



HAL
open science

Caractérisation des limites fondamentales de l'erreur quadratique moyenne pour l'estimation de signaux comportant des points de rupture

Lucien Bacharach

► **To cite this version:**

Lucien Bacharach. Caractérisation des limites fondamentales de l'erreur quadratique moyenne pour l'estimation de signaux comportant des points de rupture. Traitement du signal et de l'image [eess.SP]. Université Paris-Saclay, 2018. Français. NNT : 2018SACLS322 . tel-01957951

HAL Id: tel-01957951

<https://theses.hal.science/tel-01957951>

Submitted on 17 Dec 2018

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.

Caractérisation des limites fondamentales de l'erreur quadratique moyenne pour l'estimation de signaux comportant des points de rupture

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

École doctorale n°580 Sciences et technologies de l'information et de la
communication (STIC)
Spécialité de doctorat : Traitement du signal et des images

Thèse présentée et soutenue à Gif-sur-Yvette, le 28 septembre 2018, par

LUCIEN BACHARACH

Composition du Jury :

M. Olivier MICHEL, Professeur des Universités INP de Grenoble – GIPSA-Lab	Président
M. David BRIE, Professeur des Universités Université de Lorraine – CRAN	Rapporteur
M. Nicolas DOBIGEON, Professeur des Universités Université de Toulouse – IRIT/INP-ENSEEIH	Rapporteur
M. Claude DELPHA, Maître de Conférences Université Paris-Sud (Université Paris-Saclay) – L2S	Examineur
M. André FERRARI, Professeur des Universités Université de Nice Sophia Antipolis – Laboratoire J. L. Lagrange	Examineur
Mme Anne GÉGOUT-PETIT, Professeure des Universités Université de Lorraine – IECL	Examinatrice
M. Alexandre RENAUX, Maître de Conférences Université Paris-Sud (Université Paris-Saclay) – L2S	Directeur de thèse
M. Mohammed Nabil EL KORSO, Maître de Conférences Université Paris Nanterre – LEME	Co-directeur de thèse

UNIVERSITÉ PARIS-SACLAY

ECOLE DOCTORALE STIC

« Sciences et Technologies de l'Information et de la Communication »

THÈSE DE DOCTORAT

Spécialité : Traitement du Signal et des Images

Soutenue le 28 septembre 2018 à Gif-sur-Yvette

par

Lucien BACHARACH

Caractérisation des limites fondamentales de l'erreur quadratique moyenne pour l'estimation de signaux comportant des points de rupture

Composition du jury :

Président du jury :	M. Olivier MICHEL	Professeur des Universités GIPSA-Lab, INP de Grenoble
Rapporteurs :	M. David BRIE	Professeur des Universités CRAN, Université de Lorraine
	M. Nicolas DOBIGEON	Professeur des Universités IRIT/INP-ENSEEIH, Université de Toulouse
Examineurs :	M. Claude DELPHA	Maître de Conférences L2S, Université Paris-Sud (Université Paris-Saclay)
	M. André FERRARI	Professeur des Universités Laboratoire J. L. Lagrange, Université Nice Sophia Antipolis
	Mme Anne GÉGOUT-PETIT	Professeure des Universités IECL, Université de Lorraine
Directeurs de thèse :	M. Alexandre RENAUX	Maître de Conférences L2S, Université Paris-Sud (Université Paris-Saclay)
	M. Mohammed Nabil EL KORSO	Maître de Conférences LEME, Université Paris Nanterre

Université Paris-Sud

Laboratoire des Signaux et Systèmes
CentraleSupélec / Université Paris-Sud / CNRS / UMR 8506
3, rue Joliot-Curie, 91190 GIF-SUR-YVETTE, France

à mes grands-parents et mes parents

Remerciements

J'ai ce sentiment étrange : qu'il s'est passé énormément de choses depuis le début de cette thèse, mais aussi que tout est allé très vite, trop vite... Je souhaiterais remercier les nombreuses personnes qui m'ont accompagné tout au long de ces trois années, et qui les ont rendues aussi enrichissantes.

Mes premiers remerciements vont à ceux qui m'ont accordé leur confiance et m'ont permis de vivre cette expérience, à savoir mes encadrants de thèse, Alexandre Renaux et Mohammed Nabil El Korso. Je ne vous remercierai jamais assez pour tout ce que vous avez fait : votre implication, votre dynamisme, votre soutien en toutes circonstances... Sachez que vous êtes bien plus que mes directeurs ou encadrants. Alex, merci pour ces innombrables échanges, discussions sans fin (scientifiques ou non), et bien sûr, pour tous ces voyages incroyables ! Nabil, merci pour ton enthousiasme, ta disponibilité à *toute heure*, ainsi que tes nombreux et précieux conseils. J'ai eu une chance inestimable de travailler avec vous deux, et j'espère en avoir de nouveau l'occasion, un jour. D'ici là, j'espère qu'on se croquera, à Paris, à Toulouse, en conférence, ou ailleurs !

Je voudrais remercier tous les membres du jury de cette thèse : David Brie et Nicolas Dobigeon pour avoir accepté de relire ce manuscrit et pour leurs commentaires éclairés ; Anne Gégout-Petit, Claude Delpha ainsi qu'André Ferrari pour avoir bien voulu en être les examinateurs ; et bien sûr Olivier Michel pour avoir présidé ce jury.

Je tiens également à adresser un remerciement particulier à mes encadrants de stage au LEGOS, Marie-Claude Gennero et Stéphane Calmant, qui m'ont permis d'obtenir un premier aperçu du monde de la recherche et de mûrir le projet d'une thèse en traitement du signal.

Cette thèse s'est déroulée au sein du pôle Signaux et Statistiques du Laboratoire des Signaux et Systèmes, dans l'équipe Modélisation et Estimation. Je tiens à remercier toutes les personnes qui ont fait que c'était un plaisir d'y travailler : tout d'abord Maryvonne Giron de m'y avoir accueilli ; Silviu Niculescu ainsi que Pascal Bondon, pour leur volonté constante d'obtenir le meilleur pour le laboratoire ; sans oublier Laurence Antunes, Delphine Maucherat, Pascale Debever et Stéphanie Douesnard pour leur aide précieuse dans les diverses tâches administratives, ainsi que Céline Labrude en matière d'informatique. Merci également à Rémy Boyer pour son intérêt, au *Professeur* Frédéric Pascal pour son obligeance (et ses sarcasmes), et à Gilles Chardon pour les discussions sur les ellipsoïdes et tous les moments conviviaux.

De nombreuses personnes du L2S ont aussi grandement contribué à ce que je m'y sente chez moi, à commencer par Chengfang, thésard à l'époque, qui m'a accueilli dans notre bureau dès mon arrivée en stage ; toutes celles et ceux que j'y ai rencontrés et en ont toujours fait un lieu de bonne humeur générale : Stéphanie, Eugénie, Virginie, Yassine, Gia Thuy, Abdelhak, Violeta ; Geoffrey, ce stagiaire intrépide qui n'a pas reculé devant les bornes de Weiss-Weinstein ; et spécialement Gordana, pour son enthousiasme communicatif, en tant que co-représentante des doctorants, mais aussi et surtout en tant qu'amie. La bonne ambiance dans laquelle se sont déroulées ces trois années n'était évidemment pas confinée aux murs de ce seul bureau, et je tiens donc à remercier en particulier la joyeuse bande du 3^e : Maria, Walid, François, Yoan, Lucie, et bien sûr, cet homme

infiniment bon et généreux qu'on appelle "le Guigui", c'est-à-dire Guillaume Rance. Tous ces moments passés au sport, à discuter autour d'une bière, d'un boeuf sauté au poivre vert, ou d'un bô-bun + thé glacé au citron vont me manquer, Guigui. C'est aussi grâce à toi que j'ai rencontré un certain Hugo, avec qui nous avons passé quelques moments (plus ou moins) mémorables. Je remercie aussi Dina, Olivier, Diane, Wafae, Andreea, Ayoub, Alina, Amine, Hidayet, Mircea, Kuba, Caetano, Zohra pour les nombreux cafés, moments de détente, et repas partagés au CESFO ou ailleurs ; ainsi que Iulia, que j'admire pour son éthique et ses convictions inébranlables.

Je souhaiterais aussi remercier plusieurs personnes qu'il m'a été donné de rencontrer durant ces années, qui ont contribué à construire ma perception du métier d'enseignant-chercheur, ou qui m'ont témoigné de l'amitié : d'abord Éric Chaumette, pour son regard bienveillant sur mon travail, sa vision personnelle du monde de la recherche, et pour m'avoir ouvert ses portes à la suite de cette thèse ; et Jean-Yves Tournet, pour les discussions fructueuses qui ont permis notre collaboration, pour ses commentaires pertinents, et pour ses travaux, en particulier avec André Ferrari puis avec Nicolas Dobigeon, qui ont inspiré ceux de cette thèse. Je ne garderais pas un aussi bon souvenir de la thèse s'il n'y avait pas eu tous les bons moments passés aux quatre coins du monde — de Majorque à la Nouvelle-Orléans en passant par Peyresq — avec Joana, Guillaume Ginolhac, Dom, Arnaud, Elisa, Jean-Phi, Agnès, Frenzy, Pascal, Ammar et beaucoup d'autres, dont un bon nombre que j'ai déjà cités.

Mes remerciements vont aussi à l'équipe pédagogique de Polytech' Paris-Sud qui m'a accueilli et m'a fait confiance pour ma mission d'enseignement, c'est-à-dire en particulier Claude Delpha, Dima Rodriguez, Marc Bachelier, Antoine Diet, Patrick Ruiz, Sylvie Dugard, ainsi qu'Anthony Juton, pour son aide et ses supports sur les microcontrôleurs.

Au-delà du laboratoire, je voudrais également remercier mes amis sur le campus, ceux de SONDRRA et du GeePs, ceux de l'ADSS puis de l'ATACS pour les événements sympas organisés ; en particulier Simon Meunier, pour les nombreux délires, le week-end chez mamie et tous les matches de tennis (j'espère qu'un jour tu seras assez fort pour me battre). Et bien sûr, au-delà du campus, je remercie tous mes amis dont la connaissance remonte à plus ou moins longtemps avant la thèse, que j'ai eu la chance de revoir, même un peu, au cours des ces trois années. Alexis Krauzé : trop de choses à te raconter, on doit se programmer un bon kebab quand je repasse à Paris ! Tous mes amis rencontrés au LEGOS, en particulier Maud, Charlotte, Robin, Vincent, Sylvain : c'est tellement bon de vous revoir régulièrement, et on n'oublie pas, objectif Guadeloupe ! Mes potes de l'ENSICA, a fortiori les Ensiconnus : Pierre, Bertrand et Benjamin ; Julien, Sylvain, Lucas : ça m'a fait plaisir de vous recroiser à Toulouse à l'aube de 2018, on se revoit forcément bientôt ; Tim : ce brunch à Paris était bien sympa, vivement ton mariage ! ; et bien évidemment, le Pentanôme et ses légendaires ~~Ceschito~~ Francesco, ~~Yannou~~ Yann, ~~Zgays~~ Guillaume, et mon binôme de toujours, ~~Jøspèh~~ Joseph : un seul objectif, le Brésil... bon, et l'Argentine aussi (Ceschito ?) ; EBPL jusqu'à l'infini <3, VGGs!!! (oui ça vaut carton rouge, qu'allez-vous faire ?...) Je remercie aussi Adrien, pour son soutien constant ; Alice, pour son humour décalé et mordant ; et — comment serait-il possible de l'oublier ? — cette personne incroyable et pourtant bien réelle qu'est Amadéo, pour son savoir infini, sa grandeur d'âme, toutes ses réponses à mes questions à propos de ma thèse, et l'organisation des 23 et autres moments dantesques...

Je voudrais adresser un remerciement tout particulier à Joana, ce n'est pas grand chose face à toute la patience et tout le soutien qu'elle m'a témoignés. Merci pour ton écoute, pour tes conseils, merci d'avoir su m'aider à combattre mes doutes et me comprendre, toujours.

Enfin, je voudrais terminer en remerciant ma famille. En particulier, je souhaite dédier ce manuscrit à mes parents, Sabine et Remy, ainsi qu'à leurs parents, Simone, Jack, Jean, et Denise, qui m'a transmis le goût des maths et qui, j'espère, aurait apprécié lire ces pages. Merci pour votre gentillesse, et de m'avoir toujours soutenu dans mes études. Je tiens également à remercier

ma sœur Jeanne et mon frère Paul. Je vous aime tendrement. Je remercie aussi mes oncles et tantes, ainsi que leurs enfants, pour leur bienveillance attentionnée : Marc et Claude, François, Antoine et Vincent, Sophie et Patrick, Damien et Lucie, Corinne, et Alice.

Je tenais à vous remercier tous, chaleureusement. Et je demande pardon à ceux que j'aurais oubliés.

LUCIEN BACHARACH

Le temps doit emprunter ses ailes à l'imagination.
Ohran Pamuk

Résumé

Cette thèse porte sur l'étude des performances d'estimateurs en traitement du signal, et s'attache en particulier à étudier les bornes inférieures de l'erreur quadratique moyenne (EQM) pour l'estimation de points de rupture, afin de caractériser le comportement d'estimateurs, tels que celui du maximum de vraisemblance (dans le contexte fréquentiste), mais surtout du maximum *a posteriori* ou de la moyenne conditionnelle (dans le contexte bayésien).

La difficulté majeure provient du fait que, pour un signal échantillonné, les paramètres d'intérêt (à savoir les points de rupture) appartiennent à un espace discret. En conséquence, les résultats asymptotiques classiques (comme la normalité asymptotique du maximum de vraisemblance) ou la borne de Cramér-Rao ne s'appliquent plus. Quelques résultats sur la distribution asymptotique du maximum de vraisemblance provenant de la communauté mathématique sont actuellement disponibles, mais leur applicabilité à des problèmes pratiques de traitement du signal n'est pas immédiate.

Si l'on décide de concentrer nos efforts sur l'EQM des estimateurs comme indicateur de performance, un travail important autour des bornes inférieures de l'EQM a été réalisé ces dernières années. Plusieurs études ont ainsi permis de proposer des inégalités plus précises que la borne de Cramér-Rao. Ces dernières jouissent en outre de conditions de régularité plus faibles, et ce, même en régime non asymptotique, permettant ainsi de délimiter la plage de fonctionnement optimal des estimateurs.

Le but de cette thèse est, d'une part, de compléter la caractérisation de la zone asymptotique (en particulier lorsque le rapport signal sur bruit est élevé et/ou pour un nombre d'observations infini) dans un contexte d'estimation de points de rupture. D'autre part, le but est de donner les limites fondamentales de l'EQM d'un estimateur dans la plage non asymptotique. Les outils utilisés ici sont les bornes inférieures de l'EQM de la famille Weiss-Weinstein qui est déjà connue pour être plus précise que la borne de Cramér-Rao dans les contextes, entre autres, de l'analyse spectrale et du traitement d'antenne. Nous fournissons une forme compacte de cette famille dans le cas d'un seul et de plusieurs points de ruptures puis, nous étendons notre analyse aux cas où les paramètres des distributions sont inconnus. Nous fournissons également une analyse de la robustesse de cette famille vis-à-vis des loi a priori utilisées dans nos modèles. Enfin, nous appliquons ces bornes à plusieurs problèmes pratiques : données gaussiennes, poissonniennes et processus exponentiels.

Mots clés : Bornes inférieures de l'erreur quadratique moyenne, bornes de Weiss-Weinstein, bornes de Cramér-Rao, estimation de paramètres, estimateurs au sens du Maximum de Vraisemblance et du Maximum a Posteriori, étude des performances asymptotique et non-asymptotique, estimation de points de ruptures multiples.

Abstract

This thesis deals with the study of estimation performance in signal processing, and focuses on the analysis of lower bounds on the Mean Squared Error (MSE) for abrupt changepoint estimation. Those tools contribute to characterizing the estimation behavior for estimators such as the maximum likelihood estimator (in the frequentist context), as well as the maximum a posteriori and the conditional mean estimators (in the Bayesian context).

The main difficulty comes from the fact that, when dealing with sampled signals, the parameters of interest (i.e., the change-point locations) lie on a discrete space. Consequently, the classical large sample theory results (e.g., the asymptotic normality of the maximum likelihood estimator) or the Cramér-Rao bound do not apply. Some results concerning the asymptotic distribution of the maximum likelihood estimator are available in the mathematical literature, but their applicability to practical signal processing problems is not immediate.

Focusing on the 2nd-order moments of the estimators, lower bounds on the MSE make up essential tools, which have been the subject of many studies in the last years. As a result, new inequalities have been proposed, leading to tighter lower bounds in comparison to the Cramér-Rao bound. These new lower bounds require milder regularity conditions and enable the prediction of estimators' behavior in terms of MSE, both in asymptotic and non-asymptotic regimes.

The goal of this thesis is to complete previous results on lower bounds in the asymptotic context (i.e., when the number of samples and/or the signal-to-noise ratio is high) for change-point estimation but, also, to provide an analysis in the non-asymptotic setting. The tools used here are lower bounds from the Weiss-Weinstein family, which are already known in signal processing to outperform the Cramér-Rao bound in applications such as spectral analysis or array processing. A closed-form expression of this family is provided for a single and multiple change-points, and some extensions are given when the distribution parameters are unknown. An analysis in terms of robustness regarding the prior distribution on the change locations is also provided. Finally, we apply our results to specific problems, such as Gaussian data, Poisson data and exponentially distributed data.

Keywords: Lower bounds on the mean squared error, Weiss-Weinstein bound, Cramér-Rao bound, parameter estimation, maximum likelihood estimators, maximum a posteriori estimators, asymptotic and non-asymptotic performance, change-point estimation.

Table des matières

Résumé	xiii
1 Introduction	1
1.1 Contexte de la thèse et motivations	1
1.2 Résultats	4
2 Performances d'estimation, résultats pour des signaux comportant des ruptures	7
2.1 Estimation et performances pour des signaux réels : contexte et notions générales	7
2.1.1 Notations et cadres d'estimation	7
2.1.2 Estimation de paramètres déterministes (approche fréquentiste)	9
2.1.2.1 Performances d'estimation	9
2.1.2.1.1 Critères de performance les plus courants	9
2.1.2.1.2 Bornes de Cramér-Rao	11
2.1.2.2 Estimateurs classiques	13
2.1.2.2.1 Estimateur MVU	13
2.1.2.2.2 Estimateur du maximum de vraisemblance	14
2.1.3 Estimation de paramètres aléatoires (approche bayésienne)	16
2.1.3.1 Critères de performance en estimation bayésienne	16
2.1.3.2 Estimateurs bayésiens	17
2.1.3.2.1 Estimateur de la médiane <i>a posteriori</i> (cas scalaire)	18
2.1.3.2.2 Estimateur de la moyenne <i>a posteriori</i>	18
2.1.3.2.3 Estimateur du maximum <i>a posteriori</i>	19
2.1.3.3 Aperçu de quelques propriétés asymptotiques des estimateurs bayésiens	20
2.1.4 Estimation hybride	20
2.1.5 Bornes inférieures de l'EQM	21
2.1.5.1 L'inégalité de covariance	22
2.1.5.2 Bornes pour l'estimation de paramètres déterministes	23
2.1.5.2.1 Borne de Cramér-Rao	24
2.1.5.2.2 Bornes de Barankin	25
2.1.5.3 Bornes bayésiennes	27
2.1.5.3.1 Borne de Cramér-Rao bayésienne	27
2.1.5.3.2 Borne de Weiss-Weinstein	29
2.1.5.4 Combinaison de bornes, bornes hybrides	31
2.1.5.4.1 Combinaison de bornes bayésiennes : la borne (combinée) de Cramér-Rao/Weiss-Weinstein bayésienne (BCRWWB).	32

2.1.5.4.2	Bornes hybrides : la borne hybride de Cramér-Rao/Weiss-Weinstein	33
2.2	Estimation des paramètres de signaux comportant des points de rupture	35
2.2.1	La localisation de ruptures dans un signal échantillonné : formulation(s) du problème	36
2.2.2	Cas d'un seul point de rupture	37
2.2.3	Points de rupture multiples	40
2.3	Bornes inférieures de l'EQM (déterministes) pour l'estimation de points de ruptures	41
2.3.1	Bornes de type Cramér-Rao	41
2.3.2	Bornes de type Barankin	42
2.3.2.1	Borne de Chapman-Robbins pour la localisation d'un unique point de rupture	42
2.3.2.2	Borne de type Barankin pour la localisation de plusieurs points de rupture	43

3 Nouvelles bornes inférieures de l'EQM pour des signaux comportant des ruptures 45

3.1	Borne de Weiss-Weinstein pour l'estimation des points de rupture	46
3.1.1	Modèle multivarié	46
3.1.2	Méthode pour le calcul de la borne	47
3.1.3	Expression de la matrice $\mathbf{G}(\mathbf{H}, \mathbf{s})$	48
3.1.4	Influence de la distribution <i>a priori</i>	50
3.1.4.1	Expressions de $M(h_a, h_b, \mathbf{k})$	51
3.1.4.1.1	Cas ts : $b > a + 1$	51
3.1.4.1.2	Cas diag : $b = a$	51
3.1.4.1.3	Cas psdiag : $b = a + 1$	51
3.1.4.2	Expressions de $\nu(h_a, h_b, \mathbf{k})$	51
3.1.4.2.1	Cas diag : $b = a$	51
3.1.4.2.2	Cas psdiag : $b = a + 1$	52
3.1.4.2.3	Cas ts : $b > a + 1$	52
3.1.4.3	Expressions de $e^{\tilde{\mu}(h_a, h_b)}$ et des éléments de la matrice \mathbf{G}	53
3.1.4.3.1	Cas diag $b = a$	53
3.1.4.3.2	Cas psdiag $b = a + 1$	54
3.1.4.3.3	Cas ts $b > a + 1$	55
3.2	Bornes de Cramér-Rao/Weiss-Weinstein pour l'estimation de $\boldsymbol{\eta}$ et \mathbf{k}	55
3.2.1	Lien entre les bornes de Weiss-Weinstein et de Cramér-Rao/Weiss-Weinstein hybride et bayésienne	56
3.2.1.1	Retour sur le modèle et notations	56
3.2.1.2	Lien entre les matrices $\mathbf{G}(\mathbf{H}, \mathbf{s})$, $\mathbf{V}_{22}^{\eta^*}$, $\mathbf{C}_{22}^{\eta^*}$, \mathbf{V}_{22}^{π} et \mathbf{C}_{22}^{π}	56
3.2.1.3	Lien entre les matrices $\mathbf{V}_{11}^{\eta^*}$ et \mathbf{V}_{11}^{π}	57
3.2.1.4	Lien entre les matrices $\mathbf{V}_{12}^{\eta^*}$ et \mathbf{V}_{12}^{π}	57
3.2.2	Borne de Cramér-Rao/Weiss-Weinstein hybride	58
3.2.2.1	Expression de la matrice $\mathbf{V}_{11}^{\eta^*}$	58
3.2.2.2	Expression des matrices $\mathbf{V}_{22}^{\eta^*}$ et $\mathbf{C}_{22}^{\eta^*}$	58
3.2.2.3	Expression de la matrice $\mathbf{V}_{12}^{\eta^*}$	59
3.2.3	Borne de Cramér-Rao/Weiss-Weinstein bayésienne	60
3.2.3.1	Expression de la matrice \mathbf{V}_{11}^{π}	60
3.2.3.2	Expression des matrices \mathbf{V}_{22}^{π} et \mathbf{C}_{22}^{π}	60

3.2.3.3	Expression de la matrice \mathbf{V}_{12}^π	61
3.3	Maximisation sur un ensemble de matrices définies positives	62
4	Applications en traitement du signal	65
4.1	Cas où les paramètres de la distribution du signal observé sur chaque segment sont connus	65
4.1.1	Cas d'un signal comportant un seul point de rupture	66
4.1.1.1	Signaux gaussiens	66
4.1.1.1.1	Expression de $\rho_{q-1,q}(s)$	67
4.1.1.1.2	Résultats numériques.	68
4.1.1.2	Signaux poissonniens	71
4.1.1.2.1	Expression de $\rho_{q-1,q}(s)$	72
4.1.1.2.2	Résultats numériques.	73
4.1.1.3	Signaux suivant une loi exponentielle	73
4.1.1.3.1	Expression de $\rho_{q-1,q}(s)$	74
4.1.1.3.2	Résultats numériques.	74
4.1.2	Cas de signaux comportant de multiples points de rupture	75
4.1.2.1	Signaux gaussiens	75
4.1.2.1.1	Expressions de $r_q(s, s')$ et $R_q(s, s')$	75
4.1.2.1.2	Influence de la distribution <i>a priori</i> sur \mathbf{k}	76
4.1.2.1.3	Résultats numériques.	80
4.1.2.2	Signaux poissonniens	84
4.1.2.2.1	Expression de $r_q(s, s')$ et $R_q(s, s')$	84
4.1.2.2.2	Résultats numériques.	84
4.2	Cas où les paramètres de la distribution du signal observé sur chaque segment sont inconnus	85
4.2.1	Application à des signaux poissonniens	86
4.2.1.1	Cas de paramètres λ_q déterministes	86
4.2.1.1.1	Expressions de $\tilde{F}(\lambda_q)$ et de $\varphi_{q,\lambda_{\bar{q}}}(s)$	86
4.2.1.1.2	Résultats numériques.	86
4.2.1.2	Cas de paramètres λ_q aléatoires	86
4.2.1.2.1	Expressions des éléments de la borne de Cramér-Rao/Weiss-Weinstein bayésienne.	87
4.2.1.2.2	Résultats numériques.	87
4.2.2	Application à des signaux gaussiens	88
4.2.2.1	Cas de signaux constants sur chaque segment	89
4.2.2.1.1	Expressions de $\tilde{F}(\boldsymbol{\eta}_q)$, de $\varphi_{q,m_{\bar{q}}}(s)$ et de $\varphi_{q,\sigma_{\bar{q}}^2}(s)$	89
4.2.2.1.2	Résultats numériques.	89
4.2.2.2	Cas d'un signal sinusoïdal complexe comportant une rupture de fréquence	92
4.2.2.2.1	Modèle et hypothèses.	92
4.2.2.2.2	Expression de la borne de Weiss-Weinstein.	94
4.2.2.2.3	Résultats numériques	95
5	Conclusion et perspectives	97

A	Expression de la matrice $\mathbf{G}(H, s)$	101
A.1	Expressions de $\nu(\alpha, \beta, h_a, h_b, \mathbf{k})$	101
A.2	Expressions de $M(\alpha, \beta, h_a, h_b, \mathbf{k})$	103
A.2.1	Cas diag : $b = a$	104
A.2.2	Cas ts : $b > a + 1$	105
A.2.3	Cas psdiag : $b = a + 1$	105
A.3	Expressions de $e^{\tilde{\mu}(\alpha, \beta, h_a, h_b)}$ et de $[\mathbf{G}]_{a,b}$	107
A.3.1	Cas diag et termes diagonaux de \mathbf{G}	108
A.3.2	Cas ts et termes du triangle supérieur de \mathbf{G}	108
A.3.3	Cas psdiag et termes des premières surdiagonale et sous-diagonale de \mathbf{G}	109
A.3.3.1	Expressions de $e^{\tilde{\mu}(\alpha, \beta, h_a, h_{a+1})}$	109
A.3.3.1.1	Expression de $e^{\tilde{\mu}(\alpha, \beta, h_a, h_{a+1})}$ dans le cas sans chevauchements	109
A.3.3.1.2	Expression de $e^{\tilde{\mu}(\alpha, \beta, h_a, h_{a+1})}$ dans le cas avec chevauchements	109
A.3.3.2	Expression des termes des premières surdiagonale et sous-diagonale de \mathbf{G}	111
B	Expressions des éléments des matrices $V_{11}^{\eta^*}$ et $V_{12}^{\eta^*}$	113
B.1	Expressions des éléments de la matrice $V_{11}^{\eta^*}$	113
B.2	Expressions des éléments de la matrice $V_{12}^{\eta^*}$	115
B.2.1	Cas tsi : $\tilde{q} \neq q - 1$ and $\tilde{q} \neq q$	116
B.2.2	Cas d1 : $\tilde{q} = q - 1$	116
B.2.3	Cas d2 : $\tilde{q} = q$	118
	Notations	119
	Liste des figures	121
	Liste des tableaux	125
	Bibliographie	127

Chapitre 1

Introduction

1.1 Contexte de la thèse et motivations

En traitement du signal, il arrive qu'on se trouve confronté à des signaux susceptibles de changer de façon abrupte. Concrètement, lorsque ceux-ci sont modélisés par des processus aléatoires, cela se traduit par un changement quasi-instantané dans leur distribution, ou du moins rapide par rapport à leur fréquence d'échantillonnage. Plus précisément, c'est dans le (ou les) paramètre(s) de la distribution du signal que se produit le changement. Le terme employé dans la littérature pour faire référence à ces instants de changement abrupt est celui de *point de rupture*, soit en anglais : *changepoint* ou *abrupt change*. Les premiers travaux d'étude de tels signaux remontent à ceux de Shewhart [She31], s'appliquant au contrôle de qualité de systèmes de production. Depuis, de nombreux autres domaines d'application ont pu être investis, tels que le traitement du signal audio, le traitement de données sismiques ou astronomiques, ou encore le contrôle de systèmes de navigation. La majorité de ces domaines d'application se retrouvent dans l'ouvrage de référence [BN93]. Un autre domaine – se rapportant davantage à la surveillance de l'environnement – est celui du traitement des signaux radar, où la localisation de points de rupture peut être intéressante : c'est le cas notamment pour la segmentation de cartes de fouillis radar [BB07], la détection de cibles dans du fouillis [NNV02], ou bien pour la détection de contours dans des images SAR [TDL03]. Dans toutes ces applications, les positions des points de rupture sont généralement supposées inconnues et doivent donc être estimées/localisées le plus précisément possible.

Suivant l'application considérée et le but recherché, le problème de localisation de ces changements est formulé de différentes manières. En premier lieu, c'est le mode de réception des données qui va dicter une partie de la formulation : pour des applications en temps réel où les données doivent être traitées *au fur et à mesure*, les méthodes proposées ne pourront s'appuyer que sur les observations obtenues jusqu'à l'instant courant ; c'est ce que l'on appelle la formulation « *on-line* » du problème. À l'inverse, dans d'autres cas, on dispose d'un lot complet de données, contenant donc déjà toute l'information à traiter, on parle dans ce cas de formulation « *off-line* ».

Une autre partie de la formulation du problème est fixée par le but à atteindre : on peut d'une part chercher à *déterminer si* les données contiennent effectivement un ou plusieurs points de rupture ; d'autre part, si leur existence est supposée acquise, on peut chercher à les *localiser* avec la plus grande précision possible. En général, le premier problème (celui de la décision de l'existence d'un changement) correspond à un problème de détection se posant donc sous la forme d'un test d'hypothèses, tandis que le second (celui de la localisation) correspond davantage à un problème d'estimation. En résumé, le problème formulé reposera sur ces deux notions :

- Mode de réception des données : on-line ou off-line ;

— Objectif du traitement : détection ou estimation d'un ou plusieurs points de rupture.

Un premier type de problème, que l'on pourrait appeler celui de la détection on-line, consisterait à établir une règle de décision sur l'existence d'un changement. Cette règle de décision, qui devra reposer sur les données observées jusqu'à l'instant courant, sera généralement basée sur le dépassement d'un seuil critique qui, lorsqu'il est dépassé, implique l'existence du changement. C'est le cas de l'algorithme CUSUM (pour « *cumulative sum* »), largement décrit dans [BN93] par exemple. Ce type de formulation est particulièrement adapté pour les problèmes de contrôle de qualité ou de surveillance de systèmes [BN93].

Un autre problème serait celui de l'estimation on-line, où l'on cherche à suivre, en temps réel, l'état du système considéré. Cela s'apparente à un problème de poursuite (ou « *tracking* »), où les paramètres d'intérêt sont plutôt ceux décrivant le système (davantage que les points de rupture eux-mêmes, alors considérés comme des paramètres de nuisance). Leur estimation, ou du moins leur prise en compte, sera toutefois nécessaire pour permettre un suivi correct de l'état du système. Ce problème est discuté dans [CZ64] et étudié de manière plus approfondie dans [Gus00].

Enfin, en ce qui concerne le problème off-line, on peut d'une part s'intéresser au problème de la seule détection d'un point de rupture. Pour un signal observé $\mathbf{x} = [x_1, \dots, x_N]^\top$ décrit par une famille de distributions p_η , paramétrée par un vecteur $\eta \in \mathbb{R}^P$, le problème d'un unique point de rupture k_1 peut être posé de la façon suivante :

$$\left\{ \begin{array}{ll} \mathcal{H}_0 : & x_n \sim p_{\eta_0}(x_n | x_{n-1}, \dots, x_1) \quad \text{pour} \quad 1 \leq n \leq N \\ \mathcal{H}_1 : & \text{il existe un point de rupture } k_1 \in \llbracket 1, N-1 \rrbracket, \text{ tel que :} \\ & x_n \sim p_{\eta_0}(x_n | x_{n-1}, \dots, x_1) \quad \text{pour} \quad 1 \leq n \leq k_1 \\ & x_n \sim p_{\eta_1}(x_n | x_{n-1}, \dots, x_1) \quad \text{pour} \quad k_1 + 1 \leq n \leq N. \end{array} \right. \quad (1.1)$$

La décision s'effectue à partir d'un test entre l'hypothèse \mathcal{H}_0 et l'hypothèse \mathcal{H}_1 .

Le problème peut être étendu au cas de plusieurs points de rupture, pour lequel le problème de détection permettrait de déterminer *combien* de changements sont survenus, tandis que le problème d'estimation répond toujours à celui de savoir *où* ceux-ci se situent. Pour résoudre ces deux problèmes, Chen et Gupta proposent une méthode (sous-optimale) qui consiste à appliquer une procédure de détection et d'estimation conjointes pour un seul changement sur toutes les observations, puis de nouveau de façon itérative sur chaque sous-segment ainsi induit [CG00].

Dans les présents travaux, on s'intéresse aux performances d'*estimation* de la position d'un ou plusieurs points de rupture. Dès lors, on se place sous \mathcal{H}_1 , c'est-à-dire que l'on suppose l'existence d'un point de rupture. En d'autres termes, on délaisse le problème de la détection au profit de celui de l'estimation, en considérant qu'une étape préliminaire de détection aura déjà été effectuée. Cette démarche est classique lorsque l'on étudie des performances d'estimation. On supposera donc connu le nombre de points de rupture, mais pas leur position.

Les travaux se rapportant à ce problème se sont concentrés, pour la plupart, sur le cas d'un seul changement. Il existe plusieurs manières d'évaluer les performances d'estimation. On peut, par exemple, s'intéresser à la distribution d'un estimateur $\hat{k}_1(\mathbf{x})$ donné, c'est-à-dire trouver une expression de $\Pr(\hat{k}_1(\mathbf{x}) = k)$, pour $k \in \llbracket 1, N-1 \rrbracket$. Autrement, une alternative consiste à étudier la distribution de l'erreur, sous la forme $\Pr(\hat{k}_1(\mathbf{x}) = k_1^* \pm k)$, ou encore $\Pr(|\hat{k}_1(\mathbf{x}) - k_1^*| \leq k)$, où k_1^* représente la vraie position du point de rupture k_1 . De tels résultats fournissent une caractérisation précise des performances mais ceux-ci, quand c'est possible, sont difficiles à obtenir. Les premières études pour y parvenir remontent à celles de Page [Pag57], mais c'est Hinkley [Hin70] qui démontre dans les années 1970 la non-consistance de l'estimateur du maximum de vraisemblance $\hat{k}_1^{\text{ML}}(\mathbf{x})$ de k_1 (« ML » pour *maximum likelihood* en anglais), même dans le cas de séries gaussiennes indépendantes identiquement distribuées. Il en fournit néanmoins une expression en régime asymptotique, c'est-à-dire à grand nombre d'échantillons avant et après le

changement. Le résultat restant difficilement exploitable dans des cas pratiques, ce n'est que dans les années 2000-2010 que Fotopoulos, Jandhyala et leurs collaborateurs donnent des expressions, toujours asymptotiques, de $\hat{k}_1^{\text{ML}}(\mathbf{x})$ pour des cas particuliers de séries exponentielles [FJ01] ou gaussiennes [FJK10] indépendantes. Ces difficultés s'expliquent notamment par le fait qu'il s'agit d'un problème d'estimation de paramètre discret, pour lequel on sait que les résultats classiques de convergence de l'estimateur du maximum de vraisemblance, ou de sa normalité et efficacité asymptotiques ne s'appliquent pas : on pourra se reporter aux récents travaux de Choirat et Seri [CS12].

Pour ces raisons, on s'intéresse à une autre méthode permettant de caractériser les performances d'estimation : il s'agit de *borner* inférieurement les moments de familles d'estimateurs, en particulier les moments d'ordre 2, à savoir la variance ou l'erreur quadratique moyenne (EQM). Obtenir une borne inférieure de l'EQM permet donc de connaître *l'erreur minimale* qu'un estimateur peut espérer atteindre. Dans l'idéal, on souhaite se rapprocher, autant que possible, d'une borne inférieure de l'EQM au sens mathématique d'une borne inférieure, c'est-à-dire obtenir le plus grand des minorants de l'EQM, pour que celle-ci reflète au mieux le comportement des estimateurs qu'elle englobe. Le calcul de bornes inférieures de l'EQM offre donc un intérêt à plusieurs titres :

- elles fournissent une indication des performances optimales d'estimation ;
- elles sont valables pour une famille d'estimateurs (contrairement au calcul de la distribution d'un estimateur, comme mentionné plus haut, qui porte sur un seul estimateur, celui considéré) ;
- elles sont moins coûteuses en ressources de calcul que l'estimation de l'EQM ou de la variance d'un estimateur, qui requiert de nombreuses simulations de type Monte-Carlo ;
- elles fournissent un aperçu des performances d'estimation, et pour certaines d'entre elles, pas seulement dans le domaine asymptotique (c'est-à-dire pas seulement à fort nombre d'observations, ou à fort rapport signal à bruit) ;
- comme elles dépendent de grandeurs propres au problème considéré, elles peuvent permettre de dimensionner un système pour répondre à des exigences minimales en terme de performance d'estimation.

La borne la plus connue est la borne de Cramér-Rao (BCR) [Rao45, Cra46], car elle définit la notion d'efficacité : lorsqu'un estimateur atteint la borne, on dit qu'il est efficace. Elle est, de plus, généralement simple à calculer pour des problèmes courants de traitement du signal, comme par exemple lorsque les données sont distribuées selon une loi gaussienne. L'expression de la BCR, pour un paramètre inconnu θ^* scalaire déterministe, est donnée par :

$$\text{BCR}(\theta^*) = - \frac{1}{\mathbb{E}_{\mathbf{x}; \theta^*} \left\{ \left. \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right|_{\theta = \theta^*} \right\}}. \quad (1.2)$$

où $p(\mathbf{x}; \theta)$ est la vraisemblance des observations.

L'une des caractéristiques de la BCR est qu'elle est souvent *asymptotiquement* proche du comportement réel de l'estimation. C'est aussi une limitation car en-deçà d'un certain nombre d'observations, ou d'un certain rapport signal à bruit, la BCR est trop optimiste : elle donne une EQM minimale bien trop faible qu'aucun estimateur ne peut produire. Ce phénomène, appelé décrochement, est observé notamment pour des problèmes d'estimation de fréquence [Ren06], et de façon plus générale pour des problèmes d'estimation non-linéaire.

D'autre part, pour pouvoir la calculer, plusieurs conditions de régularité doivent être vérifiées. L'une d'elles, on peut le voir à partir de (1.2), est que la log-vraisemblance doit être dérivable par rapport au paramètre inconnu θ . Or dans notre problème de localisation de points de rupture, ce paramètre à estimer est t_0 , qui est discret. La dérivation par rapport à ce paramètre n'est donc pas

possible. Dans la littérature, ce problème a été contourné soit en employant une paramétrisation continue de t_0 [RD96], soit en approximant la rupture par une fonction continue, par exemple sigmoïdale dans le cas d'un changement de moyenne [SS99]. Ces approches amènent plusieurs problèmes, le principal étant que la BCR ainsi obtenue ne rend pas compte du comportement de l'estimation avec autant de précision qu'on pourrait le souhaiter : il existe un écart non négligeable entre l'EQM des estimateurs classiques et la borne, faisant ainsi perdre à celle-ci une partie de son caractère informatif, et donc de son intérêt.

La BCR n'est heureusement pas la seule borne inférieure de l'EQM. Il en existe en fait un très grand nombre, dont le calcul est plus ou moins complexe. On peut par exemple citer les bornes de Bhattacharyya, de Barankin ou encore de Chapman-Robbins (pour un panorama complet des différentes bornes déterministes, voir par exemple [Cha04]). En particulier, la borne de Chapman-Robbins a elle aussi été calculée pour le problème d'estimation de la localisation d'un point de rupture dans des données indépendantes, identiquement distribuées (i.i.d.) [FT03], et a même été étendue au cas de changements multiples [LRNM10], avec un écart plus faible, mais toujours important entre l'EQM de l'estimateur ML et la borne (voir chapitre 2). Ce sont ces deux travaux, parce qu'ils prenaient, de manière originale, explicitement en compte le caractère discret des paramètres à estimer, qui ont principalement motivé cette thèse.

Afin d'essayer de combler cet écart et de fournir une borne aussi précise que possible, nous nous sommes donc intéressés à la famille de bornes dite de « Weiss-Weinstein » [WW88], qui est connue pour être l'une des plus précises [NH88, NV94, XBR04, VB07]. À la différence des bornes précédemment citées, cette famille porte sur l'EQM *globale* d'un estimateur *bayésien* quelconque, c'est-à-dire qu'elle nécessite de se placer dans un cadre bayésien, et donc de supposer que le ou les paramètres inconnus sont aléatoires, suivant une loi *a priori* donnée. Dans le cas d'un paramètre inconnu θ scalaire, l'expression la plus simple de la borne de Weiss-Weinstein est la suivante :

$$\text{WWB} \triangleq \sup_{h,s} \frac{h^2 \mathbb{E}_{\mathbf{x},\theta}^2 \left\{ \left(\frac{p(\mathbf{x}, \theta + h)}{p(\mathbf{x}, \theta)} \right)^s \right\}}{\mathbb{E}_{\mathbf{x},\theta} \left\{ \left[\left(\frac{p(\mathbf{x}, \theta + h)}{p(\mathbf{x}, \theta)} \right)^s - \left(\frac{p(\mathbf{x}, \theta - h)}{p(\mathbf{x}, \theta)} \right)^{1-s} \right]^2 \right\}} \quad (1.3)$$

où $p(\mathbf{x}, \theta)$ désigne la densité de probabilité jointe des observations \mathbf{x} et du paramètre inconnu θ , h et s sont des paramètres tels que $\theta \pm h$ soient dans l'espace des paramètres et $s \in]0, 1[$, sous réserve que le dénominateur de (1.3) soit non nul. Dans le cas où l'on a plusieurs paramètres inconnus, la borne prend la forme d'une matrice dont les éléments ont une expression du type de (1.3) (voir [WW88] ainsi que le chapitre 2 du présent document).

1.2 Résultats

Après un chapitre introductif (chapitre 2) qui va permettre la mise en place des notations, des outils (estimateurs et bornes inférieures de l'erreur quadratique moyenne), et des modèles d'étude concernant l'estimation de points de rupture, les travaux présentés dans ce document s'articuleront de la manière suivante :

- dans le chapitre 3, nous présentons la borne de Weiss-Weinstein pour des signaux de paramètres connus comportant plusieurs points de rupture inconnus. Nous montrons en particulier que la structure de la matrice d'information inhérente à cette borne exhibe une structure tridiagonale. Dans un second temps, une analyse des coefficients de cette matrice montre que le calcul complet de la borne, quelle que soit la distribution des signaux considérés, ne dépend que d'un facteur lié à la divergence de Rényi entre les distributions de

deux segments consécutifs (avant et après un point de rupture). Ensuite, nous présentons deux autres bornes de la famille Weiss-Weinstein lorsque la distribution des signaux considérés dépend de paramètres inconnus. Selon la nature déterministe ou aléatoire de ces paramètres, nous proposons, respectivement, une borne hybride et une borne bayésienne combinée de Cramér-Rao/Weiss-Weinstein, qui permettent de régler un grand nombre de problèmes pratiques. Enfin, comme toutes ces bornes nécessitent une maximisation par rapport à un ensemble de matrices, nous présentons une procédure algorithmique permettant de réaliser cette opération de manière efficace, basée l'ordonnancement de Löwner.

- dans le chapitre 4, nous appliquons toutes ces bornes à des distributions fréquemment rencontrées en traitement statistique du signal : gaussienne avec changement de moyenne (paramétrée ou non) et de variance, poissonnienne et exponentielle. Une analyse de la robustesse de la borne de Weiss-Weinstein par rapport à différentes distributions *a priori* modélisant la position des points de rupture. Nous montrons que la distribution choisie au chapitre 3 offre le meilleur compromis entre complexité calculatoire et précision de la borne. Enfin, nous validons ces analyses par des résultats de simulation démontrant un gain de précision des différentes bornes proposées par rapport à la littérature.

Les conclusions et perspectives de ces travaux sont données dans le chapitre 5.

Les travaux exposés dans ce document ont donné lieu aux articles et communications suivants :

Revue internationale

- [BREC17] L. Bacharach, A. Renaux, M. N. El Korso et É. Chaumette, “Weiss-Weinstein Bound on Multiple Change-Points Estimation,” *IEEE Transactions on Signal Processing*, Volume : 65, Issue : 10, mai 2017, pp. 2686–2700.
- [BERT] L. Bacharach, M. N. El Korso, A. Renaux et J.-Y. Tournet, “A Hybrid Lower Bound for Parameter Estimation of Signals with Multiple Change-Points,” en révision mineure.

Conférences avec comités de lecture et actes

- [BREC15] L. Bacharach, A. Renaux, M. N. El Korso, et É. Chaumette, “Weiss-Weinstein bound for change-point estimation,” in *Proc. of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Cancún, Mexique, décembre 2015, pp. 477–480.
- [BER16] L. Bacharach, M. N. El Korso, et A. Renaux, “Weiss-Weinstein bound for an abrupt unknown frequency change,” in *Proc. of IEEE Workshop on Statistical Signal Processing (SSP)*, Palma de Majorque, Espagne, juin 2016, pp. 420–424.
- [BERT17] L. Bacharach, M. N. El Korso, A. Renaux, et J.-Y. Tournet, “A Bayesian Lower Bound for Parameter Estimation of Poisson Data Including Multiple Changes,” in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Nouvelle-Orléans, Louisiane, États-Unis, mars 2017, pp. 4486–4490.
- [BBRE17] L. Bacharach, G. Bibiche, A. Renaux, et M. N. El Korso, “Bornes bayésiennes pour la localisation d’un point de rupture : application à des processus exponentiels,” in *Actes du XXVI^e Colloque GRETSI sur le traitement du signal et des images*, Juan-les-Pins, France, septembre 2017.

- [BRE18] L. Bacharach, A. Renaux, et M. N. El Korso, “Prior Influence on Weiss-Weinstein Bounds for Multiple Change-Point Estimation,” accepté à *European Signal Processing Conference (EUSIPCO)*, septembre 2018, Rome, Italie.

Séminaires et présentations lors de *workshops*

- “Performances d’estimation pour des signaux comportant des points de rupture,” Journée GDR ISIS – Traitement d’antenne : signaux non gaussiens, non circulaires, non stationnaires, Paris, France, décembre 2016.
- “Bornes d’estimation bayésienne pour des signaux comportant des points de rupture,” École d’été de Peyresq en traitement du signal et des images, Peyresq, France, juin 2017.

Chapitre 2

Notions de performances d'estimation et résultats pour des signaux comportant des points de rupture

Le but de ce chapitre est de donner un aperçu des résultats déjà existants en lien avec les performances d'estimation pour des signaux comportant un ou plusieurs points de rupture. Nous commençons, dans la Section 2.1, avec des éléments généraux de théorie de l'estimation paramétrique, en distinguant les différents points de vue qu'il est possible d'adopter : fréquentiste, bayésien ou hybride. Nous y introduisons les notions, notations et outils utilisés pour l'évaluation des performances d'estimation, en particulier les bornes inférieures de l'erreur quadratique moyenne. Dans la Section 2.2, nous nous intéressons au cas de signaux pouvant comporter des points de rupture. Nous y décrivons les principaux résultats de la littérature en lien avec l'estimation des paramètres pour ce type de signaux. En particulier, nous portons notre attention sur les performances de la localisation d'une ou de plusieurs ruptures.

2.1 Estimation et performances pour des signaux réels : contexte et notions générales

2.1.1 Notations et cadres d'estimation

Comme nous l'avons vu en introduction, le signal étudié comporte généralement une part aléatoire, provenant par exemple des perturbations auxquelles il peut être soumis. Nous supposons que ce signal est échantillonné, car issu d'un système d'acquisition. Celui-ci peut être composé d'un seul ou de plusieurs capteurs. Si le signal observé n'est issu que d'un seul capteur, il consiste en un vecteur, noté \mathbf{x} , qui contient toutes les observations relevées par le système d'acquisition, i.e., $\mathbf{x} \triangleq [x_1, x_2, \dots, x_N]$. Si le signal observé est issu d'un nombre M de capteurs, il prend la forme d'une matrice, notée \mathbf{X} , dont les vecteurs colonnes contiennent les valeurs relevées par chacun des capteurs à un instant donné, i.e., $\mathbf{X} \triangleq [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ avec, pour $n = 1, \dots, N$, $\mathbf{x}_n \triangleq [x_{1,n}, x_{2,n}, \dots, x_{M,n}]^T$. On suppose, dans tout ce manuscrit, que la phase d'acquisition du signal \mathbf{X} est terminée, c'est-à-dire que sa dimension N est fixée et n'augmente donc pas au cours du temps. L'espace des observations, dans lequel \mathbf{X} prend ses valeurs, est noté Ω . Dans la majeure partie de ce manuscrit, nous supposerons que le signal étudié est réel, i.e., $\Omega \subset \mathbb{R}^{M \times N}$ (un seul cas de signal complexe sera considéré au chapitre 4, section 4.2.2.2, avec $M = 1$; dans ce cas on aura donc $\Omega \subset \mathbb{C}^N$).

Pour chaque problème étudié, le modèle considéré est différent. Ce modèle prend en général la forme d'une ou plusieurs équations qui relient le signal observé \mathbf{X} à l'information utile recherchée. Celle-ci correspond à une ou plusieurs grandeurs inconnues du problème, appelées paramètres d'intérêt, et regroupées dans un vecteur $\boldsymbol{\theta} \triangleq [\theta_1, \dots, \theta_R]^T$, de taille R (nombre total de paramètres inconnus).

Dans la suite de ce chapitre, on considère le cas de $M = 1$ capteur ; le signal observé est donc noté \mathbf{x} . Son caractère aléatoire, ainsi que l'influence de $\boldsymbol{\theta}$ sur ce signal observé, sont décrits par la distribution de \mathbf{x} (paramétrée par $\boldsymbol{\theta}$) au sens large :

- si les données observées sont des variables aléatoires absolument continues (i.e., $\Omega \subset \mathbb{R}^N$), alors la distribution de \mathbf{x} correspond à une fonction de densité de probabilité notée $p_{\mathbf{x};\boldsymbol{\theta}}(\mathbf{x};\boldsymbol{\theta})$ ou $p_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta} = \boldsymbol{\theta})$, selon que le vecteur $\boldsymbol{\theta}$ est supposé déterministe ou aléatoire (voir ci-après pour davantage de précisions) ;
- si les données observées sont des variables aléatoires discrètes (i.e., $\Omega \subset \mathbb{N}^N$), alors la distribution de \mathbf{x} correspond à une fonction de masse notée $\Pr(\mathbf{x} = \mathbf{x};\boldsymbol{\theta})$ si $\boldsymbol{\theta}$ est supposé déterministe, ou $\Pr(\mathbf{x} = \mathbf{x}|\boldsymbol{\theta} = \boldsymbol{\theta})$ si $\boldsymbol{\theta}$ est supposé aléatoire. Dans les deux cas, cette fonction de masse correspond à la probabilité que le vecteur aléatoire \mathbf{x} prenne la valeur du vecteur observé \mathbf{x} , pour tout \mathbf{x} , lorsque le vecteur de paramètres vaut $\boldsymbol{\theta}$.

Dans la suite, on s'affranchira de la distinction entre densité de probabilité et fonction de masse, et nous utiliserons toujours la notation de densité de probabilité. Pour autant, les résultats énoncés peuvent aisément se généraliser au cas d'une fonction de masse (ce cas figurera d'ailleurs parmi les exemples étudiés). De plus, pour ne pas alourdir les notations, nous omettrons l'écriture des variables aléatoires (comme \mathbf{x} par exemple) dans le nom des fonctions de densité de probabilité, dans les cas où cela ne prêterait pas à ambiguïté. Ainsi, la distribution des observations sera simplement notée $p(\mathbf{x};\boldsymbol{\theta})$, ou $p(\mathbf{x}|\boldsymbol{\theta})$, les arguments des fonctions indiquant clairement quelles sont les variables aléatoires associées. La fonction $p(\mathbf{x};\boldsymbol{\theta})$ (ou $p(\mathbf{x}|\boldsymbol{\theta})$) est communément appelée *vraisemblance des observations* (pour une valeur du vecteur de paramètres donnée $\boldsymbol{\theta}$).

En ce qui concerne la distinction entre les deux notations $p(\mathbf{x};\boldsymbol{\theta})$ et $p(\mathbf{x}|\boldsymbol{\theta})$, elle provient de la nature supposée du vecteur $\boldsymbol{\theta}$, et donc du cadre d'estimation qui est choisi : dans le cas où on note $p(\mathbf{x};\boldsymbol{\theta})$, le vecteur de paramètres $\boldsymbol{\theta}$ est supposé déterministe, avec une *vraie* valeur notée $\boldsymbol{\theta}^*$; dans l'autre cas, où on note $p(\mathbf{x}|\boldsymbol{\theta})$, le vecteur $\boldsymbol{\theta}$ est supposé aléatoire¹. On lui attribue alors une distribution dite « *a priori* » notée $\pi(\boldsymbol{\theta})$ et supposée connue, qui décrit l'ensemble des informations connues sur le vecteur $\boldsymbol{\theta}$ avant l'exécution de la procédure d'estimation (*a priori*). Cette distribution peut être continue, discrète, ou mixte (c'est-à-dire que certaines composantes de $\boldsymbol{\theta}$ sont des variables aléatoires continues, et que d'autres sont discrètes), suivant la nature des différents paramètres $\theta_1, \dots, \theta_R$ composant $\boldsymbol{\theta}$.

En résumé, le problème d'estimation peut s'énoncer de la manière suivante : on cherche à estimer, avec le plus de précision possible, la valeur du vecteur de paramètres $\boldsymbol{\theta}$ à partir des N échantillons du signal \mathbf{x} dont on dispose. Autrement dit, on cherche un estimateur de $\boldsymbol{\theta}$ qui s'exprime comme une fonction des données x_1, \dots, x_N , et qu'on notera donc $\hat{\boldsymbol{\theta}}(\mathbf{x}) \triangleq \hat{\boldsymbol{\theta}}(x_1, \dots, x_N)$. Notons que puisque les données \mathbf{x} sont aléatoires, l'estimateur $\hat{\boldsymbol{\theta}}(\mathbf{x})$ est lui-même un vecteur aléatoire. Ainsi, le comportement, et donc la précision/les performances de ce vecteur aléatoire, seront entièrement décrits par sa distribution, c'est-à-dire sa fonction de répartition. Si $\boldsymbol{\theta}$ est un vecteur de paramètres continus, on pourra se contenter de la fonction de densité de probabilité, tandis que si $\boldsymbol{\theta}$ est un vecteur de paramètres discrets, on pourra s'en tenir à la fonction de masse de $\hat{\boldsymbol{\theta}}(\mathbf{x})$. Enfin, le vecteur $\boldsymbol{\theta}$ peut tout aussi bien être composé d'un mélange de paramètres continus et discrets, auquel cas la distribution sera mixte.

1. Dans ce dernier cas, $\boldsymbol{\theta}$ est donc en réalité une réalisation du vecteur aléatoire associé, noté $\boldsymbol{\theta}$. Mais, comme expliqué plus haut, nous omettons l'écriture de ce dernier.

Dans les sections suivantes, nous présentons séparément le cas où le vecteur $\boldsymbol{\theta}$ est supposé déterministe, qui correspond au cadre de l'estimation fréquentiste, du cas où $\boldsymbol{\theta}$ est supposé aléatoire, qui correspond au cadre de l'estimation bayésienne. Bien que ceux-ci aient historiquement souvent été mis en opposition, il existe des liens entre les deux points de vue (voir [Rob07]), et ils peuvent même être combinés : pour certains problèmes, on dispose d'informations *a priori* sur certaines composantes du vecteur de paramètres $\boldsymbol{\theta}$ à estimer, mais pas sur d'autres. Ce cadre d'estimation est alors qualifié d'« hybride » [RS87, RM97, TW07, RGC⁺15, Ren15].

2.1.2 Estimation de paramètres déterministes (approche fréquentiste)

Dans un premier temps, nous nous plaçons dans le contexte de l'estimation dite « fréquentiste », où les paramètres inconnus sont supposés déterministes. Nous présentons d'abord, dans la section 2.1.2.1, quelques critères permettant d'évaluer les performances d'estimation, puis, dans la section 2.1.2.2, quelques estimateurs couramment utilisés dans le cadre fréquentiste.

2.1.2.1 Performances d'estimation

2.1.2.1.1 Critères de performance les plus courants D'une manière générale, pour un vecteur de paramètres inconnu $\boldsymbol{\theta}$ dont la vraie valeur est $\boldsymbol{\theta}^*$, on peut évaluer la qualité d'un estimateur $\hat{\boldsymbol{\theta}}$, fondé sur les observations \boldsymbol{x} , à partir de l'erreur commise, notée ε :

$$\varepsilon(\hat{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{\theta}^*) \triangleq \hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}^*, \quad (2.1)$$

où l'on met en avant la dépendance de l'erreur selon

- i) l'estimateur $\hat{\boldsymbol{\theta}}$;
- ii) les données observées \boldsymbol{x} ;
- iii) la vraie valeur du vecteur de paramètre inconnu, $\boldsymbol{\theta}^*$.

Comme mentionné dans la section 2.1.1, un estimateur $\hat{\boldsymbol{\theta}}$ est une variable aléatoire en raison de la nature elle-même aléatoire du signal observé \boldsymbol{x} . L'erreur $\varepsilon(\hat{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{\theta}^*)$ ainsi définie peut donc être très différente suivant la réalisation du signal observé \boldsymbol{x} . C'est pourquoi on s'intéressera plutôt à une erreur moyennée par rapport à toutes les valeurs que peut prendre le signal \boldsymbol{x} .

On en vient donc à définir le *biais* d'un estimateur $\hat{\boldsymbol{\theta}}$ comme la moyenne des erreurs $\varepsilon(\hat{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{\theta}^*)$ par rapport à \boldsymbol{x} , pondérée par la vraisemblance des observations $p(\boldsymbol{x}; \boldsymbol{\theta})$:

Définition 1 : Biais d'un estimateur

Le biais d'un estimateur $\hat{\boldsymbol{\theta}}$ du vecteur de paramètres $\boldsymbol{\theta}^*$ est défini comme la quantité

$$\mathbf{biais}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) \triangleq \mathbb{E}_{\boldsymbol{x}; \boldsymbol{\theta}^*} \{ \varepsilon(\hat{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{\theta}^*) \} = \mathbb{E}_{\boldsymbol{x}; \boldsymbol{\theta}^*} \{ \hat{\boldsymbol{\theta}}(\boldsymbol{x}) \} - \boldsymbol{\theta}^* \quad (2.2a)$$

$$= \int_{\Omega} (\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}^*) p(\boldsymbol{x}; \boldsymbol{\theta}^*) d\boldsymbol{x}, \quad (2.2b)$$

où Ω représente l'espace des observations (d'une manière générale, $\Omega \subset \mathbb{R}^N$). □

Le biais correspond donc à une erreur systématique commise par l'estimateur $\hat{\boldsymbol{\theta}}$: on dira qu'un estimateur est biaisé si $\mathbf{biais}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) \neq 0$.

On s'intéresse souvent au biais asymptotique, défini comme la limite du biais d'un estimateur lorsque le nombre d'observations N , ou le rapport signal sur bruit (RSB), suivant le contexte, tend vers l'infini. On parle donc d'estimateur « asymptotiquement non biaisé » lorsque $\lim_{N \rightarrow +\infty} \mathbf{biais}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) = 0$ (ou $\lim_{\text{RSB} \rightarrow +\infty} \mathbf{biais}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) = 0$).

Si l'on dispose de deux estimateurs (asymptotiquement) non biaisés, celui des deux qui aura la plus faible dispersion autour de sa valeur moyenne sera préféré. Les grandeurs qui quantifient cette dispersion autour de la valeur moyenne sont la variance (dans le cas d'un paramètre θ scalaire) et la matrice de covariance (dans le cas d'un vecteur de paramètres θ) :

Définition 2 : Matrice de covariance et variance

La matrice de covariance d'un estimateur $\widehat{\theta}$ du vecteur de paramètres θ^* (de taille R) est définie comme la matrice de taille $R \times R$

$$\mathbf{Cov}(\widehat{\theta}, \theta^*) \triangleq \mathbb{E}_{\mathbf{x}; \theta^*} \left\{ \left(\widehat{\theta}(\mathbf{x}) - \mathbb{E}_{\mathbf{x}; \theta^*} \{ \widehat{\theta}(\mathbf{x}) \} \right) \left(\widehat{\theta}(\mathbf{x}) - \mathbb{E}_{\mathbf{x}; \theta^*} \{ \widehat{\theta}(\mathbf{x}) \} \right)^\top \right\} \quad (2.3a)$$

$$= \int_{\Omega} \left(\widehat{\theta}(\mathbf{x}) - \mathbb{E}_{\mathbf{x}; \theta^*} \{ \widehat{\theta}(\mathbf{x}) \} \right) \left(\widehat{\theta}(\mathbf{x}) - \mathbb{E}_{\mathbf{x}; \theta^*} \{ \widehat{\theta}(\mathbf{x}) \} \right)^\top p(\mathbf{x}; \theta^*) d\mathbf{x}. \quad (2.3b)$$

Lorsque le paramètre à estimer est scalaire, la matrice de covariance se réduit à la variance de l'estimateur $\widehat{\theta}$:

$$\text{var}(\widehat{\theta}, \theta^*) \triangleq \mathbb{E}_{\mathbf{x}; \theta^*} \left\{ \left(\widehat{\theta}(\mathbf{x}) - \mathbb{E}_{\mathbf{x}; \theta^*} \{ \widehat{\theta}(\mathbf{x}) \} \right)^2 \right\}. \quad (2.4)$$

□

Notons que par construction, la variance est une grandeur positive, et que la matrice de covariance est symétrique, semi-définie positive.

On peut également définir l'*erreur quadratique moyenne* (EQM) d'un estimateur $\widehat{\theta}$ comme la moyenne du carré des erreurs sur toutes les valeurs de \mathbf{x} , pondérée elle aussi par la vraisemblance des observations $p(\mathbf{x}; \theta^*)$:

Définition 3 : Erreur quadratique moyenne (EQM)

L'erreur quadratique moyenne d'un estimateur $\widehat{\theta}$ du vecteur de paramètres θ^* est définie comme la matrice de taille $R \times R$

$$\mathbf{EQM}(\widehat{\theta}, \theta^*) \triangleq \mathbb{E}_{\mathbf{x}; \theta^*} \left\{ \left(\widehat{\theta}(\mathbf{x}) - \theta^* \right) \left(\widehat{\theta}(\mathbf{x}) - \theta^* \right)^\top \right\} = \mathbb{E}_{\mathbf{x}; \theta^*} \left\{ \varepsilon(\widehat{\theta}(\mathbf{x}), \theta^*) \varepsilon^\top(\widehat{\theta}(\mathbf{x}), \theta^*) \right\} \quad (2.5a)$$

$$= \int_{\Omega} \left(\widehat{\theta}(\mathbf{x}) - \theta^* \right) \left(\widehat{\theta}(\mathbf{x}) - \theta^* \right)^\top p(\mathbf{x}; \theta^*) d\mathbf{x}. \quad (2.5b)$$

Lorsque le paramètre à estimer est scalaire, l'EQM se réduit aussi à une grandeur scalaire :

$$\text{eqm}(\widehat{\theta}, \theta^*) \triangleq \mathbb{E}_{\mathbf{x}; \theta^*} \left\{ \left(\widehat{\theta}(\mathbf{x}) - \theta^* \right)^2 \right\}. \quad (2.6)$$

De la même façon que la variance, l'EQM est une grandeur positive, et de même que la matrice de covariance, la matrice des EQM est symétrique, semi-définie positive. □

De la même façon que la variance, il est clair que l'EQM est une grandeur positive, et de même que la matrice de covariance, la matrice des EQM est symétrique, définie positive.

Remarque : Les notions d'EQM, de biais et de (co)variance, sont liées par la relation

$$\mathbf{EQM}(\widehat{\theta}, \theta^*) = \mathbf{biais}(\widehat{\theta}, \theta^*) \mathbf{biais}^\top(\widehat{\theta}, \theta^*) + \mathbf{Cov}(\widehat{\theta}, \theta^*) \quad (2.7)$$

Ainsi, pour un estimateur non biaisé, les notions d'EQM et de (co)variance coïncident. □

Ces définitions permettent de rendre compte de la qualité d'un estimateur. Néanmoins l'information la plus complète est donnée par la distribution (densité de probabilité, fonction de masse ou de répartition) de l'estimateur en question, puisque c'est elle qui décrit entièrement son comportement. Si l'on connaît cette distribution, il est généralement facile d'accéder au biais,

à la variance ou à l'EQM de l'estimateur. Ainsi, lorsqu'on le peut, on cherche à calculer cette distribution. Mais à moins de s'intéresser à un estimateur particulièrement simple, le calcul de sa distribution est souvent compliqué. Cela est d'autant plus vrai que la dimension de l'espace des paramètres Θ (soit le nombre de paramètres à estimer R) est grand. On peut aboutir à une expression difficilement interprétable, ou non analytique, c'est-à-dire qui n'admet pas de forme explicite, de l'estimateur considéré. Dans de pareils cas, on peut être amené à s'intéresser uniquement à ses moments, en particulier son biais et sa variance.

Dans l'approche fréquentiste de l'estimation de paramètres, on s'intéresse particulièrement aux performances *asymptotiques* des estimateurs, c'est-à-dire lorsque le nombre d'observations N utilisées dans la procédure d'estimation tend vers l'infini. En effet, on peut intuitivement penser que plus le nombre d'échantillons sur lequel repose l'estimation est grand, plus riche sera l'information apportée sur le vecteur de paramètres à estimer θ , et meilleure sera la qualité d'estimation. Cela conduit à la notion de *consistance* d'un estimateur :

Définition 4 : Consistance d'un estimateur

Un estimateur $\hat{\theta}$ du vecteur de paramètres θ^* est dit consistant (au sens large) s'il converge en probabilité vers la vraie valeur du paramètre θ^* lorsque le nombre N d'échantillons sur lequel il se base tend vers l'infini, c'est-à-dire :

$$\forall \varepsilon > 0, \forall \theta^* \in \Theta, \lim_{N \rightarrow +\infty} \Pr\left(\|\hat{\theta}_{(N)} - \theta^*\| < \varepsilon\right) = 1 \quad (2.8)$$

où $\hat{\theta}_{(N)}$ signifie que l'estimateur $\hat{\theta}$ est basé sur N échantillons. □

Au vu de ces critères de performance, on peut attendre d'un estimateur "idéal" qu'il soit :

- non biaisé² ;
- de variance minimale (ou, de façon équivalente pour un estimateur non biaisé, d'EQM minimale).

Mais comment déterminer la variance/l'EQM minimale qu'un estimateur peut atteindre³ ? La réponse à cette question est donnée par les bornes de Cramér-Rao, qui apparaissent dans l'énoncé du théorème suivant [Fre43, Dar45, Rao45, Cra46, Van68], dont la démonstration sera donnée dans la section 2.1.5.

2.1.2.1.2 Bornes de Cramér-Rao

Théorème 1 (Bornes de Cramér-Rao). *On suppose que la fonction de log-vraisemblance, définie par $\ln p(\mathbf{x}; \theta)$, vérifie la condition de régularité suivante :*

$$\forall \theta^* \in \Theta, \mathbb{E}_{\mathbf{x}; \theta} \left\{ \left. \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right|_{\theta^*} \right\} = \mathbf{0}. \quad (2.9)$$

Alors, quel que soit l'estimateur localement⁴ non biaisé $\hat{\theta}$ de θ^* , sa matrice de covariance vérifie

$$\mathbf{Cov}(\hat{\theta}, \theta^*) \succcurlyeq \mathbf{F}^{-1}(\theta^*), \quad (2.10)$$

2. En autorisant un biais non nul, il est néanmoins possible d'obtenir un estimateur de variance plus faible que celle d'un estimateur à biais nul. Il est donc possible de jouer sur ce compromis suivant les propriétés de l'estimateur recherchées.

3. Nous avons vu que la variance est une grandeur positive, donc 0 en est un minorant. Toutefois, on ne peut raisonnablement pas attendre d'un estimateur qu'il soit de variance identiquement nulle : ce serait un estimateur qui donne systématiquement la vraie valeur du paramètre inconnu θ^* ; un tel estimateur est dit clairvoyant mais n'existe bien sûr pas en pratique.

4. La notion de biais nul *localement* est importante, et signifie mathématiquement que i) $\mathbf{biais}(\hat{\theta}, \theta^*) = 0$, et ii) $\left. \frac{\partial \mathbf{biais}(\hat{\theta}, \theta)}{\partial \theta} \right|_{\theta^*} = \mathbf{0}$ (voir [Gla72, FL02, Cha04]).

où l'inégalité « \succeq » signifie que la différence entre le premier et le second membre est une matrice semi-définie positive, et où $\mathbf{F}(\boldsymbol{\theta}^*)$ désigne la matrice d'information de Fisher, définie par

$$\mathbf{F}(\boldsymbol{\theta}^*) \triangleq \mathbb{E}_{\mathbf{x};\boldsymbol{\theta}^*} \left\{ \left. \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \left. \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right\}, \quad (2.11)$$

où les dérivées sont évaluées en la vraie valeur du paramètre $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, et sous réserve d'existence des espérances dans (2.11) de la matrice $\mathbf{F}^{-1}(\boldsymbol{\theta}^*)$.

Remarques :

- i) Si les dérivées secondes de la log-vraisemblance existent, la matrice d'information de Fisher est également donnée, grâce à (2.9), par

$$\mathbf{F}(\boldsymbol{\theta}^*) = - \mathbb{E}_{\mathbf{x};\boldsymbol{\theta}^*} \left\{ \left. \frac{\partial^2 \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right|_{\boldsymbol{\theta}^*} \right\}; \quad (2.12)$$

- ii) Comme l'estimateur $\widehat{\boldsymbol{\theta}}$ est supposé non biaisé, la matrice de covariance peut être remplacée par la matrice des EQM dans (2.10);
- iii) Ce que l'on appelle la borne de Cramér-Rao (BCR) correspond donc finalement à l'inverse de la matrice d'information de Fisher :

$$\mathbf{BCR}(\boldsymbol{\theta}^*) \triangleq \mathbf{F}^{-1}(\boldsymbol{\theta}^*) \quad (2.13)$$

et ce que l'on appelle les bornes de Cramér-Rao correspond en fait aux éléments diagonaux de la matrice $\mathbf{BCR}(\boldsymbol{\theta}^*)$, puisque l'inégalité (2.10) implique en particulier, pour $i = 1, \dots, R$,

$$[\mathbf{Cov}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*)]_{i,i} \geq [\mathbf{BCR}(\boldsymbol{\theta}^*)]_{i,i}, \quad (2.14a)$$

ou

$$[\mathbf{EQM}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*)]_{i,i} \geq [\mathbf{BCR}(\boldsymbol{\theta}^*)]_{i,i}; \quad (2.14b)$$

- iv) Dans le cas scalaire ($R = 1$), l'inégalité (2.10) se réduit à

$$\text{var}(\widehat{\theta}, \theta^*) \geq \frac{1}{\mathbb{E}_{\mathbf{x};\boldsymbol{\theta}^*} \left\{ \left[\left. \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right|_{\theta^*} \right]^2 \right\}} = - \frac{1}{\mathbb{E}_{\mathbf{x};\boldsymbol{\theta}^*} \left\{ \left. \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right|_{\theta^*} \right\}} \quad (2.15)$$

- v) Dans le cas où le signal \mathbf{x} consiste en N observations x_n , $n = 1, \dots, N$, indépendantes, identiquement distribuées (i.i.d.), alors la matrice d'information de Fisher pour le signal observé \mathbf{x} peut s'écrire

$$\mathbf{F}(\boldsymbol{\theta}^*) = N \widetilde{\mathbf{F}}(\boldsymbol{\theta}^*) \quad (2.16)$$

où $\widetilde{\mathbf{F}}(\boldsymbol{\theta}^*)$ désigne la matrice d'information de Fisher sur le vecteur $\boldsymbol{\theta}^*$ pour un seul échantillon x du signal observé, c'est-à-dire

$$\widetilde{\mathbf{F}}(\boldsymbol{\theta}^*) = \mathbb{E}_{x;\boldsymbol{\theta}^*} \left\{ \left. \frac{\partial \ln p(x; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}^*} \left. \frac{\partial \ln p(x; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right|_{\boldsymbol{\theta}^*} \right\}, \quad (2.17)$$

où x est ici une variable aléatoire scalaire. □

L'un des intérêts des bornes de Cramér-Rao est que l'on peut souvent en obtenir des formes explicites pour des problèmes courants en traitement du signal. Par exemple, un cas extrêmement courant est celui de signaux gaussiens, dont la moyenne et/ou la matrice de covariance sont des fonctions des paramètres à estimer. Considérons par exemple un signal à valeurs complexes, distribué selon une loi gaussienne complexe circulaire : $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}(\boldsymbol{\theta}^*), \mathbf{C}(\boldsymbol{\theta}^*))$. La fonction densité de probabilité de ces observations est donnée par

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{\pi^N \det(\mathbf{C}(\boldsymbol{\theta}))} \exp\left\{-\left(\mathbf{x} - \mathbf{m}(\boldsymbol{\theta})\right)^H \mathbf{C}^{-1}(\boldsymbol{\theta}) \left(\mathbf{x} - \mathbf{m}(\boldsymbol{\theta})\right)\right\}. \quad (2.18)$$

Les éléments de la matrice d'information de Fisher sont alors donnés par la formule dite de « Slepian-Bangs » [Sle54, Ban71], [Kay93, pp.563-565], soit pour $1 \leq i, j \leq R$:

$$[\mathbf{F}(\boldsymbol{\theta}^*)]_{i,j} = 2 \operatorname{Re} \left[\frac{\partial \mathbf{m}^H(\boldsymbol{\theta})}{\partial \theta_i} \Big|_{\boldsymbol{\theta}^*} \mathbf{C}^{-1}(\boldsymbol{\theta}^*) \frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \theta_j} \Big|_{\boldsymbol{\theta}^*} \right] + \operatorname{tr} \left[\mathbf{C}^{-1}(\boldsymbol{\theta}^*) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i} \Big|_{\boldsymbol{\theta}^*} \mathbf{C}^{-1}(\boldsymbol{\theta}^*) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_j} \Big|_{\boldsymbol{\theta}^*} \right]. \quad (2.19)$$

Les bornes de Cramér-Rao fournissent donc, parfois facilement, l'EQM ou la variance minimale qu'un estimateur non biaisé peut espérer atteindre. Notons donc qu'elles ne dépendent pas d'un estimateur en particulier, mais qu'elles sont valables pour un ensemble d'estimateurs, en l'occurrence, pour les BCR, les estimateurs non biaisés. On verra aussi qu'elles peuvent dans certains cas permettre de déterminer un bon estimateur. De plus, si la théorie permet d'affirmer qu'un estimateur atteint la borne, alors le calcul de celle-ci fournit directement une valeur de la variance de cet estimateur. Un estimateur non biaisé qui atteint la BCR est dit « efficace » (en traitement du signal). Lorsqu'un estimateur est seulement asymptotiquement (i.e., à grand nombre d'observations ou grand RSB) non-biaisé, et atteint seulement asymptotiquement la BCR, on parle d'estimateur « asymptotiquement efficace ».

2.1.1.2 Estimateurs classiques

2.1.1.2.1 Estimateur MVU Nous avons donc vu deux critères possibles pour caractériser un bon estimateur : un biais nul et une variance minimale, c'est-à-dire égale à la borne de Cramér-Rao. Même si dans la pratique, on ne pourra pas toujours satisfaire ces deux critères, ils permettent néanmoins de définir un estimateur optimal, à savoir l'estimateur non-biaisé à minimum de variance, qui sera dans la suite abrégé par MVU (pour *Minimum Variance Unbiased*).

Définition 5 : Estimateur MVU

S'il existe, l'estimateur MVU est celui qui vérifie les deux propriétés :

- i) $\mathbb{E}_{\mathbf{x}; \boldsymbol{\theta}^*} \{\hat{\boldsymbol{\theta}}^{\text{MVU}}(\mathbf{x})\} = \boldsymbol{\theta}^*$;
- ii) $\forall \hat{\boldsymbol{\theta}} \neq \hat{\boldsymbol{\theta}}^{\text{MVU}}, \forall \boldsymbol{\theta}^* \in \Theta, \mathbf{Cov}(\hat{\boldsymbol{\theta}}^{\text{MVU}}, \boldsymbol{\theta}^*) \preceq \mathbf{Cov}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*)$.

□

Le théorème suivant [Kay93, p.44] donne une condition nécessaire et suffisante pour que la BCR soit atteinte et pour en déduire une expression de l'estimateur MVU. Une version plus complète de ce théorème peut être trouvée dans [LC03, p.121].

Théorème 2. *On peut trouver un estimateur non biaisé qui atteint la borne de Cramér-Rao (c'est-à-dire qui vérifie $\mathbf{Cov}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) = \mathbf{F}^{-1}(\boldsymbol{\theta}^*)$) si et seulement si*

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}^*} = \mathbf{F}(\boldsymbol{\theta}^*) (\mathbf{g}(\mathbf{x}) - \boldsymbol{\theta}^*) \quad (2.20)$$

pour une fonction vectorielle $\mathbf{g} : \Omega \rightarrow \mathbb{R}^R$. Cet estimateur est l'estimateur MVU et est donné par $\hat{\boldsymbol{\theta}}(\mathbf{x}) = \mathbf{g}(\mathbf{x})$.

Ce théorème fournit un outil puissant pour déterminer l'estimateur MVU, et illustre l'utilité des BCR pour cela. Toutefois, la condition (2.20) n'est satisfaite en pratique que dans le cas de modèles linéaires. Mentionnons en outre que même si la condition (2.20) n'est pas satisfaite, l'estimateur MVU peut tout de même exister, sans que celui-ci ne soit efficace. Cela passe alors par la recherche d'une statistique exhaustive. Pour plus de détails sur ce type d'approche, nous reportons le lecteur à [Kay93, Chapitre 5].

2.1.2.2.2 Estimateur du maximum de vraisemblance Lorsque l'estimateur MVU n'existe pas, ou qu'il est très difficile à dériver, une alternative très souvent utilisée dans la pratique est l'estimateur du *maximum de vraisemblance*, que nous abrègerons dans la suite par ML (pour *Maximum Likelihood*, en anglais). Son principe est assez intuitif, puisqu'il s'agit de rechercher la valeur du paramètre $\boldsymbol{\theta}$ qui rend l'observation du signal \mathbf{x} la plus vraisemblable, c'est-à-dire qui maximise la fonction de vraisemblance $p(\mathbf{x}; \boldsymbol{\theta})$. Notons bien ici que même si $p(\mathbf{x}; \boldsymbol{\theta})$ est une fonction de \mathbf{x} (ce qui implique bien sûr, comme on l'a vu, la dépendance de l'estimateur en \mathbf{x}), la valeur de \mathbf{x} est ici fixée puisqu'il s'agit du signal que l'on a observé. Dans ce cas, la vraisemblance correspond à $p(\mathbf{x}; \boldsymbol{\theta})$ vue comme une fonction de $\boldsymbol{\theta}$. Formellement, il peut donc être défini comme suit :

Définition 6 : Estimateur du maximum de vraisemblance (ML)

L'estimateur du maximum de vraisemblance, noté $\hat{\boldsymbol{\theta}}^{\text{ML}}$, est défini comme

$$\hat{\boldsymbol{\theta}}^{\text{ML}}(\mathbf{x}) \triangleq \arg \max_{\boldsymbol{\theta}} p(\mathbf{x}; \boldsymbol{\theta}). \quad (2.21)$$

□

L'une des propriétés importantes de l'estimateur ML est que, sous certaines hypothèses, il est asymptotiquement efficace. Plus précisément, on peut énoncer le théorème suivant.

Théorème 3 (Propriétés asymptotiques de l'estimateur ML). *Si la fonction de vraisemblance vérifie les conditions de régularité énoncées dans [LC03, Théorème 5.10, p.463], alors l'estimateur du maximum de vraisemblance $\hat{\boldsymbol{\theta}}_{(N)}^{\text{ML}}$ (basé sur N observations) du vecteur de paramètres $\boldsymbol{\theta}$ est asymptotiquement efficace, dans le sens où il vérifie*

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{(N)}^{\text{ML}} - \boldsymbol{\theta}^*) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \tilde{\mathbf{F}}^{-1}(\boldsymbol{\theta}^*)), \quad (2.22)$$

lorsque N tend vers l'infini [Ser09].

Remarque : Parmi les conditions de régularité énoncées dans [LC03, Théorème 5.10, p.463], on notera que figure l'existence d'au moins trois dérivées successives de la log-vraisemblance $\ln p(\mathbf{x}; \boldsymbol{\theta})$ par rapport aux paramètres. Ainsi, pour des paramètres discrets, ces dérivées n'existent pas, donc cette condition n'est pas satisfaite. Le théorème 3 ne s'applique donc pas dans ce cas. □

Le théorème 3 énoncé ci-dessus concerne la convergence de l'estimateur ML en terme du nombre d'échantillons N . Nous pouvons mentionner qu'il existe des résultats de convergence de l'estimateur ML en terme du rapport signal sur bruit (RSB). C'est le cas en particulier pour un signal \mathbf{x} gaussien, distribué selon le modèle⁵ $\mathbf{x} \sim \mathcal{N}(\mathbf{m}(\boldsymbol{\theta}^*), \sigma^2 \mathbf{C})$, dont la moyenne est une fonction du vecteur de paramètres inconnu $\boldsymbol{\theta}$. Il a été montré que lorsque le RSB tend vers l'infini

5. Le résultat est ici énoncé pour un signal \mathbf{x} réel; il a également été démontré dans le cas complexe circulaire.

(i.e., lorsque σ^2 tend vers 0), on obtient $(1/\sigma)(\widehat{\boldsymbol{\theta}}^{\text{ML}} - \boldsymbol{\theta}^*) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, (1/\sigma^2)\mathbf{F}^{-1}(\boldsymbol{\theta}^*))$, c'est-à-dire que l'estimateur ML est asymptotiquement gaussien et efficace [RFCL06]. Cela n'est par contre plus valable par exemple dans le cas où la matrice de covariance est elle aussi fonction du vecteur de paramètres à estimer $\boldsymbol{\theta}^*$ [RFBL07].

Ainsi, dès lors que l'on dispose de la fonction de vraisemblance, il semble possible d'implémenter l'estimateur ML, puisqu'il s'agit de maximiser une fonction, ce qui correspond à un problème relativement standard, en tout cas largement documenté. En effet, s'il est parfois possible d'en obtenir une expression analytique, on doit en revanche recourir à des méthodes numériques d'optimisation multidimensionnelle lorsque ce n'est pas le cas. De plus, on ne dispose pas non plus toujours d'expressions analytiques de son biais ou de sa variance. Pour en obtenir des valeurs approchées, on recourt en général à des méthodes de simulation dites « de Monte-Carlo ». Elles consistent à générer indépendamment un certain nombre de réalisations N_{MC} du signal \mathbf{x} (réalisations que l'on notera $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N_{\text{MC}})}$) selon le modèle spécifié, et pour chacune d'elle, à procéder à l'estimation (ML ou autre), donnant ainsi lieu à une séquence d'estimées $\widehat{\boldsymbol{\theta}}(\mathbf{x}^{(1)}), \widehat{\boldsymbol{\theta}}(\mathbf{x}^{(2)}), \dots, \widehat{\boldsymbol{\theta}}(\mathbf{x}^{(N_{\text{MC}})})$. À partir de cette dernière, on peut obtenir une estimation de l'EQM (ou de la même façon, du biais, de la variance...) par la moyenne empirique des erreurs au carré, soit

$$\widehat{\text{EQM}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} (\widehat{\boldsymbol{\theta}}(\mathbf{x}^{(i)}) - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}(\mathbf{x}^{(i)}) - \boldsymbol{\theta})^{\text{T}}. \quad (2.23)$$

Cette estimation a loi des grands nombres, qui stipule que le membre de droite de (2.23) tend vers le membre de gauche lorsque N_{MC} tend vers l'infini. De plus, pour un nombre de simulations N_{MC} donné, l'écart-type de l'estimateur de l'EQM ainsi mis en place est proportionnel à $1/N_{\text{MC}}$. Cela implique une vitesse de convergence relativement faible, et un nombre de tirages N_{MC} important sera donc nécessaire pour obtenir une estimation fiable de la grandeur approchée.

La complexité de la procédure d'estimation mise en place (qui peut correspondre à la résolution d'un problème d'optimisation dans le cas de l'estimateur ML) combinée au nombre N_{MC} souvent important de simulations de Monte-Carlo nécessaire à l'obtention d'une bonne estimation de l'EQM se solde par un coût calculatoire qui peut être prohibitif. C'est une autre justification de l'intérêt de disposer de bornes inférieures de l'EQM.

En résumé, plusieurs raisons nous poussent à nous intéresser au calcul de bornes inférieures de l'EQM : premièrement, elles dépendent uniquement du problème d'estimation considéré, et pas d'un estimateur en particulier (elles sont valables pour une famille d'estimateurs, par exemple non biaisés). Deuxièmement, on peut dans certains cas prouver qu'une borne peut être atteinte, éventuellement pour un nombre d'observations qui tend vers l'infini (voir Théorème 3) [Kay93, LC03] ou à fort rapport signal à bruit (RSB) [RFCL06]. Cela signifie alors que cette borne fournit une valeur précise de l'EQM minimale théorique dans ces régimes asymptotiques. Enfin, de telles bornes peuvent souvent s'exprimer sous une forme analytique (ou, comme on le verra pour d'autres bornes que la BCR, semi-analytique), par exemple pour des distributions des observations gaussiennes [Sle54, Ban71, Kay93], elliptiques [PR10, BA13, GG13] ou autres [TRB⁺14]. Cela signifie qu'une évaluation de l'EQM minimale théorique peut être effectuée avec un faible coût de calcul. Dans tous les cas, elles forment une référence grâce à laquelle on peut comparer différents estimateurs.

Avant de présenter d'autres bornes inférieures de l'EQM que la BCR, nous présentons les autres cadres d'estimation considérés dans cette thèse : les points de vue bayésiens et hybrides.

2.1.3 Estimation de paramètres aléatoires (approche bayésienne)

Dans l'approche bayésienne, on suppose que le vecteur de paramètres inconnus θ est un vecteur aléatoire. Le fait de « probabiliser l'inconnu » implique ici d'attribuer à θ une loi *a priori* $\pi(\theta)$ qui résume l'information disponible sur θ avant la procédure d'estimation.

2.1.3.1 Critères de performance en estimation bayésienne

En estimation bayésienne, les critères de performance se basent sur une fonction positive dite « de coût », que l'on pourra noter $C(\theta, \hat{\theta})$; elle peut correspondre à un type d'erreur de l'estimateur $\hat{\theta}$ lorsque le paramètre à estimer vaut θ , mais pas seulement. Plus généralement, c'est une fonction qui pénalise une erreur de l'estimée $\hat{\theta}(x)$. Comme cette dernière est une variable aléatoire qui dépend du signal observé x , on peut comme on l'a fait en utilisant le point de vue fréquentiste, moyenner ce coût par rapport à la vraisemblance des observations (cette fois notée $p(x|\theta)$, puisque θ est ici supposé aléatoire) sur toutes les observations possibles x , pour obtenir une évaluation plus objective de l'estimateur $\hat{\theta}$ qui ne dépendrait ainsi pas de l'expérience menée. Cela aboutit à la notion de « risque », que l'on peut qualifier de « fréquentiste », et noté $R(\theta, \hat{\theta})$:

$$R(\theta, \hat{\theta}) \triangleq \mathbb{E}_{x|\theta} \{C(\theta, \hat{\theta}(x))\} = \int_{\Omega} C(\theta, \hat{\theta}(x)) p(x|\theta) dx. \quad (2.24)$$

On peut retrouver par exemple la notion d'EQM (souvent appelée « risque quadratique » dans la littérature bayésienne) que nous avons vue dans la section 2.1.2.1 en définissant le coût comme l'erreur quadratique : si $C(\theta, \hat{\theta}(x)) = (\hat{\theta}(x) - \theta)(\hat{\theta}(x) - \theta)^T$, alors $R(\theta, \hat{\theta}) = \mathbf{EQM}(\hat{\theta}, \theta)$.

Précisons maintenant qu'il s'agit de l'EQM *locale*, c'est-à-dire pour une valeur du vecteur de paramètre égale à θ . Ainsi, on voit que pour un estimateur $\hat{\theta}$ donné, l'EQM est une fonction de la valeur prise par le vecteur de paramètre θ . Ainsi, si l'on souhaite comparer l'EQM de deux estimateurs $\hat{\theta}_1$ et $\hat{\theta}_2$, on peut très bien avoir $\mathbf{EQM}(\hat{\theta}_1, \theta) \preceq \mathbf{EQM}(\hat{\theta}_2, \theta)$ pour un vecteur de paramètre égal à θ , et $\mathbf{EQM}(\hat{\theta}_1, \theta') \succeq \mathbf{EQM}(\hat{\theta}_2, \theta')$ pour un vecteur de paramètre égal à θ' . En d'autres termes, on ne dispose pas d'une relation d'ordre totale pour toutes les valeurs possibles de θ .

Dans le cadre bayésien, comme on dispose d'une loi *a priori* sur le paramètre θ , on peut utiliser la règle de Bayes pour relier la loi *a priori* $\pi(\theta)$ ainsi que la vraisemblance des observations $p(x|\theta)$ à la distribution *a posteriori* du vecteur de paramètres θ , c'est-à-dire conditionnée par l'observation du signal x , qui sera notée $\pi(\theta|x)$:

$$\pi(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{p(x)} = \frac{p(x|\theta)\pi(\theta)}{\int_{\Theta} p(x|\theta)\pi(\theta)d\theta}, \quad (2.25)$$

et sur laquelle se basera l'inférence concernant θ .

En fait, l'approche bayésienne se distingue de l'approche fréquentiste en ce qu'elle vise non pas à minimiser le risque fréquentiste $R(\theta, \hat{\theta})$ (qui moyenne le coût sur toutes les valeurs possibles des observations x , par rapport à la vraisemblance $p(x|\theta)$), mais à minimiser un risque *a posteriori*, noté $\tilde{R}(\hat{\theta}|x)$, défini comme une moyenne du coût par rapport à la loi *a posteriori* $\pi(\theta|x)$, sur toutes les valeurs possibles de θ :

$$\tilde{R}(\hat{\theta}|x) \triangleq \mathbb{E}_{\theta|x} \{C(\theta, \hat{\theta}(x))\} = \int_{\Theta} C(\theta, \hat{\theta}(x)) \pi(\theta|x) d\theta, \quad (2.26)$$

qui donne donc une valeur *conditionnée* par celle observée du signal, x . Cette définition se justifie par le fait que θ est aléatoire et inconnu, il est donc préférable que le critère de qualité de $\hat{\theta}$ n'en

dépende pas. Il dépend en revanche de \mathbf{x} , qui lui est connu puisque c'est le signal observé. Ainsi, l'approche bayésienne vise à mieux prendre en compte l'observation \mathbf{x} dont on dispose, à l'inverse de l'approche fréquentiste dont le risque $\mathbf{R}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ évalue les estimateurs pour toutes les valeurs observées *possibles*, et pas seulement celle effectivement observée.

Cependant, on peut aussi tout à fait s'intéresser aux performances *globales*, c'est-à-dire à la fois pour toutes les valeurs de \mathbf{x} et toutes les valeurs de $\boldsymbol{\theta}$ possibles. Cela nous conduit à définir le risque *intégré*.

Définition 7 : Risque intégré

Le risque intégré correspond au coût moyen, pondéré par la distribution jointe du signal observé \mathbf{x} et du vecteur de paramètres $\boldsymbol{\theta}$, notée $p(\mathbf{x}, \boldsymbol{\theta})$, pour toutes leurs valeurs possibles :

$$\bar{\mathbf{R}}(\hat{\boldsymbol{\theta}}) \triangleq \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \{ \mathbf{C}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) \} = \int_{\Theta \times \Omega} \mathbf{C}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) p(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{x}. \quad (2.27)$$

□

Cette mesure de performance offre l'avantage d'associer un nombre réel à un estimateur $\hat{\boldsymbol{\theta}}$, au lieu d'une fonction de $\boldsymbol{\theta}$ ou de \mathbf{x} . Cela induit donc d'une relation d'ordre total pour comparer les performances globales de différents estimateurs.

Avant de présenter quelques estimateurs bayésiens classiques, remarquons le lien entre les trois types de risque définis ci-dessus en (2.24), (2.26) et (2.27) : en faisant apparaître la loi *a priori* du vecteur $\boldsymbol{\theta}$, on voit que le risque intégré correspond au risque fréquentiste moyenné par rapport à cette loi *a priori* $\pi(\boldsymbol{\theta})$:

$$\begin{aligned} \bar{\mathbf{R}}(\hat{\boldsymbol{\theta}}) &= \int_{\Theta} \int_{\Omega} \mathbf{C}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) p(\mathbf{x} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\mathbf{x} d\boldsymbol{\theta} \\ &= \int_{\Theta} \pi(\boldsymbol{\theta}) \left(\int_{\Omega} \mathbf{C}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) p(\mathbf{x} | \boldsymbol{\theta}) d\mathbf{x} \right) d\boldsymbol{\theta} \\ &= \mathbb{E}_{\boldsymbol{\theta}} \{ \mathbb{E}_{\mathbf{x} | \boldsymbol{\theta}} \{ \mathbf{C}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) \} \} \\ &= \mathbb{E}_{\boldsymbol{\theta}} \{ \mathbf{R}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) \}. \end{aligned} \quad (2.28)$$

Enfin, en utilisant la règle de Bayes et le théorème de Fubini (la fonction de coût $\mathbf{C}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x}))$ est positive), on montre qu'un estimateur qui minimise le risque intégré correspond à un estimateur obtenu en sélectionnant les estimées $\hat{\boldsymbol{\theta}}(\mathbf{x})$, pour toutes les valeurs de \mathbf{x} parcourant l'espace Ω , d'un estimateur qui minimise le coût moyen *a posteriori*. En effet,

$$\begin{aligned} \bar{\mathbf{R}}(\hat{\boldsymbol{\theta}}) &= \int_{\Theta} \int_{\Omega} \mathbf{C}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) p(\mathbf{x} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\mathbf{x} d\boldsymbol{\theta} \\ &= \int_{\Omega} \int_{\Theta} \mathbf{C}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) \pi(\boldsymbol{\theta} | \mathbf{x}) p(\mathbf{x}) d\boldsymbol{\theta} d\mathbf{x} \\ &= \int_{\Omega} \left(\int_{\Theta} \mathbf{C}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) \pi(\boldsymbol{\theta} | \mathbf{x}) d\boldsymbol{\theta} \right) p(\mathbf{x}) d\mathbf{x} \\ &= \int_{\Omega} \tilde{\mathbf{R}}(\hat{\boldsymbol{\theta}} | \mathbf{x}) p(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (2.29)$$

2.1.3.2 Estimateurs bayésiens

À partir des notions de risque définis ci-dessus, l'inférence bayésienne propose des estimateurs optimaux par rapport à ces risques, en particulier le risque intégré. Pour une fonction de coût $\mathbf{C}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ et une distribution *a priori* $\pi(\boldsymbol{\theta})$ données, un estimateur qui minimise le risque intégré

$\tilde{\mathbf{R}}(\hat{\boldsymbol{\theta}})$ est appelé « estimateur de Bayes ». Comme on l'a vu, cet estimateur minimise, pour tout $\mathbf{x} \in \Omega$, le coût moyen *a posteriori* $\tilde{\mathbf{R}}(\hat{\boldsymbol{\theta}} | \mathbf{x})$. Nous énumérons ici quelques estimateurs de Bayes courants associés à différentes fonctions de coût.

2.1.3.2.1 Estimateur de la médiane *a posteriori* (cas scalaire) Nous ne considérons dans cette section que le cas d'un paramètre à estimer θ scalaire réel ($\Theta = \mathbb{R}$). On considère comme fonction de coût l'erreur absolue, c'est-à-dire $\mathbf{C}(\theta, \hat{\theta}(\mathbf{x})) = |\hat{\theta}(\mathbf{x}) - \theta|$. Le coût moyen *a posteriori* s'écrit

$$\begin{aligned} \tilde{\mathbf{R}}(\hat{\theta} | \mathbf{x}) &= \int_{\Theta} |\hat{\theta}(\mathbf{x}) - \theta| \pi(\theta | \mathbf{x}) d\theta \\ &= \int_{-\infty}^{\hat{\theta}(\mathbf{x})} (\hat{\theta}(\mathbf{x}) - \theta) \pi(\theta | \mathbf{x}) d\theta + \int_{\hat{\theta}(\mathbf{x})}^{+\infty} (\theta - \hat{\theta}(\mathbf{x})) \pi(\theta | \mathbf{x}) d\theta \\ &= \int_{-\infty}^{\hat{\theta}(\mathbf{x})} \Pr(\theta < \theta | \mathbf{x} = \mathbf{x}) d\theta + \int_{\hat{\theta}(\mathbf{x})}^{+\infty} \Pr(\theta > \theta | \mathbf{x} = \mathbf{x}) d\theta \end{aligned} \quad (2.30)$$

où (2.30) est obtenue par intégration par parties. En dérivant (2.30) par rapport à $\hat{\theta}(\mathbf{x})$ et en annulant la dérivée obtenue, on aboutit à la condition

$$\Pr(\theta < \hat{\theta}(\mathbf{x}) | \mathbf{x} = \mathbf{x}) - \Pr(\theta > \hat{\theta}(\mathbf{x}) | \mathbf{x} = \mathbf{x}) = 0, \quad (2.31)$$

soit, comme $\Pr(\theta > \hat{\theta}(\mathbf{x}) | \mathbf{x} = \mathbf{x}) = 1 - \Pr(\theta < \hat{\theta}(\mathbf{x}) | \mathbf{x} = \mathbf{x})$,

$$\Pr(\theta < \hat{\theta}(\mathbf{x}) | \mathbf{x} = \mathbf{x}) = \frac{1}{2}. \quad (2.32)$$

L'estimateur de Bayes pour la fonction de coût « erreur absolue » est donc défini comme la médiane de la distribution *a posteriori* du paramètre θ ; nous pourrions le noter $\hat{\theta}^{\text{MedAP}}$.

Notons qu'il est difficile de généraliser cet estimateur au cas multidimensionnel, en particulier parce que la notion de médiane multivariée peut être définie de plusieurs manières (voir par exemple [NO99]).

Par ailleurs, même en dimension 1, cet estimateur peut être difficile à mettre en place en pratique, car il nécessite l'échantillonnage de la distribution *a posteriori* $\pi(\theta | \mathbf{x})$, et donc le calcul de la loi marginale $p(\mathbf{x})$. Cela implique l'emploi de plusieurs méthodes numériques d'intégration puisqu'on a

$$\Pr(\theta < \theta | \mathbf{x} = \mathbf{x}) = \frac{\int_{-\infty}^{\theta} p(\mathbf{x} | \theta') \pi(\theta') d\theta'}{\int_{-\infty}^{+\infty} p(\mathbf{x} | \theta') \pi(\theta') d\theta'}. \quad (2.33)$$

De plus, même si l'on sait que l'estimateur en question est optimal pour le risque considéré (en l'occurrence l'erreur absolue), on peut chercher à connaître son comportement dans des conditions données, et on est alors amené à utiliser des méthodes de Monte-Carlo décrites dans la section 2.1.2.2.2. Cela peut rapidement mener à un coût de calculs rédhibitoire.

2.1.3.2.2 Estimateur de la moyenne *a posteriori* On considère ici un vecteur de paramètres $\boldsymbol{\theta}$ à estimer, avec comme fonction de coût l'erreur quadratique $\mathbf{C}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) = (\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})^{\top}$ (nous avons vu qu'alors, $\mathbf{R}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \mathbf{EQM}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$). Le coût moyen *a posteriori* peut s'écrire

$$\begin{aligned} \tilde{\mathbf{R}}(\hat{\boldsymbol{\theta}} | \mathbf{x}) &= \mathbb{E}_{\boldsymbol{\theta} | \mathbf{x}} \{ (\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})^{\top} \} \\ &= \hat{\boldsymbol{\theta}}(\mathbf{x}) \hat{\boldsymbol{\theta}}^{\top}(\mathbf{x}) - 2 \mathbb{E}_{\boldsymbol{\theta} | \mathbf{x}} \{ \boldsymbol{\theta} \} \hat{\boldsymbol{\theta}}^{\top}(\mathbf{x}) + \mathbb{E}_{\boldsymbol{\theta} | \mathbf{x}} \{ \boldsymbol{\theta} \boldsymbol{\theta}^{\top} \}. \end{aligned} \quad (2.34)$$

En dérivant (2.34) par rapport à $\hat{\boldsymbol{\theta}}(\mathbf{x})$, et en annulant cette dérivée, on obtient la condition

$$\hat{\boldsymbol{\theta}}(\mathbf{x}) = \mathbb{E}_{\boldsymbol{\theta}|\mathbf{x}}\{\boldsymbol{\theta}\}, \quad (2.35)$$

qui définit l'estimateur de Bayes pour le coût quadratique comme la moyenne de la distribution *a posteriori* du vecteur de paramètres $\boldsymbol{\theta}$. Cet estimateur sera noté « estimateur MMSE », comme dans la littérature anglo-saxonne (pour « *Minimum Mean Square Error* »).

Comme dans le cas de l'estimateur de la médiane *a posteriori*, l'estimateur MMSE nécessite l'échantillonnage de la loi *a posteriori*, et par là même le calcul de la loi marginale $p(\mathbf{x})$. Ainsi l'emploi de plusieurs méthodes numériques d'intégration peut s'avérer nécessaire, qui, conjuguées à des simulations de Monte-Carlo, risquent de faire exploser le coût calculatoire.

2.1.3.2.3 Estimateur du maximum *a posteriori* Enfin, nous présentons ici un autre estimateur bayésien de référence qui est l'estimateur du maximum *a posteriori* (MAP), défini comme le mode de la distribution *a posteriori* du vecteur de paramètres $\boldsymbol{\theta}$, soit

$$\hat{\boldsymbol{\theta}}^{\text{MAP}}(\mathbf{x}) \triangleq \arg \max_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta} | \mathbf{x}). \quad (2.36)$$

Notons que puisque la distribution $p(\mathbf{x})$ ne dépend pas de $\boldsymbol{\theta}$, nous avons donc aussi, par la règle de Bayes, $\hat{\boldsymbol{\theta}}^{\text{MAP}}(\mathbf{x}) = \arg \max_{\boldsymbol{\theta}} p(\mathbf{x} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) = \arg \max_{\boldsymbol{\theta}} p(\mathbf{x}, \boldsymbol{\theta})$, de sorte que l'estimateur du MAP ne nécessite pas de calculer la loi marginale $p(\mathbf{x})$. C'est cette dernière relation qui est utilisée en pratique pour le calcul du MAP. De plus, on peut remarquer que, pour une distribution *a priori* uniforme, l'estimateur de MAP coïncide précisément avec l'estimateur du maximum de vraisemblance.

L'estimateur du MAP peut être conçu comme un cas limite d'estimateurs de Bayes associés à la fonction de coût dite « 0-1 », qui vaut 0 lorsque $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$, et 1 lorsque $\hat{\boldsymbol{\theta}} \neq \boldsymbol{\theta}$. Plus précisément, on peut définir, pour $\varepsilon > 0$, une suite de fonctions de coût comme

$$C_{\varepsilon}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) \triangleq \mathbb{1}_{]-\infty, -\varepsilon[\cup]\varepsilon, +\infty[}(\|\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\|) = \begin{cases} 1 & \text{si } \|\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\| > \varepsilon \\ 0 & \text{si } \|\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\| \leq \varepsilon, \end{cases} \quad (2.37)$$

où $\mathbb{1}_{\mathcal{A}}(x)$ est la fonction indicatrice de l'ensemble \mathcal{A} (elle vaut 1 si $x \in \mathcal{A}$, 0 sinon). On peut alors écrire le coût moyen *a posteriori* :

$$\tilde{\mathbf{R}}(\hat{\boldsymbol{\theta}} | \mathbf{x}) = \int_{\Theta} C_{\varepsilon}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) \pi(\boldsymbol{\theta} | \mathbf{x}) d\boldsymbol{\theta} = 1 - \int_{\mathcal{B}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \varepsilon)} \pi(\boldsymbol{\theta} | \mathbf{x}) d\boldsymbol{\theta}, \quad (2.38)$$

où $\mathcal{B}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \varepsilon) \triangleq \{\boldsymbol{\theta} \in \Theta \mid \|\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\| \leq \varepsilon\}$ indique la boule de centre $\hat{\boldsymbol{\theta}}(\mathbf{x})$ et de rayon ε . Pour un paramètre scalaire, elle correspond à l'intervalle $[\hat{\boldsymbol{\theta}}(\mathbf{x}) - \varepsilon, \hat{\boldsymbol{\theta}}(\mathbf{x}) + \varepsilon]$, et on voit que minimiser $\tilde{\mathbf{R}}(\hat{\boldsymbol{\theta}} | \mathbf{x})$ revient à maximiser $\int_{\hat{\boldsymbol{\theta}}(\mathbf{x}) - \varepsilon}^{\hat{\boldsymbol{\theta}}(\mathbf{x}) + \varepsilon} \pi(\theta | \mathbf{x}) d\theta$, ce qui, lorsque ε tend vers 0, est équivalent à maximiser $\pi(\theta | \mathbf{x})$ par rapport à θ . L'estimateur du MAP correspond alors à la limite des estimateurs de Bayes pour la fonction de coût C_{ε} lorsque ε tend vers 0.

Ajoutons enfin que dans le cas où Θ est un ensemble discret (dénombrable), l'estimateur du MAP correspond exactement (i.e., ce n'est pas un cas limite) à l'estimateur de Bayes pour la fonction de coût 0-1, $C(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) = \mathbb{1}_{\{\boldsymbol{\theta}\}}(\hat{\boldsymbol{\theta}}(\mathbf{x}))$.

2.1.3.3 Aperçu de quelques propriétés asymptotiques des estimateurs bayésiens

Dans la problématique bayésienne, on s'intéresse à l'estimation d'un vecteur $\boldsymbol{\theta}$ conditionnée par le signal observé \mathbf{x} . Il est donc assez rare d'analyser les propriétés asymptotiques (i.e., lorsque le nombre d'observations N tend vers l'infini) des estimateurs bayésiens. Néanmoins, de telles études existent, et démontrent en général de bonnes propriétés asymptotiques pour ces estimateurs [IH81, Van98, LC03, Rob07]. Intuitivement, ces bonnes performances asymptotiques s'expliquent par le fait que lorsque le nombre d'observations augmente (et que le nombre de paramètres reste fixe), l'influence de la vraisemblance devient prépondérante devant celle de la loi *a priori* (qui ne dépend pas de N). Ainsi, la distribution *a posteriori*, sur laquelle est basée l'inférence bayésienne, aura tendance à se comporter comme la fonction de vraisemblance, et donc l'estimation à se rapprocher de l'estimateur ML.

En particulier, il est démontré (voir [LC03, Théorème 8.3, p.490]) que sous certaines conditions de régularité, l'estimateur MMSE vérifie $\sqrt{N}(\hat{\boldsymbol{\theta}}_{(N)}^{\text{MMSE}} - \boldsymbol{\theta}) \mid \boldsymbol{\theta} \xrightarrow{L} \mathcal{N}(\mathbf{0}, \mathbf{F}^{-1}(\boldsymbol{\theta}))$. Insistons sur le fait qu'il s'agit là d'une convergence *conditionnellement à la valeur de $\boldsymbol{\theta}$* , et donc d'un résultat *local*.

Des résultats d'efficacité pour une large classe d'estimateurs de Bayes ont également été prouvés (voir [IH81]), notamment ceux associés aux fonctions de coût de la forme $\mathcal{C}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = |\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}|^\alpha$ ($\alpha \geq 1$), qui englobe les erreurs absolue et quadratique.

Mentionnons enfin le théorème de Bernstein-von Mises [Van98, p.141], qui énonce des conditions suffisantes pour la convergence, au sens de la norme de variation totale, de la distribution *a posteriori* $\pi(\boldsymbol{\theta} \mid \mathbf{x})$ vers une loi normale centrée dont la matrice de covariance est l'inverse de la matrice d'information de Fisher.

Une remarque importante pour la suite de ce manuscrit est que, parmi les conditions nécessaires à l'application de tous ces théorèmes de convergence, figure toujours la différentiabilité de la fonction de vraisemblance $p(\mathbf{x} \mid \boldsymbol{\theta})$ par rapport à $\boldsymbol{\theta}$. Par conséquent, si celle-ci n'est pas vérifiée, les théorèmes en question sont mis en défaut.

2.1.4 Estimation hybride

Dans certains problèmes d'estimation, on peut disposer d'information *a priori* sur certains des paramètres composant le vecteur $\boldsymbol{\theta}$ inconnu, mais pas sur tous [Roc86, RS87, NK95, RM97, Yer00, TW07, VB07, VVRB12, RGC⁺15]. On scinde alors le vecteur $\boldsymbol{\theta}$ en deux sous-vecteurs $\boldsymbol{\theta}_d$ et $\boldsymbol{\theta}_a$, où $\boldsymbol{\theta}_d$ est supposé déterministe et $\boldsymbol{\theta}_a$ est supposé aléatoire et regroupe les paramètres pour lesquels on dispose d'information *a priori*. En traitement du signal, l'estimation est alors qualifiée de « hybride », puisqu'elle fait intervenir les deux points de vue évoqués précédemment (fréquentiste et bayésien). Même si cette approche est moins omniprésente dans la littérature que les deux précédentes, l'étude de ses performances d'estimation présente un intérêt certain, notamment dans des problèmes d'estimation où interviennent des paramètres de nuisance.

Cette problématique a été introduite dans [Roc86] et [RS87], pour résoudre un problème de calibration d'un réseau d'antennes à partir de sources de positions inconnues, en champ lointain. La position des capteurs n'étant pas connue précisément, un *a priori* gaussien leur est attribué, tandis qu'aucune information n'est disponible *a priori* pour la position des sources. L'étude se base donc sur le calcul d'une borne de Cramér-Rao adaptée à ce contexte, c'est-à-dire pour l'estimation conjointe de paramètres déterministes (dans ce cas, les positions des sources) et de variables aléatoires (les positions des capteurs de l'antenne).

Il existe assez peu de résultats théoriques concernant les estimateurs hybrides. Notons qu'ils ne consistent pas naïvement en la juxtaposition d'un estimateur de paramètres déterministes et d'un estimateur bayésien, puisqu'il s'agit d'un problème d'estimation conjointe. Pour autant, les estimateurs mis en place se rapprochent en général du maximum de vraisemblance. En particulier,

l'estimateur conjoint du maximum de vraisemblance/*maximum a posteriori* (noté MLMAP dans la suite) a pu être utilisé dans divers contextes [Bar72, LO78, Mus82]. Une étude approfondie en est dressée dans [Yer00], établissant en particulier sa non consistance, mais démontrant, pour certains cas, sa supériorité par rapport à l'estimateur ML. Ses liens avec la méthode des moindres carrés étendus et l'algorithme espérance-maximisation (EM) y sont également étudiés. S'inspirant de la formulation des deux estimateurs ML et MAP sur lesquels il est basé, l'estimateur MLMAP se fonde sur la distribution $p(\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d)$. Plus précisément, il s'exprime donc comme

$$\hat{\boldsymbol{\theta}}^{\text{MLMAP}} \triangleq [\hat{\boldsymbol{\theta}}_d^{\text{MLMAP}}; \hat{\boldsymbol{\theta}}_a^{\text{MLMAP}}] \triangleq \arg \max_{\boldsymbol{\theta}_d, \boldsymbol{\theta}_a} p(\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d) \quad (2.39)$$

où la notation $[\mathbf{y}; \mathbf{z}]$ traduit la concaténation verticale des deux vecteurs \mathbf{y} et \mathbf{z} , i.e., $[\mathbf{y}; \mathbf{z}] \triangleq [\mathbf{y}^\top, \mathbf{z}^\top]^\top$.

Pour qualifier les performances d'estimation dans un contexte hybride, il est apparu plus commode de s'intéresser (encore une fois) à des bornes inférieures de l'erreur quadratique moyenne. La plus évidente d'entre elles est bien sûr la borne de Cramér-Rao hybride (BCRH) mentionnée plus haut [RS87, NK95, TW07, BGR⁺08, VVRB12]. Il en existe d'autres comme la borne de Barankin hybride [RM95, RM97], la borne de Barankin/Weiss-Weinstein [RGC⁺15], ou encore la borne de Barankin/Ziv-Zakai hybride, qui s'avèrent particulièrement utiles dans des conditions non asymptotiques (faible rapport signal à bruit, faible nombre d'observations). Même s'il n'existe pas de preuve formelle que ces bornes sont atteintes en règle générale, elles permettent néanmoins d'obtenir un aperçu fidèle des limites de performance d'estimation pour un problème donné. Il a tout de même été démontré que la BCRH est atteinte par l'estimateur MLMAP dans le cas d'un modèle linéaire en les paramètres et gaussien, c'est-à-dire du type $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{b}$, où \mathbf{H} est une matrice connue (indépendante de $\boldsymbol{\theta}$) et \mathbf{b} est un bruit additif gaussien [Ren15].

2.1.5 Bornes inférieures de l'EQM

Comme nous venons de le voir, l'évaluation des performances d'estimation pour un problème donné est indissociable en pratique de la détermination de bornes inférieures de l'erreur quadratique moyenne, puisque celles-ci fournissent une indication précise des limites de performance d'estimation. La détermination de ces bornes peut être vue comme une étape préliminaire à la mise en place d'une procédure d'estimation : avant de savoir précisément quelle sera cette procédure, il faut savoir quelle précision dans l'estimation on peut envisager d'atteindre.

On se concentre ici⁶ sur une famille de bornes inférieures de l'erreur quadratique moyenne : celles issues de l'inégalité de covariance. Cette dernière est une généralisation en plusieurs dimensions de l'inégalité de Cauchy-Schwarz ; elle est rappelée dans la sous-section qui suit. Elle permet d'obtenir des bornes inférieures de l'EQM dans les trois contextes d'estimation évoqués précédemment (fréquentiste, bayésienne, hybride), que nous détaillerons séparément ensuite (voir Sections 2.1.5.2 à 2.1.5.4.2).

Mentionnons enfin que l'inégalité de covariance n'est pas le seul principe qui permet d'accéder à de telles bornes inférieures. Que ce soit dans le cadre de l'estimation de paramètres déterministes ou aléatoires, il a été montré qu'elles peuvent être définies comme la solution d'un problème de minimisation de l'EQM (ou plus généralement, d'une matrice de Gram) sous contraintes

6. Comme évoqué en introduction, il existe parmi les bornes bayésiennes une autre famille de bornes inférieures de l'EQM que nous n'évoquons que brièvement maintenant. Il s'agit des bornes de la famille dite « de Ziv-Zakai », en référence à la borne originale du même nom. Elles reposent non pas sur l'inégalité de Cauchy-Schwarz, mais sur l'inégalité de Kotelnikov, et elles relient l'EQM à la probabilité d'erreur dans un test d'hypothèses (binaire, ou plus généralement M -aire). Leur unification a été établie par Kristine Bell dans ses travaux de thèse, débouchant ensuite sur des versions étendues de ces bornes [Bel95, BSEV97].

linéaires [Ren06, RFL⁺08, Cha04, TT10a, TT10b]. Cette approche a l'avantage de mettre plus clairement en évidence les conditions de régularité nécessaires à la validité de la borne, puisque celles-ci correspondent aux contraintes du problème de minimisation. Il s'agit en réalité d'une forme duale de l'inégalité de Cauchy-Schwarz.

Nous nous en tenons ici toutefois à l'établissement de bornes à l'aide de l'inégalité de covariance, pour la raison que c'est la méthode généralement retenue dans la littérature pour présenter les bornes hybrides.

2.1.5.1 L'inégalité de covariance

L'inégalité de covariance est énoncée dans le théorème suivant.

Théorème 4 (Inégalité de covariance). *Soit un problème d'estimation quelconque, où le vecteur de paramètres inconnu, noté $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^R$, peut être supposé déterministe, aléatoire ou bien hybride. Soit $\widehat{\boldsymbol{\theta}}(\mathbf{x})$ un estimateur de $\boldsymbol{\theta}$, i.e., une fonction mesurable de Ω à valeurs dans Θ . Soit $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ une fonction mesurable à valeurs dans \mathbb{R}^K ($K \geq R$) telle que*

- i) la matrice de covariance $\mathbb{E}\{\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})\mathbf{v}^\top(\mathbf{x}, \boldsymbol{\theta})\}$ soit définie positive ;
- ii) tous les éléments des matrices $\mathbb{E}\{\widehat{\boldsymbol{\theta}}(\mathbf{x})\mathbf{v}^\top(\mathbf{x}, \boldsymbol{\theta})\}$ et $\mathbb{E}\{\boldsymbol{\theta}\mathbf{v}^\top(\mathbf{x}, \boldsymbol{\theta})\}$ sont finis.

Alors nous avons l'inégalité matricielle suivante [LC03] :

$$\mathbb{E}\left\{(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})^\top\right\} \succcurlyeq \mathbf{C}\mathbf{V}^{-1}\mathbf{C}^\top \quad (2.40)$$

où $\mathbf{C} \triangleq \mathbb{E}\{(\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})\mathbf{v}^\top(\mathbf{x}, \boldsymbol{\theta})\}$, $\mathbf{V} \triangleq \mathbb{E}\{\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})\mathbf{v}^\top(\mathbf{x}, \boldsymbol{\theta})\}$ et on rappelle que le signe d'inégalité « \succcurlyeq » signifie que la différence entre le premier et le second membre est une matrice semi-définie positive.

Remarque : Avant de prouver ce résultat, on peut remarquer que le membre de droite dans (2.40), tel qu'il est écrit pour le moment, dépend de $\widehat{\boldsymbol{\theta}}(\mathbf{x})$, en général. Or, nous avons vu que l'un des intérêts pratiques des bornes inférieures de l'EQM est leur indépendance vis-à-vis de la procédure d'estimation. Comme on le verra, c'est grâce à un choix judicieux de la fonction $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ que l'on pourra obtenir une matrice \mathbf{C} qui ne dépend pas de $\widehat{\boldsymbol{\theta}}$. \square

Démonstration. Introduisons ici la mesure de probabilité $\nu(\mathbf{x}, \boldsymbol{\theta})$ appropriée au contexte d'estimation considéré⁷. Soit l'ensemble $\mathcal{L}^2(\nu)$ des fonctions de carré mesurable par rapport à la mesure ν , c'est-à-dire $\mathcal{L}^2(\nu) \triangleq \left\{f \text{ à valeurs réelles et mesurable } \left| \int \sqrt{|f|^2} d\nu < \infty \right.\right\}$. Soit le produit scalaire sur $\mathcal{L}^2(\nu)$ défini comme suit : pour toutes fonctions $g(\mathbf{x}, \boldsymbol{\theta})$ et $h(\mathbf{x}, \boldsymbol{\theta})$ à valeurs réelles, mesurables par rapport à ν et de carré intégrable,

$$\langle g | h \rangle \triangleq \int g(\mathbf{x}, \boldsymbol{\theta}) h(\mathbf{x}, \boldsymbol{\theta}) d\nu(\mathbf{x}, \boldsymbol{\theta}) = \mathbb{E}\{g(\mathbf{x}, \boldsymbol{\theta}) h(\mathbf{x}, \boldsymbol{\theta})\}. \quad (2.41)$$

La norme euclidienne associée à ce produit scalaire sera notée $\|\cdot\|$, i.e., $\|g\| = \sqrt{\langle g | g \rangle}$.

Soient $\mathbf{a} \in \mathbb{R}^R$ et $\mathbf{b} \in \mathbb{R}^K$ deux vecteurs non nuls. Considérons deux fonctions $e(\mathbf{x}, \boldsymbol{\theta})$ et $\psi(\mathbf{x}, \boldsymbol{\theta})$ à valeurs réelles, intégrables et appartenant à $\mathcal{L}^2(\nu)$, définies comme

$$e(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{a}^\top (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \quad \text{et} \quad \psi(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{b}^\top \mathbf{v}(\mathbf{x}, \boldsymbol{\theta}). \quad (2.42)$$

7. Dans le cadre fréquentiste, ce sera la mesure de probabilité induite par la distribution $p(\mathbf{x}; \boldsymbol{\theta})$, dans le cadre bayésien celle induite par $p(\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_a)$.

D'après l'inégalité de Cauchy-Schwarz, on a

$$\|e(\mathbf{x}, \boldsymbol{\theta})\|^2 \geq \frac{|\langle e(\mathbf{x}, \boldsymbol{\theta}) | \psi(\mathbf{x}, \boldsymbol{\theta}) \rangle|^2}{\|\psi(\mathbf{x}, \boldsymbol{\theta})\|^2}, \quad (2.43)$$

c'est-à-dire

$$\|\mathbf{a}^\top (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})\|^2 \geq \frac{|\langle \mathbf{a}^\top (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) | \mathbf{b}^\top \mathbf{v}(\mathbf{x}, \boldsymbol{\theta}) \rangle|^2}{\|\mathbf{b}^\top \mathbf{v}(\mathbf{x}, \boldsymbol{\theta})\|^2}, \quad (2.44)$$

qui peut se réécrire, par bilinéarité du produit scalaire, comme

$$\mathbf{a}^\top \langle (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) | (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \rangle \mathbf{a} \geq \frac{|\mathbf{a}^\top \langle (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) | \mathbf{v}^\top(\mathbf{x}, \boldsymbol{\theta}) \rangle \mathbf{b}|^2}{\mathbf{b}^\top \langle \mathbf{v}(\mathbf{x}, \boldsymbol{\theta}) | \mathbf{v}^\top(\mathbf{x}, \boldsymbol{\theta}) \rangle \mathbf{b}}. \quad (2.45)$$

En définissant les deux matrices $\mathbf{C} \triangleq \langle (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) | \mathbf{v}^\top(\mathbf{x}, \boldsymbol{\theta}) \rangle$ et $\mathbf{V} \triangleq \langle \mathbf{v}(\mathbf{x}, \boldsymbol{\theta}) | \mathbf{v}^\top(\mathbf{x}, \boldsymbol{\theta}) \rangle$, on obtient

$$\mathbf{a}^\top \langle (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) | (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \rangle \mathbf{a} \geq \frac{|\mathbf{a}^\top \mathbf{C} \mathbf{b}|^2}{\mathbf{b}^\top \mathbf{V} \mathbf{b}}. \quad (2.46)$$

Comme la matrice \mathbf{V} est symétrique définie positive, on peut définir le produit scalaire associé à \mathbf{V} comme $\langle \mathbf{y} | \mathbf{z} \rangle_{\mathbf{V}} \triangleq \mathbf{y}^\top \mathbf{V} \mathbf{z}$, $\forall \mathbf{y}, \mathbf{z} \in \mathbb{R}^K$. Ainsi, on a $\mathbf{a}^\top \mathbf{C} \mathbf{b} = \langle \mathbf{V}^{-1} \mathbf{C}^\top \mathbf{a} | \mathbf{b} \rangle_{\mathbf{V}}$. De plus, comme l'inégalité (2.46) est valable quel que soit $\mathbf{b} \in \mathbb{R}^K$, on a également

$$\mathbf{a}^\top \langle (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) | (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \rangle \mathbf{a} \geq \sup_{\mathbf{b} \in \mathbb{R}^M} \frac{|\langle \mathbf{V}^{-1} \mathbf{C}^\top \mathbf{a} | \mathbf{b} \rangle_{\mathbf{V}}|^2}{\langle \mathbf{b} | \mathbf{b} \rangle_{\mathbf{V}}}. \quad (2.47)$$

Or, en utilisant une fois encore l'inégalité de Cauchy-Schwarz, on a

$$|\langle \mathbf{V}^{-1} \mathbf{C}^\top \mathbf{a} | \mathbf{b} \rangle_{\mathbf{V}}|^2 \leq \|\mathbf{V}^{-1} \mathbf{C}^\top \mathbf{a}\|_{\mathbf{V}}^2 \|\mathbf{b}\|_{\mathbf{V}}^2 \quad (2.48)$$

avec égalité si et seulement si $\exists k \in \mathbb{R}$, $\mathbf{b} = k \mathbf{V}^{-1} \mathbf{C}^\top \mathbf{a}$. Ainsi, pour cette valeur de \mathbf{b} , (2.47) devient

$$\begin{aligned} \mathbf{a}^\top \langle (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) | (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \rangle \mathbf{a} &\geq \|\mathbf{V}^{-1} \mathbf{C}^\top \mathbf{a}\|_{\mathbf{V}}^2 \\ &\Downarrow \\ \mathbf{a}^\top \langle (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) | (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \rangle \mathbf{a} &\geq \mathbf{a}^\top \mathbf{C} \mathbf{V}^{-1} \mathbf{V} \mathbf{V}^{-1} \mathbf{C}^\top \mathbf{a} \\ &\Downarrow \\ \mathbf{a}^\top \langle (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) | (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \rangle \mathbf{a} &\geq \mathbf{a}^\top \mathbf{C} \mathbf{V}^{-1} \mathbf{C}^\top \mathbf{a}. \end{aligned} \quad (2.49)$$

Et comme la dernière inégalité est vraie pour tout vecteur $\mathbf{a} \in \mathbb{R}^R$ non nul, on obtient finalement l'inégalité de covariance énoncée en (2.40). \square

À partir du théorème 4, nous sommes en mesure d'obtenir de nombreuses bornes inférieures de l'EQM. Comme nous l'avons vu, cela passe par la définition de fonctions $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ adéquates.

2.1.5.2 Bornes pour l'estimation de paramètres déterministes

Dans cette section, on se place dans le cadre de l'estimation fréquentiste, c'est-à-dire que $\boldsymbol{\theta}$ est un vecteur de paramètres déterministes, défini sur un ensemble Θ . Par conséquent, on considère que la mesure ν introduite dans la section 2.1.5.1 est la mesure induite par la distribution $p(\mathbf{x}; \boldsymbol{\theta})$: si l'on note \mathcal{T} la tribu associée à l'espace des observations Ω , on a, $\forall A \in \mathcal{T}$, $\nu(A) = \int_A p(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x}$.

2.1.5.2.1 Borne de Cramér-Rao [Fis22, Fre43, Dar45, Rao45, Cra46, Van68]

On peut retrouver la borne de Cramér-Rao classique en définissant les composantes de la fonction⁸ $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$, pour $1 \leq r \leq R$, comme

$$[\mathbf{v}(\mathbf{x}, \boldsymbol{\theta}^*)]_r = \left. \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_r} \right|_{\boldsymbol{\theta}^*}. \quad (2.50)$$

Mais sans hypothèses supplémentaires, la matrice \mathbf{C} dans (2.40) dépend toujours de $\widehat{\boldsymbol{\theta}}(\mathbf{x})$, ce qui, comme on l'a vu, ne doit pas être le cas pour une borne inférieure de l'EQM. Nous imposons donc la condition que l'estimateur $\widehat{\boldsymbol{\theta}}$ soit localement non biaisé⁹ :

$$\mathbb{E}_{\mathbf{x}; \boldsymbol{\theta}^*} \{ \widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}^* \} = \mathbf{0}. \quad (2.51)$$

$$\left. \frac{\partial \mathbb{E}_{\mathbf{x}; \boldsymbol{\theta}^*} \{ \widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}^* \}}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}^*} = \mathbf{0} \quad (2.52)$$

En général, cette condition est satisfaite au moins en régime asymptotique, c'est-à-dire pour un grand nombre d'échantillons ou à fort RSB.

On peut alors en déduire l'expression des colonnes de la matrice \mathbf{C} , notées \mathbf{c}_r , pour $1 \leq r \leq R$:

$$\begin{aligned} \mathbf{c}_r &= \mathbb{E}_{\mathbf{x}; \boldsymbol{\theta}} \left\{ (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_r} \right\} \\ &= \int_{\Omega} \widehat{\boldsymbol{\theta}}(\mathbf{x}) \frac{\partial p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_r} d\mathbf{x} - \boldsymbol{\theta} \mathbb{E}_{\mathbf{x}; \boldsymbol{\theta}} \left\{ \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_r} \right\} \\ &= \frac{\partial}{\partial \theta_r} \int_{\Omega} \widehat{\boldsymbol{\theta}}(\mathbf{x}) p(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} - \boldsymbol{\theta} \mathbb{E}_{\mathbf{x}; \boldsymbol{\theta}} \left\{ \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_r} \right\} \\ &= \frac{\partial \boldsymbol{\theta}}{\partial \theta_r} - \boldsymbol{\theta} \mathbb{E}_{\mathbf{x}; \boldsymbol{\theta}} \left\{ \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_r} \right\}. \end{aligned}$$

Si nous imposons la condition supplémentaire

$$\forall \boldsymbol{\theta} \in \Theta, \mathbb{E}_{\mathbf{x}; \boldsymbol{\theta}} \left\{ \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\} = \mathbf{0} \quad (2.53)$$

qui n'est autre que la condition de régularité que nous avons énoncée dans le théorème 1, voir (2.9), alors nous obtenons simplement

$$\mathbf{c}_r = \mathbf{e}_r \quad (2.54)$$

où \mathbf{e}_r est le vecteur de \mathbb{R}^R dont toutes les composantes sont nulles, sauf la r -ième, qui vaut 1. Cela signifie que la matrice \mathbf{C} est en fait la matrice identité de taille $R \times R$, i.e., $\mathbf{C} = \mathbf{I}$.

Enfin, nous obtenons que la matrice \mathbf{V} dans (2.40) n'est autre que la matrice d'information de Fisher :

$$\begin{aligned} \mathbf{V} &= \mathbb{E}_{\mathbf{x}; \boldsymbol{\theta}^*} \{ \mathbf{v}(\mathbf{x}, \boldsymbol{\theta}^*) \mathbf{v}^\top(\mathbf{x}, \boldsymbol{\theta}^*) \} \\ &= \mathbb{E}_{\mathbf{x}; \boldsymbol{\theta}^*} \left\{ \left. \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}^*} \left. \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right|_{\boldsymbol{\theta}^*} \right\} = \mathbf{F}(\boldsymbol{\theta}^*). \end{aligned} \quad (2.55)$$

Nous retrouvons donc exactement la borne de Cramér-Rao donnée par (2.13), présentée dans le théorème 1.

8. Ici la dimension K de la fonction $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ du théorème 4 est réduite à R .

9. Notons que la borne de Cramér-Rao existe aussi pour des estimateurs biaisés [Cha04], cas que nous n'envisageons pas ici.

2.1.5.2.2 Bornes de Barankin [Bar49, Ham50, CR51, MS69, MH71]

Nous avons vu que la borne de Cramér-Rao est atteinte en particulier par l'estimateur du maximum de vraisemblance, dans des conditions asymptotiques, c'est-à-dire à fort rapport signal-à-bruit (RSB) et/ou grand nombre d'observations. Dès lors, la BCR fournit un aperçu très fidèle du comportement de l'estimateur ML. En revanche, on peut remarquer en simulations que lorsque le RSB/le nombre d'observations diminuent, les performances d'estimation peuvent se dégrader brusquement et de façon très marquée, comme c'est le cas en analyse spectrale (voir par exemple [RB74]). Ce phénomène, appelé « décrochement », ne peut être mis en évidence par le seul tracé de la BCR. C'est une des limitations de la BCR : à faible RSB/nombre d'observations, elle est imprécise, dans le sens où les valeurs de l'EQM minimale qu'elle fournit sont irréalistes car trop optimistes. Dans ces conditions, la BCR est impossible à atteindre, elle ne reflète donc pas le comportement de l'estimation de manière assez réaliste. Elle peut également être invalide en toute rigueur, car dans ces conditions, les conditions de biais nul local ne sont pas satisfaites.

Or, nous avons vu que la BCR est valide pour une famille d'estimateurs non biaisés localement, c'est-à-dire pour une valeur bien précise du vecteur $\boldsymbol{\theta}$, à savoir sa vraie valeur $\boldsymbol{\theta}^*$. Pour obtenir une borne plus précise, une idée est d'ajouter des contraintes sur la famille d'estimateurs à laquelle la borne est applicable. Barankin a proposé de s'intéresser à des estimateurs non biaisés, non pas pour une seule valeur $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, mais sur tout un ensemble Π comprenant $\boldsymbol{\theta}^*$, c'est-à-dire les estimateurs $\hat{\boldsymbol{\theta}}$ tels que¹⁰

$$\forall \boldsymbol{\theta} \in \Pi, \mathbb{E}_{\mathbf{x};\boldsymbol{\theta}}\{\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\} = \mathbf{0}. \quad (2.56)$$

On notera bien la différence de cette condition avec (2.51) qui n'impose le biais nul qu'en la seule valeur $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Il s'avère toutefois que la borne de Barankin proprement dite, pour les estimateurs vérifiant (2.56), s'exprime comme la solution d'une équation intégrale qui n'admet que très rarement une forme analytique (voir par exemple [Mar97]).

Une manière d'approcher la borne de Barankin, et ainsi de rendre son calcul tractable est de discrétiser l'ensemble Π dans (2.56), c'est-à-dire que (2.56) devient une condition de biais nul en un ensemble discret de points $\{\boldsymbol{\theta}_k\}_{k=1,\dots,K}$ appelés¹¹ « points test », pouvant être choisis arbitrairement sur l'espace des paramètres Θ , en plus du biais nul en $\boldsymbol{\theta}^*$:

$$\begin{cases} \mathbb{E}_{\mathbf{x};\boldsymbol{\theta}^*}\{\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}^*\} = \mathbf{0}, & (2.57a) \\ \mathbb{E}_{\mathbf{x};\tilde{\boldsymbol{\theta}}_k}\{\hat{\boldsymbol{\theta}}(\mathbf{x}) - \tilde{\boldsymbol{\theta}}_k\} = \mathbf{0}, \quad \forall k = 1, \dots, K. & (2.57b) \end{cases}$$

Afin d'obtenir la borne associée à ces contraintes, on définit les composantes de la fonction $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$, pour $1 \leq k \leq K$, comme

$$[\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})]_k = \frac{p(\mathbf{x}; \tilde{\boldsymbol{\theta}}_k)}{p(\mathbf{x}; \boldsymbol{\theta})} - 1. \quad (2.58)$$

Ainsi, les colonnes de la matrice \mathbf{C} dans (2.40) sont données, pour $1 \leq k \leq K$, par

$$\begin{aligned} \mathbf{c}_k &= \mathbb{E}_{\mathbf{x};\boldsymbol{\theta}}\left\{(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})\left(\frac{p(\mathbf{x}; \tilde{\boldsymbol{\theta}}_k)}{p(\mathbf{x}; \boldsymbol{\theta})} - 1\right)\right\} \\ &= \mathbb{E}_{\mathbf{x};\boldsymbol{\theta}}\left\{\frac{(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})p(\mathbf{x}; \tilde{\boldsymbol{\theta}}_k)}{p(\mathbf{x}; \boldsymbol{\theta})}\right\} - \mathbb{E}_{\mathbf{x};\boldsymbol{\theta}}\{(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})\} \\ &= \mathbb{E}_{\mathbf{x};\tilde{\boldsymbol{\theta}}_k}\{(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})\} - \mathbb{E}_{\mathbf{x};\boldsymbol{\theta}}\{(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})\} \end{aligned} \quad (2.59)$$

10. Cette condition est à comparer avec la condition de biais localement nul de la BCR, donnée par (2.51) et (2.52). Il est clair que la condition énoncée ici est bien plus contraignante.

11. De l'anglais « *test points* ».

En évaluant (2.59) en θ^* et en utilisant les conditions (2.57a) et (2.57b), on obtient

$$\mathbf{c}_k = \tilde{\theta}_k - \theta^*. \quad (2.60)$$

Dans la suite on écrira $\tilde{\theta}_k = \theta^* + \mathbf{h}_k$, $k = 1, \dots, K$, où $\mathbf{h}_k \in \mathcal{H} \triangleq \{\mathbf{h} \in \mathbb{R}^R \mid p(\mathbf{x}; \theta^* + \mathbf{h}) > 0\}$, qui définit l'ensemble des « valeurs tests ¹² ».

D'autre part, les éléments de la matrice \mathbf{V} dans (2.40) sont donnés, pour $1 \leq i, j \leq K$, par

$$\begin{aligned} [\mathbf{V}]_{i,j} &= \mathbb{E}_{\mathbf{x}; \theta^*} \left\{ \left(\frac{p(\mathbf{x}; \tilde{\theta}_i)}{p(\mathbf{x}; \theta^*)} - 1 \right) \left(\frac{p(\mathbf{x}; \tilde{\theta}_j)}{p(\mathbf{x}; \theta^*)} - 1 \right) \right\} \\ &= \mathbb{E}_{\mathbf{x}; \theta^*} \left\{ \frac{p(\mathbf{x}; \theta^* + \mathbf{h}_i) p(\mathbf{x}; \theta^* + \mathbf{h}_j)}{p^2(\mathbf{x}; \theta^*)} \right\} - 1. \end{aligned}$$

La « borne de Barankin avec K points test » est donc communément exprimée sous la forme

$$\mathbf{B}\mathbf{B}(\mathbf{H}) = \mathbf{H}(\mathbf{J} - \mathbf{1}\mathbf{1}^\top)^{-1} \mathbf{H}^\top \quad (2.61)$$

où \mathbf{H} est la matrice des points test : $\mathbf{H} \triangleq [\mathbf{h}_1, \dots, \mathbf{h}_K]$; $\mathbf{1}$ représente le vecteur (ici de taille K) dont toutes les composantes valent 1; et \mathbf{J} la matrice de taille $K \times K$ dont les éléments valent

$$[\mathbf{J}]_{i,j} = \mathbb{E}_{\mathbf{x}; \theta^*} \left\{ \frac{p(\mathbf{x}; \theta^* + \mathbf{h}_i) p(\mathbf{x}; \theta^* + \mathbf{h}_j)}{p^2(\mathbf{x}; \theta^*)} \right\}. \quad (2.62)$$

Notons que la matrice $\mathbf{B}\mathbf{I}\mathbf{M} \triangleq \mathbf{J} - \mathbf{1}\mathbf{1}^\top$ est parfois appelée « matrice d'information de Barankin » [LRNM10] car elle joue le même rôle que la matrice d'information de Fisher. Il est également important de remarquer que la borne de Cramér-Rao (2.13) peut être vue comme un cas limite de la borne dans (2.61) en prenant $\mathbf{H} = h\mathbf{I}$, où \mathbf{I} est la matrice identité, et en faisant tendre h vers zéro (voir par exemple [MH71]) ¹³. Mais l'un des grands avantages de la borne de Barankin sur la borne de Cramér-Rao réside dans ses conditions de régularité particulièrement faibles : en particulier, elle ne nécessite pas la différentiabilité de la log-vraisemblance par rapport au vecteur de paramètres θ . Cela permet notamment de la calculer dans le cas d'un problème d'estimation de paramètres discrets [CS12].

Ainsi, pour chaque ensemble de points test $\mathbf{H} \in \mathcal{H}^K$ choisi, on obtient une borne inférieure de l'EQM donnée par (2.61). Si \mathbf{H} est mal choisi, on peut obtenir une borne moins précise que la BCR. Cependant, l'inégalité (2.40) reste valable si l'on maximise (2.61) par rapport à \mathbf{H} . La meilleure borne de Barankin avec K points test est donc donnée par

$$\mathbf{B}\mathbf{B} = \sup_{\mathbf{H} \in \mathcal{H}^K} \mathbf{H}(\mathbf{J}(\mathbf{H}) - \mathbf{1}\mathbf{1}^\top)^{-1} \mathbf{H}^\top \quad (2.63)$$

où la dépendance de \mathbf{J} en \mathbf{H} est mise en avant.

Cette borne a en réalité été introduite par Hammersley [Ham50], Chapman et Robbins [CR51] dans le cas de l'estimation d'un paramètre scalaire ($R = 1$) et d'un seul point test ($K = 1$), auquel cas son expression se réduit à

$$\text{BHamChR} \triangleq \sup_{h \in \mathcal{H}} \frac{h^2}{\mathbb{E}_{\mathbf{x}; \theta^*} \left\{ \left(\frac{p(\mathbf{x}; \theta^* + h)}{p(\mathbf{x}; \theta^*)} \right)^2 \right\} - 1}. \quad (2.64)$$

La borne (2.63) – cas d'un vecteur de paramètres ($R > 1$) et/ou plus d'un point test par paramètre ($K > 1$) – a pour sa part été introduite par McAulay, Seidman et Hofstetter [MS69, MH71]. Dans la littérature, elles sont toutefois appelées simplement « bornes de Barankin », la plupart du temps.

¹². Dans la littérature, on désigne souvent, par abus de langage, les \mathbf{h}_k comme les « points test », même si en toute rigueur, ce sont les $\tilde{\theta}_k$. Dans la suite, cet abus de langage sera largement commis.

¹³. La BCR est même en fait un cas particulier de la borne originelle de Barankin [Bar49].

2.1.5.3 Bornes bayésiennes

On se place dans le cadre bayésien, c'est-à-dire que le vecteur de paramètres inconnus $\boldsymbol{\theta} \in \Theta$ est supposé aléatoire avec une distribution *a priori* $\pi(\boldsymbol{\theta})$. Ici, on considère que la mesure ν introduite à la section 2.1.5.1 correspond à la mesure associée à la distribution jointe $p(\mathbf{x}, \boldsymbol{\theta})$: en notant \mathcal{T}' la tribu associée à l'espace $\Omega \times \Theta$, on a, $\forall(A \times B) \in \mathcal{T}'$, $\nu(A \times B) = \int_{A \times B} p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} d\boldsymbol{\theta}$. Ainsi, dans ce contexte, les bornes bayésiennes portent sur l'EQM *globale* (EQMG), qui correspond au risque intégré pour le coût quadratique (voir section 2.1.3.1), c'est-à-dire

$$\begin{aligned} \mathbf{EQMG}(\widehat{\boldsymbol{\theta}}) &\triangleq \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})^\top \right\} = \int_{\Omega} \int_{\Theta} (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta})^\top p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} d\boldsymbol{\theta} \\ &= \mathbb{E}_{\boldsymbol{\theta}} \{ \mathbf{EQM}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \} = \int_{\Theta} \mathbf{EQM}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned} \quad (2.65)$$

Comme expliqué en préambule de la section 2.1.5, les bornes de Ziv-Zakai sont des bornes bayésiennes mais ne seront pas exposées ici. Nous ne présentons ici que la borne de Cramér-Rao bayésienne et la borne de Weiss-Weinstein¹⁴.

2.1.5.3.1 Borne de Cramér-Rao bayésienne [Van68, VB07]

La borne de Cramér-Rao bayésienne a été introduite par Van Trees [Van68]. Elle est obtenue en définissant la fonction $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ du théorème 4 (en fixant $K = R$) comme

$$\mathbf{v}(\mathbf{x}, \boldsymbol{\theta}) = \begin{cases} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} & , \text{ si } \boldsymbol{\theta} \in \Theta', \\ \mathbf{0} & , \text{ sinon.} \end{cases} \quad (2.66)$$

où $\Theta' \triangleq \{\boldsymbol{\theta} \in \Theta \mid p(\mathbf{x}, \boldsymbol{\theta}) > 0\}$. Ici encore, sans imposer de conditions supplémentaires, le membre de droite dans l'inégalité (2.40) dépend de $\boldsymbol{\theta}$. Ici, on n'utilise pas la condition de biais nul comme en (2.51), mais les deux conditions suivantes :

1. la fonction $\boldsymbol{\theta} \mapsto p(\mathbf{x}, \boldsymbol{\theta})$ est absolument continue, pour tout $\mathbf{x} \in \Omega$ p.p.¹⁵ ;
2. $\lim_{\theta_r \rightarrow \pm\infty} \theta_r p(\mathbf{x}, \boldsymbol{\theta}) = 0$, pour $r = 1, \dots, R$, et pour tout $\mathbf{x} \in \Omega$ p.p.

Ainsi, on obtient pour les colonnes \mathbf{c}_r , $1 \leq r \leq R$, de la matrice \mathbf{C} :

$$\mathbf{c}_r = \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_r} \right\} = \int_{\Omega} \int_{\Theta} (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \frac{\partial p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_r} d\mathbf{x} d\boldsymbol{\theta}. \quad (2.67)$$

En intégrant par parties, on obtient

$$\begin{aligned} \int_{\Theta} (\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \frac{\partial p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_r} d\boldsymbol{\theta} &= \widehat{\boldsymbol{\theta}}(\mathbf{x}) [p(\mathbf{x}, \boldsymbol{\theta})]_{-\infty}^{+\infty} - [p(\mathbf{x}, \boldsymbol{\theta})]_{-\infty}^{+\infty} - \int_{\Theta} \left(\frac{\partial \widehat{\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_r} - \frac{\partial \boldsymbol{\theta}}{\partial \theta_r} \right) p(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \mathbf{e}_r \int_{\Theta} p(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta} \end{aligned} \quad (2.68)$$

avec \mathbf{e}_r les vecteurs dont toutes les composantes sont nulles sauf la r -ième, égale à 1, après avoir utilisé les conditions 1 et 2 énoncées ci-dessus, ainsi que le fait que $\widehat{\boldsymbol{\theta}}(\mathbf{x})$ est indépendant de θ_r . En intégrant sur Ω , on retrouve donc $\mathbf{c}_r = \mathbf{e}_r$, soit $\mathbf{C} = \mathbf{I}$.

14. Il existe une multitude d'autres bornes bayésiennes, comme la borne de Bhattacharyya, la borne de Bobrovsky-Zakai, la borne d'Abel bayésienne... Voir [WW88, VB07].

15. p.p. : presque partout, c'est-à-dire (ici) pour tout $\mathbf{x} \in \Omega$ sauf éventuellement sur un ensemble de mesure nulle.

Enfin, la matrice \mathbf{V} est donnée par

$$\mathbf{V} = \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right\} \quad (2.69)$$

et si les dérivées secondes de la log-vraisemblance existent et que leur espérance est finie et non-nulle, on a aussi

$$\mathbf{V} = - \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ \frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right\}. \quad (2.70)$$

Remarque : On ne dispose pas des mêmes résultats sur la borne de Cramér-Rao bayésienne (BCRB) que sur la BCR (du cadre déterministe). En particulier, la BCRB n'est pas atteinte dans le cas général, même asymptotiquement. Pour le voir, on considère le cas d'un paramètre θ scalaire, et de données \mathbf{x} i.i.d. La BCRB est donc donnée par

$$\text{BCRB} = - \frac{1}{\mathbb{E}_{\mathbf{x}, \theta} \left\{ \frac{\partial^2 \ln p(\mathbf{x}, \theta)}{\partial \theta^2} \right\}} = \frac{1}{\mathbb{E}_\theta \{F(\theta)\} - \mathbb{E}_\theta \left\{ \frac{\partial^2 \ln \pi(\theta)}{\partial \theta^2} \right\}} = \frac{1}{N \mathbb{E}_\theta \{\tilde{F}(\theta)\} - \mathbb{E}_\theta \left\{ \frac{\partial^2 \ln \pi(\theta)}{\partial \theta^2} \right\}} \quad (2.71)$$

où $\tilde{F}(\theta)$ est l'information de Fisher pour une seule observation. Lorsque N devient infiniment grand, le terme $N \mathbb{E}_\theta \{\tilde{F}(\theta)\}$ devient prépondérant devant le terme $\mathbb{E}_\theta \left\{ \frac{\partial^2 \ln \pi(\theta)}{\partial \theta^2} \right\}$. Ceci implique que $\text{BCRB} \sim_{N \rightarrow +\infty} \frac{1}{N \mathbb{E}_\theta \{\tilde{F}(\theta)\}}$.

D'autre part, si l'on prend l'espérance par rapport à $\boldsymbol{\theta}$ des deux membres de l'inégalité (2.15), on obtient $\text{EQMG}(\hat{\boldsymbol{\theta}}) \geq \frac{1}{N} \mathbb{E}_\theta \left\{ \frac{1}{\tilde{F}(\theta)} \right\}$, c'est-à-dire également une borne inférieure de l'EQM, couramment appelée EBCR (« espérance de la borne de Cramér-Rao ») [VB07]. Or, par l'inégalité de Jensen, nous obtenons l'ordre suivant :

$$\text{EQMG}(\hat{\boldsymbol{\theta}}) \geq \frac{1}{N} \mathbb{E}_\theta \left\{ \frac{1}{\tilde{F}(\theta)} \right\} \geq \frac{1}{N \mathbb{E}_\theta \{\tilde{F}(\theta)\}} \quad (2.72)$$

soit

$$\text{EBCR} \geq \text{BCRB} \quad (2.73)$$

avec égalité si et seulement si l'information de Fisher \tilde{F} ne dépend pas explicitement de θ . Ainsi, lorsque l'inégalité est stricte, la BCRB n'est pas atteinte. Un exemple simple peut être trouvé dans [VB07, p.38].

Il apparaît donc que la EBCR serait plus intéressante à calculer que la BCRB. Pour autant, il faut se rappeler que la BCR n'est valable que pour un estimateur $\hat{\boldsymbol{\theta}}$ localement non biaisé. Ainsi, comme la EBCR porte sur l'erreur quadratique moyenne globale, elle n'est valable que pour un estimateur $\hat{\boldsymbol{\theta}}$ non biaisé *globalement*, c'est-à-dire pour toutes les valeurs de $\boldsymbol{\theta} \in \Theta$. De plus, dans des cas plus complexes, par exemple lorsque $\boldsymbol{\theta}$ est un vecteur, le calcul de la EBCR est généralement bien plus compliqué, puisqu'elle est donnée par $\mathbf{EBCR} = \mathbb{E}_\theta \{\mathbf{F}^{-1}(\boldsymbol{\theta})\}$, i.e., il est nécessaire de connaître l'expression de l'inverse de la matrice d'information de Fisher. Ce n'est pas le cas de la BCRB, qui nécessite le calcul de la matrice d'information de Fisher bayésienne, donnée par (2.70), et dont le calcul est en général bien plus aisé. Son inverse peut ensuite être calculée numériquement. De plus, la BCRB est souvent assez précise ; elle permet donc d'avoir un bon aperçu des performances d'estimation. \square

2.1.5.3.2 Borne de Weiss-Weinstein [WW88, VB07]

De la même façon que la borne de Barankin, la borne de Weiss-Weinstein (BWW) fait intervenir des points test $\tilde{\boldsymbol{\theta}}_k$, $k = 1, \dots, K$, qui peuvent être choisis arbitrairement sur l'espace des paramètres Θ . La BWW est obtenue en définissant les composantes de la fonction $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ du théorème 4 comme

$$[\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})]_k = \begin{cases} \left(\frac{p(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_k)}{p(\mathbf{x}, \boldsymbol{\theta})} \right)^{s_k} - \left(\frac{p(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_k)}{p(\mathbf{x}, \boldsymbol{\theta})} \right)^{1-s_k}, & \text{si } \boldsymbol{\theta} \in \Theta', \\ 0 & \text{sinon,} \end{cases} \quad (2.74)$$

où les $\mathbf{h}_k \in \mathcal{H} \triangleq \{\mathbf{h} \in \mathbb{R}^R \mid p(\mathbf{x}, \boldsymbol{\theta} \pm \mathbf{h}) > 0\}$, $k = 1, \dots, K$, désignent les valeurs tests¹⁶ et où les $s_k \in]0, 1[$ sont des valeurs tests supplémentaires pouvant elles aussi être arbitrairement choisies. On en déduit l'expression des colonnes \mathbf{c}_k de la matrice \mathbf{C} (qui est de taille $R \times K$), pour $1 \leq k \leq K$:

$$\begin{aligned} \mathbf{c}_k &= \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ (\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \left[\frac{p^{s_k}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_k)}{p^{s_k}(\mathbf{x}, \boldsymbol{\theta})} - \frac{p^{1-s_k}(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_k)}{p^{1-s_k}(\mathbf{x}, \boldsymbol{\theta})} \right] \right\} \\ &= \iint_{\Omega \times (\Theta' \cap (\Theta' - \mathbf{h}_k))} (\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) p^{s_k}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_k) p^{1-s_k}(\mathbf{x}, \boldsymbol{\theta}) \, d\mathbf{x} \, d\boldsymbol{\theta} \\ &\quad - \iint_{\Omega \times ((\Theta' + \mathbf{h}_k) \cap \Theta')} (\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) p^{1-s_k}(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_k) p^{s_k}(\mathbf{x}, \boldsymbol{\theta}) \, d\mathbf{x} \, d\boldsymbol{\theta} \\ &= \iint_{\Omega \times (\Theta' \cap (\Theta' - \mathbf{h}_k))} (\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) p^{s_k}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_k) p^{1-s_k}(\mathbf{x}, \boldsymbol{\theta}) \, d\mathbf{x} \, d\boldsymbol{\theta} \\ &\quad - \iint_{\Omega \times (\Theta' \cap (\Theta' - \mathbf{h}_k))} (\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta} - \mathbf{h}_k) p^{1-s_k}(\mathbf{x}, \boldsymbol{\theta}) p^{s_k}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_k) \, d\mathbf{x} \, d\boldsymbol{\theta} \\ &= \mathbf{h}_k \iint_{\Omega \times (\Theta' \cap (\Theta' - \mathbf{h}_k))} p^{s_k}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_k) p^{1-s_k}(\mathbf{x}, \boldsymbol{\theta}) \, d\mathbf{x} \, d\boldsymbol{\theta} \end{aligned} \quad (2.75)$$

en utilisant le changement de variables $\boldsymbol{\theta}' = \boldsymbol{\theta} - \mathbf{h}_k$ dans la deuxième intégrale. Dans les lignes de calcul conduisant à (2.75), nous avons utilisé la notation $\Theta' + \mathbf{h}_k$ pour désigner l'ensemble Θ' translaté de \mathbf{h}_k , c'est-à-dire $\Theta' + \mathbf{h}_k \triangleq \{\boldsymbol{\theta} + \mathbf{h}_k \mid \boldsymbol{\theta} \in \Theta'\}$. Notons que ces lignes montrent que l'ensemble Θ' , qui comprend le support de la loi *a priori* sur $\boldsymbol{\theta}$, peut être fini. Finalement, on obtient donc

$$\mathbf{c}_k = \mathbf{h}_k \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ \left(\frac{p(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_k)}{p(\mathbf{x}, \boldsymbol{\theta})} \right)^{s_k} \right\}. \quad (2.76)$$

D'autre part, les éléments de la matrice \mathbf{V} sont donnés, pour $1 \leq i, j \leq K$, par

$$[\mathbf{V}]_{i,j} = \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ \left[\frac{p^{s_i}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_i)}{p^{s_i}(\mathbf{x}, \boldsymbol{\theta})} - \frac{p^{1-s_i}(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_i)}{p^{1-s_i}(\mathbf{x}, \boldsymbol{\theta})} \right] \left[\frac{p^{s_j}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_j)}{p^{s_j}(\mathbf{x}, \boldsymbol{\theta})} - \frac{p^{1-s_j}(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_j)}{p^{1-s_j}(\mathbf{x}, \boldsymbol{\theta})} \right] \right\}. \quad (2.77)$$

On peut réagencer les termes des matrices \mathbf{C} et \mathbf{V} pour aboutir à l'expression la plus répandue des « bornes de Weiss-Weinstein avec K points test » :

$$\mathbf{W}(\mathbf{H}, \mathbf{s}) = \mathbf{H}(\mathbf{G}(\mathbf{H}, \mathbf{s}))^{-1} \mathbf{H}^\top \quad (2.78)$$

16. De la même façon que dans la section 2.1.5.2.2, les points test peuvent être définis comme les $\tilde{\boldsymbol{\theta}}_k = \boldsymbol{\theta} \pm \mathbf{h}_k$, $k = 1, \dots, K$.

où $\mathbf{H} \triangleq [\mathbf{h}_1, \dots, \mathbf{h}_K]$, $\mathbf{s} \triangleq [s_1, \dots, s_K]^\top$, et \mathbf{G} (dont la dépendance en \mathbf{H} et \mathbf{s} ne sera pas toujours marquée, afin de ne pas alourdir les écritures) est la matrice de taille $K \times S$ dont les éléments sont donnés, pour $1 \leq i, j \leq K$, par

$$[\mathbf{G}]_{i,j} = \frac{\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}} \left\{ \left[\frac{p^{s_i}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_i)}{p^{s_i}(\mathbf{x}, \boldsymbol{\theta})} - \frac{p^{1-s_i}(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_i)}{p^{1-s_i}(\mathbf{x}, \boldsymbol{\theta})} \right] \left[\frac{p^{s_j}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_j)}{p^{s_j}(\mathbf{x}, \boldsymbol{\theta})} - \frac{p^{1-s_j}(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_j)}{p^{1-s_j}(\mathbf{x}, \boldsymbol{\theta})} \right] \right\}}{\mathbb{E}_{\mathbf{x},\boldsymbol{\theta}} \left\{ \frac{p^{s_i}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_i)}{p^{s_i}(\mathbf{x}, \boldsymbol{\theta})} \right\} \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}} \left\{ \frac{p^{s_j}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_j)}{p^{s_j}(\mathbf{x}, \boldsymbol{\theta})} \right\}}. \quad (2.79)$$

En développant (2.79), on peut réexprimer les éléments de la matrice \mathbf{G} sous la forme

$$[\mathbf{G}]_{i,j} = \frac{e^{\tilde{\mu}(s_i, s_j, \mathbf{h}_i, \mathbf{h}_j)} + e^{\tilde{\mu}(1-s_i, 1-s_j, -\mathbf{h}_i, -\mathbf{h}_j)} - e^{\tilde{\mu}(s_i, 1-s_j, \mathbf{h}_i, -\mathbf{h}_j)} - e^{\tilde{\mu}(1-s_i, s_j, -\mathbf{h}_i, \mathbf{h}_j)}}{e^{\mu(s_i, \mathbf{h}_i)} e^{\mu(s_j, \mathbf{h}_j)}}, \quad (2.80)$$

où l'on définit, pour $0 < \alpha, \beta < 1$ et $(\mathbf{h}_a, \mathbf{h}_b) \in \mathcal{H}^2$

$$\tilde{\mu}(\alpha, \beta, \mathbf{h}_a, \mathbf{h}_b) \triangleq \ln \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}} \left\{ \frac{p^\alpha(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_a) p^\beta(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_b)}{p^{\alpha+\beta}(\mathbf{x}, \boldsymbol{\theta})} \right\} \quad (2.81a)$$

$$= \ln \int_{\Theta} \int_{\Omega} p^\alpha(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_a) p^\beta(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_b) p^{1-\alpha-\beta}(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{x}, \quad (2.81b)$$

et

$$\mu(\alpha, \mathbf{h}_a) \triangleq \tilde{\mu}(\alpha, 0, \mathbf{h}_a, \mathbf{0}) = \ln \mathbb{E}_{\mathbf{x},\boldsymbol{\theta}} \left\{ \left(\frac{p(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_a)}{p(\mathbf{x}, \boldsymbol{\theta})} \right)^\alpha \right\} \quad (2.82a)$$

$$= \ln \int_{\Theta} \int_{\Omega} p^\alpha(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_a) p^{1-\alpha}(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{x}. \quad (2.82b)$$

Il est donc clair que pour calculer la borne de Weiss-Weinstein, nous avons seulement besoin de calculer l'expression de $\tilde{\mu}(\alpha, \beta, \mathbf{h}_a, \mathbf{h}_b)$, ou de $e^{\tilde{\mu}(\alpha, \beta, \mathbf{h}_a, \mathbf{h}_b)}$. Des formes analytiques de ces quantités peuvent être trouvées pour de nombreux problèmes, notamment pour des signaux gaussiens à moyenne et/ou matrice de covariance paramétrées [Van01, VRBM14].

Remarque : Comme indiqué dans [WW88], une fonction similaire à $\mu(\alpha, \mathbf{h}_a)$ a été utilisée par Chernoff pour obtenir une borne sur la probabilité d'erreur minimale atteignable dans un problème de test d'hypothèse binaire [Che52]. Des propriétés intéressantes de la fonction $\mu(\alpha, \mathbf{h}_a)$ sont détaillées dans [Che52], ainsi que dans [Van68, Chapitre 2, pp.116–133]. \square

De la même manière que la borne de Barankin, nous obtenons donc, pour chaque valeur de points test $\mathbf{H} \in \mathcal{H}^K$ et $\mathbf{s} \in]0, 1[^K$ choisie, une borne inférieure de l'erreur quadratique moyenne (globale) $\mathbf{W}(\mathbf{H}, \mathbf{s})$. Ainsi, la borne la plus précise peut être obtenue en maximisant $\mathbf{W}(\mathbf{H}, \mathbf{s})$ par rapport à \mathbf{H} et \mathbf{s} . C'est celle-ci qui est généralement admise comme étant la borne de Weiss-Weinstein :

$$\mathbf{BWW} \triangleq \sup_{\substack{\mathbf{H} \in \mathcal{H}^K \\ \mathbf{s} \in]0, 1[^K}} \mathbf{W}(\mathbf{H}, \mathbf{s}). \quad (2.83)$$

L'opération de maximisation dans (2.83) (et dans (2.63) pour la borne de Barankin) permet d'obtenir une borne inférieure de l'EQM très précise. En effet, ces bornes rendent compte du phénomène de « décrochement » de l'estimation (voir Section 2.1.5.2.2) rencontré notamment en analyse spectrale [MS69, RB74, RFL⁺08] (nous avons que l'une des limites de la borne de Cramér-Rao est qu'elle ne met pas en évidence ce phénomène).

Néanmoins, il est important de noter ici que l'opérateur « sup » dans (2.83) porte sur l'ensemble $\mathcal{W} \triangleq \{\mathbf{W}(\mathbf{H}, \mathbf{s}) \mid \mathbf{H} \in \mathcal{H}^K, \mathbf{s} \in]0, 1[^K\}$, qui est un sous-ensemble de \mathbb{S}^R , qui désigne l'ensemble

des matrices carrées de taille $R \times R$, symétriques, semi-définies positives. Cet ensemble \mathbb{S}^R est ici muni de la relation d'ordre « \succcurlyeq » (vue aux théorèmes 1 et 4 notamment), communément appelée « relation d'ordre de Löwner » (ou simplement « ordre de Löwner »). Or, cette relation n'est pas une relation d'ordre *totale*, c'est-à-dire que deux matrices $\mathbf{A}, \mathbf{B} \in \mathbb{S}^R$ ne sont pas toujours mutuellement comparables. Cela peut poser un certain nombre de problèmes, en particulier que la notion de borne supérieure de \mathcal{W} peut ne pas être définie de manière unique. Cette problématique est traitée dans le Chapitre 3, Section 3.3. Elle est valable aussi pour la borne de Barankin pour l'estimation d'un vecteur de paramètres (2.63), et pour toutes les bornes qui font intervenir des points test d'une manière générale.

Une autre remarque importante concerne la complexité calculatoire de la borne de Weiss-Weinstein. En premier lieu, revenons sur le nombre K de points test. Celui-ci peut être choisi librement, en théorie. Afin d'obtenir la borne la plus précise possible, il est souhaitable de prendre K le plus grand possible, puisque cela permet d'échantillonner plus finement l'espace des paramètres Θ . En particulier, on sait que choisir $K < R$ ne peut permettre d'obtenir une borne précise [LRNM10, Lemme 1]. Toutefois, une augmentation du nombre de points test K se fait au prix d'une complexité calculatoire accrue, du fait notamment i) de la nécessité d'inverser la matrice \mathbf{G} de taille $K \times K$, et ii) de la dimension de l'espace de recherche de la meilleure BWW, qui est $\mathcal{H}^K \times]0, 1[^K \subset \mathbb{R}^{(R+1) \times K}$.

Plusieurs approches visant à régler ces problèmes existent dans la littérature. L'une d'elles consiste à choisir un ensemble de points tests raisonnables et à effectuer l'optimisation apparaissant dans (2.83) sur cet ensemble [CGQL08, Cha14, Ren15]. Cela revient à effectuer un compromis entre la complexité de calcul et la précision de la borne. Il existe également d'autres approches fondées sur le choix d'un nombre important de points test uniformément répartis sur Θ , et sur l'utilisation d'une transformation linéaire, permettant de s'affranchir de l'étape d'optimisation dans (2.83), voir [TT09, TT10a, Ren15].

Nous nous limiterons dans la suite à la première approche. En particulier, nous choisirons $K = R$. De plus, afin de pouvoir obtenir des expressions analytiques pour les éléments de la matrice \mathbf{G} , nous limiterons à une matrice \mathbf{H} diagonale, i.e., $\mathbf{H} = [h_1 \mathbf{e}_1, \dots, h_R \mathbf{e}_R] = \text{diag}(h_1, \dots, h_R)$. Enfin, une simplification supplémentaire couramment adoptée dans la littérature consiste à choisir $s_1 = s_2 = \dots = s_R = 1/2$, car il a été remarqué que dans de nombreux cas, c'est la valeur qui conduit à la borne la plus précise (et accessoirement des calculs plus simples) [WW88, VRBM14, XMGM16].

Pour résumer, la borne de Weiss-Weinstein possède les mêmes avantages sur la borne de Cramér-Rao bayésienne que ceux de la borne de Barankin sur la borne de Cramér-Rao déterministe : i) de faibles conditions de régularité, qui permettent en particulier de la calculer pour un problème d'estimation de paramètres discrets, mais également pour une loi *a priori* uniforme sur les paramètres ; ii) sa plus grande précision, notamment dans la zone de non information où elle rend compte du phénomène de décrochement. Elle requiert en revanche un temps de calcul plus important que la BCRB, du fait de l'étape de maximisation nécessaire. Néanmoins, à titre de comparaison, ce coût calculatoire supplémentaire ne dépasse généralement pas celui des performances d'un estimateur comme celui du maximum de vraisemblance qui, en plus d'une étape de maximisation, nécessite aussi de nombreuses (et coûteuses) simulations de Monte-Carlo (d'autant plus coûteuses que l'étape de maximisation doit être exécutée lors de chaque simulation de Monte-Carlo).

2.1.5.4 Combinaison de bornes, bornes hybrides

Nous avons vu que les bornes citées et présentées précédemment n'ont pas toutes les mêmes propriétés. Il peut donc être intéressant de les combiner, et cela peut être fait de différentes

manières. L'une d'elles consiste à combiner les propriétés de plusieurs bornes, indifféremment pour toutes les composantes du vecteur de paramètres inconnus $\boldsymbol{\theta}$. Pour cela, il suffit de définir la fonction $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ du théorème 4 comme une combinaison linéaire des différentes fonctions $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ associées à chacune des bornes que l'on souhaite combiner. C'est le cas de la borne de McAulay-Hofstetter [MH71], qui combine la borne de Cramér-Rao et la borne de McAulay-Seidman (Barankin); c'est le cas également de la borne d'Abel [Abe93, RFLR06], qui combine les bornes de McAulay-Seidman et de Bhattacharyya; enfin un dernier exemple est celui de la borne de Quinlan-Chaumette-Larzabal [QCL06], qui est également une combinaison des bornes précédentes.

Une autre manière de combiner des bornes inférieures de l'EQM est particulièrement intéressante lorsque les paramètres qui composent le vecteur $\boldsymbol{\theta}$ sont de nature très différente. En supposant qu'on scinde $\boldsymbol{\theta}$ en deux sous-vecteurs $\boldsymbol{\theta}_1$ et $\boldsymbol{\theta}_2$ de sorte que $\boldsymbol{\theta} = [\boldsymbol{\theta}_1; \boldsymbol{\theta}_2]$ (où $[\boldsymbol{\theta}_1; \boldsymbol{\theta}_2]$ désigne la concaténation verticale de deux vecteurs, i.e., $[\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T]^T$), on peut chercher à obtenir une borne inférieure de l'EQM pour $\boldsymbol{\theta}$ qui correspond à une certaine borne pour le sous-vecteur $\boldsymbol{\theta}_1$, combinée à une autre borne pour le sous-vecteur $\boldsymbol{\theta}_2$. Une telle borne peut être construite en définissant la fonction $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta})$ du théorème 4 comme la concaténation des fonctions $\mathbf{v}_1(\mathbf{x}, \boldsymbol{\theta})$ et $\mathbf{v}_2(\mathbf{x}, \boldsymbol{\theta})$ associées à chacune des bornes que l'on souhaite combiner, i.e., $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta}) = [\mathbf{v}_1(\mathbf{x}, \boldsymbol{\theta}); \mathbf{v}_2(\mathbf{x}, \boldsymbol{\theta})]$. Néanmoins, la borne obtenue ne consistera pas en la concaténation naïve des bornes combinées, car des termes dits « croisés » apparaissent du fait de l'estimation conjointe des vecteurs $\boldsymbol{\theta}_1$ et $\boldsymbol{\theta}_2$.

Nous donnons maintenant deux exemples de ce type de borne combinée : d'abord un exemple de combinaison de deux bornes bayésiennes, puis un exemple de combinaison d'une borne déterministe avec une borne bayésienne, c'est-à-dire de borne hybride, utile pour des problèmes d'estimation hybride.

2.1.5.4.1 Combinaison de bornes bayésiennes : la borne (combinée) de Cramér-Rao/Weiss-Weinstein bayésienne (BCRWWB). Supposons que nous cherchons à estimer un vecteur de paramètres aléatoires $\boldsymbol{\theta} = [\boldsymbol{\theta}_1; \boldsymbol{\theta}_2]$, où $\boldsymbol{\theta}_1 \in \mathbb{R}^P$, $\boldsymbol{\theta}_2 \in \mathbb{R}^Q$ et $\boldsymbol{\theta} \in \mathbb{R}^R$ ($P + Q = R$), et que nous souhaitons obtenir une borne inférieure combinée de l'EQM sur $\boldsymbol{\theta}$ qui correspond à la borne de Cramér-Rao bayésienne (BCRB) pour l'erreur sur $\boldsymbol{\theta}_1$, et à la borne de Weiss-Weinstein (BWW) pour l'erreur sur $\boldsymbol{\theta}_2$. Comme nous venons de l'expliquer, nous définissons donc $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta}) = [\mathbf{v}_1(\mathbf{x}, \boldsymbol{\theta}); \mathbf{v}_2(\mathbf{x}, \boldsymbol{\theta})]$, où $\mathbf{v}_1(\mathbf{x}, \boldsymbol{\theta})$ est définie, comme dans (2.66), par

$$\mathbf{v}_1(\mathbf{x}, \boldsymbol{\theta}) = \begin{cases} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} & , \text{ si } \boldsymbol{\theta} \in \Theta', \\ \mathbf{0} & , \text{ sinon,} \end{cases} \quad (2.84)$$

et $\mathbf{v}_2(\mathbf{x}, \boldsymbol{\theta})$ est définie, comme dans (2.74), par

$$[\mathbf{v}_2(\mathbf{x}, \boldsymbol{\theta})]_k = \begin{cases} \left(\frac{p(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_k^{(P+K)})}{p(\mathbf{x}, \boldsymbol{\theta})} \right)^{s_k} - \left(\frac{p(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_k^{(P+K)})}{p(\mathbf{x}, \boldsymbol{\theta})} \right)^{1-s_k} & , \text{ si } \boldsymbol{\theta} \in \Theta', \\ 0 & , \text{ sinon,} \end{cases} \quad (2.85)$$

où on définit le vecteur $\mathbf{h}_k^{(P+K)} \in \mathbb{R}^{P+K}$ comme $\mathbf{h}_k^{(P+K)} \triangleq [\mathbf{0}_{P \times 1}; \mathbf{h}_k]$, avec $\mathbf{h}_k \in \mathbb{R}^K$ et K le nombre de points-tests. Dans la suite, nous nous en tiendrons au cas où $K = Q$, pour les mêmes raisons que celles évoquées dans la section 2.1.5.3.2.

Sous réserve que les deux conditions de régularité suivantes soient respectées :

1. les fonctions $\theta_{1,1}, \dots, \theta_{1,P} \mapsto p(\mathbf{x}, \boldsymbol{\theta})$ sont absolument continues, pour tout $\mathbf{x} \in \Omega$ p.p.,
2. $\lim_{\theta_{1,r} \rightarrow \pm\infty} \theta_{1,r} p(\mathbf{x}, \boldsymbol{\theta}) = 0$, pour $r = 1, \dots, P$, et pour tout $\mathbf{x} \in \Omega$ p.p.,

nous obtenons la matrice \mathbf{C} suivante, notée \mathbf{C}^π (pour souligner la dépendance selon les distributions *a priori* sur les paramètres) :

$$\mathbf{C}^\pi = \begin{bmatrix} \mathbf{I}_P & \mathbf{0}_{P \times Q} \\ \mathbf{0}_{Q \times P} & \mathbf{C}_{22}^\pi \end{bmatrix} \quad (2.86)$$

avec les colonnes de la matrice \mathbf{C}_{22}^π données, pour $1 \leq q \leq Q$, par

$$\mathbf{c}_q^\pi = \mathbf{h}_q \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ \left(\frac{p(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_q^{(R)})}{p(\mathbf{x}, \boldsymbol{\theta})} \right)^{s_q} \right\}. \quad (2.87)$$

Notons que les valeurs des autres blocs de la matrice \mathbf{C}^π , en particulier celles des blocs de gauche (supérieur et inférieur) d'une part, et du bloc supérieur droit d'autre part, sont justifiées par le respect des conditions de régularité 1 et 2 et par la forme du vecteur $\mathbf{h}_k^{(R)}$, respectivement (les manipulations sont très similaires à celles aboutissant aux équations (2.68) et (2.75)).

D'autre part nous obtenons l'expression de la matrice \mathbf{V} , notée \mathbf{V}^π , comme

$$\mathbf{V}^\pi = \begin{bmatrix} \mathbf{V}_{11}^\pi & \mathbf{V}_{12}^\pi \\ (\mathbf{V}_{12}^\pi)^\top & \mathbf{V}_{22}^\pi \end{bmatrix} \quad (2.88)$$

où \mathbf{V}_{11}^π est la matrice de taille $P \times P$ donnée par

$$\mathbf{V}_{11}^\pi = \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1^\top} \right\} = - \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ \frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^\top} \right\}, \quad (2.89)$$

où la dernière égalité est vraie si les dérivées secondes de la log-vraisemblance existent et si leur espérance est finie. \mathbf{V}_{12}^π est la matrice de taille $P \times Q$ dont les éléments sont donnés par

$$[\mathbf{V}_{12}^\pi]_{p,q} = \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_{1,p}} \left(\frac{p^{s_q}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_q^{(R)})}{p^{s_q}(\mathbf{x}, \boldsymbol{\theta})} - \frac{p^{1-s_q}(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_q^{(R)})}{p^{1-s_q}(\mathbf{x}, \boldsymbol{\theta})} \right) \right\}, \quad (2.90)$$

et \mathbf{V}_{22}^π est la matrice de taille $Q \times Q$ dont les éléments sont donnés, pour $1 \leq i, j \leq Q$, par

$$[\mathbf{V}_{22}^\pi]_{i,j} = \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} \left\{ \left[\frac{p^{s_i}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_i^{(R)})}{p^{s_i}(\mathbf{x}, \boldsymbol{\theta})} - \frac{p^{1-s_i}(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_i^{(R)})}{p^{1-s_i}(\mathbf{x}, \boldsymbol{\theta})} \right] \left[\frac{p^{s_j}(\mathbf{x}, \boldsymbol{\theta} + \mathbf{h}_j^{(R)})}{p^{s_j}(\mathbf{x}, \boldsymbol{\theta})} - \frac{p^{1-s_j}(\mathbf{x}, \boldsymbol{\theta} - \mathbf{h}_j^{(R)})}{p^{1-s_j}(\mathbf{x}, \boldsymbol{\theta})} \right] \right\}. \quad (2.91)$$

Nous obtenons ainsi la borne bayésienne combinée de Cramér-Rao/Weiss-Weinstein (BBCRWW) [BV06, VB07, TT10b]. Celle-ci a été introduite en premier par Bell et Van Trees [BV06] sous une forme récursive, pour un problème de poursuite de cible.

2.1.5.4.2 Borne hybrides : la borne hybride de Cramér-Rao/Weiss-Weinstein On suppose maintenant que l'on cherche à estimer un vecteur de paramètres $\boldsymbol{\theta} = [\boldsymbol{\theta}_d; \boldsymbol{\theta}_a]$, où $\boldsymbol{\theta}_d \in \mathbb{R}^P$ est un vecteur de paramètres supposés déterministes dont la vraie valeur est notée $\boldsymbol{\theta}_d^*$, et $\boldsymbol{\theta}_a \in \mathbb{R}^Q$ est un vecteur de paramètres supposés aléatoires. De la même manière que précédemment, on peut combiner des bornes déterministes (pour l'erreur par rapport à $\boldsymbol{\theta}_d$) avec des bornes bayésiennes (pour l'erreur par rapport à $\boldsymbol{\theta}_a$). De telles bornes ont été étudiées dans la littérature, principalement en traitement d'antennes [RS87, RM97, TW07, VB07, BGR⁺08, RGC⁺15].

On présente ici la borne hybride de Cramér-Rao/Weiss-Weinstein (BHCRWW). Elle est proche de la BBCRWW de la section précédente dans la mesure où, pour l'obtenir, il suffit de remplacer la distribution $p(\mathbf{x}, \boldsymbol{\theta})$ dans les équations de la section précédente par $p(\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d)$, et les espérances $\mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}}\{\cdot\}$ par $\mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d^*}\{\cdot\}$. Plus explicitement, on définit donc, de la même manière

que précédemment, $\mathbf{v}(\mathbf{x}, \boldsymbol{\theta}_a, \boldsymbol{\theta}_d^*) = [\mathbf{v}_1(\mathbf{x}, \boldsymbol{\theta}_a, \boldsymbol{\theta}_d^*); \mathbf{v}_2(\mathbf{x}, \boldsymbol{\theta}_a, \boldsymbol{\theta}_d^*)]$, où $\mathbf{v}_1(\mathbf{x}, \boldsymbol{\theta}_a, \boldsymbol{\theta}_d^*)$ est défini, comme dans (2.50), par

$$\mathbf{v}_1(\mathbf{x}, \boldsymbol{\theta}_a, \boldsymbol{\theta}_d^*) = \begin{cases} \left. \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d)}{\partial \boldsymbol{\theta}_d} \right|_{\boldsymbol{\theta}_d^*} & , \text{ si } (\boldsymbol{\theta}_a, \boldsymbol{\theta}_d^*) \in \Theta', \\ \mathbf{0} & , \text{ sinon,} \end{cases} \quad (2.92)$$

où $\Theta' = \{(\boldsymbol{\theta}_a, \boldsymbol{\theta}_d) \mid p(\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d) > 0\}$, et les composantes de $\mathbf{v}_2(\mathbf{x}, \boldsymbol{\theta}_a, \boldsymbol{\theta}_d^*)$ sont données, comme dans (2.74), pour $1 \leq k \leq K$, par

$$[\mathbf{v}_2(\mathbf{x}, \boldsymbol{\theta}_a, \boldsymbol{\theta}_d^*)]_k = \begin{cases} \left(\frac{p(\mathbf{x}, \boldsymbol{\theta}_a + \mathbf{h}_k; \boldsymbol{\theta}_d^*)}{p(\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d^*)} \right)^{s_k} - \left(\frac{p(\mathbf{x}, \boldsymbol{\theta}_a - \mathbf{h}_k; \boldsymbol{\theta}_d^*)}{p(\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d^*)} \right)^{1-s_k} & , \text{ si } (\boldsymbol{\theta}_a, \boldsymbol{\theta}_d^*) \in \Theta', \\ 0 & , \text{ sinon,} \end{cases} \quad (2.93)$$

où $\mathbf{h}_k \in \mathcal{H} \triangleq \{\mathbf{h} \in \mathbb{R}^Q \mid p(\mathbf{x}, \boldsymbol{\theta}_a \pm \mathbf{h}; \boldsymbol{\theta}_d) > 0\} \subset \mathbb{R}^Q$, pour $k = 1, \dots, K$. En supposant que l'on prend $K = Q$ points test, et que les conditions de régularité suivantes sont satisfaites :

1. On considère une classe d'estimateurs $\widehat{\boldsymbol{\theta}}$ non biaisés par rapport à $\boldsymbol{\theta}_d$ (en moyenne par rapport à $\boldsymbol{\theta}_a$), c'est-à-dire tels que

$$\mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d^*} \{\widehat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}_{|\boldsymbol{\theta}_d^*}\} = [\mathbf{0}; \mathbf{d}] \quad (2.94)$$

où \mathbf{d} est un vecteur quelconque de taille Q , indépendant de $\boldsymbol{\theta}$;

2. De la même manière qu'en (2.53), on suppose que, quel que soit $\boldsymbol{\theta}_d^*$, on a

$$\mathbb{E}_{\mathbf{x}|\boldsymbol{\theta}_a; \boldsymbol{\theta}_d^*} \left\{ \left. \frac{\partial \ln p(\mathbf{x} \mid \boldsymbol{\theta}_a; \boldsymbol{\theta}_d)}{\partial \boldsymbol{\theta}_d} \right|_{\boldsymbol{\theta}_d^*} \right\} = \mathbf{0}, \quad (2.95)$$

on retrouve la matrice \mathbf{C} , notée $\mathbf{C}^{\boldsymbol{\theta}_d^*}$ (pour souligner sa dépendance en la vraie valeur du vecteur de paramètres $\boldsymbol{\theta}_d^*$), donnée par

$$\mathbf{C}^{\boldsymbol{\theta}_d^*} = \begin{bmatrix} \mathbf{I}_P & \mathbf{0}_{P \times Q} \\ \mathbf{0}_{Q \times P} & \mathbf{C}_{22}^{\boldsymbol{\theta}_d^*} \end{bmatrix} \quad (2.96)$$

avec les colonnes de la matrice $\mathbf{C}_{22}^{\boldsymbol{\theta}_d^*}$ données, pour $1 \leq q \leq Q$, par

$$\mathbf{c}_q^{\boldsymbol{\theta}_d^*} = \mathbf{h}_q \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d^*} \left\{ \left(\frac{p(\mathbf{x}, \boldsymbol{\theta}_a + \mathbf{h}_q; \boldsymbol{\theta}_d^*)}{p(\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d^*)} \right)^{s_q} \right\}, \quad (2.97)$$

et la matrice \mathbf{V} , notée $\mathbf{V}^{\boldsymbol{\theta}_d^*}$, donnée par

$$\mathbf{V}^{\boldsymbol{\theta}_d^*} = \begin{bmatrix} \mathbf{V}_{11}^{\boldsymbol{\theta}_d^*} & \mathbf{V}_{12}^{\boldsymbol{\theta}_d^*} \\ (\mathbf{V}_{12}^{\boldsymbol{\theta}_d^*})^\top & \mathbf{V}_{22}^{\boldsymbol{\theta}_d^*} \end{bmatrix} \quad (2.98)$$

où $\mathbf{V}_{11}^{\boldsymbol{\theta}_d^*}$ est la matrice donnée par

$$\mathbf{V}_{11}^{\boldsymbol{\theta}_d^*} = \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d^*} \left\{ \left. \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d)}{\partial \boldsymbol{\theta}_d} \right|_{\boldsymbol{\theta}_d^*} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d)}{\partial \boldsymbol{\theta}_d^\top} \right|_{\boldsymbol{\theta}_d^*} \right\} = - \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d^*} \left\{ \left. \frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta}_a; \boldsymbol{\theta}_d)}{\partial \boldsymbol{\theta}_d \partial \boldsymbol{\theta}_d^\top} \right|_{\boldsymbol{\theta}_d^*} \right\}, \quad (2.99)$$

où la dernière égalité est vraie si les dérivées secondes de la log-vraisemblance existent et si leur espérance est finie. Remarquons que l'inverse du bloc $\mathbf{V}_{11}^{\boldsymbol{\theta}_d^*}$ correspond exactement à la « borne de

Cramér-Rao modifiée » au sens de [GRM98], c'est-à-dire la borne de Cramér-Rao sur le vecteur de paramètres θ_d^* , lorsque les paramètres du vecteur θ_a sont considérés comme des paramètres de nuisance. $V_{12}^{\theta_d^*}$ est la matrice de taille $P \times Q$ dont les éléments sont donnés par

$$[V_{12}^{\theta_d^*}]_{p,q} = \mathbb{E}_{\mathbf{x}, \theta_a; \theta_d^*} \left\{ \frac{\partial \ln p(\mathbf{x}, \theta_a; \theta_d)}{\partial \theta_{d,p}} \bigg|_{\theta_d^*} \left(\frac{p^{s_q}(\mathbf{x}, \theta_a + \mathbf{h}_q; \theta_d^*)}{p^{s_q}(\mathbf{x}, \theta_a; \theta_d^*)} - \frac{p^{1-s_q}(\mathbf{x}, \theta_a - \mathbf{h}_q; \theta_d^*)}{p^{1-s_q}(\mathbf{x}, \theta_a; \theta_d^*)} \right) \right\}, \quad (2.100)$$

et $V_{22}^{\theta_d^*}$ est la matrice de taille $Q \times Q$ dont les éléments sont donnés, pour $1 \leq i, j \leq Q$, par

$$[V_{22}^{\theta_d^*}]_{i,j} = \mathbb{E}_{\mathbf{x}, \theta_a; \theta_d^*} \left\{ \left[\frac{p^{s_i}(\mathbf{x}, \theta_a + \mathbf{h}_i; \theta_d^*)}{p^{s_i}(\mathbf{x}, \theta_a; \theta_d^*)} - \frac{p^{1-s_i}(\mathbf{x}, \theta_a - \mathbf{h}_i; \theta_d^*)}{p^{1-s_i}(\mathbf{x}, \theta_a; \theta_d^*)} \right] \times \left[\frac{p^{s_j}(\mathbf{x}, \theta_a + \mathbf{h}_j; \theta_d^*)}{p^{s_j}(\mathbf{x}, \theta_a; \theta_d^*)} - \frac{p^{1-s_j}(\mathbf{x}, \theta_a - \mathbf{h}_j; \theta_d^*)}{p^{1-s_j}(\mathbf{x}, \theta_a; \theta_d^*)} \right] \right\}. \quad (2.101)$$

Nous obtenons donc ainsi la borne hybride de Cramér-Rao/Weiss-Weinstein (BHCRWW). Bien que son expression semble fortement similaire à celle de la BBCCRWW présentée dans la section précédente, il faut garder à l'esprit qu'il est généralement plus simple d'obtenir des expressions analytiques pour la BHCRWW que pour la BBCCRWW. En effet, les espérances $\mathbb{E}_{\mathbf{x}, \theta} \{ \cdot \}$ de la BBCCRWW contiennent $N + P + Q$ intégrales (nombre d'observations+taille du vecteur θ_1 +taille du vecteur θ_2), tandis que les espérances $\mathbb{E}_{\mathbf{x}, \theta_a; \theta_d^*} \{ \cdot \}$ n'en contiennent que $N + Q$ (nombre d'observations+taille du vecteur θ_a). Comme nous le verrons notamment au Chapitre 3, certaines des intégrales supplémentaires dans la BBCCRWW peuvent ne pas admettre de forme analytique, ce qui contraint à recourir à des méthodes numériques pour les calculer. Cela complique d'autant plus l'obtention de la borne, et contribue à privilégier le calcul de la BHCRWW.

2.2 Estimation des paramètres de signaux comportant des points de rupture

La section précédente nous a donné l'occasion de passer en revue un bon nombre de notions se rapportant à l'estimation de paramètres pour le traitement du signal. Nous avons, dans un premier temps, insisté sur l'importance de connaître la distribution (au moins asymptotique) d'un estimateur, afin d'étudier ses performances. Dans un second temps, nous avons expliqué l'intérêt de calculer des bornes inférieures sur les moments de ces estimateurs (en particulier d'ordre 2, comme l'EQM). Rappelons que de telles bornes ont l'avantage d'être valables pour toute une famille d'estimateurs, et qu'elles fournissent un aperçu instructif des performances d'estimation, en particulier lorsque l'on n'a pas accès à la distribution des estimateurs.

Dans cette section et dans toute la suite, nous nous intéressons à un cas particulier de signaux : celui des signaux pouvant comprendre un ou plusieurs points de rupture, dont la localisation est généralement inconnue. Nous chercherons donc à estimer les localisations de ces ruptures. Nous avons vu dans le Chapitre 1 qu'il existe diverses manières de modéliser et de traiter ce type de signaux, par exemple si le nombre d'observations augmente au fur et à mesure (traitement « on-line ») ou si celui-ci est fixé (traitement « off-line », ou analyse rétrospective). Nous avons également vu que le but de ce traitement peut être la détection et/ou l'estimation des ruptures dans le signal. Les travaux présentés dans cette thèse se focalisent uniquement sur le problème de l'estimation dans un contexte off-line. Toutefois, dans cette section, nous mentionnerons aussi des travaux traitant du problème de détection, car il est parfois traité de manière indissociable du problème d'estimation. Pour une vue d'ensemble de toutes les problématiques afférentes aux données comportant des points de rupture dans un contexte off-line et des avancées récentes dans ce domaine, on peut également se reporter à l'étude de Jandhyala et al. [JFML13].

2.2.1 La localisation de ruptures dans un signal échantillonné : formulation(s) du problème

Pour définir formellement le modèle de signaux comportant des ruptures, nous reprenons les notations de la Section 2.1 : on considère $\mathbf{x} \triangleq [x_1, \dots, x_N]^T$ le vecteur des observations (ou signal observé). Pour les raisons évoquées dans le Chapitre 1, nous considérons le signal observé aléatoire ; ainsi chaque observation x_n , $n = 1, \dots, N$, correspond à une variable aléatoire, dont la distribution, notée¹⁷ $p(x_n | x_{n-1}, \dots, x_1; \boldsymbol{\eta}(n))$, est paramétrée par un vecteur noté $\boldsymbol{\eta}(n)$. Par exemple, dans le cas où la distribution de chaque observation est indépendante et gaussienne, i.e., $x_n \sim \mathcal{N}(m(n), \sigma^2(n))$, on a $\boldsymbol{\eta}(n) = [m(n), \sigma^2(n)]^T$. Pour un signal qui ne comporte pas de point de rupture, $\boldsymbol{\eta}(n)$ ne varie pas en fonction de n , il est égal à un vecteur constant noté¹⁸ $\boldsymbol{\eta}_0^{(0)}$:

$$\boldsymbol{\eta}(1) = \boldsymbol{\eta}(2) = \dots = \boldsymbol{\eta}(N) \triangleq \boldsymbol{\eta}_0^{(0)}. \quad (2.102)$$

À l'inverse, pour un signal qui comporte un point de rupture à l'instant $n = k_1$, on a

$$\boldsymbol{\eta}(1) = \dots = \boldsymbol{\eta}(k_1) \triangleq \boldsymbol{\eta}_0 \neq \boldsymbol{\eta}(k_1 + 1) = \dots = \boldsymbol{\eta}(N) \triangleq \boldsymbol{\eta}_1, \quad (2.103)$$

où $\boldsymbol{\eta}_q$, $q = 0, 1$, indique la valeur du vecteur de paramètre $\boldsymbol{\eta}(\cdot)$ sur le $(q + 1)$ -ième segment. Ainsi, le problème à résoudre, dans un premier temps, est celui de la détection d'un point de rupture, qui consiste à prendre correctement une décision entre les deux hypothèses \mathcal{H}_0 et \mathcal{H}_1 suivantes :

$$\mathcal{H}_0 : x_n \sim p(x_n | x_{n-1}, \dots, x_1; \boldsymbol{\eta}_0^{(0)}) \quad \text{pour} \quad 1 \leq n \leq N \quad (2.104a)$$

\mathcal{H}_1 : il existe un point de rupture $k_1 \in \llbracket 1, N - 1 \rrbracket$ inconnu, tel que :

$$\begin{cases} x_n \sim p(x_n | x_{n-1}, \dots, x_1; \boldsymbol{\eta}_0) & \text{pour} \quad 1 \leq n \leq k_1, \\ x_n \sim p(x_n | x_{n-1}, \dots, x_1; \boldsymbol{\eta}_1) & \text{pour} \quad k_1 + 1 \leq n \leq N. \end{cases} \quad (2.104b)$$

Les premiers travaux se rapportant à la détection et la localisation d'un changement abrupt remontent à ceux de Shewhart [She31], s'appliquant au contrôle de qualité de systèmes de production. Ils ont été suivis par ceux de Page [Pag54a, Pag55], pour la détection d'un changement dans la moyenne d'observations indépendantes, distribuées selon une loi normale.

Notons que dans (2.104b), la position du point de rupture k_1 est bien inconnue. L'estimation de k_1 s'effectue en général une fois que le test de décision entre \mathcal{H}_0 et \mathcal{H}_1 a rejeté l'hypothèse \mathcal{H}_0 , et que l'on se place donc sous \mathcal{H}_1 . Toutefois, des méthodes existent qui permettent de traiter les problèmes de détection et d'estimation de façon simultanée [CG00].

Lorsque le signal est observé sur une durée particulièrement longue, on peut s'attendre à ce qu'il comporte plus d'un seul point de rupture. Sous l'hypothèse que le nombre total de changements abrupts compris dans le signal observé \mathbf{x} est $Q (> 1)$, et en notant les instants des ruptures comme k_1, \dots, k_Q , le modèle (2.104b) devient

$$\begin{cases} x_n \sim p(x_n | x_{n-1}, \dots, x_1; \boldsymbol{\eta}_0) & \text{pour} \quad 1 \leq n \leq k_1 \\ x_n \sim p(x_n | x_{n-1}, \dots, x_1; \boldsymbol{\eta}_1) & \text{pour} \quad k_1 + 1 \leq n \leq k_2 \\ \vdots & \vdots \\ x_n \sim p(x_n | x_{n-1}, \dots, x_1; \boldsymbol{\eta}_Q) & \text{pour} \quad k_Q + 1 \leq n \leq N. \end{cases} \quad (2.105)$$

Dans le cas le plus général, Q , $\boldsymbol{\eta}_0, \boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_Q$ sont également des paramètres inconnus, en plus bien sûr de k_1, k_2, \dots, k_Q . Le problème posé ici est donc celui de la segmentation du signal \mathbf{x} en

17. Nous nous plaçons d'abord dans le cas général où les observations ne sont pas indépendantes.

18. L'exposant (0) sert ici à signifier qu'il s'agit de la valeur du vecteur de paramètres $\boldsymbol{\eta}_0$ lorsqu'il n'y a aucun (zéro) point de rupture dans le signal observé.

Q segments délimités par les points de rupture k_{q-1} et k_q , pour $q = 1, \dots, Q + 1$, en définissant $k_0 \triangleq 0$ et $k_{Q+1} \triangleq N$.

La détermination du nombre Q de points de rupture est un problème en soi, pour lequel diverses méthodes ont été proposées dans la littérature (voir section 2.2.3). On comprend donc que l'analyse de signaux comportant de multiples changements est bien plus complexe que lorsqu'on en considère au plus un seul, tant du point de vue méthodologique que calculatoire. Dès lors, de nombreux travaux se sont davantage intéressés au problème de détection/estimation d'un seul changement. Dans la section suivante, nous passons en revue quelques résultats de la littérature pour ce problème. Nous évoquerons ensuite, dans la section 2.2.3, des méthodes permettant de traiter le cas de plusieurs ruptures.

2.2.2 Cas d'un seul point de rupture

Dans cette section, nous nous intéressons au problème (2.104) (cas d'un seul point de rupture). Nous focaliserons notre attention sur les résultats concernant l'estimation de la position du point de rupture k_1 , lorsque l'on se place sous l'hypothèse \mathcal{H}_1 (2.104b). En tout premier lieu, comme on considère un signal \mathbf{x} échantillonné, il est important de noter que le paramètre k_1 à estimer est par nature discret. L'estimation de paramètres discrets est un problème particulier qui, de façon assez surprenante, a fait l'objet d'un nombre de travaux assez restreint. Cela peut se comprendre en notant qu'un outil aussi important que la borne de Cramér-Rao n'est pas défini pour un tel type de paramètre, puisque celle-ci fait intervenir la dérivée de la log-vraisemblance par rapport aux paramètres. Néanmoins, une étude assez récente [CS12] permet d'avoir une vue d'ensemble des résultats existants pour des modèles à paramètres discrets. Cette étude fournit un certain nombre de résultats concernant le comportement d'estimateurs de type maximum de vraisemblance (notamment l'estimateur ML lui-même) comme les conditions de leur consistance, l'expression de leur distribution asymptotique, ainsi que leur taux de convergence. Enfin, les auteurs proposent une borne inférieure sur le risque 0-1 de type Chapman-Robbins¹⁹, et montrent qu'aucun estimateur de paramètre discret n'atteint cette borne. Ces résultats ne sont pas directement transposables au cas de l'estimation d'un point de rupture, car le modèle de signaux avec ruptures ne vérifie pas les hypothèses requises dans [CS12]. On peut néanmoins s'attendre à observer ce type de comportement pour l'estimateur ML du point de rupture k_1 , en particulier le fait de ne pas atteindre de borne inférieure, même par rapport à un autre risque que le risque 0-1 comme le risque quadratique (l'EQM).

La nature singulière du paramètre k_1 (c'est un paramètre de changement de modèle), en tant que paramètre inconnu, nécessite donc une étude qui lui est propre. Après les travaux pionniers de Page en ce sens [Pag54a, Pag54b, Pag55, Pag57], les premiers à se focaliser sur le problème de l'estimation seule d'un point de rupture (c'est-à-dire sous l'hypothèse \mathcal{H}_1 (2.104b) exclusivement) sont ceux de Hinkley [Hin70, Hin72], pour des données indépendantes. Il s'intéresse au comportement de l'estimateur du maximum de vraisemblance de la localisation du point de rupture k_1 , noté \hat{k}_1^{ML} . Sa distribution à nombre d'observations N fini, i.e., $\Pr(\hat{k}_1^{\text{ML}}(\mathbf{x}) = k)$, n'étant pas tractable, il se place dans un cadre asymptotique consiste à considérer un nombre d'observations avant et après le point de rupture k_1 qui tend vers l'infini, c'est-à-dire

$$\begin{aligned} k_1(N) &\xrightarrow{N \rightarrow +\infty} +\infty \\ N - k_1(N) &\xrightarrow{N \rightarrow +\infty} +\infty \end{aligned} \tag{2.106}$$

en introduisant la dépendance de k_1 en N . Il ressort de ces premières analyses que la distribution

19. Nous avons vu dans la section 2.1.5.2.2, en la formulant par rapport au risque quadratique (i.e., l'EQM), que cette borne est bien définie pour un paramètre de nature discrète; il n'est donc pas surprenant de la retrouver ici.

de l'estimateur ML, même asymptotique (pourtant plus facile à obtenir, en théorie), n'admet pas de forme explicite, mais s'exprime sous la forme compliquée de la solution d'une équation intégrale [Hin70]. Par ailleurs, Hinkley remarque aussi que l'estimateur ML de k_1 n'est pas un estimateur consistant. En appliquant ses résultats au cas de données gaussiennes [Hin70] et suivant une loi de Bernoulli [HH70], il parvient à obtenir des approximations de la distribution asymptotique de \hat{k}_1^{ML} . Il observe que ces distributions asymptotiques ne fournissent pas toujours des approximations précises du comportement de l'estimateur \hat{k}_1^{ML} à N fini, et exprime donc le besoin de trouver davantage d'informations sur la distribution de cet estimateur pour des échantillons de taille finie.

Un autre résultat important obtenu par Hinkley concerne le fait de supposer les vecteurs $\boldsymbol{\eta}_0$ et $\boldsymbol{\eta}_1$ connus ou inconnus. Il prouve que, en supposant $\boldsymbol{\eta}_0$ et $\boldsymbol{\eta}_1$ inconnus, et sous réserve que leurs estimateurs ML soient consistants (conditionnellement à la valeur de k_1), la distribution asymptotique de \hat{k}_1^{ML} est inchangée par rapport au cas où $\boldsymbol{\eta}_0$ et $\boldsymbol{\eta}_1$ sont supposés connus [Hin72]. Notons toutefois que cette équivalence de comportement n'est pas garantie en régime non asymptotique.

Dans la lignée de ces travaux, Yao [Yao87] considère un régime asymptotique correspondant à (2.106), en rajoutant également le comportement limite $\lim_{N \rightarrow +\infty} \boldsymbol{\eta}_0(N) = \lim_{N \rightarrow +\infty} \boldsymbol{\eta}_1(N) \triangleq \boldsymbol{\eta}_\infty$, autrement dit

$$\boldsymbol{\delta}(N) \triangleq \boldsymbol{\eta}_1(N) - \boldsymbol{\eta}_0(N) \xrightarrow{N \rightarrow +\infty} \mathbf{0}. \quad (2.107)$$

Sous quelques conditions de régularité supplémentaires (voir [Yao87]), il obtient

$$\boldsymbol{\delta}^\top(N) \mathbf{F}(\boldsymbol{\eta}_\infty^*) \boldsymbol{\delta}(N) (\hat{k}_{1(\infty)}^{\text{ML}} - k_1^*) \xrightarrow{\mathcal{L}} Y \quad (2.108)$$

où $\boldsymbol{\eta}_\infty^*$ et k_1^* représentent les vraies valeurs des paramètres $\boldsymbol{\eta}_\infty$ et k_1 respectivement, $\hat{k}_{1(\infty)}^{\text{ML}}$ représente l'estimateur ML de k_1 dans le régime asymptotique précédemment décrit, et Y représente une variable aléatoire dont la fonction de répartition $F_Y(y) \triangleq \Pr(Y \leq y)$ est donnée par $F_Y(y) = 1 - F_Y(-y)$ et pour $y > 0$,

$$F_Y(y) = 1 + \sqrt{\frac{y}{2\pi}} \exp\left\{-\frac{y}{8}\right\} - \frac{y+5}{2} \Phi\left(-\frac{\sqrt{y}}{2}\right) + \frac{3}{2} \exp\{y\} \Phi\left(-\frac{3\sqrt{y}}{2}\right), \quad (2.109)$$

avec $\Phi(\cdot)$ la fonction de répartition de la loi normale centrée réduite, i.e., pour $y \in \mathbb{R}$,

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y \exp\left\{-\frac{t^2}{2}\right\} dt. \quad (2.110)$$

Par ailleurs, il obtient également une borne supérieure de la quantité $\Pr(\hat{k}_{1(\infty)}^{\text{ML}} \neq \hat{k}_{1(N)}^{\text{ML}})$ en $O(N^{-1/2})$, pour un $\|\boldsymbol{\delta}(N)\|$ suffisamment faible, où $\hat{k}_{1(N)}^{\text{ML}}$ représente l'estimateur ML de k_1 pour un échantillon de taille N finie. Ce type de résultat permet en particulier de savoir dans quelles conditions l'approximation du comportement de l'estimateur $\hat{k}_{1(N)}^{\text{ML}}$ à N fini par sa version asymptotique $\hat{k}_{1(\infty)}^{\text{ML}}$ est valable. Tous ces résultats s'avèrent valides seulement dans le cas d'observations gaussiennes changeant de moyenne, et pour des vecteurs $\boldsymbol{\eta}_0$ et $\boldsymbol{\eta}_1$ connus.

Plus récemment, Fotopoulos, Jandhyala et leurs collaborateurs ont également obtenu des résultats importants concernant la distribution asymptotique de l'estimateur ML de k_1 , en supposant (2.106) mais pas (2.107). En premier lieu, ils obtiennent des bornes inférieures et supérieures de cette distribution, c'est-à-dire de la quantité $\Pr(\hat{k}_{1(\infty)}^{\text{ML}} - k_1^* = i)$, pour $i \in \mathbb{Z}$. Ils obtiennent également un taux de convergence exponentiel de la distribution de $\hat{k}_{1(N)}^{\text{ML}}$ vers celle de $\hat{k}_{1(\infty)}^{\text{ML}}$: ils considèrent la distance uniforme entre ces deux distributions, i.e., la quantité

$$\sup_{\mathcal{A} \in \mathfrak{B}} |\Pr(\hat{k}_{1(N)}^{\text{ML}} - k_1^* \in \mathcal{A}) - \Pr(\hat{k}_{1(\infty)}^{\text{ML}} - k_1^* \in \mathcal{A})|$$

où \mathfrak{B} est la tribu de Borel sur \mathbb{R} , et démontrent que celle-ci peut être approximée par une fonction en $O(N^{-3/2})$ [JF99]. Leurs résultats leur permettent par la suite d'obtenir des expressions analytiques exactes et exploitables de la distribution asymptotique de \hat{k}_1^{ML} dans le cas d'un changement de paramètre λ dans des données distribuées selon une loi exponentielle (i.e., $p(x; \lambda) = \lambda e^{-\lambda x}$) [FJ01] et d'un changement de moyenne dans des données distribuées selon une loi normale [FJK10] – ce dernier cas est étendu par Liu dans sa thèse à des données multivariées et un changement de matrice de covariance [Liu10].

Une approche connexe à celle que l'on vient de décrire consiste à estimer non pas la localisation du point de rupture k_1 proprement dite, mais plutôt la fraction de la taille des données $\frac{k_1}{N}$. Remarquons qu'on n'estime donc plus dans ce cas un paramètre discret. Une différence avec le cas précédent est qu'il est possible de démontrer, sous certaines conditions [DP86], la consistance de l'estimateur ML de $\tau \triangleq \frac{k_1}{N}$. Le calcul de sa distribution asymptotique, dans les conditions (2.106) et (2.107), est également plus aisé que dans le cas précédent : Deshayes et Picard ont montré que [DP84] (voir également [CH97])

$$\psi(N) (\hat{\tau}_{(\infty)}^{\text{ML}} - \tau^*) \xrightarrow{\mathcal{L}} \xi \quad (2.111)$$

où

$$\psi(N) \triangleq \frac{k_1(N)(N - k_1(N))}{N} \boldsymbol{\delta}^\top(N) \mathbf{F}(\boldsymbol{\eta}_\infty^*) \boldsymbol{\delta}(N), \quad (2.112)$$

et ξ est une variable aléatoire réelle de densité

$$p_\xi(\xi) = \frac{3}{4} \exp\{|\xi|\} \operatorname{erfc}\left\{\frac{3}{2}\sqrt{|\xi|}\right\} - \frac{1}{4} \operatorname{erfc}\left\{\frac{1}{2}\sqrt{\frac{|\xi|}{2}}\right\}, \quad (2.113)$$

avec, pour $\xi \in \mathbb{R}$,

$$\operatorname{erfc}\{\xi\} \triangleq \frac{2}{\sqrt{\pi}} \int_\xi^{+\infty} \exp\{-u^2\} du = 2(1 - \Phi(\xi\sqrt{2})). \quad (2.114)$$

Bien qu'ils portent sur des quantités différentes, on pourra tout de même relever la proximité du résultat (2.111) obtenu par Deshayes et Picard [DP84] avec celui obtenu quelques années plus tard par Yao, énoncé en (2.108) [Yao87]. Un lien plus approfondi entre ces deux résultats n'a pour autant pas été établi dans la littérature à notre connaissance.

L'avantage de cette dernière approche est qu'elle ne nécessite pas de traitement particulier pour chaque famille de distribution des observations, comme c'est le cas de celle proposée par Hinkley. Elle peut même s'étendre à des changements plus complexes, dits « non additifs », comme des changements de matrice de covariance dans des données gaussiennes multivariées, ou bien des changements dans la moyenne et la covariance de processus auto-régressifs [Pic85].

Dans leur monographie [CG00], Chen et Gupta, en plus de reprendre les principaux résultats pour la résolution du problème de détection (2.104), proposent une approche pour traiter les problèmes de détection et d'estimation de manière conjointe, en se basant sur divers critères d'information, tels que celui d'Akaike (noté AIC) [Aka74] ou de Schwarz (noté SIC, également appelé « critère d'information bayésien » et donc également noté BIC) [Sch78], ainsi que diverses variantes qui en sont inspirées.

Nous n'avons traité ici que de travaux employant le point de vue fréquentiste. Pourtant, de nombreuses études ont été menées dans un cadre bayésien, et ce très tôt, pour le problème de détection des changements abrupts [Shy61], et dans un contexte on-line [CZ64]. La démarche bayésienne pour l'inférence sur la présence et la position d'une éventuelle rupture dans le cadre off-line a été pleinement décrite par Smith [Smi75]. Mais ce n'est que plus tard, avec le développement des méthodes de Monte-Carlo par chaînes de Markov (abrégées par l'acronyme anglais « MCMC »), qu'aura lieu un réel essor des méthodes bayésiennes pour la localisation de ruptures, en particulier pour traiter le cas de ruptures multiples, comme nous allons le voir dans la prochaine section.

2.2.3 Points de rupture multiples

Nous avons mentionné plus haut la difficulté particulière de traiter des données comportant plusieurs ruptures, particulièrement en raison de la nécessité d'estimer le nombre de ruptures Q correctement. Il existe bien sûr des applications où une erreur sur Q s'avère moins cruciale que dans d'autre : en segmentation de signaux audio par exemple, il n'est en général pas plus coûteux de surestimer le nombre de segments (si ce n'est que les traitements de chacun des segments ainsi relevés seront plus nombreux).

Le problème de la détermination de Q peut en fait être vu davantage comme un problème de détection : la résolution du test d'hypothèses (2.104) revient à décider $Q = 0$ ou $Q = 1$. Une procédure pour traiter le cas de points de ruptures multiples consisterait donc en l'application successive d'un test de détection sur chaque sous-segment mis en évidence par ce test :

- i) On effectue le test de détection d'une rupture sur le signal complet $\mathbf{x} = [x_1, \dots, x_N]$: si aucune rupture n'est détectée, alors la procédure se termine, sinon il existe un point de rupture et on note sa localisation \hat{k}_1 .
- ii) On effectue le test de détection d'une rupture séparément sur les deux sous-segments induits par l'étape précédente, c'est-à-dire $\mathbf{x}_{1:\hat{k}_1} \triangleq [x_1, \dots, x_{\hat{k}_1}]$ d'une part, et $\mathbf{x}_{\hat{k}_1+1:N} \triangleq [x_{\hat{k}_1+1}, \dots, x_N]$ d'autre part.
- iii) On répète ainsi de suite la procédure sur chaque sous-segment ainsi induit jusqu'à ce qu'on ne détecte plus de rupture sur aucun d'entre eux.
- iv) On obtient donc finalement l'ensemble des points de rupture trouvé aux étapes précédentes $\hat{k}_1, \hat{k}_2, \dots, \hat{k}_Q$.

Cette procédure, communément appelée « procédure de segmentation binaire », a été proposée par Vostrikova [Vos81], qui a démontré sa consistance. Malgré ce résultat, Hawkins a montré que cette procédure est seulement sous-optimale [Haw01].

Une méthode pour estimer le nombre de changements Q a été proposée par Yao, fondée sur l'utilisation du critère d'information de Schwarz (SIC), rencontré plus haut [Yao88]. Celui-ci se définit comme

$$\text{SIC}(Q) \triangleq -2 \ln p(\mathbf{x}; \hat{\boldsymbol{\theta}}_{|Q}) + L(Q - 1) \ln N \quad (2.115)$$

où $p(\mathbf{x}; \hat{\boldsymbol{\theta}}_{|Q})$ désigne la vraisemblance des observations, avec $\boldsymbol{\theta}_{|Q} \triangleq [k_1, \dots, k_Q, \boldsymbol{\eta}_0^T, \dots, \boldsymbol{\eta}_Q^T]^T$ le vecteur des paramètres inconnus et $\hat{\boldsymbol{\theta}}_{|Q}$ son estimée, lorsqu'on suppose la présence de Q ruptures (k_1, \dots, k_Q) , et où L désigne la taille des vecteurs $\boldsymbol{\eta}_q$, $q = 0, \dots, Q$. On voit que l'emploi d'une telle méthode nécessite l'estimation préalable de $\boldsymbol{\theta}_{|Q}$, pour chaque valeur de Q testée. Il peut donc se solder par un coût calculatoire assez élevé, même si des méthodes de programmation dynamique existent pour le contrôler.

Avec l'amélioration de la puissance de calcul des ordinateurs dans les années 1990 et 2000, les méthodes MCMC rendent le cadre bayésien beaucoup plus attractif, en permettant l'échantillonnage des lois *a posteriori* du vecteur de paramètres inconnus. Carlin et al. [CGS92] proposent l'utilisation d'un échantillonneur de Gibbs pour un modèle bayésien hiérarchique. De façon analogue, Barry et Hartigan [BH93] utilisent un échantillonneur de Gibbs pour obtenir les estimées MMSE des paramètres inconnus de signaux comportant des ruptures. Une solution pour obtenir les estimées MAP est, à la même période, proposée par Djurić [Dju94]. Parmi les développements qui font suite à ces travaux, on peut mentionner l'idée de combiner l'échantillonneur de Gibbs avec l'algorithme de Metropolis-Hastings [Chi98], ainsi que celle d'introduire une nouvelle paramétrisation du problème : la détection et la localisation des changements peut s'effectuer en estimant un processus aléatoire $\mathbf{r} \triangleq [r_1, \dots, r_N]^T \in \mathbb{R}^N$ dont la composante r_n vaut 1 si n est un point de rupture et 0 sinon [Lav98]. Sans connaissance particulière sur la répartition des ruptures *a priori*,

ce processus peut être défini comme une suite de variables de Bernoulli indépendantes. Cette paramétrisation se révèle très commode, puisqu'elle fournit facilement une estimée du nombre de ruptures comme $\hat{Q} = \sum_{n=1}^N \hat{r}_n = \hat{\mathbf{r}}^T \mathbf{1}$, avec $\mathbf{1}$ le vecteur de taille N dont toutes les composantes valent 1. L'approche proposée par Fearnhead [Fea06] permet également d'obtenir une loi *a priori* jointe sur le nombre de changements et sur leur localisation.

Toutes ces avancées permettent alors de nombreuses applications en traitement du signal, en particulier bien sûr à la segmentation de signaux [TC97, LL01, PADF02, MF06]. Dobigeon et al. étendent ces méthodes à des modèles bayésiens hiérarchiques pour des données multivariées poissonniennes (notamment rencontrées en astronomie) [DTS07] ou des processus autorégressifs d'ordre donné (pour le traitement de la parole par exemple) [DTD07]. Une loi *a priori* qui contrôle la corrélation des localisations des ruptures, entre chaque série temporelle considérée, est introduite, ce qui permet un traitement conjoint de plusieurs séries de mesures issues de différents capteurs.

En résumé, nous avons vu les difficultés inhérentes à l'analyse des performances d'estimateurs d'un point de rupture, comme celui du maximum de vraisemblance. Bien que des résultats théoriques existent, ceux-ci sont généralement difficiles à obtenir puisque les conditions d'application des théorèmes classiques vus dans la section 2.1 ne sont pas vérifiées. De plus, ces résultats ne s'avèrent valables que dans un domaine asymptotique idéal, où le nombre d'observations avant et après la rupture tend vers l'infini, et où les paramètres des distributions sur chaque segment doivent souvent être connus. Et bien que des analyses aient montré que la distribution asymptotique de l'estimateur ML approxime parfois correctement son comportement à nombre d'observations fini, cela ne nous garantit pas qu'il s'agit du meilleur estimateur. Ainsi, quelles sont les performances ultimes d'estimation dans ce contexte particulier ? Nous avons également entrevu que, dans le contexte de points de rupture multiples, les récentes avancées en matière de puissance calculatoire ont permis la mise en place de nombreuses méthodes, dont beaucoup emploient le point de vue bayésien. Dès lors, comment comparer ces méthodes ? Et quelles en sont les meilleures performances possibles ? Nous l'avons vu, le calcul de bornes inférieures de l'EQM peut en général permettre de répondre à ces questions.

2.3 Bornes inférieures de l'EQM (déterministes) pour l'estimation de points de ruptures

2.3.1 Bornes de type Cramér-Rao

Nous avons vu dans la section 2.1 que la borne de Cramér-Rao est un outil important pour l'analyse de performances d'estimation. Or, comme mentionné dans la section 2.2.2, il n'est pas possible de la calculer à proprement parler pour la localisation du point de rupture k_1 . En effet, le caractère discret de k_1 implique que les conditions de régularité requises pour le calcul de la BCR ne sont pas vérifiées.

Pour contourner ces restrictions, plusieurs approches ont été proposées dans la littérature. Tout d'abord, un signal continu (non échantillonné), et donc une paramétrisation continue de la localisation du changement, combinée à l'approximation de la rupture par une fonction régulière (par exemple, une fonction de Heaviside filtrée par une gaussienne), ont conduit à une première approximation de la BCR pour modéliser des contours sur des images [KH92]. De façon similaire, une autre BCR approchée a pu être obtenue en approximant la rupture par la sortie d'un filtrage à bande passante finie d'une fonction de Heaviside par une spline cubique [BM96]. Une autre façon d'approximer la rupture est d'utiliser une fonction dite sigmoïdale [RD96, SS98, SS99, TFS04],

c'est-à-dire en modélisant le signal avant échantillonnage comme

$$x(t) = g(t) s_\alpha(t, t_1^*) + v(t), \quad (2.116)$$

où t_1^* est la vraie localisation de la rupture, α est le paramètre de temps de montée, $g(t)$ et $v(t)$ sont des bruits multiplicatifs et additifs, respectivement, et $s_\alpha(\cdot, t_1^*)$ est la fonction modélisant la rupture, qui est définie par

$$s_\alpha(t, t_1^*) \triangleq \frac{1 + A + e^{-\alpha(t-t_1^*)}}{1 + e^{-\alpha(t-t_1^*)}}, \quad (2.117)$$

représentant ainsi un changement dans l'amplitude qui passe de 1 à $1 + A$. Le signal est ensuite échantillonné avec une période fixée T_s en posant $t = nT_s$. Reza et Doroodchi [RD96] ont calculé une BCR « modifiée » (dans le sens de [GRM98]) en prenant une loi *a priori* uniforme ou triangulaire pour la localisation de la rupture. La forme de la matrice d'information de Fisher obtenue a par la suite été simplifiée par Swami et Sadler [SS98], puis ceux-ci ont étendu ces résultats au cas de signaux noyés dans des bruits additifs et multiplicatifs [SS99]. Tourneret et al. ont notamment montré, en analysant la structure de la matrice d'information de Fisher pour le modèle (2.116), que l'estimation de la position du point de rupture est asymptotiquement (i.e., pour $N \rightarrow +\infty$) découplée de celle des autres paramètres du problème.

Les inconvénients des approches évoquées ici sont qu'il s'agit de résultats approchés, pour certains valables seulement pour des changements de moyennes [KH92, BM96, RD96], mais surtout, la BCR obtenue tend vers zéro lorsque l'on fait tendre le paramètre α vers l'infini, c'est-à-dire lorsque l'on fait tendre le signal observé vers un signal comportant une vraie rupture. Dans ce cas, la BCR n'apporte plus aucune information ; on parle de borne « triviale ». De plus, ces résultats ont été obtenus uniquement en considérant une unique rupture, et prive donc d'étudier l'influence de l'estimation de plusieurs ruptures successives. Enfin, ajoutons à cela le fait que la borne de Cramér-Rao n'est en général valable qu'en régime asymptotique, ce qui sensiblement réduit son domaine de validité.

Mentionnons enfin une dernière approche permettant d'obtenir des résultats de type « bornes de Cramér-Rao » : elle consiste à interpoler linéairement le signal discret pour obtenir un signal continu et ainsi s'affranchir du caractère discret de la localisation du changement [JPP14]. L'analyse des résultats obtenus ne semble pas assurer la conservation du caractère de borne inférieure, même s'ils semblent permettre d'approcher avec une bonne précision l'EQM asymptotique de l'estimateur ML.

Les limitations de ces approches conduisent à considérer des bornes de type Barankin, que nous avons introduites à la section 2.1.5.2.2.

2.3.2 Bornes de type Barankin

2.3.2.1 Borne de Chapman-Robbins pour la localisation d'un unique point de rupture

Dans cette section, on s'intéresse au cas d'un unique point de rupture ; on considère donc le modèle (2.104b). Nous résumons le résultat général obtenu par Ferrari et Tourneret [FT03] pour l'expression de la borne de (Hammersley-)Chapman-Robbins pour l'estimation du point de rupture k_1 en

$$\text{BHamChR}(k_1^*) = \max\{B^-, B^+\}, \quad (2.118)$$

où

$$B^+ \triangleq \sup_{1 \leq h \leq N - k_1^* - 1} \frac{h^2}{\varrho_{0,1}^h - 1} \quad \text{et} \quad B^- \triangleq \sup_{1 \leq h \leq k_1^* - 1} \frac{h^2}{\varrho_{1,0}^h - 1} \quad (2.119)$$

avec, pour $(p, q) \in \{(0, 1), (1, 0)\}$,

$$\varrho_{p,q} \triangleq \int_{\Omega'} \frac{p^2(x; \boldsymbol{\eta}_p)}{p(x; \boldsymbol{\eta}_q)} dx, \quad (2.120)$$

où Ω' désigne l'espace des observations en un seul instant, c'est-à-dire qu'on a $\Omega = (\Omega')^N = \Omega' \times \dots \times \Omega'$. Il est important de noter que l'intégrale dans la définition de $\varrho_{p,q}$ (2.120) est une intégrale simple, et se calcule aisément pour la plupart des distributions usuelles. Nous renvoyons le lecteur intéressé à l'article original [FT03] pour son expression dans les cas de données gaussiennes (changeant de moyenne et/ou de variance), exponentielles ou poissonniennes.

2.3.2.2 Borne de type Barankin pour la localisation de plusieurs points de rupture

Cette section résume les résultats de La Rosa et al. [LRN07,LRNM10], étendant les résultats de la section précédente au cas de points de rupture multiples. Le modèle considéré est donc (2.105), avec le nombre de ruptures Q et les paramètres $\boldsymbol{\eta}_0, \dots, \boldsymbol{\eta}_Q$ supposés connus. Les paramètres à estimer sont donc seulement les points de rupture k_1, \dots, k_Q rassemblés dans le vecteur \mathbf{k} , de vraie valeur notée \mathbf{k}^* .

La borne calculée est donc celle présentée en (2.63). Il est démontré que le choix d'un nombre de points tests K égal à Q permet d'obtenir un bon compromis entre précision de la borne de Barankin et un coût calculatoire de celle-ci raisonnable. On choisit donc $K = Q$, et on considère une matrice \mathbf{H} des points tests diagonale, i.e., $\mathbf{H} = \text{diag}(h_1, \dots, h_Q)$. La matrice d'information de Barankin (voir section 2.1.5.2.2) obtenue est tridiagonale, c'est-à-dire qu'elle s'écrit, en mettant en avant sa dépendance en \mathbf{k}^* et en \mathbf{H} ,

$$\mathbf{BIM}(\mathbf{k}^*, \mathbf{H}) = \begin{bmatrix} A_1 & B_1 & 0 & \dots & 0 \\ B_1 & A_2 & B_2 & \ddots & \vdots \\ 0 & B_2 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & B_{Q-1} \\ 0 & \dots & 0 & B_{Q-1} & A_Q \end{bmatrix}, \quad (2.121)$$

où, pour $1 \leq q \leq Q$,

$$A_q \triangleq \begin{cases} \varrho_{q-1,q}^{h_q} - 1 & \text{si } h_q > 0, \\ \varrho_{q,q-1}^{-h_q} - 1 & \text{si } h_q < 0, \end{cases} \quad (2.122)$$

avec, pour $(p, q) \in \llbracket 0, Q \rrbracket$, $\varrho_{p,q}$ donné par (2.120), et, pour $1 \leq q \leq Q - 1$

$$B_q \triangleq \begin{cases} 0 & \text{si } \widetilde{\Delta k}_q^* < 0, \\ \left(\int_{\Omega'} \frac{p(x; \boldsymbol{\eta}_{q-1})p(x; \boldsymbol{\eta}_{q+1})}{p(x; \boldsymbol{\eta}_q)} dx \right)^{\widetilde{\Delta k}_q^*} - 1 & \text{si } \widetilde{\Delta k}_q^* > 0, \end{cases} \quad (2.123)$$

avec $\widetilde{\Delta k}_q^* \triangleq (k_q^* + h_q) - (k_{q+1}^* + h_{q+1})$.

Rappelons que la borne de Barankin est donnée par (2.63) que l'on redonne :

$$\mathbf{BB}(\mathbf{k}^*) = \sup_{\mathbf{H} \in \mathcal{H}^K} \mathbf{H} [\mathbf{BIM}(\mathbf{k}^*, \mathbf{H})]^{-1} \mathbf{H}^\top. \quad (2.124)$$

Dans [LRNM10], La Rosa et al. trouvent même une expression analytique de l'inverse de la matrice d'information de Barankin $\mathbf{BIM}(\mathbf{k}^*, \mathbf{H})$ qui intervient dans (2.124), permettant ainsi de

réduire le coût calculatoire de la borne. De plus, une méthode pour calculer la borne supérieure de l'ensemble des matrices $\mathcal{B} \triangleq \{\mathbf{H}[\mathbf{BIM}(\mathbf{k}^*, \mathbf{H})]^{-1}\mathbf{H} \mid \mathbf{H} \in \mathcal{H}\}$ est discutée, sur laquelle nous reviendrons dans le Chapitre 3, section 3.3.

La borne est calculée pour des données gaussiennes changeant de moyenne, de variance, ainsi que pour des données poissonniennes changeant de paramètre, et elle est comparée au comportement de l'estimateur ML. La borne de Barankin est ici assez optimiste en ce qui concerne les performances ultimes en matière d'EQM.

Ces résultats laissent penser qu'il est possible d'obtenir des bornes plus précises concernant les performances ultimes d'estimation de ruptures. C'est pourquoi, dans cette thèse, nous nous sommes intéressés à la borne de Weiss-Weinstein et quelques unes de ses variantes, puisqu'elle est connue pour fournir dans le contexte bayésien un aperçu des plus précis quant aux meilleures performances d'estimation.

Chapitre 3

Nouvelles bornes inférieures de l'erreur quadratique moyenne pour des signaux comportant des points de rupture

Nous avons vu dans le chapitre précédent l'intérêt d'obtenir des bornes inférieures de l'erreur quadratique moyenne en estimation paramétrique. Dans le cas de signaux comportant des points de rupture, cet intérêt est même encore plus grand, en raison de la grande difficulté rencontrée pour évaluer les performances d'estimation de la position des ruptures de manière satisfaisante. Nous avons notamment évoqué, pour ce problème, les limitations de la borne de Cramér-Rao (voir Chapitre 2, section 2.3.1), et le caractère optimiste de la borne de Barankin, dans l'aperçu du comportement de l'estimation qu'elle fournit (Chapitre 2, section 2.3.2).

Un des points rappelés dans le chapitre précédent est également les nuances qui existent entre les cadres d'estimation fréquentiste (non bayésien) et bayésien. En ce qui concerne l'évaluation des performances d'estimation, il ressort que dans le cadre fréquentiste, on s'intéresse davantage aux performances d'estimation *locales* – c'est-à-dire pour *une* valeur du paramètre inconnu θ^* *fixée* dans l'espace des paramètres Θ , tandis que dans le cadre bayésien, on s'intéresse plutôt aux performances d'estimation *globales* – c'est-à-dire considérant *toutes* les valeurs du paramètre inconnu θ possibles de l'espace des paramètres Θ . Il est frappant de constater que pour le problème de la localisation de ruptures, du point de vue des bornes inférieures de l'EQM, seules des bornes non bayésiennes (Cramér-Rao et Barankin), qui portent donc sur l'EQM locale, ont été étudiées. Pourtant, le point de vue bayésien avait été employé dans les tout premiers travaux sur l'analyse des signaux avec ruptures [Shy61, CZ64]. Nous proposons donc, dans ce chapitre, de s'intéresser à l'une des bornes bayésiennes (qui porte donc sur l'EQM globale) connues pour être l'une des plus précises, à savoir la borne de Weiss-Weinstein [WW88, VB07, RFL⁺08, TT10b]. Celle-ci a été présentée dans le Chapitre 2, section 2.1.5.3.2, où nous avons en particulier noté le peu de conditions de régularité qu'elle nécessite. Ces qualités en font une borne tout indiquée pour son application à l'estimation des paramètres de signaux comportant des changements abrupts.

Ce chapitre est donc consacré à l'obtention d'expressions générales de la borne de Weiss-Weinstein pour l'estimation des paramètres de signaux modélisés selon (2.105), en supposant les observations indépendantes, et le nombre de changements connu. Nous focaliserons notre attention sur les positions des ruptures elles-mêmes, notées k_1, \dots, k_Q . Suivant les hypothèses faites sur la connaissance et la nature des autres paramètres du signal, i.e., $\boldsymbol{\eta}_0, \dots, \boldsymbol{\eta}_Q$, nous considérons successivement :

- **en section 3.1** : la borne de Weiss-Weinstein (BWW) pour l'estimation de $\mathbf{k} \triangleq [k_1, \dots, k_Q]^T$,

avec les paramètres $\boldsymbol{\eta}_0, \dots, \boldsymbol{\eta}_Q$ supposés connus ;

- **en section 3.2.3** : la borne bayésienne combinée de Cramér-Rao/Weiss-Weinstein (BBC-CRWW, présentée au Chapitre 2, section 2.1.5.4.1), pour l'estimation de $\boldsymbol{\theta} \triangleq [\boldsymbol{\eta}^\top, \mathbf{k}^\top]^\top$, où $\boldsymbol{\eta} \triangleq [\boldsymbol{\eta}_0^\top, \dots, \boldsymbol{\eta}_Q^\top]^\top$ est supposé aléatoire ;
- **en section 3.2.2** : la borne de Cramér-Rao/Weiss-Weinstein hybride (BCRWWH, présentée au Chapitre 2, section 2.1.5.4.2), pour l'estimation de $\boldsymbol{\theta} \triangleq [\boldsymbol{\eta}^\top, \mathbf{k}^\top]^\top$, où $\boldsymbol{\eta}$ est supposé déterministe.

On rappelle que le nombre de ruptures Q est toujours supposé connu. En effet, nous avons vu à la section 2.2.3 que l'estimation de Q est un problème en soi, qui correspond davantage à un problème de détection ou d'estimation d'ordre de modèle. Les bornes calculées ici sont donc des bornes conditionnées au nombre Q de rupture, où Q a été estimé préalablement, par exemple en utilisant le critère d'information de Schwarz [Yao88]. Cette hypothèse de nombre de ruptures connu est comparable à celle habituellement faite en estimation de directions d'arrivées pour le traitement d'antenne [WK85], où les performances d'estimation sont évaluées pour un nombre de sources fixé, supposé connu [SN89, SN90].

3.1 Borne de Weiss-Weinstein pour l'estimation des points de rupture

Dans cette section, on cherche à obtenir la borne de Weiss-Weinstein pour l'estimation de points de ruptures k_1, \dots, k_Q , regroupés dans le vecteur \mathbf{k} .

3.1.1 Modèle multivarié

Dans cette section, on considère un signal modélisé par (2.105), dans le cas (légèrement) plus général où celui-ci est multivarié, c'est-à-dire qu'on le note $\mathbf{X} \triangleq [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{M \times N}$, où N représente toujours le nombre d'observations, et M la taille de chaque vecteur d'observation \mathbf{x}_n , qui peut correspondre au nombre de capteurs dans un contexte de traitement d'antennes par exemple. Nous supposons que le signal \mathbf{X} est temporellement indépendant, et que celui-ci comprend un nombre total de changements abrupts Q supposé connu. Ces changements surviennent à des instants notés k_1, \dots, k_Q supposés inconnus. La distribution d'un vecteur d'observations \mathbf{x}_n sur le q -ième segment $[[k_{q-1} + 1, k_q]]$, $q = 1, \dots, Q + 1$ (en définissant $k_0 \triangleq 0$ et $k_{Q+1} \triangleq N$) est paramétrée par un vecteur $\boldsymbol{\eta}_{q-1} = [\eta_{q-1,1}, \eta_{q-1,2}, \dots, \eta_{q-1,L}]^\top \in \mathbb{R}^L$ supposé connu (L représente le nombre de paramètre de la distribution) ; par conséquent elle est notée¹ $p_{\boldsymbol{\eta}_{q-1}}(\mathbf{x}_n)$.

On réécrit donc le modèle (2.105) comme

$$\begin{cases} \mathbf{x}_n \sim p_{\boldsymbol{\eta}_0}(\mathbf{x}_n) & \text{pour } 1 \leq n \leq k_1 \\ \mathbf{x}_n \sim p_{\boldsymbol{\eta}_1}(\mathbf{x}_n) & \text{pour } k_1 + 1 \leq n \leq k_2 \\ \vdots & \vdots \\ \mathbf{x}_n \sim p_{\boldsymbol{\eta}_Q}(\mathbf{x}_n) & \text{pour } k_Q + 1 \leq n \leq N. \end{cases} \quad (3.1)$$

Le vecteur de paramètres à estimer dans ce modèle est donc $\mathbf{k} \triangleq [k_1, \dots, k_Q]$, et celui-ci est estimé dans un cadre bayésien. Ainsi, les k_q , $q = 1, \dots, Q$, sont des variables aléatoires dont la loi *a priori* est supposée connue, qui dépend en général de l'application étudiée. Différentes lois *a priori* peuvent être attribuées au vecteur \mathbf{k} (voir par exemple [Fea06] pour quelques exemples).

1. La notation du vecteur $\boldsymbol{\eta}_{q-1}$ en indice de la fonction densité de probabilité, et non plus en argument de celle-ci (comme c'est le cas dans (2.105)), sert à indiquer que $\boldsymbol{\eta}_{q-1}$ est connu, et qu'il ne fait donc pas partie, pour le moment, du vecteur de paramètres inconnus (à estimer) $\boldsymbol{\theta}$.

Dans notre cas, comme Q est supposé connu, nous choisissons une loi *a priori* qui est compatible avec cette hypothèse, et qui offre un bon compromis entre exploration de l'espace des paramètres $\Theta \triangleq \{k_1, \dots, k_Q \in \llbracket 1, N \rrbracket \mid 0 < k_1 < \dots < k_Q < N\}$ et complexité des calculs mis en jeu. Nous considérons donc, dans un premier temps, la marche aléatoire uniforme qui définit la position *a priori* des points de rupture, pour $q = 1, \dots, Q$, comme

$$k_q = k_{q-1} + l_q, \quad (3.2)$$

où l_q est une variable aléatoire suivant une loi uniforme discrète $\mathcal{U}(\llbracket d, D \rrbracket)$, avec d et D des nombres entiers devant vérifier $1 \leq d < D \leq \lfloor N/(Q+1) \rfloor$, où $\lfloor \cdot \rfloor$ représente la fonction partie entière. En notant $\Delta \triangleq D - d + 1$, la distribution *a priori* jointe du le vecteur \mathbf{k} s'écrit donc

$$\pi_{\mathbf{d}}(\mathbf{k}) = \frac{1}{\Delta^Q} \prod_{q=1}^Q \mathbb{1}_{\mathcal{K}_{\mathbf{d}}^{(q)}}(k_q) = \frac{1}{\Delta^Q} \mathbb{1}_{\mathcal{K}_{\mathbf{d}}}(\mathbf{k}) \quad (3.3)$$

où $\mathbf{d} \triangleq [d, D]^\top$ et $\mathcal{K}_{\mathbf{d}}^{(q)} \triangleq \llbracket k_{q-1} + d, k_{q-1} + D \rrbracket$. Le support de cette distribution *a priori* de \mathbf{k} est noté $\mathcal{K}_{\mathbf{d}} \triangleq \mathcal{K}_{\mathbf{d}}^{(1)} \times \mathcal{K}_{\mathbf{d}}^{(2)} \times \dots \times \mathcal{K}_{\mathbf{d}}^{(Q)} \subset \Theta$. Il correspond à l'ensemble des segmentations possibles de la fenêtre d'observation $\llbracket 1, N \rrbracket$ autorisées par la loi *a priori*², c'est-à-dire les segmentations contenant exactement $Q + 1$ segments, chacun de longueur maximale Δ .

Dans un second temps, nous étudierons la borne de Weiss-Weinstein (BWW) pour le problème (3.1) avec d'autres distributions *a priori* pour \mathbf{k} (voir la section 3.1.4 de ce chapitre). Pour l'instant, nous nous en tenons à la distribution (3.3) pour simplifier l'exposé du calcul de la BWW, dans le cas général $\mathbf{s} \in]0, 1[^Q$ quelconque (où \mathbf{s} est celui de l'équation (2.78)).

La vraisemblance des observations s'écrit, d'après le modèle (3.1),

$$p(\mathbf{X} | \mathbf{k}) = \prod_{n=1}^{k_1} p_{\eta_0}(\mathbf{x}_n) \prod_{n=k_1+1}^{k_2} p_{\eta_1}(\mathbf{x}_n) \dots \prod_{n=k_{Q-1}+1}^N p_{\eta_Q}(\mathbf{x}_n) = \prod_{q=0}^Q \prod_{n=k_q+1}^{k_{q+1}} p_{\eta_q}(\mathbf{x}_n). \quad (3.4)$$

En la combinant avec (3.3), on obtient la distribution jointe des observations \mathbf{X} et du vecteur de paramètres \mathbf{k} par

$$p(\mathbf{X}, \mathbf{k}) = p(\mathbf{X} | \mathbf{k}) \pi(\mathbf{k}). \quad (3.5)$$

3.1.2 Méthode pour le calcul de la borne

Nous exposons ici la méthode utilisée pour obtenir les expressions des éléments de la matrice \mathbf{G} de la BWW (voir (2.83)), pour le modèle (3.1) et la distribution *a priori* (3.3). Comme les éléments de la matrice \mathbf{G} sont donnés par (2.80), il suffit de trouver les expressions de $e^{\tilde{\mu}(\alpha, \beta, \mathbf{h}_a, \mathbf{h}_b)}$, pour $0 < \alpha, \beta < 1$, $(a, b) \in \llbracket 1, Q \rrbracket^2$ et $(\mathbf{h}_a, \mathbf{h}_b) \in \mathcal{H}^2$, à partir de (2.81b). En prenant en compte la nature discrète des paramètres à estimer, (2.81b) se réécrit

$$\begin{aligned} e^{\tilde{\mu}(\alpha, \beta, \mathbf{h}_a, \mathbf{h}_b)} &= \sum_{\mathbf{k} \in \Theta} \int_{\Omega} p^\alpha(\mathbf{X}, \mathbf{k} + \mathbf{h}_a) p^\beta(\mathbf{X}, \mathbf{k} + \mathbf{h}_b) p^{1-\alpha-\beta}(\mathbf{X}, \mathbf{k}) d\mathbf{X} \\ &= \sum_{\mathbf{k} \in \Theta} \nu(\alpha, \beta, \mathbf{h}_a, \mathbf{h}_b, \mathbf{k}) M(\alpha, \beta, \mathbf{h}_a, \mathbf{h}_b, \mathbf{k}) \end{aligned} \quad (3.6)$$

où l'on définit

$$\nu(\alpha, \beta, \mathbf{h}_a, \mathbf{h}_b, \mathbf{k}) \triangleq \pi_{\mathbf{d}}^\alpha(\mathbf{k} + \mathbf{h}_a) \pi_{\mathbf{d}}^\beta(\mathbf{k} + \mathbf{h}_b) \pi_{\mathbf{d}}^{1-\alpha-\beta}(\mathbf{k}) \quad (3.7)$$

2. On notera bien qu'il s'agit d'un ensemble de segmentations restreint par rapport à toutes les segmentations possibles de $\llbracket 1, N \rrbracket$ en $Q + 1$ segments exactement, qui n'est autre que $\Theta \triangleq \{\mathbf{k} = [k_1, \dots, k_Q]^\top \mid 0 < k_1 < k_2 < \dots < k_Q < N\}$.

d'une part, et

$$M(\alpha, \beta, \mathbf{h}_a, \mathbf{h}_b, \mathbf{k}) \triangleq \int_{\Omega} p^{\alpha}(\mathbf{X} | \mathbf{k} + \mathbf{h}_a) p^{\beta}(\mathbf{X} | \mathbf{k} + \mathbf{h}_b) p^{1-\alpha-\beta}(\mathbf{X} | \mathbf{k}) d\mathbf{X} \quad (3.8a)$$

$$= \mathbb{E}_{\mathbf{x}|\mathbf{k}} \left\{ \frac{p^{\alpha}(\mathbf{X} | \mathbf{k} + \mathbf{h}_a) p^{\beta}(\mathbf{X} | \mathbf{k} + \mathbf{h}_b)}{p^{\alpha+\beta}(\mathbf{X} | \mathbf{k})} \right\}, \quad (3.8b)$$

d'autre part.

Nous rappelons que, comme expliqué dans la partie 2.1.5.3.2, nous nous limitons à des matrices de points-test \mathbf{H} (voir équation (2.78)) diagonales. Ainsi, les vecteurs \mathbf{h}_a et \mathbf{h}_b sont de la forme $\mathbf{h}_a = [0, \dots, 0, h_a, 0, \dots, 0]^T$ dont tous les éléments sont nuls sauf le a -ième, et $\mathbf{h}_b = [0, \dots, 0, h_b, 0, \dots, 0]^T$, où h_b peut être nul, comme au dénominateur de (2.80). De ce fait, nous commettons parfois l'abus de notation de remplacer \mathbf{h}_a et \mathbf{h}_b par leur composante non nulle (h_a et h_b respectivement) dans l'argument de la fonction $\tilde{\mu}(\cdot)$, ainsi que partout où aucune confusion n'est possible.

Comme la matrice \mathbf{G} est symétrique, on se restreint aux termes du triangle supérieur (i.e., tels que $b \geq a$), et déduisons les éléments restants par symétrie. La méthodologie utilisée pour obtenir les expressions des éléments de \mathbf{G} est la suivante :

1. obtenir les expressions de $\nu(\alpha, \beta, h_a, h_b, \mathbf{k})$;
2. obtenir les expressions de $M(\alpha, \beta, h_a, h_b, \mathbf{k})$;
3. en déduire les expressions de $e^{\tilde{\mu}(\alpha, \beta, h_a, h_b)}$, par (3.6) ;
4. et finalement en déduire les expressions de $[\mathbf{G}]_{a,b}$, par (2.80).

Bien sûr, toutes ces expressions varient suivant les éléments $[\mathbf{G}]_{a,b}$ que l'on considère ; on distingue donc les trois cas suivants :

- i. **diag** : $b = a$ (correspondant aux termes de la diagonale de \mathbf{G})
- ii. **ts** : $b > a + 1$ (termes du triangle supérieur de \mathbf{G})
- iii. **psdiag** : $b = a + 1$ (termes de la première surdiagonale de \mathbf{G}).

Les calculs de $\nu(\alpha, \beta, h_a, h_b, \mathbf{k})$, $M(\alpha, \beta, h_a, h_b, \mathbf{k})$, $e^{\tilde{\mu}(\alpha, \beta, h_a, h_b)}$ et $[\mathbf{G}]_{a,b}$ dans chacun des cas mentionnés ci-dessus sont fournis dans l'Annexe A. Nous donnons maintenant le résultat principal de cette section, à savoir l'expression de la matrice \mathbf{G} .

3.1.3 Expression de la matrice $\mathbf{G}(\mathbf{H}, \mathbf{s})$

D'après les calculs fournis dans l'Annexe A, on trouve que la matrice \mathbf{G} est tridiagonale, c'est-à-dire qu'elle s'écrit

$$\mathbf{G}(\mathbf{H}, \mathbf{s}) = \begin{bmatrix} A'_1 & B'_1 & 0 & \cdots & 0 \\ B'_1 & A'_2 & B'_2 & \ddots & \vdots \\ 0 & B'_2 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & B'_{Q-1} \\ 0 & \cdots & 0 & B'_{Q-1} & A'_Q \end{bmatrix}, \quad (3.9)$$

avec les termes diagonaux donnés, pour $q \in \llbracket 1, Q \rrbracket$, par

$$A'_q \triangleq \frac{u(\Delta, h_q) \cdot \left(\rho_{q-1,q}^{|h_q|}(\varepsilon_{h_q}(2s_q)) + \rho_{q-1,q}^{|h_q|}(\varepsilon_{h_q}(2s_q - 1)) \right) - 2u(\Delta, 2h_q) \cdot \rho_{q-1,q}^{2|h_q|}(\varepsilon_{h_q}(s_q))}{u^2(\Delta, h_q) \cdot \rho_{q-1,q}^{2|h_q|}(\varepsilon_{h_q}(s_q))} \quad (3.10)$$

où l'on définit les fonctions intermédiaires suivantes :

- *i)* pour $\Delta \in \mathbb{N}$, $q \in \llbracket 1, Q \rrbracket$ et $h_q \in \mathbb{Z}$,

$$u(\Delta, h_q) \triangleq \begin{cases} \frac{(\Delta - |h_q|)^2}{\Delta^2} & \text{si } q < Q \text{ et } |h_q| < \Delta, \\ \frac{\Delta - |h_q|}{\Delta} & \text{si } q = Q \text{ et } |h_q| < \Delta, \\ 0 & \text{si } |h_q| \geq \Delta; \end{cases} \quad (3.11)$$

- *ii)* pour $q \in \llbracket 1, Q \rrbracket$ et $s \in]0, 1[$,

$$\rho_{q-1,q}(s) \triangleq \int_{\Omega'} p_{\boldsymbol{\eta}_{q-1}}^s(\mathbf{x}) p_{\boldsymbol{\eta}_q}^{1-s}(\mathbf{x}) d\mathbf{x}, \quad (3.12)$$

où Ω' désigne, comme dans (2.120), l'espace des observations en un seul instant, c'est-à-dire qu'on a $\Omega = (\Omega')^N = \Omega' \times \dots \times \Omega'$, et pour $q, r \in \llbracket 0, Q \rrbracket$, $\boldsymbol{\eta}_{q:r}$ représente la concaténation verticale des vecteurs $\boldsymbol{\eta}_q$ à $\boldsymbol{\eta}_r$, i.e., $\boldsymbol{\eta}_{q:r} \triangleq [\boldsymbol{\eta}_q; \boldsymbol{\eta}_{q+1}; \dots; \boldsymbol{\eta}_r]$;

- *iii)* pour $h \in \mathbb{Z}$ et $s \in]0, 1[$,

$$\varepsilon_h(s) \triangleq \begin{cases} s & \text{si } h > 0 \\ 1 - s & \text{si } h < 0. \end{cases} \quad (3.13)$$

D'autre part, les termes de la surdiagonale (égaux par symétrie à ceux de la sous-diagonale) sont donnés, pour $q \in \llbracket 1, Q - 1 \rrbracket$, par

$$B'_q \triangleq \text{sign}(h_q h_{q+1}) \frac{\Delta}{(\Delta - |h_q|)(\Delta - |h_{q+1}|)} v(\mathbf{d}, \mathbf{s}_{q:q+1}, \mathbf{h}_{q:q+1}, R_q) \quad (3.14)$$

où l'on définit, pour $q \in \llbracket 1, Q - 1 \rrbracket$, $\mathbf{s}_{q:q+1} \triangleq (s_q; s_{q+1}) \in]0, 1[^2$, et $\mathbf{h}_{q:q+1} \triangleq (h_q; h_{q+1}) \in \mathcal{H}^2$,

$v(\mathbf{d}, \mathbf{s}_{q:q+1}, \mathbf{h}_{q:q+1}, R_q) \triangleq$

$$\begin{cases} 2(\Delta - |h_q| - |h_{q+1}|) - 2(\Delta - \max\{|h_q|, |h_{q+1}|\}), & \text{si } \min\{|h_q|, |h_{q+1}|\} \leq d \\ \left. \begin{aligned} & 2(\Delta - |h_q| - |h_{q+1}|) - (\Delta - \max\{|h_q|, |h_{q+1}|\}) \\ & - \max\{D - |h_q| - |h_{q+1}| + 1, 0\} \\ & + \frac{1 - R_q^{d - \min\{|h_q|, |h_{q+1}|\}}(s_q, s_{q+1})}{1 - R_q(s_q, s_{q+1})} \end{aligned} \right| & \text{si } \min\{|h_q|, |h_{q+1}|\} \geq d + 1. \end{cases} \quad (3.15)$$

avec, pour $q \in \llbracket 1, Q - 1 \rrbracket$ et $(s, s') \in]0, 1[^2$,

$$R_q(s, s') \triangleq \frac{\rho_{q-1,q}(s) \rho_{q,q+1}(s')}{r_q(s, s' - s)}, \quad (3.16)$$

où l'on définit enfin, pour $(s, s') \in]0, 1[^2$,

$$r_q(s, s') \triangleq \int_{\Omega'} p_{\eta_{q-1}}^s(\mathbf{x}) p_{\eta_q}^{s'}(\mathbf{x}) p_{\eta_{q+1}}^{1-s-s'}(\mathbf{x}) d\mathbf{x}. \quad (3.17)$$

On peut noter la similarité de la structure de la matrice \mathbf{G} dans (3.9) avec celle de la matrice d'information de Barankin \mathbf{BIM} , rappelée dans (2.121). Le fait que les termes sous-diagonaux soient non nuls confirme l'idée que l'estimation conjointe de Q changements abrupts n'est pas équivalente à l'estimation séparée d'un unique changement répétée Q fois. En revanche, une différence notable entre cette matrice \mathbf{G} et la matrice \mathbf{BIM} est que les termes sous-diagonaux de \mathbf{G} ne sont généralement jamais nuls. Ainsi \mathbf{G} , contrairement à \mathbf{BIM} , n'est pas diagonale par blocs. Bien que son inversion soit donc moins simple que celle de \mathbf{BIM} , celle-ci peut tout de même être effectuée de manière efficace (voir par exemple [Meu92]).

Ces résultats ont été énoncés quelle que soit la distribution (dont le type a été spécifié au préalable) des observations sur chaque segment. Dans le chapitre 4, nous appliquerons ces résultats à des distributions couramment rencontrées en traitement du signal. Au vu des équations (3.10) et (3.14), le seul effort requis pour le calcul des termes de \mathbf{G} , pour une distribution donnée, sera d'obtenir les expressions de $\rho_{q-1,q}(s)$ (défini en (3.12)) et de $r_q(s, s')$ (défini en (3.17)), l'expression de $R_q(s, s')$ étant immédiatement déduite à partir de (3.16).

3.1.4 Influence de la distribution *a priori*

Dans la section précédente, nous avons donné l'expression des éléments de la matrice \mathbf{G} pour le modèle (3.1), et pour la distribution *a priori* (3.3). Cette distribution *a priori* bien précise nous a permis d'obtenir des expressions analytiques relativement simples pour les éléments de \mathbf{G} . Toutefois cette distribution ne correspond pas forcément à l'information dont on dispose sur les localisations *a priori* des ruptures dans le signal étudié. En effet, la loi (3.3) indique que la longueur d'un segment entre deux ruptures successives est strictement comprise entre deux valeurs fixées d et D , ce qui peut paraître trop restrictif.

Dans une tentative de surmonter cette limitation, nous nous intéressons dans cette section à une famille de distributions *a priori* plus large sur la localisation des ruptures. Concrètement, nous donnons les expressions des éléments de la matrice \mathbf{G} , toujours pour le modèle (3.1), pour des distributions *a priori* de type « chaîne de Markov », i.e.,

$$\pi(\mathbf{k}) = \prod_{q=1}^Q \pi(k_q | k_{q-1}). \quad (3.18)$$

Notons que cette distribution sera toujours conditionnée au nombre de ruptures Q dans la fenêtre d'observation $\llbracket 1, N \rrbracket$. Par conséquent, son support, noté \mathcal{K} , sera également fini et inclus dans $\Theta \triangleq \{k_1, \dots, k_Q \in \llbracket 1, N \rrbracket \mid 0 < k_1 < \dots < k_Q < N\} \subset \llbracket 1, N \rrbracket^Q$.

Afin de ne pas alourdir excessivement l'exposé, nous adoptons la simplification supplémentaire consistant à fixer $s_1 = s_2 = \dots = s_Q = 1/2$ pour les éléments de \mathbf{G} . Comme indiqué à la section 2.1.5.3.2, cette simplification est souvent suggérée et adoptée pour le calcul de la borne de Weiss-Weinstein [WW88, VRBM14, XMGM16].

Dans la méthodologie présentée à la section 3.1.2, seule l'expression de $\nu(\alpha, \beta, h_a, h_b, \mathbf{k})$ va se trouver modifiée pour le calcul de $e^{\tilde{\mu}(\alpha, \beta, h_a, h_b)}$; celle de $M(\alpha, \beta, h_a, h_b, \mathbf{k})$ sera bien sûr inchangée puisqu'elle ne dépend pas de la distribution *a priori*. Comme on considère ici que $\alpha = \beta = 1/2$, on notera ces quantités simplement $\nu(h_a, h_b, \mathbf{k})$, $M(h_a, h_b, \mathbf{k})$ et $e^{\tilde{\mu}(h_a, h_b)}$.

3.1.4.1 Expressions de $M(h_a, h_b, \mathbf{k})$

Comme ces expressions ne dépendent pas de la distribution *a priori*, on les obtient simplement en remplaçant α et β par $1/2$ dans les équations (A.12), (A.13), (A.14), (A.15), (A.24) et (A.27b) de l'annexe A. On notera simplement $\rho_{a-1,a} \triangleq \rho_{a-1,a}(1/2) = \int_{\Omega'} \sqrt{p_{\eta_{q-1}}(\mathbf{x}) p_{\eta_q}(\mathbf{x})} d\mathbf{x}$.

3.1.4.1.1 Cas ts : $b > a + 1$. À partir de (A.15), on obtient simplement

$$M(\mathbf{h}_a, \mathbf{h}_b, \mathbf{k}) = M(\mathbf{h}_a, \mathbf{h}_b) = \rho_{a-1,a}^{|\mathbf{h}_a|} \rho_{b-1,b}^{|\mathbf{h}_b|}. \quad (3.19)$$

3.1.4.1.2 Cas diag : $b = a$. À partir de (A.12), (A.13), (A.14), on obtient respectivement

$$M(h_a, h_a, \mathbf{k}) = M(h_a, h_a) = 1, \quad (3.20)$$

$$M(h_a, -h_a, \mathbf{k}) = M(h_a, -h_a) = \rho_{a-1,a}^{2|h_a|}, \quad (3.21)$$

$$M(h_a, 0, \mathbf{k}) = M(h_a, 0) = \rho_{a-1,a}^{|\mathbf{h}_a|}. \quad (3.22)$$

3.1.4.1.3 Cas psdiag : $b = a + 1$. L'expression de $M(\cdot)$ dans ce cas dépend, comme expliqué dans l'annexe A, de l'éventuel « chevauchement » entre segments successifs (voir section A.2.3). On obtient donc, à partir de (A.24) et (A.27b),

$$M(h_a, h_{a+1}, \mathbf{k}) = \begin{cases} \rho_{a-1,a}^{|\mathbf{h}_a|} \rho_{a,a+1}^{|\mathbf{h}_{a+1}|} \triangleq M(h_a, h_{a+1}), & \text{si } k_a + h_a \leq k_{a+1} + h_{a+1}, \quad (3.23) \\ \frac{\rho_{a-1,a}^{|\mathbf{h}_a|} \rho_{a,a+1}^{|\mathbf{h}_{a+1}|}}{R_a^{(k_a+h_a)-(k_{a+1}+h_{a+1})}} \triangleq \tilde{M}(h_a, h_{a+1}, \mathbf{k}), & \text{si } k_a + h_a > k_{a+1} + h_{a+1}, \quad (3.24) \end{cases}$$

où

$$R_a \triangleq R_a(1/2, 1/2) = \frac{\rho_{a-1,a} \rho_{a,a+1}}{r_a(1/2, 0)} = \frac{\rho_{a-1,a} \rho_{a,a+1}}{\rho_{a-1,a+1}}. \quad (3.25)$$

3.1.4.2 Expressions de $\nu(h_a, h_b, \mathbf{k})$

Contrairement à $M(h_a, h_b, \mathbf{k})$, les expressions de $\nu(h_a, h_b, \mathbf{k}) = \sqrt{\pi(\mathbf{k} + \mathbf{h}_a) \pi(\mathbf{k} + \mathbf{h}_b)}$ dépendent uniquement de la distribution *a priori* choisie. On donne ici ces expressions pour des distributions du type (3.18). On distingue ici aussi les trois cas **diag**, **psdiag** et **ts**.

3.1.4.2.1 Cas diag : $b = a$. On a, d'une part, pour $a < Q$,

$$\nu(h_a, h_a, \mathbf{k}) = \pi(\mathbf{k} + \mathbf{h}_a) = \prod_{q=1}^Q \pi(k_q + \delta_{q,a} h_a | k_{q-1} + \delta_{q-1,a} h_a) \quad (3.26a)$$

$$= \left(\prod_{\substack{q=1 \\ q \neq a, a+1}}^Q \pi(k_q | k_{q-1}) \right) \pi(k_a + h_a | k_{a-1}) \pi(k_{a+1} | k_a + h_a), \quad (3.26b)$$

et, pour $a = Q$,

$$\nu(h_Q, h_Q, \mathbf{k}) = \pi(\mathbf{k} + \mathbf{h}_Q) = \left(\prod_{q=1}^{Q-1} \pi(k_q | k_{q-1}) \right) \pi(k_Q + h_Q | k_{Q-1}). \quad (3.27)$$

D'autre part, à l'aide du changement d'indice $\mathbf{k}' = \mathbf{k} - \mathbf{h}_a$, on a, pour $a < Q$,

$$\begin{aligned} \nu(h_a, -h_a, \mathbf{k}) &= \nu(2h_a, 0, \mathbf{k}') = \sqrt{\pi(\mathbf{k}' + 2\mathbf{h}_a) \pi(\mathbf{k}')} \\ &= \left(\prod_{\substack{q=1 \\ q \neq a, a+1}}^Q \pi(k_q | k_{q-1}) \right) \sqrt{\pi(k'_a + 2h_a | k'_{a-1}) \pi(k'_a | k'_{a-1})} \\ &\quad \times \sqrt{\pi(k'_{a+1} | k'_a + 2h_a) \pi(k'_{a+1} | k'_a)}, \end{aligned} \quad (3.28)$$

et, pour $a = Q$,

$$\nu(h_Q, -h_Q, \mathbf{k}) = \nu(2h_Q, 0, \mathbf{k}') = \left(\prod_{q=1}^{Q-1} \pi(k_q | k_{q-1}) \right) \sqrt{\pi(k'_Q + 2h_Q | k'_{Q-1}) \pi(k'_Q | k'_{Q-1})}. \quad (3.29)$$

3.1.4.2.2 Cas psdiag : $b = a + 1$. Pour $a + 1 < Q$, on a

$$\begin{aligned} \nu(h_a, h_{a+1}, \mathbf{k}) &= \sqrt{\pi(\mathbf{k} + \mathbf{h}_a) \pi(\mathbf{k} + \mathbf{h}_b)} \\ &= \left(\prod_{\substack{q=1 \\ q \neq a, a+1, a+2}}^Q \pi(k_q | k_{q-1}) \right) \sqrt{\pi(k_a + h_a | k_{a-1}) \pi(k_a | k_{a-1})} \\ &\quad \times \sqrt{\pi(k_{a+1} | k_a + h_a) \pi(k_{a+1} + h_{a+1} | k_a)} \\ &\quad \times \sqrt{\pi(k_{a+2} | k_{a+1} + h_{a+1}) \pi(k_{a+2} | k_{a+1})} \end{aligned} \quad (3.30)$$

et pour $a + 1 = Q$, on a

$$\begin{aligned} \nu(h_{Q-1}, h_Q, \mathbf{k}) &= \left(\prod_{q=1}^{Q-2} \pi(k_q | k_{q-1}) \right) \sqrt{\pi(k_{Q-1} + h_{Q-1} | k_{Q-2}) \pi(k_{Q-1} | k_{Q-2})} \\ &\quad \times \sqrt{\pi(k_Q | k_{Q-1} + h_{Q-1}) \pi(k_Q + h_Q | k_{Q-1})}. \end{aligned} \quad (3.31)$$

3.1.4.2.3 Cas ts : $b > a + 1$. De la même façon que dans les cas précédents ,pour $b < Q$, on a

$$\begin{aligned} \nu(h_a, h_b, \mathbf{k}) &= \sqrt{\pi(\mathbf{k} + \mathbf{h}_a) \pi(\mathbf{k} + \mathbf{h}_b)} \\ &= \left(\prod_{\substack{q=1 \\ q \neq a, a+1, b, b+1}}^Q \pi(k_q | k_{q-1}) \right) \sqrt{\pi(k_a + h_a | k_{a-1}) \pi(k_a | k_{a-1})} \sqrt{\pi(k_{a+1} | k_a + h_a) \pi(k_{a+1} | k_a)} \\ &\quad \times \sqrt{\pi(k_b + h_b | k_{b-1}) \pi(k_b | k_{b-1})} \sqrt{\pi(k_{b+1} | k_b + h_b) \pi(k_{b+1} | k_b)}, \end{aligned} \quad (3.32)$$

et, pour $b = Q$,

$$\begin{aligned} \nu(h_a, h_Q, \mathbf{k}) &= \left(\prod_{\substack{q=1 \\ q \neq a, a+1}}^{Q-1} \pi(k_q | k_{q-1}) \right) \sqrt{\pi(k_a + h_a | k_{a-1}) \pi(k_a | k_{a-1})} \sqrt{\pi(k_{a+1} | k_a + h_a) \pi(k_{a+1} | k_a)} \\ &\quad \times \sqrt{\pi(k_Q + h_Q | k_{Q-1}) \pi(k_Q | k_{Q-1})}. \end{aligned} \quad (3.33)$$

3.1.4.3 Expressions de $e^{\tilde{\mu}(h_a, h_b)}$ et des éléments de la matrice \mathbf{G}

Les résultats précédents permettent d'obtenir les expressions de $e^{\tilde{\mu}(h_a, h_b)}$, et *in fine*, les éléments de la matrice \mathbf{G} . Il suffit pour cela d'injecter dans (3.6) les expressions de $M(\cdot)$ (obtenues par (3.19), (3.20), (3.21), (3.22), (3.23) et (3.24)) et les expressions de $\nu(\cdot)$ (obtenues par (3.26b), (3.27), (3.28), (3.29), (3.30), (3.31), (3.32) et (3.33)).

Comme dans l'annexe A pour la distribution *a priori* (3.3), une attention particulière doit être accordée aux bornes de la somme discrète Q -uple dans (3.6), car le support \mathcal{K} de la distribution *a priori* (3.18) est également fini (inclus dans $\Theta \triangleq \{k_1, \dots, k_Q \in \llbracket 1, N \rrbracket \mid 0 < k_1 < \dots < k_Q < N\}$), mais pas nécessairement égal à celui obtenu dans l'annexe A. On pourra néanmoins obtenir des bornes du type de celles obtenues dans les tableaux A.1, A.2 et A.3 de l'annexe A. On reprend donc les notations utilisées dans l'annexe A, section A.1, à savoir que l'on notera

$$\mathcal{K}'_{h_a, h_b} \triangleq (\mathcal{K} - \mathbf{h}_a) \cap (\mathcal{K} - \mathbf{h}_b) \cap \mathcal{K} = \prod_{q=1}^Q \llbracket (k_q)_{\min}, (k_q)_{\max} \rrbracket \quad (3.34)$$

le domaine de la sommation sur \mathbf{k} dans (3.6), avec $(k_q)_{\min}$ et $(k_q)_{\max}$ représentant les bornes de chacune des Q sommes sous-jacentes. Comme vu dans l'annexe A, l'ensemble $\mathcal{K} - \mathbf{h}_a$ dans (3.34) représente, avec un abus de notation, l'ensemble \mathcal{K} translaté de $-\mathbf{h}_a$, c'est-à-dire $\mathcal{K} - \mathbf{h}_a \triangleq \{\mathbf{k} - \mathbf{h}_a \mid \mathbf{k} \in \mathcal{K}\}$. On obtient donc pour $e^{\tilde{\mu}(h_a, h_b)}$ les expressions générales suivantes.

3.1.4.3.1 Cas diag $b = a$. D'après (3.20) et (3.26b), on a d'une part, pour $a < Q$,

$$e^{\tilde{\mu}(h_a, h_a)} = \sum_{\mathbf{k} \in \mathcal{K}'_{h_a, 0}} \pi(\mathbf{k} + \mathbf{h}_a) \quad (3.35a)$$

$$\begin{aligned} &= \sum_{k_1} \pi(k_1) \sum_{k_2} \pi(k_2 \mid k_1) \dots \sum_{k_a} \sqrt{\pi(k_a + h_a \mid k_{a-1}) \pi(k_a \mid k_{a-1})} \\ &\quad \times \sum_{k_{a+1}} \sqrt{\pi(k_{a+1} \mid k_a + h_a) \pi(k_{a+1} \mid k_a)} \\ &\quad \times \sum_{k_{a+2}} \pi(k_{a+2} \mid k_{a+1}) \dots \sum_{k_Q} \pi(k_Q \mid k_{Q-1}), \end{aligned} \quad (3.35b)$$

qui est généralement inférieur à 1. On peut noter que les $(Q - a - 1)$ dernières sommes dans (3.35b) sont égales à 1. Lorsque $a = Q$, cette simplification ne peut se faire, et on a donc seulement

$$e^{\tilde{\mu}(h_Q, h_Q)} = \sum_{\mathbf{k} \in \mathcal{K}'_{h_Q, 0}} \pi(\mathbf{k} + \mathbf{h}_Q) = \sum_{k_1} \pi(k_1) \sum_{k_2} \pi(k_2 \mid k_1) \dots \sum_{k_Q} \sqrt{\pi(k_Q + h_Q \mid k_{Q-1}) \pi(k_Q \mid k_{Q-1})}. \quad (3.36)$$

D'autre part, on obtient de la même manière, à partir de (3.21) et (3.28), et pour $a < Q$,

$$e^{\tilde{\mu}(h_a, -h_a)} = e^{\tilde{\mu}(-h_a, h_a)} = \rho_{a-1, a}^{2|h_a|} \sum_{\mathbf{k}' \in \mathcal{K}'_{-2h_a, 0}} \nu(2h_a, 0, \mathbf{k}') \quad (3.37a)$$

$$\begin{aligned} &= \rho_{a-1, a}^{2|h_a|} \sum_{k_1} \pi(k_1) \sum_{k_2} \pi(k_2 \mid k_1) \dots \sum_{k_a} \sqrt{\pi(k_a + h_a \mid k_{a-1}) \pi(k_a \mid k_{a-1})} \\ &\quad \times \sum_{k_{a+1}} \sqrt{\pi(k_{a+1} \mid k_a + h_a) \pi(k_{a+1} \mid k_a)}, \end{aligned} \quad (3.37b)$$

après simplification des $(Q - a - 1)$ dernières sommes égales à 1. Pour $a = Q$, la même expression est obtenue en supprimant la dernière somme (par rapport à k_{a+1}).

Alors, en injectant (3.35) et (3.37) dans (2.80), on obtient pour le numérateur de $[\mathbf{G}]_{a,a}$

$$\text{num}([\mathbf{G}]_{a,a}) = \sum_{\mathbf{k} \in \mathcal{K}'_{-h_a,0}} \pi(\mathbf{k} + \mathbf{h}_a) + \sum_{\mathbf{k} \in \mathcal{K}'_{h_a,0}} \pi(\mathbf{k} - \mathbf{h}_a) - 2\rho_{a-1,a}^{2|h_a|} \sum_{\mathbf{k}' \in \mathcal{K}'_{-2h_a,0}} \nu(2\mathbf{h}_a, \mathbf{0}, \mathbf{k}'). \quad (3.38)$$

De manière analogue, on obtient les termes du dénominateur de $[\mathbf{G}]_{a,a}$ comme

$$\text{den}([\mathbf{G}]_{a,a}) = \rho_{a-1,a}^{2|h_a|} \left(\sum_{\mathbf{k}' \in \mathcal{K}'_{-h_a,0}} \nu(\mathbf{h}_a, \mathbf{0}, \mathbf{k}') \right)^2. \quad (3.39)$$

3.1.4.3.2 Cas psdiag $b = a + 1$. Nous avons vu dans l'annexe A que pour la distribution *a priori* (3.3), l'expression de $e^{\tilde{\mu}(h_a, h_{a+1})}$ est conditionnée par l'éventuelle présence de termes avec chevauchement entre segments successifs. Cette présence est déterminée par les valeurs respectives de h_a et h_{a+1} , et dépend de la distribution *a priori* considérée.

S'il n'y a pas de termes de chevauchement, on obtient simplement

$$e^{\tilde{\mu}(h_a, h_{a+1})} = M(h_a, h_{a+1}) \sum_{\mathbf{k} \in \mathcal{K}'_{-h_a, -h_{a+1}}} \nu(h_a, h_{a+1}, \mathbf{k}) = \rho_{a-1,a}^{|h_a|} \rho_{a,a+1}^{|h_{a+1}|} \sum_{\mathbf{k} \in \mathcal{K}'_{-h_a, -h_{a+1}}} \nu(h_a, h_{a+1}, \mathbf{k}) \quad (3.40)$$

en utilisant (3.23), et l'expression de $\nu(h_a, h_{a+1}, \mathbf{k})$ donnée par (3.30) ou (3.31).

S'il y a des termes de chevauchement, on devra, comme dans l'annexe A, section A.3.3.1.2, scinder la $(a + 1)$ -ième somme dans (3.6) entre :

- les termes « avec chevauchement » pour lesquels $M(h_a, h_{a+1}, \mathbf{k}) = \tilde{M}(h_a, h_{a+1}, \mathbf{k})$ est donné par (3.24), pour $k_{a+1} < k_a + h_a - h_{a+1}$,
- et les termes « sans chevauchement » pour lesquels $M(h_a, h_{a+1}, \mathbf{k}) = M(h_a, h_{a+1})$ est donné par (3.23), pour $k_{a+1} \geq k_a + h_a - h_{a+1}$.

On obtient donc, en posant $h_a = |h_a| > 0$ et $h_{a+1} = -|h_{a+1}| < 0$,

$$e^{\tilde{\mu}(|h_a|, -|h_{a+1}|)} = \rho_{a,a+1}^{|h_a|} \rho_{a+1,a+2}^{|h_{a+1}|} (S_1 + S_2) \quad (3.41)$$

avec, pour $a + 1 < Q$,

$$\begin{aligned} S_1 \triangleq & \sum_{k_1} \pi(k_1) \sum_{k_2} \pi(k_2 | k_1) \dots \sum_{k_a} \frac{\sqrt{\pi(k_a + h_a | k_{a-1}) \pi(k_a | k_{a-1})}}{R_a^{k_a + |h_a| + |h_{a+1}|}} \\ & \times \sum_{k_{a+1} < k_a + |h_a| + |h_{a+1}|} R_a^{k_{a+1}} \sqrt{\pi(k_{a+1} | k_a + h_a) \pi(k_{a+1} + h_{a+1} | k_a)} \\ & \times \sum_{k_{a+2}} \sqrt{\pi(k_{a+2} | k_{a+1} + h_{a+1}) \pi(k_{a+2} | k_{a+1})}, \end{aligned} \quad (3.42)$$

et

$$\begin{aligned} S_2 \triangleq & \sum_{k_1} \pi(k_1) \sum_{k_2} \pi(k_2 | k_1) \dots \sum_{k_a} \sqrt{\pi(k_a + h_a | k_{a-1}) \pi(k_a | k_{a-1})} \\ & \times \sum_{k_{a+1} \geq k_a + |h_a| + |h_{a+1}|} \sqrt{\pi(k_{a+1} | k_a + h_a) \pi(k_{a+1} + h_{a+1} | k_a)} \\ & \times \sum_{k_{a+2}} \sqrt{\pi(k_{a+2} | k_{a+1} + h_{a+1}) \pi(k_{a+2} | k_{a+1})}, \end{aligned} \quad (3.43)$$

en simplifiant là encore les $(Q - a - 2)$ dernières sommes égales à 1. Si $a + 1 = Q$, on obtient des expressions similaires en supprimant simplement la dernière somme (par rapport à k_{a+2}) dans (3.42) et (3.43).

En utilisant (3.40) et (3.41), on obtient

$$\begin{aligned} \text{num}([\mathbf{G}]_{a,a+1}) = \\ \text{sign}(h_a h_{a+1}) \left(\sum_{\mathbf{k} \in \mathcal{K}'_{-|h_a|, -|h_{a+1}|}} \nu(|h_a|, |h_{a+1}|, \mathbf{k}) + \sum_{\mathbf{k} \in \mathcal{K}'_{|h_a|, |h_{a+1}|}} \nu(-|h_a|, -|h_{a+1}|, \mathbf{k}) - \sum_{\mathbf{k} \in \mathcal{K}'_{|h_a|, -|h_{a+1}|}} \nu(-|h_a|, |h_{a+1}|, \mathbf{k}) - S_1 - S_2 \right), \end{aligned} \quad (3.44)$$

et d'autre part,

$$\text{den}([\mathbf{G}]_{a,a+1}) = \left(\sum_{\mathbf{k}' \in \mathcal{K}'_{-h_a, 0}} \nu(\mathbf{h}_a, \mathbf{0}, \mathbf{k}') \right) \left(\sum_{\mathbf{k}' \in \mathcal{K}'_{-h_{a+1}, 0}} \nu(\mathbf{h}_{a+1}, \mathbf{0}, \mathbf{k}') \right). \quad (3.45)$$

3.1.4.3.3 Cas ts $b > a + 1$. D'après (3.32) (ou (3.33)) et (3.19), on peut montrer que $e^{\tilde{\mu}(h_a, h_b)} = -e^{\tilde{\mu}(h_a, -h_b)}$ en utilisant le changement de variables $k'_b = k_b - h_b$ dans le second membre. Ainsi, on a également $e^{\tilde{\mu}(-h_a, -h_b)} = -e^{\tilde{\mu}(-h_a, h_b)}$, ce qui conduit de nouveau, à partir de (3.6), à

$$[\mathbf{G}]_{a,b} = 0 \quad (3.46)$$

pour $|b - a| > 1$, soit à la structure tridiagonale de la matrice \mathbf{G} . Notons que ce résultat est ici démontré pour une classe plus large de distributions *a priori* (celles données par (3.18)).

3.2 Bornes de Cramér-Rao/Weiss-Weinstein pour l'estimation de $\boldsymbol{\eta}$ et \mathbf{k}

Dans la section précédente, nous nous sommes intéressé à la borne de Weiss-Weinstein pour l'estimation de Q points de ruptures, en considérant le modèle (3.1). Nous avons supposé que les vecteurs $(\boldsymbol{\eta}_q)_{q=0, \dots, Q}$ paramétrant la distribution des observations sur chaque segment étaient connus. En pratique, ces vecteurs sont généralement inconnus, et doivent par conséquent eux aussi être estimés. Ainsi, les performances d'estimation pour les ruptures k_1, \dots, k_Q prédites par la borne de Weiss-Weinstein calculée à la section précédente forment des performances théoriques optimistes. Spécifiquement, on peut s'attendre à ce que ces performances soient optimistes par rapport à celles prédites par une borne où les vecteurs $(\boldsymbol{\eta}_q)_{q=0, \dots, Q}$ seraient supposés inconnus, du moins en régime non asymptotique, i.e., à faibles nombres d'observations sur chaque segment et faibles quantités de changement d'un segment à un autre.

Nous considérons dans cette section deux alternatives pour évaluer les performances d'estimation des vecteurs de paramètres $\boldsymbol{\eta} = [\boldsymbol{\eta}_0; \boldsymbol{\eta}_1; \dots; \boldsymbol{\eta}_Q]$ et \mathbf{k} , suivant l'hypothèse faite sur la nature des vecteurs $(\boldsymbol{\eta}_q)_{q=0, \dots, Q}$ (le vecteur \mathbf{k} étant pour sa part toujours supposé aléatoire).

Si les vecteurs $(\boldsymbol{\eta}_q)_{q=0, \dots, Q}$ sont supposés déterministes, alors on s'intéresse à la borne de Cramér-Rao/Weiss-Weinstein hybride (qui a été présentée au chapitre 2, section 2.1.5.4.2). Plus précisément on cherchera à obtenir des expressions analytiques pour les matrices $\mathbf{V}_{11}^{\boldsymbol{\eta}^*}$, $\mathbf{V}_{12}^{\boldsymbol{\eta}^*}$, $\mathbf{V}_{22}^{\boldsymbol{\eta}^*}$ et $\mathbf{C}_{22}^{\boldsymbol{\eta}^*}$. Celles-ci sont fournies dans la section 3.2.2.

Si les vecteurs $(\boldsymbol{\eta}_q)_{q=0, \dots, Q}$ sont supposés aléatoires, on leur attribue une loi *a priori* $\pi(\boldsymbol{\eta})$ indépendante de la loi *a priori* sur la localisation des ruptures $\pi(\mathbf{k})$, et on s'intéresse alors à la borne de Cramér-Rao/Weiss-Weinstein bayésienne (qui a été présentée au chapitre 2, section

2.1.5.4.1). Plus précisément, on cherchera à obtenir les expressions des matrices V_{11}^π , V_{12}^π , V_{22}^π et C_{22}^π . Celles-ci sont fournies dans la section 3.2.3.

Avant cela, comme les bornes considérées sont très proches, il est utile d'établir un lien entre elles.

3.2.1 Lien entre les bornes de Weiss-Weinstein et de Cramér-Rao/Weiss-Weinstein hybride et bayésienne

3.2.1.1 Retour sur le modèle et notations

Dans le cadre complètement bayésien, où l'on suppose les vecteurs $(\boldsymbol{\eta}_q)_{q=0,\dots,Q}$ inconnus et de nature aléatoire, nous attribuons à ceux-ci une loi *a priori* notée $\pi(\boldsymbol{\eta})$ du type

$$\pi(\boldsymbol{\eta}) = \prod_{q=0}^Q \pi(\boldsymbol{\eta}_q), \quad (3.47)$$

c'est-à-dire que l'on suppose que les vecteurs $\boldsymbol{\eta}_q = [\eta_{q,1}, \eta_{q,2}, \dots, \eta_{q,L}]^\top$ sont indépendants d'un segment à l'autre. On notera $\Upsilon \triangleq \Upsilon_0 \times \Upsilon_1 \times \dots \times \Upsilon_Q \subseteq \mathbb{R}^{L(Q+1)}$ le support de la loi *a priori* $\pi(\boldsymbol{\eta})$, avec $\Upsilon_q \subseteq \mathbb{R}^L$, $q = 0, \dots, Q$ les supports des lois *a priori* sur chacun des vecteurs $\boldsymbol{\eta}_q$. On supposera également que la distribution $\pi(\boldsymbol{\eta})$ est indépendante de celle sur les localisations des ruptures, i.e., qu'on aura $\pi(\mathbf{k}, \boldsymbol{\eta}) = \pi(\mathbf{k}) \pi(\boldsymbol{\eta})$. Le type de distribution *a priori* assigné à $\boldsymbol{\eta}_q$ dépend en général de la famille de la distribution des observations sur chaque segment³.

En ce qui concerne la distribution *a priori* sur les ruptures $\pi(\mathbf{k})$, nous supposons dans la suite qu'il s'agit d'une marche aléatoire uniforme comme celle donnée par (3.3).

Ainsi, la loi jointe entre les observations et les vecteurs de paramètres inconnus (à estimer) peut s'écrire

$$p(\mathbf{X}, \mathbf{k}, \boldsymbol{\eta}) = p(\mathbf{X} | \mathbf{k}, \boldsymbol{\eta}) \pi(\mathbf{k}) \pi(\boldsymbol{\eta}). \quad (3.48)$$

Il faut ici noter que la distribution $p(\mathbf{X} | \mathbf{k}, \boldsymbol{\eta})$ correspond à la vraisemblance notée $p(\mathbf{X} | \mathbf{k})$ à l'équation (3.4), où les vecteurs $(\boldsymbol{\eta}_q)_{q=0,\dots,Q}$ étaient supposés connus. Dans le contexte de l'estimation hybride où les vecteurs $(\boldsymbol{\eta}_q)_{q=0,\dots,Q}$ sont supposés déterministes, on note cette même fonction de vraisemblance comme $p(\mathbf{X} | \mathbf{k}; \boldsymbol{\eta})$. De la même manière, suivant le contexte d'estimation, les distributions des observations sur chaque segment qui apparaissent dans le modèle (3.1) sont notées respectivement

- $p_{\boldsymbol{\eta}_q}(\cdot)$ lorsque les vecteurs $(\boldsymbol{\eta}_q)_{q=0,\dots,Q}$ sont supposés connus (comme écrit dans (3.1));
- $p(\cdot; \boldsymbol{\eta}_q)$ lorsque les vecteurs $(\boldsymbol{\eta}_q)_{q=0,\dots,Q}$ sont supposés inconnus et de nature déterministe (i.e., on ne dispose d'aucune information *a priori* dessus);
- $p(\cdot | \boldsymbol{\eta}_q)$ lorsque les vecteurs $(\boldsymbol{\eta}_q)_{q=0,\dots,Q}$ sont supposés inconnus et de nature aléatoire (i.e., on dispose de l'information *a priori* donnée par $\pi(\boldsymbol{\eta})$).

Ces remarques nous permettent d'établir les liens suivants entre les bornes de Weiss-Weinstein, de Cramér-Rao/Weiss-Weinstein hybride et bayésienne.

3.2.1.2 Lien entre les matrices $G(\mathbf{H}, \mathbf{s})$, $V_{22}^{\boldsymbol{\eta}^*}$, $C_{22}^{\boldsymbol{\eta}^*}$, V_{22}^π et C_{22}^π

Intéressons-nous d'abord à l'expression des éléments de la matrice $V_{22}^{\boldsymbol{\eta}^*}$ de la borne de Cramér-Rao/Weiss-Weinstein hybride, donnée par (2.101), en remplaçant $\boldsymbol{\theta}_a$ par \mathbf{k} et $\boldsymbol{\theta}_d$ par $\boldsymbol{\eta}$. Elle correspond exactement à l'expression du numérateur des éléments de la matrice $G(\mathbf{H}, \mathbf{s})$ donnés dans (2.79). De même, les éléments du bloc $C_{22}^{\boldsymbol{\eta}^*}$ donnés dans (2.97) correspondent au

3. Il est d'usage de choisir une distribution *a priori* dite « conjuguée » (voir [Rob07]).

dénominateur des éléments de $\mathbf{G}(\mathbf{H}, \mathbf{s})$. Notons que du fait de la forme des vecteurs test \mathbf{h}_q choisie, la matrice $\mathbf{C}_{22}^{\boldsymbol{\eta}^*}$ est nécessairement diagonale. Ainsi, les blocs $\mathbf{V}_{22}^{\boldsymbol{\eta}^*}$ et $\mathbf{C}_{22}^{\boldsymbol{\eta}^*}$ de la BCRWWH concordent précisément avec la borne de Weiss-Weinstein pour l'estimation de \mathbf{k} , pour des vecteurs $\boldsymbol{\eta}_0, \dots, \boldsymbol{\eta}_Q$ donnés. On notera⁴, pour $a, b \in \llbracket 1, Q \rrbracket$,

$$[\mathbf{V}_{22}^{\boldsymbol{\eta}^*}]_{a,b} \equiv \text{num}([\mathbf{G}]_{a,b}), \quad (3.49)$$

$$[\mathbf{C}_{22}^{\boldsymbol{\eta}^*}]_{a,a} \equiv h_a \sqrt{\text{den}([\mathbf{G}]_{a,a})}, \quad (3.50)$$

signifiant que ces relations sont valables, modulo une légère adaptation dans les notations des distributions mentionnées à la section 3.2.1.1. De manière analogue, en utilisant (3.48) dans l'expression des éléments de \mathbf{C}_{22}^π et \mathbf{V}_{22}^π donnés en (2.87) et (2.91), on peut aboutir, pour $a, b \in \llbracket 1, Q \rrbracket$, à

$$[\mathbf{C}_{22}^\pi]_{a,b} \equiv \mathbb{E}_\boldsymbol{\eta} \{ [\mathbf{C}_{22}^{\boldsymbol{\eta}}]_{a,b} \}, \quad (3.51)$$

$$[\mathbf{V}_{22}^\pi]_{a,b} \equiv \mathbb{E}_\boldsymbol{\eta} \{ [\mathbf{V}_{22}^{\boldsymbol{\eta}}]_{a,b} \}. \quad (3.52)$$

3.2.1.3 Lien entre les matrices $\mathbf{V}_{11}^{\boldsymbol{\eta}^*}$ et \mathbf{V}_{11}^π

De même que précédemment, en utilisant (3.48) dans l'expression de la matrice \mathbf{V}_{11}^π donnée en (2.89), on obtient

$$\mathbf{V}_{11}^\pi \equiv \mathbb{E}_\boldsymbol{\eta} \{ \mathbf{V}_{11}^{\boldsymbol{\eta}} \} - \mathbb{E}_\boldsymbol{\eta} \left\{ \frac{\partial^2 \ln \pi(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top} \right\}, \quad (3.53)$$

relation analogue à celle reliant la borne de Cramér-Rao bayésienne et la borne de Cramér-Rao classique (écrite en (2.71) dans le cas scalaire). On relèvera toutefois qu'ici la matrice $\mathbf{V}_{11}^{\boldsymbol{\eta}}$ correspond à l'information de Fisher sur $\boldsymbol{\eta}$ où les ruptures k_1, \dots, k_Q sont vues comme des paramètres de nuisance ; la borne de Cramér-Rao qui y est associée correspond donc à la borne dite de « Cramér-Rao modifiée » introduite dans [GRM98].

3.2.1.4 Lien entre les matrices $\mathbf{V}_{12}^{\boldsymbol{\eta}^*}$ et \mathbf{V}_{12}^π

Enfin, toujours en utilisant (3.48) dans l'expression des éléments de la matrice \mathbf{V}_{12}^π donnée en (2.90), on obtient, pour $p \in \llbracket 1, L(Q+1) \rrbracket$, $q \in \llbracket 1, Q \rrbracket$, et en posant $p = L\tilde{q} + \ell$ avec $\tilde{q} \in \llbracket 0, Q \rrbracket$ et $\ell \in \llbracket 1, L \rrbracket$,

$$\begin{aligned} [\mathbf{V}_{12}^\pi]_{p,q} &\equiv \mathbb{E}_\boldsymbol{\eta} \{ \mathbf{V}_{12}^{\boldsymbol{\eta}} \} + \mathbb{E}_\boldsymbol{\eta} \left\{ \frac{\partial \ln \pi(\boldsymbol{\eta})}{\partial \eta_{\tilde{q}, \ell}} (e^{\mu(s_q, h_q)} - e^{\mu(1-s_q, -h_q)}) \right\} \\ &\equiv \mathbb{E}_\boldsymbol{\eta} \{ \mathbf{V}_{12}^{\boldsymbol{\eta}} \}, \end{aligned} \quad (3.54)$$

car, d'après (A.32) obtenue dans l'annexe A, on a $e^{\mu(1-s_q, -h_q)} = e^{\mu(s_q, h_q)}$.

Ces relations montrent que l'on peut obtenir la borne de Cramér-Rao/Weiss-Weinstein bayésienne à partir de la borne de Cramér-Rao/Weiss-Weinstein hybride. D'une manière générale, les espérances par rapport au vecteur $\boldsymbol{\eta}$ n'ont pas de forme analytique, et doivent donc être évaluées numériquement.

Nous fournissons donc, dans la section suivante, les expressions des éléments de la BCRWWH. Nous en déduisons ensuite, dans la section 3.2.3, les éléments de la BCRWWB.

4. Le signe « \equiv » utilisé ici signifie que les deux membres sont égaux, modulo une légère modification dans la notation des densités : il suffit de remplacer $p_{\boldsymbol{\eta}_q}(\mathbf{x}_n)$ par $p(\mathbf{x}_n; \boldsymbol{\eta}_q)$, ou $p(\mathbf{x}_n; \boldsymbol{\eta}_q)$ par $p(\mathbf{x}_n | \boldsymbol{\eta}_q)$, selon les objets comparés.

3.2.2 Borne de Cramér-Rao/Weiss-Weinstein hybride

Nous donnons d'abord l'expression de la matrice $\mathbf{V}_{11}^{\eta^*}$, ensuite celles des matrices $\mathbf{V}_{22}^{\eta^*}$ et $\mathbf{C}_{22}^{\eta^*}$, puis enfin celle de la matrice $\mathbf{V}_{12}^{\eta^*}$. Les calculs permettant de parvenir à ces expressions sont fournis dans l'annexe B.

3.2.2.1 Expression de la matrice $\mathbf{V}_{11}^{\eta^*}$

En reformulant l'équation (2.99) avec les notations de la section 3.2.1, et en introduisant les changements de variable $p = Lq + \ell$ et $p' = Lq' + \ell'$, avec $(q, q') \in \llbracket 0, Q \rrbracket^2$ et $(\ell, \ell') \in \llbracket 1, L \rrbracket^2$, on a

$$[\mathbf{V}_{11}^{\eta^*}]_{p,p'} = -\mathbb{E}_{\mathbf{x}, \mathbf{k}; \eta^*} \left\{ \frac{\partial^2 \ln p(\mathbf{x}, \mathbf{k}; \eta)}{\partial \eta_{q,\ell} \partial \eta_{q',\ell'}} \Big|_{\eta^*} \right\}. \quad (3.55)$$

Dans l'annexe B, on montre que la matrice $\mathbf{V}_{11}^{\eta^*}$ est diagonale par blocs, c'est-à-dire

$$\boxed{\mathbf{V}_{11}^{\eta^*} = \text{bdiag} \left(\frac{d+D}{2} \tilde{\mathbf{F}}(\eta_0^*), \dots, \frac{d+D}{2} \tilde{\mathbf{F}}(\eta_{Q-1}^*), \left(N - \frac{Q(d+D)}{2} \right) \tilde{\mathbf{F}}(\eta_Q^*) \right)} \quad (3.56)$$

où $\tilde{\mathbf{F}}(\eta_q^*)$ représente la matrice d'information de Fisher de taille $(L \times L)$ pour $\eta_q \in \mathbb{R}^L$, et pour un seul instant d'observation sur le q -ième segment, c'est-à-dire que pour $q \in \llbracket 0, Q \rrbracket$ et $(\ell, \ell') \in \llbracket 1, L \rrbracket^2$,

$$[\tilde{\mathbf{F}}(\eta_q)]_{\ell,\ell'} = -\mathbb{E}_{x; \eta_q} \left\{ \frac{\partial^2 \ln p(x; \eta)}{\partial \eta_{q,\ell} \partial \eta_{q,\ell'}} \right\} = -\int_{\Omega'} \frac{\partial^2 \ln p(x; \eta_q)}{\partial \eta_{q,\ell} \partial \eta_{q,\ell'}} p(x; \eta_q) dx. \quad (3.57)$$

La structure diagonale par blocs de $\mathbf{V}_{11}^{\eta^*}$ indique le découplage entre les vecteurs $(\eta_q)_{q=0,\dots,Q}$, d'un segment à l'autre.

3.2.2.2 Expression des matrices $\mathbf{V}_{22}^{\eta^*}$ et $\mathbf{C}_{22}^{\eta^*}$

Comme mentionné dans la section 3.2.1.2, ces matrices coïncident exactement avec la borne de Weiss-Weinstein calculée dans la section 3.1 (voir (3.49) et (3.50)). On obtient donc, en adaptant les résultats de la section 3.1.3, les résultats suivants.

La matrice $\mathbf{V}_{22}^{\eta^*}$ est tridiagonale, i.e., quel que soit $(q, q') \in \llbracket 1, Q \rrbracket^2$ tel que $|q - q'| > 1$, on a

$$\boxed{[\mathbf{V}_{22}^{\eta^*}]_{q,q'} = 0.} \quad (3.58)$$

Les termes diagonaux de $\mathbf{V}_{22}^{\eta^*}$ correspondent au numérateur de (3.10), ils sont donc donnés, pour $q \in \llbracket 1, Q \rrbracket$, par

$$\boxed{[\mathbf{V}_{22}^{\eta^*}]_{q,q} = u(\Delta, h_q) \cdot \left(\rho_{q-1,q}^{|h_q|}(\varepsilon_{h_q}(2s_q)) + \rho_{q-1,q}^{|h_q|}(\varepsilon_{h_q}(2s_q - 1)) \right) - 2u(\Delta, 2h_q) \cdot \rho_{q-1,q}^{2|h_q|}(\varepsilon_{h_q}(s_q))} \quad (3.59)$$

où $u(\Delta, h_q)$ est défini dans (3.11), $\rho_{q-1,q}(s)$ est défini dans (3.12) (où il suffit de remplacer $p_{\eta_q}(\cdot)$ par $p(\cdot; \eta_q)$), et $\varepsilon_h(s)$ est défini dans (3.13).

De la même manière, les éléments de $\mathbf{C}_{22}^{\boldsymbol{\eta}^*}$ (qui est diagonale) sont donnés, pour $q \in \llbracket 1, Q \rrbracket$, par

$$[\mathbf{C}_{22}^{\boldsymbol{\eta}^*}]_{q,q} = h_q u(\Delta, h_q) \cdot \rho_{q-1,q}^{|h_q|}(\varepsilon_{h_q}(s_q)). \quad (3.60)$$

Enfin, les termes des surdiagonales de $[\mathbf{V}_{22}^{\boldsymbol{\eta}^*}]$ sont donnés, pour $q \in \llbracket 1, Q-1 \rrbracket$, par

$$[\mathbf{V}_{22}^{\boldsymbol{\eta}^*}]_{q,q+1} \triangleq \text{sign}(h_q h_{q+1}) \tilde{u}(\Delta, \mathbf{h}_{q:q+1}) v(\mathbf{d}, \mathbf{s}_{q:q+1}, \mathbf{h}_{q:q+1}, R_q) \rho_{q-1,q}^{|h_q|}(\varepsilon_{h_q}(s_q)) \rho_{q,q+1}^{|h_{q+1}|}(\varepsilon_{h_{q+1}}(s_{q+1})) \quad (3.61)$$

où, pour $q \in \llbracket 1, Q-1 \rrbracket$ et $\mathbf{h}_{q:q+1} \triangleq (h_q; h_{q+1}) \in \mathcal{H}^2$,

$$\tilde{u}(\Delta, \mathbf{h}_{q:q+1}) \triangleq \begin{cases} \frac{(\Delta - |h_q|)(\Delta - |h_{q+1}|)}{\Delta^3} & \text{si } q < Q \text{ et } \max(|h_q|, |h_{q+1}|) < \Delta \\ \frac{\Delta - |h_Q|}{\Delta^2} & \text{si } q = Q \text{ et } |h_Q| < \Delta \\ 0 & \text{si } \max(|h_q|, |h_{q+1}|) \geq \Delta, \end{cases} \quad (3.62)$$

et $v(\mathbf{d}, \mathbf{s}_{q:q+1}, \mathbf{h}_{q:q+1}, R_q)$ est donné par (3.15).

3.2.2.3 Expression de la matrice $\mathbf{V}_{12}^{\boldsymbol{\eta}^*}$

Pour $p \in \llbracket 1, L(Q+1) \rrbracket$ et $q \in \llbracket 1, Q \rrbracket$, en posant $p = L\tilde{q} + \ell$ avec $\tilde{q} \in \llbracket 0, Q \rrbracket$ et $\ell \in \llbracket 1, L \rrbracket$, en reformulant l'équation (2.100), on obtient

$$[\mathbf{V}_{12}^{\boldsymbol{\eta}^*}]_{p,q} = \mathbb{E}_{\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}^*} \left\{ \frac{\partial \ln p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{\tilde{q}, \ell}} \Bigg|_{\boldsymbol{\eta}^*} \left[\frac{p^{s_q}(\mathbf{x}, \mathbf{k} + \mathbf{h}_q; \boldsymbol{\eta}^*)}{p^{s_q}(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}^*)} - \frac{p^{1-s_q}(\mathbf{x}, \mathbf{k} - \mathbf{h}_q; \boldsymbol{\eta}^*)}{p^{1-s_q}(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}^*)} \right] \right\}. \quad (3.63)$$

Comme montré dans l'annexe B, la matrice $\mathbf{V}_{12}^{\boldsymbol{\eta}^*}$ est de la forme

$$\mathbf{V}_{12}^{\boldsymbol{\eta}^*} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \tilde{\mathbf{w}}_1 & \mathbf{w}_2 & \ddots & \vdots \\ \mathbf{0} & \tilde{\mathbf{w}}_2 & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{w}_Q \\ \mathbf{0} & \dots & \mathbf{0} & \tilde{\mathbf{w}}_Q \end{bmatrix}, \quad (3.64)$$

où, pour $q \in \llbracket 1, Q \rrbracket$, les vecteurs \mathbf{w}_q et $\tilde{\mathbf{w}}_q$ sont de taille $L \times 1$ et leurs éléments peuvent être écrits, pour $\ell \in \llbracket 1, L \rrbracket$, d'une part, comme

$$w_{q,\ell} = [\mathbf{V}_{12}^{\boldsymbol{\eta}^*}]_{L(q-1)+\ell,q} = -h_q u(\Delta, h_q) \rho_{q-1,q}^{|h_q|-1}(\varepsilon_{h_q}(s_q)) \varphi_{q,\eta_{q-1,\ell}}(\varepsilon_{h_q}(s_q)) \quad (3.65)$$

et, d'autre part, comme

$$\tilde{w}_{q,\ell} = [\mathbf{V}_{12}^{\boldsymbol{\eta}^*}]_{Lq+\ell,q} = h_q u(\Delta, h_q) \rho_{q-1,q}^{|h_q|-1}(\varepsilon_{h_q}(s_q)) \varphi_{q,\eta_{q,\ell}}(\varepsilon_{h_q}(s_q)) \quad (3.66)$$

avec, pour $j \in \llbracket 1, Q \rrbracket$, $\tilde{j} \in \{j-1, j\}$, $k \in \llbracket 1, L \rrbracket$, et $s \in]0, 1[$, la quantité $\varphi_{j, \eta_{\tilde{j}, k}}(s)$ définie par

$$\varphi_{j, \eta_{\tilde{j}, k}}(s) \triangleq \int_{\Omega'} \frac{\partial \ln p(x; \boldsymbol{\eta}_{\tilde{j}})}{\partial \eta_{\tilde{j}, k}} \Big|_{\boldsymbol{\eta}^*} p^s(x; \boldsymbol{\eta}_{\tilde{j}-1}^*) p^{1-s}(x; \boldsymbol{\eta}_{\tilde{j}}^*) dx. \quad (3.67)$$

Pour conclure, les équations (3.56), (3.58)–(3.61), et (3.64)–(3.66) fournissent toutes les expressions nécessaires pour déterminer les éléments de la matrice $\mathbf{C}^{\boldsymbol{\eta}^*} (\mathbf{V}^{\boldsymbol{\eta}^*})^{-1} (\mathbf{C}^{\boldsymbol{\eta}^*})^\top$. Il est intéressant de noter que vu les structures des matrices $\mathbf{V}_{11}^{\boldsymbol{\eta}^*}$, $\mathbf{V}_{22}^{\boldsymbol{\eta}^*}$ et $\mathbf{V}_{12}^{\boldsymbol{\eta}^*}$, l'inversion de $\mathbf{V}^{\boldsymbol{\eta}^*}$ ne s'avère pas très difficile d'un point de vue calculatoire.

Par ailleurs, on notera bien que le fait que la matrice $\mathbf{V}_{12}^{\boldsymbol{\eta}^*}$ soit non nulle en général prouve que la borne de Cramér-Rao/Weiss-Weinstein hybride ne consiste pas en la juxtaposition naïve de ces deux bornes, et qu'il existe donc bien, en général, un couplage entre les positions des ruptures k_1, \dots, k_Q et les paramètres des distributions sur chaque segment, rassemblés dans les vecteurs $(\boldsymbol{\eta}_q)_{q=0, \dots, Q}$.

3.2.3 Borne de Cramér-Rao/Weiss-Weinstein bayésienne

Les expressions des éléments de la borne de Cramér-Rao/Weiss-Weinstein bayésienne peuvent être obtenus directement à partir des résultats de la section précédente, avec les relations (3.51)–(3.54) de la section 3.2.1. On liste ces résultats ci-dessous.

3.2.3.1 Expression de la matrice \mathbf{V}_{11}^π

En utilisant les relations (3.53) et (3.56), on obtient

$$\mathbf{V}_{11}^\pi = \text{bdiag} \left(\begin{aligned} & \frac{d+D}{2} \mathbb{E}_{\boldsymbol{\eta}_0} \{ \tilde{\mathbf{F}}(\boldsymbol{\eta}_0) \} - \mathbb{E}_{\boldsymbol{\eta}_0} \left\{ \frac{\partial^2 \ln \pi(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}_0 \partial \boldsymbol{\eta}_0^\top} \right\}, \dots, \\ & \frac{d+D}{2} \mathbb{E}_{\boldsymbol{\eta}_{Q-1}} \{ \tilde{\mathbf{F}}(\boldsymbol{\eta}_{Q-1}) \} - \mathbb{E}_{\boldsymbol{\eta}_{Q-1}} \left\{ \frac{\partial^2 \ln \pi(\boldsymbol{\eta}_{Q-1})}{\partial \boldsymbol{\eta}_{Q-1} \partial \boldsymbol{\eta}_{Q-1}^\top} \right\}, \\ & \left(N - \frac{Q(d+D)}{2} \right) \mathbb{E}_{\boldsymbol{\eta}_Q} \{ \tilde{\mathbf{F}}(\boldsymbol{\eta}_Q) \} - \mathbb{E}_{\boldsymbol{\eta}_Q} \left\{ \frac{\partial^2 \ln \pi(\boldsymbol{\eta}_Q)}{\partial \boldsymbol{\eta}_Q \partial \boldsymbol{\eta}_Q^\top} \right\} \end{aligned} \right) \quad (3.68)$$

où, pour $q \in \llbracket 0, Q \rrbracket$ et $(\ell, \ell') \in \llbracket 1, L \rrbracket^2$,

$$\begin{aligned} [\mathbb{E}_{\boldsymbol{\eta}_q} \{ \tilde{\mathbf{F}}(\boldsymbol{\eta}_q) \}]_{\ell, \ell'} &= - \mathbb{E}_{\boldsymbol{\eta}_q} \left\{ \mathbb{E}_{x | \boldsymbol{\eta}_q} \left\{ \frac{\partial^2 \ln p(x | \boldsymbol{\eta})}{\partial \eta_{q, \ell} \partial \eta_{q, \ell'}} \right\} \right\} \\ &= - \int_{\Upsilon_q} \pi(\boldsymbol{\eta}_q) \int_{\Omega'} \frac{\partial^2 \ln p(x | \boldsymbol{\eta}_q)}{\partial \eta_{q, \ell} \partial \eta_{q, \ell'}} p(x | \boldsymbol{\eta}_q) dx d\boldsymbol{\eta}_q. \end{aligned} \quad (3.69)$$

3.2.3.2 Expression des matrices \mathbf{V}_{22}^π et \mathbf{C}_{22}^π

En utilisant les relations (3.49) et (3.50) d'une part, et (3.58)–(3.61) d'autre part, on obtient les résultats suivants.

La matrice \mathbf{V}_{22}^π est tridiagonale, i.e., quel que soit $(q, q') \in \llbracket 1, Q \rrbracket^2$ tel que $|q - q'| > 1$, on a

$$[\mathbf{V}_{22}^\pi]_{q, q'} = 0. \quad (3.70)$$

Les termes diagonaux de \mathbf{V}_{22}^π sont donnés, pour $q \in \llbracket 1, Q \rrbracket$, par

$$\boxed{[\mathbf{V}_{22}^\pi]_{q,q} = u(\Delta, h_q) \cdot \left(\mathbb{E}_\boldsymbol{\eta} \left\{ \rho_{q-1,q}^{|h_q|}(\varepsilon_{h_q}(2s_q)) \right\} + \mathbb{E}_\boldsymbol{\eta} \left\{ \rho_{q-1,q}^{|h_q|}(\varepsilon_{h_q}(2s_q - 1)) \right\} \right) - 2u(\Delta, 2h_q) \cdot \mathbb{E}_\boldsymbol{\eta} \left\{ \rho_{q-1,q}^{2|h_q|}(\varepsilon_{h_q}(s_q)) \right\}} \quad (3.71)$$

où, pour $q \in \llbracket 1, Q \rrbracket$, $s \in]0, 1[$ et $h \in \mathcal{H}$,

$$\begin{aligned} \mathbb{E}_\boldsymbol{\eta} \left\{ \rho_{q-1,q}^{|h|}(s) \right\} &= \int_{\Upsilon} \rho_{q-1,q}^{|h|}(s) \pi(\boldsymbol{\eta}) \, d\boldsymbol{\eta} \\ &= \int_{\Upsilon_{q-1}} \int_{\Upsilon_q} \pi(\boldsymbol{\eta}_{q-1}) \pi(\boldsymbol{\eta}_q) \left(\int_{\Omega'} p^s(\mathbf{x} | \boldsymbol{\eta}_{q-1}) p^{1-s}(\mathbf{x} | \boldsymbol{\eta}_q) \, d\mathbf{x} \right)^h \, d\boldsymbol{\eta}_{q-1} \, d\boldsymbol{\eta}_q \end{aligned} \quad (3.72)$$

De la même manière, les éléments de \mathbf{C}_{22}^π (qui est diagonale) sont donnés, pour $q \in \llbracket 1, Q \rrbracket$, par

$$\boxed{[\mathbf{C}_{22}^\pi]_{q,q} = h_q u(\Delta, h_q) \cdot \mathbb{E}_\boldsymbol{\eta} \left\{ \rho_{q-1,q}^{|h_q|}(\varepsilon_{h_q}(s_q)) \right\}.} \quad (3.73)$$

Enfin, les termes des surdiagonales de $[\mathbf{V}_{22}^\pi]$ sont donnés, pour $q \in \llbracket 1, Q - 1 \rrbracket$, par

$$\boxed{[\mathbf{V}_{22}^\pi]_{q,q+1} \triangleq \text{sign}(h_q h_{q+1}) \tilde{u}(\Delta, h_q, h_{q+1}) \times \mathbb{E}_\boldsymbol{\eta} \left\{ v(\mathbf{d}, \mathbf{s}_{q:q+1}, \mathbf{h}_{q:q+1}, R_q) \rho_{q-1,q}^{|h_q|}(\varepsilon_{h_q}(s_q)) \rho_{q,q+1}^{|h_{q+1}|}(\varepsilon_{h_{q+1}}(s_{q+1})) \right\}} \quad (3.74)$$

où

$$\begin{aligned} \mathbb{E}_\boldsymbol{\eta} \left\{ v(\mathbf{d}, \mathbf{s}_{q:q+1}, \mathbf{h}_{q:q+1}, R_q) \rho_{q-1,q}^{|h_q|}(\varepsilon_{h_q}(s_q)) \rho_{q,q+1}^{|h_{q+1}|}(\varepsilon_{h_{q+1}}(s_{q+1})) \right\} &= \\ \int_{\Upsilon_{q-1}} \int_{\Upsilon_q} \int_{\Upsilon_{q+1}} \pi(\boldsymbol{\eta}_{q-1}) \pi(\boldsymbol{\eta}_q) \pi(\boldsymbol{\eta}_{q+1}) v(\mathbf{d}, \mathbf{s}_{q:q+1}, \mathbf{h}_{q:q+1}, R_q) & \\ \times \rho_{q-1,q}^{|h_q|}(\varepsilon_{h_q}(s_q)) \rho_{q,q+1}^{|h_{q+1}|}(\varepsilon_{h_{q+1}}(s_{q+1})) \, d\boldsymbol{\eta}_{q-1} \, d\boldsymbol{\eta}_q \, d\boldsymbol{\eta}_{q+1} & \end{aligned} \quad (3.75)$$

3.2.3.3 Expression de la matrice \mathbf{V}_{12}^π

En utilisant (3.54), (3.64) et (3.63), on obtient

$$\boxed{\mathbf{V}_{12}^\pi = \begin{bmatrix} \mathbb{E}_\boldsymbol{\eta} \{ \mathbf{w}_1 \} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbb{E}_\boldsymbol{\eta} \{ \tilde{\mathbf{w}}_1 \} & \mathbb{E}_\boldsymbol{\eta} \{ \mathbf{w}_2 \} & \ddots & \vdots \\ \mathbf{0} & \mathbb{E}_\boldsymbol{\eta} \{ \tilde{\mathbf{w}}_2 \} & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbb{E}_\boldsymbol{\eta} \{ \mathbf{w}_Q \} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbb{E}_\boldsymbol{\eta} \{ \tilde{\mathbf{w}}_Q \} \end{bmatrix},} \quad (3.76)$$

où, pour $q \in \llbracket 1, Q \rrbracket$, comme pour la BCRWWH, les vecteurs $\mathbb{E}_\boldsymbol{\eta} \{ \mathbf{w}_q \}$ et $\mathbb{E}_\boldsymbol{\eta} \{ \tilde{\mathbf{w}}_q \}$ sont de taille $L \times 1$ et leurs éléments peuvent être écrits, pour $\ell \in \llbracket 1, L \rrbracket$, d'une part, comme

$$\boxed{\mathbb{E}_\boldsymbol{\eta} \{ w_{q,\ell} \} = \mathbb{E}_\boldsymbol{\eta} \{ [\mathbf{V}_{12}^{\boldsymbol{\eta}^*}]_{L(q-1)+\ell,q} \} = -h_q u(\Delta, h_q) \mathbb{E}_\boldsymbol{\eta} \left\{ \rho_{q-1,q}^{|h_q|-1}(\varepsilon_{h_q}(s_q)) \varphi_{q,\eta_{q-1},\ell}(\varepsilon_{h_q}(s_q)) \right\}} \quad (3.77)$$

et, d'autre part, comme

$$\mathbb{E}_{\boldsymbol{\eta}}\{\tilde{w}_{q,\ell}\} = \mathbb{E}_{\boldsymbol{\eta}}\{[\mathbf{V}_{12}^{\boldsymbol{\eta}^*}]_{Lq+\ell,q}\} = h_q u(\Delta, h_q) \mathbb{E}_{\boldsymbol{\eta}}\left\{\rho_{q-1,q}^{|\mathbf{h}_q|^{-1}}(\varepsilon_{h_q}(s_q)) \varphi_{q,\eta_{q,\ell}}(\varepsilon_{h_q}(s_q))\right\} \quad (3.78)$$

avec, pour $j \in \llbracket 1, Q \rrbracket$, $\tilde{j} \in \{j-1, j\}$, $k \in \llbracket 1, L \rrbracket$, et $s \in]0, 1[$,

$$\mathbb{E}_{\boldsymbol{\eta}}\left\{\rho_{q-1,q}^{|\mathbf{h}_q|^{-1}}(\varepsilon_{h_q}(s_q)) \varphi_{j,\eta_{\tilde{j},k}}(s)\right\} = \int_{\Upsilon_{q-1}} \int_{\Upsilon_q} \pi(\boldsymbol{\eta}_{q-1}) \pi(\boldsymbol{\eta}_q) \rho_{q-1,q}^{|\mathbf{h}_q|^{-1}}(\varepsilon_{h_q}(s_q)) \varphi_{j,\eta_{\tilde{j},k}}(s) d\boldsymbol{\eta}_{q-1} d\boldsymbol{\eta}_q. \quad (3.79)$$

Pour conclure, nous disposons donc de toutes les expressions nécessaires pour déterminer les éléments de la matrice $\mathbf{C}^\pi(\mathbf{V}^\pi)^{-1}(\mathbf{C}^\pi)^\top$. Même si les structures des matrices sont préservées par rapport au cas de la BCRWWH, il subsiste toutefois dans l'expression de leurs termes des intégrales dont nous ne pourrions pas toujours obtenir des formes analytiques, comme nous le verrons au chapitre 4.

3.3 Maximisation sur un ensemble de matrices définies positives : un problème d'optimisation convexe

Nous avons vu au chapitre 2, section 2.1.5.3.2 que la borne de Weiss-Weinstein la plus précise s'obtient en maximisant l'ensemble des bornes obtenues $\mathbf{W}(\mathbf{H}, \mathbf{s}) = \mathbf{H}\mathbf{G}^{-1}\mathbf{H}^\top$ par rapport aux points tests $\mathbf{H} \in \mathcal{H}^Q$ et aux variables $s_1, \dots, s_Q \in]0, 1[$. Ceci est valable aussi pour les formes combinées de la borne de Weiss-Weinstein, en particulier celles calculées dans la section 3.2. Pour toutes les bornes considérées, on peut donc exprimer la borne la plus précise \mathbf{B}^* comme la borne supérieure d'un ensemble de matrices $\mathcal{B} \triangleq \{\mathbf{B}(\mathbf{H}, \mathbf{s}) ; \mathbf{H} \in \mathcal{H}^Q, \mathbf{s} \in]0, 1[^Q\} \subset \mathbb{R}^{R \times R}$:

$$\mathbf{B}^* = \sup_{\substack{\mathbf{H} \in \mathcal{H}^Q \\ \mathbf{s} \in]0, 1[^Q}} \mathbf{B}(\mathbf{H}, \mathbf{s}) = \sup \mathcal{B} \quad (3.80)$$

Dans cette section, nous nous intéressons à cette opération de maximisation. Nous avons déjà vu, à la section 2.1.5.3.2, que plusieurs solutions peuvent permettre de rendre cette tâche faisable, en particulier : *i*) se restreindre à \mathbf{H} diagonale, comme nous l'avons fait jusqu'ici, permettant ainsi de réduire considérablement la taille de l'espace \mathcal{H} ; *ii*) choisir $s_q = 1/2, \forall q \in \llbracket 1, Q \rrbracket$ qui est souvent optimal [WW88], [VB07, p.41], [Xu01], ce qui réduit encore la dimension de l'ensemble \mathcal{B} d'un facteur Q . Nous suggérons donc, en première approche, de choisir $s_q = 1/2, \forall q \in \llbracket 1, Q \rrbracket$, et éventuellement ensuite de rechercher d'autres valeurs de s_q possiblement optimales, par exemple en utilisant une grille suffisamment fine de l'intervalle $]0, 1[$.

Il est important de noter que, dans le contexte de l'estimation de points de rupture, l'ensemble \mathcal{H}^Q est dénombrable et fini, puisqu'il est défini à partir de l'ensemble discret \mathcal{K} , support de la loi *a priori* $\pi(\mathbf{k})$.

Pour une valeur de $\mathbf{s} \in]0, 1[^Q$ (par exemple $s_q = 1/2, \forall q \in \llbracket 1, Q \rrbracket$ comme suggéré plus haut), on cherche donc à obtenir la borne supérieure de l'ensemble $\mathcal{B}_{\mathbf{s}} \triangleq \{\mathbf{B}(\mathbf{H}, \mathbf{s}) ; \mathbf{H} \in \mathcal{H}^Q\}$, qui est un sous-ensemble (discret et fini) de \mathbb{S}^R (l'ensemble des matrices carrées symétriques de taille $R \times R$), par rapport à la relation d'ordre partiel \preceq , appelée également ordre partiel de Löwner. Du fait du caractère partiel de cette relation d'ordre, l'unicité de la borne supérieure d'un tel ensemble n'est pas garantie. Pour autant, si une borne supérieure unique existe, alors elle peut

- soit faire partie de l'ensemble $\mathcal{B}_{\mathbf{s}}$, auquel ce sera le plus grand élément ;
- soit ne pas appartenir à l'ensemble $\mathcal{B}_{\mathbf{s}}$, auquel cas il s'agit du plus petit élément de l'ensemble des majorants de $\mathcal{B}_{\mathbf{s}}$.

Autrement, une borne supérieure unique n'existe pas, mais l'ensemble \mathcal{B}_s peut toutefois comprendre plusieurs éléments maximaux⁵. Nous chercherons dès lors à obtenir un élément minimal de l'ensemble de majorants de \mathcal{B}_s qui convient. Formellement, un majorant minimal \mathbf{B} vérifie *i*) $\mathbf{B} \succcurlyeq \mathcal{B}_s$ (majorant), et *ii*) s'il l'on suppose qu'il existe un autre majorant \mathbf{B}' qui lui est inférieur, i.e., tel que $\mathbf{B} \succcurlyeq \mathbf{B}' \succcurlyeq \mathcal{B}_s$, alors on a nécessairement $\mathbf{B}' = \mathbf{B}$ (minimal). Ici encore, \mathbf{B} peut ne pas être unique sans contraintes supplémentaires (s'il est unique, il s'agit de l'unique majorant minimal de \mathcal{B}_s , c'est-à-dire sa borne supérieure).

Malgré ces remarques, en s'appuyant sur [LRNM10, Section III.D.], une manière d'obtenir un majorant minimal de \mathcal{B}_s qui convient, et défini de manière unique, est la suivante. À chaque matrice $\mathbf{A} \in \mathbb{S}_{++}^R$ (l'ensemble des matrices symétriques définies positives de taille $R \times R$), on peut associer un hyper-ellipsoïde centré, défini comme l'ensemble $\mathcal{E}(\mathbf{A}) \triangleq \{\mathbf{y} \in \mathbb{R}^R \mid \mathbf{y}^\top \mathbf{A}^{-1} \mathbf{y} \leq 1\}$. D'après [LRNM10, Lemme 3], on peut montrer que, quelles que soient les matrices définies positives \mathbf{A} et \mathbf{A}' , on a l'équivalence suivante : $\mathbf{A} \preccurlyeq \mathbf{A}' \Leftrightarrow \mathcal{E}(\mathbf{A}) \subseteq \mathcal{E}(\mathbf{A}')$. Par conséquent, pour une famille finie de matrices $\{\mathbf{A}_i\}_{i \in I}$ (où I est un ensemble fini), on peut trouver l'hyper-ellipsoïde de volume minimal, noté $\mathcal{E}(\mathbf{A}_{LJ})$, qui contient l'union de tous les hyper-ellipsoïdes $\mathcal{E}(\mathbf{A}_i)$ associés aux matrices \mathbf{A}_i . L'ellipsoïde de volume minimal $\mathcal{E}(\mathbf{A}_{LJ})$ est appelé ellipsoïde de Löwner-John pour $\bigcup_{i \in I} \mathcal{E}(\mathbf{A}_i)$, et on peut montrer que \mathbf{A}_{LJ} est un majorant minimal de l'ensemble $\{\mathbf{A}_i \mid i \in I\}$ (voir [LRNM10, Theorem 4]). Ainsi, pour approcher la borne la plus précise \mathbf{B}^* , nous calculons un majorant minimal de l'ensemble \mathcal{B}_s , la matrice \mathbf{B}_{LJ} associée à l'ellipsoïde de Löwner-John, qui contient $\bigcup_{\mathbf{H} \in \mathcal{H}^Q} \mathcal{E}(\mathbf{B}(\mathbf{H}, \mathbf{s}))$. De plus, pour réduire encore le coût de calcul, on peut préalablement restreindre l'ensemble \mathcal{B}_s à \mathcal{B}'_s uniquement composé des éléments maximaux de \mathcal{B}_s .

Or, on peut montrer que la recherche de \mathbf{B}_{LJ} correspond à la résolution du problème d'optimisation convexe suivant (voir [BV04, p. 411]) :

$$\begin{array}{ll}
 \text{minimiser} & \ln\left(\det\left(\mathbf{B}^{\frac{1}{2}}\right)\right) \\
 \text{soumis à} & b_1 \geq 0, b_2 \geq 0, \dots, b_J \geq 0, \\
 & \begin{bmatrix} \mathbf{B}^{-1} - b_i \mathbf{B}(\mathbf{H}_i, \mathbf{s})^{-1} & \mathbf{0} \\ \mathbf{0} & b_i - 1 \end{bmatrix} \preceq \mathbf{0} \\
 & (i = 1, \dots, J),
 \end{array} \tag{3.81}$$

où $\mathbf{B}(\mathbf{H}_i, \mathbf{s}) \in \mathcal{B}'_s$ et $J \triangleq \text{Card}(\mathcal{B}'_s)$. Ce problème peut être résolu efficacement en utilisant une procédure de programmation semi-définie, telle que celles fournies par le package « CVX » [GB08].

5. Un élément $\mathbf{A} \in \mathcal{B}_s$ est dit maximal s'il n'existe aucun autre élément $\mathbf{A}' \in \mathcal{B}_s$ tel que $\mathbf{A}' \succ \mathbf{A}$.

Chapitre 4

Applications en traitement du signal

Dans ce chapitre, nous séparons les résultats selon l'hypothèse faite sur le vecteur de paramètres $\boldsymbol{\eta}$ de la distribution du signal observé : connu (section 4.1) ou inconnu (section 4.2). On rappelle que ce vecteur est la concaténation des vecteurs de paramètres $\boldsymbol{\eta}_0, \boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_Q$ de la distribution du signal observé sur chaque segment. Nous donnons ici des expressions analytiques (ou quasi analytiques) des bornes inférieures de l'EQM dont nous avons obtenu la forme générale dans le chapitre précédent. Nous nous intéressons ici à deux distributions très fréquemment rencontrées en traitement du signal : la loi normale (signaux gaussiens) et la loi de Poisson (signaux poissonniens). Nous verrons aussi, plus rapidement, le cas de signaux suivant une loi exponentielle. L'idée ici est de donner des formules de « type Slepian-Bangs » pour permettre une implémentation simple et rapide des bornes considérées. Enfin, une étude de la robustesse vis-à-vis de la distribution *a priori* des positions des ruptures sera proposée.

4.1 Cas où les paramètres de la distribution du signal observé sur chaque segment sont connus

Nous nous intéressons d'abord au cas (idéal) où les vecteurs des paramètres du signal $(\boldsymbol{\eta}_q)_{q=1,\dots,Q}$ sur chaque segment sont supposés connus. Dans ce cas seulement, nous pouvons comparer la borne de Weiss-Weinstein calculée au chapitre 3, section 3.1, à la borne de Barankin de [FT03] (dans le cas d'une seule rupture, présentée au chapitre 2, section 2.3.2.1) et [LRNM10] (dans le cas de plusieurs ruptures, présentée au chapitre 2, section 2.3.2.2).

Rappelons-nous toutefois que la borne de Weiss-Weinstein est une borne bayésienne, tandis que la borne de Barankin est une borne « déterministe » (i.e., pour l'estimation de paramètres supposés déterministes). Elles ne portent donc pas toutes les deux exactement sur la même mesure de performance : nous avons vu que les bornes bayésiennes portent sur l'EQM *globale* (EQMG) tandis que les bornes déterministes reposent sur l'EQM *locale*. Or, nous avons également vu que l'EQMG correspond au risque intégré pour le coût quadratique (voir chapitre 2, section 2.1.3.1), c'est-à-dire à la moyenne de l'EQM locale par rapport à la loi *a priori* sur le vecteur de paramètres inconnu. L'EQMG, et donc les bornes bayésiennes, prennent ainsi en compte toutes les valeurs possibles du vecteur de paramètres, et par conséquent ne dépendent pas de la vraie valeur du vecteur de paramètres, d'où l'adjectif « global ». En revanche, les bornes déterministes dépendent en général de la vraie valeur du paramètre, c'est pourquoi on peut parler de bornes « locales ».

Néanmoins, en termes de simulation, il n'est pas difficile d'utiliser une borne déterministe afin de refléter les performances en terme d'EQM globale : on y parvient par la même opération que celle qui permet de passer de l'EQM locale à l'EQM globale, à savoir en moyennant la première

par rapport à la loi *a priori*.

Dans le cas qui nous intéresse, à savoir le problème de l'estimation de ruptures, on peut donc obtenir la borne de Barankin dite « globale », ou « moyennée », que l'on notera **BBG**, en moyennant la borne de Barankin classique (celle exposée au chapitre 2, section 2.1.5.2.2) par rapport à la loi *a priori* sur \mathbf{k} :

$$\mathbf{BBG} \triangleq \sup_{\mathbf{H} \in \mathcal{H}^Q} \mathbf{H} \mathbb{E}_{\mathbf{k}} \{ [\mathbf{BIM}(\mathbf{k}, \mathbf{H})]^{-1} \} \mathbf{H}^\top, \quad (4.1)$$

où $\mathbf{BIM}(\mathbf{k}, \mathbf{H})$ est donnée au chapitre 2, section 2.3.2.2.

4.1.1 Cas d'un signal comportant un seul point de rupture

Dans cette section, nous supposons que le signal considéré ne comporte qu'un seul point de rupture. Le modèle (3.1) se réduit donc à

$$\begin{cases} \mathbf{x}_n \sim p_{\boldsymbol{\eta}_0}(\mathbf{x}_n) & \text{pour } 1 \leq n \leq k_1, \\ \mathbf{x}_n \sim p_{\boldsymbol{\eta}_1}(\mathbf{x}_n) & \text{pour } k_1 + 1 \leq n \leq N, \end{cases} \quad (4.2)$$

où le seul paramètre à estimer est k_1 . Dans ce cas, la distribution *a priori* sur le paramètre k_1 se réduit à la distribution uniforme discrète sur $[[d, D]]$. Dans la suite de cette section, nous prendrons $d = 1$ et $D = N - 1$.

4.1.1.1 Signaux gaussiens

Supposons que les observations \mathbf{x}_n soient gaussiennes et spatialement décorréliées avant et après la rupture k_1 , i.e., en notant leur moyenne \mathbf{m}_0 (avant la rupture) et \mathbf{m}_1 (après la rupture) et leur matrice de covariance $\sigma_0^2 \mathbf{I}_M$ (avant la rupture) et $\sigma_1^2 \mathbf{I}_M$ (après la rupture), leur distribution est donnée, pour $q = 0$ ou 1 , par

$$p_{\mathbf{m}_q, \sigma_q^2}(\mathbf{x}_n) = \frac{1}{(2\pi\sigma_q^2)^{M/2}} \exp\left\{-\frac{\|\mathbf{x}_n - \mathbf{m}_q\|^2}{2\sigma_q^2}\right\}. \quad (4.3)$$

Le vecteur $\boldsymbol{\eta}_q$, $q \in \{0, 1\}$, est donc composé de la moyenne et de la variance des observations, soit $\boldsymbol{\eta}_q = (\mathbf{m}_q; \sigma_q^2)$. On peut donc avoir un changement dans la moyenne \mathbf{m}_q , ou dans la variance σ_q^2 , ou bien dans les deux. Pour chaque type de changement, on définit une grandeur appelée « quantité de changement », ou parfois « rapport signal sur bruit » qui permet de quantifier la différence entre les distributions du signal sur deux segments distincts. Ainsi, pour $q \in [[1, Q]]$, si l'on définit la quantité de changement dans la moyenne d'un signal gaussien univarié (i.e., pour $M = 1$) $\omega_{q-1, q}^m$, entre le q -ième segment (de moyenne m_{q-1}) et le $(q + 1)$ -ième segment (de moyenne m_q), comme

$$\omega_{q-1, q}^m \triangleq \frac{(m_q - m_{q-1})^2}{\sigma_{q-1}^2}, \quad (4.4)$$

on peut également définir la version pour un signal multivarié (i.e., $M > 1$), $\omega_{q-1, q}^{\mathbf{m}}$, entre le q -ième segment (de moyenne $\mathbf{m}_{q-1} = [m_{1, q-1}, m_{2, q-1}, \dots, m_{M, q-1}]^\top$) et le $(q + 1)$ -ième segment (de moyenne $\mathbf{m}_q = [m_{1, q}, m_{2, q}, \dots, m_{M, q}]^\top$), comme

$$\omega_{q-1, q}^{\mathbf{m}} \triangleq \frac{\|\mathbf{m}_q - \mathbf{m}_{q-1}\|^2}{\sigma_{q-1}^2} = \frac{\sum_{i=1}^M (m_{i, q} - m_{i, q-1})^2}{\sigma_{q-1}^2} = \sum_{i=1}^M \omega_{q-1, q}^{m_i}, \quad (4.5)$$

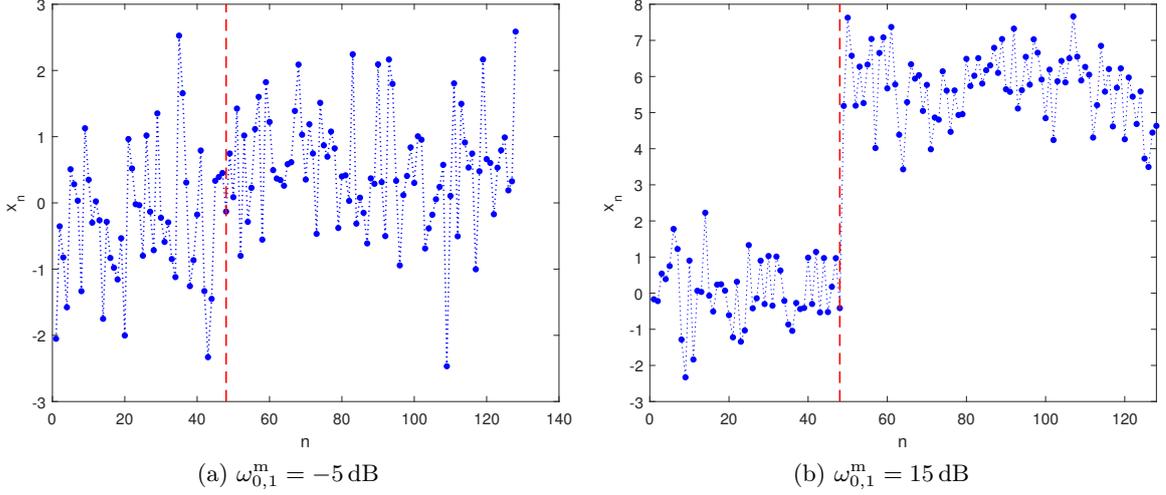


FIGURE 4.1 – Deux exemples de signaux gaussiens comportant un point de rupture dans la moyenne à $n = 48$ (ligne verticale en pointillés rouges), pour $N = 128$ observations.

où $\omega_{q-1,q}^{m_i}$ représente la quantité de changement univariée sur la i -ième composante de \mathbf{x}_n uniquement. On remarque que (4.5) illustre l'apport, en terme de quantité de changement, de chaque composante supplémentaire du signal.

En ce qui concerne un signal gaussien dont la variance peut changer d'un segment à un autre, on définit la quantité de changement en variance $\omega_{q-1,q}^v$, entre le q -ième et le $(q+1)$ -ième segment, comme

$$\omega_{q-1,q}^v \triangleq \frac{\sigma_q^2}{\sigma_{q-1}^2}. \quad (4.6)$$

Dans le cas présent, on suppose que $Q = 1$, donc on s'intéresse aux quantités $\omega_{0,1}^m$ et $\omega_{0,1}^v$.

Sur la figure 4.1, nous montrons deux exemples de signaux gaussiens comportant un point de rupture dans la moyenne à $n = 48$, dans le cas $M = 1$, pour un nombre total d'observations $N = 128$, l'un pour une quantité de changement de $\omega_{0,1}^m = -5$ dB, l'autre pour une quantité de changement $\omega_{0,1}^m = 15$ dB. Pour ces exemples nous avons choisi $m_0 = 0$ et $\sigma_0^2 = \sigma_1^2 = 1$.

Sur la figure 4.2, nous montrons deux exemples de signaux gaussiens comportant un point de rupture dans la variance à $n = 48$, dans le cas $M = 1$, pour un nombre total d'observations $N = 128$, l'un pour une quantité de changement de $\omega_{0,1}^v = 5$ dB, l'autre pour une quantité de changement $\omega_{0,1}^v = 20$ dB. Pour ces exemples nous avons choisi $m_0 = m_1 = 1$ et $\sigma_0^2 = 1$.

4.1.1.1.1 Expression de $\rho_{q-1,q}(s)$. L'expression de $\rho_{q-1,q}(s)$ pour des données gaussiennes changeant de moyenne et/ou de variance s'obtient en injectant les distributions des observations sur chaque segment, données par (4.3), dans (3.12). Après des manipulations calculatoires, impliquant notamment l'intégrale d'une densité gaussienne de moyenne $\frac{s\sigma_1^2\mathbf{m}_0 + (1-s)\sigma_0^2\mathbf{m}_1}{s\sigma_1^2 + (1-s)\sigma_0^2}$ et de matrice de covariance $\frac{\sigma_0^2\sigma_1^2}{s\sigma_1^2 + (1-s)\sigma_0^2} \mathbf{I}_M$, il vient

$$\rho_{q-1,q}(s) = \left[\frac{(\omega_{q-1,q}^v)^s}{s\omega_{q-1,q}^v + 1 - s} \right]^{\frac{M}{2}} \exp \left\{ -\frac{s(1-s)\omega_{q-1,q}^m}{2(s\omega_{q-1,q}^v + 1 - s)} \right\}. \quad (4.7)$$

Dans le cas présent où $Q = 1$, seul $\rho_{0,1}$ intervient : il suffit donc de remplacer q par 1 dans (4.7).

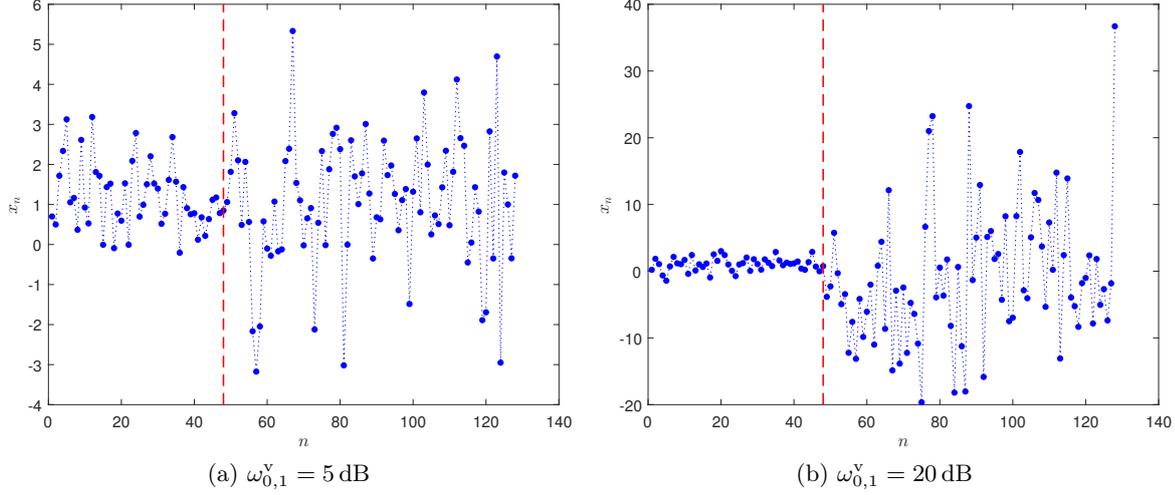


FIGURE 4.2 – Deux exemples de signaux gaussiens comportant un point de rupture dans la variance à $n = 48$ (ligne verticale en pointillés rouges), pour $N = 128$ observations.

4.1.1.1.2 Résultats numériques. Nous cherchons maintenant à illustrer la précision de la borne calculée précédemment par des résultats numériques. On étudie séparément les cas du changement dans la moyenne uniquement, et du changement dans la variance uniquement. On s'intéresse ici à des séries univariées ($M = 1$).

Changement de moyenne. Dans le cas d'un changement de moyenne, on a $\omega_{0,1}^v = 1$, et $\rho_{0,1}(s)$ se réduit donc à

$$\rho_{0,1}(s) = \exp\left\{-\frac{s(1-s)\omega_{0,1}^m}{2}\right\}. \quad (4.8)$$

Par conséquent, la borne de Weiss-Weinstein calculée dans le chapitre 3, section 3.1 se réduit à

$$\begin{aligned} \text{BWW} &= \sup_{h,s} W(h,s) \\ &= \sup_{h,s} \frac{h^2 \left(1 - \frac{|h|}{N-1}\right)^2 e^{-|h|s(1-s)\omega_{0,1}^m}}{\left(1 - \frac{|h|}{N-1}\right) \left[e^{-|h|s(1-2s)\omega_{0,1}^m} + e^{-|h|(2s-1)(1-s)\omega_{0,1}^m} \right] - 2 \left(1 - \frac{2|h|}{N-1}\right) e^{-|h|s(1-s)\omega_{0,1}^m}}. \end{aligned} \quad (4.9)$$

En différentiant cette dernière expression, on peut obtenir la condition nécessaire pour un maximum local par rapport à s en résolvant $\frac{\partial W(h,s)}{\partial s}\Big|_{s^*} = 0$. Après calcul, cette condition conduit à

$$\begin{aligned} &(2s^* - 1) \left(1 - \frac{2|h|}{N-1}\right) \exp\{-|h|s^{*2}\omega_{0,1}^m\} \\ &- \left[(2s^* - 1) \left(1 - \frac{2|h|}{N-1}\right) + (s^* - 1) \left(1 - \frac{|h|}{N-1}\right) \right] \exp\{-|h|(2s^* - 1)\omega_{0,1}^m\} - s^* \left(1 - \frac{|h|}{N-1}\right) = 0. \end{aligned} \quad (4.10)$$

On peut voir que $s^* = 1/2$ est solution de cette équation. Sur la figure 4.3, nous avons tracé le membre de gauche de (4.10) en fonction de s , pour $|h| = 1$ et $\omega_{0,1}^m = 0 \text{ dB}$. Dans ce cas, il est

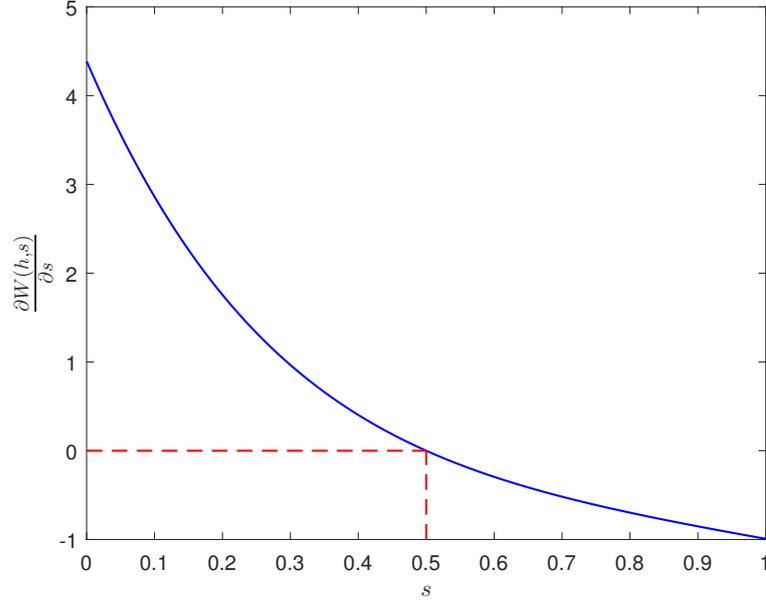


FIGURE 4.3 – Tracé de $\frac{\partial W(h,s)}{\partial s}$ en fonction de s , dans le cas $|h| = 1$ et $\omega_{0,1}^m = 0$ dB (trait bleu continu). Il est clair que la seule solution d'annulation de la fonction est $s^* = 1/2$ (pointillés rouges).

donc clair que $\frac{\partial W(h,s)}{\partial s}$ ne s'annule qu'une seule fois, et que $s^* = 1/2$ est bien la seule solution qui conduit à l'optimum.

Sur la figure 4.4, nous traçons l'évolution de la borne de Weiss-Weinstein et de la borne de Barankin globale (obtenue à partir des résultats de [FT03] et des explications fournies au début de la section 4.1), en fonction de la quantité de changement $\omega_{0,1}^m$, dans le cas de $N = 128$ observations. Nous traçons également les racines (carrées) des EQM globales (REQMG) approchées de deux estimateurs bayésiens courants, à savoir les estimateurs du MAP et MMSE (que nous avons présentés à la section 2.1.3.2), obtenues après $N_{MC} = 1000$ simulations de Monte-Carlo.

On constate, comme attendu que la REQMG de l'estimateur MMSE est inférieure à celle du MAP, en particulier pour des quantités de changement faibles. Plus cette dernière augmente, plus ces deux estimateurs tendent à être équivalents en terme d'EQMG.

On remarque également que la borne de Weiss-Weinstein est plus précise que la borne de Barankin globale : par exemple, à $\omega_{0,1}^m = -5$ dB, la BWW indique une REQMG minimale d'environ 7 échantillons, tandis que la BBG indique une REQMG minimale d'environ 2,5 échantillons. Or, la véritable REQMG minimale pour cette quantité de changement, qui est approchée par la REQMG empirique de l'estimateur MMSE, est d'environ 10 échantillons.

Pour autant, il subsiste un léger écart entre la BWW et la REQMG des estimateurs de Bayes considérés, même pour de grandes quantités de changement. Nous rapprochons ce phénomène des résultats de non-consistance et de non efficacité des estimateurs de paramètres discrets obtenus par Choirat et Seri [CS12], pour des modèles plus réguliers que ceux considérés ici. Il n'est donc pas étonnant d'obtenir ce type de résultats dans le cas de signaux comportant une ou plusieurs ruptures.

À faible quantité de changement, on remarque que la borne de Barankin globale devient supérieure aux REQMG des estimateurs de Bayes, et perd donc son caractère de borne *inférieure* de l'EQM. Ceci est dû au fait que les conditions d'application de la borne, comme (2.57b), ne sont pas satisfaites dans cette situation. Au contraire, la borne de Weiss-Weinstein, qui a des conditions

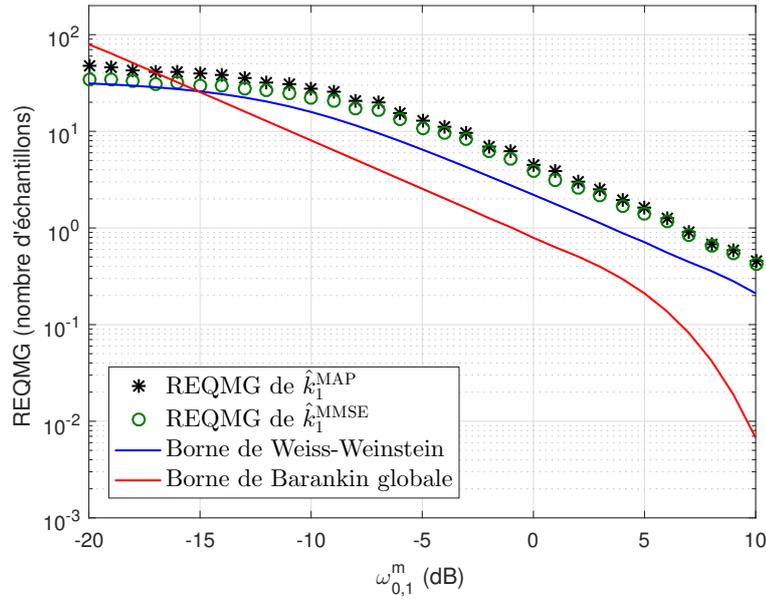


FIGURE 4.4 – Évolution de la borne de Weiss-Weinstein (trait noir continu), de la borne de Barankin globale (trait rouge continu) et des racines (carrés) des EQM globales (REQMG) des estimateurs du MAP \hat{k}_1^{MAP} (étoiles bleues) et MMSE \hat{k}_1^{MMSE} (cercles verts) en fonction de la quantité de changement $\omega_{0,1}^m$, dans le cas de $Q = 1$ unique changement dans la moyenne de $N = 128$ données gaussiennes.

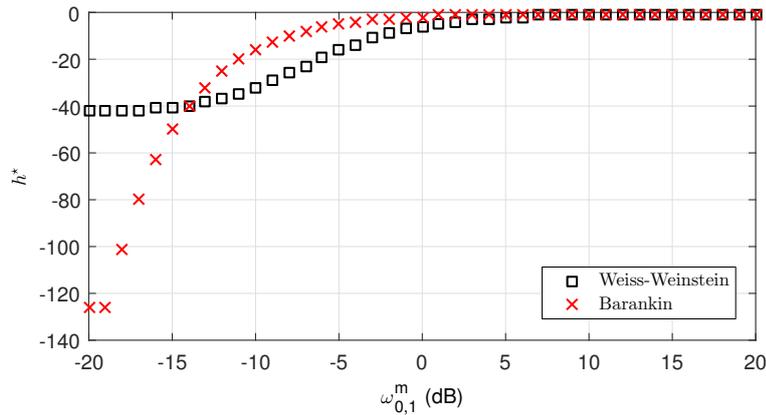


FIGURE 4.5 – Évolution de la valeur h^* optimale des points test conduisant aux bornes de Weiss-Weinstein (carrés noirs), et de Barankin globale (croix rouges) les plus précises, en fonction de la quantité de changement $\omega_{0,1}^m$, dans le cas de $Q = 1$ unique changement dans la moyenne de $N = 128$ données gaussiennes.

de régularité moins stricte, conserve son caractère de borne inférieure, et rend même très bien compte du comportement de l'estimation dans la région de faible quantité de changement.

Dans la figure 4.5, nous montrons l'évolution des valeurs des points test h^* conduisant aux bornes de Weiss-Weinstein et de Barankin globale les plus précises. On constate que ces valeurs dépendent fortement de la valeur de la quantité de changement, mais également qu'elles se rapprochent de zéro au fur et à mesure que la quantité de changement augmente. Dans le cas présent, comme l'ensemble des points test \mathcal{H} est discret, les valeurs de h^* se rapprochent en

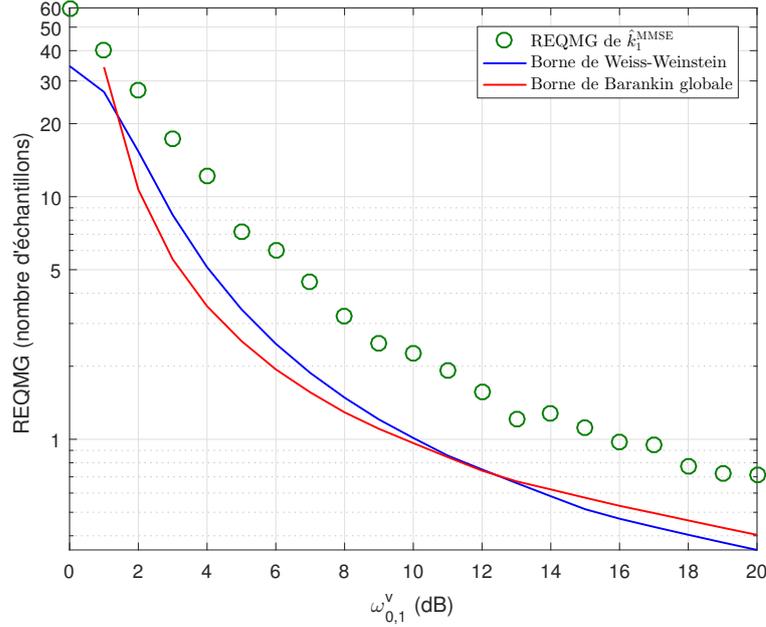


FIGURE 4.6 – Évolution de la borne de Weiss-Weinstein (trait noir continu), de la borne de Barankin globale (trait rouge continu) et des EQM globales des estimateurs du MAP \hat{k}_1^{MAP} (étoiles bleues) et MMSE \hat{k}_1^{MMSE} (cercles verts) en fonction de la quantité de changement $\omega_{0,1}^v$, dans le cas de $Q = 1$ unique changement dans la variance de $N = 128$ données gaussiennes.

réalité de 1, en valeur absolue (rappelons également que $0 \notin \mathcal{H}$).

Changement de variance. Dans le cas d'un changement dans la variance d'un signal univarié ($M = 1$), on a $\alpha_{0,1}^m = 0$, et d'après (4.7), $\rho_{0,1}(s)$ se réduit donc à

$$\rho_{0,1}(s) = \sqrt{\frac{(\omega_{0,1}^v)^s}{s\omega_{0,1}^v + 1 - s}}. \quad (4.11)$$

De la même façon que précédemment, on obtient directement la borne de Weiss-Weinstein associée en injectant la valeur de $\rho_{0,1}(s)$ ainsi obtenue dans l'expression de A_1 donnée par (3.10).

Les résultats numériques se rapportant à ce cas sont donnés en figure 4.6, où nous ne traçons que la REQMG empirique de l'estimateur MMSE de k_1 (toujours obtenue après $N_{\text{MC}} = 1000$ simulations de Monte-Carlo), que nous comparons aux bornes de Barankin globale et de Weiss-Weinstein associées. L'écart de précision entre la borne de Weiss-Weinstein et la borne de Barankin globale est ici moins flagrant que dans le cas d'un changement de moyenne. On observe en effet que pour des quantités de changement $1 \text{ dB} < \omega_{q-1,q}^v < 12 \text{ dB}$, on a $\text{BWW} > \text{BBG}$, tandis que l'ordre s'inverse pour des quantités de changement $\omega_{q-1,q}^v$ plus élevées que 12 dB. On relève qu'ici aussi, de la même manière que dans le cas d'un changement de moyenne, aucune des deux bornes n'est atteinte, même pour de fortes quantités de changement.

4.1.1.2 Signaux poissonniens

Nous nous intéressons maintenant au cas de signaux distribués selon une loi de Poisson, dont le paramètre connaît un saut abrupt à l'instant k_1 , et passe ainsi de la valeur λ_0 à la valeur λ_1 .

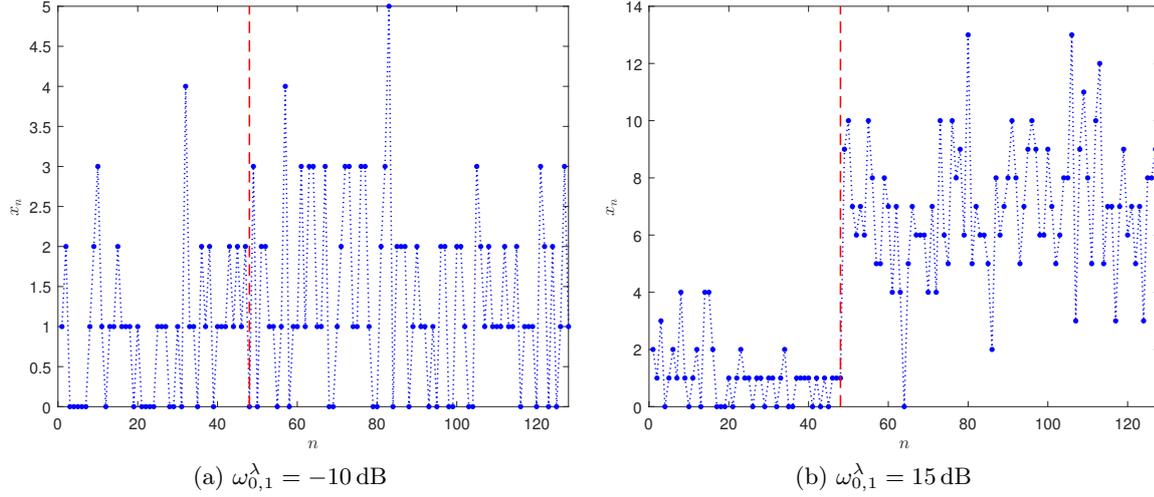


FIGURE 4.7 – Deux exemples de signaux poissonniens comportant un point de rupture dans le paramètre λ_q à $n = 48$ (ligne verticale en pointillés rouges), pour $N = 128$ observations.

On considère dans ce cas une série temporelle univariée (i.e., $M = 1$), et les distributions des observations sur chaque segment sont donc données, pour $q \in \{0, 1\}$, par

$$p_{\lambda_q}(x_n) = \Pr(x_n = x_n) = \frac{\lambda_q^{x_n}}{x_n!} \exp\{-\lambda_q\}. \quad (4.12)$$

Nous n'avons donc cette fois qu'un seul paramètre (λ_q) susceptible de varier d'un segment à l'autre. On définit la quantité de changement $\omega_{q-1,q}^\lambda$ associée à ce type de rupture comme

$$\omega_{q-1,q}^\lambda \triangleq \frac{(\lambda_q - \lambda_{q-1})^2}{\lambda_{q-1}^2}, \quad (4.13)$$

avec $q \in \{0, 1\}$ dans le cas de $Q = 1$ seul point de rupture. Nous donnons en figures 4.7a et 4.7b deux exemples de ce type de données pour deux valeurs de la quantité de changement $\omega_{0,1}^\lambda$ différentes, à savoir -10 dB et 15 dB, respectivement.

4.1.1.2.1 Expression de $\rho_{q-1,q}(s)$. L'expression de $\rho_{q-1,q}(s)$, pour des données distribuées selon une loi de Poisson, s'obtient de la même manière que dans le cas gaussien, c'est-à-dire en injectant l'expression de la distribution des observations sur chaque segment, donnée par (4.12), dans (3.12). Notons que dans ce cas, les observations x_n représentent des variables aléatoires discrètes, ce qui signifie que dans (3.12), l'intégrale sur Ω' par rapport aux observations se transforme ici en somme discrète sur \mathbb{N} . On a donc

$$\begin{aligned} \rho_{q-1,q}(s) &= \sum_{x=0}^{+\infty} p_{\lambda_{q-1}}^s(x) p_{\lambda_q}^{1-s}(x) \\ &= \exp\{-s\lambda_{q-1} - (1-s)\lambda_q\} \sum_{x=0}^{+\infty} \frac{(\lambda_{q-1}^s \lambda_q^{1-s})^x}{x!} \\ &= \exp\{-s\lambda_{q-1} - (1-s)\lambda_q + \lambda_{q-1}^s \lambda_q^{1-s}\}. \end{aligned} \quad (4.14)$$

Cette quantité ne peut pas s'exprimer directement en fonction de la quantité de changement $\omega_{q-1,q}^\lambda$ définie dans (4.13), comme c'est le cas pour des signaux gaussiens (section 4.1.1.1).

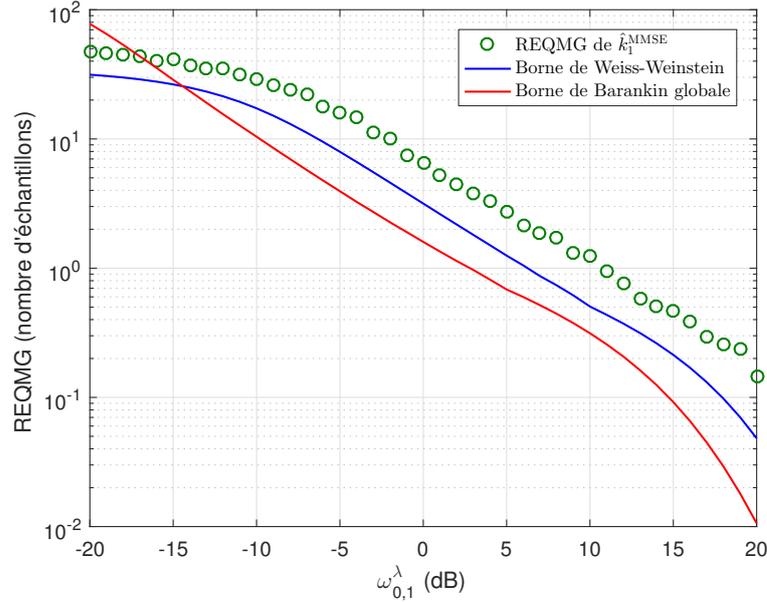


FIGURE 4.8 – Cas de $Q = 1$ changement dans le paramètre λ_0 de $N = 128$ observations poissonniennes. Évolution de la borne de Weiss-Weinstein (trait bleu continu), de la borne de Barankin globale (trait rouge continu) et de l'EQM globale de l'estimateur MMSE \hat{k}_1^{MMSE} (cercles verts) en fonction de la quantité de changement $\omega_{0,1}^\lambda$.

4.1.1.2.2 Résultats numériques. Les résultats numériques obtenus dans le cas de signaux poissonniens sont représentés en figure 4.8. Les conditions de simulation sont les mêmes que dans le cas gaussien, à savoir $N = 128$ observations, $N_{\text{MC}} = 1000$ simulations de Monte-Carlo pour obtenir la REQMG empirique de l'estimateur MMSE de k_1 .

On observe que la borne de Weiss-Weinstein est, ici également, plus précise que la borne de Barankin globale, et ce sur une grande plage de quantité de changement $\omega_{q-1,q}^\lambda > -14$ dB. Dans ce cas aussi, cet écart de précision entre les deux bornes n'est pas négligeable, en particulier pour des faibles quantités de changement. Par exemple, pour $\omega_{q-1,q}^\lambda = -10$ dB, la BWW donne une REQMG minimale plus proche de celle du MMSE, d'environ 7 échantillons, par rapport à celle indiquée par la BBG. L'écart entre la BWW et la REQMG du MMSE est inférieur à 3 échantillons pour $\omega_{q-1,q}^\lambda > 0$ dB, et inférieur à 1 échantillon pour $\omega_{q-1,q}^\lambda > 10$ dB. La BBG perd sa validité pour de très faibles quantités de changement, inférieures à -17 dB.

4.1.1.3 Signaux suivant une loi exponentielle

Pour terminer cette analyse du cas d'un unique point de rupture dans la fenêtre d'observation, on s'intéresse au cas de données suivant une loi exponentielle, dont le paramètre Λ_q change de façon abrupte à l'instant k_1 . Nous nous plaçons dans le cas univarié ($M = 1$) ici également. La distribution des observations sur chaque segment est par conséquent donnée, pour $q \in \{0, 1\}$, par

$$p_{\Lambda_q}(x_n) = \Lambda_q \exp\{-\Lambda_q x_n\}. \quad (4.15)$$

La quantité de changement $\omega_{q-1,q}^\Lambda$ associée à une rupture dans la valeur du paramètre Λ_q est dans ce cas définie comme le rapport

$$\omega_{q-1,q}^\Lambda \triangleq \frac{\Lambda_{q-1}}{\Lambda_q}. \quad (4.16)$$

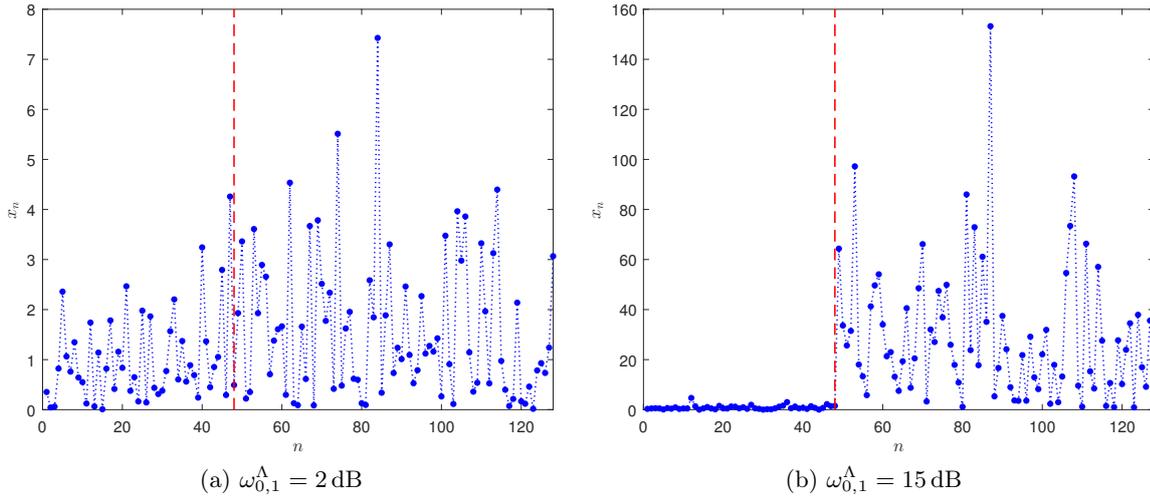


FIGURE 4.9 – Deux exemples de signaux distribués selon une loi exponentielle comportant un point de rupture dans le paramètre Λ_q à $n = 48$ (ligne verticale en pointillés rouges), pour $N = 128$ observations.

Comme dans les cas précédents, nous donnons deux exemples de signaux ayant ce type de rupture en figure 4.9, pour des valeurs de quantité de changement $\omega_{q-1,q}^\Lambda$ égales à 2 dB (figure 4.9a) et 15 dB (figure 4.9b).

4.1.1.3.1 Expression de $\rho_{q-1,q}(s)$. De la même manière que dans les cas précédents, en injectant (4.15) dans (3.12), on obtient l'expression du terme clé $\rho_{q-1,q}(s)$ intervenant dans l'expression de la borne de Weiss-Weinstein. On obtient, dans le cas de données suivant une loi exponentielle,

$$\rho_{q-1,q}(s) = \frac{\Lambda_{q-1}^s \Lambda_q^{1-s}}{s\Lambda_{q-1} + (1-s)\Lambda_q} = \frac{(\omega_{q-1,q}^\Lambda)^s}{s\omega_{q-1,q}^\Lambda + 1 - s}. \quad (4.17)$$

4.1.1.3.2 Résultats numériques. En se plaçant dans les mêmes conditions opératoires que dans les cas précédents ($N = 128$ observations, $N_{MC} = 1000$ simulations de Monte-Carlo) mais cette fois pour des données suivant une loi exponentielle, on obtient la borne de Weiss-Weinstein, la borne de Barankin globale, et la REQMG approchée de l'estimateur MMSE représentées sur la figure 4.10. L'allure des courbes est similaire à celui du cas d'un changement de variance dans des données gaussiennes. Cependant, ici, la BBG se comporte bien différemment pour de fortes quantités de changement, et ne passe dans ce cas pas au-dessus de la BWW. Au contraire, le comportement numérique de la BBG la fait s'éloigner progressivement du comportement réel de l'estimation (ici donné par le MMSE) à mesure que $\omega_{0,1}^\Lambda$ augmente, et la rend même impossible à calculer pour $\omega_{0,1}^\Lambda > 17$ dB. La borne de Weiss-Weinstein conserve pour sa part un comportement numérique plus régulier, et rend compte de l'EQM minimale de manière plus précise, sur une plage plus large de quantités de changement.

Dans le cas $Q = 1$ d'une unique rupture, l'estimateur MMSE possède une complexité de calcul largement raisonnable, et ne rendent pas prohibitives les nombreuses simulations de Monte-Carlo nécessaires pour obtenir son EQM associée. En revanche, lorsque le nombre de ruptures s'accroît, il devient plus coûteux, et les bornes inférieures de l'EQM telles que celles proposées ici prennent alors tout leur intérêt.

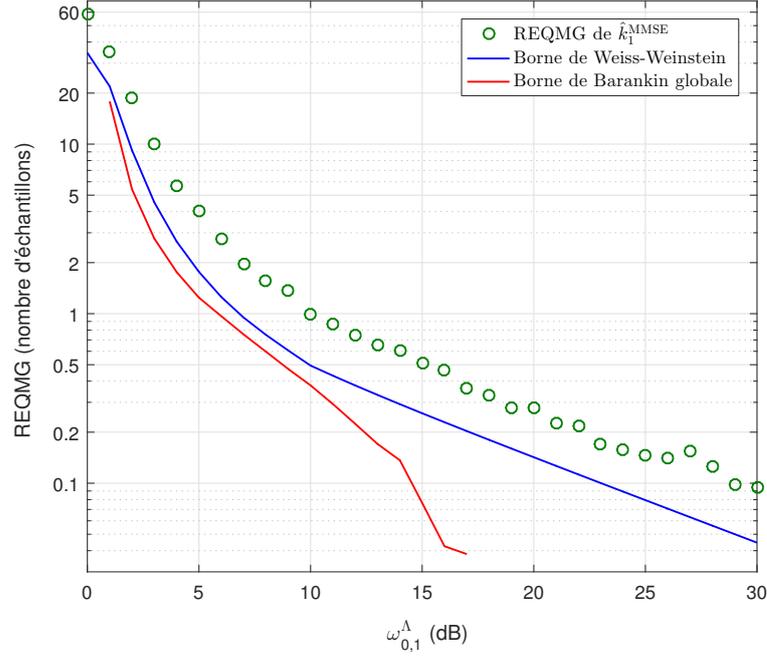


FIGURE 4.10 – Évolution de la borne de Weiss-Weinstein (trait bleu continu), de la borne de Barankin globale (trait rouge continu) et de l'EQM globale de l'estimateur MMSE \hat{k}_1^{MMSE} (cercles verts) en fonction de la quantité de changement $\omega_{0,1}^\Lambda$.

4.1.2 Cas de signaux comportant de multiples points de rupture

Nous revenons maintenant au cas d'un nombre de ruptures Q quelconque donné. Le modèle considéré est donc celui donné par (3.1). Nous nous intéressons à des signaux gaussiens dans la section 4.1.2.1 et étudions l'influence de diverses lois *a priori* du type de (3.18). Dans la section 4.1.2.2, nous considérons des signaux poissonniens.

4.1.2.1 Signaux gaussiens

4.1.2.1.1 Expressions de $r_q(s, s')$ et $R_q(s, s')$ La seule quantité restant à calculer pour obtenir la matrice $\mathbf{G}(\mathbf{H}, \mathbf{s})$ de la borne de Weiss-Weinstein donnée à la section 3.1.3 est $R_q(s, s')$. D'après (3.16), celle-ci s'obtient à partir de $r_q(s, s')$. En injectant (4.3) dans (3.17) et après des manipulations du même type que celles menant à (4.7), on obtient¹, pour $q \in \llbracket 1, Q-1 \rrbracket$,

$$r_q(s, s') = \left[\frac{(\omega_{q-1,q+1}^v)^s (\omega_{q,q+1}^v)^{s'}}{s\omega_{q-1,q+1}^v + s'\omega_{q,q+1}^v + 1 - s - s'} \right]^{\frac{M}{2}} \times \exp \left\{ -\frac{s s' \omega_{q,q+1}^v \omega_{q-1,q}^m + s(1-s-s')\omega_{q-1,q+1}^m + s'(1-s-s')\omega_{q,q+1}^m}{2(s\omega_{q-1,q+1}^v + s'\omega_{q,q+1}^v + 1 - s - s')} \right\}, \quad (4.18)$$

puis, à partir de (4.7) et (4.18) que l'on injecte dans (3.16), il vient finalement

$$R_q(s, s') = \left[\frac{s\omega_{q-1,q+1}^v + (s' - s)\omega_{q,q+1}^v + 1 - s'}{(s\omega_{q-1,q}^v + 1 - s)(s'\omega_{q,q+1}^v + 1 - s')} \right]^{\frac{M}{2}}$$

1. L'expression de $R_q(s, s')$ donnée ici corrige la typo dans le papier [BREC17].

$$\begin{aligned} & \times \exp \left\{ - \frac{s(1-s')}{2(s\omega_{q-1,q+1}^v + (s'-s)\omega_{q,q+1}^v + 1-s')} \right. \\ & \quad \left. \times \left(\frac{s\omega_{q-1,q+1}^v + 1-s}{s\omega_{q-1,q}^v + 1-s} \omega_{q-1,q}^m + \frac{s'\omega_{q-1,q+1}^v + 1-s'}{s'\omega_{q,q+1}^v + 1-s'} \omega_{q,q+1}^m - \omega_{q-1,q+1}^m \right) \right\}. \end{aligned} \quad (4.19)$$

4.1.2.1.2 Influence de la distribution *a priori* sur \mathbf{k} . Nous étudions l'influence de la distribution *a priori* $\pi(\mathbf{k})$ sur la borne de Weiss-Weinstein dans le cas de changements multiples dans la moyenne d'un signal gaussien. Spécifiquement, nous considérons quatre distributions différentes du type décrit par (3.18). Nous donnons ici les expressions de ces distributions ainsi que leur influence sur la forme des éléments de la matrice $\mathbf{G}(\mathbf{H})$ obtenus au chapitre 3, section 3.1.4.3, dans le cas $\mathbf{s} = (1/2; 1/2; \dots; 1/2)$ et pour une distribution du type (3.18) quelconque. Plus précisément, nous donnons les expressions des quantités restant à déterminer avec une loi *a priori* fixée, soit, d'après les résultats de la section 3.1.4.3, les quantités suivantes :

- i. $\sum_{\mathbf{k} \in \mathcal{K}'_{-h_a,0}} \pi(\mathbf{k} + \mathbf{h}_a)$,
- ii. $\sum_{\mathbf{k} \in \mathcal{K}'_{-h_a,0}} \nu(\mathbf{h}_a, \mathbf{0}, \mathbf{k})$,
- iii. $\sum_{\mathbf{k} \in \mathcal{K}'_{-h_a,-h_{a+1}}} \nu(h_a, h_{a+1}, \mathbf{k})$,
- iv. S_1 et S_2 (ou bien $S_1 + S_2$) définies en (3.42) et (3.43), respectivement.

Marche aléatoire uniforme. Nous avons vu que les points de rupture distribués selon la marche aléatoire uniforme (3.2) conduisent à la distribution *a priori* jointe $\pi_{\mathbf{d}}(\mathbf{k})$ donnée par (3.3), qui est bien du type (3.18), et que l'on rappelle ici :

$$\pi_{\mathbf{d}}(\mathbf{k}) = \frac{1}{\Delta^Q} \prod_{q=1}^Q \mathbb{1}_{\mathcal{K}'_{\mathbf{d}}(q)}(k_q) = \frac{1}{\Delta^Q} \mathbb{1}_{\mathcal{K}_{\mathbf{d}}}(\mathbf{k}). \quad (4.20)$$

Il s'agit de la distribution *a priori* utilisée préférentiellement tout au long de ce manuscrit. Ainsi en reprenant les résultats obtenus au chapitre 3, section 3.1.3 et dans l'annexe A, on a (i et ii)

$$\sum_{\mathbf{k} \in (\mathcal{K}'_{\mathbf{d}})_{-h_a,0}} \pi(\mathbf{k} \pm \mathbf{h}_a) = \sum_{\mathbf{k} \in (\mathcal{K}'_{\mathbf{d}})_{-h_a,0}} \nu(\mathbf{h}_a, \mathbf{0}, \mathbf{k}) = u(\Delta, h_a) \quad (4.21)$$

où $u(\Delta, h_a)$ est définie en (3.11), (iii)

$$\sum_{\mathbf{k} \in (\mathcal{K}'_{\mathbf{d}})_{-h_a,-h_{a+1}}} \nu(h_a, h_{a+1}, \mathbf{k}) = \begin{cases} \frac{(\Delta - |h_a|)(\Delta - |h_{a+1}|)}{\Delta^3}, & \text{si } a < Q, \\ \frac{\Delta - |h_Q|}{\Delta^2}, & \text{si } a = Q, \\ 0, & \text{sinon,} \end{cases} \quad (4.22)$$

et (iv)

$$S_1 + S_2 = \left(\sum_{\mathbf{k} \in (\mathcal{K}'_{\mathbf{d}})_{-h_a,-h_{a+1}}} \nu(\mathbf{h}_a, \mathbf{h}_{a+1}, \mathbf{k}) \right) \left(D - |h_a| - |h_{a+1}| + 1 - \frac{1 - R_a^{d - \min\{|h_a|, |h_{a+1}|\}}}{1 - R_a} \right). \quad (4.23)$$

Distribution uniforme sur des segments distincts. Nous considérons également une distribution *a priori* des ruptures uniformes sur des segments distincts, *fixés*, i.e., on a, $\forall q \in \llbracket 1, Q \rrbracket$,

$$k_q \sim \mathcal{U}(\llbracket \tau_{q-1} + 1, \tau_q \rrbracket), \quad (4.24)$$

avec $\tau_0 \triangleq 0 < \tau_1 < \dots < \tau_Q < N$, les entiers $(\tau_q)_{q=1, \dots, Q}$ pouvant être choisis arbitrairement sur $\llbracket 1, N \rrbracket$, indépendamment des valeurs de k_{q-1}, \dots, k_1 . Dans ce cas, tous les points de rupture $(k_q)_{q=1, \dots, Q}$ sont *a priori* supposés indépendants deux à deux, i.e., dans (3.18), on a $\pi(k_q | k_{q-1}) = \pi(k_q)$. En notant $\boldsymbol{\tau} = (\tau_0; \tau_1; \dots; \tau_Q)$, la distribution *a priori* jointe $\pi_{\boldsymbol{\tau}}(\mathbf{k})$ s'écrit donc

$$\pi_{\boldsymbol{\tau}}(\mathbf{k}) = \prod_{q=1}^Q \frac{\mathbb{1}_{\llbracket \tau_{q-1}+1, \tau_q \rrbracket}(k_q)}{\tau_q - \tau_{q-1}}. \quad (4.25)$$

Avant de donner les expressions des quantités clés pour la borne de Weiss-Weinstein, il faut remarquer que cette distribution *a priori* ne permet pas à la « situation de chevauchement » (dont il est question dans l'annexe A ainsi qu'au chapitre 3, section 4.1.2.1.2) de se produire. En effet, la plage de valeurs admissibles pour h_a , i.e., telle que $\pi_{\boldsymbol{\tau}}(\mathbf{k} + \mathbf{h}_a) > 0$, est donné par $\llbracket 1 - (\tau_a - \tau_{a-1}), (\tau_a - \tau_{a-1}) - 1 \rrbracket \setminus \{0\}$. Or, pour $h_a > 0$ et $h_{a+1} < 0$, on trouve que $(k_a)^{\max} = \tau_a - h_a$, tandis que $(k_{a+1})_{\min} = \tau_a + h_{a+1} + 1$. Ceci est incompatible avec la plage de valeurs de h_a pour que la condition d'existence d'un chevauchement, à savoir $t_a + h_a > t_{a+1} + h_{a+1}$, puisse être satisfaite.

Ceci implique que pour cette loi *a priori*, le cas **psdiag** est traité exactement de la même manière que le cas **ts**, ce qui conduit à

$$[\mathbf{G}]_{a, a+1} = 0, \quad (4.26)$$

autrement dit la matrice \mathbf{G} est diagonale. On trouve que ses termes diagonaux sont donnés par

$$[\mathbf{G}]_{a, a} = \frac{2 \left(1 - \frac{|h_a|}{\tau_a - \tau_{a-1}} - \rho_{a, a+1} \left(1 - \frac{2|h_a|}{\tau_a - \tau_{a-1}} \right) \right)}{\rho_{a, a+1}^2 \left(1 - \frac{|h_a|}{\tau_a - \tau_{a-1}} \right)^2}. \quad (4.27)$$

Distribution globalement uniforme sur l'ensemble des segmentations possibles. L'ensemble des segmentations possibles est en fait l'espace des paramètres lui-même $\Theta = \{k_1, \dots, k_Q \in \llbracket 1, N \rrbracket \mid 0 < k_1 < \dots < k_Q < N\}$. Par conséquent, la loi *a priori* uniforme sur Θ est donnée par

$$\pi_{\mathbf{u}}(\mathbf{k}) = \frac{1}{\text{Card}(\Theta)} \mathbb{1}_{\Theta}(\mathbf{k}) = \frac{1}{\binom{N-1}{Q}} \prod_{q=1}^Q \mathbb{1}_{\llbracket k_{q-1}+1, N-Q+q-1 \rrbracket}(k_q). \quad (4.28)$$

Notons qu'il s'agit de la distribution *a priori* la moins informative de celles considérées. De plus, on peut démontrer que cette loi correspond à celle du vecteur indicateur de changements² $\mathbf{r} \in \{0, 1\}^N$ (introduit dans [Lav98]) conditionnée par $\sum_{n=1}^N r_n = Q$ (i.e., par le fait que le nombre total de changements est égal à Q). On obtient donc, avec cette distribution *a priori*, les expressions suivantes des quantités énoncées en début de section : (i et ii)

$$\sum_{\mathbf{k} \in \Theta'_{-h_a, 0}} \pi_{\mathbf{u}}(\mathbf{k} \pm \mathbf{h}_a) = \sum_{\mathbf{k} \in \Theta'_{-h_a, 0}} \nu(\mathbf{h}_a, \mathbf{0}, \mathbf{k}) = \frac{\binom{N-1-|h_a|}{Q}}{\binom{N-1}{Q}}, \quad (4.29)$$

2. Nous avons vu au chapitre 2, section 2.2.3, que les composantes r_n de ce vecteur valent 1 si l'instant n est un point de rupture et 0 sinon.

où, pour $(a, b) \in \llbracket 1, Q \rrbracket^2$, $\Theta'_{-h_a, -h_b} \triangleq \Theta \cap (\Theta - \mathbf{h}_a) \cap (\Theta - \mathbf{h}_b)$, (iii)

$$\sum_{\mathbf{k} \in \Theta'_{-h_a, -h_{a+1}}} \nu(\mathbf{h}_a, \mathbf{h}_{a+1}, \mathbf{k}) = \frac{\binom{N-1-|h_a|-|h_{a+1}|}{Q}}{\binom{N-1}{Q}}, \quad (4.30)$$

et (iv)

$$S_1 + S_2 = \frac{\binom{N-|h_a|-|h_{a+1}|}{Q}}{\binom{N-1}{Q}} + \sum_{k_1} \dots \sum_{k_{a-1}} \sum_{k_a} \left(R_a^{-k_a-|h_a|-|h_{a+1}|} \sum_{k_{a+1}} \frac{\binom{N-1-k_{a+1}}{Q-(a+1)}}{\binom{N-1}{Q}} R_a^{k_{a+1}} \right) \quad (4.31)$$

où les bornes des sommes sont données :

- pour $q \in \llbracket 1, a-1 \rrbracket$, par $(k_q)_{\min} = k_{q-1} + 1$ et $(k_q)_{\max} = N - Q + q - 1$,
- pour $q = a$, par $(k_a)_{\min} = k_{a-1} + 1$ et $(k_a)_{\max} = N - Q + a - |h_a| - 1$,
- pour $q = a + 1$, par $(k_{a+1})_{\min} = k_a + 1 + \max\{|h_a|, |h_{a+1}|\}$ et $(k_{a+1})_{\max} = k_a + |h_a| + |h_{a+1}| - 1$.

On n'obtient donc pas, dans ce cas, d'expression complètement analytique des termes de la matrice \mathbf{G} . Pour autant, les sommes intervenant dans (4.31) sont discrètes et de bornes finies, ce qui n'implique aucune difficulté d'implémentation.

Marche aléatoire de Poisson tronquée. Enfin, nous considérons une marche aléatoire de Poisson tronquée pour la distribution *a priori* de la localisation des ruptures, i.e., $\forall q \in \llbracket 1, Q \rrbracket$,

$$k_q = k_{q-1} + \ell_q + 1, \quad (4.32)$$

où ℓ_q suit une loi de Poisson tronquée $\mathcal{P}^{N-Q+q-1}(\lambda_\pi)$ de paramètre λ_π et de valeur maximale $N - Q + q - 1$. Ce choix de distribution *a priori* est inspiré par [MF06], en l'adaptant au cas où le nombre total de changements est supposé égal à Q , d'où la nécessité d'utiliser une loi tronquée. Le paramètre de la loi λ_π représente la longueur *a priori* des segments $\llbracket k_{q-1} + 1, k_q \rrbracket$, $q = 1, \dots, Q$ ($k_0 \triangleq 0$). Une valeur de λ_π qui favorise une répartition uniforme de la longueur des segments est donc $\lambda_\pi = N/(Q + 1) - 1$. Ainsi, la distribution *a priori* du vecteur \mathbf{k} est obtenue par

$$\begin{aligned} \pi_\lambda(\mathbf{k}) &= \prod_{q=1}^Q \Pr(k_q = k_q \mid k_{q-1} = k_{q-1} \text{ et } k_{q-1} + 1 \leq k_q \leq N - Q + q - 1) \\ &= \prod_{q=1}^Q \frac{\Pr(l_q = \ell_q, l_q \leq N - Q + q - 1)}{\Pr(l_q \leq N - Q + q - 2 - k_{q-1})} \\ &= \prod_{q=1}^Q \frac{\exp(-\lambda_\pi) \frac{\lambda_\pi^{k_q - k_{q-1} - 1}}{(k_q - k_{q-1} - 1)!}}{\sum_{i=0}^{N-Q+q-2-k_{q-1}} \exp(-\lambda_\pi) \frac{(\lambda_\pi)^i}{i!}} \mathbb{1}_{\llbracket k_{q-1}+1, N-Q+q-1 \rrbracket}(k_q) \\ &= \frac{\lambda_\pi^{k_Q - Q}}{\prod_{q=1}^Q \left((k_q - k_{q-1} - 1)! \sum_{i=0}^{N-Q+q-2-k_{q-1}} \frac{(\lambda_\pi)^i}{i!} \right)} \mathbb{1}_{\llbracket k_{q-1}+1, N-Q+q-1 \rrbracket}(k_q) \\ &= \frac{\lambda_\pi^{k_Q - Q}}{\prod_{q=1}^Q \left((k_q - k_{q-1} - 1)! \Sigma_{N-Q+q-2-k_{q-1}} \right)} \mathbb{1}_{\llbracket k_{q-1}+1, N-Q+q-1 \rrbracket}(k_q) \end{aligned} \quad (4.33)$$

en notant, pour $k \in \mathbb{N}$, $\Sigma_k \triangleq \sum_{i=0}^k \frac{(\lambda_\pi)^i}{i!}$. Il n'est donc pas étonnant, au vu de l'expression de $\pi_\lambda(\mathbf{k})$, déjà bien plus complexe que celle de $\pi_u(\mathbf{k})$ donnée en (4.28), qu'on n'obtienne pas de

formes analytiques intéressantes pour les éléments de la matrice \mathbf{G} . On trouve seulement des expressions comprenant plusieurs sommes discrètes, à la manière de (4.31). Pour ne pas alourdir le manuscrit, nous ne fournissons les expressions obtenues que dans le cas $Q = 2$, permettant bien de se rendre compte de la complexité calculatoire mise en jeu. On obtient donc :

(i) pour $a = 1$,

$$\sum_{\mathbf{k} \in \Theta'_{-h_1, 0}} \pi_\lambda(\mathbf{k} \pm \mathbf{h}_1) = \frac{1}{\Sigma_{N-3}} \sum_{\ell_1 = \max\{-h_1, 0\}}^{N-3-\max\{h_1, 0\}} \frac{\lambda_\pi^{\ell_1}}{(\ell_1 + h_1)! \Sigma_{N-3-\ell_1-h_1}} \sum_{\ell_2 = \max\{h_1, 0\}}^{N-3-\ell_1} \frac{\lambda_\pi^{\ell_2}}{(\ell_2 - h_1)!}, \quad (4.34)$$

(i)(bis) pour $a = 2$,

$$\sum_{\mathbf{k} \in \Theta'_{-h_2, 0}} \pi_\lambda(\mathbf{k} \pm \mathbf{h}_2) = \frac{1}{\Sigma_{N-3}} \sum_{\ell_1 = 0}^{N-3} \frac{\lambda_\pi^{\ell_1}}{\ell_1! \Sigma_{N-3-\ell_1}} \sum_{\ell_2 = \max\{-h_2, 0\}}^{N-3-\ell_1-\max\{h_2, 0\}} \frac{\lambda_\pi^{\ell_2}}{(\ell_2 + h_2)!}, \quad (4.35)$$

(ii) pour $a = 1$,

$$\begin{aligned} \sum_{\mathbf{k} \in \Theta'_{-h_1, 0}} \nu(h_1, 0, \mathbf{k}) &= \frac{1}{\Sigma_{N-3}} \sum_{\ell_1 = \max\{-h_1, 0\}}^{N-3-\max\{h_1, 0\}} \frac{\lambda_\pi^{\ell_1}}{\ell_1! \Sigma_{N-3-\ell_1}} \sqrt{\frac{\Sigma_{N-3-\ell_1}}{\Sigma_{N-3-\ell_1-h_1}} \left(P_{|h_1|}^{\ell_1 + \max\{h_1, 0\}} \right)^{-\text{sign}(h_1)}} \\ &\quad \times \sum_{\ell_2 = \max\{h_1, 0\}}^{N-3-\ell_1} \frac{\lambda_\pi^{\ell_2}}{\ell_2!} \sqrt{\left(P_{|h_1|}^{\ell_2 + \max\{-h_1, 0\}} \right)^{\text{sign}(h_1)}}, \end{aligned} \quad (4.36)$$

où, pour deux entiers k et n tels que $0 < k < n$, P_k^n représente la factorielle décroissante de n , i.e., $P_k^n \triangleq \frac{n!}{(n-k)!} = \prod_{i=n-k+1}^n i$ (parfois également appelée « symbole de Pochhammer ») ;

(ii)(bis) pour $a = 2$,

$$\sum_{\mathbf{k} \in \Theta'_{-h_2, 0}} \nu(h_2, 0, \mathbf{k}) = \frac{\sqrt{\lambda_\pi^{h_2}}}{\Sigma_{N-3}} \sum_{\ell_1 = 0}^{N-3} \frac{\lambda_\pi^{\ell_1}}{\ell_1! \Sigma_{N-3-\ell_1}} \sum_{\ell_2 = \max\{-h_2, 0\}}^{N-3-\ell_1-\max\{h_2, 0\}} \frac{\lambda_\pi^{\ell_2}}{\ell_2!} \sqrt{\left(P_{|h_2|}^{\ell_2 + \max\{h_2, 0\}} \right)^{-\text{sign}(h_2)}}, \quad (4.37)$$

(iii)

$$\begin{aligned} \sum_{\mathbf{k} \in \Theta'_{-h_1, -h_2}} \nu(h_1, h_2, \mathbf{k}) &= \frac{\sqrt{\lambda_\pi^{h_2}}}{\Sigma_{N-3}} \sum_{\ell_1 = \max\{-h_1, 0\}}^{N-3-\max\{h_1, 0\}} \frac{\lambda_\pi^{\ell_1}}{\ell_1! \Sigma_{N-3-\ell_1}} \sqrt{\frac{\Sigma_{N-3-\ell_1}}{\Sigma_{N-3-\ell_1-h_1}} \left(P_{|h_1|}^{\ell_1 + \max\{h_1, 0\}} \right)^{-\text{sign}(h_1)}} \\ &\quad \times \sum_{\ell_2 = \max\{h_1, -h_2\}}^{N-3-\ell_1-\max\{h_2, 0\}} \frac{\lambda_\pi^{\ell_2}}{\ell_2!} \sqrt{\frac{\left(P_{|h_1|}^{\ell_2 + \max\{-h_1, 0\}} \right)^{\text{sign}(h_1)}}{\left(P_{|h_2|}^{\ell_2 + \max\{h_2, 0\}} \right)^{\text{sign}(h_2)}}}, \end{aligned} \quad (4.38)$$

et enfin (iv), où l'on rappelle que l'on a $h_1 > 0$, $h_2 < 0$, et $\min\{|h_1|, |h_2|\} > 1$,

$$S_1 = \frac{\sqrt{\lambda_\pi^{h_2}}}{\Sigma_{N-3}} R_1^{1-|h_1|-|h_2|} \sum_{\ell_1 = 0}^{N-3-h_1} \frac{\lambda_\pi^{\ell_1}}{\ell_1! \Sigma_{N-3-\ell_1}} \sqrt{\frac{\Sigma_{N-3-\ell_1}}{\Sigma_{N-3-\ell_1-h_1}} \left(P_{|h_1|}^{\ell_1 + h_1} \right)} \sum_{\ell_2 = \max\{|h_1|, |h_2|\}}^{|h_1|+|h_2|-2} \frac{(R_1 \lambda_\pi)^{\ell_2}}{\ell_2!} \sqrt{P_{|h_1|}^{\ell_2} P_{|h_2|}^{\ell_2}} \quad (4.39)$$

et

$$S_2 = \frac{\sqrt{\lambda_\pi^{h_2}}}{\Sigma_{N-3}} \sum_{\ell_1 = 0}^{N-3-h_1} \frac{\lambda_\pi^{\ell_1}}{\ell_1! \Sigma_{N-3-\ell_1}} \sqrt{\frac{\Sigma_{N-3-\ell_1}}{\Sigma_{N-3-\ell_1-h_1}} \left(P_{|h_1|}^{\ell_1 + h_1} \right)} \sum_{\ell_2 = |h_1|+|h_2|-1}^{N-3-\ell_1} \frac{\lambda_\pi^{\ell_2}}{\ell_2!} \sqrt{P_{|h_1|}^{\ell_2} P_{|h_2|}^{\ell_2}}. \quad (4.40)$$

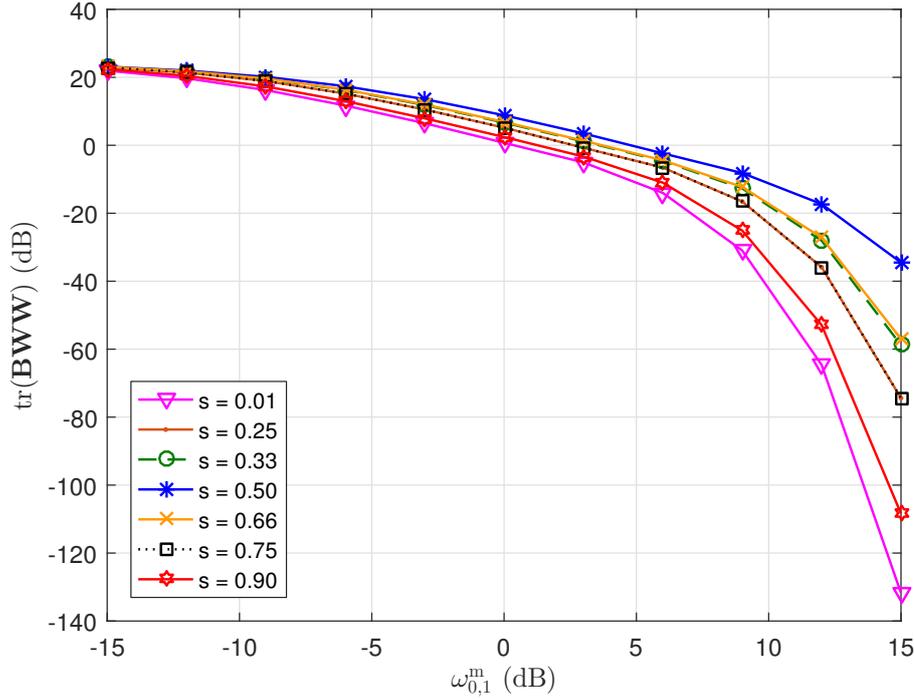


FIGURE 4.11 – Évolution de $\text{tr}(\mathbf{BWW})$, pour plusieurs valeurs de $\mathbf{s} = [s, s, \dots, s]^T$, $s \in]0, 1[$, dans le cas $Q = 2$ changements dans la moyenne de $N = 80$ observations gaussiennes, en fonction de la quantité de changement $\omega_{0,1}^m$.

4.1.2.1.3 Résultats numériques. Nous illustrons maintenant les résultats théoriques obtenus précédemment en les appliquant à des données simulées.

Ruptures distribuées *a priori* selon la marche aléatoire uniforme. Dans un premier temps, nous étudions le comportement de la borne de Weiss-Weinstein avec la distribution *a priori* sur les positions des points de rupture préférentiellement utilisée tout au long de ce manuscrit, à savoir la marche aléatoire uniforme (voir (3.3)) pour $d = 1$ et $D = D_{\max} = \lfloor (N - 1)/Q \rfloor$, pour des données gaussiennes comportant un changement dans la moyenne ou dans la variance. Il s'agit donc exactement de la BWW dont la matrice \mathbf{G} est donnée par (3.9), avec $\rho_{q-1,q}(s)$ et $R_q(s, s')$ donnés par (4.7) et (4.19), respectivement. La borne la plus précise est calculée en suivant la procédure décrite à la section 3.3, en utilisant le package CVX [GB08].

Nous considérons d'abord le cas de changements dans la moyenne, i.e., $\omega_{q,q'}^v = 1, \forall q, q'$. Nous choisissons $\sigma^2 = 1$, $m_0 = 1$, et nous fixons les moyennes sur les segments suivants de sorte que $\omega_{0,1}^m = \omega_{1,2}^m = \omega_{2,3}^m = \dots = \omega^m$, et, pour $q \in \llbracket 1, Q - 1 \rrbracket$, $m_{q+1} = m_q + (-1)^q \sqrt{\sigma^2} \omega^m$.

Sur la figure 4.11, nous étudions d'abord l'influence des paramètres s_q dans le cas $Q = 2$ ruptures, $N = 80$ observations, et pour un grand nombre de valeurs du couple $(s_1, s_2) \in]0, 1]^2$. Nous ne représentons que les évolutions de $\text{tr}(\mathbf{BWW})$ en fonction de ω^m , pour les couples (s_1, s_2) tels que $s_1 = s_2$ pour les valeurs 0.01, 0.25, 0.33, 0.5, 0.66, 0.75, 0.90. Il semble clair que dans ce cas, la valeur $s_1 = s_2 = 0.5$ conduit à la BWW la plus précise.

Sur la figure 4.12, nous considérons le cas de $Q = 3$ changements dans la moyenne de $N = 80$ observations. On représente les REQM globales approchées de l'estimateur du MAP de chacune des ruptures, ainsi que les bornes de Weiss-Weinstein associées, en prenant les racines carrées des éléments diagonaux des matrices \mathbf{EQMG} et \mathbf{BWW} , respectivement. On peut noter que l'écart

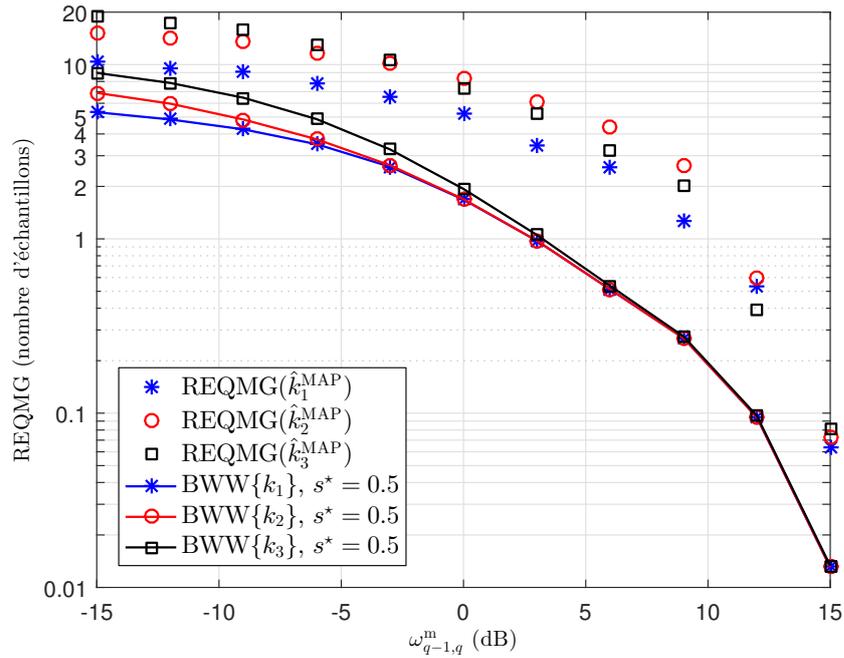


FIGURE 4.12 – Cas de $Q = 3$ changements dans la moyenne de $N = 80$ observations gaussiennes. Évolution des bornes de Weiss-Weinstein (traits pleins continus) et des EQM globales associées aux estimations par le MAP des points de rupture k_1 (étoiles bleues), k_2 (cercles rouges), et k_3 (carrés noirs), en fonction de la quantité de changement $\omega_{q-1,q}^m$.

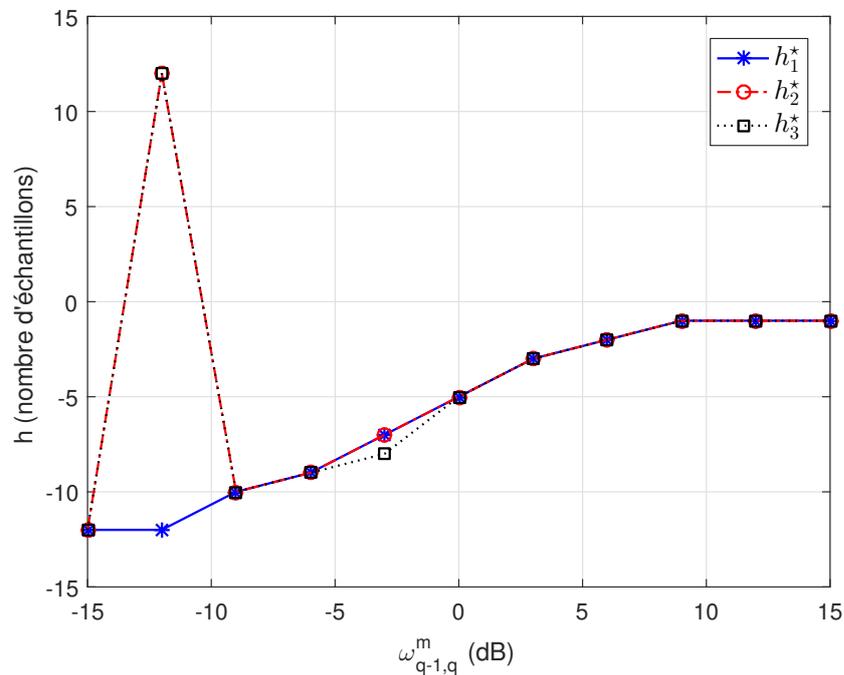


FIGURE 4.13 – Cas de $Q = 3$ changements dans la moyenne de $N = 80$ observations gaussiennes. Évolution des valeurs test optimales pour la borne de Weiss-Weinstein, associées aux paramètres k_1 (étoiles bleues), k_2 (cercles rouges), et k_3 (carrés noirs), en fonction de la quantité de changement $\omega_{q-1,q}^m$.

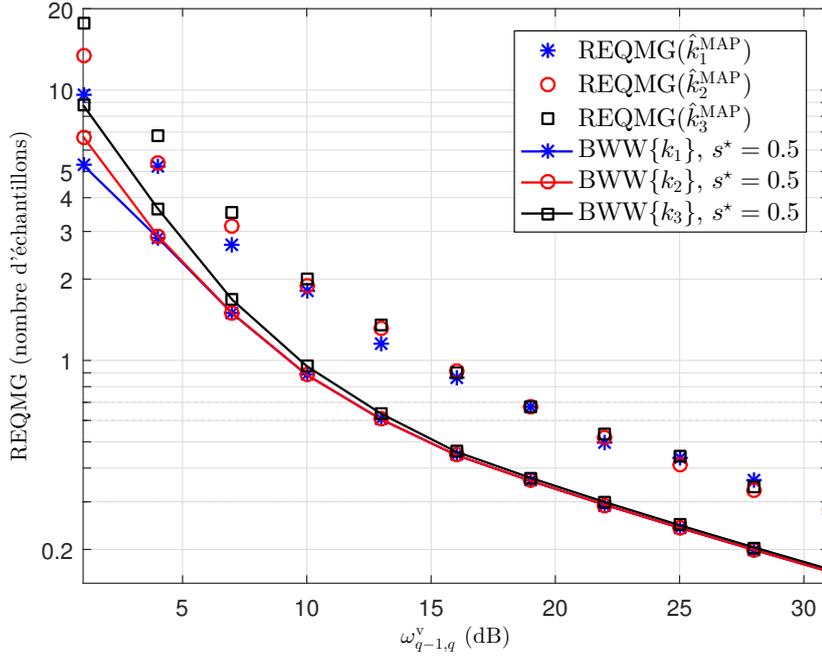


FIGURE 4.14 – Cas de $Q = 3$ changements dans la variance de $N = 80$ observations gaussiennes. Évolution des bornes de Weiss-Weinstein (traits pleins continus) et des EQM globales associées aux estimations par le MAP des points de rupture k_1 (étoiles bleues), k_2 (cercles rouges), et k_3 (carrés noirs), en fonction de la quantité de changement $\omega_{q-1,q}^m$.

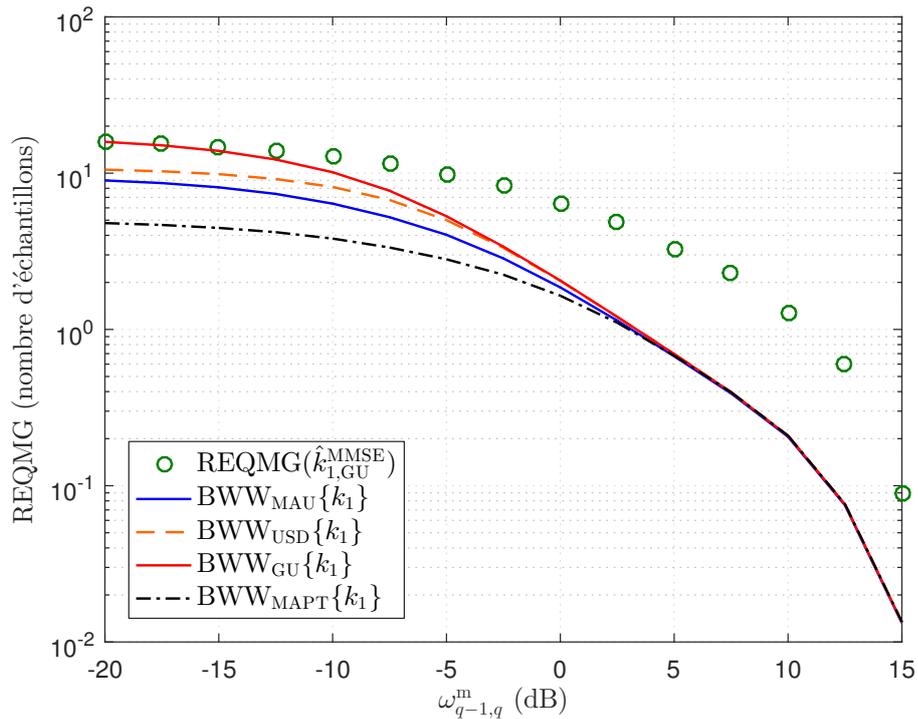
entre les EQMG du MAP et les BWW pour chaque rupture diminue à mesure que la quantité de changement ω^m augmente. En particulier, pour $\omega^m > 10$ dB, l'écart est inférieur à 1 échantillon.

Nous traçons, sur la figure 4.13, l'évolution des valeurs optimales des points test h_1^* , h_2^* , h_3^* conduisant à la borne de Weiss-Weinstein la plus précise, en fonction de ω^m . Le comportement est quasiment identique pour chacun d'eux, et comme dans le cas $Q = 1$, on constate que ces valeurs se rapprochent de 0 à mesure que ω^m augmente.

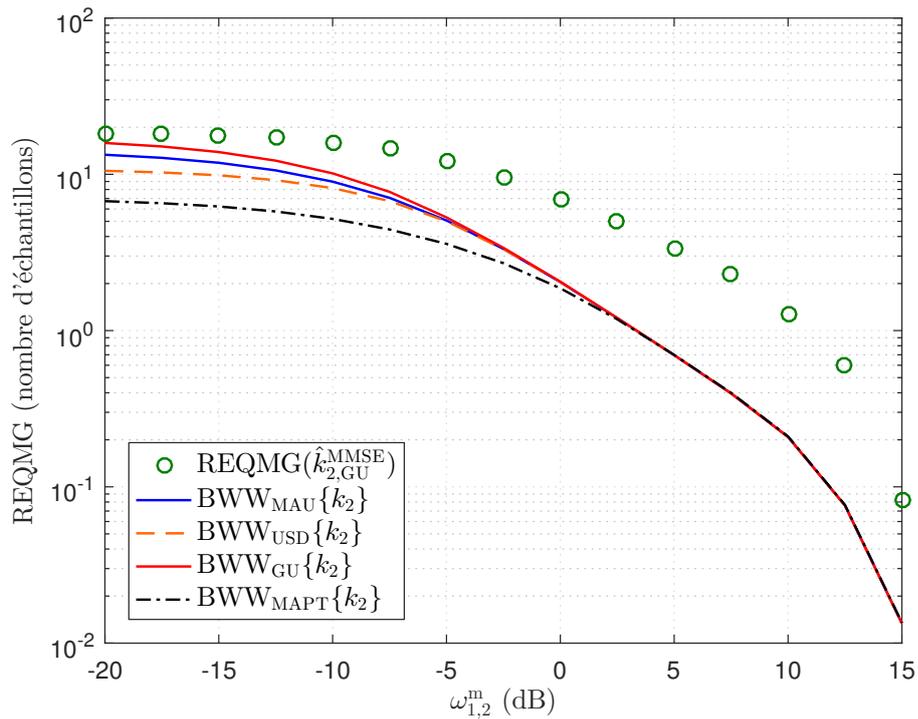
Dans le cas d'un changement de variance, on a $\omega_{q-1,q}^m = 0$, $\forall q \in \llbracket 1, Q \rrbracket$. On choisit ici de fixer $m_q = 0$, $\forall q \in \llbracket 0, Q \rrbracket$, $\sigma_0^2 = 1$, et les variances sur les segments suivants sont fixées de sorte que $\sigma_{q+1}^2 = \sigma_q^2 \cdot \omega^v$ pour $q \in \llbracket 0, Q-1 \rrbracket$, où $\omega^v = \omega_{0,1}^v = \omega_{1,2}^v = \dots = \omega_{Q-1,Q}^v$. Nous représentons le cas de $Q = 3$ ruptures dans $N = 80$ observations sur la figure 4.14. Le comportement, pour chacune des ruptures, est comparable à celui observé dans le cas $Q = 1$, tant du point de vue des REQMG du MAP que des bornes associées. Les trois REQMG d'une part, et les trois BWW associées d'autre part, convergent asymptotiquement (à forte quantité de changement) vers une valeur commune. Lorsque ω^v diminue, on observe que la REQMG associée aux dernières ruptures (k_2 , k_3) sont plus élevées que celles associées aux premières. Ce comportement est également mis en évidence par le tracé des BWW.

Influence de la distribution *a priori*. En s'intéressant toujours à des données gaussiennes, nous illustrons maintenant, via des simulations numériques, l'influence de la distribution *a priori* des positions des ruptures.

Plus précisément, nous comparons les bornes de Weiss-Weinstein associées aux quatre distributions *a priori* considérées dans la section 4.1.2.1.2, dans le cas de $Q = 2$ changements dans la moyenne de $N = 100$ observations. Les résultats sont présentés sur la figure 4.15, pour chacun des deux points de rupture. On trace également la REQMG associée à l'estimateur MMSE de ces



(a) Premier point de rupture k_1



(b) Second point de rupture k_2

FIGURE 4.15 – Évolutions de l'EQMG globale de l'estimateur MMSE des points de rupture k_1 (à gauche) et k_2 (à droite), pour la loi *a priori* globalement uniforme (cercles verts) et des bornes de Weiss-Weinstein pour des distributions *a priori* diverses : marche aléatoire uniforme (MAU, trait plein bleu), uniforme sur des segments distincts (USD, pointillés oranges), globalement uniforme (GU, trait plein rouge), et marche aléatoire de Poisson tronquée (MAPT, pointillés noirs).

ruptures, pour la distribution *a priori* globalement uniforme sur les ruptures. Pour des conserver la clarté des figures, nous ne traçons pas les REQMG de ce même estimateur pour les autres distributions *a priori*.

Principalement, on observera que quelle que soit la distribution *a priori* considérée sur la localisation des points de rupture, les bornes de Weiss-Weinstein associées convergent vers une valeur commune lorsque la quantité de changement augmente : à partir de $\omega^m = 3$ dB, on est incapable de distinguer les courbes ; l'écart avec la REQMG de l'estimateur MMSE pour l'*a priori* globalement uniforme est alors de l'ordre de 4 échantillons. Par contre, l'influence est bien plus importante pour de faibles quantités de changement. En effet, dans ces conditions, c'est principalement la loi *a priori* qui régit le résultat de l'estimation, et cela est aussi mis en évidence par les bornes de Weiss-Weinstein, qui diffèrent sensiblement suivant l'*a priori* considéré.

4.1.2.2 Signaux poissonniens

Nous terminons cette analyse du cas où les paramètres du signal sont connus sur chaque segment avec le cas d'observations suivant une loi de Poisson. Nous donnons d'abord, dans la section 4.1.2.2.1, les expressions des quantités restant à déterminer pour obtenir la borne de Weiss-Weinstein pour l'estimation de ruptures multiples, pour ce type de signaux. Nous verrons ensuite des résultats issus de simulation dans la section 4.1.2.2.2, permettant d'évaluer la précision de la BWW.

4.1.2.2.1 Expression de $r_q(s, s')$ et $R_q(s, s')$. Pour obtenir les expressions de $r_q(s, s')$ et $R_q(s, s')$ dans le cas de données distribuées selon une loi de Poisson, nous injectons (4.12) dans (3.17) et (3.16), ce qui donne, après des manipulations du même type que celles menant à (4.14) (en prenant en particulier bien compte la nature discrète des observations) : pour $q \in \llbracket 1, Q - 1 \rrbracket$ et $(s, s') \in]0, 1]^2$,

$$r_q(s, s') = \exp\left\{-s\lambda_{q-1} - s'\lambda_q - (1 - s - s')\lambda_{q+1} + \lambda_{q-1}^s \lambda_q^{s'} \lambda_{q+1}^{1-s-s'}\right\} \quad (4.41)$$

et

$$R_q(s, s') = \exp\left\{-\lambda_q \left[1 - \left(\frac{\lambda_{q-1}}{\lambda_q}\right)^s - \left(\frac{\lambda_{q+1}}{\lambda_q}\right)^{1-s'} + \left(\frac{\lambda_{q-1}}{\lambda_q}\right)^s \left(\frac{\lambda_{q+1}}{\lambda_q}\right)^{1-s'}\right]\right\}. \quad (4.42)$$

Comme dans le cas $Q = 1$ d'une unique rupture, ces quantités ne s'expriment pas simplement en fonction de la quantité de changement $\omega_{q-1,q}^\lambda$ définie en (4.13).

4.1.2.2.2 Résultats numériques. Nous considérons le cas de $Q = 3$ changement dans le paramètre λ_0 de $N = 80$ observations suivant une loi de Poisson. On fixe $\lambda_0 = 1$, et les paramètres sur les segments suivants sont définis de sorte que $\omega_{0,1}^\lambda = \omega_{1,2}^\lambda = \omega_{2,3}^\lambda = \omega^\lambda$, et, pour $q \in \llbracket 0, Q - 1 \rrbracket$, $\lambda_{q+1} = \lambda_q(1 + \sqrt{\omega^\lambda})$. Le comportement de la borne de Weiss-Weinstein, en comparaison de celui des REQMG des estimateurs du MAP de chacun des points de rupture, en fonction de la quantité de changement ω^λ , est illustré sur la figure 4.16. Par opposition aux cas étudiés précédemment, on constate que ni les REQMG du MAP, ni les BWW associées pour chacun des points de rupture ne convergent vers une valeur commune. On observe un point de croisement à $\omega^\lambda = -6$ dB, en-dessous duquel le comportement est similaire à celui du cas gaussien (les premières ruptures sont mieux estimées que les dernières), et au-dessus duquel l'ordre des REQMG pour chacune des ruptures s'inverse (i.e., le dernières ruptures sont mieux estimées que les premières). Ce comportement peut être expliqué par deux causes :

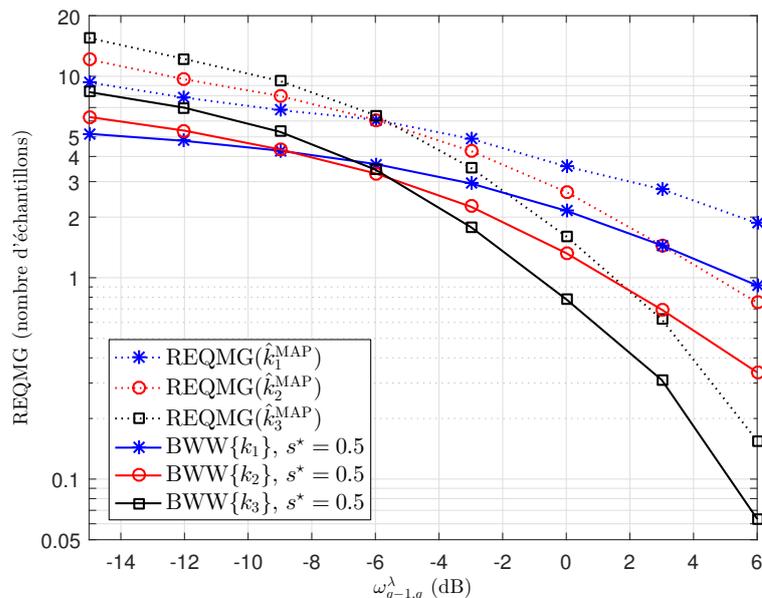


FIGURE 4.16 – Cas de $Q = 3$ changements dans le paramètre λ_q de $N = 80$ observations poissonniennes. Évolution des bornes de Weiss-Weinstein (traits pleins continus) et des EQM globales associées à l'estimateur MAP des points de rupture k_1 (étoiles bleues), k_2 (cercles rouges), et k_3 (carrés noirs), en fonction de la quantité de changement $\omega_{q-1,q}^\lambda$.

- Dans le cas de données poissonniennes, un seul paramètre (λ_q) est susceptible de changer d'un segment à un autre, mais celui-ci régit à la fois la moyenne et la variance, contrairement au cas gaussien où ces deux grandeurs correspondent à deux paramètres différents. Or, on peut comprendre intuitivement qu'un changement à la fois dans la moyenne et dans la variance sera plus facilement estimé qu'un changement dans la moyenne seulement (une justification mathématique de cet argument peut être trouvée dans [BN93]).
- La définition des paramètres λ_q utilisée pour cette simulation, à savoir $\lambda_{q+1} = \lambda_q(1 + \sqrt{\omega^\lambda})$, implique une augmentation géométrique de $(\lambda_q)_{q \geq 0}$, ce qui provoque une différence de moyenne entre deux segments, pour les dernières ruptures, plus importante que pour les premières ruptures.

4.2 Cas où les paramètres de la distribution du signal observé sur chaque segment sont inconnus

Dans la section précédente, nous avons obtenu des bornes inférieures de l'EQM associées à l'estimation d'un ou plusieurs points de rupture, en supposant les paramètres des signaux sur chaque segment connus. En pratique, ces paramètres sont eux aussi inconnus. Les bornes ainsi calculées fournissent donc un aperçu de performances d'estimation idéales. Dorénavant, nous nous intéressons au cas où les paramètres des signaux sur chaque segment sont inconnus. Nous chercherons à caractériser les performances de ces paramètres supplémentaires inconnus, en plus de celles des points de rupture. Or, les bornes de Cramér-Rao/Weiss-Weinstein hybride et bayésienne présentées dans la section 3.2 répondent justement à cette problématique. Nous étudions donc, dans la section 4.2.1, le cas de données suivant une loi de Poisson sur chaque segment, et dans la section 4.2.2, le cas de données gaussiennes comportant des changements dans la moyenne ou dans la variance, et dont les moyennes et/ou variances sur chaque segment

sont supposées inconnues.

4.2.1 Application à des signaux poissonniens

Nous nous plaçons donc d'abord dans le cas où les observations sont supposées suivre une loi de Poisson, dont les paramètres λ_q , sur chaque segment, sont supposés inconnus. Dans la section 4.2.1.1, on suppose que les paramètres de Poisson sont déterministes, on s'intéresse alors à la version hybride de la borne de Cramér-Rao/Weiss-Weinstein (voir chapitre 3, section 3.2.2). Dans la section 4.2.1.2, on supposera que ces paramètres sont aléatoires, et on s'intéressera donc à la version bayésienne de la borne de Cramér-Rao/Weiss-Weinstein.

4.2.1.1 Cas de paramètres λ_q déterministes

Nous supposons, dans cette section, que les paramètres $(\lambda_q)_{q=0,\dots,Q}$ sont déterministes, et on cherche à obtenir la borne de Cramér-Rao/Weiss-Weinstein hybride, dont l'expression est donnée au chapitre 3, section 3.2.2. Nous donnons d'abord les expressions des quantités restant à déterminer (section 4.2.1.1.1), avant de présenter des résultats de simulations (section 4.2.1.1.2).

4.2.1.1.1 Expressions de $\tilde{F}(\lambda_q)$ et de $\varphi_{q,\lambda_{\tilde{q}}}(s)$. L'expression de l'information de Fisher pour le paramètre λ_q d'une loi de Poisson est donné, pour $q \in \llbracket 0, Q \rrbracket$, par

$$\tilde{F}(\lambda_q) = \frac{1}{\lambda_q}. \quad (4.43)$$

Par ailleurs, en injectant (4.12) dans (3.67), on obtient après calculs, pour $q \in \llbracket 0, Q \rrbracket$, $\tilde{q} \in \{q-1, q\}$, et $s \in]0, 1[$,

$$\varphi_{q,\lambda_{\tilde{q}}}(s) = \rho_{q-1,q}(s) \left(\frac{\lambda_{q-1}^s \lambda_q^{1-s}}{\lambda_{\tilde{q}}} - 1 \right). \quad (4.44)$$

Ainsi, tous les éléments de la BCRWWH, dans ce cas, admettent une expression analytique, ce qui permet une implémentation facile de la borne.

4.2.1.1.2 Résultats numériques. On se place ici dans le cas de $Q = 3$ points de ruptures dans le paramètre λ_0 de $N = 100$ observations suivant une loi de Poisson. Les valeurs successives sur chaque segment de λ_q sont fixées de la même façon qu'expliqué à la section 4.1.2.2.2. Les résultats de simulation sont présentés à la figure 4.17. En observant le comportement des EQMG de l'estimateur du MLMAP pour les $(\lambda_q)_{q=1,\dots,3}$ (figure 4.17a), on constate que la BCRWWH fournit un bon aperçu des performances d'estimation en terme d'EQMG lorsque la quantité de changement tend vers l'infini : la BCRWWH est très proche de l'EQMG de l'estimateur MLMAP pour chacun des paramètres λ_q . La quantité de changement $\omega^\lambda \approx 0$ dB constitue un seuil en-dessous duquel les EQMG du MLMAP s'éloignent nettement de la BCRWWH : ce phénomène se rapproche du « décrochement » observé en analyse spectrale, et la BCRWWH (qui correspond à la BCR pour les paramètres considérés) ne permet pas d'en rendre compte. En ce qui concerne l'estimation des ruptures (figure 4.17b), on observe un comportement similaire à celui où les λ_q sont supposés connus.

4.2.1.2 Cas de paramètres λ_q aléatoires

Nous nous plaçons ici dans un cadre complètement bayésien, c'est-à-dire que les paramètres $(\lambda_q)_{q=0,\dots,Q}$ de la loi de Poisson sur chacun des segments sont supposés aléatoires. On leur

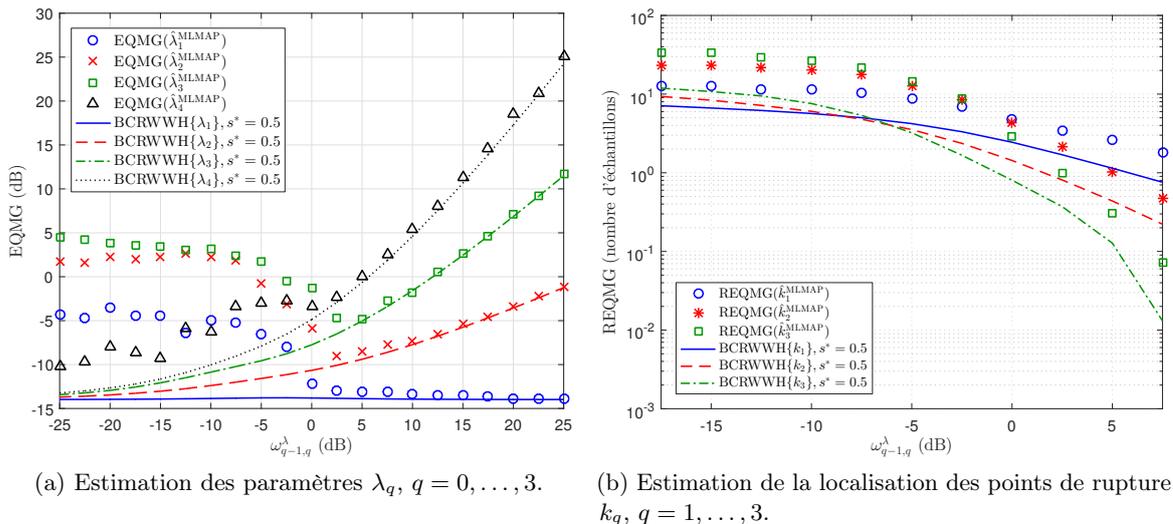


FIGURE 4.17 – Évolutions des EQMG globales de l'estimateur MLMAP et des bornes de Weiss-Weinstein associées en fonction de la quantité de changement $\omega_{q-1,q}^\lambda$, dans le cas de données poissonniennes avec $N = 100$ observations et $Q = 3$ ruptures, pour les paramètres $(\lambda_q)_{q=0,\dots,3}$ (à gauche), et les points de rupture $(k_q)_{q=1,\dots,3}$ (à droite).

attribue la distribution *a priori* conjuguée, qui pour ce type de paramètres est une loi Gamma de paramètres α_q et β , i.e., sur chaque segment, on a $\lambda_q \sim \mathcal{G}(\alpha_q, \beta)$, $q = 0, \dots, Q$. Ainsi, en supposant que les paramètres λ_q sont indépendants d'un segment à l'autre, la loi jointe sur le vecteur $\boldsymbol{\lambda} \triangleq [\lambda_0, \lambda_1, \dots, \lambda_Q]^\top$ est bien du type de (3.47), et donnée par

$$\pi(\boldsymbol{\lambda}) = \prod_{q=0}^Q \frac{\beta^{\alpha_q}}{\Gamma(\alpha_q)} \lambda_q^{\alpha_q-1} \exp\{-\beta \lambda_q\} \mathbb{1}_{\mathbb{R}_+}(\lambda_q), \quad (4.45)$$

où $\Gamma(\cdot)$ représente la fonction gamma usuelle, c'est-à-dire, pour $\alpha > 0$, $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} \exp\{-t\} dt$.

4.2.1.2.1 Expressions des éléments de la borne de Cramér-Rao/Weiss-Weinstein bayésienne. Pour répondre au cadre purement bayésien employé ici, nous nous intéressons à la version bayésienne de la borne de Cramér-Rao/Weiss-Weinstein, présentée au chapitre 3, section 3.2.3. Comme nous avons vu que les valeurs optimales de $(s_q)_{q=1,\dots,Q}$ de la BCRWWH sont $s_q^* = 0.5$ dans le cas de données poissonniennes (voir section 4.2.1.1.2), nous choisissons ici de calculer la BCRWWB directement avec ces valeurs. Les expressions des éléments de la BCRWWB données à la section 3.2.3 se trouvent donc légèrement simplifiées, mais ne peuvent pas s'exprimer sous une forme analytique, à part seulement les éléments de $[\mathbf{V}_{11}^\pi]$ (qui devient diagonale puisque $L = 1$ dans le cas poissonnien), alors donnés, pour $q \in \llbracket 0, Q \rrbracket$ et $\alpha_q > 2$, par

$$[\mathbf{V}_{11}^\pi]_{q,q} = \begin{cases} \frac{(d+D)\beta}{2(\alpha_q-1)} + \frac{\beta^2}{\alpha_q-2}, & \text{pour } q < Q, \\ \left(N - \frac{Q(d+D)}{2}\right) \frac{\beta}{(\alpha_Q-1)} + \frac{\beta^2}{\alpha_Q-2}, & \text{pour } q = Q. \end{cases} \quad (4.46)$$

4.2.1.2.2 Résultats numériques. Comme on suppose ici que les paramètres λ_q des lois sur chaque segment sont inconnus et aléatoires, on ne peut pas cette fois tracer les évolutions

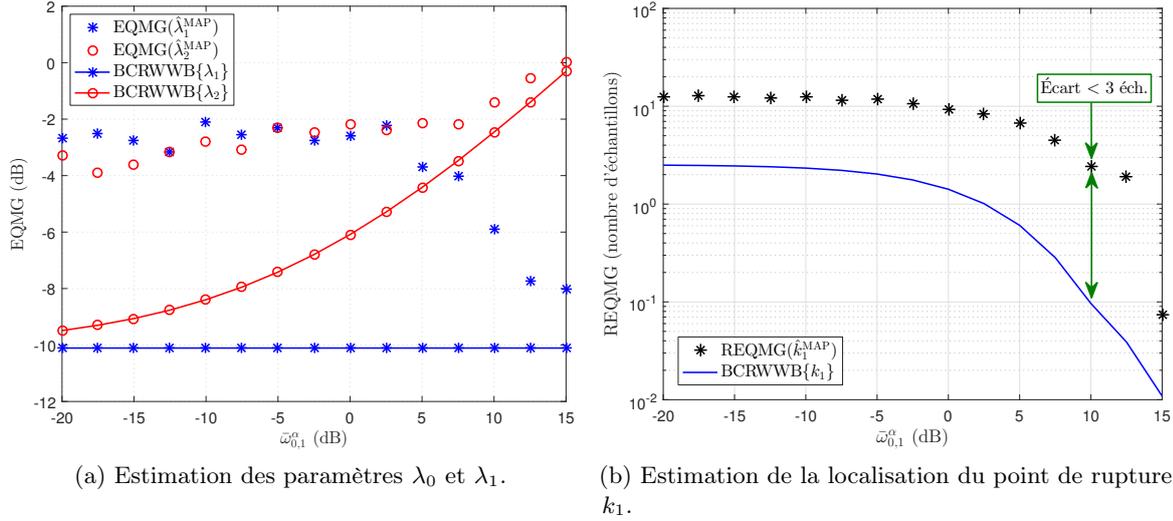


FIGURE 4.18 – Évolutions des EQMG globales de l'estimateur MAP et des éléments de la borne de Cramér-Rao/Weiss-Weinstein bayésienne associée, en fonction de la quantité de changement moyenne $\bar{\omega}_{q-1,q}^\alpha$, dans le cas de données poissonniennes avec $N = 80$ observations et $Q = 1$ rupture, pour les paramètres λ_0 et λ_1 (a), et le point de rupture k_1 (b).

de l'EQMG et de bornes en fonction de la quantité de changement $\omega^\lambda = (\lambda_{q+1} - \lambda_q)^2 / \lambda_q^2$, $\forall q \in \llbracket 0, Q - 1 \rrbracket$. Toutefois, la borne de Cramér-Rao/Weiss-Weinstein bayésienne dépend ici des paramètres $(\alpha_q)_{q=0,\dots,Q}$ et β de la loi *a priori* sur les λ_q . Étant donné que ceux-ci gouvernent la valeur moyenne $\mathbb{E}_{\lambda_q}\{\lambda_q\} = \alpha_q / \beta$ des paramètres de Poisson générés par la loi *a priori* gamma, on peut définir une quantité de changement *moyenne* comme $\bar{\omega}_{q-1,q}^\alpha = (\alpha_q - \alpha_{q-1})^2 / \alpha_{q-1}^2$. Une telle définition implique que plus $\bar{\omega}_{q-1,q}^\alpha$ est grand, plus la quantité de changement entre deux segments successifs est grande, *en moyenne*.

Sur la figure 4.18, nous présentons le cas de $Q = 1$ changement dans $N = 80$ données poissonniennes. Nous ne présentons pas de cas $Q \geq 2$ du fait de la grande complexité calculatoire de la borne, même dans le cas $Q = 1$. Pour autant, la figure 4.18a illustre un comportement assez semblable à celui obtenu en supposant les λ_q déterministes. Sur la figure 4.18b, on observe un écart décroissant entre la REQMG de l'estimateur MAP et la BCRWWB associées au seul point de rupture présent, à mesure que $\bar{\omega}_{0,1}^\alpha$ augmente. La borne proposée fournit donc un aperçu plutôt réaliste du comportement de l'estimation. Pour de faibles $\bar{\omega}_{0,1}^\alpha$ (inférieurs à 0 dB), l'écart entre la BCRWWB et la REQMG de l'estimateur MAP est maximal et de l'ordre de 10 échantillons. Pour $\bar{\omega}_{0,1}^\alpha = 10$ dB, il est de l'ordre de 2.5 échantillons. Pour $\bar{\omega}_{0,1}^\alpha > 13$ dB, l'écart est inférieur à 1 échantillon.

4.2.2 Application à des signaux gaussiens

Nous considérons maintenant l'exemple de signaux gaussiens, en supposant les moyennes et/ou les variances sur chaque segment inconnues. Nous considérons d'abord, dans la section 4.2.2.1, le cas de signaux constants par morceaux, comme nous l'avons fait jusqu'à présent (dans les sections 4.1.1.1 et 4.1.2.1). En d'autres termes, le signal considéré est dans ce cas supposé être une constante noyée dans du bruit additif, i.e., on a, sur chaque segment, pour $n \in \llbracket k_{q-1}, k_q \rrbracket$, $x_n = m_q + b_n$, où $b_n \sim \mathcal{N}(0, \sigma_q^2)$. Nous nous intéressons enfin, dans la section 4.2.2.2, à un cas de signal dont la moyenne sur chaque segment est une fonction (connue) du paramètre à estimer η_q ,

i.e., $m_q = f(\eta_q)$.

4.2.2.1 Cas de signaux constants sur chaque segment

Comme vu dans la section 4.2.1.2, l'emploi de la borne de Cramér-Rao/Weiss-Weinstein bayésienne nécessite de recourir à des techniques d'évaluation numériques d'intégrales. Ceci la rend donc plus complexe à calculer, et diminue quelque peu son intérêt vis-à-vis de la version hybride de la borne de Cramér-Rao/Weiss-Weinstein. Nous ne considérons donc, dans cette section, que la BCRWW hybride. Rappelons que dans le cas gaussien, nous nous intéressons aux deux types de changement possibles, à savoir les changements dans la moyenne et dans la variance.

4.2.2.1.1 Expressions de $\tilde{\mathbf{F}}(\eta_q)$, de $\varphi_{q,m_{\tilde{q}}}(s)$ et de $\varphi_{q,\sigma_{\tilde{q}}^2}(s)$. Pour des signaux gaussiens, l'expression de l'information de Fisher pour la moyenne et la variance m_q et σ_q^2 sur chaque segment est donnée, pour $q \in \llbracket 0, Q \rrbracket$, par

$$\tilde{\mathbf{F}}(\eta_q) = \text{diag} \left(\frac{1}{\sigma_q^2}, \frac{1}{2(\sigma_q^2)^2} \right), \quad (4.47)$$

Concernant les expressions de $\varphi_{q,m_{\tilde{q}}}(s)$ et de $\varphi_{q,\sigma_{\tilde{q}}^2}(s)$, après des manipulations du même type que celles menant à (4.7) et en reconnaissant des moments d'ordre 1 et 2 de distributions gaussiennes, on obtient, pour $q \in \llbracket 0, Q \rrbracket$, $\ell \in \llbracket 1, L \rrbracket$, $\tilde{q} \in \{q-1, q\}$ et $s \in]0, 1[$,

$$\varphi_{q,m_{\tilde{q}}}(s) = \frac{\rho_q(s)}{\sigma_{\tilde{q}}^2} \left(\frac{sm_{q-1}\omega_{q-1,q}^v + (1-s)m_q}{s\omega_{q-1,q}^v + 1-s} - m_{\tilde{q}} \right) \quad (4.48)$$

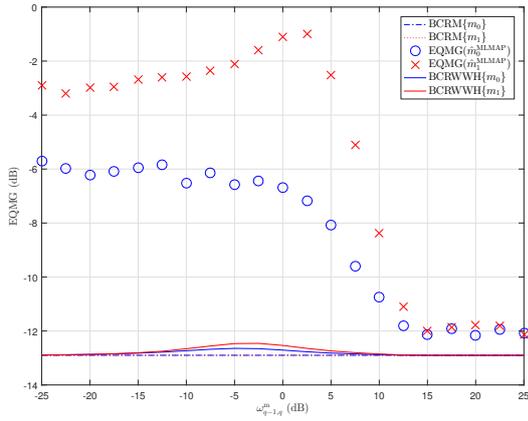
et

$$\varphi_{q,\sigma_{\tilde{q}}^2}(s) = \frac{\rho_{q-1,q}(s)}{2\sigma_{\tilde{q}}^2} \left(\frac{(sm_{q-1}\omega_{q-1,q}^v + (1-s)m_q)^2}{\sigma_{\tilde{q}}^2(s\omega_{q-1,q}^v + 1-s)^2} + \frac{\sigma_{\tilde{q}}^2 - 2m_{\tilde{q}}(sm_{q-1}\omega_{q-1,q}^v + (1-s)m_q) + \frac{m_{\tilde{q}}^2}{\sigma_{\tilde{q}}^2}}{\sigma_{\tilde{q}}^2(s\omega_{q-1,q}^v + 1-s)} - 1 \right). \quad (4.49)$$

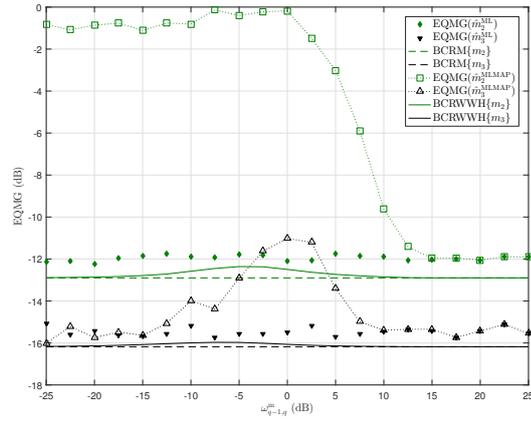
Ces expressions se simplifient dans les cas d'un changement dans la moyenne seulement (i.e., $\omega_{q-1,q}^v = 1$) ou d'un changement dans la variance seulement (i.e., $\omega_{q-1,q}^m = 0$). Elles ne s'annulent pas pour autant, ce qui traduit, dans ces cas aussi, les couplages toujours existant entre les paramètres du signal sur chaque segment et les point de rupture.

4.2.2.1.2 Résultats numériques. On s'intéresse d'abord au cas de changements dans la moyenne seulement (i.e., $\omega_{q-1,q}^v = 1$), puis au cas de changements dans la variance seulement (i.e., $\omega_{q-1,q}^m = 0$).

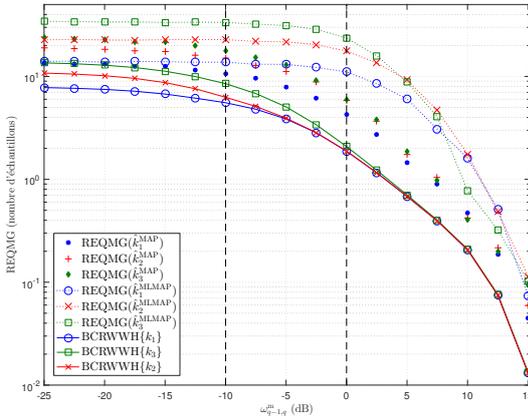
Changement de moyenne. On considère, dans le cas de signaux comportant des changements dans la moyenne, un scénario semblable à celui décrit dans la section 4.1.2.1.3, dont les résultats sont fournis à la figure 4.19. L'évolution des EQMG associées aux estimées des moyennes par l'estimateur MLMAP (figures 4.19a et 4.19b) montrent un effet semblable à un phénomène de décrochement : la quantité de changement $\omega_{q-1,q}^m \approx 15$ dB semble constituer un seuil en-dessous duquel les performances d'estimation des moyennes se dégradent. La valeur de ce seuil semble être plus faible pour la moyenne du dernier segment, autour de 10 dB. La borne de Cramér-Rao/Weiss-Weinstein hybride ne rend quasiment pas compte de ce phénomène, on peut néanmoins observer un très léger renflement dans la forme de la BCRWWH pour $-15 \text{ dB} < \omega_{q-1,q}^m < 10 \text{ dB}$. Pour des



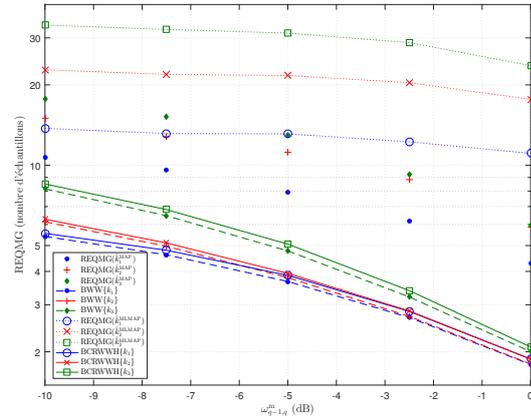
(a) Évolutions des EQMG de l'estimateur ML-MAP des moyennes m_0 et m_1 (cercles bleus pour m_0 , croix rouges pour m_1), des BCRWWH associées (trait plein bleu pour m_0 , trait plein rouge pour m_1), et des bornes de Cramér-Rao modifiées (BCRM) associées (pointillés bleus pour m_0 , pointillés rouges pour m_1). Ces deux dernières courbes sont confondues.



(b) Évolutions des EQMG de l'estimateur ML-MAP des moyennes m_2 et m_3 (pointillés fins, verts avec carrés vides pour m_2 , noirs avec triangles vides pour m_3), des BCRWWH associées (trait plein, vert pour m_2 , noir pour m_3), et des bornes de Cramér-Rao modifiées (BCRM) associées (pointillés longs, verts pour m_2 , noirs pour m_3). Pour comparaison, évolution des EQMG de l'estimateur ML des moyennes, lorsque les ruptures sont supposées connues (losanges verts pleins pour m_2 , triangles noirs pleins pour m_3).



(c) Évolutions des REQMG des estimateurs MAP, lorsque les moyennes et les variances sur chaque segment sont supposées connues (points bleus pour k_1 , croix "+" rouges pour k_2 , losanges pleins verts pour k_3), et MLMAP (cercles bleus pour k_1 , croix rouges pour k_2 , carrés verts pour k_3) des points de rupture, et BCRWWH associées (traits pleins, bleu pour k_1 , rouge pour k_2 , vert pour k_3).



(d) Zoom de la figure (c) entre les deux lignes pointillées noires verticales, soit sur la zone de quantité de changement $-10 \text{ dB} < \omega_{q-1,q}^m < 0 \text{ dB}$. En pointillés longs sont ajoutées, pour comparaison, les bornes de Weiss-Weinstein pour les points de rupture (i.e., correspondant au cas où les moyennes sur chaque segment sont connues; voir section 4.1).

FIGURE 4.19 – Évolutions des EQM globales des estimateurs des moyennes $(m_q)_{q=0,1}$ (figure (a)) et $(m_q)_{q=2,3}$ (figure (b)), et des estimateurs des points de ruptures $(k_q)_{q=1,\dots,Q}$ (figures (c) et (d)), comparées aux bornes associées, dans le cas de données gaussiennes changeant de moyenne, avec $N = 100$ observations et $Q = 3$ ruptures, en fonction de la quantité de changement $\omega_{q-1,q}^m$.

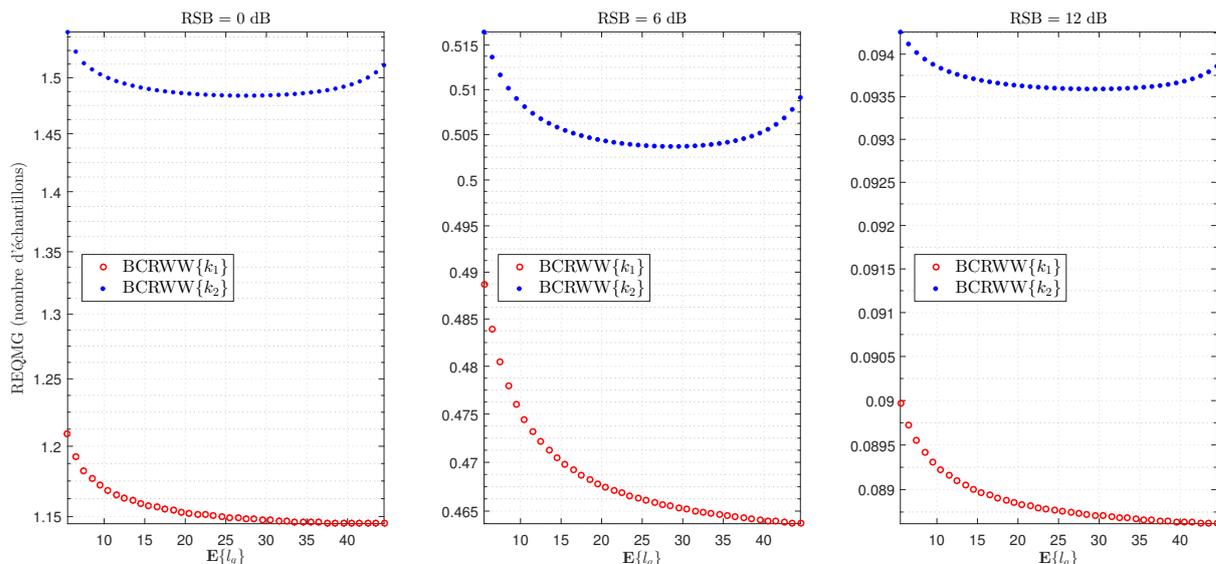


FIGURE 4.20 – Cas de $Q = 2$ changements dans la moyenne m_q de $N = 100$ observations gaussiennes. Évolution des bornes de Cramér-Rao/Weiss-Weinstein associées aux deux ruptures k_1 et k_2 (cercles rouges pour k_1 , points bleus pour k_2), en fonction de la longueur moyenne *a priori* des segments $[k_{q-1}, k_q]$, $q = 1, 2$, autrement dit $\mathbb{E}\{l_q\}$, pour trois valeurs de la quantité de changement différentes.

quantités de changement supérieures à 15 dB, on constate que l'EQMG du MLMAP se rapproche subitement, pour chaque moyenne, de la BCRWWH associée. Il subsiste asymptotiquement un écart perceptible entre les deux, qui s'explique par le fait que la « partie Cramér-Rao » de la BCRWWH, à savoir $\mathbf{V}_{11}^{\boldsymbol{\eta}^*}$ et qui est celle représentée sur les figures 4.19a et 4.19b, correspond en fait, comme on l'a vu à la section 2.1.5.4.2, à la « borne de Cramér-Rao modifiée, » (BCRM) au sens de [GRM98], sur le vecteur $\boldsymbol{\eta}^*$. Pour comparaison, nous traçons également, sur les figures 4.19a et 4.19b, les BCRM associées aux estimations des moyennes sur chaque segment : il s'agit donc des éléments diagonaux de $(\mathbf{V}_{11}^{\boldsymbol{\eta}^*})^{-1}$, soit la borne de Cramér-Rao sur les moyennes lorsque les points de rupture sont inconnus mais considérés comme des paramètres de nuisance, et ne font donc pas partie du vecteur de paramètres à estimer. Or, on ne peut pas s'attendre à ce que la BCRM offre la même précision³ que la BCR classique ou la EBCR, car l'information de Fisher de la BCRM est moyennée par rapport à la loi *a priori* sur les paramètres de nuisance, en l'occurrence \mathbf{k} .

En ce qui concerne le comportement de l'estimation des points de rupture (figures 4.19c et 4.19d), on constate que, bien que les REQMG de l'estimateur MLMAP (i.e., lorsque les moyennes sont supposées inconnues) des ruptures soit plus élevées que celles du MAP (i.e., lorsque les moyennes sont connues), la BCRWWH n'en rend quasiment pas compte : l'écart observé entre la BCRWWH et la BWW (qui correspond à la borne pour les points de rupture avec les moyennes et variances sur chaque segment connues, i.e., celle proposée à la section 4.1.2.1) est extrêmement faible (figure 4.19d). Le comportement de la BCRWWH, concernant les points de rupture, est donc en tout point similaire à celui de la BWW, analysé à la section 4.1.2.1.3.

Sur la figure 4.20, on considère le cas de $Q = 2$ points de rupture, où l'on représente les BCRWWH associées à k_1 et k_2 , en fonction de la longueur moyenne des segments $\llbracket k_q, k_{q+1} \rrbracket$,

3. On pourrait le montrer par un raisonnement similaire à celui aboutissant sur (2.73) (voir chapitre 2, section 2.1.5.3.1).

$q \in \{0, 1, 2\}$ (avec les définitions conventionnelles habituelles $k_0 \triangleq 0$ et $k_3 = N$), c'est-à-dire $\mathbb{E}\{l_q\} = (d + D)/2$. Nous avons considéré le cas de $N = 100$ observations, et fixé $\Delta = 10$. Nous pouvons ainsi observer la forme de la BCRWWH en fonction de l'écart moyen entre deux points de rupture consécutifs. On peut observer que lorsque k_1 est plus proche en moyenne du bord de la fenêtre d'observation (faibles valeurs de $\mathbb{E}\{l_1\}$), la valeur de la BCRWWH associée augmente. De même pour k_2 , lorsqu'il est en moyenne proche de k_1 (faibles valeurs de $\mathbb{E}\{l_2\}$), ou lorsqu'il est proche de N (grandes valeurs de $\mathbb{E}\{l_2\}$, ou de manière équivalente, faibles valeurs de $\mathbb{E}\{l_3\} = N - \mathbb{E}\{l_2\} - \mathbb{E}\{l_1\}$). Le comportement de la BCRWWH rend donc compte de l'effet intuitif selon lequel la qualité de l'estimation se dégrade lorsque le nombre d'observations sur chaque segment diminue.

Changement de variance. Nous fournissons, sur la figure 4.21, les résultats de simulation dans le cas de $Q = 3$ changements dans la variance de $N = 100$ observations gaussiennes. Les autres paramètres sont identiques à ceux du scénario décrit à la section 4.2.2.1.2 dans le cas du changement de variance. Les simulations ont été menées dans le cas où les moyennes sur chaque segment sont également inconnues. Le comportement de leur estimation n'est pas représenté ici, mais est en tout point similaire à celui des variances (figure 4.21a). Les mêmes remarques concernant la précision de la BCRWWH pour les variances (figure 4.21a), et son comportement similaire au cas où les variances sont connues pour les points de rupture (figure 4.21b) que celles faites dans le cas de données poissonniennes (voir section 4.2.1.1) s'appliquent. Les seules différences concernent la zone de décrochement pour les variances : elle s'étend de 0 dB à 25 dB pour σ_0^2 , de 0 dB à 27 dB pour σ_1^2 et de 0 dB à 20 dB pour σ_2^2 . Pour σ_3^2 , on n'est pas en mesure de déterminer l'existence d'une telle zone.

4.2.2.2 Cas d'un signal sinusoïdal complexe comportant une rupture de fréquence

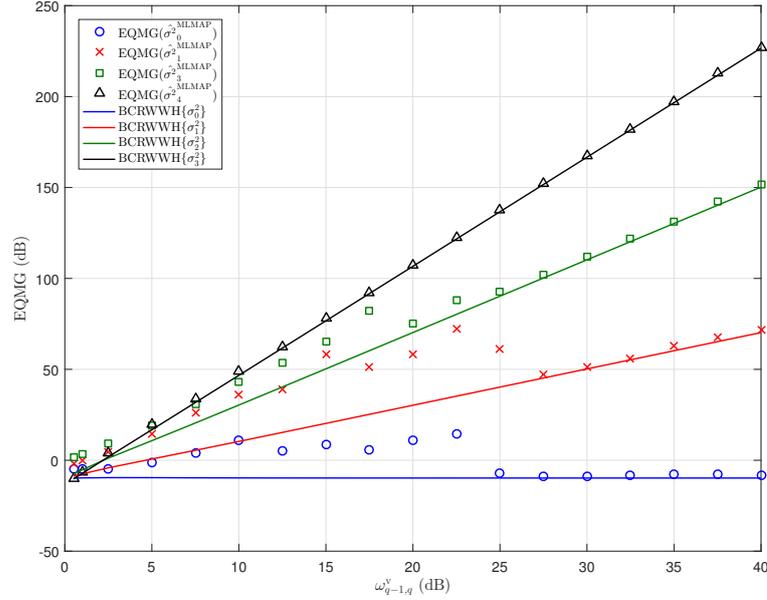
Dans les sections précédentes, nous avons présenté les résultats obtenus pour des signaux constants sur chaque segment uniquement. Dans un effort d'illustrer les diverses formes de signaux pouvant être considérées, nous envisageons ici le cas de signaux variant dans le temps sur chaque segment. Spécifiquement, nous nous intéressons au cas d'une rupture dans la fréquence d'un signal sinusoïdal complexe. Dans ce cas précis, il y a lieu de s'intéresser à la borne de Weiss-Weinstein, à la fois pour l'estimation du point de rupture mais également pour l'estimation des fréquences avant et après la rupture. Nous savons en effet que le problème d'estimation d'une fréquence implique un phénomène de décrochement, qui implique une dégradation importante des performances d'estimation en-dessous d'un certain rapport signal à bruit (RSB) [RB74]. Or, nous savons également que la borne de Weiss-Weinstein, contrairement à celle de Cramér-Rao, permet de rendre compte de ce phénomène [VB07].

Nous revenons donc ici à un cadre d'estimation complètement bayésien, où les fréquences seront donc supposées aléatoires.

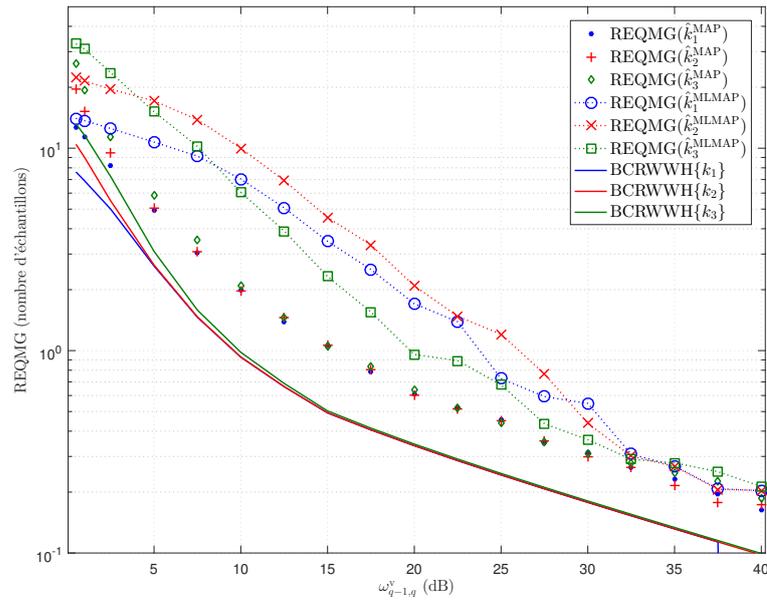
4.2.2.2.1 Modèle et hypothèses. Nous considérons le modèle d'une sinusoïde complexe dont la fréquence f_0 (normalisée, i.e., $f_0 \in [-\frac{1}{2}, \frac{1}{2}]$), supposée inconnue, passe à une valeur f_1 inconnue aussi, à un instant k_1 lui aussi inconnu. Ce modèle peut s'écrire de la manière suivante :

$$\begin{cases} x_n = e^{i2\pi f_0 n} + b_n & \text{pour } n = 1, \dots, k_1, \\ x_n = e^{i2\pi f_1 n} + b_n & \text{pour } n = k_1 + 1, \dots, N, \end{cases} \quad (4.50)$$

où $b_n \sim \mathcal{CN}(0, \sigma^2)$, i.e., il s'agit d'un bruit complexe, gaussien, circulaire, indépendant, identiquement distribué (i.i.d.), de moyenne nulle et de variance σ^2 . En concaténant les observations dans



(a) Évolutions des EQMG de l'estimateur MLMAP des variances (σ_q^2) $_{q=0,\dots,Q}$ (cercles bleus pour σ_0^2 , croix rouges pour σ_1^2 , carrés verts pour σ_2^2 , triangles noirs pour σ_3^2), et des BCRWWH associées (traits pleins, bleu pour σ_0^2 , rouge pour σ_1^2 , vert pour σ_2^2 , noir pour σ_3^2).



(b) Évolutions des REQMG des estimateurs MAP, lorsque les moyennes et les variances sur chaque segment sont supposées connues (points bleus pour k_1 , croix "+" rouges pour k_2 , losanges verts pour k_3), et MLMAP (cercles bleus pour k_1 , croix "x" rouges pour k_2 , carrés verts pour k_3) des points de rupture, et BCRWWH associées (traits pleins, bleu pour k_1 , rouge pour k_2 , vert pour k_3).

FIGURE 4.21 – Cas de $Q = 3$ changements dans la variance de $N = 100$ observations gaussiennes. Évolution des EQM globales des estimateurs des variances sur chaque segment (a) et des points de rupture (b), et des bornes de Cramér-Rao/Weiss-Weinstein hybrides associées, en fonction de la quantité de changement $\omega_{0,1}^m$.

le vecteur $\mathbf{x} = [x_1, \dots, x_n]^T$, la modèle (4.50) se réécrit

$$\mathbf{x} = \mathbf{m}(\boldsymbol{\theta}) + \mathbf{b}, \quad (4.51)$$

où $\boldsymbol{\theta} \triangleq [f_0, f_1, k_1]^T$ est le vecteur de paramètres à estimer, et $\mathbf{m}(\boldsymbol{\theta}) \triangleq [\mathbf{m}_{1:k_1}(f_0), \mathbf{m}_{k_1+1:N}(f_1)]^T$, avec, pour $(n, n') \in \llbracket 1, N \rrbracket^2$ et $f \in [-\frac{1}{2}, \frac{1}{2}]$, $\mathbf{m}_{n:n'}(j) \triangleq [e^{i2\pi f n}, e^{i2\pi f(n+1)}, \dots, e^{i2\pi f n'}]^T$.

Suivant une approche bayésienne, on suppose que les fréquences f_0 et f_1 sont aléatoires, et sont distribuées uniformément sur l'intervalle $[-\frac{1}{2}, \frac{1}{2}]$, i.e., $(f_q)_{q=0,1} \sim \mathcal{U}([-\frac{1}{2}, \frac{1}{2}])$, et que le point de rupture k_1 suit une distribution uniforme discrète entre 1 et $N - 1$, i.e., $k_1 \sim \mathcal{U}(\llbracket 1, N - 1 \rrbracket)$. Ceci aboutit à la distribution *a priori* jointe sur le vecteur $\boldsymbol{\theta}$ suivante :

$$\pi(\boldsymbol{\theta}) = \pi(f_0, f_1, k_1) = \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(f_0) \cdot \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(f_1) \cdot \frac{1}{N-1} \mathbb{1}_{\llbracket 1, N-1 \rrbracket}(k_1). \quad (4.52)$$

Nous calculons la borne de Weiss-Weinstein associée au modèle (4.50), pour l'estimation du paramètre $\boldsymbol{\theta} = [f_0, f_1, k_1]^T$, avec la loi *a priori* (4.52).

4.2.2.2 Expression de la borne de Weiss-Weinstein. Nous donnons dans cette section l'expression des éléments de la matrice \mathbf{G} de taille 3×3 , définis dans (2.80). Nous optons pour les simplifications usuelles discutées dans la section 2.1.5.3.2, c'est-à-dire i) $s_1 = s_2 = s_3 = 1/2$, et ii) $\mathbf{H} = \text{diag}(h_1, h_2, h_3)$. Pour plus de clarté sur les quantités manipulées, nous renommons h_1 en h_{f_0} , h_2 en h_{f_1} , et h_3 en h_{k_1} . Nous avons donc $\mathbf{H} = \text{diag}(h_{f_0}, h_{f_1}, h_{k_1})$. Comme dans la section 3.1.4, on notera les quantités associées aux éléments de la matrice \mathbf{G} simplement comme $\nu(h_a, h_b, k_1) = \nu(1/2, 1/2, h_a, h_b, k_1)$, $M(h_a, h_b, k_1) = M(1/2, 1/2, h_a, h_b, k_1)$, $e^{\tilde{\mu}(1/2, 1/2, h_a, h_b)} = e^{\tilde{\mu}(h_a, h_b)}$ et $e^{\mu(h_a)} = e^{\tilde{\mu}(h_a, 0)}$.

On obtient, du fait de l'indépendance supposée des fréquences une matrice \mathbf{G} de la forme

$$\mathbf{G} = \begin{bmatrix} G_{f_0 f_0} & 0 & G_{f_0 k_1} \\ 0 & G_{f_1 f_1} & G_{f_1 k_1} \\ G_{f_0 k_1} & G_{f_1 k_1} & G_{k_1 k_1} \end{bmatrix}, \quad (4.53)$$

avec, pour $q \in \{0, 1\}$,

$$G_{f_q f_q} = \frac{2(1 - |h_{f_q}| - e^{\mu(2h_{f_q})})}{e^{2\mu(h_{f_q})}}, \quad (4.54)$$

$$G_{k_1 k_1} = \frac{2(1 - \frac{|h_{k_1}|}{N-1} - e^{\mu(2h_{k_1})})}{e^{2\mu(h_{k_1})}}, \quad (4.55)$$

et les autres termes croisés ne s'exprimant pas autrement, pour $q \in \{0, 1\}$, que par (2.80), c'est-à-dire, avec prise en compte des simplifications,

$$G_{f_q k_1} = \frac{e^{\tilde{\mu}(h_{f_q}, h_{k_1})} + e^{\tilde{\mu}(-h_{f_q}, -h_{k_1})} - e^{\tilde{\mu}(h_{f_q}, -h_{k_1})} - e^{\tilde{\mu}(-h_{f_q}, h_{k_1})}}{e^{\mu(h_{f_q})} e^{\mu(h_{k_1})}}, \quad (4.56)$$

de sorte qu'il suffit finalement d'avoir les expressions, pour $q \in \{0, 1\}$, de $e^{\mu(h_{f_q})}$, $e^{\mu(h_{k_1})}$ et $e^{\tilde{\mu}(h_{f_q}, h_{k_1})}$. Après calculs, on obtient les expressions suivantes pour chacune de ces quantités :

— pour $e^{\mu(h_{f_0})}$:

$$e^{\mu(h_{f_0})} = \frac{1 - |h_{f_0}|}{N-1} \sum_{k=1}^{N-1} \exp \left\{ -\frac{1}{4\sigma^2} \left(2k+1 - \frac{\sin(\pi h_{f_0}(2k+1))}{\sin(\pi h_{f_0})} \right) \right\}, \quad (4.57)$$

— pour $e^{\mu(h_{f_1})}$:

$$e^{\mu(h_{f_1})} = \frac{1 - |h_{f_1}|}{N - 1} \sum_{k=1}^{N-1} \exp \left\{ -\frac{1}{2\sigma^2} \left(N - k - \cos(\pi h_{f_1}(N + k + 1)) \frac{\sin(\pi h_{f_1}(N - k))}{\sin(\pi h_{f_1})} \right) \right\}, \quad (4.58)$$

— pour $e^{\mu(h_{k_1})}$:

$$e^{\mu(h_{k_1})} = \frac{\exp \left\{ -\frac{|h_{k_1}|}{2\sigma^2} \right\}}{N - 1} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{k=1}^{N-1-|h_{k_1}|} \exp \left\{ \frac{1}{2\sigma^2} \sum_{n=k+1}^{k+|h_{k_1}|} \cos(2\pi(f_1 - f_0)n) \right\} df_0 df_1, \quad (4.59)$$

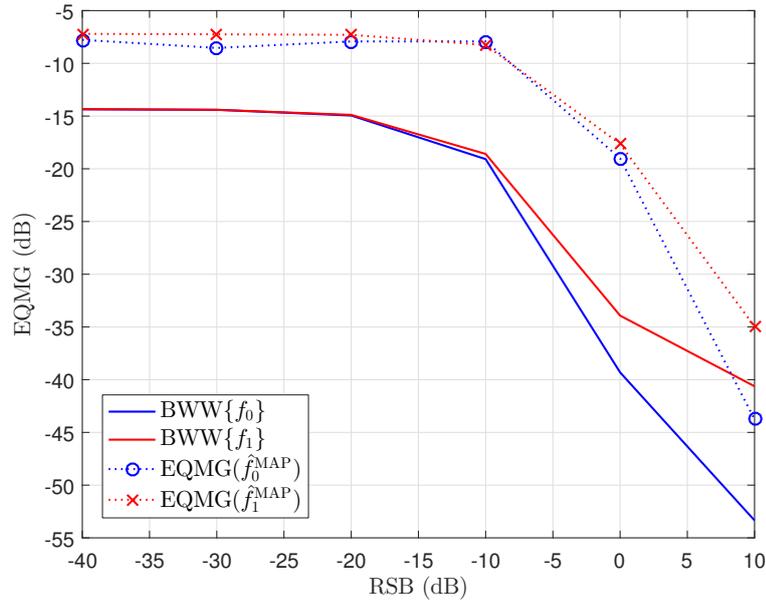
— et enfin, pour $e^{\tilde{\mu}(h_{f_0}, h_{k_1})}$:

$$\begin{aligned} e^{\tilde{\mu}(h_{f_0}, h_{k_1})} &= \frac{1}{N - 1} \exp \left\{ -\frac{|h_{k_1}|}{2\sigma^2} \right\} \\ &\times \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2} + \max\{-h_{f_0}, 0\}}^{\frac{1}{2} - \max\{h_{f_0}, 0\}} \sum_{k=1}^{N-1-|h_{k_1}|} \exp \left\{ -\frac{1}{2\sigma^2} \left[k - \sum_{n=1}^k \cos(2\pi h_{f_0} n) \right. \right. \\ &\quad \left. \left. - \sum_{n=k+1}^{k+|h_{k_1}|} \cos \left(2\pi (f_1 - f_0 - h_{f_0} \mathbb{1}_{\mathbb{R}_+}(h_{k_1})) n \right) \right] \right\} df_0 df_1, \end{aligned} \quad (4.60)$$

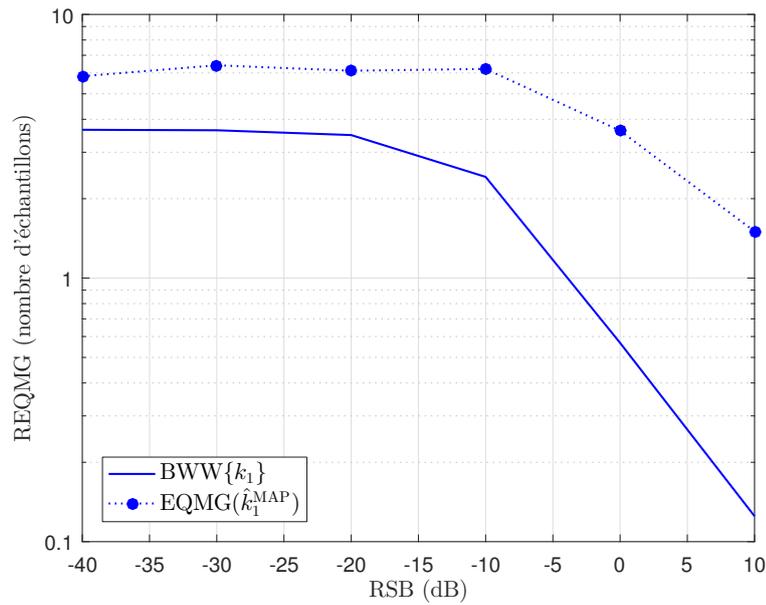
— pour $e^{\tilde{\mu}(h_{f_1}, h_{k_1})}$:

$$\begin{aligned} e^{\tilde{\mu}(h_{f_1}, h_{k_1})} &= \frac{1}{N - 1} \exp \left\{ -\frac{N}{2\sigma^2} \right\} \\ &\times \int_{-\frac{1}{2} + \max\{-h_{f_1}, 0\}}^{\frac{1}{2} - \max\{h_{f_1}, 0\}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{k=1}^{N-1-|h_{k_1}|} \exp \left\{ \frac{1}{2\sigma^2} \left[k + \sum_{n=k+|h_{k_1}|+1}^N \cos(2\pi h_{f_1} n) \right. \right. \\ &\quad \left. \left. + \sum_{n=k+1}^{k+|h_{k_1}|} \cos \left(2\pi (f_1 + h_{f_1} \mathbb{1}_{\mathbb{R}_+}(h_{k_1}) - f_0) n \right) \right] \right\} df_0 df_1, \end{aligned} \quad (4.61)$$

4.2.2.2.3 Résultats numériques Les résultats numériques fournis à la figure 4.22 correspondent au cas de $N = 20$ observations, où les EQMG et REQMG approchées de l'estimateur MAP ont été obtenues après $N_{MC} = 1000$ simulations de Monte-Carlo. La plage de rapport signal-sur-bruit $RSB = 1/\sigma^2$ considérée s'étend de -40 dB à 10 dB. Pour des RSB plus élevés, des problèmes numériques empêchent de calculer la borne de Weiss-Weinstein. On remarque néanmoins que celle-ci est plutôt en accord avec le comportement de l'EQM globale pour chacun des paramètres : en ce qui concerne les fréquences (figure 4.22a), la zone de non-information pour des RSB plus faibles que -20 dB est bien retranscrite. L'écart entre la BWW et les EQMG des fréquences tend même à diminuer à mesure que le RSB augmente. Concernant le point de rupture (figure 4.22b), le même type de comportement que celui déjà observé, par exemple dans le cas de données gaussiennes ou poissonniennes avec $Q = 1$ (voir figures 4.4, 4.8, ou 4.18b). L'écart entre la BWW et la REQMG de l'estimateur MAP de k_1 est d'environ 2 échantillons pour un RSB de 0 dB, et diminue progressivement à mesure que les RSB augmente.



(a) Évolutions des EQMG de l'estimateur MAP des fréquences f_0 et f_1 (cercles bleus pour f_0 , croix rouges pour f_1), et des BWW associées (traits pleins, bleu pour f_0 , rouge pour f_1).



(b) Évolutions de la REQMG de l'estimateur MAP du point de rupture k_1 (points bleus), et de la BWW associée (trait plein bleu).

FIGURE 4.22 – Cas de $Q = 1$ changement dans la fréquence d'une sinusoïde complexe noyée dans bruit gaussien, avec $N = 20$ observations. Évolution des EQM globales des estimateurs MAP des fréquences avant et après rupture (a) et du point de rupture (b), avec dans chaque cas les bornes de Weiss-Weinstein associées, en fonction du rapport signal-sur-bruit $\text{RSB} = 1/\sigma^2$.

Chapitre 5

Conclusion et perspectives

Le travail rapporté dans ce manuscrit a concerné l'étude des limites de performances d'estimation, en terme d'erreur quadratique moyenne, pour des signaux comportant des points de rupture inconnus. Chronologiquement, notre démarche s'est organisée de la manière suivante :

- Dans un premier temps, nous avons étudié le problème simple d'un seul point de rupture contenu dans des données gaussiennes, poissonniennes et exponentielles, en supposant les paramètres de ces distributions connus, avant et après le point de rupture. Nous avons montré que la borne de Weiss-Weinstein se réduisait au calcul de la distance de Bhattacharyya entre les deux distributions. Les résultats obtenus ont montré un écart significatif entre la borne proposée et celles existant dans la littérature [BREC15, BBRE17].
- Ces résultats ont été étendus à une famille plus générale de bornes de Weiss-Weinstein et au cas de plusieurs points de rupture. Nous avons montré que l'obtention de cette borne ne se réduisait pas à la concaténation des bornes associées à un seul point de rupture. En particulier, nous avons mis en évidence l'influence du couplage entre deux points de rupture consécutifs. Nous avons montré que la borne ainsi calculée se réduit cette fois-ci au calcul de la divergence de Rényi entre les distributions de segments consécutifs. Du fait de la nature multivariée du problème, l'obtention de la borne la plus précise fait intervenir une relation d'ordre (ordre partiel de Löwner) entre matrices semi-définies positives. Afin de résoudre ce problème, nous nous sommes ramenés à un problème d'optimisation convexe, résolu par une procédure de programmation semi-définie positive [BREC17].
- Sur la base de ces résultats encourageants, nous avons étudié le cas plus réaliste où les paramètres des signaux sur chaque segment sont inconnus. Une première analyse a concerné un problème potentiellement utile en traitement du signal radar, et en télécommunications numériques. Celui-ci consiste en un unique saut abrupt de fréquence dans des données gaussiennes, lorsque les deux fréquences (avant et après rupture) sont inconnues. Ce problème étant connu pour faire intervenir un décrochement de l'EQM des estimateurs des fréquences, nous avons choisi d'étudier la borne de Weiss-Weinstein pour l'estimation des fréquences et du point de rupture [BER16].
- Dans la lignée de l'étude précédente, mais dans le cas de points de rupture multiples, nous avons étendu les résultats de [BREC17] au cas de paramètres inconnus des distributions sur chaque segment. Nous avons étudié le cas des données gaussiennes avec changement de moyenne et/ou de variance ainsi que le cas de données poissonniennes avec changement de paramètre. Dans un contexte complètement bayésien, nous avons proposé la borne de Cramér-Rao/Weiss-Weinstein bayésienne [BERT17], et dans un contexte hybride où les paramètres des distributions sont supposés inconnus et déterministes, nous avons proposé la borne de Cramér-Rao/Weiss-Weinstein hybride [BERT].

- Enfin, notre étude étant largement menée dans un contexte bayésien, nous nous sommes intéressés à l'influence de la distribution *a priori* sur la position des points de rupture : nous avons pour cela considéré des marches aléatoires uniforme et de Poisson, ainsi que des lois uniforme sur des segments distincts, et globalement uniforme. Nous avons montré que la marche aléatoire uniforme, que nous avons privilégiée, offre un bon compromis entre complexité calculatoire et précision en terme d'EQM [BRE18].

A l'issue de ce travail, les perspectives pouvant être envisagées, classées de court à long-terme, sont les suivantes :

- Nous nous sommes focalisés sur les bornes de la famille Weiss-Weinstein, mais il existe dans la littérature une autre famille de bornes : la famille dite de « Ziv-Zakai » [Bel95, BEV96], connue pour fournir une précision similaire à la famille de Weiss-Weinstein. Cependant, cette famille repose sur un concept différent, à savoir un test d'hypothèse binaire et l'inégalité de Kotelnikov, et nécessitera par conséquent des techniques de calcul adaptées.
- Les travaux proposés ici se sont intéressés au cas important de données indépendantes. Cependant, bon nombre de signaux aléatoires ne vérifient pas cette propriété. Une solution pour les modéliser est par exemple d'utiliser des processus auto-régressifs, GARCH, FARIMA, etc. Ce type de signaux est susceptible de comporter des points de rupture [DTD07, SB13]. Une étude préliminaire pourrait concerner la mise en place de la borne de Weiss-Weinstein pour un processus auto-régressif d'ordre 1 avec un unique point de rupture.
- Jusqu'à maintenant, les bornes étudiées se fondaient sur un nombre de ruptures Q supposé connu et différent de zéro (par exemple grâce à une détection préalable). Une perspective intéressante, serait de nous intéresser à une borne dont l'expression tient compte de l'issue du test de détection (présence ou non d'une rupture). Celle-ci pourrait ensuite être étendue au cas de points de rupture multiples. Une telle borne serait donc d'un type différent de celles précédemment présentées, puisqu'elle prendrait en compte les divers ordres (ou dimensions) possibles du modèle. La prise en compte du test de détection pourrait également être complétée, à plus long terme, par une analyse des performances d'estimation pour le problème « on-line », où l'utilisation des bornes séquentielles pourrait s'avérer prometteuse [TMN98, RO04, RN05].
- Une autre perspective consisterait à obtenir des bornes inférieures sur d'autres risques que le risque quadratique, comme le risque 0-1, particulièrement adapté aux paramètres discrets. Ces bornes pourraient ainsi offrir un meilleur aperçu des performances de l'estimation de points de rupture. Deux versions d'une borne « de type Chapman-Robbins » sur le risque 0-1, l'une locale et l'autre globale, mais pour des modèles stationnaires, ont notamment été proposées dans [CS12]. Il est démontré en particulier que l'estimateur ML atteint asymptotiquement la version globale de cette borne. Ces résultats ne s'appliquent pas à des modèles de signaux comportant des ruptures, par essence non-stationnaires. Il serait donc intéressant d'étudier dans quelle mesure les travaux de [CS12] pourraient être modifiés et adaptés au cas de signaux comportant d'éventuelles ruptures.
- Le principal inconvénient qui a motivé ces travaux est le fait que les conditions de régularité concernant la borne de Cramér-Rao ne sont pas respectées du fait de la nature discrète des paramètres à estimer. Il existe des problèmes autres que celui de la localisation de points de rupture où les paramètres à estimer sont, par essence même, discrets. On peut penser au GNSS [GRM16], où il est intéressant d'estimer des entiers multiples des longueurs d'ondes lors d'une triangulation, ou encore à des problèmes de télécommunication numérique, où les signaux transmis proviennent d'une constellation discrète [GRM98]. Il serait donc intéressant de proposer une étude pour ce type de problèmes, où les résultats en termes

performance ne sont obtenus qu'à partir de simulations de type Monte-Carlo.

Annexe A

Expression de la matrice $G(\mathbf{H}, \mathbf{s})$

Nous donnons ici les détails des calculs permettant d'aboutir au résultat du chapitre 3, section 3.1.3, en utilisant la méthode décrite dans la section 3.1.2.

A.1 Expressions de $\nu(\alpha, \beta, h_a, h_b, \mathbf{k})$

Nous exposons ici la méthodologie permettant d'obtenir les expressions de $\nu(\alpha, \beta, h_a, h_b, \mathbf{k})$ pour la distribution *a priori* présentée en (3.3). En réexprimant cette distribution évaluée au point test $\mathbf{k} + \mathbf{h}_a$, on a

$$\pi_{\mathbf{d}}(\mathbf{k} + \mathbf{h}_a) = \frac{1}{\Delta^Q} \mathbb{1}_{\mathcal{K}_{\mathbf{d}}}(\mathbf{k} + \mathbf{h}_a) = \frac{1}{\Delta^Q} \mathbb{1}_{\mathcal{K}_{\mathbf{d}} - \mathbf{h}_a}(\mathbf{k}) \quad (\text{A.1})$$

où $\mathcal{K}_{\mathbf{d}} - \mathbf{h}_a$ représente, avec un abus de notation, l'ensemble $\mathcal{K}_{\mathbf{d}}$ translaté de $-\mathbf{h}_a$, c'est-à-dire

$$\mathcal{K}_{\mathbf{d}} - \mathbf{h}_a \triangleq \{\mathbf{k} - \mathbf{h}_a \mid \mathbf{k} \in \mathcal{K}_{\mathbf{d}}\}. \quad (\text{A.2})$$

Ainsi, l'expression de $\nu(\alpha, \beta, h_a, h_b, \mathbf{k})$ s'obtient en injectant (3.3) et (A.1) dans (3.7). On obtient donc

$$\nu(\alpha, \beta, h_a, h_b, \mathbf{k}) = \frac{1}{\Delta^Q} \mathbb{1}_{\mathcal{K}_{\mathbf{d}} - \mathbf{h}_a}(\mathbf{k}) \mathbb{1}_{\mathcal{K}_{\mathbf{d}} - \mathbf{h}_b}(\mathbf{k}) \mathbb{1}_{\mathcal{K}_{\mathbf{d}}}(\mathbf{k}) = \frac{1}{\Delta^Q} \mathbb{1}_{(\mathcal{K}_{\mathbf{d}} - \mathbf{h}_a) \cap (\mathcal{K}_{\mathbf{d}} - \mathbf{h}_b) \cap \mathcal{K}_{\mathbf{d}}}(\mathbf{k}) \quad (\text{A.3})$$

où l'intersection $(\mathcal{K}_{\mathbf{d}} - \mathbf{h}_a) \cap (\mathcal{K}_{\mathbf{d}} - \mathbf{h}_b) \cap \mathcal{K}_{\mathbf{d}}$, notée $(\mathcal{K}'_{\mathbf{d}})_{h_a, h_b}$ dans la suite, s'exprime comme un produit cartésien de Q ensembles d'entiers consécutifs de la forme $\llbracket (k_q)_{\min}, (k_q)^{\max} \rrbracket$, $q = 1, \dots, Q$, i.e.,

$$(\mathcal{K}'_{\mathbf{d}})_{h_a, h_b} \triangleq (\mathcal{K}_{\mathbf{d}} - \mathbf{h}_a) \cap (\mathcal{K}_{\mathbf{d}} - \mathbf{h}_b) \cap \mathcal{K}_{\mathbf{d}} = \prod_{q=1}^Q \llbracket (k_q)_{\min}, (k_q)^{\max} \rrbracket. \quad (\text{A.4})$$

Cet ensemble représente le domaine de la sommation sur \mathbf{k} dans (3.6), et $(k_q)_{\min}$ et $(k_q)^{\max}$, $q = 1, \dots, Q$ représentent ainsi les bornes des Q sommes sous-jacentes. Les expressions de $(k_q)_{\min}$ et $(k_q)^{\max}$ varient selon les cas *diag*, *psdiag* et *ts*, et sont donc données dans les tableaux A.1, A.2 et A.3.

q	$(k_q)_{\min}$	$(k_q)_{\max}$
a	$k_{a-1} + d + \max(-h_a, -h_b, 0)$	$h_{a-1} + D - \max(h_a, h_b, 0)$
$a + 1$ (si $a + 1 \leq Q$)	$k_a + d + \max(h_a, h_b, 0)$	$k_a + D - \max(-h_a, -h_b, 0)$
$q \neq a, a + 1$	$k_{q-1} + d$	$k_{q-1} + D$

TABLEAU A.1 – Expressions des bornes $(k_q)_{\min}$ et $(k_q)_{\max}$ des Q sommes sur k_1, \dots, k_Q dans l'équation (3.6), dans le cas **diag** ($b = a$). Dans ce tableau, on a $h_b = \pm h_a$.

q	$(k_q)_{\min}$	$(k_q)_{\max}$
a	$k_{a-1} + d + \max(-h_a, 0)$	$k_{a-1} + D - \max(h_a, 0)$
$a + 1$	$k_a + d + \max(h_a, -h_{a+1}, 0)$	$k_a + D - \max(-h_a, h_{a+1}, 0)$
$a + 2$ (si $a + 2 \leq Q$)	$k_{a+1} + d + \max(h_{a+1}, 0)$	$k_{a+1} + D - \max(-h_{a+1}, 0)$
$q \neq a, a + 1, a + 2$	$k_{q-1} + d$	$k_{q-1} + D$

TABLEAU A.2 – Expressions des bornes $(k_q)_{\min}$ et $(k_q)_{\max}$ des Q sommes sur k_1, \dots, k_Q dans l'équation (3.6), dans le cas **psdiag** ($b = a + 1$).

q	$(k_q)_{\min}$	$(k_q)_{\max}$
a	$k_{a-1} + d + \max(-h_a, 0)$	$k_{a-1} + D - \max(h_a, 0)$
$a + 1$	$k_a + d + \max(h_a, 0)$	$k_a + D - \max(-h_a, 0)$
b	$k_{b-1} + d + \max(-h_b, 0)$	$k_{b-1} + D - \max(h_b, 0)$
$b + 1$ (si $b + 1 \leq Q$)	$k_b + d + \max(h_b, 0)$	$k_b + D - \max(-h_b, 0)$
$q \neq a, a + 1, b, b + 1$	$k_{q-1} + d$	$k_{q-1} + D$

TABLEAU A.3 – Expressions des bornes $(k_q)_{\min}$ et $(k_q)_{\max}$ des Q sommes sur k_1, \dots, k_Q dans l'équation (3.6), dans le cas **ts** ($b > a + 1$).

A.2 Expressions de $M(\alpha, \beta, h_a, h_b, \mathbf{k})$

En écrivant la vraisemblance des observations donnée en (3.4) comme

$$p(\mathbf{X} | \mathbf{k}) = \prod_{q=0}^Q \prod_{n=k_q+1}^{k_{q+1}} p_{\eta_q}(\mathbf{x}_n), \quad (\text{A.5a})$$

$$= \left(\prod_{\substack{q=0 \\ q \neq a-1, a}}^Q \prod_{n=k_q+1}^{k_{q+1}} p_{\eta_q}(\mathbf{x}_n) \right) \cdot \prod_{n=k_{a-1}+1}^{k_a} p_{\eta_{a-1}}(\mathbf{x}_n) \prod_{n=k_a+1}^{k_{a+1}} p_{\eta_a}(\mathbf{x}_n) \quad (\text{A.5b})$$

on peut de la même façon écrire celle évaluée au point test $\mathbf{k} + \mathbf{h}_a$, avec $^1 (\mathbf{h}_a)_i = h_a \delta_{i,a}$, où $\delta_{i,a}$ est le symbole de Krönecker (i.e., $\delta_{i,a} = 1$ si $i = a$ et $\delta_{i,a} = 0$ si $i \neq a$). On a donc

$$p(\mathbf{X} | \mathbf{k} + \mathbf{h}_a) = \prod_{n=1}^{k_1} p_{\eta_0}(\mathbf{x}_n) \prod_{n=k_1+1}^{k_2} p_{\eta_1}(\mathbf{x}_n) \cdots \prod_{n=k_{a-1}+1}^{k_a+h_a} p_{\eta_{a-1}}(\mathbf{x}_n) \prod_{n=k_a+h_a+1}^{k_{a+1}} p_{\eta_a}(\mathbf{x}_n) \cdots \prod_{n=k_Q+1}^N p_{\eta_Q}(\mathbf{x}_n) \quad (\text{A.6a})$$

$$= \prod_{q=0}^Q \prod_{n=k_q+\delta_{q,a-1}h_a}^{k_{q+1}+\delta_{q,a-1}h_a} p_{\eta_q}(\mathbf{x}_n) \quad (\text{A.6b})$$

$$= \left(\prod_{\substack{q=0 \\ q \neq a-1, a}}^Q \prod_{n=k_q+1}^{k_{q+1}} p_{\eta_q}(\mathbf{x}_n) \right) \cdot \prod_{n=k_{a-1}+1}^{k_a+h_a} p_{\eta_{a-1}}(\mathbf{x}_n) \prod_{n=k_a+h_a+1}^{k_{a+1}} p_{\eta_a}(\mathbf{x}_n) \quad (\text{A.6c})$$

que l'on peut réécrire, en explorant ses différentes expressions suivant le signe de h_a ,

$$p(\mathbf{X} | \mathbf{k} + \mathbf{h}_a) = \left(\prod_{\substack{q=0 \\ q \neq a-1, a}}^Q \prod_{n=k_q+1}^{k_{q+1}} p_{\eta_q}(\mathbf{x}_n) \right) \times \begin{cases} \prod_{n=k_{a-1}+1}^{k_a} p_{\eta_{a-1}}(\mathbf{x}_n) \prod_{n=k_a+1}^{k_a+h_a} p_{\eta_{a-1}}(\mathbf{x}_n) \prod_{n=k_a+h_a+1}^{k_{a+1}} p_{\eta_a}(\mathbf{x}_n), & \text{si } h_a > 0 \\ \prod_{n=k_{a-1}+1}^{k_a+h_a} p_{\eta_{a-1}}(\mathbf{x}_n) \prod_{n=k_a+h_a+1}^{k_a} p_{\eta_a}(\mathbf{x}_n) \prod_{n=k_a+1}^{k_{a+1}} p_{\eta_a}(\mathbf{x}_n), & \text{si } h_a < 0 \end{cases}$$

$$= \left(\prod_{q=0}^Q \prod_{n=k_q+\delta_{q,a} \max\{h_a, 0\}+1}^{k_{q+1}-\delta_{q,a-1} \max\{-h_a, 0\}} p_{\eta_q}(\mathbf{x}_n) \right) \cdot \prod_{n=k_a-\max\{-h_a, 0\}+1}^{k_a+\max\{h_a, 0\}} p_{\epsilon_{h_a}(\eta_{a-1}, \eta_a)}(\mathbf{x}_n) \quad (\text{A.6d})$$

où l'on définit, pour $\mathbf{y}, \mathbf{z} \in \mathbb{R}^L$, et $h \in \mathbb{R}^*$,

$$\epsilon_h(\mathbf{y}, \mathbf{z}) \triangleq \mathbf{y} \mathbb{1}_{]0, +\infty[}(h) + \mathbf{z} \mathbb{1}_{]-\infty, 0[}(h) = \begin{cases} \mathbf{y}, & \text{si } h > 0, \\ \mathbf{z}, & \text{si } h < 0. \end{cases} \quad (\text{A.7})$$

La dernière écriture (A.6d) de $p(\mathbf{X} | \mathbf{k} + \mathbf{h}_a)$, bien qu'apparemment plus compliquée, nous permet dans la suite d'éviter une analyse de cas fastidieuse supplémentaire sur le signe de h_a . On peut d'ailleurs réécrire $p(\mathbf{X} | \mathbf{k})$ de cette même manière à partir de (A.5b), comme

$$p(\mathbf{X} | \mathbf{k}) = \left(\prod_{q=0}^Q \prod_{n=k_q+\delta_{q,a} \max\{h_a, 0\}+1}^{k_{q+1}-\delta_{q,a-1} \max\{-h_a, 0\}} p_{\eta_q}(\mathbf{x}_n) \right) \cdot \prod_{n=k_a-\max\{-h_a, 0\}+1}^{k_a+\max\{h_a, 0\}} p_{\epsilon_{h_a}(\eta_a, \eta_{a-1})}(\mathbf{x}_n), \quad (\text{A.8})$$

1. Rappelons-nous que nous avons supposé la matrice $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_Q]$ diagonale.

en prenant garde aux positions de $\boldsymbol{\eta}_{a-1}$ et $\boldsymbol{\eta}_a$ dans la fonction $\epsilon_{h_a}(\cdot, \cdot)$, inversées par rapport à celles dans (A.6d).

Par conséquent, on obtient les expressions de $M(\alpha, \beta, \mathbf{h}_a, \mathbf{h}_b, \mathbf{k})$ en injectant (A.8) et (A.6d) (deux fois, en remplaçant une fois a par b) dans (3.8a). Ces expressions dépendent bien sûr des cas **diag**, **psdiag** et **ts** évoqués ci-dessus.

A.2.1 Cas **diag** : $b = a$

Il faut tout d'abord remarquer que, lorsque $b = a$, on peut seulement avoir i) $h_b = h_a$ ou ii) $h_b = -h_a$ ou iii) $h_b = 0$. On donne la méthodologie utilisée pour le calcul de $M(\alpha, \beta, h_a, h_b, \mathbf{k})$ dans le premier cas i), elle s'adapte ensuite directement pour tous les autres cas.

On se place donc dans le cas i) $h_b = h_a$: en injectant (A.8) et (A.6d) dans (3.8a), on obtient

$$\begin{aligned}
 M(\alpha, \alpha, h_a, h_a, \mathbf{k}) &= \int_{\Omega} p^{2\alpha}(\mathbf{X} | \mathbf{k} + \mathbf{h}_a) p^{1-2\alpha}(\mathbf{X} | \mathbf{k}) d\mathbf{X} \\
 &= \int_{\Omega} \frac{p^{2\alpha}(\mathbf{X} | \mathbf{k} + \mathbf{h}_a)}{p^{2\alpha-1}(\mathbf{X} | \mathbf{k})} d\mathbf{X} \\
 &= \int_{\Omega} \frac{\left(\prod_{q=0}^Q \prod_{n=k_q+\delta_{q,a} \max\{h_a, 0\}+1}^{k_{q+1}-\delta_{q,a-1} \max\{-h_a, 0\}} p_{\boldsymbol{\eta}_q}^{2\alpha}(\mathbf{x}_n) \right) \cdot \prod_{n=k_a-\max\{-h_a, 0\}+1}^{k_a+\max\{h_a, 0\}} p_{\boldsymbol{\epsilon}_{h_a}(\boldsymbol{\eta}_{a-1}, \boldsymbol{\eta}_a)}^{2\alpha}(\mathbf{x}_n)}{\left(\prod_{q=0}^Q \prod_{n=k_q+\delta_{q,a} \max\{h_a, 0\}+1}^{k_{q+1}-\delta_{q,a-1} \max\{-h_a, 0\}} p_{\boldsymbol{\eta}_q}^{2\alpha-1}(\mathbf{x}_n) \right) \cdot \prod_{n=k_a-\max\{-h_a, 0\}+1}^{k_a+\max\{h_a, 0\}} p_{\boldsymbol{\epsilon}_{h_a}(\boldsymbol{\eta}_a, \boldsymbol{\eta}_{a-1})}^{2\alpha-1}(\mathbf{x}_n)} d\mathbf{X} \\
 &= \int_{\Omega} \left(\prod_{q=0}^Q \prod_{n=k_q+\delta_{q,a} \max\{h_a, 0\}+1}^{k_{q+1}-\delta_{q,a-1} \max\{-h_a, 0\}} p_{\boldsymbol{\eta}_q}(\mathbf{x}_n) \right) \cdot \prod_{n=k_a-\max\{-h_a, 0\}+1}^{k_a+\max\{h_a, 0\}} p_{\boldsymbol{\eta}_{a-1}}^{\boldsymbol{\epsilon}_{h_a}(2\alpha)}(\mathbf{x}_n) p_{\boldsymbol{\eta}_a}^{\boldsymbol{\epsilon}_{h_a}(1-2\alpha)}(\mathbf{x}_n) d\mathbf{X}
 \end{aligned} \tag{A.9}$$

où l'on définit, pour $s \in \mathbb{R}$ et $h \in \mathbb{R}^*$,

$$\epsilon_h(s) \triangleq \epsilon_h(s, 1-s) = \begin{cases} s, & \text{si } h > 0, \\ 1-s, & \text{si } h < 0 \end{cases} \tag{A.10}$$

correspondant à la définition donnée en (3.13). Dans (A.9), on a une intégrale N -uple d'un produit de fonctions dont les variables $\mathbf{x}_1, \dots, \mathbf{x}_N$ sont complètement séparées. L'intégrale du produit est donc égale au produit des intégrales, et on obtient donc

$$\begin{aligned}
 M(\alpha, \alpha, h_a, h_a, \mathbf{k}) &= \prod_{q=0}^Q \prod_{n=k_q+\delta_{q,a} \max\{h_a, 0\}+1}^{k_{q+1}-\delta_{q,a-1} \max\{-h_a, 0\}} \left(\int_{\Omega'} p_{\boldsymbol{\eta}_q}(\mathbf{x}_n) d\mathbf{x}_n \right) \\
 &\quad \times \prod_{n=k_a-\max\{-h_a, 0\}+1}^{k_a+\max\{h_a, 0\}} \left(\int_{\Omega'} p_{\boldsymbol{\eta}_{a-1}}^{\boldsymbol{\epsilon}_{h_a}(2\alpha)}(\mathbf{x}_n) p_{\boldsymbol{\eta}_a}^{\boldsymbol{\epsilon}_{h_a}(1-2\alpha)}(\mathbf{x}_n) d\mathbf{x}_n \right) \tag{A.11}
 \end{aligned}$$

En remarquant que les intégrales par rapport aux variables $\mathbf{x}_1, \dots, \mathbf{x}_{k_a-\max\{-h_a, 0\}}, \mathbf{x}_{k_a+\max\{h_a, 0\}+1}, \dots, \mathbf{x}_N$ valent toutes 1, on obtient finalement

$$M(\alpha, \alpha, h_a, h_a, \mathbf{k}) = \prod_{n=k_a-\max\{-h_a, 0\}+1}^{k_a+\max\{h_a, 0\}} \left(\int_{\Omega'} p_{\boldsymbol{\eta}_{a-1}}^{\boldsymbol{\epsilon}_{h_a}(2\alpha)}(\mathbf{x}_n) p_{\boldsymbol{\eta}_a}^{\boldsymbol{\epsilon}_{h_a}(1-2\alpha)}(\mathbf{x}_n) d\mathbf{x}_n \right)$$

$$\begin{aligned}
&= \left(\int_{\Omega'} p_{\eta_{a-1}}^{\varepsilon_{h_a}(2\alpha)}(\mathbf{x}) p_{\eta_a}^{\varepsilon_{h_a}(1-2\alpha)}(\mathbf{x}) d\mathbf{x} \right)^{\max\{h_a, 0\} + \max\{-h_a, 0\}} \\
&= \left(\int_{\Omega'} p_{\eta_{a-1}}^{\varepsilon_{h_a}(2\alpha)}(\mathbf{x}) p_{\eta_a}^{\varepsilon_{h_a}(1-2\alpha)}(\mathbf{x}) d\mathbf{x} \right)^{|h_a|} \\
&= \rho_{a-1, a}^{|h_a|}(\varepsilon_{h_a}(2\alpha))
\end{aligned} \tag{A.12}$$

où, pour $q \in \llbracket 1, Q \rrbracket$ et $s \in]0, 1[$, $\rho_{q-1, q}(s)$ est défini en (3.12).

En utilisant exactement la même méthodologie pour les cas ii) $h_b = -h_a$ et iii) $h_b = 0$, on obtient respectivement, de façon analogue,

$$\begin{aligned}
M(\alpha, 1 - \alpha, h_a, -h_a, \mathbf{k}) &= \rho_{a-1, a}(\varepsilon_{h_a}(\alpha)) \cdot \rho_{a-1, a}(\varepsilon_{-h_a}(1 - \alpha)) \\
&= \rho_{a-1, a}^{2|h_a|}(\varepsilon_{h_a}(\alpha))
\end{aligned} \tag{A.13}$$

et

$$M(\alpha, 0, h_a, \mathbf{0}, \mathbf{k}) = \rho_{a-1, a}^{|h_a|}(\varepsilon_{h_a}(\alpha)). \tag{A.14}$$

Les termes calculés en (A.12) et (A.13) interviennent au numérateur de A'_q dans (3.10), tandis que celui obtenu en (A.14) y intervient au dénominateur.

Les cas suivants sont eux aussi une application la méthodologie que l'on vient de présenter.

A.2.2 Cas ts : $b > a + 1$

En appliquant la même méthodologie que celle présentée dans la section précédente, on obtient simplement

$$M(\alpha, \beta, h_a, h_b, \mathbf{k}) = \rho_{a-1, a}^{|h_a|}(\varepsilon_{h_a}(\alpha)) \cdot \rho_{b-1, b}^{|h_b|}(\varepsilon_{h_b}(\beta)). \tag{A.15}$$

A.2.3 Cas psdiag : $b = a + 1$

Ce cas requiert une attention particulière car il est plus complexe que les deux précédents. On peut en effet rencontrer ce qu'on appelle des « chevauchements² » entre des segments associés à différents points test. Cela correspond concrètement aux situations où, pour $a < b$, $k_a + h_a > k_b + h_b$.

Une analyse attentive des tableaux A.2 et A.3 montre que cette situation ne peut en fait se produire que si (i) $b = a + 1$, (ii) $h_a > 0$ et $h_{a+1} < 0$ et (iii) $\min\{|h_a|, |h_{a+1}|\} \geq d + 1$: en effet, en supposant $b = a + 2$, en considérant les lignes “ $a + 1$ ” et “ b ” du tableau A.3, on a en particulier, respectivement,

$$k_a + d + \max\{h_a, 0\} \leq k_{a+1} \tag{A.16}$$

d'une part, et

$$k_{a+1} + d + \max\{-h_{a+2}, 0\} \leq k_{a+2} \tag{A.17}$$

d'autre part. Ceci implique que

$$k_a + h_a < k_{a+1} < k_{a+2} + h_{a+2} \tag{A.18}$$

car, $\forall x \in \mathbb{R}$, $\max\{x, 0\} \geq x$, et $d > 0$. Ceci prouve (i). Par ailleurs, si on suppose $b = a + 1$, en considérant la ligne “ $a + 1$ ” du tableau A.2, on obtient en particulier

$$k_a + d + \max\{h_a, -h_{a+1}, 0\} \leq k_{a+1}. \tag{A.19}$$

2. On parle en anglais d'*overlap* [LRNM10].

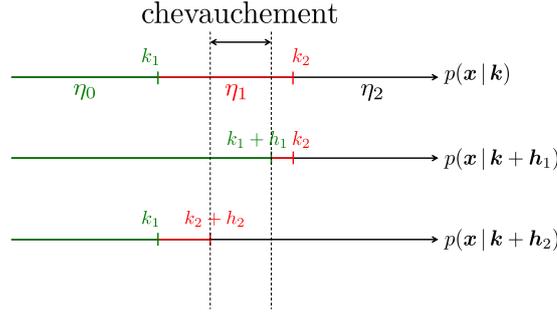


FIGURE A.1 – Illustration d’une situation de chevauchement entre 3 segments consécutifs.

Si on suppose $h_a < 0$, on obtient

$$k_{a+1} + h_{a+1} \geq k_a + d + \max\{-h_{a+1}, 0\} + h_{a+1} > k_a > k_a + h_a, \quad (\text{A.20})$$

et si on suppose $h_{a+1} > 0$, obtient

$$k_{a+1} + h_{a+1} \geq k_a + d + \max\{h_a, 0\} + h_{a+1} > k_a + h_a, \quad (\text{A.21})$$

ce qui prouve (ii). Enfin, supposons que $h_a > 0$ et $h_{a+1} < 0$. La condition d’existence du chevauchement est $k_a + h_a > k_{a+1} + h_{a+1}$, soit $k_{a+1} \leq k_a + h_a - h_{a+1} - 1$. Par conséquent, comme la plage de valeur de k_{a+1} autorisées par la loi *a priori* choisie dépend de k_a (d’après le tableau A.2), la condition d’existence d’au moins un terme de chevauchement s’écrit

$$(k_{a+1})_{\min} \leq k_a + h_a - h_{a+1} - 1, \quad (\text{A.22})$$

soit, d’après le tableau A.2,

$$k_a + d + \max\{h_a, -h_{a+1}\} \leq k_a + h_a - h_{a+1} - 1, \quad (\text{A.23})$$

qui est exactement équivalent à la condition (iii).

On suppose donc maintenant que (i) $b = a + 1$. Si l’une des deux autres conditions (ii) ou (iii) n’est pas remplie, on vient de montrer qu’il n’y a pas de chevauchement. Dans ce cas, $M(\alpha, \beta, h_a, h_{a+1}, \mathbf{k})$ s’écrit de la même manière que dans le cas ts précédemment traité, c’est-à-dire que l’on a

$$M(\alpha, \beta, h_a, h_{a+1}, \mathbf{k}) = \rho_{a-1,a}^{|h_a|}(\varepsilon_{h_a}(\alpha)) \cdot \rho_{a,a+1}^{|h_{a+1}|}(\varepsilon_{h_{a+1}}(\beta)). \quad (\text{A.24})$$

Supposons enfin que les trois conditions (i) $b = a + 1$, (ii) $h_a > 0$ et $h_{a+1} < 0$ et (iii) $\min\{|h_a|, |h_{a+1}|\} \geq d + 1$ sont vérifiées. Comme illustré sur la figure A.1, cette situation se traduit par un « chevauchement » du $(a - 1)$ -ième segment dans $p(\mathbf{x} | \mathbf{k} + \mathbf{h}_a)$ avec le $(a + 1)$ -ième segment dans $p(\mathbf{x} | \mathbf{k} + \mathbf{h}_{a+1})$, ce qui va impliquer des modifications dans l’écriture de $M(\alpha, \beta, h_a, h_{a+1}, \mathbf{k})$.

Pour le voir, on reprend la méthodologie utilisée dans les cas précédents, en remarquant cette fois que, comme $k_a < k_{a+1} + h_{a+1} < k_a + h_a < k_{a+1}$, on peut exprimer les trois fonctions de

vraisemblance qui composent $M(\alpha, \beta, h_a, h_{a+1}, \mathbf{k})$ comme

$$p(\mathbf{X} | \mathbf{k}) = \left(\prod_{\substack{q=0 \\ q \neq a}}^Q \prod_{n=k_q+1}^{k_{q+1}} p_{\eta_q}(\mathbf{x}_n) \right) \cdot \prod_{n=k_{a+1}}^{k_{a+1}+h_{a+1}} p_{\eta_a}(\mathbf{x}_n) \cdot \prod_{n=k_{a+1}+h_{a+1}+1}^{k_a+h_a} p_{\eta_a}(\mathbf{x}_n) \cdot \prod_{n=k_a+h_a+1}^{k_{a+1}} p_{\eta_a}(\mathbf{x}_n) \quad (\text{A.25a})$$

$$p(\mathbf{X} | \mathbf{k} + \mathbf{h}_a) = \left(\prod_{\substack{q=0 \\ q \neq a}}^Q \prod_{n=k_q+1}^{k_{q+1}} p_{\eta_q}(\mathbf{x}_n) \right) \cdot \prod_{n=k_{a+1}}^{k_{a+1}+h_{a+1}} p_{\eta_{a-1}}(\mathbf{x}_n) \cdot \prod_{n=k_{a+1}+h_{a+1}+1}^{k_a+h_a} p_{\eta_{a-1}}(\mathbf{x}_n) \cdot \prod_{n=k_a+h_a+1}^{k_{a+1}} p_{\eta_a}(\mathbf{x}_n) \quad (\text{A.25b})$$

$$p(\mathbf{X} | \mathbf{k} + \mathbf{h}_{a+1}) = \left(\prod_{\substack{q=0 \\ q \neq a}}^Q \prod_{n=k_q+1}^{k_{q+1}} p_{\eta_q}(\mathbf{x}_n) \right) \cdot \prod_{n=k_{a+1}}^{k_{a+1}+h_{a+1}} p_{\eta_a}(\mathbf{x}_n) \cdot \prod_{n=k_{a+1}+h_{a+1}+1}^{k_a+h_a} p_{\eta_{a+1}}(\mathbf{x}_n) \cdot \prod_{n=k_a+h_a+1}^{k_{a+1}} p_{\eta_{a+1}}(\mathbf{x}_n). \quad (\text{A.25c})$$

Alors, en injectant ces expressions dans (3.8a), on trouve

$$\begin{aligned} M(\alpha, \beta, h_a, h_{a+1}, \mathbf{k}) &= \int_{\Omega} \frac{p^{\alpha}(\mathbf{X} | \mathbf{k} + \mathbf{h}_a) p^{\beta}(\mathbf{X} | \mathbf{k} + \mathbf{h}_{a+1})}{p^{\alpha+\beta-1}(\mathbf{X} | \mathbf{k})} d\mathbf{X} \\ &= \int_{\Omega} \left(\prod_{\substack{q=0 \\ q \neq a}}^Q \prod_{n=k_q+1}^{k_{q+1}} p_{\eta_q}(\mathbf{x}_n) \right) \times \prod_{n=k_{a+1}}^{k_{a+1}+h_{a+1}} p_{\eta_{a-1}}^{\alpha}(\mathbf{x}_n) p_{\eta_a}^{1-\alpha}(\mathbf{x}_n) \\ &\quad \times \prod_{n=k_{a+1}+h_{a+1}+1}^{k_a+h_a} \frac{p_{\eta_{a-1}}^{\alpha}(\mathbf{x}_n) p_{\eta_{a+1}}^{\beta}(\mathbf{x}_n)}{p_{\eta_a}^{\alpha+\beta-1}(\mathbf{x}_n)} \\ &\quad \times \prod_{n=k_a+h_a+1}^{k_{a+1}} p_{\eta_{a+1}}^{\beta}(\mathbf{x}_n) p_{\eta_a}^{1-\beta}(\mathbf{x}_n) d\mathbf{X}. \quad (\text{A.26}) \end{aligned}$$

De même que précédemment, en constatant que nous avons une intégrale N -uple d'un produit de fonctions dont les variables $\mathbf{x}_1, \dots, \mathbf{x}_N$ sont complètement séparées, on obtient, en éliminant les termes égaux à 1,

$$M(\alpha, \beta, h_a, h_{a+1}, \mathbf{k}) = \frac{\rho_{a-1,a}^{k_{a+1}+h_{a+1}-k_a}(\alpha) \rho_{a,a+1}^{k_{a+1}-(k_a+h_a)}(1-\beta)}{r_a^{(k_{a+1}+h_{a+1})-(k_a+h_a)}(\alpha, 1-\alpha-\beta)} \quad (\text{A.27a})$$

$$= \rho_{a-1,a}^{|h_a|}(\alpha) \rho_{a,a+1}^{|h_{a+1}|}(1-\beta) R_a^{(k_{a+1}+h_{a+1})-(k_a+h_a)}(\alpha, 1-\beta), \quad (\text{A.27b})$$

où, pour $q \in \llbracket 1, Q \rrbracket$ et $s, s' \in]0, 1[$, $R_q(s, s')$ et $r_q(s, s')$ sont définis en (3.16) et (3.17) respectivement.

On dispose donc maintenant des expressions de $M(\alpha, \beta, h_a, h_b, \mathbf{k})$ dans tous les cas, donnés par les équation (A.12), (A.13), (A.14) pour le cas **diag**, (A.15) pour le cas **ts**, (A.24) pour le cas **psdiag** sans chevauchement, et enfin (A.27b) pour le cas (A.27b) pour le cas **psdiag** avec chevauchement. Remarquons que les expressions de $M(\alpha, \beta, h_a, h_b, \mathbf{k})$ ne dépendent jamais de \mathbf{k} sauf dans le dernier cas (**psdiag** avec chevauchement). Dans la suite, pour que ces dépendances (ou indépendances) en \mathbf{k} soient plus visibles, on notera $M(\alpha, \beta, h_a, h_b, \mathbf{k})$ comme $\tilde{M}(\alpha, \beta, h_a, h_b, \mathbf{k})$ dans le cas **psdiag** avec chevauchement (où $M(\cdot)$ dépend bien de \mathbf{k}) et comme $M(\alpha, \beta, h_a, h_b)$ dans tous les autres cas (où $M(\cdot)$ ne dépend pas de \mathbf{k}). Nous sommes donc en mesure de calculer les expressions de $e^{\tilde{\mu}(\alpha, \beta, h_a, h_b)}$.

A.3 Expressions de $e^{\tilde{\mu}(\alpha, \beta, h_a, h_b)}$ et de $[G]_{a,b}$

Les expressions de $e^{\tilde{\mu}(\alpha, \beta, h_a, h_b)}$ s'obtiennent à partir de (3.6), où le domaine de la somme discrète Q -uple est en réalité $(\mathcal{K}'_d)_{h_a, h_b}$, avec les expressions de $\nu(\cdot)$ et $M(\cdot)$ obtenues précédemment. Les

résultats obtenus sont différents suivant les cas diag, ts et psdiag.

A.3.1 Cas diag et termes diagonaux de \mathbf{G}

On suppose que $b = a$. En injectant (A.3) et (A.12) dans (3.6), on obtient

$$e^{\tilde{\mu}(\alpha, \alpha, h_a, h_a)} = \rho_{a-1, a}^{|h_a|} (\varepsilon_{h_a}(2\alpha)) \cdot \frac{1}{\Delta^Q} \sum_{\mathbf{k} \in (\mathcal{K}'_d)_{h_a, h_a}} \dots \sum 1. \quad (\text{A.28})$$

En notant, d'après le tableau A.1, que

$$\begin{aligned} & \frac{1}{\Delta^Q} \sum_{\mathbf{k} \in (\mathcal{K}'_d)_{h_a, h_a}} \dots \sum 1 \\ &= \begin{cases} \frac{1}{\Delta^Q} \sum_{k_1=d}^D \sum_{k_2=k_1+d}^{k_1+D} \dots \sum_{k_a=k_{a-1}+\max\{-h_a, 0\}+1}^{k_{a-1}+D-\max\{h_a, 0\}} \sum_{k_{a+1}=k_a+1+\max\{h_a, 0\}}^{k_a+D-\max\{-h_a, 0\}} \dots \sum_{k_Q=k_{Q-1}+d}^{k_{Q-1}+D} 1 & \text{si } a < Q, \\ \frac{1}{\Delta^Q} \sum_{k_1=d}^D \sum_{k_2=k_1+d}^{k_1+D} \dots \sum_{k_Q=k_{Q-1}+1+\max\{-h_a, 0\}}^{k_{Q-1}+D-\max\{h_a, 0\}} 1 & \text{si } a = Q \end{cases} \\ &= u(\Delta, h_a), \end{aligned} \quad (\text{A.29})$$

on obtient finalement

$$e^{\tilde{\mu}(\alpha, \alpha, h_a, h_a)} = u(\Delta, h_a) \cdot \rho_{a-1, a}^{|h_a|} (\varepsilon_{h_a}(2\alpha)) \quad (\text{A.30})$$

De la même manière, à partir de (A.3) et (A.13), on aboutit à

$$e^{\tilde{\mu}(\alpha, 1-\alpha, h_a, -h_a)} = \rho_{a-1, a}^{2|h_a|} (\varepsilon_{h_a}(\alpha)) \cdot u(\Delta, 2h_a) \quad (\text{A.31})$$

et à partir de (A.3) et (A.14),

$$e^{\mu(\alpha, h_a)} = e^{\tilde{\mu}(\alpha, 0, h_a, \mathbf{0})} = \rho_{a-1, a}^{|h_a|} (\varepsilon_{h_a}(\alpha)) \cdot u(\Delta, h_a). \quad (\text{A.32})$$

En injectant (A.30) et (A.31) au numérateur de (2.80) d'une part, et (A.32) au dénominateur de (2.80) d'autre part, on obtient bien (3.10).

A.3.2 Cas ts et termes du triangle supérieur de \mathbf{G}

On suppose que $b > a + 1$. De la même manière que précédemment, comme $M(\cdot)$ ne dépend pas de \mathbf{k} , on obtient

$$e^{\tilde{\mu}(\alpha, \beta, h_a, h_b)} = \rho_{a-1, a}^{|h_a|} (\varepsilon_{h_a}(\alpha)) \cdot \rho_{b-1, b}^{|h_b|} (\varepsilon_{h_b}(\beta)) \cdot u(\Delta, h_a) \cdot u(\Delta, h_b). \quad (\text{A.33})$$

En exploitant les propriétés de symétrie des fonctions mises en jeu, on remarque que

$$e^{\tilde{\mu}(\alpha, \beta, h_a, h_b)} = e^{\tilde{\mu}(1-\alpha, \beta, -h_a, h_b)} = e^{\tilde{\mu}(\alpha, 1-\beta, h_a, -h_b)}, \quad (\text{A.34})$$

ce qui implique, en injectant (A.33) dans (2.80), que

$$[\mathbf{G}]_{a, b} = 0 \quad (\text{A.35})$$

pour $|b - a| > 1$. Cela prouve la structure tridiagonale de la matrice \mathbf{G} .

A.3.3 Cas psdiag et termes des premières surdiagonale et sous-diagonale de G

Supposons que $b = a + 1$. On calcule dans un premier temps les expressions de $e^{\tilde{\mu}(\alpha,\beta,h_a,h_{a+1})}$ (section A.3.3.1), d'abord dans le cas sans chevauchement entre segments successifs (sous-section A.3.3.1.1), puis dans le cas avec chevauchement (sous-section A.3.3.1.2). Nous en déduisons ensuite l'expression des termes des premières surdiagonale et sous-diagonale de G (section A.3.3.2).

A.3.3.1 Expressions de $e^{\tilde{\mu}(\alpha,\beta,h_a,h_{a+1})}$

A.3.3.1.1 Expression de $e^{\tilde{\mu}(\alpha,\beta,h_a,h_{a+1})}$ dans le cas sans chevauchements Rappelons que ce cas est rencontré si

- i) $h_a > 0$ et $h_{a+1} > 0$,
ou
- ii) $h_a < 0$ et $h_{a+1} < 0$,
ou
- iii) $h_a < 0$ et $h_{a+1} > 0$,
ou
- iv) $h_a > 0$ et $h_{a+1} < 0$ et $\min\{|h_a|, |h_{a+1}|\} \leq d$.

Dans ce cas, on a vu que $M(\cdot)$ ne dépend pas de \mathbf{k} , donc on procède de la même manière que dans les cas précédents. D'après le tableau A.2, on obtient que

$$e^{\tilde{\mu}(\alpha,\beta,h_a,h_{a+1})} = \rho_{a-1,a}^{|h_a|}(\varepsilon_{h_a}(\alpha)) \cdot \rho_{a,a+1}^{|h_{a+1}|}(\varepsilon_{h_{a+1}}(\beta)) \cdot u_{\text{psd}}(\Delta, h_a, h_{a+1}) \quad (\text{A.36})$$

où l'on définit, pour $q \in \llbracket 1, Q-1 \rrbracket$,

$$u_{\text{psd}}(\Delta, h_q, h_{q+1}) \triangleq \begin{cases} \frac{(\Delta - |h_q|)(\Delta - |h_{q+1}|) \max\{\Delta'_{h_q, h_{q+1}}, 0\}}{\Delta^3} & \text{si } q+1 < Q \\ \frac{(\Delta - |h_{Q-1}|) \max\{\Delta'_{h_{Q-1}, h_Q}, 0\}}{\Delta^2} & \text{si } q+1 = Q \end{cases} \quad (\text{A.37})$$

avec

$$\Delta'_{h_q, h_{q+1}} \triangleq \Delta - \max\{-h_q, h_{q+1}, 0\} - \max\{h_q, -h_{q+1}, 0\}. \quad (\text{A.38})$$

A.3.3.1.2 Expression de $e^{\tilde{\mu}(\alpha,\beta,h_a,h_{a+1})}$ dans le cas avec chevauchements Supposons maintenant qu'il y ait des chevauchements entre segments successifs, c'est-à-dire que

$$h_a > 0 \quad \text{et} \quad h_{a+1} < 0 \quad \text{et} \quad \min\{|h_a|, |h_{a+1}|\} \geq d + 1.$$

Nous avons vu que dans ce cas, $M(\cdot)$ dépend explicitement de \mathbf{k} si $k_a + h_a > k_{a+1} + h_{a+1}$. Le calcul de $e^{\tilde{\mu}(\alpha,\beta,h_a,h_{a+1})}$ n'est donc pas aussi simple que dans les cas précédents : dans la somme Q -uple sur \mathbf{k} , la somme par rapport à k_{a+1} doit être scindée en deux suivant que $k_a + h_a \gtrless k_{a+1} + h_{a+1}$. Plus précisément :

- pour $k_{a+1} \in \llbracket k_a + \max\{|h_a|, |h_{a+1}|\} + d, k_a + |h_a| + |h_{a+1}| - 1 \rrbracket$, il y a chevauchement, donc $M(\cdot)$ s'écrit selon (A.27b) ;
- pour $k_{a+1} \in \llbracket k_a + |h_a| + |h_{a+1}|, k_a + D \rrbracket$, il n'y a pas de chevauchement, donc $M(\cdot)$ s'écrit selon (A.24).

Par conséquent, on a, pour $a + 1 < Q$,

$$e^{\tilde{\mu}(\alpha, \beta, h_a, h_{a+1})} = \frac{1}{\Delta^Q} \sum_{\substack{k_q = k_{q-1} + d \\ q \in \llbracket 1, a-1 \rrbracket}} \dots \sum_{k_{a-1} = k_{a-1} + d}^{k_{a-1} + D} \sum_{k_a = k_{a-1} + d}^{k_{a-1} + D - h_a} [S_1(\alpha, \beta, h_a, h_{a+1}, \mathbf{k}_{1:a}) + S_2(\alpha, \beta, h_a, h_{a+1}, \mathbf{k}_{1:a})] \quad (\text{A.39})$$

où $\mathbf{k}_{1:a}$ représente le vecteur \mathbf{k} tronqué $[k_1, \dots, k_a]^\top$, et

$$S_1(\alpha, \beta, h_a, h_{a+1}, \mathbf{k}_{1:a}) \triangleq \sum_{k_{a+1} = k_a + \max\{|h_a|, |h_{a+1}|\} + d}^{k_a + |h_a| + |h_{a+1}| - 1} \sum_{\substack{k_q = (k_q)_{\min} \\ q \in \llbracket a+2, Q \rrbracket}}^{(k_q)_{\max}} \tilde{M}(\alpha, \beta, h_a, h_{a+1}, \mathbf{k}) \quad (\text{A.40})$$

contient les termes de chevauchement, et

$$S_2(\alpha, \beta, h_a, h_{a+1}, \mathbf{k}_{1:a}) \triangleq \sum_{k_{a+1} = k_a + |h_a| + |h_{a+1}|}^{k_a + D} \sum_{\substack{k_q = (k_q)_{\min} \\ q \in \llbracket a+2, Q \rrbracket}}^{(k_q)_{\max}} M(\alpha, \beta, h_a, h_{a+1}) \quad (\text{A.41})$$

contient les termes sans chevauchement. Dans (A.24), nous avons vu que $M(\alpha, \beta, h_a, h_{a+1})$ ne dépend pas de k_{a+1}, \dots, k_Q . Ainsi $S_2(\cdot)$ s'obtient, avec l'aide du tableau A.2, comme

$$\begin{aligned} S_2(\alpha, \beta, h_a, h_{a+1}, \mathbf{k}_{1:a}) &= M(\alpha, \beta, h_a, h_{a+1}) \sum_{k_{a+1} = k_a + |h_a| + |h_{a+1}|}^{k_a + D} \sum_{\substack{k_q = (k_q)_{\min} \\ q \in \llbracket a+2, Q \rrbracket}}^{(k_q)_{\max}} 1 \\ &= \rho_{a-1, a}^{|h_a|}(\varepsilon_{h_a}(\alpha)) \cdot \rho_{a, a+1}^{|h_{a+1}|}(\varepsilon_{h_{a+1}}(\beta)) \\ &\quad \times \max\{D - |h_a| - |h_{a+1}| + 1, 0\} \cdot (\Delta - |h_{a+1}|) \cdot \Delta^{Q-a-2}, \end{aligned} \quad (\text{A.42})$$

qui ne dépend en fait pas de $\mathbf{k}_{1:a}$. D'autre part, d'après (A.27b), on développe $S_1(\cdot)$ comme

$$\begin{aligned} S_1(\alpha, \beta, h_a, h_{a+1}, \mathbf{k}_{1:a}) &= \rho_{a-1, a}^{|h_a|}(\alpha) \cdot \rho_{a, a+1}^{|h_{a+1}|}(1 - \beta) \cdot R_a^{h_{a+1} - (k_a + h_a)}(\alpha, 1 - \beta) \\ &\quad \times \sum_{k_{a+1} = k_a + \max\{|h_a|, |h_{a+1}|\} + d}^{k_a + |h_a| + |h_{a+1}| - 1} R_a^{k_{a+1}}(\alpha, 1 - \beta) \sum_{\substack{k_q = (k_q)_{\min} \\ q \in \llbracket a+2, Q \rrbracket}}^{(k_q)_{\max}} 1 \end{aligned} \quad (\text{A.43})$$

où l'on a

$$\sum_{\substack{k_q = (k_q)_{\min} \\ q \in \llbracket a+2, Q \rrbracket}}^{(k_q)_{\max}} 1 = (\Delta - |h_{a+1}|) \cdot \Delta^{Q-a-2} \quad (\text{A.44})$$

qui ne dépend pas de k_{a+1} . Il vient donc, dans (A.43),

$$\begin{aligned} S_1(\alpha, \beta, h_a, h_{a+1}, \mathbf{k}_{1:a}) &= \rho_{a-1, a}^{|h_a|}(\alpha) \cdot \rho_{a, a+1}^{|h_{a+1}|}(1 - \beta) \cdot R_a^{h_{a+1} - (k_a + h_a)}(\alpha, 1 - \beta) \cdot (\Delta - |h_{a+1}|) \cdot \Delta^{Q-a-2} \\ &\quad \times R_a^{k_a + \max\{|h_a|, |h_{a+1}|\} + d}(\alpha, 1 - \beta) \frac{1 - R_a^{\min\{|h_a|, |h_{a+1}|\} - d}(\alpha, 1 - \beta)}{1 - R_a(\alpha, 1 - \beta)} \end{aligned} \quad (\text{A.45})$$

en reconnaissant une somme de termes d'une suite géométrique de raison $R_a(\alpha, 1 - \beta)$. En recombinaut et simplifiant les termes restants, on obtient finalement, pour $a + 1 < Q$,

$$\begin{aligned} S_1(\alpha, \beta, h_a, h_{a+1}, \mathbf{k}_{1:a}) &= \rho_{a-1, a}^{|h_a|}(\alpha) \cdot \rho_{a, a+1}^{|h_{a+1}|}(1 - \beta) \cdot (\Delta - |h_{a+1}|) \cdot \Delta^{Q-a-2} \\ &\quad \times \frac{R_a^{d - \min\{|h_a|, |h_{a+1}|\}}(\alpha, 1 - \beta) - 1}{1 - R_a(\alpha, 1 - \beta)}. \end{aligned} \quad (\text{A.46})$$

Pour $a + 1 = Q$, on obtient, de manière similaire,

$$S_2(\alpha, \beta, h_{Q-1}, h_Q, \mathbf{k}_{1:Q-1}) = \rho_{Q-2, Q-1}^{|h_{Q-1}|}(\varepsilon_{h_{Q-1}}(\alpha)) \cdot \rho_{Q-1, Q}^{|h_Q|}(\varepsilon_{h_Q}(\beta)) \cdot (D - |h_{Q-1}| - |h_Q| + 1) \quad (\text{A.47})$$

et

$$S_1(\alpha, \beta, h_{Q-1}, h_Q, \mathbf{k}_{1:Q-1}) = \rho_{Q-2, Q-1}^{|h_{Q-1}|}(\alpha) \cdot \rho_{Q-1, Q}^{|h_Q|}(1 - \beta) \cdot \frac{R_{Q-1}^{d - \min\{|h_{Q-1}|, |h_Q|\}}(\alpha, 1 - \beta) - 1}{1 - R_{Q-1}(\alpha, 1 - \beta)}. \quad (\text{A.48})$$

En injectant les expressions obtenues de $S_1(\cdot)$ et $S_2(\cdot)$ dans (A.39), on obtient finalement les expressions de $e^{\tilde{\mu}(\alpha,\beta,h_a,h_{a+1})}$ dans le cas avec chevauchement :

$$e^{\tilde{\mu}(\alpha,\beta,h_a,h_{a+1})} = \frac{(\Delta - |h_a|)(\Delta - |h_{a+1}|)}{\Delta^3} \cdot \rho_{a-1, a}^{|h_a|}(\alpha) \cdot \rho_{a, a+1}^{|h_{a+1}|}(1 - \beta) \\ \times \left[\max\{D - |h_a| - |h_{a+1}| + 1, 0\} - \frac{1 - R_a^{d - \min\{|h_a|, |h_{a+1}|\}}(\alpha, 1 - \beta)}{1 - R_a(\alpha, 1 - \beta)} \right], \quad (\text{A.49a})$$

si $a + 1 < Q$, et

$$e^{\tilde{\mu}(\alpha,\beta,h_{Q-1},h_Q)} = \frac{(\Delta - |h_{Q-1}|)}{\Delta^2} \cdot \rho_{Q-2, Q-1}^{|h_{Q-1}|}(\alpha) \cdot \rho_{Q-1, Q}^{|h_Q|}(1 - \beta) \\ \times \left[\max\{D - |h_{Q-1}| - |h_Q| + 1, 0\} - \frac{1 - R_{Q-1}^{d - \min\{|h_{Q-1}|, |h_Q|\}}(\alpha, 1 - \beta)}{1 - R_{Q-1}(\alpha, 1 - \beta)} \right], \quad (\text{A.49b})$$

si $a + 1 = Q$.

A.3.3.2 Expression des termes des premières surdiagonale et sous-diagonale de \mathbf{G}

Pour éviter une analyse de cas fastidieuse sur les signes respectifs de h_a et h_{a+1} , nous réécrivons (2.80) comme

$$[\mathbf{G}]_{a, a+1} = \text{sign}(h_a h_{a+1}) \\ \times \frac{\left(\begin{aligned} &e^{\tilde{\mu}(\varepsilon_{h_a}(s_a), \varepsilon_{h_{a+1}}(s_{a+1}), |h_a|, |h_{a+1}|)} + e^{\tilde{\mu}(1 - \varepsilon_{h_a}(s_a), 1 - \varepsilon_{h_{a+1}}(s_{a+1}), -|h_a|, -|h_{a+1}|)} \\ &- e^{\tilde{\mu}(\varepsilon_{h_a}(s_a), 1 - \varepsilon_{h_{a+1}}(s_{a+1}), |h_a|, -|h_{a+1}|)} - e^{\tilde{\mu}(1 - \varepsilon_{h_a}(s_a), \varepsilon_{h_{a+1}}(s_{a+1}), -|h_a|, |h_{a+1}|)} \end{aligned} \right)}{e^{\mu(s_a, h_a)} e^{\mu(s_{a+1}, h_{a+1})}}. \quad (\text{A.50})$$

Ainsi, il est clair que seul le troisième terme au numérateur de (A.50) peut correspondre au cas avec chevauchement. Lorsque $\min\{|h_a|, |h_{a+1}|\} \leq d$, ce terme s'écrit selon (A.36), tandis que lorsque $\min\{|h_a|, |h_{a+1}|\} \geq d + 1$, il s'écrit selon (A.49a) (si $a + 1 < Q$) ou (A.49b) (si $a + 1 = Q$). Les trois autres termes au numérateur de (A.50) s'écrivent selon (A.36), et les termes du dénominateur selon (A.32). En injectant ces expressions dans (A.50), on obtient bien l'expression des termes sur- et sous-diagonaux de \mathbf{G} donnés en (3.14).

Annexe B

Expressions des éléments des matrices

$V_{11}^{\eta^*}$ et $V_{12}^{\eta^*}$

Cette annexe fournit les détails techniques permettant d'aboutir aux résultats présentés au chapitre 3, section 3.2.2, à savoir l'expression des éléments des matrices qui forment la borne de Cramér-Rao/Weiss-Weinstein hybride.

Nous avons vu dans la section 3.2.1.2 que les éléments des blocs $V_{22}^{\eta^*}$ et $C_{22}^{\eta^*}$ s'expriment simplement à partir de ceux de la matrice $\mathbf{G}(\mathbf{H}, \mathbf{s})$ de la borne de Weiss-Weinstein, dont l'expression pour le problème des points de rupture a été donnée à la section 3.1.3. Les détails techniques pour l'obtention des expressions des matrices $V_{22}^{\eta^*}$ et $C_{22}^{\eta^*}$ ont donc déjà été fournis dans l'annexe A. On pourra donc s'y référer pour expliquer les résultats de la section 3.2.2.2, moyennant de légères adaptations dans les notations des densités qui interviennent (voir section 3.2.1.1).

En conséquence, nous ne développons ici que les détails aboutissant aux expressions des éléments des matrices $V_{11}^{\eta^*}$ et $V_{12}^{\eta^*}$.

B.1 Expressions des éléments de la matrice $V_{11}^{\eta^*}$

Comme $p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}) = p(\mathbf{x} | \mathbf{k}; \boldsymbol{\eta}) \pi(\mathbf{k})$, on a

$$\frac{\partial^2 \ln p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{q,\ell} \partial \eta_{q',\ell'}} = \frac{\partial^2 \ln p(\mathbf{x} | \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{q,\ell} \partial \eta_{q',\ell'}} + \frac{\partial^2 \ln \pi(\mathbf{k})}{\partial \eta_{q,\ell} \partial \eta_{q',\ell'}} = \frac{\partial^2 \ln p(\mathbf{x} | \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{q,\ell} \partial \eta_{q',\ell'}} \quad (\text{B.1})$$

du fait de l'hypothèse d'indépendance de la distribution *a priori* sur \mathbf{k} par rapport à $\boldsymbol{\eta}$. D'après (3.4), la log-vraisemblance se déduit directement :

$$\ln p(\mathbf{x} | \mathbf{k}; \boldsymbol{\eta}) = \sum_{q=0}^Q \sum_{n=k_q+1}^{k_{q+1}} \ln p(x_n; \boldsymbol{\eta}_q), \quad (\text{B.2})$$

et sa dérivée première par rapport à $\eta_{q,\ell}$, pour $(q, q') \in \llbracket 0, Q \rrbracket^2$ et $(\ell, \ell') \in \llbracket 1, L \rrbracket^2$, donne donc

$$\frac{\partial \ln p(\mathbf{x} | \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{q,\ell}} = \sum_{n=k_q+1}^{k_{q+1}} \frac{\partial \ln p(x_n; \boldsymbol{\eta}_q)}{\partial \eta_{q,\ell}}. \quad (\text{B.3})$$

Ainsi, on peut directement remarquer que

$$\forall q' \neq q, \frac{\partial^2 \ln p(\mathbf{x} | \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{q,\ell} \partial \eta_{q',\ell'}} = 0, \quad (\text{B.4})$$

ce qui conduit à la structure diagonale par blocs de $V_{11}^{\eta^*}$.

Supposons maintenant que $q' = q$. En dérivant (B.3) par rapport à $\eta_{q,\ell}$, et en prenant l'espérance du tout par rapport à la vraisemblance jointe $p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})$, on obtient, en posant $p = Lq + \ell$ et $p' = Lq + \ell'$,

$$\begin{aligned} [\mathbf{V}_{11}]_{p,p'} &= - \sum_{\mathbf{k} \in \mathbb{Z}^Q} \int_{\Omega} \sum_{n=k_q+1}^{k_{q+1}} \frac{\partial^2 \ln p(x_n; \boldsymbol{\eta}_q)}{\partial \eta_{q,\ell} \partial \eta_{q,\ell'}} p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}) \, d\mathbf{x} \\ &= - \sum_{\mathbf{k} \in \mathbb{Z}^Q} \pi(\mathbf{k}) \int_{\Omega} \sum_{n=k_q+1}^{k_{q+1}} \frac{\partial^2 \ln p(x_n; \boldsymbol{\eta}_q)}{\partial \eta_{q,\ell} \partial \eta_{q,\ell'}} \prod_{i=0}^Q \prod_{m=k_i+1}^{k_{i+1}} p(x_m; \boldsymbol{\eta}_i) \, d\mathbf{x} \\ &= - \sum_{\mathbf{k} \in \mathbb{Z}^Q} \pi(\mathbf{k}) \sum_{n=k_q+1}^{k_{q+1}} \int_{\Omega} \left[\prod_{\substack{i=0 \\ i \neq q}}^Q \prod_{m=k_i+1}^{k_{i+1}} p(x_m; \boldsymbol{\eta}_i) \right] \left[\left(\prod_{\substack{m=k_q+1 \\ m \neq n}}^{k_{q+1}} p(x_m; \boldsymbol{\eta}_q) \right) \cdot \frac{\partial^2 \ln p(x_n; \boldsymbol{\eta}_q)}{\partial \eta_{q,\ell} \partial \eta_{q,\ell'}} p(x_n; \boldsymbol{\eta}_q) \right] d\mathbf{x} \end{aligned}$$

en scindant le double produit. Comme les variables sont séparées, on peut intervertir les signes de produit et d'intégrale, ce qui nous mène à

$$\begin{aligned} [\mathbf{V}_{11}]_{p,p'} &= - \sum_{\mathbf{k} \in \mathbb{Z}^Q} \pi(\mathbf{k}) \sum_{n=k_q+1}^{k_{q+1}} \left[\prod_{\substack{i=0 \\ i \neq q}}^Q \prod_{m=k_i+1}^{k_{i+1}} \int_{\Omega'} p(x_m; \boldsymbol{\eta}_i) \, dx_m \right] \\ &\quad \times \left[\left(\prod_{\substack{m=k_q+1 \\ m \neq n}}^{k_{q+1}} \int_{\Omega'} p(x_m; \boldsymbol{\eta}_q) \, dx_m \right) \int_{\Omega'} \frac{\partial^2 \ln p(x_n; \boldsymbol{\eta}_q)}{\partial \eta_{q,\ell} \partial \eta_{q,\ell'}} p(x_n; \boldsymbol{\eta}_q) \, dx_n \right] \\ &= - \frac{1}{\Delta^Q} \sum_{\substack{k_j=k_{j-1}+d \\ j \in \llbracket 1, Q \rrbracket}}^{k_{j-1}+D} \sum_{n=k_q+1}^{k_{q+1}} \int_{\Omega'} \frac{\partial^2 \ln p(x; \boldsymbol{\eta}_q)}{\partial \eta_{q,\ell} \partial \eta_{q,\ell'}} p(x; \boldsymbol{\eta}_q) \, dx, \end{aligned} \quad (\text{B.5})$$

après simplification des intégrales égales à 1. Notons que la dernière intégrale ne dépend en fait ni de n (ce qui explique la suppression de n en indice de la variable x), ni de \mathbf{k} , et qu'elle correspond à l'élément (ℓ, ℓ') de la matrice d'information de Fisher sur le vecteur $\boldsymbol{\eta}_q$, pour une seule observation sur le q -ième segment, comme définie dans (3.57). Nous obtenons donc, pour $q < Q$,

$$\begin{aligned} [\mathbf{V}_{11}]_{p,p'} &= \frac{[\mathbf{F}(\boldsymbol{\eta}_q)]_{\ell,\ell'}}{\Delta^Q} \sum_{\substack{k_j=k_{j-1}+d \\ j \in \llbracket 1, q \rrbracket}}^{k_{j-1}+D} \sum_{k_{q+1}=k_q+d}^{k_q+D} \Delta^{Q-q-1} (k_{q+1} - k_q) \\ &= \frac{[\mathbf{F}(\boldsymbol{\eta}_q)]_{\ell,\ell'}}{\Delta^{q+1}} \sum_{\substack{k_j=k_{j-1}+d \\ j \in \llbracket 1, q \rrbracket}}^{k_{j-1}+D} \left(\frac{\Delta(2k_q + d + D)}{2} - \Delta k_q \right) \\ &= \frac{d + D}{2} [\mathbf{F}(\boldsymbol{\eta}_q)]_{\ell,\ell'}, \end{aligned}$$

et, pour $q = Q$, en utilisant la même méthode,

$$[\mathbf{V}_{11}]_{p,p'} = \left(N - \frac{Q(d + D)}{2} \right) [\mathbf{F}(\boldsymbol{\eta}_q)]_{\ell,\ell'}. \quad (\text{B.6})$$

B.2 Expressions des éléments de la matrice $V_{12}^{\eta^*}$

Nous cherchons ici à obtenir l'expression membre de droite dans (3.63). Pour cela, en posant $\tilde{q} \in \llbracket 0, Q \rrbracket$, $\ell \in \llbracket 1, L \rrbracket$ et $q \in \llbracket 1, Q \rrbracket$, on développe d'abord la quantité

$$\begin{aligned}
 & \mathbb{E}_{\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}} \left\{ \frac{\partial \ln p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{\tilde{q}, \ell}} \frac{p^{s_q}(\mathbf{x}, \mathbf{k} + \mathbf{h}_q; \boldsymbol{\eta})}{p^{s_q}(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})} \right\} \\
 &= \sum_{\mathbf{k} \in \mathbb{Z}^Q} \int_{\Omega} \frac{\partial \ln p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{\tilde{q}, \ell}} p^{s_q}(\mathbf{x}, \mathbf{k} + \mathbf{h}_q; \boldsymbol{\eta}) p^{1-s_q}(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}) d\mathbf{x} \\
 &= \sum_{\mathbf{k} \in \mathbb{Z}^Q} \int_{\Omega} \frac{\partial \ln p(\mathbf{x} | \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{\tilde{q}, \ell}} (p(\mathbf{x} | \mathbf{k} + \mathbf{h}_q; \boldsymbol{\eta}) \pi(\mathbf{k} + \mathbf{h}_q))^{s_q} (p(\mathbf{x} | \mathbf{k}; \boldsymbol{\eta}) \pi(\mathbf{k}))^{1-s_q} d\mathbf{x} \\
 &= \sum_{\mathbf{k} \in \mathbb{Z}^Q} \pi^{s_q}(\mathbf{k} + \mathbf{h}_q) \pi^{1-s_q}(\mathbf{k}) \int_{\Omega} \sum_{n=k_{\tilde{q}}+1}^{k_{\tilde{q}}+1} \frac{\partial \ln p(x_n; \boldsymbol{\eta}_{\tilde{q}})}{\partial \eta_{\tilde{q}, \ell}} p^{s_q}(\mathbf{x} | \mathbf{k} + \mathbf{h}_q; \boldsymbol{\eta}) p^{1-s_q}(\mathbf{x} | \mathbf{k}; \boldsymbol{\eta}) d\mathbf{x}.
 \end{aligned} \tag{B.7}$$

D'une part, en utilisant les relations (A.6d) et (A.8), et le même type de manipulations que celles conduisant à (A.9) (en ignorant les intégrales sur Ω), on obtient, quel que soit le signe de h_q ,

$$\begin{aligned}
 p^{s_q}(\mathbf{x} | \mathbf{k} + \mathbf{h}_q; \boldsymbol{\eta}) p^{1-s_q}(\mathbf{x} | \mathbf{k}; \boldsymbol{\eta}) &= \left(\prod_{i=0}^Q \prod_{m=k_i+\delta_{i,q} \max\{h_q, 0\}+1}^{k_{i+1}-\delta_{i,q-1} \max\{-h_q, 0\}} p(x_m; \boldsymbol{\eta}_i) \right) \\
 &\quad \times \prod_{m=k_q-\max\{-h_q, 0\}+1}^{k_q+\max\{h_q, 0\}} p^{\varepsilon_{h_q}(s_q)}(x_m; \boldsymbol{\eta}_{q-1}) p^{1-\varepsilon_{h_q}(s_q)}(x_m; \boldsymbol{\eta}_q) \tag{B.8}
 \end{aligned}$$

avec $\varepsilon_{h_q}(s_q)$ défini en (3.13).

D'autre part, d'après (3.3) et (A.1), on a

$$\begin{aligned}
 \pi^{s_q}(\mathbf{k} + \mathbf{h}_q) \pi^{1-s_q}(\mathbf{k}) &= \frac{1}{\Delta^Q} \mathbb{1}_{\mathcal{K}}(\mathbf{k} + \mathbf{h}_q) \mathbb{1}_{\mathcal{K}}(\mathbf{k}) \\
 &= \frac{1}{\Delta^Q} \mathbb{1}_{(\mathcal{K} - \mathbf{h}_q) \cap \mathcal{K}}(\mathbf{k}) \\
 &= \frac{1}{\Delta^Q} \mathbb{1}_{\mathcal{K}'_{h_q, 0}}(\mathbf{k}) \tag{B.9}
 \end{aligned}$$

où, comme dans l'annexe A, $\mathcal{K} - \mathbf{h}_q$ représente, avec un abus de notation, l'ensemble \mathcal{K} translaté de $-\mathbf{h}_q$, c'est-à-dire

$$\mathcal{K} - \mathbf{h}_q \triangleq \{\mathbf{k} - \mathbf{h}_q \mid \mathbf{k} \in \mathcal{K}\}, \tag{B.10}$$

et $\mathcal{K}'_{h_q, 0}$ se déduit de la définition (A.4).

On peut alors poursuivre le calcul de (B.7) en considérant les trois cas :

1. tsi : $\tilde{q} \neq q - 1$ et $\tilde{q} \neq q$ (termes des triangles supérieur et inférieur de $V_{12}^{\eta^*}$, hors première et deuxième diagonales) ;
2. d1 : $\tilde{q} = q - 1$ (termes de la première diagonale) ;
3. d2 : $\tilde{q} = q$ (termes de la deuxième diagonale).

Nous donnons les détails pour chacun de ces trois cas dans les sections qui suivent.

B.2.1 Cas tsi : $\tilde{q} \neq q - 1$ and $\tilde{q} \neq q$

En injectant (B.8) et (B.9) dans (B.7), et en utilisant le même type de manipulations que celles menant à (B.5), on obtient

$$\begin{aligned} & \mathbb{E}_{\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}} \left\{ \frac{\partial \ln p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{\tilde{q}, \ell}} \frac{p^{s_q}(\mathbf{x}, \mathbf{k} + \mathbf{h}_q; \boldsymbol{\eta})}{p^{s_q}(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})} \right\} \\ &= \frac{1}{\Delta^Q} \sum_{\mathbf{k} \in \mathbb{Z}^Q} \mathbb{1}_{\mathcal{K}'_{h_q, 0}}(\mathbf{k}) \sum_{n=k_{\tilde{q}+1}}^{k_{\tilde{q}+1}} \left[\prod_{\substack{i=0 \\ i \neq \tilde{q}}}^Q \prod_{m=k_i + \delta_{i, q-1} \max\{h_q, 0\} + 1}^{k_{i+1} - \delta_{i, q-1} \max\{-h_q, 0\}} \int_{\Omega'} p(x_m; \boldsymbol{\eta}_i) dx_m \right. \\ & \quad \times \left(\prod_{\substack{m=k_{\tilde{q}+1} \\ m \neq n}}^{k_{\tilde{q}+1}} \int_{\Omega'} p(x_m; \boldsymbol{\eta}_{\tilde{q}}) dx_m \right) \int_{\Omega'} \frac{\partial \ln p(x_n; \boldsymbol{\eta}_{\tilde{q}})}{\partial \eta_{\tilde{q}, \ell}} p(x_n; \boldsymbol{\eta}_q) dx_n \\ & \quad \left. \times \prod_{m=k_q - \max\{-h_q, 0\} + 1}^{k_q + \max\{h_q, 0\}} \int_{\Omega'} p^{\varepsilon_{h_q}(s_q)}(x_m; \boldsymbol{\eta}_{q-1}) p^{1-\varepsilon_{h_q}(s_q)}(x_m; \boldsymbol{\eta}_q) dx_m \right]. \end{aligned} \quad (\text{B.11})$$

Or, du fait de la condition de régularité (2.95), qui est satisfaite dans le cas présent, et de l'indépendance des observations, on déduit que l'intégrale $\int_{\Omega'} \frac{\partial \ln p(x_n; \boldsymbol{\eta}_{\tilde{q}})}{\partial \eta_{\tilde{q}, \ell}} \Big|_{\boldsymbol{\eta}^*} p(x_n; \boldsymbol{\eta}_q^*) dx_n$ dans (B.11) est en fait nulle. Ainsi, pour $\tilde{q} \in \llbracket 0, Q \rrbracket$, $\ell \in \llbracket 1, L \rrbracket$ et $q \in \llbracket 1, Q - 1 \rrbracket$, tels que $\tilde{q} \neq q - 1$ et $\tilde{q} \neq q$, en posant $p = L\tilde{q} + \ell$, on obtient

$$[\mathbf{V}_{12}^{\eta^*}]_{p, q} = 0. \quad (\text{B.12})$$

B.2.2 Cas d1 : $\tilde{q} = q - 1$

Dans ce cas, le développement de (B.7) dépend du signe de h_q .

- Supposons d'abord $h_q > 0$. Ce cas est traité exactement de la même manière que dans la section précédente, i.e., on peut réécrire (B.11) de façon similaire en remplaçant seulement \tilde{q} par $q - 1$, et on trouve alors également

$$\mathbb{E}_{\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}^*} \left\{ \frac{\partial \ln p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{q-1, \ell}} \frac{p^{s_q}(\mathbf{x}, \mathbf{k} + \mathbf{h}_q; \boldsymbol{\eta})}{p^{s_q}(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})} \right\} = 0. \quad (\text{B.13})$$

- Supposons maintenant $h_q < 0$ (pour plus de clarté, on remplace " h_q " par " $-|h_q|$ "). Dans l'équation (B.11), nous devons scinder en deux la somme discrète par rapport à n , de sorte que

(B.11) se réécrit

$$\begin{aligned}
 & \mathbb{E}_{\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}} \left\{ \frac{\partial \ln p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{q-1, \ell}} \frac{p^{s_q}(\mathbf{x}, \mathbf{k} + \mathbf{h}_q; \boldsymbol{\eta})}{p^{s_q}(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})} \right\} \\
 &= \frac{1}{\Delta^Q} \sum_{\mathbf{k} \in \mathbb{Z}^Q} \mathbb{1}_{\mathcal{K}'_{h_q, 0}}(\mathbf{k}) \left[\sum_{n=k_{q-1}+1}^{k_q - |h_q|} \left(\prod_{\substack{i=0 \\ i \neq q-1}}^Q \prod_{m=k_i+1}^{k_{i+1}} \int_{\Omega'} p(x_m; \boldsymbol{\eta}_i) dx_m \right. \right. \\
 & \quad \times \left(\prod_{\substack{m=k_{q-1}+1 \\ m \neq n}}^{k_q - |h_q|} \int_{\Omega'} p(x_m; \boldsymbol{\eta}_{q-1}) dx_m \right) \int_{\Omega'} \frac{\partial \ln p(x_n; \boldsymbol{\eta}_{q-1})}{\partial \eta_{q-1, \ell}} p(x_n; \boldsymbol{\eta}_q) dx_n \\
 & \quad \times \left. \prod_{m=k_q - |h_q| + 1}^{k_q} \int_{\Omega'} p^{1-s_q}(x_m; \boldsymbol{\eta}_{q-1}) p^{s_q}(x_m; \boldsymbol{\eta}_q) dx_m \right) \\
 & \quad + \sum_{n=k_q - |h_q| + 1}^{k_q} \left(\prod_{i=0}^Q \prod_{m=k_i+1}^{k_{i+1} - \delta_{i, q-1} |h_q|} \int_{\Omega'} p(x_m; \boldsymbol{\eta}_i) dx_m \right. \\
 & \quad \times \prod_{\substack{m=k_q - |h_q| + 1 \\ m \neq n}}^{k_q} \int_{\Omega'} p^{1-s_q}(x_m; \boldsymbol{\eta}_{q-1}) p^{s_q}(x_m; \boldsymbol{\eta}_q) dx_m \\
 & \quad \times \left. \int_{\Omega'} \frac{\partial \ln p(x_n; \boldsymbol{\eta}_{q-1})}{\partial \eta_{q-1, \ell}} p^{1-s_q}(x_n; \boldsymbol{\eta}_{q-1}) p^{s_q}(x_n; \boldsymbol{\eta}_q) dx_n \right) \Bigg]. \tag{B.14}
 \end{aligned}$$

Notons que la première somme par rapport à n dans (B.14) (i.e., pour $n \in \llbracket k_{q-1} + d, k_q - |h_q| \rrbracket$) est écrite de la même manière que dans (B.11). Par conséquent, toujours du fait de la condition de régularité (2.95), cette partie est nulle. La seconde somme par rapport à n (i.e., pour $n \in \llbracket k_q - |h_q| + 1, k_q \rrbracket$) n'est pas nulle, et en utilisant les définitions (3.12) et (3.67), et en évaluant (B.14) en $\boldsymbol{\eta}^*$, on obtient

$$\mathbb{E}_{\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}^*} \left\{ \frac{\partial \ln p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{q-1, \ell}} \Bigg|_{\boldsymbol{\eta}^*} \frac{p^{s_q}(\mathbf{x}, \mathbf{k} + \mathbf{h}_q; \boldsymbol{\eta}^*)}{p^{s_q}(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}^*)} \right\} = \frac{|h_q|}{\Delta^Q} \rho_{q-1, q}^{|h_q|-1} (1-s_q) \varphi_{q, \eta_{q-1, \ell}} (1-s_q) \sum_{\mathbf{k} \in \mathbb{Z}^Q} \mathbb{1}_{\mathcal{K}'_{h_q, 0}}(\mathbf{k}). \tag{B.15}$$

On termine le calcul en remarquant que

$$\begin{aligned}
 & \frac{1}{\Delta^Q} \sum_{\mathbf{k} \in \mathbb{Z}^Q} \mathbb{1}_{\mathcal{K}'_{h_q, 0}}(\mathbf{k}) = \frac{1}{\Delta^Q} \sum_{\mathbf{k} \in \mathcal{K}'_{h_q, 0}} \dots \sum 1 \\
 &= \begin{cases} \frac{1}{\Delta^Q} \sum_{k_1=d}^D \sum_{k_2=k_1+d}^{k_1+D} \dots \sum_{k_q=k_{q-1}+|h_q|+d}^{k_{q-1}+D} \sum_{k_{q+1}=k_q+d}^{k_q-|h_q|+D} \dots \sum_{k_Q=k_{Q-1}+d}^{k_{Q-1}+D} 1 & \text{si } q < Q, \\ \frac{1}{\Delta^Q} \sum_{k_1=d}^D \sum_{k_2=k_1+d}^{k_1+D} \dots \sum_{k_Q=k_{Q-1}+|h_Q|+d}^{k_{Q-1}+D} 1 & \text{si } q = Q \end{cases} \\
 &= u(\Delta, h_q), \tag{B.16}
 \end{aligned}$$

et on obtient finalement

$$\mathbb{E}_{\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}^*} \left\{ \frac{\partial \ln p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{q-1, \ell}} \Big|_{\boldsymbol{\eta}^*} \frac{p^{s_q}(\mathbf{x}, \mathbf{k} + \mathbf{h}_q; \boldsymbol{\eta}^*)}{p^{s_q}(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}^*)} \right\} = |h_q| u(\Delta, h_q) \rho_{q-1, q}^{|h_q|^{-1}}(1 - s_q) \varphi_{q, \eta_{q-1, \ell}}(1 - s_q). \quad (\text{B.17})$$

Alors, en injectant correctement (B.13) et (B.17) dans (3.63), quel que soit le signe de h_q , on obtient (3.65).

B.2.3 Cas d2 : $\tilde{q} = q$

Dans ce cas également, le développement de (B.7) dépend du signe de h_q . Les manipulations sont analogues à celles du cas d1, de sorte qu'on obtient finalement, de la même manière que dans la section B.2.2,

$$\mathbb{E}_{\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}^*} \left\{ \frac{\partial \ln p(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta})}{\partial \eta_{q, \ell}} \Big|_{\boldsymbol{\eta}^*} \frac{p^{s_q}(\mathbf{x}, \mathbf{k} + \mathbf{h}_q; \boldsymbol{\eta}^*)}{p^{s_q}(\mathbf{x}, \mathbf{k}; \boldsymbol{\eta}^*)} \right\} = \begin{cases} |h_q| u(\Delta, h_q) \rho_{q-1, q}^{|h_q|^{-1}}(s_q) \varphi_{q, \eta_{q, \ell}}(s_q) & \text{si } h_q > 0, \\ 0 & \text{si } h_q < 0. \end{cases} \quad (\text{B.18})$$

Alors, en injectant (B.18) dans (3.63), on obtient, quel que soit le signe de h_q , (3.65).

Notations

Acronymes

Notation	Signification
BB	Borne de Barankin.
BBG	Borne de Barankin globale.
BCR	Borne de Cramér-Rao.
BCRB	Borne de Cramér-Rao bayésienne.
BCRWWB	Borne (combinée) de Cramér-Rao/Weiss-Weinstein bayésienne.
BCRWWH	Borne de Cramér-Rao/Weiss-Weinstein hybride.
BWW	Borne de Weiss-Weinstein.
CUSUM	<i>Cumulative sum</i> : somme cumulée.
EQM	Erreur quadratique moyenne.
EQMG	Erreur quadratique moyenne globale.
MAP	Maximum <i>a posteriori</i> .
ML	<i>Maximum likelihood</i> : maximum de vraisemblance.
MLMAP	Maximum de vraisemblance-maximum <i>a posteriori</i> .
MVU	<i>Minimum variance unbiased</i> : non biaisé à minimum de variance.
MMSE	<i>Minimum mean squared error</i> : erreur quadratique moyenne minimale.
REQM	Racine (carrée) de l'erreur quadratique moyenne.
REQMG	Racine (carrée) de l'erreur quadratique moyenne globale.
RSB	Rapport signal sur bruit.

Symboles mathématiques généraux

Notation	Signification
\mathbb{N}	Ensemble des nombres entiers naturels.
\mathbb{Z}	Ensemble des nombres entiers relatifs.
\mathbb{R}	Corps des nombres réels.
\mathbb{C}	Corps des nombres complexes.
$\text{Re}\{z\}$	Partie réelle d'un nombre complexe z .
$\text{Im}\{z\}$	Partie imaginaire d'un nombre complexe z .
z^*	Nombre complexe conjugué du nombre complexe z .
$ z $	Valeur absolue de z si $z \in \mathbb{R}$, ou module de z si $z \in \mathbb{C}$.

$\delta_{i,j}$	Symbole de Krönecker.
$\min_{\mathbf{y}} f(\mathbf{y}), \max_{\mathbf{y}} f(\mathbf{y})$	Valeur minimale ou maximale de l'expression $f(\mathbf{y})$
$\arg \min_{\mathbf{y}} f(\mathbf{y}), \arg \max_{\mathbf{y}} f(\mathbf{y})$	Valeur du vecteur \mathbf{y} minimisant ou maximisant l'expression $f(\mathbf{y})$.
$\min\{a_1, \dots, a_N\}$	Plus petit élément de l'ensemble $\{a_1, \dots, a_N\}$.
$\max\{a_1, \dots, a_N\}$	Plus grand élément de l'ensemble $\{a_1, \dots, a_N\}$.
$\Gamma(\alpha)$	Fonction Gamma, i.e., pour $\alpha > 0$: $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$.
$\mathbb{1}_{\mathcal{A}}(x)$	Fonction indicatrice de l'ensemble \mathcal{A} , i.e., $\mathbb{1}_{\mathcal{A}}(x) = 1$ si $x \in \mathcal{A}$, 0 sinon.
$\text{Card}(\mathcal{A})$	Cardinal de l'ensemble \mathcal{A} .
$\llbracket n, N \rrbracket$	Ensemble des entiers successifs de n à N : $\{n, n+1, \dots, N\}$.

Symboles et opérateurs matriciels

Notation	Signification
a, A	Les lettres en italique représentent une quantité scalaire.
\mathbf{a}	Les lettres en gras minuscules représentent une quantité vectorielle (vecteur colonne).
\mathbf{A}	Les lettres majuscules en gras représentent une quantité matricielle.
$\mathbb{K}^{M \times N}$	Ensemble des matrices de dimension $M \times N$ à coefficients sur le corps \mathbb{K} ($\mathbb{K} = \mathbb{R}$ ou \mathbb{C}).
\mathbb{S}^N	Ensemble des matrices carrées de taille $N \times N$ symétriques.
\mathbb{S}_+^N	Ensemble des matrices carrées de taille $N \times N$ symétriques, semi-définies positives.
\mathbb{S}_{++}^N	Ensemble des matrices carrées de taille $N \times N$ symétriques, définies positives.
\mathbf{A}^T	Transposée de la matrice \mathbf{A} .
\mathbf{A}^H	Transposée conjuguée de la matrice \mathbf{A} (à coefficients complexes).
\mathbf{I}_N	Matrice identité de taille $N \times N$.
$\det(\mathbf{A})$	Déterminant de la matrice \mathbf{A} .
$[\mathbf{A}]_{i,j}$	Élément correspondant à la i ligne et j colonne de la matrice \mathbf{A} .
$\text{tr}(\mathbf{A}) = \sum_{i=1}^N A_{i,i}$	Trace de la matrice carrée \mathbf{A} de taille $N \times N$.
$\mathbf{A} \succcurlyeq \mathbf{B}$	Relation d'ordre de Löwner, i.e., pour \mathbf{A} et \mathbf{B} deux matrices carrées symétriques, la matrice $\mathbf{A} - \mathbf{B}$ est semi-définie positive.
$\text{diag}(a_1, a_2, \dots, a_N)$	Matrice diagonale de taille $N \times N$, dont la diagonale est composée des éléments a_1, \dots, a_N .

Symboles relatifs aux probabilités

Notation	Signification
$\Pr(k = k)$	Probabilité que la variable aléatoire (discrète) k prenne la valeur k .

$p_x(x)^*, p(x)$	Densité de probabilité de la variable aléatoire (absolument continue) x .
$p_{x,y}(x, y)^*, p(x, y)$	Densité de probabilité jointe des variables aléatoires (absolument continues) x et y .
$p_{x y=y}(x y = y)^*, p(x y)$	Densité de probabilité conditionnelle ou vraisemblance (selon le contexte) de la variable aléatoire x conditionnée à l'évènement « la variable aléatoire y prend la valeur y ».
$\mathbb{E}\{\cdot\}$	Opérateur d'espérance mathématique.
$\mathbb{E}_{\mathbf{x}, \theta}\{\cdot\}$	Opérateur d'espérance mathématique par rapport à la loi jointe $p(\mathbf{x}, \theta)$.
$\mathbb{E}_{\mathbf{x} \theta}\{\cdot\}$	Opérateur d'espérance mathématique par rapport à la loi conditionnelle $p(\mathbf{x} \theta)$.
$x \sim \mathcal{L}$	La variable aléatoire x est distribuée selon la loi \mathcal{L} .
$x_N \stackrel{N \rightarrow +\infty}{\sim} \mathcal{L}$	La suite de variables aléatoires x_N est asymptotiquement (i.e., quand N tend vers l'infini) distribuée selon la loi \mathcal{L} .
$\mathcal{U}([a, b])$	Loi uniforme continue sur l'intervalle $[a, b]$.
$\mathcal{U}(\llbracket n, N \rrbracket)$	Loi uniforme discrète sur l'ensemble des entiers successifs $\llbracket n, N \rrbracket$.
$\mathcal{N}(\mathbf{m}, \Sigma)$	Loi normale multivariée réelle de vecteur moyen \mathbf{m} et de matrice de covariance Σ .
$\mathbb{CN}(\mathbf{m}, \Sigma)$	Loi normale multivariée complexe, circulaire, de vecteur moyen \mathbf{m} et de matrice de covariance Σ .
$\mathcal{E}(\Lambda)$	Loi exponentielle de paramètre Λ .
$\mathcal{P}(\lambda)$	Loi de Poisson de paramètre λ .
$\mathcal{P}^N(\lambda)$	Loi de Poisson tronquée, de paramètre λ et de valeur maximale N .
$\mathcal{G}(\alpha, \beta)$	Loi Gamma de paramètres α et β .

* Nous faisons ici référence à la manière classique de noter la densité de probabilité d'une variable aléatoire x , c'est-à-dire en notant le nom de la variable aléatoire associée en indice de la densité. Dans le manuscrit, nous nous affranchissons de cette notation, et notons simplement $p(x)$, la variable en argument de la fonction indiquant clairement la densité de quelle variable aléatoire il s'agit. Ainsi, lorsque nous notons une quantité η en indice de la densité, n'ayant pas de lien avec l'argument de la fonction, comme dans « $p_\eta(x)$ », cela signifie qu'il s'agit d'un paramètre (supposé connu) de la fonction de densité.

Table des figures

4.1	Deux exemples de signaux gaussiens comportant un point de rupture dans la moyenne à $n = 48$ (ligne verticale en pointillés rouges), pour $N = 128$ observations.	67
4.2	Deux exemples de signaux gaussiens comportant un point de rupture dans la variance à $n = 48$ (ligne verticale en pointillés rouges), pour $N = 128$ observations.	68
4.3	Tracé de $\frac{\partial W(h,s)}{\partial s}$ en fonction de s , dans le cas $ h = 1$ et $\omega_{0,1}^m = 0$ dB (trait bleu continu). Il est clair que la seule solution d'annulation de la fonction est $s^* = 1/2$ (pointillés rouges).	69
4.4	Évolution de la borne de Weiss-Weinstein (trait noir continu), de la borne de Barankin globale (trait rouge continu) et des racines (carrées) des EQM globales (REQMG) des estimateurs du MAP \hat{k}_1^{MAP} (étoiles bleues) et MMSE \hat{k}_1^{MMSE} (cercles verts) en fonction de la quantité de changement $\omega_{0,1}^m$, dans le cas de $Q = 1$ unique changement dans la moyenne de $N = 128$ données gaussiennes.	70
4.5	Évolution de la valeur h^* optimale des points test conduisant aux bornes de Weiss-Weinstein (carrés noirs), et de Barankin globale (croix rouges) les plus précises, en fonction de la quantité de changement $\omega_{0,1}^m$, dans le cas de $Q = 1$ unique changement dans la moyenne de $N = 128$ données gaussiennes.	70
4.6	Évolution de la borne de Weiss-Weinstein (trait noir continu), de la borne de Barankin globale (trait rouge continu) et des EQM globales des estimateurs du MAP \hat{k}_1^{MAP} (étoiles bleues) et MMSE \hat{k}_1^{MMSE} (cercles verts) en fonction de la quantité de changement $\omega_{0,1}^v$, dans le cas de $Q = 1$ unique changement dans la variance de $N = 128$ données gaussiennes.	71
4.7	Deux exemples de signaux poissonniens comportant un point de rupture dans le paramètre λ_q à $n = 48$ (ligne verticale en pointillés rouges), pour $N = 128$ observations.	72
4.8	Cas de $Q = 1$ changement dans le paramètre λ_0 de $N = 128$ observations poissonniennes. Évolution de la borne de Weiss-Weinstein (trait bleu continu), de la borne de Barankin globale (trait rouge continu) et de l'EQM globale de l'estimateur MMSE \hat{k}_1^{MMSE} (cercles verts) en fonction de la quantité de changement $\omega_{0,1}^\lambda$.	73
4.9	Deux exemples de signaux distribués selon une loi exponentielle comportant un point de rupture dans le paramètre Λ_q à $n = 48$ (ligne verticale en pointillés rouges), pour $N = 128$ observations.	74
4.10	Évolution de la borne de Weiss-Weinstein (trait bleu continu), de la borne de Barankin globale (trait rouge continu) et de l'EQM globale de l'estimateur MMSE \hat{k}_1^{MMSE} (cercles verts) en fonction de la quantité de changement $\omega_{0,1}^\Lambda$.	75
4.11	Évolution de $\text{tr}(\mathbf{BWW})$, pour plusieurs valeurs de $\mathbf{s} = [s, s, \dots, s]^\top$, $s \in]0, 1[$, dans le cas $Q = 2$ changements dans la moyenne de $N = 80$ observations gaussiennes, en fonction de la quantité de changement $\omega_{0,1}^m$.	80

4.12	Cas de $Q = 3$ changements dans la moyenne de $N = 80$ observations gaussiennes. Évolution des bornes de Weiss-Weinstein (traits pleins continus) et des EQM globales associées aux estimations par le MAP des points de rupture k_1 (étoiles bleues), k_2 (cercles rouges), et k_3 (carrés noirs), en fonction de la quantité de changement $\omega_{q-1,q}^m$	81
4.13	Cas de $Q = 3$ changements dans la moyenne de $N = 80$ observations gaussiennes. Évolution des valeurs test optimales pour la borne de Weiss-Weinstein, associées aux paramètres k_1 (étoiles bleues), k_2 (cercles rouges), et k_3 (carrés noirs), en fonction de la quantité de changement $\omega_{q-1,q}^m$	81
4.14	Cas de $Q = 3$ changements dans la variance de $N = 80$ observations gaussiennes. Évolution des bornes de Weiss-Weinstein (traits pleins continus) et des EQM globales associées aux estimations par le MAP des points de rupture k_1 (étoiles bleues), k_2 (cercles rouges), et k_3 (carrés noirs), en fonction de la quantité de changement $\omega_{q-1,q}^m$	82
4.15	Évolutions de l'EQMG globale de l'estimateur MMSE des points de rupture k_1 (à gauche) et k_2 (à droite), pour la loi <i>a priori</i> globalement uniforme (cercles verts) et des bornes de Weiss-Weinstein pour des distributions <i>a priori</i> diverses : marche aléatoire uniforme (MAU, trait plein bleu), uniforme sur des segments distincts (USD, pointillés oranges), globalement uniforme (GU, trait plein rouge), et marche aléatoire de Poisson tronquée (MAPT, pointillés noirs).	83
4.16	Cas de $Q = 3$ changements dans le paramètre λ_q de $N = 80$ observations poissonniennes. Évolution des bornes de Weiss-Weinstein (traits pleins continus) et des EQM globales associées à l'estimateur MAP des points de rupture k_1 (étoiles bleues), k_2 (cercles rouges), et k_3 (carrés noirs), en fonction de la quantité de changement $\omega_{q-1,q}^\lambda$	85
4.17	Évolutions des EQMG globales de l'estimateur MLMAP et des bornes de Weiss-Weinstein associées en fonction de la quantité de changement $\omega_{q-1,q}^\lambda$, dans le cas de données poissonniennes avec $N = 100$ observations et $Q = 3$ ruptures, pour les paramètres $(\lambda_q)_{q=0,\dots,3}$ (à gauche), et les points de rupture $(k_q)_{q=1,\dots,3}$ (à droite).	87
4.18	Évolutions des EQMG globales de l'estimateur MAP et des éléments de la borne de Cramér-Rao/Weiss-Weinstein bayésienne associée, en fonction de la quantité de changement moyenne $\bar{\omega}_{q-1,q}^\alpha$, dans le cas de données poissonniennes avec $N = 80$ observations et $Q = 1$ rupture, pour les paramètres λ_0 et λ_1 (a), et le point de rupture k_1 (b).	88
4.19	Évolutions des EQM globales des estimateurs des moyennes $(m_q)_{q=0,1}$ (figure (a)) et $(m_q)_{q=2,3}$ (figure (b)), et des estimateurs des points de ruptures $(k_q)_{q=1,\dots,Q}$ (figures (c) et (d)), comparées aux bornes associées, dans le cas de données gaussiennes changeant de moyenne, avec $N = 100$ observations et $Q = 3$ ruptures, en fonction de la quantité de changement $\omega_{q-1,q}^m$	90
4.20	Cas de $Q = 2$ changements dans la moyenne m_q de $N = 100$ observations gaussiennes. Évolution des bornes de Cramér-Rao/Weiss-Weinstein associées aux deux ruptures k_1 et k_2 (cercles rouges pour k_1 , points bleus pour k_2), en fonction de la longueur moyenne <i>a priori</i> des segments $[k_{q-1}, k_q]$, $q = 1, 2$, autrement dit $\mathbb{E}\{l_q\}$, pour trois valeurs de la quantité de changement différentes.	91
4.21	Cas de $Q = 3$ changements dans la variance de $N = 100$ observations gaussiennes. Évolution des EQM globales des estimateurs des variances sur chaque segment (a) et des points de rupture (b), et des bornes de Cramér-Rao/Weiss-Weinstein hybrides associées, en fonction de la quantité de changement $\omega_{0,1}^m$	93

4.22	Cas de $Q = 1$ changement dans la fréquence d'une sinusoïde complexe noyée dans bruit gaussien, avec $N = 20$ observations. Évolution des EQM globales des estimateurs MAP des fréquences avant et après rupture (a) et du point de rupture (b), avec dans chaque cas les bornes de Weiss-Weinstein associées, en fonction du rapport signal-sur-bruit $RSB = 1/\sigma^2$	96
A.1	Illustration d'une situation de chevauchement entre 3 segments consécutifs. . . .	106

Liste des tableaux

A.1	Expressions des bornes $(k_q)_{\min}$ et $(k_q)^{\max}$ des Q sommes sur k_1, \dots, k_Q dans l'équation (3.6), dans le cas diag ($b = a$). Dans ce tableau, on a $h_b = \pm h_a$	102
A.2	Expressions des bornes $(k_q)_{\min}$ et $(k_q)^{\max}$ des Q sommes sur k_1, \dots, k_Q dans l'équation (3.6), dans le cas psdiag ($b = a + 1$).	102
A.3	Expressions des bornes $(k_q)_{\min}$ et $(k_q)^{\max}$ des Q sommes sur k_1, \dots, k_Q dans l'équation (3.6), dans le cas ts ($b > a + 1$).	102

Bibliographie

- [Abe93] J. S. Abel. A bound on mean square estimate error. *IEEE Transactions on Information Theory*, 39(5) :1675–1680, September 1993.
- [Aka74] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6) :716–723, December 1974.
- [BA13] O. Besson and Y. Abramovich. On the Fisher information matrix for multivariate elliptically contoured distributions. *IEEE Signal Processing Letters*, 20(11) :1130–1133, November 2013.
- [Ban71] G. W. Bangs. *Array processing with generalized beamformers*. PhD thesis, Yale University, New Haven, CT, USA, 1971.
- [Bar49] E. W. Barankin. Locally best unbiased estimates. *Annals of Mathematical Statistics*, 20(4) :477–501, December 1949.
- [Bar72] Y. Bar-Shalom. Optimal simultaneous state estimation and parameter identification in linear discrete-time systems. *IEEE Transactions on Automatic Control*, AC-17 :308–319, 1972.
- [BB07] N. Brunel and F. Barbaresco. Doppler and Polarimetric Statistical Segmentation for Radar Clutter map based on Pairwise Markov Chains. In *Proc. of IEEE RADAR*, October 2007.
- [BBRE17] L. Bacharach, G. Bibiche, A. Renaux, and M. N. El Korso. Bornes bayésiennes pour la localisation d’un point de rupture : application à des processus exponentiels. In *Proc. du colloque GRETSI sur le traitement du signal et des images*, Juan-les-Pins, September 2017. in french.
- [Bel95] K. L. Bell. *Performance bounds in parameter estimation with application to bearing estimation*. PhD thesis, George Mason University, Fairfax, VA, USA, 1995.
- [BER16] L. Bacharach, M. N. El Korso, and A. Renaux. Weiss-Weinstein bound for an abrupt unknown frequency change. In *Proc. of IEEE Workshop on Statistical Signal Processing (SSP)*, pages 420–424, Palma de Mallorca, Spain, June 2016.
- [BERT] L. Bacharach, M. N. El Korso, A. Renaux, and J.-Y. Tournéret. A hybrid lower bound for parameter estimation of signals with multiple change-points. soumis à *IEEE Transactions on Signal Processing*.
- [BERT17] L. Bacharach, M. N. El Korso, A. Renaux, and J.-Y. Tournéret. A Bayesian Lower Bound for Parameter Estimation of Poisson Data Including Multiple Changes. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 4486–4490, New Orleans, LA, USA, March 2017.
- [BEV96] K. L. Bell, Y. Ephraim, and H. L. Van Trees. Explicit Ziv-Zakai lower bound for bearing estimation. *IEEE Transactions on Signal Processing*, 44(11) :2810–2824, November 1996.

- [BGR⁺08] S. Bay, B. Geller, A. Renaux, J.-P. Barbot, and J.-M. Brossier. On the hybrid Cramér-Rao bound and its application to dynamical phase estimation. *IEEE Signal Processing Letters*, 15 :453–456, 2008.
- [BH93] D. Barry and J. A. Hartigan. A Bayesian analysis for change-point problems. *Journal of the American Statistical Association*, 88(421) :309–319, March 1993.
- [BM96] A. Bartov and H. Messer. Analysis of inherent limitations in localizing step-like singularities in a continuous signal. In *Proc. IEEE SP Int. Symp. Time-Frequency Time-Scale Analysis*, pages 21–24, Paris, France, June 1996.
- [BN93] M. Basseville and I. V. Nikiforov. *Detection of Abrupt Changes, Theory and Application*. Englewood Cliffs, NJ : Prentice-Hall, 1993.
- [BRE18] L. Bacharach, A. Renaux, and M. N. El Korso. Prior influence on weiss-Weinstein bounds for multiple change-point estimation. In *Proc. of European Signal Processing Conference (EUSIPCO)*, pages –, Rome, Italy, September 2018.
- [BREC15] L. Bacharach, A. Renaux, M. N. El Korso, and E. Chaumette. Weiss-Weinstein bound for change-point estimation. In *Proc. of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 477–480, Cancún, Mexico, December 2015.
- [BREC17] L. Bacharach, A. Renaux, M. N. El Korso, and E. Chaumette. Weiss-Weinstein bound on multiple change-points estimation. *IEEE Transactions on Signal Processing*, 65(10) :2686–2700, May 2017.
- [BSEV97] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees. Extended Ziv-Zakai lower bound for vector parameter estimation. *IEEE Transactions on Information Theory*, 43(2) :624–637, March 1997.
- [BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge Univ. Press, Cambridge, U.K., 2004.
- [BV06] K. L. Bell and H. L. Van Trees. Combined Cramér-Rao/Weiss-Weinstein bound for tracking target bearing. In *Proc. IEEE Workshop Sensor Array Multi-channel Process.*, pages 273–277, Waltham, MA, USA, July 2006.
- [CG00] J. Chen and A. K. Gupta. *Parametric Statistical Change Point Analysis*. Birkhäuser Basel, 2000.
- [CGQL08] E. Chaumette, J. Galy, A. Quinlan, and P. Larzabal. A new Barankin bound approximation for the prediction of the threshold region performance of maximum likelihood estimators. *IEEE Transactions on Signal Processing*, 56(11) :5319–5333, November 2008.
- [CGS92] B. P. Carlin, A. E. Gelfand, and A. F. M. Smith. Hierarchical Bayesian analysis of changepoint problems. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(2) :389–405, 1992.
- [CH97] M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, New-York, NY (USA), Chichester (UK), December 1997.
- [Cha04] E. Chaumette. *Contribution à la caractérisation des performances des problèmes conjoints de détection et d’estimation*. PhD thesis, Ecole Normale Supérieure de Cachan, Cachan, FR, December 2004.
- [Cha14] É. Chaumette. *Caractérisation des problèmes conjoints de détection et d’estimation*. Habilitation à Diriger des Recherches, January 2014.

- [Che52] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. *Annals of Mathematical Statistics*, 23(4) :493–507, December 1952.
- [Chi98] S. Chib. Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86(2) :221–241, October 1998.
- [CR51] D. G. Chapman and H. Robbins. Minimum variance estimation without regularity assumptions. *Annals of Mathematical Statistics*, 22(4) :581–586, December 1951.
- [Cra46] H. Cramér. *Mathematical Methods of Statistics*, volume 9 of *Princeton Mathematics*. Princeton University Press, New-York, September 1946.
- [CS12] C. Choirat and R. Seri. Estimation in discrete parameter models. *Statistical Science*, 27(2) :278–293, 2012.
- [CZ64] H. Chernoff and S. Zacks. Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics*, 35(3) :999–1018, 1964.
- [Dar45] G. Darmois. Sur les limites de la dispersion de certaines estimations. *Revue de l'Institut International de Statistique*, 13(1/4) :9–15, 1945.
- [Dju94] P. M. Djurić. A MAP solution to off-line segmentation of signals. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 505–508, April 1994.
- [DP84] J. Deshayes and D. Picard. Lois asymptotiques des tests et estimateurs de rupture dans un modèle statistique classique. *Annales de l'Institut Henri Poincaré, section B*, 20(4) :309–327, January 1984.
- [DP86] J. Deshayes and D. Picard. Off-line statistical analysis of change-point models using non parametric and likelihood methods. In M. Basseville and A. Benveniste, editors, *Detection of Abrupt Changes in Signals and Dynamical Systems*, pages 103–168, Berlin, Heidelberg, 1986. Springer Berlin Heidelberg.
- [DTD07] N. Dobigeon, J.-Y. Tournet, and M. Davy. Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a bayesian sampling approach. *IEEE Transactions on Signal Processing*, 55(4) :1251–1263, April 2007.
- [DTS07] N. Dobigeon, J.-Y. Tournet, and J. D. Scargle. Joint segmentation of multivariate astronomical time series : Bayesian sampling with a hierarchical model. *IEEE Transactions on Signal Processing*, 55(2) :414–423, February 2007.
- [Fea06] P. Fearnhead. Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing*, 16(2) :203–213, June 2006.
- [Fis22] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222 :309–368, 1922.
- [FJ01] S. B. Fotopoulos and S. K. Jandhyala. Maximum likelihood estimation of a change-point for exponentially distributed random variables. *Statistics and Probability Letters*, 51 :423–429, 2001.
- [FJK10] S. B. Fotopoulos, S. K. Jandhyala, and E. Khapalova. Exact asymptotic distribution of change-point MLE for change in the mean of Gaussian sequences. *The Annals of Applied Statistics*, 4(2) :1081–1104, November 2010.

- [FL02] Ph. Forster and P. Larzabal. On lower bounds for deterministic parameter estimation. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 1137–1140, Orlando, FL, USA, May 2002.
- [Fre43] M. Frechet. Sur l’extension de certaines évaluations statistiques au cas de petit échantillons. *Revue de l’Institut International de Statistique*, 11(3/4) :182–205, 1943.
- [FT03] A. Ferrari and J.-Y. Tournet. Barankin lower bound for change points in independent sequences. In *Proc. IEEE Workshop Statist. Signal Process.*, pages 557–560, St. Louis, MO, USA, September 2003.
- [GB08] M. Grant and S. Boyd. *CVX : Matlab software for disciplined convex programming*. [Online]. Available : <http://stanford.edu/boyd/cvx>, Web page and software, December 2008.
- [GG13] M. S. Greco and F. Gini. Cramér-rao lower bounds on covariance matrix estimation for complex elliptically symmetric distributions. *IEEE Transactions on Signal Processing*, 61(24) :6401–6409, December 2013.
- [Gla72] F. E. Glave. A new look at the Barankin lower bound. *IEEE Transactions on Information Theory*, 18(3) :349–356, May 1972.
- [GRM98] F. Gini, R. Reggiannini, and U. Mengali. The modified Cramér-Rao bound in vector parameter estimation. *IEEE Transactions on Communications*, 46(1) :52–60, January 1998.
- [GRM16] J. G. Garcia, P. A. Roncaglio, and C. H. Muravchik. A Bayesian technique for real and integer parameters estimation in linear models and its application to GNSS high precision positioning. *IEEE Transactions on Signal Processing*, 64(4) :923–933, February 2016.
- [Gus00] F. Gustafsson. *Adaptive Filtering and Change Detection*. John Wiley Sons Ltd, 2000.
- [Ham50] J. M. Hammersley. On estimating restricted parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(2) :192–240, 1950.
- [Haw01] D. M. Hawkins. Fitting multiple change-point models to data. *Computational Statistics and Data Analysis*, 37(3) :323–341, September 2001.
- [HH70] D. V. Hinkley and E. A. Hinkley. Inference about the change-point in a sequence of binomial random variables. *Biometrika*, 57(1) :477–488, December 1970.
- [Hin70] D. V. Hinkley. Inference about the change-point in a sequence of random variables. *Biometrika*, 57(1) :1–18, April 1970.
- [Hin72] D. V. Hinkley. Time-ordered classification. *Biometrika*, 59(3) :509–523, 1972.
- [IH81] I. A. Ibragimov and R. Z. Has’minskii. *Statistical Estimation : Asymptotic Theory*. Springer-Verlag, New-York, 1981.
- [JF99] V. Jandhyala and S. Fotopoulos. Capturing the distributional behaviour of the maximum likelihood estimator of a changepoint. *Biometrika*, 86(1) :129–140, March 1999.
- [JFML13] V. Jandhyala, S. Fotopoulos, I. MacNeill, and P. Liu. Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, 34(4) :423–446, July 2013.
- [JPP14] C. Jauffret and A.-C. Pérez-Pignol. Cramér-Rao-type lower bound for a time of abrupt change. In *Proc. of International Conference on Information Fusion (FUSION)*, pages 1–7, Salamanca, Spain, July 2014.

- [Kay93] S. M. Kay. *Fundamentals of Statistical Signal Processing : Estimation Theory*, volume 1. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, March 1993.
- [KH92] R. Kakarala and A. O. Hero. On achievable accuracy in edge localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(7) :777–781, July 1992.
- [Lav98] M. Lavielle. Optimal segmentation of random processes. *IEEE Transactions on Signal Processing*, 46(5) :1365–1373, May 1998.
- [LC03] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer, New-York, NY, USA, 2nd edition, September 2003.
- [Liu10] P. Liu. *Maximum likelihood estimation of an unknown change-point in the parameters of a multivariate gaussian series with applications to environmental monitoring*. PhD thesis, Washington State University, Pullman, WA (USA), May 2010.
- [LL01] M. Lavielle and E. Lebarbier. An application of MCMC methods for the multiple change-points problem. *Signal Processing*, 81 :39–53, January 2001.
- [LO78] J. S. Lim and A. V. Oppenheim. All-pole modeling of degraded speech. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-26 :197–210, May 1978.
- [LRN07] P. S. La Rosa, A. Renaux, and A. Nehorai. Barankin bounds for multiple change points estimation : computational aspects. In *Proc. of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages –, St. Thomas, U.S. Virgin Islands, USA, December 2007.
- [LRNM10] P. S. La Rosa, A. Renaux, A. Nehorai, and C. H. Muravchik. Barankin-type lower bound on multiple change-point estimation. *IEEE Transactions on Signal Processing*, 58(11) :5534–5549, November 2010.
- [Mar97] T. L. Marzetta. Computing the Barankin bound by solving an unconstrained quadratic optimization problem. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages 3829–3832, Munich, DE, April 1997.
- [Meu92] G. Meurant. A review on the inverse of symmetric tridiagonal and block tridiagonal matrices. *SIAM Journal on Matrix Analysis and Applications*, 13(3) :707–728, July 1992.
- [MF06] A. Mohammad-Djafari and O. Féron. Bayesian approach to change points detection in time series. *Int. J. Imag. Syst. Tech.*, 16(5) :215–221, January 2006.
- [MH71] R. J. McAulay and E. M. Hofstetter. Barankin bounds on parameter estimation. *IEEE Transactions on Information Theory*, 17(6) :669–676, November 1971.
- [MS69] R. J. McAulay and L. P. Seidman. A useful form of the Barankin lower bound and its application to PPM threshold analysis. *IEEE Transactions on Information Theory*, 15(2) :273–279, March 1969.
- [Mus82] B. Musicus. *Iterative algorithms for optimal signal reconstruction and parameter identification given noisy and incomplete data*. PhD thesis, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA (USA), May 1982.
- [NH88] T. J. Nohara and S. Haykin. Application of the Weiss-Weinstein bound to a two dimensional antenna array. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(9) :1533–1534, September 1988.

- [NK95] S. Narasimhan and J. L. Krolik. Fundamental limits on acoustic source range estimation performance in uncertain ocean channels. *The Journal of the Acoustical Society of America*, 97(1) :215–226, January 1995.
- [NNV02] N. A. Nechval, K. N. Nechval, and E. Vasermanis. Detection of target signals in clutter using change point statistics. In *Proc. SPIE 4541, Image and Signal Processing for Remote Sensing VII*, 111, January 2002.
- [NO99] A. Niinimaa and H. Oja. Multivariate median. In S. Kotz, C. B. Read, and D. L. Banks, editors, *Encyclopedia of Statistical Sciences (Update Vol. 3)*, pages 497–505. Wiley, New-York, August 1999.
- [NV94] H. Nguyen and H. L. Van Trees. Comparison of performance bounds for DOA estimation. In *Proc. of IEEE Workshop on Statistical Signal and Array Processing (SSAP)*, volume 1, pages 313–316, June 1994.
- [PADF02] E. Punskeya, C. Andrieu, A. Doucet, and W. Fitzgerald. Bayesian curve fitting using MCMC with applications to signal segmentation. *IEEE Transactions on Signal Processing*, 50(3) :747–758, March 2002.
- [Pag54a] E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1-2) :100–115, June 1954.
- [Pag54b] E. S. Page. Control charts for the mean of a normal population. *Journal of the Royal Statistical Society. Series B (Methodological)*, 16(1) :131–135, 1954.
- [Pag55] E. S. Page. A test for change in a parameter occurring at an unknown point. *Biometrika*, 42(3-4) :523–527, December 1955.
- [Pag57] E. S. Page. On problems in which a change in a parameter occurs at an unknown time point. *Biometrika*, 44(1-2) :248–252, June 1957.
- [Pic85] D. Picard. Testing and estimating change-points in time series. *Advances in Applied Probability*, 17(4) :841–867, December 1985.
- [PR10] F. Pascal and A. Renaux. Statistical analysis of the covariance matrix MLE in K-distributed clutter. *Signal Processing*, 90(4) :1165–1175, April 2010.
- [QCL06] A. Quinlan, E. Chaumette, and P. Larzabal. A direct method to generate approximations of the Barankin bound. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 808–811, Toulouse, FR, May 2006.
- [Rao45] C. R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37 :81–91, 1945.
- [RB74] D. C. Rife and R. R. Boorstyn. Single tone parameter estimation from discrete time observations. *IEEE Transactions on Information Theory*, 20(5) :591–598, September 1974.
- [RD96] A. M. Reza and M. Doroodchi. Cramér-Rao lower bound on locations of sudden changes in a steplike signal. *IEEE Transactions on Signal Processing*, 44(10) :2551–2556, October 1996.
- [Ren06] A. Renaux. *Contribution à l’analyse des performances d’estimation en traitement statistique du signal*. PhD thesis, Ecole Normale Supérieure de Cachan, 2006.
- [Ren15] C. Ren. *Caractérisation des performances minimales d’estimation pour des modèles d’observations non-standards*. PhD thesis, Université Paris Sud - Paris XI, Orsay, France, September 2015.

- [RFBL07] A. Renaux, P. Forster, E. Boyer, and P. Larzabal. Unconditional maximum likelihood performance at finite number of samples and high signal-to-noise ratio. *IEEE Transactions on Signal Processing*, 55(5, Part 2) :2358–2364, May 2007.
- [RFCL06] A. Renaux, P. Forster, E. Chaumette, and P. Larzabal. On the high-SNR conditional maximum-likelihood estimator full statistical characterization. *IEEE Transactions on Signal Processing*, 54(12) :4840–4843, December 2006.
- [RFL⁺08] A. Renaux, P. Forster, P. Larzabal, C. D. Richmond, and A. Nehorai. A fresh look at the Bayesian bounds of the Weiss-Weinstein family. *IEEE Transactions on Signal Processing*, 56(11) :5334–5352, November 2008.
- [RFLR06] A. Renaux, P. Forster, P. Larzabal, and C. D. Richmond. The Bayesian Abel bound on the mean square error. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 9–12, Toulouse, FR, May 2006.
- [RGC⁺15] C. Ren, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux. Hybrid Barankin–Weiss-Weinstein bounds. *IEEE Signal Processing Letters*, 22(11) :2064–2068, November 2015.
- [RM95] I. Reuven and H. Messer. The use of the Barankin bound for determining the threshold SNR in estimating the bearing of a source in the presence of another. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 1645–1648, Detroit, MI, USA, May 1995.
- [RM97] I. Reuven and H. Messer. A Barankin-type lower bound on the estimation error of a hybrid parameter vector. *IEEE Transactions on Information Theory*, 43(3) :1084–1093, May 1997.
- [RN05] S. Reece and D. Nicholson. Tighter alternatives to the cramér-rao lower bound for discrete-time filtering. In *Proc. of 8th International Conference on Information Fusion (FUSION)*, pages 101–106, July 2005. Philadelphia, PA, USA.
- [RO04] I. Rapoport and Y. Oshman. A new estimation error lower bound for interruption indicators in systems with uncertain measurements. *IEEE Transactions on Information Theory*, 50(12) :3375–3384, 2004.
- [Rob07] C. P. Robert. *The Bayesian Choice : From Decision Theoretic Foundations to Computational Implementation*. Springer-Verlag, New-York, NY, USA, January 2007.
- [Roc86] Y. Rockah. *Array processing in the presence of uncertainty*. PhD thesis, Yale University, New Haven, CT (USA), January 1986.
- [RS87] Y. Rockah and P. Schultheiss. Array shape calibration using sources in unknown locations—part I : Far-field sources. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(3) :286–299, March 1987.
- [SB13] L. Song and P. Bondon. Structural changes estimation for strongly dependent processes. *Journal of Statistical Computation and Simulation*, 83(10) :1783–1806, 2013.
- [Sch78] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2) :461–464, March 1978.
- [Ser09] R. J. Serfling. *Approximation theorems of mathematical statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, New-York, NY (USA), Chichester (UK), December 2009.

- [She31] W. A. Shewhart. *Economic Control of Quality Manufactured Products*. Van Nostrand, New York, 1931.
- [Shy61] A. N. Shyriaev. The problem of the most rapid detection of a disturbance in a stationary process. *Soviet Math. Dokl.*, (2) :795–799, 1961.
- [Sle54] D. Slepian. Estimation of signal parameters in the presence of noise. *Trans. IRE Professional Group Inf. Theory*, 3(3) :68–69, 1954.
- [Smi75] A. F. M. Smith. A bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, 62(2) :407–416, August 1975.
- [SN89] P. Stoica and A. Nehorai. MUSIC, maximum likelihood and the Cramér-Rao bound. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(5) :720–741, May 1989.
- [SN90] P. Stoica and A. Nehorai. MUSIC, maximum likelihood and the Cramér-Rao bound : further results and comparisons. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(12) :2140–2150, December 1990.
- [SS98] A. Swami and B. Sadler. Cramér-Rao bounds for step-change localization in additive and multiplicative noise. In *Proc. of IEEE Workshop on Statistical Signal and Array Processing (SSAP)*, pages 403–406, Portland, OR, USA, September 1998.
- [SS99] B. Sadler and A. Swami. Analysis of multiscale products for step detection and estimation. *IEEE Transactions on Information Theory*, 45(3) :1043–1051, April 1999.
- [TC97] J.-Y. Tournet and M. Chabert. Off-line detection and estimation of abrupt changes corrupted by multiplicative colored gaussian noise. In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3693–3696, Munich, Germany, April 1997.
- [TDL03] J.-Y. Tournet, M. Doisy, and M. Lavielle. Bayesian off-line detection of multiple change-points corrupted by multiplicative noise : application to SAR image edge detection. *Signal Processing*, 83(9) :1871–1887, September 2003.
- [TFS04] J.-Y. Tournet, A. Ferrari, and A. Swami. Cramér-rao lower bounds for change-points in additive and multiplicative noise. *Signal Processing*, 84(7) :1071–1088, July 2004.
- [TMN98] P. Tichavský, C. H. Muravchik, and A. Nehorai. Posterior cramér-rao bounds for discrete-time nonlinear filtering. *IEEE Transactions on Signal Processing*, 46(5) :1386–1396, May 1998.
- [TRB⁺14] N. D. Tran, A. Renaux, R. Boyer, S. Marcos, and P. Larzabal. Performance bounds for the pulse-phase estimation of x-ray pulsars. *IEEE Transactions on Aerospace and Electronic Systems*, 50(1) :786–793, January 2014.
- [TT09] K. Todros and J. Tabrikian. Hybrid lower bound via compression of the sampled CLR function. In *Proc. IEEE Workshop Statist. Signal Process.*, pages 602–605, Cardiff, Wales, UK, August 2009.
- [TT10a] K. Todros and J. Tabrikian. General classes of performance lower bounds for parameter estimation - part I : non-Bayesian bounds for unbiased estimators. *IEEE Transactions on Information Theory*, 56(10) :5045–5063, October 2010.
- [TT10b] K. Todros and J. Tabrikian. General classes of performance lower bounds for parameter estimation - part II : Bayesian bounds. *IEEE Transactions on Information Theory*, 56(10) :5064–5082, October 2010.

- [TW07] P. Tichavský and K. Wong. Quasi-fluid-mecanics based quasi-Bayesian Cramér-Rao bounds for towed-array direction finding. *IEEE Transactions on Signal Processing*, 52(1) :36–47, January 2007.
- [Van68] H. L. Van Trees. *Detection, Estimation and Modulation Theory*, volume 1. John Wiley & Sons, New-York, NY, USA, 1968.
- [Van98] A. W. Van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.
- [Van01] H. L. Van Trees. *Detection, Estimation and Modulation Theory : Radar-Sonar Signal Processing and Gaussian Signals in Noise*, volume 3. John Wiley & Sons, New-York, NY, USA, September 2001.
- [VB07] H. L. Van Trees and K. L. Bell, editors. *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Wiley/IEEE Press, New-York, NY, USA, September 2007.
- [Vos81] L. J. Vostrikova. Detecting “disorder” in multidimensional random processes. *Soviet Mathematics Doklady*, 24 :55–59, 1981.
- [VRBM14] D. T. Vu, A. Renaux, R. Boyer, and S. Marcos. Some results on the Weiss-Weinstein bound for conditional and unconditional signal models in array processing. *Signal Processing*, 95(2) :126–148, February 2014.
- [VVRB12] J. Vila-Valls, L. Ros, and J. M. Brossier. Joint oversampled carrier and time-delay synchronization in digital communications with large excess bandwidth. *Signal Processing*, 92(1) :76–88, January 2012.
- [WK85] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-33(2) :387–392, April 1985.
- [WW88] E. Weinstein and A. J. Weiss. A general class of lower bounds in parameter estimation. *IEEE Transactions on Information Theory*, 34(2) :338–342, March 1988.
- [XBR04] W. Xu, A. B. Baggeroer, and C. D. Richmond. Bayesian bounds for matched-field parameter estimation. *IEEE Transactions on Signal Processing*, 52(12) :3293–3305, December 2004.
- [XMGM16] F. Xavier, C. F. Mecklenbrauker, P. Gerstoft, and G. Matz. Weiss-Weinstein bounds for various priors. In *Proc. IEEE Workshop Sensor Array Multi-channel Process.*, pages 1–5, Rio de Janeiro, Brazil, July 2016.
- [Xu01] W. Xu. *Performance bounds on matched-field methods for source localization and estimation of ocean environmental parameters*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, June 2001.
- [Yao87] Y.-C. Yao. Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *The Annals of Statistics*, 15(3) :1321–1328, 1987.
- [Yao88] Y.-C. Yao. Estimating the number of change-points via Schwarz’ criterion. *Statistics and Probability Letters*, 6 :181–189, February 1988.
- [Yer00] A. Yeredor. The joint MAP-ML criterion and its relation to ML and to extended least-squares. *IEEE Transactions on Signal Processing*, 48(12) :3484–3492, December 2000.

Titre : Caractérisation des limites fondamentales de l'erreur quadratique moyenne pour l'estimation de signaux comportant des points de rupture

Mots clés : Bornes inférieures de l'erreur quadratique moyenne, bornes de Weiss-Weinstein, bornes de Cramér-Rao, maximum de vraisemblance (MV), maximum a posteriori (MAP), performances d'estimation asymptotiques et non asymptotiques, points de ruptures

Résumé : Cette thèse porte sur l'analyse des performances d'estimateurs en traitement du signal, à travers l'étude des bornes inférieures de l'erreur quadratique moyenne (EQM) appliquées à l'estimation de points de rupture. Ces outils permettent de caractériser le comportement d'estimateurs, tels que celui du maximum de vraisemblance (dans le contexte fréquentiste), ou ceux du maximum a posteriori et de la moyenne conditionnelle (dans le contexte bayésien).

La difficulté majeure provient du fait que, pour un signal échantillonné, les paramètres d'intérêt (à savoir les points de rupture) appartiennent à un espace discret. En conséquence, les résultats asymptotiques classiques, comme la normalité asymptotique du maximum de vraisemblance, ou la borne de Cramér-Rao ne s'appliquent plus. Quelques résultats sur la distribution asymptotique du maximum de vraisemblance provenant de la communauté mathématique sont actuellement disponibles, mais leur applicabilité à des problèmes pratiques de traitement du signal n'est pas immédiate.

En revanche, si l'on se concentre sur les moments d'ordre 2 des estimateurs, les bornes inférieures de l'EQM forment des outils pertinents, qui ont fait l'objet d'importants travaux ces dernières années. Ceux-ci ont notamment abouti à des

inégalités plus précises que la borne de Cramér-Rao, qui jouissent en outre de conditions de régularité plus faibles, et ce, même en régime non asymptotique, permettant ainsi de délimiter la plage de fonctionnement optimal des estimateurs.

Le but de cette thèse est, d'une part, de compléter la caractérisation de la zone asymptotique (en particulier lorsque le rapport signal sur bruit est élevé et/ou pour un nombre d'observations infini) pour des problèmes d'estimation de points de rupture. D'autre part, le but est de donner les limites fondamentales de l'EQM d'un estimateur dans la plage non asymptotique. Les outils utilisés ici sont les bornes inférieures de l'EQM de la famille Weiss-Weinstein, qui est déjà connue pour être plus précise que la borne de Cramér-Rao dans les contextes, entre autres, de l'analyse spectrale et du traitement d'antenne. Nous fournissons une forme compacte de cette famille dans le cas d'un seul et de plusieurs points de ruptures, puis nous étendons notre analyse aux cas où les paramètres des distributions sont inconnus. Nous fournissons également une analyse de la robustesse de cette famille vis-à-vis de la distribution a priori sur la localisation des ruptures. Enfin, nous appliquons ces bornes à plusieurs problèmes pratiques : données gaussiennes, poissonniennes et processus exponentiels.

Title : Characterization of mean squared error fundamental limitations in parameter estimation of signals with change-points

Keywords : Lower bounds on the mean squared error, Weiss-Weinstein bounds, Cramér-Rao bounds, maximum likelihood (ML), maximum a posteriori (MAP), asymptotic and non-asymptotic estimation performance, change-points

Abstract : This thesis deals with the study of estimation performance in signal processing, and focuses on the analysis of lower bounds on the Mean Squared Error (MSE) for abrupt changepoint estimation. Those tools contribute to characterizing the estimation behavior for estimators such as the maximum likelihood estimator (in the frequentist context), as well as the maximum a posteriori and the conditional mean estimators (in the Bayesian context).

The main difficulty comes from the fact that, when dealing with sampled signals, the parameters of interest (i.e., the change-point locations) lie on a discrete space. Consequently, the classical large sample theory results (e.g., the asymptotic normality of the maximum likelihood estimator) or the Cramér-Rao bound do not apply. Some results concerning the asymptotic distribution of the maximum likelihood estimator are available in the mathematical literature, but their applicability to practical signal processing problems is not immediate.

Focusing on the 2nd-order moments of the estimators, lower bounds on the MSE make up essential tools, which have been the subject of many studies in the last years.

As a result, new inequalities have been proposed, leading to tighter lower bounds in comparison to the Cramér-Rao bound. These new lower bounds require milder regularity conditions and enable the prediction of estimators' behavior in terms of MSE, both in asymptotic and non-asymptotic regimes.

The goal of this thesis is to complete previous results on lower bounds in the asymptotic context (i.e., when the number of samples and/or the signal-to-noise ratio is high) for change-point estimation but, also, to provide an analysis in the non-asymptotic setting. The tools used here are lower bounds from the Weiss-Weinstein family, which are already known in signal processing to outperform the Cramér-Rao bound in applications such as spectral analysis or array processing. A closed-form expression of this family is provided for a single and multiple change-points, and some extensions are given when the distribution parameters are unknown. An analysis in terms of robustness regarding the prior distribution on the change locations is also provided. Finally, we apply our results to specific problems, such as Gaussian data, Poisson data and exponentially distributed data.

