\mathbf{H}}}_{\hat{\mathcal{S}}}$ .
5. Calculer le projecteur oblique  $\mathbf{E}_{\tilde{\mathbf{H}}_{\hat{\mathcal{S}}}\dot{\tilde{\mathbf{H}}}_{\hat{\mathcal{S}}}}$  suivant l'expression (2.23) :

$$\mathbf{E}_{\tilde{\mathbf{H}}_{\hat{\mathcal{S}}}\dot{\tilde{\mathbf{H}}}_{\hat{\mathcal{S}}}} = \tilde{\mathbf{H}}_{\hat{\mathcal{S}}} \left( \tilde{\mathbf{H}}_{\hat{\mathcal{S}}}^T \mathbf{P}_{\dot{\tilde{\mathbf{H}}}_{\hat{\mathcal{S}}}}^{\perp} \tilde{\mathbf{H}}_{\hat{\mathcal{S}}} \right)^{-1} \tilde{\mathbf{H}}_{\hat{\mathcal{S}}}^T \mathbf{P}_{\dot{\tilde{\mathbf{H}}}_{\hat{\mathcal{S}}}}^{\perp}. \quad (4.76)$$

6. Effectuer l'estimation des amplitudes non nulles projetées par :

$$\hat{\mathbf{x}}_{\hat{\mathcal{S}}} = \tilde{\mathbf{H}}_{\hat{\mathcal{S}}}^{\dagger} \mathbf{E}_{\tilde{\mathbf{H}}_{\hat{\mathcal{S}}}\dot{\tilde{\mathbf{H}}}_{\hat{\mathcal{S}}}} \mathbf{y}. \quad (4.77)$$

7. Remplacer les amplitudes non nulles de  $\hat{\mathbf{s}}$  par les estimées  $\hat{\mathbf{x}}_{\hat{\mathcal{S}}}$  pour obtenir le vecteur parcimonieux estimé  $\hat{\mathbf{s}}$

**Algorithme 2 :** Algorithme BiCE- $\mathcal{A}$

La première étape de cet algorithme n'est pas nécessaire suivant les applications, si l'on dispose directement des matrices  $\tilde{\mathbf{H}}$  et  $\dot{\tilde{\mathbf{H}}}$ .

### Coût de calcul des post traitement BiCE et OGEC

Le post-traitement BiCE est très simple à mettre en place puisqu'il est composé simplement de deux étapes : une projection oblique des mesures, et ensuite une minimisa-

tion des moindres carrés des mesures projetées, résumé dans l'expression (4.77) qui peut être réécrit comme :

$$\hat{\mathbf{x}}_{\hat{\mathcal{S}}} = \left( \tilde{\mathbf{H}}_{\hat{\mathcal{S}}}^T \mathbf{P}_{\tilde{\mathbf{H}}_{\hat{\mathcal{S}}}}^{\perp} \tilde{\mathbf{H}}_{\hat{\mathcal{S}}} \right)^{-1} \tilde{\mathbf{H}}_{\hat{\mathcal{S}}}^T \mathbf{P}_{\tilde{\mathbf{H}}_{\hat{\mathcal{S}}}}^{\perp} \dot{\tilde{\mathbf{H}}}_{\hat{\mathcal{S}}} \mathbf{y} \quad (4.78)$$

$$\Leftrightarrow \tilde{\mathbf{H}}_{\hat{\mathcal{S}}}^T \mathbf{P}_{\tilde{\mathbf{H}}_{\hat{\mathcal{S}}}}^{\perp} \tilde{\mathbf{H}}_{\hat{\mathcal{S}}} \hat{\mathbf{x}}_{\hat{\mathcal{S}}} = \tilde{\mathbf{H}}_{\hat{\mathcal{S}}}^T \mathbf{P}_{\tilde{\mathbf{H}}_{\hat{\mathcal{S}}}}^{\perp} \dot{\tilde{\mathbf{H}}}_{\hat{\mathcal{S}}} \mathbf{y} \quad (4.79)$$

$$\Leftrightarrow \mathbf{Z}^T \mathbf{Z} \hat{\mathbf{x}}_{\hat{\mathcal{S}}} = \mathbf{Z} \mathbf{y} \quad (4.80)$$

où  $\mathbf{Z} = \mathbf{P}_{\tilde{\mathbf{H}}_{\hat{\mathcal{S}}}}^{\perp} \tilde{\mathbf{H}}_{\hat{\mathcal{S}}}$ . La formulation (4.80) est une minimisation des moindres carrés classique dont le coût de calcul peut être évalué en  $O(N^2L)$ . On peut comparer la complexité du calcul du projecteur orthogonal de rang  $(N-L)$ ,  $\mathbf{P}_{\tilde{\mathbf{H}}_{\hat{\mathcal{S}}}}^{\perp}$  avec la complexité d'une décomposition en valeur singulière de rang réduit, qui est en  $O((N-L)NL)$ . Le calcul de la projection  $\mathbf{Z}$  en elle même a une complexité en  $O(N^2L)$ . Ces trois étapes ont quasiment le même poids, on peut donc en conclure que la complexité de l'algorithme BiCE est en  $O(N^2L)$ .

L'étape d'estimation de l'erreur de grille  $\hat{\mathbf{e}}$  de l'algorithme OGEC (4.32) repose sur les mêmes calculs que l'étape d'estimation des amplitudes de l'algorithme BiCE (4.77) et a donc le même coût. Dans le cas de l'algorithme OGEC il faut ajouter la complexité de l'étape de ré-estimation des amplitudes si on a besoin d'estimer ce paramètre. On peut comparer ce coût à celui de l'OMP, qui est l'un des estimateurs parcimonieux connu pour sa rapidité d'exécution. Son coût de calcul est de l'ordre de  $O(LNK)$  [128], ce qui est supérieur à celui des deux algorithmes proposés. En effet, par définition on a  $K \gg N$ , l'absence de ce paramètre dans le post-traitement de par la réduction à seulement les amplitudes non nulles permet de réduire fortement le coût de calcul de cette deuxième étape. L'OMP étant un des algorithmes de reconstruction parcimonieuse le plus léger d'un point de vue calculatoire, on peut considérer que de manière générale le coût du post traitement BiCE et OGEC est inférieur à celui de  $\mathcal{A}$ .

Un exemple de temps de calcul sous Matlab comparant les estimateurs parcimonieux OMP, CoSAMP, et SPGL1 avec le post traitement BiCE est donné dans le tableau 4.1. Le temps indiqué dans le tableau est le temps moyen de calcul mesuré avec les fonctions `tic` et `toc` sur 50 réalisations, pour  $N = 300$ ,  $L = 10$ ,  $K = 100$ ,  $\sigma^2 = 1 \cdot 10^{-4}$  et  $\sigma_e^2 = 1 \cdot 10^{-2}$ . Le contexte applicatif que nous utilisons est le modèle de l'estimation des retards des signaux impulsionsnels, introduits dans la section 2.4 et que nous allons présenter plus en détail dans la partie simulation. Le temps de calcul du post-traitement BiCE est du même ordre de grandeur que celui utilisé par l'algorithme OMP pour les paramètres considérés, tandis que les autres méthodes d'estimation parcimonieuses sont beaucoup plus coûteuses. Le temps de calcul du BiCE est similaire à l'OMP dans les simulations effectuées, mais le programme n'est pas du tout optimisé et utilise des méthodes de calcul et d'inversion matricielle par défaut. Le programme suit les étapes données dans l'algorithme 2 et fait la minimisation au moindre carré par un produit matriciel alors qu'il est possible d'optimiser cette étape en suivant le modèle (4.80).

TABLEAU 4.1 – Temps de calcul de différents algorithmes d'estimation.

|        |         |
|--------|---------|
| OMP    | 6.4ms   |
| COSAMP | 146.5ms |
| SPGL1  | 238.4ms |
| BiCE   | 5.7ms   |

### 4.2.3 Performances théoriques de l'algorithme

#### Etude du biais et de l'erreur quadratique moyenne

Nous cherchons à présent à caractériser le biais théorique sur l'estimation des amplitudes non nulles en sortie de l'algorithme BiCE, défini par

$$\mathbf{b}(\mathbf{x}) = \mathbb{E}_{\mathbf{w}, \mathbf{e} | \mathbf{x}} [\hat{\mathbf{x}} - \mathbf{x}]. \quad (4.81)$$

Nous nous concentrons uniquement sur la partie BiCE, étant donné que les performances d'estimation du support sont celles de l'algorithme  $\mathcal{A}$  utilisé, c'est pourquoi nous nous concentrons sur l'erreur d'estimation du vecteur  $\mathbf{x}$ . On fait de ce fait l'hypothèse que le support est correctement estimé, c'est à dire que  $\hat{\mathbf{x}}_{\mathcal{S}} = \hat{\mathbf{x}}$ . De plus nous faisons l'hypothèse que le bruit  $\mathbf{w}$  est un bruit blanc gaussien centré de variance  $\sigma^2$  tandis que l'erreur est également centrée, de variance  $\sigma_e^2$  et son moment d'ordre 4 est  $m_e^4$ . On s'intéresse au contexte bayésien, c'est à dire que les paramètres à estimer sont également des variables aléatoires, on considère que les amplitudes suivent une distribution gaussienne de variance  $\sigma_x^2$  de telle sorte que  $\mathbb{E}_{\mathbf{x}}(\mathbf{x}\mathbf{x}^T) = \sigma_x^2 \mathbf{I}$ .

On peut exprimer le résultat de l'algorithme BiCE (4.77) à l'aide de l'expression développée des échantillons, donnée en (4.71) :

$$\hat{\mathbf{x}} \approx \tilde{\mathbf{H}}_{\mathcal{S}}^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_{\mathcal{S}} \dot{\mathbf{H}}_{\mathcal{S}}} \left( (\tilde{\mathbf{H}}_{\mathcal{S}} - \dot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\mathbf{e})) \mathbf{x} + \ddot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\mathbf{e})^2 \mathbf{x} + \mathbf{w} \right) \quad (4.82)$$

$$\approx \mathbf{x} + \tilde{\mathbf{H}}_{\mathcal{S}}^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_{\mathcal{S}} \dot{\mathbf{H}}_{\mathcal{S}}} (\ddot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\mathbf{e})^2 \mathbf{x} + \mathbf{w}). \quad (4.83)$$

Ce résultat est obtenu directement par les propriétés (2.21) et (2.22) du projecteur oblique. L'erreur d'estimation est donc égale à :

$$\hat{\mathbf{x}} - \mathbf{x} \approx \tilde{\mathbf{H}}_{\mathcal{S}}^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_{\mathcal{S}} \dot{\mathbf{H}}_{\mathcal{S}}} (\ddot{\mathbf{H}}_{\mathcal{S}} \text{diag}(\mathbf{e})^2 \mathbf{x} + \mathbf{w}). \quad (4.84)$$

L'erreur quadratique moyenne est définie par :

$$\text{BEQM}_{\text{BiCE}} = \mathbb{E}_{\mathbf{w}, \mathbf{e}, \mathbf{x}} \left[ \|\hat{\mathbf{x}} - \mathbf{x}\|^2 \right] = \mathbb{E}_{\mathbf{w}, \mathbf{e}, \mathbf{x}} \left[ \text{Tr} \left[ (\hat{\mathbf{x}} - \mathbf{x}) (\hat{\mathbf{x}} - \mathbf{x})^T \right] \right]. \quad (4.85)$$

Ce qui devient en introduisant (4.84) :

$$\begin{aligned}
 \text{BEQM}_{\text{BiCE}} &\approx \text{Tr} \left[ \mathbb{E}_{\mathbf{w}, \mathbf{e}, \mathbf{x}} \left[ \tilde{\mathbf{H}}_S^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S} \left( \ddot{\mathbf{H}}_S \text{diag}(\mathbf{e})^2 \mathbf{x} + \mathbf{w} \right) \left( \mathbf{x}^T \ddot{\mathbf{H}}_S^T \text{diag}(\mathbf{e})^2 + \mathbf{w}^T \right) \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S}^T \tilde{\mathbf{H}}_S^{\dagger T} \right] \right] \\
 &= \text{Tr} \left[ \mathbb{E}_{\mathbf{w}, \mathbf{e}, \mathbf{x}} \left[ \text{diag}(\mathbf{e})^4 \tilde{\mathbf{H}}_S^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S} \ddot{\mathbf{H}}_S \mathbf{x} \mathbf{x}^T \ddot{\mathbf{H}}_S^T \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S}^T \tilde{\mathbf{H}}_S^{\dagger T} \right] \right] \\
 &\quad + \text{Tr} \left[ \mathbb{E}_{\mathbf{w}, \mathbf{e}, \mathbf{x}} \left[ \text{diag}(\mathbf{e})^2 \tilde{\mathbf{H}}_S^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S} \left( \ddot{\mathbf{H}}_S \mathbf{x} \mathbf{w}^T + \mathbf{w} \mathbf{x}^T \ddot{\mathbf{H}}_S^T \right) \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S}^T \tilde{\mathbf{H}}_S^{\dagger T} \right] \right] \\
 &\quad + \text{Tr} \left[ \mathbb{E}_{\mathbf{w}, \mathbf{e}, \mathbf{x}} \left[ \tilde{\mathbf{H}}_S^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S} \mathbf{w} \mathbf{w}^T \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S}^T \tilde{\mathbf{H}}_S^{\dagger T} \right] \right] \\
 &= \sigma_x^2 \text{Tr} \left[ \mathbb{E}_{\mathbf{e}} \left[ \text{diag}(\mathbf{e})^4 \tilde{\mathbf{H}}_S^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S} \ddot{\mathbf{H}}_S \ddot{\mathbf{H}}_S^T \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S}^T \tilde{\mathbf{H}}_S^{\dagger T} \right] \right] \\
 &\quad + \sigma^2 \text{Tr} \left[ \mathbb{E}_{\mathbf{e}} \left[ \tilde{\mathbf{H}}_S^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S} \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S}^T \tilde{\mathbf{H}}_S^{\dagger T} \right] \right] \\
 &= \sigma_x^2 \mathbb{E}_{\mathbf{e}} \left[ \text{Tr} \left[ \text{diag}(\mathbf{e})^4 \left( \tilde{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \tilde{\mathbf{H}}_S \right)^{-1} \tilde{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \ddot{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \tilde{\mathbf{H}}_S \left( \tilde{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \tilde{\mathbf{H}}_S \right)^{-1} \right] \right] \\
 &\quad + \sigma^2 \text{Tr} \left[ \mathbb{E}_{\mathbf{e}} \left[ \left( \tilde{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \tilde{\mathbf{H}}_S \right)^{-1} \right] \right] \\
 &= \sigma_x^2 \mathbb{E}_{\mathbf{e}} \left[ \text{Tr} \left[ \text{diag}(\mathbf{e})^4 \ddot{\mathbf{H}}_S \ddot{\mathbf{H}}_S^T \left( \tilde{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \tilde{\mathbf{H}}_S \right)^{-1} \right] \right] + \sigma^2 \text{Tr} \left[ \mathbb{E}_{\mathbf{e}} \left[ \left( \tilde{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \tilde{\mathbf{H}}_S \right)^{-1} \right] \right].
 \end{aligned} \tag{4.87}$$

Le deuxième terme de la somme de la deuxième égalité s'annule puisque le bruit est centré.

La BEQM ainsi que le biais moyen peuvent être calculés numériquement à partir de ces expressions mais cela demande de calculer des espérances sur le paramètre d'erreur à l'aide d'un procédé de type Monte Carlo. Cela est compliqué à utiliser en pratique, de plus l'expression ainsi obtenue ne permet pas de saisir le comportement de l'algorithme. On souhaite au contraire obtenir une expression analytique des mesures de performances ne dépendant pas d'une variable aléatoire telle que l'erreur. On cherche à exprimer ces mesures en fonction des matrices déterministes  $\mathbf{H}_S$  et  $\dot{\mathbf{H}}_S$  au lieu des matrices aléatoires  $\tilde{\mathbf{H}}_S$  et  $\dot{\tilde{\mathbf{H}}}_S$ . S'affranchir de l'espérance est possible, en utilisant une approximation de Taylor au premier ordre du projecteur orthogonal  $\mathbf{P}_{\dot{\mathbf{H}}_S}^\perp$ , donnée par KRIM, FORSTER et PROAKIS [79] :

$$\mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \approx \mathbf{P}_{\mathbf{H}_S}^\perp - \dot{\mathbf{H}}_S^{\dagger T} \text{diag}(\mathbf{e}) \dot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp + \mathbf{P}_{\mathbf{H}_S}^\perp \ddot{\mathbf{H}}_S \text{diag}(\mathbf{e}) \dot{\mathbf{H}}_S^\dagger. \tag{4.88}$$

En se basant sur cette approximation, on peut donner les approximations des grandeurs

suivantes :

$$\begin{aligned}
 \tilde{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \tilde{\mathbf{H}}_S &= (\mathbf{H}_S + \dot{\mathbf{H}}_S \text{diag}(\mathbf{e}))^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp (\mathbf{H}_S + \dot{\mathbf{H}}_S \text{diag}(\mathbf{e})) \\
 &\quad - (\mathbf{H}_S + \dot{\mathbf{H}}_S \text{diag}(\mathbf{e}))^T \dot{\mathbf{H}}_S^T \text{diag}(\mathbf{e}) \ddot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp (\mathbf{H}_S + \dot{\mathbf{H}}_S \text{diag}(\mathbf{e})) \\
 &\quad - (\mathbf{H}_S + \dot{\mathbf{H}}_S \text{diag}(\mathbf{e}))^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \text{diag}(\mathbf{e}) \dot{\mathbf{H}}_S^T (\mathbf{H}_S + \dot{\mathbf{H}}_S \text{diag}(\mathbf{e})) \\
 &= \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \mathbf{H}_S + \text{diag}(\mathbf{e}) \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \dot{\mathbf{H}}_S + \text{diag}(\mathbf{e}) \dot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \mathbf{H}_S + \text{diag}(\mathbf{e})^2 \dot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \dot{\mathbf{H}}_S \\
 &\quad - \mathbf{H}_S^T \dot{\mathbf{H}}_S^T \text{diag}(\mathbf{e}) \ddot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \mathbf{H}_S - \text{diag}(\mathbf{e})^2 \mathbf{H}_S^T \dot{\mathbf{H}}_S^T \ddot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \dot{\mathbf{H}}_S \\
 &\quad - \text{diag}(\mathbf{e})^2 \dot{\mathbf{H}}_S^T \dot{\mathbf{H}}_S^T \ddot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \mathbf{H}_S - \text{diag}(\mathbf{e})^2 \dot{\mathbf{H}}_S^T \dot{\mathbf{H}}_S^T \text{diag}(\mathbf{e}) \ddot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \dot{\mathbf{H}}_S \\
 &\quad - \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \text{diag}(\mathbf{e}) \dot{\mathbf{H}}_S^T \mathbf{H}_S - \text{diag}(\mathbf{e})^2 \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^T \mathbf{H}_S \\
 &\quad - \text{diag}(\mathbf{e})^2 \dot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^T \mathbf{H}_S - \text{diag}(\mathbf{e})^3 \dot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^T \mathbf{H}_S \\
 &\approx \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \mathbf{H}_S,
 \end{aligned}$$

où seul le premier terme ne dépendant pas de l'erreur est conservé tandis qu'on considère comme négligeables les termes en  $\text{diag}(\mathbf{e})$ ,  $\text{diag}(\mathbf{e})^2$  et  $\text{diag}(\mathbf{e})^3$ . Cette approximation est importante, et peut sembler brutale, mais nous montrerons dans la partie simulation qu'elle permet de bien approcher le comportement de l'erreur, notamment pour des grandes matrices aléatoires qui sont le type de matrices rencontré le plus souvent en pratique.

Le projecteur oblique est approximé en :

$$\begin{aligned}
 \tilde{\mathbf{H}}_S^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_S \tilde{\mathbf{H}}_S} &= \left( \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1} \tilde{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \\
 &\approx \left( \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1} (\mathbf{H}_S + \dot{\mathbf{H}}_S \text{diag}(\mathbf{e}))^T \left( \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp - \dot{\mathbf{H}}_S^T \text{diag}(\mathbf{e}) \ddot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp - \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \text{diag}(\mathbf{e}) \dot{\mathbf{H}}_S^T \right) \\
 &\approx \left( \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1} \left( \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp - \text{diag}(\mathbf{e}) \ddot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp - \text{diag}(\mathbf{e}) \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^T \right. \\
 &\quad \left. + \text{diag}(\mathbf{e}) \dot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp - \text{diag}(\mathbf{e})^2 \ddot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp - \text{diag}(\mathbf{e})^2 \dot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^T \right) \\
 &\approx \mathbf{H}^\dagger \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S} + \text{diag}(\mathbf{e}) \left( \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1} \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp - \ddot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp - \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^T \right),
 \end{aligned}$$

où cette fois-ci seuls les termes de second ordre sont négligés. Cette approximation permet de donner une expression plus simple de l'erreur d'estimation :

$$\begin{aligned}
 \hat{\mathbf{x}} - \mathbf{x} &\approx \mathbf{H}^\dagger \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S} (\ddot{\mathbf{H}}_S \text{diag}(\mathbf{e})^2 \mathbf{x} + \mathbf{w}) \\
 &\quad + \text{diag}(\mathbf{e}) \left( \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1} \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp - \ddot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp - \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^T \right) (\ddot{\mathbf{H}}_S \text{diag}(\mathbf{e})^2 \mathbf{x} + \mathbf{w}) \\
 &\approx \mathbf{H}^\dagger \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S} \mathbf{w} + \text{diag}(\mathbf{e}) \left( \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1} \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp - \ddot{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp - \mathbf{H}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^T \right) \mathbf{w} \\
 &\quad + \text{diag}(\mathbf{e})^2 \mathbf{H}^\dagger \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S} \ddot{\mathbf{H}}_S \mathbf{x},
 \end{aligned}$$

où le terme en  $e_l^3$  a été négligé.

De cette erreur on peut déterminer les différentes métriques de performance, premièrement le biais est donné par :

$$\begin{aligned}
 \mathbf{b}(\mathbf{x}) &= \mathbb{E}_{\mathbf{w}, \mathbf{e} | \mathbf{x}} \left[ \text{diag}(\mathbf{e})^2 \mathbf{H}^\dagger \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S} \ddot{\mathbf{H}}_S \mathbf{x} \right] \\
 &= \sigma_e^2 \mathbf{H}^\dagger \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S} \ddot{\mathbf{H}}_S \mathbf{x},
 \end{aligned}$$

en utilisant le fait que la moyenne sur le bruit de  $\mathbf{w}$  est nulle.

En second lieu on peut également donner une expression analytique de la BEQM :

$$\begin{aligned}
 \text{BEQM}_{\text{Biac}} &\approx \text{Tr} \left[ \mathbb{E}_{\mathbf{w}, \mathbf{e}, \mathbf{x}} \left[ (\hat{\mathbf{x}} - \mathbf{x}) (\hat{\mathbf{x}} - \mathbf{x})^T \right] \right] \\
 &= \text{Tr} \left[ \mathbb{E}_{\mathbf{w}, \mathbf{e}, \mathbf{x}} \left[ \mathbf{H}^T \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S} \mathbf{w} \mathbf{w}^T \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S}^T \mathbf{H}^{\dagger T} \right] \right] \\
 &\quad + \text{Tr} \left[ \mathbb{E}_{\mathbf{w}, \mathbf{e}, \mathbf{x}} \left[ \text{diag}(\mathbf{e})^2 \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp - \ddot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp - \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^\dagger \right) \mathbf{w} \mathbf{w}^T \right. \right. \\
 &\quad \cdot \left. \left. \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp - \ddot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp - \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^\dagger \right)^T \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \right] \right] \\
 &\quad + \text{Tr} \left[ \mathbb{E}_{\mathbf{w}, \mathbf{e}, \mathbf{x}} \left[ \text{diag}(\mathbf{e})^4 \mathbf{H}^T \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S} \ddot{\mathbf{H}}_S \mathbf{x} \mathbf{x}^T \ddot{\mathbf{H}}_S^T \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S}^T \mathbf{H}^{\dagger T} \right] \right] \\
 &= \sigma^2 \text{Tr} \left[ \mathbf{H}^T \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S} \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S}^T \mathbf{H}^{\dagger T} \right] \\
 &\quad + \sigma^2 \sigma_e^2 \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp - \ddot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp - \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^\dagger \right) \right. \\
 &\quad \cdot \left. \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp - \ddot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp - \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^\dagger \right)^T \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \right] \\
 &\quad + m_e^4 \sigma_x^2 \text{Tr} \left[ \mathbf{H}^T \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S} \ddot{\mathbf{H}}_S \ddot{\mathbf{H}}_S^T \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S}^T \mathbf{H}^{\dagger T} \right] \\
 &\approx \sigma^2 \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \right] \\
 &\quad + m_e^4 \sigma_x^2 \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \ddot{\mathbf{H}}_S \ddot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \right].
 \end{aligned}$$

Premièrement les termes proportionnels à  $\mathbf{n}$  n'ont pas été détaillés puisque leur espérance est nulle, et pour obtenir l'expression finale nous considérons que le terme croisé  $\sigma_e^2 \sigma^2$  est négligeable puisque à la fois  $\sigma^2$  et  $\sigma_e^2$  sont petits.

**Résultat 24.** Pour une estimation du support parfaite, et  $\sigma_e^2$  suffisamment petit, le vecteur de biais conditionnel ainsi que la BEQM de l'estimateur BiCE sont donnés par :

$$\mathbf{b}(\mathbf{x}) \approx \sigma_e^2 \mathbf{H}^T \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S} \ddot{\mathbf{H}}_S \mathbf{x}, \quad (4.89)$$

$$\text{BEQM}_{\text{BiCE}} \approx \sigma^2 F_S + \sigma_x^2 m_e^4 \bar{F}_S, \quad (4.90)$$

où  $m_e^4 = \mathbb{E}(e_l^4)$  et les quantités  $F_S$  et  $\bar{F}_S$  sont données respectivement dans les expressions (4.91) et (4.92).

$$F_S = \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \right], \quad (4.91)$$

$$\bar{F}_S = \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \ddot{\mathbf{H}}_S \ddot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \right], \quad (4.92)$$

où  $\ddot{\mathbf{H}}$  est la dérivée au second ordre de la matrice de dictionnaire utilisant la grille exacte.

On peut observer que le biais conditionnel tend vers zéro quand l'erreur sur le dictionnaire tend vers zéro également. On peut en conclure qu'en l'absence d'erreur l'algorithme BiCE est conditionnellement non biaisé, ce qui assure que les performances de l'algorithme  $\mathcal{A}$  ne sont pas dégradées en l'absence d'erreur de dictionnaire par le post-traitement BiCE.

**Résultat 25.** Quand  $\sigma_e^2$ , bien que supposé petit domine le bruit  $\sigma^2$ , la BEQM de l'algorithme BiCE est :

$$\text{BEQM}_{\text{BiCE}} \approx \sigma_x^2 m_e^4 \bar{F}_S = O(m_e^4). \quad (4.93)$$

Le résultat précédent est obtenu simplement en négligeant le terme lié uniquement au bruit.

### Efficacité statistique de l'algorithme

Pour étudier l'efficacité de l'algorithme BiCE il faut se placer dans le contexte des bornes biaisées, puisque nous avons vu en (4.89) que l'estimateur BiCE est un estimateur biaisé. Etant donné qu'on se focalise sur le comportement du post-traitement, les bornes sont considérées uniquement sur le vecteur non parcimonieux  $\mathbf{x}$ . De plus on remarque que la BEQM dépend de l'observation projetée  $\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}$ , on va donc tout d'abord dériver la borne moyennée de l'estimateur biaisé sur les mesures projetées, qui correspond au comportement de l'algorithme BiCE. Cette borne va ensuite être comparée à la borne idéale pour un estimateur non biaisé en l'absence d'erreur de grille. La BEQM sur les mesures projetées est bornée inférieurement par la BCRM :

$$\mathbb{E} \left( \|\hat{\mathbf{x}}(\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}) - \mathbf{x}\|^2 \right) \geq \mathcal{C}_{S|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}}, \quad (4.94)$$

où la BCRM sur l'estimateur biaisé des mesures projetée est définie par :

$$\mathcal{C}_{S|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}} = \text{Tr} \left[ \mathbb{E}_{\mathbf{x}} \left( \left( \mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}} \right) \mathbf{F}(\mathbf{x}|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y})^{-1} \left( \mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}} \right)^T \right) \right] + \mathbb{E}_{\mathbf{x}} (\|\mathbf{b}(\mathbf{x})\|^2), \quad (4.95)$$

où  $\mathbf{F}(\mathbf{x}|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y})$  est la MIF et  $\mathbf{b}(\mathbf{x})$  le biais de l'estimation (4.89). Cette borne est nommée en anglais la borne I-ECRB biaisée [29], il s'agit de la borne moyennée sur le vecteur  $\mathbf{x}$  de la MIF biaisée donnée en (1.62) définie par rapport aux mesures projetées  $\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}$  au lieu de directement les mesures  $\mathbf{y}$ . En utilisant la définition du biais de l'estimateur BiCE on a :

$$\frac{\partial \mathbf{b}(\boldsymbol{\theta}_S)}{\partial \boldsymbol{\theta}_S} \approx \sigma_e^2 \mathbf{H}_S^\dagger \mathbf{E}_{\mathbf{H}_S \mathbf{H}_S} \ddot{\mathbf{H}}_S = \sigma_e^2 \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \mathbf{H}_S \mathbf{P}_{\mathbf{H}_S}^\perp \ddot{\mathbf{H}}_S, \quad (4.96)$$

qui ne dépend pas du vecteur d'amplitude  $\mathbf{x}$ .

Les mesures projetées sont définies par :

$$\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y} = \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \mathbf{x} + \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{w}. \quad (4.97)$$

On a donc

$$\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y} | \mathbf{x} \sim \mathcal{N} \left( \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \mathbf{x}, \sigma^2 \mathbf{P}_{\mathbf{H}_S}^\perp \right), \quad (4.98)$$

ce qui permet d'obtenir la MIF déterministe selon la formule de Slepian-Bang :

$$\mathbf{F}(\mathbf{x}|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}) = \frac{1}{\sigma^2} \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp (\mathbf{P}_{\mathbf{H}_S}^\perp)^{-1} \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \quad (4.99)$$

$$= \frac{1}{\sigma^2} \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S. \quad (4.100)$$

Celle-ci ne dépend pas de  $\mathbf{x}$ , ce qui permet d'ôter l'espérance sur  $\mathbf{x}$  dans le premier terme de la borne.

L'espérance de la norme du biais au carré est :

$$\mathbb{E}_{\mathbf{x}}(\|\mathbf{b}(\mathbf{x})\|^2) \approx \sigma_e^4 \sigma_x^2 \bar{F}_S, \quad (4.101)$$

où  $\bar{F}_S$  est donné en (4.92).

Le terme produit quant à lui est :

$$\begin{aligned} \text{Tr} \left[ \left( \mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}} \right) \mathbf{F}(\mathbf{x} | \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y})^{-1} \left( \mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}} \right)^T \right] &= \sigma^2 \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \right] \\ &+ \sigma^2 \sigma_e^4 \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \mathbf{H}_S \mathbf{P}_{\mathbf{H}_S}^\perp \ddot{\mathbf{H}}_S \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \ddot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S^T \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \right] \\ &+ \sigma^2 \sigma_e^2 \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \left( \mathbf{H}_S \mathbf{P}_{\mathbf{H}_S}^\perp \ddot{\mathbf{H}}_S + \ddot{\mathbf{H}}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S^T \right) \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \right] \end{aligned}$$

Les variances d'erreur et de bruit étant des quantités faibles, on considère que les termes en  $O(\sigma_e^2 \sigma^2)$  et en  $O(\sigma_e^4 \sigma^2)$  sont négligeables par rapport au terme en  $O(\sigma^2)$  pour un  $\sigma_e^2 \sigma^2$  suffisamment petit, ce qui permet d'obtenir que :

$$\text{Tr} \left[ \left( \mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}} \right) \mathbf{F}(\mathbf{x} | \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y})^{-1} \left( \mathbf{I} + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}} \right)^T \right] \approx \sigma^2 F_S \quad (4.102)$$

où  $F_S$  est donné en (4.91).

**Résultat 26.** La borne moyennée biaisée sur les erreurs projetées est donnée par :

$$\mathcal{C}_{S|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}} \approx \sigma^2 F_S + \sigma_x^2 \sigma_e^4 \bar{F}_S \quad (4.103)$$

**Résultat 27.** En présence d'erreur de dictionnaire ( $\sigma_e^2 \neq 0$ ), l'efficacité statistique de l'algorithme BiCE par rapport aux mesures projetées est :

$$\text{BEQM}_{\text{BiCE}} - \mathcal{C}_{S|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}} = \sigma_x^2 (m_e^4 - \sigma_e^4) \bar{F}_S. \quad (4.104)$$

L'inégalité de Jensen [75] pour les fonctions convexes permet de prouver que  $m_e^4 \leq \sigma_e^4$ , ce qui confirme que la borne  $\mathcal{C}_{S|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}}$  est bien une borne inférieure de la BEQM.

L'erreur quadratique moyenne du BiCE est plus aisée à saisir avec un exemple pratique de distribution. Nous prenons l'exemple d'un vecteur d'erreur  $\mathbf{e}$  suivant une distribution gaussienne généralisée centrée [88] telle que

$$e_l \sim \mathcal{GN}(0, \sigma_e^2 \frac{\Gamma(1/s)}{\Gamma(3/s)}, s). \quad (4.105)$$

Cette distribution est utilisée car elle est assez générale, et peut prendre des paramètres menant d'une distribution Laplacienne ( $s = 1$ ), à la distribution gaussienne ( $s = 2$ ) ou même uniforme ( $s \rightarrow \infty$ ). Le  $k^e$  moment d'une gaussienne généralisée centrée est donné par [88]

$$m_e^k = \sigma_e^k \sqrt{\frac{\Gamma(1/s)}{\Gamma(3/s)}}^k \frac{\Gamma((k+1)/s)}{\Gamma(1/s)}. \quad (4.106)$$

On a donc  $m_e^2 = \sigma_e^2$  et  $m_e^4 = \sigma_e^4 k_s$  où  $k_s = \frac{\Gamma(1/s)\Gamma(5/s)}{\Gamma(3/s)^2}$ . Notons que la grandeur  $k_s > 1$  quel que soit  $s$ . Prenons à présent le paramètre  $\sigma_e = d/6$  de telle sorte que dans le cas d'une distribution uniforme où  $s \rightarrow \infty$ , le support de la distribution corresponde à la largeur de la

grille discrète  $d = \bar{\theta}_j - \bar{\theta}_{j+1}$ . Dans ce cas la variance de l'erreur est  $\sigma_e^2 = d^2/36$  et le moment d'ordre 4 est donné par  $m_e^4 = k_s d^4/(36)^2$ . Nous avons vu que pour un bruit additif négligeable devant la variance de l'erreur, l'erreur quadratique moyenne de l'algorithme BiCE est de l'ordre de  $m_e^4$ , qui étant donné que  $k_s > 1$  est très petit pour des grilles espacées de  $d < 1$ . Cela montre que l'ajout du BiCE peut améliorer nettement les performances d'estimation pour un coût raisonnable dans les contexte où  $m_e^4$  est petit. Pour la distribution gaussienne généralisée, l'efficacité statistique du BiCE est

$$\text{BEQM}_{\text{BiCE}} - \mathcal{C}_{S|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}} = \sigma_x^2 \sigma_e^4 (k_s - 1) \bar{F}_S \quad (4.107)$$

qui est une fonction décroissante par rapport à  $s$ , comme illustré dans la figure 4.5. Sur cette figure nous avons tracé (4.107) en fonction de  $s$ , pour différentes valeurs d'erreur de grille  $\sigma_e^2$ . Plus  $s$  est grand, meilleure est l'efficacité de l'algorithme. Les distributions pointues de l'erreur sont donc plus difficiles à traiter pour l'algorithme, contrairement à des distributions plates telle que la distribution uniforme. Cela peut intuitivement s'expliquer par le fait que l'algorithme se base sur une projection de la partie liée à l'erreur sur les mesures pour en réduire les effets. Une distribution pointue centrée en zéro sera moins projetée qu'une distribution plus plate, et l'algorithme sera donc moins pertinent pour ce type d'erreur. L'efficacité étant proportionnelle à  $\sigma_e^2$ , augmenter cette erreur augmente la différence entre la BEQM et la borne biaisée projetée. On peut en déduire que l'algorithme proposé est adapté principalement pour corriger les faibles erreurs de grille. De plus augmenter la variance sur les amplitudes augmente également la différence entre la borne et les performances de l'algorithme.

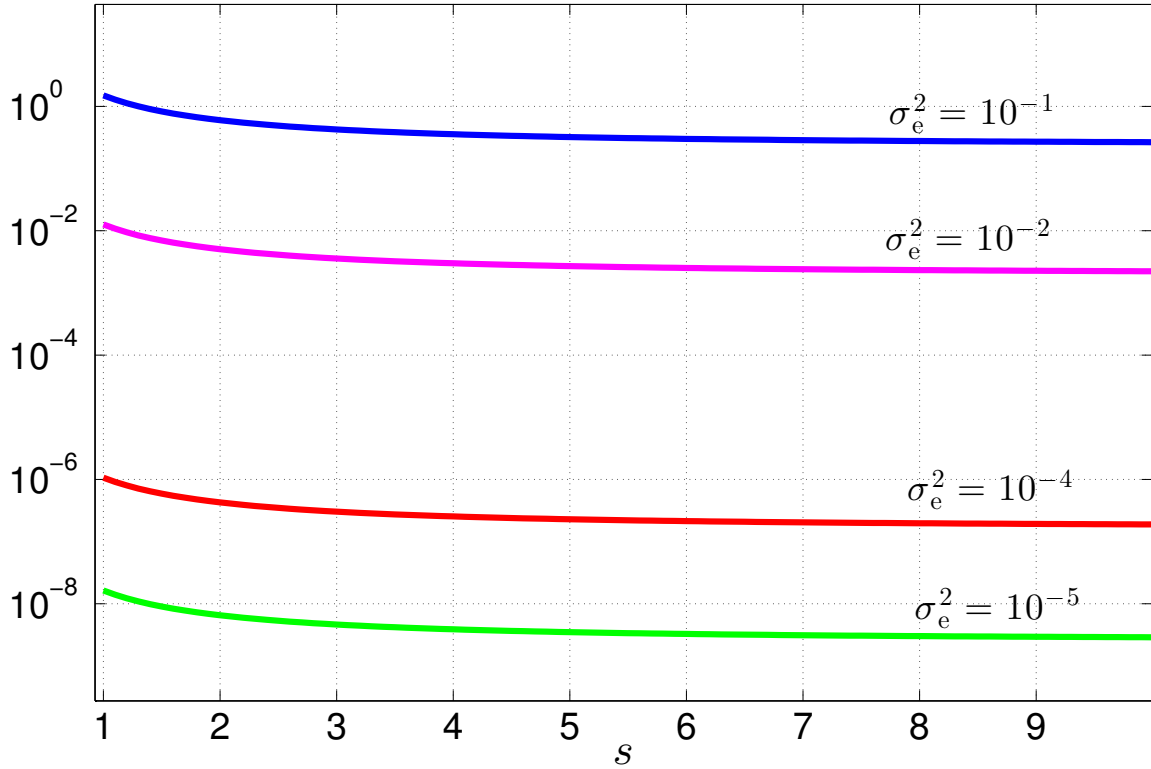


FIGURE 4.5 – Efficacité statistique de l'algorithme BiCE (4.107) en fonction du paramètre de forme  $s$  pour  $N = 50$ ,  $K = 200$ ,  $L = 10$ ,  $\sigma_\theta^2 = 1e - 1$  et différentes valeurs de  $\sigma_e^2$ .

Pour compléter la figure 4.5, on trace à présent la BEQM et la BCRM toujours en fonction du paramètre de forme  $s$  sur la figure 4.6. La première chose à noter est que la borne ne dépend pas de  $s$  puisque seul le moment d'ordre 2 de la distribution intervient et non le moment d'ordre 4 qui lui dépend de  $s$ . De ce fait la différence observée figure 4.5 vient de la diminution des performances de l'algorithme BiCE quand  $s$  augmente.

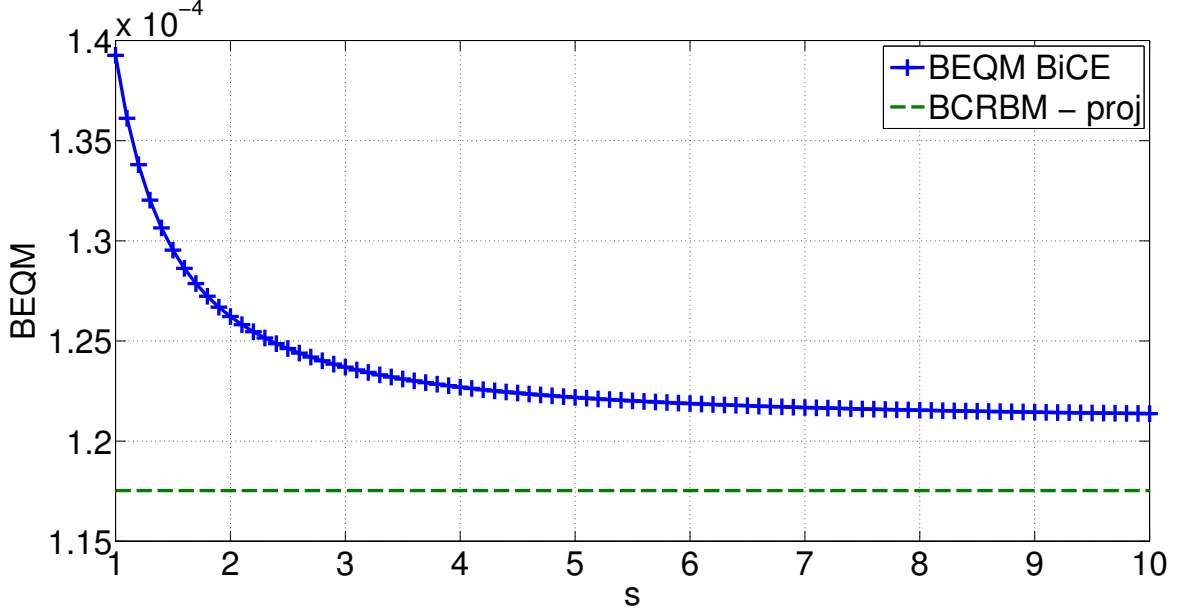


FIGURE 4.6 – Performance du BiCE en fonction du paramètre de forme  $s$ ,  $N = 200$ ,  $L = 10$ ,  $K = 1000$ ,  $\sigma^2 = 10^{-5}$  et  $\sigma_e^2 = 10^{-3}$ .

Le cas de  $\sigma_e^2 \rightarrow 0$  est important à étudier pour vérifier que l'ajout du post-traitement proposé ne dégrade pas les performances de l'algorithme d'estimation parcimonieuse  $\mathcal{A}$ . Cela permet d'utiliser l'algorithme BiCE- $\mathcal{A}$  dans le cas plus général où l'on peut être soumis à un bruit sur le dictionnaire sans que celui-ci soit tout le temps présent.

**Résultat 28.** Quand l'erreur s'annule ( $\sigma_e^2 \rightarrow 0$ ) les performances de l'algorithme BiCE sont :

$$\mathbf{b}(\mathbf{x}) \rightarrow \mathbf{0}, \quad (4.108)$$

$$\text{BEQM}_{\text{BiCE}} \rightarrow \mathcal{C}_{S|\mathbf{P}_{\mathbf{H}_S}^\perp} \mathbf{y}. \quad (4.109)$$

Ce qui signifie que l'algorithme est statistiquement efficace en l'absence d'erreur de modèle, par rapport aux mesures projetées.

On vient de montrer que le BiCE est efficace concernant les mesures projetées, cependant il peut être sous-optimal par rapport au modèle parfait de mesures sans erreur. En l'absence d'erreur la borne BCRM dans un cas idéal est dérivée à partir du modèle :

$$\mathbf{y} = \mathbf{H}_S \mathbf{x} + \mathbf{w}, \quad (4.110)$$

où  $\mathbf{w}$  un bruit blanc gaussien. Dans ce contexte la BCRM sans erreur est donnée par (voir section 3.3) :

$$\mathbb{E}(\|\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x}\|^2) \geq \mathcal{C}_{S|\mathbf{y}} = \sigma^2 \text{Tr} \left[ (\mathbf{H}_S^T \mathbf{H}_S)^{-1} \right]. \quad (4.111)$$

alors que le BEQM sans erreur est

$$\mathbb{E}(\|\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x}\|^2) = \sigma^2 \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S \right)^{-1} \right]. \quad (4.112)$$

On a la propriété que  $\mathcal{C}_{S|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}} > \mathcal{C}_{S|\mathbf{y}}$  étant donné que  $\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{H}_S < \mathbf{H}_S$  au sens des matrices positives définies donc en théorie l'algorithme BiCE est sous-optimal même en l'absence d'erreur, bien qu'il soit non biaisé. En pratique nous allons montrer dans la partie simulation que les bornes  $\mathcal{C}_{S|\mathbf{P}_{\mathbf{H}_S}^\perp \mathbf{y}}$  et  $\mathcal{C}_{S|\mathbf{y}}$  sont très proches.

#### 4.2.4 Résultat de l'algorithme pour l'estimation parcimonieuse des signaux impulsionnels

Pour illustrer les performances théoriques et démontrer l'utilisation pratique de l'algorithme BiCE, nous nous plaçons dans le même contexte des signaux à modèles paramétriques que dans la section 4.1.4.

Nous nous intéressons tout d'abord aux bornes et à la pertinence de l'approximation faites sur la BEQM basée sur le développement de Taylor du projecteur oblique, qui faisait des hypothèses importantes au niveau de l'ordre de grandeur de  $\sigma_e^2$ . Sur la courbe 4.7 on trace tout d'abord l'erreur quadratique moyenne sur 100 réalisations se basant sur l'expression exacte de la BEQM (4.87) en bleu, puis la borne approximative mais ne dépendant pas d'un paramètre aléatoire (4.90) en vert. Le point le plus important à tirer de cette courbe est que l'approximation de la BEQM est valable pour des  $\sigma_e^2 < 10^{-3}$  dans le contexte de la simulation c'est à dire  $N = 200$ ,  $K = 1000$ ,  $\sigma_x^2 = 10^{-1}$  et  $\sigma^2 = 10^{-4}$ . Cette contrainte semble donc forte et peut sembler invalider les expressions données, cependant on a observé dans les simulations que quel que soit le contexte et les paramètres choisis, l'hypothèse concernant la bonne reconstruction du support est bien plus forte que l'hypothèse effectuée pour dériver les calculs des performances. En effet si on choisit une variance d'erreur plus importante que  $\sigma_e^2 = 10^{-3}$ , les algorithmes de reconstruction que nous avons testés, même placés dans les meilleures conditions que sont des retards bien espacés, des dictionnaires incohérents et un bruit additif faible, n'obtiennent pas une estimation du support exacte, permettant de mesurer ensuite les performances de l'ajout de post traitement BiCE. Une chose intéressante à noter est que pour des erreurs importantes, la BEQM exacte finit par donner de meilleures performances que la BEQM théorique qui sous-estime donc les performances de l'algorithme dans ce contexte là. Les bornes projetées et idéales ont également été tracées sur la figure 4.7, on peut noter que pour les mêmes  $\sigma_e^2$  que ceux pour lesquels l'approximation est valable, la BEQM et les deux bornes sont confondues et rejoignent les résultats de la borne idéale, qui considère l'absence d'erreur.

Pour observer le comportement de l'algorithme BiCE en pratique, on utilise les trois algorithmes de reconstruction parcimonieuse que nous avons également utilisés dans la section 3.4.4 (OMP, COSAMP et SPGL1) pour faire la première estimation du support, puis nous comparons les résultats sur l'estimation des amplitudes avec et sans correction BiCE. Nous considérons un nombre de source fixe et connu des algorithmes  $L = 10$ , et nous utilisons des paramètres pour lesquels nous observons que les résultats des différents algorithmes permettent d'extraire le bon support. Cette condition est importante pour visualiser uniquement le comportement du post traitement BiCE et le comparer à

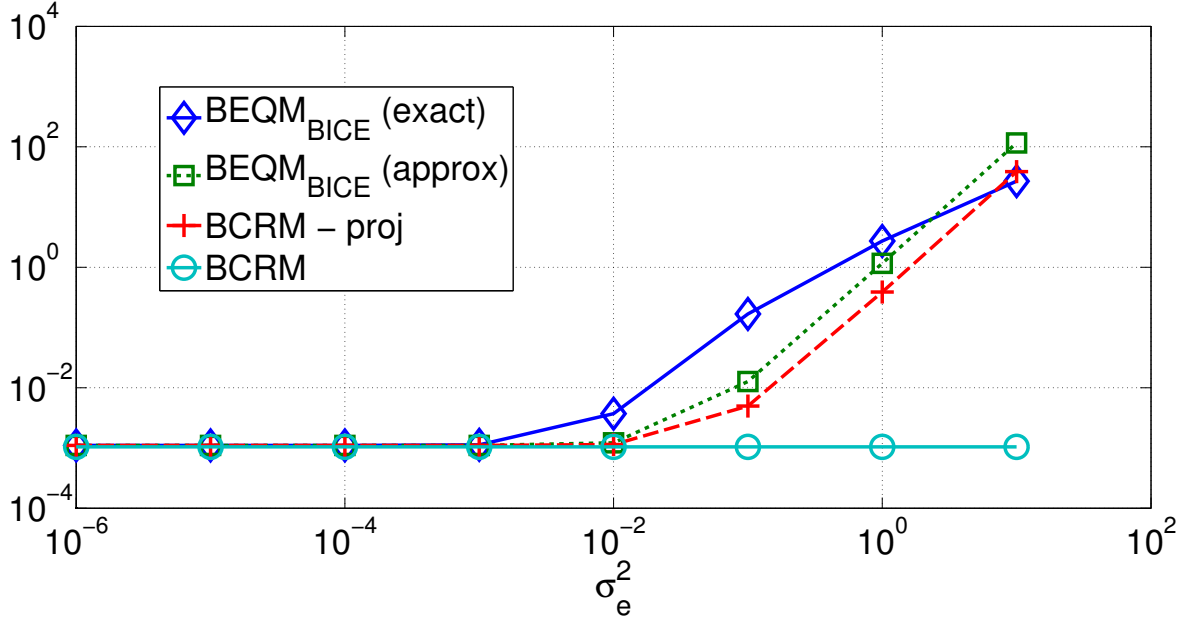


FIGURE 4.7 – Performance exacte et approximée comparée avec les bornes BCRM, en fonction de  $\sigma_e^2$ ,  $N = 200$ ,  $L = 10$ ,  $K = 1000$  et  $\sigma^2 = 10^{-4}$ .

la BEQM théorique qui considère qu'on connaît le support. Quand le support est mal estimé, on a observé que le post-traitement BiCE ne modifiait pas le résultat de l'estimation de l'amplitude et n'améliore pas les performances. Cependant quand le support est correctement estimé, on peut observer que le post traitement BiCE permet de pratiquement atteindre la borne idéale. Les résultats sont montrés pour deux jeux de paramètres différentes, la figure 4.8 considérant une matrice avec un degré de parcimonie plus important que la figure 4.9 mais pour qui les algorithmes sont donc plus sensibles à l'erreur de modèle et au bruit. C'est pourquoi on a choisi  $\sigma_e^2 = 10^{-6}$  dans la figure 4.8 et  $\sigma_e^2 = 10^{-5}$  dans la figure 4.9. Les courbes tracées correspondent aux résultats des trois estimateurs, premièrement sans aucune correction en ligne continue, et en pointillé avec la correction BiCE. Ces courbes sont tracées en fonction du RSB qui est défini par

$$\text{RSB} = \frac{\mathbb{E} \|\mathbf{x}\|^2}{\mathbb{E} \|\mathbf{w}\|^2} = \frac{\sigma_x^2}{\sigma^2} \quad (4.113)$$

où l'on a fait varier le bruit  $\sigma^2$ . Quel que soit l'algorithme utilisé en première étape, l'algorithme BiCE donne la même estimation de l'amplitude, c'est pourquoi on peut envisager de l'utiliser pour estimer les amplitudes d'un vecteur parcimonieux dont le support a été auparavant estimé par un algorithme spécialisé dans l'estimation du support quand on est en présence d'un bruit de dictionnaire structuré.

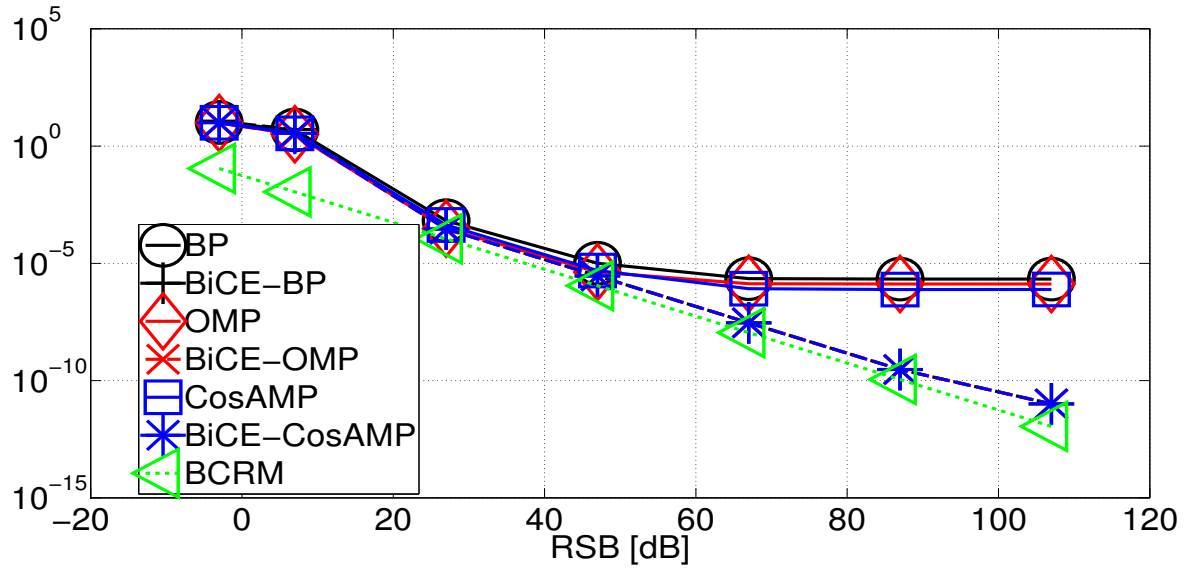


FIGURE 4.8 – Résultat de l'algorithme BiCE- $\mathcal{A}$  pour différents  $\mathcal{A}$ ,  $N = 200$ ,  $L = 10$ ,  $K = 1000$  et  $\sigma_e^2 = 10^{-6}$ .

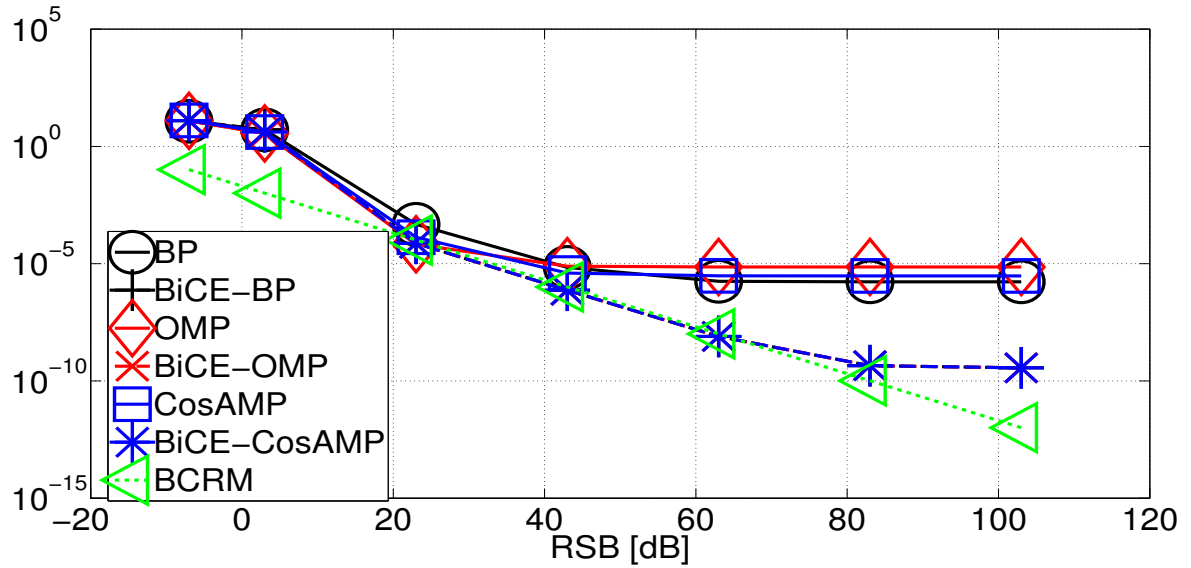


FIGURE 4.9 – Résultat de l'algorithme BiCE- $\mathcal{A}$  pour différents  $\mathcal{A}$ ,  $N = 500$ ,  $L = 10$ ,  $K = 1000$  et  $\sigma_e^2 = 10^{-5}$ .



# Chapitre 5

## Conclusion et Perspectives

Dans cette thèse nous nous sommes intéressés aux nouvelles méthodes d'échantillonnage qui, s'appuyant sur la parcimonie de la grande majorité des signaux pratiques, permettent de s'affranchir des limitations inhérentes au théorème de Shannon. Les deux méthodes utilisent deux manières différentes de modéliser la parcimonie des signaux, dépendant des connaissances à priori que l'on dispose, pour le reconstruire malgré l'utilisation d'un nombre d'échantillons réduit pris sur une fenêtre finie.

### 5.1 Échantillonnage de signaux impulsionnels

Les signaux impulsionnels, à bande non limitée, ne peuvent être échantillonnés au sens de Shannon. L'utilisation d'un noyau analogique avant l'échantillonnage permet de passer outre ce problème et d'obtenir des mesures permettant d'estimer les paramètres du modèle du signal. Ces paramètres étant d'un nombre fini, et le modèle étant parfaitement connu, il est alors possible de reconstruire parfaitement le signal, malgré le nombre réduit d'échantillons. Il a été démontré [127, 56] qu'en l'absence de bruit, l'estimation des paramètres du modèle pouvait être parfaite avec un nombre d'échantillons à peine supérieur au nombre de paramètres à estimer, sous la condition que le noyau de pré-filtrage utilisé respecte des conditions bien précises. Cette méthode d'échantillonnage peut se généraliser à toutes les formes de signaux à modèles paramétriques. Dans cette thèse nous avons étudié les performances d'estimation en dérivant les bornes de Cramer-Rao qui sont des bornes inférieures de l'erreur quadratique moyenne de tout estimateur non biaisé. Nous avons tout d'abord élargi la borne déterministe existante au cas à plusieurs impulsions, puis nous avons étudié les bornes bayésiennes pour deux modèles de signaux paramétriques différents. Premièrement nous avons considéré le modèle d'école des signaux à somme d'impulsions de Dirac, puis nous avons élargi notre étude à des formes d'impulsions quelconques échantillonnées avec un noyau de la famille des somme de sinus cardinaux (SoS).

La famille de noyaux SoS est un ensemble de noyaux paramétrables assurant de rester sous les conditions de reconstruction des signaux à modèle paramétrique, tout en permettant une variété importante de noyaux allant du noyau de Dirichlet à une sinusoïde parfaite. L'intérêt de cette famille de noyau réside dans sa formulation simple permettant d'isoler la contribution du paramètre choisi par l'utilisateur. Nous avons mis à profit cette propriété pour trouver le noyau SoS minimisant les bornes de Cramer Rao Bayésiennes. Minimiser les bornes inférieures sur l'erreur d'estimation permet de trouver le schéma

d'échantillonnage promettant les meilleures performances. Dans la suite de nos travaux il serait intéressant d'étudier les différents algorithmes d'estimation des signaux impulsionnels en utilisant les noyaux optimaux obtenus, et de développer un nouvel algorithme permettant d'exploiter au mieux les propriétés de ces noyaux. L'objectif est alors de trouver un algorithme atteignant les bornes optimisées, pour obtenir ainsi les meilleures performances possible pour le problème de l'estimation des signaux impulsionnels.

## 5.2 Robustesse des méthodes par dictionnaire

La reconstruction de signaux discrets à partir d'un nombre d'échantillons faible a fait l'objet d'un intérêt important de la communauté scientifique ces dernières années. Des algorithmes efficaces promouvant des solutions parcimonieuses ont été développés, la contrainte de parcimonie permettant de contre-balancer le faible nombre d'échantillons en choisissant la solution la plus réaliste du problème sur-complet. En effet, les signaux réalistes sont la plupart du temps parcimonieux, au moins dans une base donnée. L'inconvénient majeur de cette théorie est que ce soit une théorie portant sur des signaux discrets, alors que dans la pratique les signaux que l'on cherche à retrouver sont continus. Cette discrétisation induit une méconnaissance du dictionnaire au niveau de l'algorithme de reconstruction, qui induit une erreur d'estimation même pour des algorithmes parfaits et à haut rapport signal sur bruit. En effet, nous avons montré que les erreurs de modèle dégradent les bornes inférieures sur les performances que l'on peut attendre, et de ce fait une erreur d'estimation persiste même pour un algorithme optimal. Nous avons mis en évidence l'existence de deux modèles d'erreur différents dans le contexte Bayésien, souvent confondus car équivalents dans un cas déterministe. Le modèle que nous avons nommé "dictionnaire bruité" permet de prendre en compte les erreurs apparaissant uniquement au moment de la reconstruction, le signal échantillonné étant discret. Le problème de l'erreur de grille quant à lui modélise une méconnaissance dans la modélisation même du signal, qui peut être par exemple dû au fait de modéliser l'estimation d'un paramètre continu avec un dictionnaire discret. Nous avons mis en évidence que le problème de dictionnaire bruité est moins problématique que celui de l'erreur de grille, le biais induit par ce dernier étant inférieur. Nous avons également comparé les bornes bayésiennes de Van Trees et la borne de Cramer-Rao moyennée, pour le modèle de "l'erreur de grille" et montré que la BCRM permet de donner une bonne idée des résultats des algorithmes d'estimation pour des bruits où la BCRB décroche.

Il serait intéressant par la suite d'étudier la robustesse pour une erreur suivant une distribution uniforme plutôt que gaussienne, la distribution uniforme représentant mieux l'erreur de grille telle qu'on l'entends en général. Une autre perspective intéressante pour continuer d'étudier la robustesse des algorithmes de reconstruction parcimonieuse au problème des erreurs de modèle serait de lever l'hypothèse du support correctement estimé. Nous avons observé dans nos simulations que dans ce contexte l'erreur dû au support provient du bruit additif et ne dépend donc plus de l'erreur de modèle, cette intuition devrait être confirmée. Un dernier point à étudier concernant les bornes minimales est le niveau de bruit à partir duquel l'hypothèse de haut RSB n'est plus vérifiée, ce point semblant dépendre du nombre d'échantillons mesurés d'après nos simulations.

### 5.3 Correction du biais

En dernier lieu nous avons proposé deux algorithmes d'estimation post-traitement permettant de réduire l'effet de l'erreur de modèle à haut RSB. Dans la littérature, les différents algorithmes proposés pour corriger l'erreur de grille se concentrent sur une modification du coeur même des algorithmes de reconstruction parcimonieuse, pour en proposer des versions plus robustes au modèle erroné. De notre côté nous nous plaçons dans le cas où l'estimation du signal parcimonieux est déjà faite et nous corrigeons cette estimation pour réduire l'erreur due à la méconnaissance du modèle. Nous avons proposés deux algorithmes correspondant aux deux type d'erreurs de modèle rencontrés. Nous avons développés dans les deux cas les expressions des performances théoriques des algorithmes, et montrés leur efficacité en terme des mesures projetées. Nous avons également montré en pratique que ces algorithmes permettent d'améliorer l'estimation de l'amplitude des éléments non nuls du vecteur parcimonieux à estimer. Dans le contexte de l'estimation par dictionnaire structurée nous avons également montré que l'erreur due à la discrétisation était réduite par l'algorithme OGEC. Nous avons démontré que les algorithmes étaient efficaces au sens de la borne moyennée projetée, mais cette borne est supérieure à la BCRB et à la borne non projetée. De ce fait il reste un potentiel d'amélioration de ces algorithmes, pour développer une méthode permettant d'atteindre la borne. Les travaux sur les bornes sur l'erreur structurée et sur l'algorithme OGEC feront l'objet d'une publication de journal prochainement.

### 5.4 Perspectives générales

Nous avons étudié premièrement les noyaux de pré-filtrage permettant d'échantillonner des signaux paramétriques continus avec un faible nombre d'échantillons, puis nous nous sommes intéressés aux méthodes d'estimation et d'échantillonnage par dictionnaire. Dans le contexte de l'acquisition comprimée, les signaux que l'on cherche à reconstruire sont déjà échantillonnés puisque discrets. La théorie de l'acquisition comprimée a montré, tout comme celle de l'échantillonnage des signaux à taux d'innovation finie, que l'on pouvait réduire grandement le nombre d'échantillons et obtenir tout de même une reconstruction parfaite du signal. L'acquisition comprimée nécessite pour cela une matrice de mesure respectant des conditions précise, tandis que pour échantillonner des signaux à modèles paramétriques on utilise un noyau de pré-filtrage devant également respecter un certain nombre de conditions. Une perspective de recherche intéressante est de lier ces deux méthodes, en utilisant par exemple un pendant analogique de la matrice de mesure aléatoire pour déterminer un noyau de pré-filtrage général, permettant de mélanger aléatoirement les échantillons d'un signal analogique.

# Annexe A

## Liste des acronymes

**BCR** Borne de Cramer Rao (en anglais *Cramer Rao Bound*). 22, 24, 31, 33, 34, 36–39, 45, 46, 48, 49, 52, 93, 94, 97–99

**BCRB** Borne de Cramer Rao Bayésienne. 23, 55, 57, 63, 67, 83–85, 131

**BCRM** Borne de Cramer Rao Moyennée, en anglais ECRB pour *Expected Cramer Rao Bound*. iii, 23, 24, 93, 94, 96–99, 102, 128, 129, 131, 141–143, 146

**BEQM** Erreur Quadratique Moyenne Bayésienne (en anglais BMSE pour *Bayesian Mean Square Error*). 24, 26, 63, 93, 98, 103, 106–109, 113, 127, 128, 140

**BiCE** *Bias Correction Estimator*. 120, 133–135

**DB** Dictionnaire bruité. 81

**BP** *Basis Pursuit*. 18, 26, 120, 130, 146

**COSAMP** *Compressive Sensing Approximate Message Passing*. 130, 146

**EIV** *Error in Variables* ou en français erreur dans les variables. 27, 82, 83

**EQM** Erreur Quadratique Moyenne (en anglais MSE pour *Mean Square Error*). 22

**ISB** Rapport Interférence sur Bruit. 130

**LM MSE** *Linear Minimum Mean Square Estimator* ou en français Estimateur de l'erreur quadratique moyenne minimum linéaire. 26, 124–126, 130

**MIB** Matrice d'Information de Fisher Bayésienne. 24, 50, 51, 55, 58, 60, 61, 63, 83–85, 87, 88, 122, 123

**MIF** Matrice d'Information de Fisher déterministe (FIM en anglais). 24, 25, 49, 50, 58, 59, 65, 67, 69, 84, 85, 88, 93, 94, 101, 128, 129, 141

**EG** Erreur de grille. 81

**OGEC** *Off Grid Error Correction*. 120, 123

**OMP** *Orthogonal Matching Pursuit*. 10, 17, 26, 77, 120, 130, 134, 137, 146

**RIP** *Restricted Isometry Property*. 16, 19, 26, 81, 130

**RSB** Rapport Signal sur Bruit. 26, 32, 46, 55, 64, 72, 78, 80, 120, 121, 123, 130, 146

**SoS** noyau de sommes de sinus cardinaux, *Sum of Sincs* en anglais. 11, 12

# Bibliographie

- [1] M. AHARON, M. ELAD et A. BRUCKSTEIN. « K -SVD : An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation ». Dans : *IEEE Transactions on Signal Processing* 54.11 (nov. 2006), p. 4311–4322.
- [2] A. ALDROUBI, X CHEN et A. M. POWELL. « Perturbations of measurement matrices and dictionaries in compressed sensing ». Dans : *Applied and Computational Harmonic Analysis* 33.2 (2012), p. 282–291.
- [3] A. AMINI, U. KAMILOV et M. UNSER. « Bayesian denoising of generalized poisson processes with finite rate of innovation ». Dans : *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mar. 2012, p. 3629–3632.
- [4] A. ANGIERSKI et V. KUEHN. « Cramer-Rao bound for sampling and reconstruction of FRI signals ». Dans : *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mai 2014, p. 1793–1797.
- [5] C. D. AUSTIN, J. N. ASH et R. L. MOSES. « Dynamic Dictionary Algorithms for Model Order and Parameter Estimation ». Dans : *IEEE Transactions on Signal Processing* 61.20 (oct. 2013), p. 5117–5130.
- [6] B. BABADI, N. KALOUPSIDIS et V. TAROKH. « Asymptotic achievability of the Cramér-Rao bound for noisy compressive sampling ». Dans : *Signal Processing, IEEE Transactions on* 57.3 (2009), p. 1233–1236.
- [7] R. BARANIUK, M. DAVENPORT, R. DEVORE et M. WAKIN. « A simple proof of the restricted isometry property for random matrices ». Dans : *Constructive Approximation* 28.3 (2008), p. 253–263.
- [8] R. G. BARANIUK. « Compressive Sensing [Lecture Notes] ». Dans : *IEEE Signal Processing Magazine* 24.4 (juil. 2007), p. 118–121.
- [9] A. BECK et M. TEBOULLE. « A fast iterative shrinkage-thresholding algorithm for linear inverse problems ». Dans : *SIAM journal on imaging sciences* 2.1 (2009), p. 183–202.
- [10] R. T BEHRENS et L. L SCHARF. « Signal processing applications of oblique projection operators ». Dans : *Signal Processing, IEEE Transactions on* 42.6 (1994), p. 1413–1424.
- [11] Z. BEN-HAIM et Y.C. ELDAR. « A Lower Bound on the Bayesian MSE Based on the Optimal Bias Function ». Dans : *Information Theory, IEEE Transactions on* 55.11 (2009), p. 5179–5196.
- [12] Z. BEN-HAIM et Y.C. ELDAR. « The Cramér-Rao Bound for Estimating a Sparse Parameter Vector ». Dans : *Signal Processing, IEEE Transactions on* 58.6 (juin 2010), p. 3384–3389.

- [13] Z. BEN-HAIM, T. MICHAELI et Y. C. ELDAR. « Performance bounds and design criteria for estimating finite rate of innovation signals ». Dans : *Information Theory, IEEE Transactions on* 58.8 (2012), p. 4993–5015.
- [14] T. BENDORY, S. DEKEL et A. FEUER. « Robust recovery of stream of pulses using convex optimization ». Dans : *Journal of Mathematical Analysis and Applications* 442.2 (2016), p. 511–536.
- [15] T. BENDORY, A. BAR-ZION, D. ADAM, S. DEKEL et A. FEUER. « Stable Support Recovery of Stream of Pulses With Application to Ultrasound Imaging ». Dans : *IEEE Transactions on Signal Processing* 64.14 (juil. 2016), p. 3750–3759.
- [16] J. BERENT, P.-L. DRAGOTTI et T. BLU. « Sampling piecewise sinusoidal signals with finite rate of innovation methods ». Dans : *Signal Processing, IEEE Transactions on* 58.2 (2010), p. 613–625.
- [17] J. M. BERNARDO et A. F. M. SMITH. *Bayesian Theory*. John Wiley & Sons, Inc., 2008. ISBN : 9780470316870.
- [18] S. BERNHARDT et R. BOYER. « Caractérisation théorique du biais d'estimation du à l'erreur de grille dans le cadre de l'acquisition comprimée ». Dans : *Journée GDR" Représentations parcimonieuses, échantillonnage compressé et imagerie médicale"*. 2014.
- [19] S. BERNHARDT, R. BOYER, S. MARCOS et P. LARZABAL. « Compressed Sensing with Basis Mismatch : Performance Bounds and Sparse-Based Estimator ». Dans : *IEEE Transactions on Signal Processing* 64.13 (juil. 2016), p. 3483–3494.
- [20] S. BERNHARDT, R. BOYER, S. MARCOS et P. LARZABAL. « Compressed Sensing with uncertainty - the Bayesian estimation perspective ». Dans : *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*. Déc. 2015, p. 421–424.
- [21] S. BERNHARDT, R. BOYER, S. MARCOS, Y. C. ELDAR et P. LARZABAL. « Cramer-Rao Bound for finite streams of pulses ». Dans : *2014 22nd European Signal Processing Conference (EUSIPCO)*. Sept. 2014, p. 984–988.
- [22] S. BERNHARDT, R. BOYER, S. MARCOS, Y. ELDAR et P. LARZABAL. « Nouveaux noyaux pour l'échantillonnage de signaux impulsionnels à bande non-limitée ». Dans : *25ème édition du colloque GRETSI*. 2015.
- [23] S. BERNHARDT, R. BOYER, B. ZHANG, S. MARCOS et P. LARZABAL. « Performance analysis for sparse based biased estimator : Application to line spectra analysis ». Dans : *2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. Invited paper. Juin 2014, p. 365–368.
- [24] S. BERNHARDT, R. BOYER, S. MARCOS, Y. C. ELDAR et P. LARZABAL. « Sampling FRI signals with the SOS kernel : Bounds and optimal kernel ». Dans : *Signal Processing Conference (EUSIPCO), 2015 23rd European*. Août 2015, p. 2172–2176.
- [25] S. BERNHARDT, R. BOYER, S. MARCOS et P. LARZABAL. « Sparse-based estimators improvement in case of Basis mismatch ». Dans : *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*. Déc. 2015, p. 417–420.

- [26] M. A. BISCH, S. MIRON, D. BRIE et D. LEJEUNE. « A sparse approach for DOA estimation with a multiple spatial invariance sensor array ». Dans : *2016 IEEE Statistical Signal Processing Workshop (SSP)*. Juin 2016, p. 1–5.
- [27] T. BLU, P.-L. DRAGOTTI, M. VETTERLI, P. MARZILIANO et L. COULOT. « Sparse sampling of signal innovations ». Dans : *Signal Processing Magazine, IEEE* 25.2 (2008), p. 31–40.
- [28] G. BOULEUX, P. STOICA et R. BOYER. « An optimal Prior-knowLEDGE-based DOA estimation method ». Dans : *17th European Signal Processing Conf.* Glasgow, United Kingdom, 2009, p. 869–873.
- [29] R. BOYER. « Oblique projection for source estimation in a competitive environment : algorithm and statistical analysis ». Dans : *Signal Processing* 89.12 (2009), p. 2547–2554.
- [30] R. BOYER et G. BOULEUX. « Oblique projections for direction-of-arrival estimation with prior knowledge ». Dans : *IEEE Trans. Signal Processing* 56.4 (2008), p. 1374–1387.
- [31] R. BOYER et P. LARZABAL. « Sparsity in array processing : methods and performances ». Dans : *Special Session, 2014. IEEE/SP SAM Workshop*. IEEE. 2014.
- [32] R. BOYER, P. LARZABAL et B. H. FLEURY. « Oracle performance estimation of Bernoulli-distributed sparse vectors ». Dans : *2016 IEEE Statistical Signal Processing Workshop (SSP)*. Juin 2016, p. 1–4.
- [33] R. BOYER, B. BABADI, N. KALOUPSIDIS et V. TAROKH. « Errata to " Asymptotic Achievability of the Cramér-Rao Bound for Noisy Compressive Sampling " ». working paper or preprint. Avr. 2016. URL : <https://hal-supelec.archives-ouvertes.fr/hal-01299004>.
- [34] R. BOYER, R. COUILLET, B. H. FLEURY et P. LARZABAL. « Large-System Estimation Performance in Noisy Compressed Sensing With Random Support of Known Cardinality ; A Bayesian Analysis ». Dans : *IEEE Transactions on Signal Processing* 64.21 (nov. 2016), p. 5525–5535.
- [35] M. BROSSARD, M. N. EL KORSO, M. PESAVENTO, R. BOYER et P. LARZABAL. « Calibration of Radio Interferometers Using a Sparse DoA Estimation Framework ». Dans : *European Signal Processing Conference (EUSIPCO)* (2016).
- [36] E. J. CANDÉS et T. TAO. « Decoding by linear programming ». Dans : *Information Theory, IEEE Transactions on* 51.12 (2005), p. 4203–4215.
- [37] E. J. CANDÉS et T. TAO. « Near-Optimal Signal Recovery From Random Projections : Universal Encoding Strategies ? » Dans : *IEEE Transactions on Information Theory* 52.12 (déc. 2006).
- [38] E. J. CANDÉS et M. B. WAKIN. « An Introduction To Compressive Sampling ». Dans : *IEEE Signal Processing Magazine* 25.2 (mar. 2008), p. 21–30.
- [39] E.J. CANDÉS, J. ROMBERG et T. TAO. « Robust uncertainty principles : exact signal reconstruction from highly incomplete frequency information ». Dans : *Information Theory, IEEE Transactions on* 52.2 (2006), p. 489–509.
- [40] I. CARRON. *Compressive sensing : the Big Picture*. <https://sites.google.com/site/igorcarron2/cs#measurement>.

- [41] L. CHAARI, H. BATATIA, N. DOBIGEON et J. Y. TOURNERET. « A hierarchical sparsity-smoothness Bayesian model for l0-l1-l2 regularization ». Dans : *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mai 2014, p. 1901–1905.
- [42] D.H. CHAE, P. SADEGHI et R.A. KENNEDY. « Effects of basis-mismatch in compressive sampling of continuous sinusoidal signals ». Dans : *Future Computer and Communication (ICFCC), 2010 2nd International Conference on*. T. 2. 2010, p. V2–739–V2–743.
- [43] E. CHAUMETTE, J. GALY, A. QUINLAN et P. LARZABAL. « A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum Likelihood Estimators ». Dans : *IEEE Transactions on Signal Processing* 56.11 (nov. 2008), p. 5319–5333.
- [44] S. CHEN, D. DONOHO et M. SAUNDERS. « Atomic Decomposition by Basis Pursuit ». Dans : *SIAM Journal on Scientific Computing* 20.1 (1998), p. 33–61.
- [45] Y. CHI, L.L. SCHARF, A. PEZESHIKI et A.R. CALDERBANK. « Sensitivity to Basis Mismatch in Compressed Sensing ». Dans : *Signal Processing, IEEE Transactions on* 59.5 (2011), p. 2182–2195.
- [46] N. CHU, A. MOHAMMAD-DJAFARI et J. PICHERAL. « Robust Bayesian super-resolution approach via sparsity enforcing a priori for near-field aeroacoustic source imaging ». Dans : *Journal of Sound and Vibration* 332.18 (2013), p. 4369–4389.
- [47] C. R. COMSA et A. M. HAIMOVICH. « Performance bound for time delay and amplitude estimation from low rate samples of pulse trains ». Dans : *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. IEEE. 2012, p. 455–459.
- [48] L. CONDAT, A. HIRABAYASHI et Y. HIRONAGA. « Recovery of nonuniformdirac pulses from noisy linear measurements ». Dans : *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. Mai 2013, p. 6014–6018.
- [49] R. COUILLET et M. DEBBAH. *Random matrix methods for wireless communications*. Cambridge University Press, 2011.
- [50] Z. DOĞAN, C. GILLIAM, T. BLU et D. VAN DE VILLE. « Reconstruction of Finite Rate of Innovation Signals with Model-Fitting Approach ». Dans : *IEEE Transactions on Signal Processing* 63.22 (nov. 2015), p. 6024–6036.
- [51] N. DOBIGEON, A. BASARAB, D. KOUAMÉ et J. Y. TOURNERET. « Regularized Bayesian compressed sensing in ultrasound imaging ». Dans : *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. Août 2012, p. 2600–2604.
- [52] Y. DODGE. *Statistique : Dictionnaire encyclopédique*. Mathematics and Statistics. Springer-Verlag, 2007. ISBN : 9782287720949.
- [53] D. DONOHO et J. TANNER. « Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing ». Dans : *Philosophical Transactions of the Royal Society of London A : Mathematical, Physical and Engineering Sciences* 367.1906 (2009), p. 4273–4293.
- [54] D. L. DONOHO, A. MALEKI et A. MONTANARI. « Message-passing algorithms for compressed sensing ». Dans : *Proceedings of the National Academy of Sciences* 106.45 (2009), p. 18914–18919.

- [55] P. L. DRAGOTTI et F. HOMANN. « Sampling signals with finite rate of innovation in the presence of noise ». Dans : *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. Avr. 2009, p. 2941–2944.
- [56] P. L. DRAGOTTI, M. VETTERLI et T. BLU. « Sampling Moments and Reconstructing Signals of Finite Rate of Innovation : Shannon Meets Strang-Fix ». Dans : *IEEE Transactions on Signal Processing* 55.5 (mai 2007), p. 1741–1757.
- [57] M.F. DUARTE et Y.C. ELDAR. « Structured Compressed Sensing : From Theory to Applications ». Dans : *Signal Processing, IEEE Transactions on* 59.9 (2011), p. 4053–4085.
- [58] M. N. EL KORSO, R. BOYER, P. LARZABAL et B. H. FLEURY. « Estimation Performance for the Bayesian Hierarchical Linear Model ». Dans : *IEEE Signal Processing Letters* 23.4 (avr. 2016), p. 488–492.
- [59] Y. C. ELDAR. « Compressed Sensing of Analog Signals in Shift-Invariant Spaces ». Dans : *IEEE Transactions on Signal Processing* 57.8 (août 2009), p. 2986–2997.
- [60] Y. C. ELDAR et T. MICHAELI. « Beyond bandlimited sampling ». Dans : *IEEE Signal Processing Magazine* 26.3 (mai 2009), p. 48–68.
- [61] V. EMIYA, A. BONNEFOY, L. DAUDET et R. GRIBONVAL. « Compressed sensing with unknown sensor permutation ». Dans : *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mai 2014, p. 1040–1044.
- [62] A. ERDOZAIN et P. M. CRESPO. « Reconstruction of streams of impulses from quantized samples using a stochastic algorithm based on Genetic Algorithms ». Dans : *Sarnoff Symposium, 2009. SARNOFF '09. IEEE*. Mar. 2009, p. 1–5.
- [63] X.-B. GAO, G. H GOLUB et L.-Z. LIAO. « Continuous methods for symmetric generalized eigenvalue problems ». Dans : *Linear Algebra and its Applications* 428.2 (2008), p. 676–696.
- [64] C. GILLIAM et T. BLU. « Fitting instead of annihilation : Improved recovery of noisy FRI signals ». Dans : *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mai 2014, p. 51–55.
- [65] S. GLEICHMAN et Y. C. ELDAR. « Blind Compressed Sensing ». Dans : *IEEE Transactions on Information Theory* 57.10 (oct. 2011), p. 6958–6975.
- [66] G. H GOLUB et C. F VAN LOAN. *Matrix computations*. T. 3. JHU Press, 2012.
- [67] J. H. de M. GOULART, M. BOIZARD, R. BOYER, G. FAVIER et P. COMON. « Tensor CP Decomposition With Structured Factor Matrices : Algorithms and Performance ». Dans : *IEEE Journal of Selected Topics in Signal Processing* 10.4 (juin 2016), p. 757–769.
- [68] M. GRANT et S. BOYD. *CVX : Matlab Software for Disciplined Convex Programming, version 2.1*. <http://cvxr.com/cvx>. Mar. 2014.
- [69] M. GRANT et S. BOYD. « Graph implementations for nonsmooth convex programs ». Dans : *Recent Advances in Learning and Control*. Sous la dir. de V. BLONDEL, S. BOYD et H. KIMURA. Lecture Notes in Control and Information Sciences. Springer-Verlag Limited, 2008, p. 95–110.
- [70] M. A. HERMAN et T. STROHMER. « High-Resolution Radar via Compressed Sensing ». Dans : *IEEE Transactions on Signal Processing* 57.6 (juin 2009), p. 2275–2284.

- [71] M.A. HERMAN et T. STROHMER. « General Deviants : An Analysis of Perturbations in Compressed Sensing ». Dans : *Selected Topics in Signal Processing, IEEE Journal of* 4.2 (2010), p. 342–349.
- [72] G. HUANG, N. FU, J. ZHANG et L. QIAO. « Sparsity-based reconstruction method for signals with finite rate of innovation ». Dans : *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mar. 2016, p. 4503–4507.
- [73] T. HUANG, Y. LIU, H. MENG et X. WANG. « Adaptive Compressed Sensing via Minimizing Cramer-Rao Bound ». Dans : *IEEE Signal Processing Letters* 21.3 (mar. 2014), p. 270–274.
- [74] R. JAGANNATH et K.V.S. HARI. « Block Sparse Estimator for Grid Matching in Single Snapshot DoA Estimation ». Dans : *Signal Processing Letters, IEEE* 20.11 (2013), p. 1038–1041.
- [75] J. L. W. V. JENSEN. « Sur les fonctions convexes et les inégalités entre les valeurs moyennes ». Dans : *Acta mathematica* 30.1 (1906), p. 175–193.
- [76] S. JI, Y. XUE et L. CARIN. « Bayesian Compressive Sensing ». Dans : *IEEE Transactions on Signal Processing* 56.6 (juin 2008), p. 2346–2356.
- [77] A. JUNG, Z. BEN-HAIM, F. HLAWATSCH et Y. C. ELDAR. « On unbiased estimation of sparse vectors corrupted by Gaussian noise ». Dans : *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. Mar. 2010, p. 3990–3993.
- [78] S. M. KAY. *Fundamentals of Statistical Signal Processing : Detection Theory*. T. II. Upper Saddle River, New Jersey, 07458 : Prentice-Hall, 1998.
- [79] H. KRIM, P. FORSTER et J. G. PROAKIS. « Operator approach to performance analysis of root-MUSIC and root-min-norm ». Dans : *Signal Processing, IEEE Transactions on* 40.7 (1992), p. 1687–1696.
- [80] F. KRZAKALA, M. MÉZARD et L. ZDEBOROVÁ. « Compressed sensing under matrix uncertainty : Optimum thresholds and robust approximate message passing ». Dans : *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. Mai 2013, p. 5519–5523.
- [81] F. LE CHEVALIER. *Principes de traitement des signaux radar et sonar*. Masson, 1989.
- [82] D. MALIOUTOV, M. CETIN et A. S. WILLSKY. « A sparse signal reconstruction perspective for source localization with sensor arrays ». Dans : *IEEE Transactions on Signal Processing* 53.8 (août 2005), p. 3010–3022.
- [83] V. A MARČENKO et L. A. PASTUR. « Distribution of eigenvalues for some sets of random matrices ». Dans : *Sbornik : Mathematics* 1.4 (1967), p. 457–483.
- [84] T. MICHAELI et Y. C. ELDAR. « Optimization-based recovery from rate of innovation samples ». Dans : *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mar. 2012, p. 3649–3652.
- [85] T. MICHAELI et Y. C. ELDAR. « Xampling at the Rate of Innovation ». Dans : *IEEE Transactions on Signal Processing* 60.3 (mar. 2012), p. 1121–1133.
- [86] A. MOHAMMAD-DJAFARI et M. DUMITRU. « Bayesian sparse solutions to linear inverse problems with non-stationary noise with Student-t priors ». Dans : *Digital Signal Processing* 47 (2015). Special Issue in Honour of William J. (Bill) Fitzgerald, p. 128–156.

- [87] S. MULLETI, B. A. SHENOY et C. S. SEELAMANTULA. « FRI Sampling on Structured Nonuniform Grids ;Application to Super-Resolved Optical Imaging ». Dans : *IEEE Transactions on Signal Processing* 64.15 (août 2016), p. 3841–3853.
- [88] S. NADARAJAH. « A generalized normal distribution ». Dans : *Journal of Applied Statistics* 32.7 (2005), p. 685–694.
- [89] D. NEEDELL et R. VERSHYNIN. « Signal Recovery From Incomplete and Inaccurate Measurements Via Regularized Orthogonal Matching Pursuit ». Dans : *Selected Topics in Signal Processing, IEEE Journal of* 4.2 (2010), p. 310–316.
- [90] R. NIAZADEH, M. BABAIE-ZADEH et C. JUTTEN. « On the achievability of Cramer–Rao bound in noisy compressed sensing ». Dans : *Signal Processing, IEEE Transactions on* 60.1 (2012), p. 518–526.
- [91] J. OÑATIVIA, Y. M. LU et P. L. DRAGONI. « Finite dimensional FRI ». Dans : *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mai 2014, p. 1808–1812.
- [92] V. OLLIER, R. BOYER, M. N. EL KORSO et P. LARZABAL. « Bayesian Lower Bounds for Dense or Sparse (Outlier) Noise in the RMT Framework ». Dans : *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*. invited article, 2016.
- [93] P. PAKROOH, L. L. SCHARF et A. PEZESHKI. « Threshold Effects in Parameter Estimation From Compressed Data ». Dans : *IEEE Transactions on Signal Processing* 64.9 (mai 2016), p. 2345–2354.
- [94] P. PAKROOH, A. PEZESHKI, L. L. SCHARF, D. COCHRAN et S. D. HOWARD. « Analysis of Fisher Information and the Cramer-Rao Bound for Nonlinear Parameter Estimation After Random Compression ». Dans : *IEEE Transactions on Signal Processing* 63.23 (déc. 2015), p. 6423–6428.
- [95] S. PARIS, D. MARY et A. FERRARI. « Detection Tests Using Sparse Models, With Application to Hyperspectral Data ». Dans : *IEEE Transactions on Signal Processing* 61.6 (mar. 2013), p. 1481–1494.
- [96] J. T. PARKER, V. CEVHER et P. SCHNITER. « Compressive sensing under matrix uncertainties : An Approximate Message Passing approach ». Dans : *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. Nov. 2011, p. 804–808.
- [97] Y. C. PATI, R. REZAIEFAR, Y. C. PATI et P. S. KRISHNAPRASAD. « Orthogonal Matching Pursuit : Recursive Function Approximation with Applications to Wavelet Decomposition ». Dans : *Proceedings of the 27 th Annual Asilomar Conference on Signals, Systems, and Computers*. 1993, p. 40–44.
- [98] K. B. PETERSEN et M. S. PEDERSEN. « The matrix cookbook ». Dans : *Technical University of Denmark* 7 (2008), p. 15.
- [99] M. A. PINSKY. *Introduction to Fourier analysis and wavelets*. T. 102. American Mathematical Society, 2002.
- [100] D. RAMASAMY, S. VENKATESWARAN et U. MADHOW. « Compressive Parameter Estimation in AWGN ». Dans : *IEEE Transactions on Signal Processing* 62.8 (avr. 2014), p. 2012–2027.

- [101] C. REN, J. GALY, E. CHAUMETTE, P. LARZABAL et A. RENAUX. « Hybrid Barankin-Weiss-Weinstein Bounds ». Dans : *IEEE Signal Processing Letters* 22.11 (nov. 2015), p. 2064–2068.
- [102] RICEDSP. *Compressive imagine : a new single-pixel camera*. <http://dsp.rice.edu/cscamera>.
- [103] M. ROSENBAUM et A. B. TSYBAKOV. « Sparse recovery under matrix uncertainty ». Dans : *The Annals of Statistics* 38.5 (2010), p. 2620–2651.
- [104] S. SAHNOUN, E.-H. DJERMOUNE, D. BRIE et P. COMON. « A simultaneous sparse approximation method for multidimensional harmonic retrieval ». Dans : *Signal Processing* 131 (2017), p. 36–48.
- [105] S. SAHNOUN, E. H. DJERMOUNE, C. SOUSSEN et D. BRIE. « Sparse multiresolution modal estimation ». Dans : *2011 IEEE Statistical Signal Processing Workshop (SSP)*. Juin 2011, p. 309–312.
- [106] G. SCHWARZ. « Estimating the Dimension of a Model ». Dans : *The Annals of Statistics* 6.2 (1978), p. 461–464.
- [107] M. SHAGHAGHI et S. A. VOROBYOV. « Cramer-Rao Bound for Sparse Signals Fitting the Low-Rank Model with Small Number of Parameters ». Dans : *IEEE Signal Processing Letters* 22.9 (sept. 2015), p. 1497–1501.
- [108] S. SHAHBAZPANAH, A. B. GERSHMAN, Z.-Q. LUO et K. M. WONG. « Robust adaptive beamforming for general-rank signal models ». Dans : *Signal Processing, IEEE Transactions on* 51.9 (2003), p. 2257–2269.
- [109] C. E. SHANNON. « A Mathematical Theory of Communication ». Dans : *The Bell System Technical Journal* 27 (1948), p. 379–423 ; 623–656.
- [110] C. SOUSSEN, J. IDIER, D. BRIE et J. DUAN. « From Bernoulli-Gaussian Deconvolution to Sparse Signal Restoration ». Dans : *IEEE Transactions on Signal Processing* 59.10 (oct. 2011), p. 4572–4584.
- [111] P. STOICA et R.L. MOSES. *Spectral analysis of signals*. Pearson Prentice Hall, 2005. ISBN : 9780131139565.
- [112] G. STRANG et G. FIX. « A Fourier analysis of the finite element variational method ». Dans : *Constructive Aspect of Functional Analysis* (1971), p. 796–830.
- [113] V. Y. F. TAN et V. K. GOYAL. « Estimating Signals With Finite Rate of Innovation From Noisy Samples : A Stochastic Algorithm ». Dans : *IEEE Transactions on Signal Processing* 56.10 (oct. 2008), p. 5135–5146.
- [114] Z. TAN, P. YANG et A. NEHORAI. « Joint Sparse Recovery Method for Compressed Sensing With Structured Dictionary Mismatches ». Dans : *Signal Processing, IEEE Transactions on* 62.19 (oct. 2014), p. 4997–5008.
- [115] Y. TANG, L. CHEN et Y. GU. « On the Performance Bound of Sparse Estimation With Sensing Matrix Perturbation ». Dans : *Signal Processing, IEEE Transactions on* 61.17 (sept. 2013), p. 4372–4386.
- [116] O. TEKE, A. C. GURBUZ et O. ARIKAN. « Perturbed Orthogonal Matching Pursuit ». Dans : *IEEE Transactions on Signal Processing* 61.24 (déc. 2013).

- [117] R. TIBSHIRANI. « Regression shrinkage and selection via the lasso ». Dans : *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), p. 267–288.
- [118] J. A. TROPP et S. J. WRIGHT. « Computational Methods for Sparse Solution of Linear Inverse Problems ». Dans : *Proceedings of the IEEE* 98.6 (juin 2010), p. 948–958.
- [119] A. M. TULINO et S. VERDÚ. « Random Matrix Theory and Wireless Communications ». Dans : *Commun. Inf. Theory* 1.1 (juin 2004), p. 1–182. ISSN : 1567-2190.
- [120] R. TUR, Y. C. ELDAR et Z. FRIEDMAN. « Innovation rate sampling of pulse streams with application to ultrasound imaging ». Dans : *Signal Processing, IEEE Transactions on* 59.4 (2011), p. 1827–1842.
- [121] M. UNSER. « Sampling-50 years after Shannon ». Dans : *Proceedings of the IEEE* 88.4 (avr. 2000), p. 569–587.
- [122] J. A. URIGÜEN, Y. C. ELDAR, P. L. DRAGOTTI et Zvika B.-H. « Sampling at the rate of innovation : theory and applications ». Dans : *Compressed Sensing: Theory and Applications*. Sous la dir. d'Y. C. ELDAR et G. KUTYNIOK. Cambridge University Press, 2012. Chap. 4, p. 148.
- [123] J. A. URIGUEN, T. BLU et P.-L. DRAGOTTI. « FRI sampling with arbitrary kernels ». Dans : *Signal Processing, IEEE Transactions on* 61.21 (2013), p. 5310–5323.
- [124] E. VAN DEN BERG et M. P. FRIEDLANDER. « Probing the Pareto frontier for basis pursuit solutions ». Dans : *SIAM Journal on Scientific Computing* 31.2 (2008), p. 890–912.
- [125] E. VAN DEN BERG et M. P. FRIEDLANDER. *SPGL1 : A solver for large-scale sparse reconstruction*. Juin 2007. URL : <http://www.cs.ubc.ca/labs/scl/spgl1>.
- [126] H. L. VAN TREES. *Detection, estimation, and modulation theory*. John Wiley & Sons, 2004.
- [127] M. VETTERLI, P. MARZILIANO et T. BLU. « Sampling signals with finite rate of innovation ». Dans : *Signal Processing, IEEE Transactions on* 50.6 (2002), p. 1417–1428.
- [128] J. WANG, S. KWON et B. SHIM. « Generalized orthogonal matching pursuit ». Dans : *Signal Processing, IEEE Transactions on* 60.12 (2012), p. 6202–6216.
- [129] X. WEI, T. BLU et P.-L. DRAGOTTI. « Finite rate of innovation with non-uniform samples ». Dans : *Signal Processing, Communication and Computing (ICSPCC), 2012 IEEE International Conference on*. IEEE. 2012, p. 369–372.
- [130] X. WEI et P. L. DRAGOTTI. « Guaranteed Performance in the FRI Setting ». Dans : *IEEE Signal Processing Letters* 22.10 (oct. 2015), p. 1661–1665.
- [131] A. WIESEL, Y.C. ELDAR et A. YEREDOR. « Linear Regression With Gaussian Model Uncertainty : Algorithms and Bounds ». Dans : *Signal Processing, IEEE Transactions on* 56.6 (juin 2008), p. 2194–2205.
- [132] E. P WIGNER. « Characteristic Vectors of Bordered Matrices With Infinite Dimensions ». Dans : *Annals of Mathematics* (1955), p. 548–564.
- [133] J. WISHART. « The generalised product moment distribution in samples from a normal multivariate population ». Dans : *Biometrika* (1928), p. 32–52.

- [134] Z. YANG, C. ZHANG et L. XIE. « Robustly Stable Signal Recovery in Compressed Sensing With Structured Matrix Perturbation ». Dans : *Signal Processing, IEEE Transactions on* 60.9 (2012), p. 4658–4671.
- [135] H. ZHU, G. LEUS et G.B. GIANNAKIS. « Sparsity-Cognizant Total Least-Squares for Perturbed Compressive Sampling ». Dans : *Signal Processing, IEEE Transactions on* 59.5 (2011), p. 2002–2016.
- [136] S. ZHU, A. MOHAMMAD-DJAFARI, H. WANG, B. DENG, X. LI et J. MAO. « Parameter estimation for SAR micromotion target based on sparse signal representation ». Dans : *EURASIP Journal on Advances in Signal Processing* Special issue on Sparse Signal Processing (18 January 2012) (2012), p. 2012–13.