UNIVERSITÉ TOULOUSE III – PAUL SABATIER

U.F.R. PHYSIQUE CHIMIE AUTOMATIQUE

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ TOULOUSE III

(SCIENCES)

Discipline : Traitement du Signal

par

Sébastien BOURGUIGNON

ANALYSE SPECTRALE A HAUTE RÉSOLUTION DE SIGNAUX IRRÉGULIÈREMENT ÉCHANTILLONNÉS : APPLICATION A L'ASTROPHYSIQUE

Soutenue le 14 décembre 2005 devant le jury composé de :

Jean-Yves Tourneret	Président
Thierry Chonavel	Rapporteur
Jérôme Idier	Rapporteur
Albert Bijaoui	Examinateur
Yannick Deville	Examinateur
Hervé Carfantan	Directeur de thèse
Sylvie Roques	Invitée

Laboratoire d'Astrophysique de Toulouse et de Tarbes – UMR 5572

Université Paul Sabatier Toulouse 3 – CNRS

Observatoire Midi–Pyrénées

14, avenue Edouard Belin

31400 Toulouse

FRANCE







À Irene

À ma mère

Remerciements

À tout seigneur... l'honneur d'ouvrir le bal des remerciements. Étant désormais affranchi de toute servitude, je dois reconnaître avoir eu un réel plaisir à travailler en compagnie d'Hervé, qui a su se montrer à la fois en directeur de thèse toujours disponible, intègre et enthousiaste, en collègue enseignant motivé et motivant, en psychologue compréhensif, en cuisinier remarquable et surtout en véritable ami.

J'aimerais particulièrement remercier Thierry Chonavel et Jérôme Idier pour leur examen minutieux de ce manuscrit ainsi que pour leurs remarques constructives dont, je l'espère, ce document aura pu tirer profit. Jean-Yves Tourneret et Yannick Deville ont apporté leurs compétences théoriques pour rendre la soutenance intéressante du point de vue du traiteur de signal ; qu'ils en soient également remerciés. Je suis enfin reconnaissant à Sylvie Roques et Albert Bijaoui d'avoir étoffé ce jury de leur expertise astronomique. Sur le plan scientifique, je tiens encore à saluer Jean-Jacques Fuchs, dont les interventions auront permis à ce travail de rebondir vers des perspectives prometteuses.

Je souhaite également remercier l'Observatoire Midi-Pyrénées pour m'avoir permis de préparer cette thèse dans des conditions confortables. Plus particulièrement, je remercie Anne-Marie, Christine, Dominique, Josiane et Éric pour leur sympathie permanente. L'équipe Signal, Image, Instrumentation du laboratoire d'Astrophysique de Toulouse et de Tarbes m'a accueilli très chaleureusement et je souhaite leur témoigner toute ma gratitude. Salut à vous, donc, Jean, Jean-François, Sylvie, Éric, Laurent, Jean-Louis, José, Yannick, Shahram, Christophe, Denis, Rima, Matthieu, Hicham et Johan. Je remercie encore les structures espagnoles de la Biblioteca Nacional de España et de la bibliothèque de la Universidad Nacional de Educación a Distancia, dont les ressources ont également contribué à la réalisation de ce travail.

Dans un contexte moins académique mais tout aussi essentiel, ces années passées au laboratoire d'Astrophysique auront occasionné de nombreuses et enrichissantes rencontres. Les précaires, ex-précaires et assimilés retrouveront dans ces lignes une occasion d'être rassemblés autre que la dégustation du café des montagnes. Alors, en vrac, salut à vous, Audrey, Magali, Céline & Seb, Virginie & Marceau, Béné, Pat, Mika, Matthieu, Sandrine & Marc, Nassim, Philippe, Erika, Johan, Vicky, Amaro, Magalie, Davide, Ruben, Nico et Herald. Salut à toi, Loïc, se séparer après de trois ans de vie de bureau commune, et qui plus est heureuse, on n'en ressort pas indemne! Salut à toi, Seb, qui as pris la relève pour supporter et apaiser mon humeur variable des mois de rédaction et les diverses gueulantes de fin de thèse. En souvenir des moments épiques que furent, pour ne citer qu'eux, le barbecue du frigo, les apéros de la ferme ou encore les soirées internationale à Arbas.

Ayant également eu une vie extérieure au laboratoire durant ces quatre années, je me dois d'enchaîner par une pensée pour mes amis d'ailleurs. Salut à vous, Clarita, Perrine, Florent,

Angéline, Bénédicte, Cyril, Ben, Touss, Flo, Lo, Fred, Guillaume, Tim, François, Ana, Martin, Daff et Marlon. Mention spéciale pour Anne-Sophie et Norbert, notamment pour l'exercice de leurs talents lors de la préparation de l'après-thèse et pour m'avoir entouré de toute leur gentillesse lors des périodes plus mouvementées qui l'ont précédée. Merci aussi à Armand Guy. Lecteur, si tu t'estimes lésé de n'avoir pas trouvé ta place ici, passe donc à la maison et on arrangera ça.

Naturellement, mes remerciements les plus chaleureux s'adressent à ma famille, qui a toujours soutenu mes projets universitaires et privés. Il y a une part de vous tous dans l'aboutissement de cette étape. Je remercie notamment mon père et ma sœur pour leur encouragement permanent. Mike, bro, merci aussi pour m'avoir permis d'écrire un anglais un peu moins « awkward ». Une personne ne lira malheureusement pas ces mots, étant partie beaucoup trop tôt. D'où tu es, c'est pourtant toi qui m'as redonné la motivation pour aller jusqu'au bout.

Irene, c'est avec toi que je termine ces pensées. Hé oui, c'est également pendant la préparation de cette thèse que nous nous sommes rencontrés, et ton arrivée dans ma vie a été à coup sûr mon soutien le plus fort. Tu sais pourtant bien que le mot merci atteint vite ses limites. Tu sais aussi, pour reprendre les mots du vieux Georges, que pour te rendre hommage, il est d'autres moyens...

Introduction

Le travail de thèse présenté dans ce manuscrit a été réalisé au sein de l'équipe « Signal, Image et Instrumentation » du Laboratoire d'Astrophysique de l'Observatoire Midi-Pyrénées et la thématique abordée est l'analyse de séries temporelles issues d'observations astronomiques. Nous nous intéressons précisément aux méthodes d'étude de la variabilité temporelle présente dans certains types d'objets stellaires. En raison de leur éloignement, ces étoiles sont perçues par l'observateur comme des objets ponctuels ; leur étude s'effectue alors de manière indirecte en analysant l'évolution temporelle d'un paramètre caractéristique. Il s'agit par exemple de la variation du flux lumineux émis par une étoile ou encore de fluctuations de la vitesse radiale d'une étoile présentant des mouvements en surface. L'analyse de ces *courbes de lumière* ou *courbes de vitesse radiale* est donc un outil précieux pour l'astrophysicien. En particulier, la recherche de variations périodiques, voire sinusoïdales, représente un enjeu majeur. En astérosismologie, par exemple, la détermination des modes de vibration d'une étoile variable permet d'en déduire des paramètres essentiels comme sa masse, sa température de surface ou encore sa composition chimique. Dans un domaine différent, la connaissance de la période de révolution d'un système d'étoile double rend possible le calcul des différents éléments orbitaux. Un autre champ d'application très récent concerne la détection indirecte d'exoplanètes. Du point de vue du traitement du signal, l'étude de telles courbes peut être abordée sous l'angle de l'estimation du *spectre* qui leur est associé, dans lequel nous recherchons des raies traduisant des variations temporelles sinusoïdales.

La disponibilité des instruments d'observation étant limitée – il y a beaucoup trop d'objets à étudier devant le nombre de télescopes! – les données recueillies sont souvent en nombre restreint, d'autant plus si des conditions météorologiques défavorables perturbent les campagnes d'observation. Par ailleurs, l'alternance du jour et de la nuit ainsi que les variations saisonnières de l'hémisphère céleste observable rendent les objets étudiés périodiquement invisibles. De plus les mesures au cours d'une même série d'observations sont généralement irrégulièrement espacées dans le temps. On comprend alors aisément que l'impossibilité d'obtenir des couvertures observationnelles satisfaisantes est particulièrement pénalisante lorsque l'on s'intéresse à des phénomènes variables au cours du temps. En termes d'analyse spectrale, la contrainte d'un *échantillonnage irrégulier*, comportant de larges périodes sans mesures, rend le spectre de Fourier délicat à exploiter : la convolution des raies spectrales théoriques par la réponse en fréquence de la fenêtre d'observation, à laquelle les irrégularités de l'échantillonnage confèrent une forme complexe, peut générer des alias indésirables et masquer la présence de composantes fréquentielles de faible amplitude. De même, les méthodes d'analyse spectrale paramétrique apparues dans les années 1980 [Kay, 1988], particulièrement efficaces comparées à l'analyse de Fourier pour des données à temps court, ne sont pas directement transposables au cas de l'échantillonnage irrégulier. Depuis les années 1990, de nouvelles techniques d'analyse spectrale non paramétrique ont été développées qui abordent l'estimation spectrale sous la forme d'un problème inverse linéaire, le spectre étant discrétisé sur une grille fréquentielle arbitrairement fine. Le manque de données et l'objectif d'une

analyse à haute résolution confèrent à ce problème un caractère sous-déterminé : il s'agit alors de le *régulariser*, c'est-à-dire de contraindre le spectre recherché à vérifier des propriétés supplémentaires *a priori* attendues. Pour le problème envisagé ici, l'estimation de spectres de raies se traduit par la recherche de solutions *parcimonieuses*, c'est-à-dire de valeurs nulles excepté en un nombre restreint de positions fréquentielles.

Cette thèse aborde deux approches de la régularisation, qui ont déjà été en partie envisagées pour le problème de l'analyse spectrale. La première relève du domaine de l'optimisation et consiste à minimiser un critère des moindres carrés pénalisé par une fonction favorisant les solutions parcimonieuses. Cette démarche s'inscrit dans un cadre méthodologique fréquemment rencontré depuis les années 1990 [Idier, 2001a, chapitre 2] et des travaux récents en analyse spectrale [Sacchi *et al.*, 1998, Ciuciu *et al.*, 2001] ont proposé différentes pénalisations adaptées à l'estimation de raies spectrales ainsi que des algorithmes d'optimisation particulièrement efficaces pour des données à échantillonnage régulier.
Dans cette approche, notre contribution présente un double objectif. Nous proposons l'emploi de la pénalisation par la norme ℓ^1 qui, bien qu'universellement reconnue en termes d'information parcimonieuse dans le cas de variables réelles, n'a jamais été utilisée à notre connaissance sous la forme proposée à variables complexes du problème d'analyse spectrale. Nous présentons ensuite plusieurs solutions algorithmiques performantes adaptées au type de pénalisation employé – pénalisation $\ell^2\ell^1$ utilisée par Ciuciu *et al.* [2001], pénalisation par la norme ℓ^1 – et prenant en compte, dans la mesure du possible, les spécificités introduites par l'échantillonnage irrégulier.

La seconde approche s'appuie sur une modélisation probabiliste du spectre à reconstruire sous la forme d'un processus Bernoulli-Gaussien, structurellement adapté à des solutions parcimonieuses, et dont l'estimation des paramètres peut être réalisée dans le cadre bayésien. Cette modélisation a été initialement introduite pour la déconvolution impulsionnelle [Kormylo et Mendel, 1982] et son utilisation pour l'analyse spectrale a été proposée par Duvaut et Dublanchet [1995], Dublanchet [1996] dans le cas d'un échantillonnage régulier. Alors que ces derniers exploitent une approche de type maximum *a posteriori*, nous envisageons l'estimation alternative au sens de la moyenne *a posteriori* comme l'ont proposé Cheng *et al.* [1996] en déconvolution impulsionnelle. L'utilisation d'algorithmes d'échantillonnage stochastique rend alors possible une estimation entièrement non supervisée et permet d'associer une *probabilité* à la détection des raies spectrales, ce qui est essentiel pour les astrophysiciens. Notre apport est ici encore principalement algorithmique ; nous proposons des améliorations efficaces de l'échantillonneur de Gibbs initialement proposé par Cheng *et al.* [1996], pour lequel l'exploration de la distribution de probabilité d'intérêt s'est révélée pathologiquement lente sur ce problème.
Nous considérons également l'extension à un modèle Bernoulli-Gaussien non linéaire qui prend en compte des décalages en fréquence, lequel a été introduit par Dublanchet [1996], ce qui permet de faire évoluer les paramètres fréquentiels sur un domaine *continu*. À nouveau, nous préférons une stratégie d'estimation par la moyenne *a posteriori* plutôt que l'approche orientée optimisation initialement proposée. Le modèle Bernoulli-Gaussien avec décalages étant plus complexe, nous proposons le recours à un algorithme d'échantillonnage hybride afin d'estimer, conjointement et de manière non supervisée, l'ensemble des paramètres ainsi que les variances associées. Notons que si les algorithmes développés pour l'approche pénalisée s'efforcent de prendre en compte les spécificités de l'échantillonnage irrégulier, l'estimation par cette deuxième approche s'applique indifféremment à tous les types d'échantillonnage.

Ce manuscrit est structuré de la manière suivante.

Les spécificités du problème posé sont tout d'abord détaillées au premier chapitre où nous présentons différents types de données rencontrées et étudions les conséquences des irrégularités de l'échantillonnage pour l'analyse spectrale.

Le chapitre 2 présente des méthodes d'analyse spectrale couramment utilisées en astrophysique et spécialement conçues pour le traitement de données irrégulièrement échantillonnées. Nous décrivons notamment la généralisation du périodogramme ainsi que l'utilisation de méthodes de déconvolution de type CLEAN dont nous montrerons également les limites. Nous établissons alors le cadre formel de l'estimation dans lequel se placent les différentes approches que nous considérons par la suite, et présentons un signal artificiel qui sera utilisé systématiquement pour illustrer l'efficacité des algorithmes proposés et la qualité des estimateurs obtenus.

Nous étudions au chapitre 3 l'approche par pénalisation. Après avoir discuté du choix de la fonction de pénalisation, ce chapitre met l'accent sur l'aspect algorithmique de l'optimisation. Selon la forme de la pénalisation utilisée (nous considérons ici essentiellement des pénalisations $\ell^2\ell^1$ et ℓ^1), différentes solutions algorithmiques sont proposées et leur coût de calcul est comparé sur des données simulées.

L'approche par modélisation Bernoulli-Gaussienne fait l'objet du chapitre 4, dans lequel nous insistons sur l'intérêt de l'estimation au sens de la moyenne *a posteriori* pour le problème de la détection de raies spectrales. Différents algorithmes d'échantillonnage stochastique sont ensuite proposés et comparés sur des données simulées. La proposition d'un algorithme d'échantillonnage adéquat pour un modèle prenant en compte des décalages en fréquence, ainsi que la mise en évidence du gain obtenu sur la qualité de l'estimation, constituent la dernière partie de ce chapitre.

Le chapitre 5 est finalement dédié à l'application des différentes méthodes proposées au traitement de données réelles, pour lequel le problème d'estimation s'avère très délicat. La courbe de vitesse radiale de l'étoile de Herbig HD 104237 obtenue par Böhm *et al.* [2004] présente en effet une allure dominée par des composantes de basse fréquence qui perturbent l'estimation des raies recherchées à des fréquences plus élevées. Une étape de *simulation* de données similaires est préalablement effectuée, permettant de comparer les caractéristiques des différentes procédures d'estimation et notamment leur robustesse devant les perturbations à basse fréquence. L'application aux véritables données est ensuite présentée. Un prétraitement – souvent irréalisable en pratique puisqu'il a nécessité la fusion des données de deux campagnes d'observation à un an d'intervalle – a permis aux auteurs précités de rendre les données plus facilement exploitables par des techniques classiques et de détecter ainsi plusieurs raies spectrales. Nous comparons alors ces résultats aux estimations obtenues par les méthodes que nous avons proposées et appliquées aux données brutes, pour lesquelles aucune autre technique n'a permis d'obtenir des résultats intéressants.

Dans le dernier chapitre, nous effectuons une synthèse des résultats présentés dans ce manuscrit et proposons plusieurs pistes en perspective de ces travaux, aussi bien du point de vue méthodologique que pour le traitement de données astrophysiques.

Chapitre 1

L'estimation de spectres de raies dans un contexte astrophysique

Nous introduisons dans ce premier chapitre la problématique de l'analyse spectrale en astrophysique et les spécificités qui lui sont associées. Plus précisément, nous considérons le problème de la recherche de variations périodiques, voire sinusoïdales dans des séries temporelles décrivant par exemple l'évolution de la luminosité de l'objet au fil du temps (*courbes de lumière*), mesurée éventuellement dans plusieurs bandes de fréquences d'émission, ou encore la variation temporelle de la vitesse radiale d'un objet présentant des mouvements surfaciques ou bien se déplaçant dans la direction de l'observateur (*courbes de vitesse radiale*). Parmi les champs d'application concernés, deux disciplines retiendront particulièrement notre attention, pour lesquelles l'identification des variations représentent un enjeu majeur. Il s'agit de l'étude des étoiles variables et de celle des systèmes multiples. La description des types de données accessibles et de la nature des variations temporelles recherchées font l'objet du § 1.1, où nous présentons également le problème de leur estimation sous la forme commune de l'analyse spectrale. Deux caractéristiques majeures et propres à la nature astronomique des objets observés sont ensuite étudiées au § 1.2. Il s'agit dans un premier temps de l'espacement temporel des mesures, ou *échantillonnage*, qui présente généralement plusieurs types d'irrégularités. Nous verrons ensuite que la nature des perturbations associées aux observations confèrent également aux données astronomiques des caractéristiques spécifiques. Nous considérons enfin au § 1.3 les conséquences d'un échantillonnage irrégulier pour le problème d'analyse spectrale. Nous verrons notamment que des mesures irrégulièrement espacées présentent l'avantage d'élargir la plage fréquentielle exempte du phénomène de recouvrement, limitation bien connue à l'analyse spectrale dans le cas de l'échantillonnage régulier. L'effet de l'échantillonnage irrégulier sur le spectre de Fourier et le périodogramme est enfin étudié, et nous verrons notamment que les trous périodiques dans les observations, caractéristique très fréquente en astrophysique, provoquent l'inefficacité de ces outils de base.

1.1 Différents champs d'application

La description présentée ci-dessous provient en partie de l'ouvrage *Les Sciences du Ciel* [Léna, 1996], du cours en ligne du site public de la formation « Fenêtres sur l'Univers » de l'Observatoire de Paris[1] et surtout des nombreuses discussions menées avec les chercheurs du Laboratoire

1. Site web de la formation : `http://media4.obspm.fr/public/FSU/`

d'Astrophysique de l'Observatoire Midi-Pyrénées.

1.1.1 L'étude d'étoiles variables

Les *variables pulsantes* sont des étoiles dont le caractère oscillatoire est dû à l'équilibre métastable de leur structure. Dans une phase bien précise de leur évolution, l'équilibre hydrostatique des couches externes d'une étoile est compromis par une instabilité thermique : le gradient de température dans ces régions augmente, provoquant un fort gradient de pression sous l'effet duquel les couches externes de l'étoile se dilatent, jusqu'au point où les forces de gravité l'emportent sur les forces de pression, provoquant alors le mouvement inverse de contraction de ces couches. L'alternance de ces deux phénomènes provoque alors des variations périodiques dans la *courbe de lumière*.

Dans le cas de l'*astérosismologie*, les oscillations sont d'amplitude plus faible et de nature différente, liées aux ondes acoustiques générées par des perturbations de pression à l'intérieur d'une étoile. Celles-ci restent piégées à l'intérieur de la structure du fait des gradients de densité et de température, jouant le rôle de frontières sur lesquelles ces ondes se réfléchissent. À la différence du Soleil, situé à une distance relativement proche, la surface des disques stellaires n'est pas résolue spatialement (une étoile lointaine est perçue comme un objet ponctuel), et la détection de ces oscillations peut s'effectuer au moyen de l'analyse des variations dans la *courbe de lumière* de l'étoile. Une autre technique repose sur l'analyse des données spectroscopiques : selon que les couches externes se dilatent ou se contractent, les longueurs d'onde mesurées dans le spectre d'émission ou d'absorption du rayonnement de l'étoile seront décalées par effet Döppler-Fizeau vers les longueurs d'onde plus petites (respectivement plus grandes) selon la relation : $\frac{\Delta\lambda}{\lambda} = \sqrt{\frac{c+v}{c-v}} - 1 \sim \frac{v}{c}$ pour $v \ll c$, où v est la vitesse radiale du mouvement (*i.e.,* la projection de la vitesse le long de la ligne de visée) et c la vitesse de la lumière dans le vide. L'observation de décalages de ces raies variant de manière périodique traduit donc des mouvement périodiques des couches externes de l'étoile. Dans ce cas, les données seront donc des *courbes de vitesse radiale*, présentant des variations sinusoïdales correspondant aux modes d'oscillation de l'étoile.

1.1.2 Les systèmes multiples

Un système d'étoiles multiples est constitué de plusieurs étoiles suffisamment proches pour être liées gravitationnellement. On estime que plus de 50% des étoiles observées sont assemblées en systèmes binaires ou multiples. Dans le cas fréquent de systèmes à deux corps, les deux étoiles évoluent autour de leur centre de masse en une orbite régie par les lois de Képler. Leurs caractéristiques sont très variables ainsi que les méthodes mises en œuvre pour les détecter, puisque généralement les différents corps d'un même système ne sont pas *résolus*, c'est-à-dire que leur observation directe à l'instrument ne permet pas de les discerner. Nous considérons deux types de systèmes multiples, classés selon leur méthode d'observation [Léna, 1996].

- Les *binaires à éclipses*, ou *binaires photométriques*, sont des systèmes présentant des variations photométriques engendrées par l'occultation périodique d'une des composantes par son compagnon. Seuls les systèmes dont le plan orbital est faiblement incliné par rapport à la ligne de visée permettent d'observer de telles occultations. La détection de systèmes binaires et leur étude se font alors à partir de l'analyse des variations périodiques de la luminosité du système, *i.e.,* de sa *courbe de lumière*. Dans ce cas, l'identification de la

période du mouvement orbital permet d'en déterminer la forme par l'intermédiaire du *diagramme de phase*, qui est une représentation des données recalant toutes les observations dans un intervalle de temps ayant la taille de la période orbitale. Ce dernier permet alors de déterminer différents paramètres du système, parmi lesquels le rapport des masses des deux corps ou le demi-grand axe de l'orbite [Prieur, 2005].

– Pour les *binaires spectroscopiques*, les différents corps sont détectés au moyen d'une analyse spectroscopique. Le mouvement orbital, dans le cas où l'inclinaison de l'orbite sur le ciel n'est pas nulle, produit des variations périodiques de la vitesse radiale – la composante de la vitesse projetée le long de la ligne de visée – des deux étoiles par rapport à l'observateur, et ces variations peuvent être mesurées par spectroscopie. En effet, l'analyse de l'évolution, au fil du temps, du spectre d'émission (ou d'absorption) d'une étoile présente des décalages des raies spectrales, témoignant leur mouvement radial par effet Döppler-Fizeau : selon que l'étoile se rapproche ou s'éloigne de l'observateur, les raies spectrales sont décalées vers les longueurs d'onde plus grandes (respectivement plus petites), selon la loi déjà explicitée au paragraphe précédent : $\frac{\Delta\lambda}{\lambda} = \sqrt{\frac{c+v}{c-v}} - 1 \sim \frac{v}{c}$ pour $v \ll c$. L'observation de décalages de ces raies variant de manière périodique traduit donc un mouvement périodique de l'étoile. Dans ce cas, les données seront donc des *courbes de vitesse radiale*, permettant également la représentation sous forme d'un diagramme de phase, à partir duquel des paramètres du système peuvent être extraits, comme les demi-grand axes et les masses des deux corps, à une indétermination près due à l'inclinaison inconnue du plan orbital par rapport à la ligne de visée [Prieur, 2005]. On montre que dans le cas particulier où l'orbite est circulaire et vue par la tranche (c'est-à-dire lorsque le plan de l'orbite contient la ligne de visée), les courbes de vitesse radiale des deux corps sont sinusoïdales et oscillent en opposition de phase. Dans le cas plus général d'orbites elliptiques, les variations recherchées restent périodiques mais ne sont plus sinusoïdales.

Que ce soit dans des variations de luminosité ou de vitesse radiale, l'objectif partagé est ici la recherche d'une période contenue dans une série temporelle, le motif périodique n'étant généralement pas sinusoïdal. Dans le cas de systèmes multiples à plus de deux corps, il s'agira de manière analogue de rechercher plusieurs périodes de variation, même si le mouvement orbital du système ne peut plus être décrit de manière simple. Les périodes orbitales de tels systèmes varient, selon le type de l'objet, de quelques heures à plusieurs années. De ce fait, les mesures effectuées sont souvent éparses et ne couvrent parfois qu'une durée correspondant à quelques périodes.

Notons encore que ce principe est à la base de la détection indirecte de planètes extra-solaires, qui a constitué l'un des principaux événements astronomiques de la fin du XX$^{\text{ème}}$ siècle (la plupart des exoplanètes ont été découvertes à partir de courbes de vitesses radiales) : un système étoile-planète est de la même manière soumis aux lois gravitationnelles définissant les mouvements orbitaux. La différence entre les deux problèmes réside au niveau des amplitudes des variations périodiques à détecter dans la courbe de lumière ou de vitesse radiale de l'étoile, qui sont d'autant plus faibles que le rapport de masse entre les deux corps $M_{\text{étoile}}/M_{\text{planète}}$ est élevé.

1.1.3 Des problèmes d'analyse spectrale

L'astérosismologie et l'étude de systèmes multiples sont donc deux domaines pour lesquels la détermination précise des variations périodiques contenues dans des séries temporelles est essentielle. D'une manière générale, ces deux problèmes peuvent être vus sous le terme commun de *l'analyse spectrale*, où le terme de spectre est à entendre ici en tant que représentation duale,

dans l'espace fréquentiel, des variations temporelles des courbes de lumière ou de vitesse radiale. Pour l'astérosismologie, le problème est celui de la recherche d'un nombre inconnu d'oscillations sinusoïdales dans une certaine plage de fréquences, c'est-à-dire la recherche d'un *spectre de raies* dans le domaine de Fourier. Lors de l'étude de systèmes multiples, il s'agit de détecter un nombre restreint (typiquement un voire deux) de fluctuations périodiques non sinusoïdales, c'est-à-dire identifier les paramètres d'un signal *harmonique* (voire la somme de deux signaux harmoniques), pouvant se décomposer dans l'espace des fréquences en série de Fourier en composantes *fondamentale* et *harmoniques*. D'une manière générique, ces deux problèmes peuvent être perçus sous la forme de la recherche de raies dans l'espace de Fourier où, rappelons-le, une raie traduit une variation sinusoïdale dans la série temporelle. Nous insistons dès maintenant sur ce point afin d'éviter toute confusion entre le spectre de Fourier correspondant à des variations temporelles d'une grandeur observée et le spectre d'émission ou d'absorption, qui est relié à la nature du rayonnement du corps observé et présente également des *raies*.

1.2 Spécificités des données

Du point de vue du traitement du signal, le principe même de l'observation d'objets astronomiques confère au problème d'analyse spectrale des spécificités que nous détaillons ci-après. Il s'agit essentiellement du caractère irrégulier de l'échantillonnage et de la nature des perturbations rencontrées dans les mesures.

1.2.1 Une couverture temporelle incomplète

Lors d'observations astronomiques, il est généralement difficile d'obtenir une couverture observationnelle satisfaisante. En premier lieu, les temps d'observation disponibles sur les instruments de mesure sont limités et la monopolisation d'un télescope pour suivre un seul objet n'est souvent possible que pour une période de l'ordre de quelques nuits. Par conséquent, les données recueillies sont souvent pauvres, dans le sens où le nombre de points constituant la série temporelle peut être très faible. Ensuite, une contrainte majeure provient du fait que la plupart des objets observés ne sont visibles que de manière temporaire. La lumière du jour, ou des conditions météorologiques défavorables, rendent en effet de nombreuses mesures impossibles. De plus, certains objets ne se trouvent au-dessus de l'horizon terrestre que de manière saisonnière à cause des variations annuelles de la position de la terre. Les données recueillies présentent donc souvent un caractère incomplet, comportant des « trous » dans les périodes temporelles pour lesquelles aucune observation n'est possible. Par ailleurs, les observations réalisées sur de longues durées – pouvant atteindre parfois plusieurs dizaines d'années – sont souvent effectuées à des intervalles de temps irréguliers. Nous distinguerons donc par la suite deux types de perturbations des instants de mesure :

- le cas d'observations, à des instants réguliers ou non, pour lesquelles les mesures ne sont pas disponibles durant de larges périodes temporelles. Ces trous, généralement dus à des phénomènes périodiques comme l'alternance du jour et de la nuit ou celle des saisons, seront donc souvent également périodiques. Nous désignerons ce cas sous le terme de *données trouées*.
- le cas où les mesures sont réparties de manière totalement irrégulière sur la durée totale de l'observation. Nous parlerons alors d'*échantillonnage irrégulier*.

En général, les séries temporelles obtenues présentent conjointement ces deux sortes d'irrégularités, *i.e.,* un échantillonnage irrégulier sur lequel viennent s'ajouter des « trous » sans données disponibles. Il convient néanmoins de dissocier les deux facteurs, car leurs effets en terme de dégradation sur le spectre des données diffèrent : ce point sera l'objet du paragraphe 1.3. Un exemple typique de données à échantillonnage incomplet est représenté sur la figure 1.1, conjointement aux différences entre deux instants de mesure successifs. Il s'agit de mesures de vitesses radiales de l'étoile de Herbig HD 104237 pour cinq nuits d'observation [Böhm *et al.*, 2004], où l'alternance du jour et de la nuit provoque quatre gros « trous », et les mesures prises au cours d'une même nuit le sont à des intervalles de temps δt irréguliers.

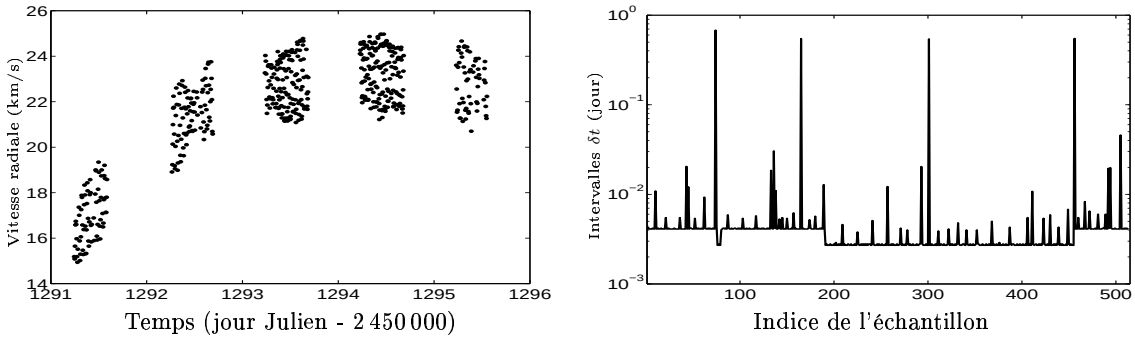
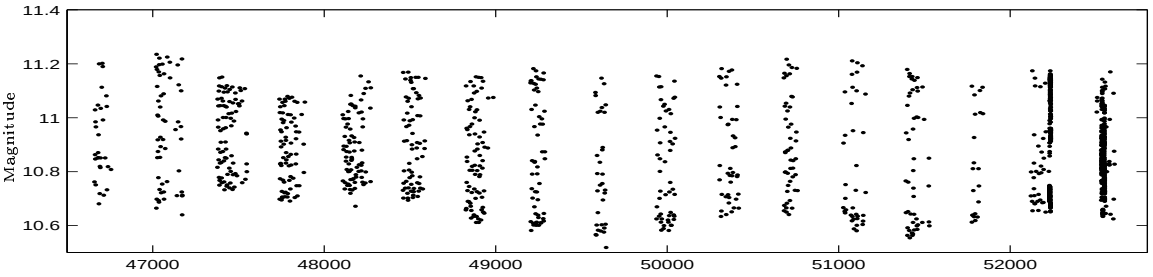
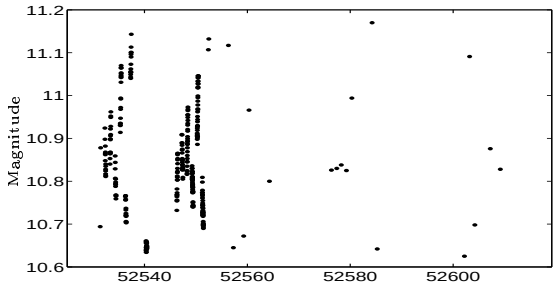


FIG. 1.1 – *Courbe de vitesse radiale (à gauche) et intervalles d'échantillonnage (à droite) pour 5 nuits d'observation de HD 104237.*

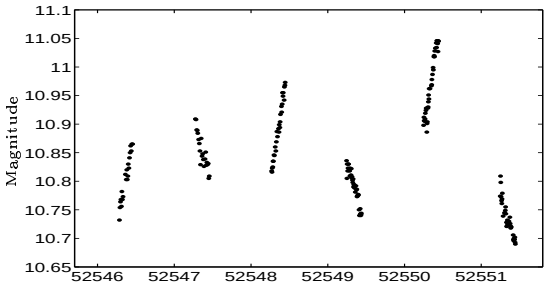
Un second exemple présentant un échantillonnage particulièrement complexe est représenté sur la figure 1.2. Il s'agit des variations de la luminosité de l'étoile T Tauri V410 Tau, dans la bande photométrique V, pendant 17 ans (une partie des observations a été publiée par Grankin [1999]). Ces données mettent en évidence plusieurs types d'irrégularités dans l'échantillonnage. La non visibilité de l'objet à certaines périodes de l'année provoque tout d'abord des trous de périodicité annuelle : les couvertures annuelles les plus étendues couvrent une période d'environ six mois (courbe (a)). De plus, les mesures effectuées au cours d'une même année le sont à des intervalles de temps très irréguliers, comme le montre la courbe (b). Enfin l'échantillonnage souffre, comme dans le cas précédent, d'une couverture journalière incomplète due à l'alternance du jour et de la nuit : la courbe (c) révèle ainsi, par tranches de 24 heures, des zones de mesure d'environ 5 heures alternées avec des trous d'environ 19 heures. On comprend alors aisément que ces différentes sortes d'irrégularités viendront perturber la recherche d'oscillations dans la courbe de lumière. L'interprétation spectrale du paragraphe 1.3 est consacrée à l'étude précise de ce point.



(a) Intégralité des mesures : 17 ans.



(b) Mesures au cours des 3 derniers mois.



(c) Mesures au cours de 6 nuits parmi les données (b).

FIG. 1.2 – *Courbe de lumière de V410 Tau à trois échelles de temps différentes [Grankin, 1999]. L'ensemble des données est constitué de 1548 points couvrant irrégulièrement une période de 17 ans.*

1.2.2 Nature des perturbations

Variabilité des caractéristiques observationnelles

Pour des observations couvrant de longues périodes temporelles, comme dans le cas des données présentées en figure 1.2, les conditions expérimentales ne sont naturellement pas identiques tout au long de la durée de mesure. En particulier, la fluctuation des conditions atmosphériques ou la variation à long terme des caractéristiques des instruments perturbent la qualité des mesures. Cette variabilité est alors la source de non stationnarités dans les données et chaque observation, ou chaque série regroupant des observations sur une courte période, peut se voir affecter de perturbations de natures et de niveaux variables.

Le cas particulier des données issues d'observations multi-sites présente des caractéristiques similaires. Afin de réduire les trous dans la couverture temporelle, ou tout simplement dans le but d'obtenir plus de données, les campagnes d'observation multi-sites ont pour but de fusionner les observations d'un même objet réalisées par différents instruments. Le réseau de téléscopes WET[2], pour *Whole Eath Telescope* [Nather *et al.*, 1990], est ainsi composé d'instruments répartis sur toute la surface du globe (actuellement, un total de 25 observatoires participent aux campagnes), visant ainsi à obtenir des séries de mesures ininterrompues pour l'étude d'étoiles variables. La figure 1.3 présente la répartition de la couverture temporelle obtenue entre 9 téléscopes lors d'une des premières campagnes d'observation WET en 1989. Les conditions propres à chaque

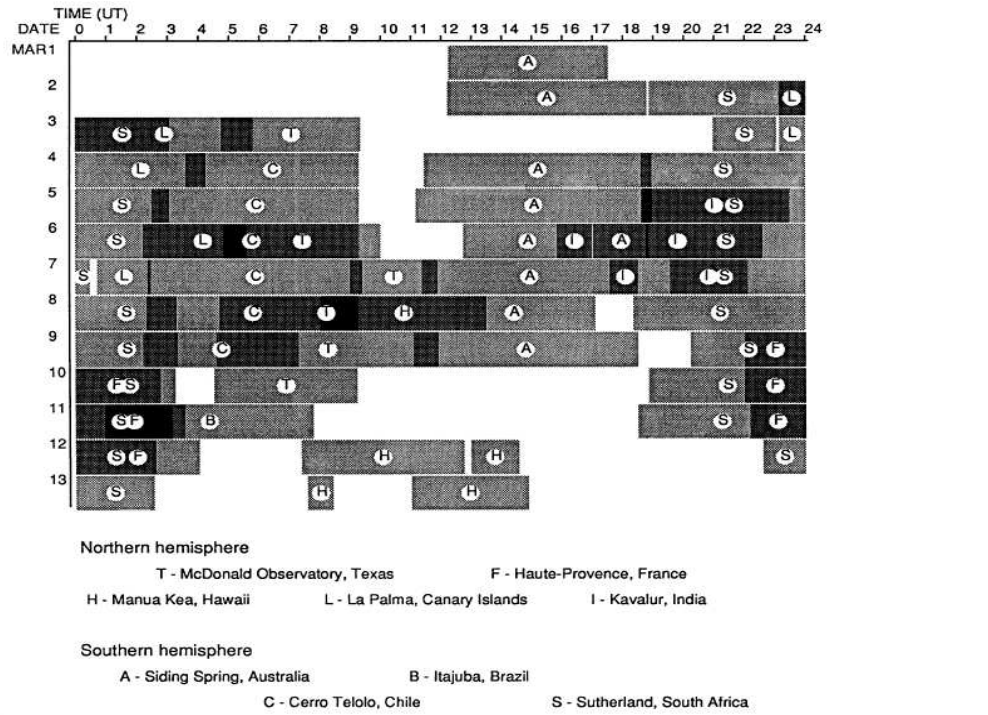


FIG. 4.—Graphical representation of the coverage obtained by the Whole Earth Telescope in 1989 March on the white dwarf PG 1159 – 035

FIG. 1.3 – *Exemple de campagne d'observation WET mettant en jeu 9 sites d'observation, tiré de [Nather et al., 1990].*

2. Site internet du projet WET : `http://wet.physics.iastate.edu`

site d'observation d'une part, et les spécificités de chaque instrument d'autre part, confèrent alors des caractéristiques différentes à chaque contribution produisant la série temporelle fusionnée, qui peut par conséquent souffrir d'une certaine hétérogénéité.

Présence de basses fréquences

Un second type de perturbations rencontrées concerne des données comme la courbe de vitesse radiale présentée sur la figure 1.1. L'étude de l'étoile HD 104237 Böhm *et al.* [2004] montrent qu'il s'agit d'un système multiple complexe incluant une binaire spectroscopique, induisant un mouvement orbital périodique d'une période de l'ordre de 20 jours, et dont le premier composant est une étoile variable pulsante, avec des oscillations de fréquences bien plus élevées entre 28.5 et 35.6 j^{-1}. Lorsque l'on s'intéresse à l'étude de la variabilité, les éléments orbitaux du système, d'amplitude dominant celle des oscillations, génèrent des variations à basse fréquence dans la courbe de vitesse radiale nettement visibles sur la figure 1.1, qui viennent perturber la recherche des oscillations à plus haute fréquence. Le couplage de deux phénomènes à l'origine de périodicités, qui nous intéressent par ailleurs de manière séparée, est donc générateur de complications pour l'étude des variations présentes dans les données. Dans le cas du système multiple HD 104237, Böhm *et al.* [2004] ont recours à une approximation binaire pour décrire le mouvement orbital global du système, ce qui permet d'extraire une partie des basses fréquences[3]. Les données ainsi corrigées sont représentées sur la figure 1.4 conjointement aux données initiales, par rapport auxquelles la plupart du comportement à basse fréquence a effectivement été supprimée. Néanmoins, cette correction peut se révéler délicate à maîtriser et l'approximation inexacte des mouvements orbitaux peut induire des erreurs sur la courbe résultante. Comme nous l'avons vu au paragraphe 1.1, on considère qu'environ une étoile sur deux est multiple. Ce type de perturbation est donc fréquemment rencontré.

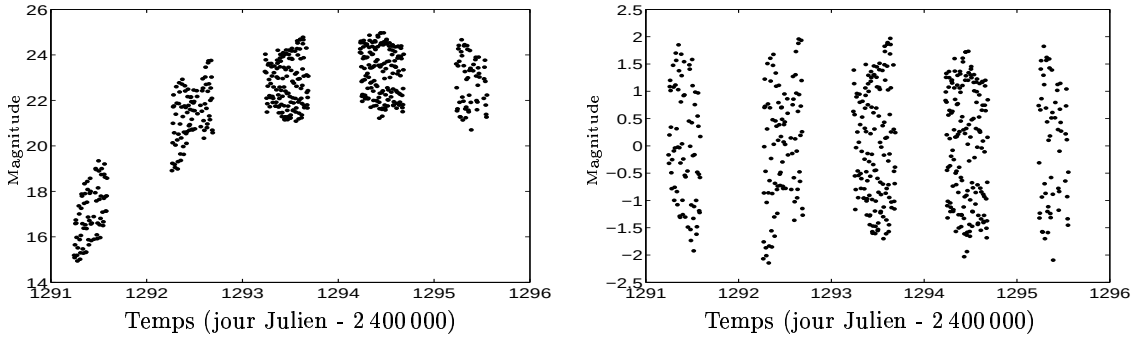


FIG. 1.4 – *Courbe de vitesse radiale de HD 104237: données brutes (a) et après extraction du mouvement orbital par approximation binaire (b).*

1.3 Conséquences pour l'analyse spectrale

L'influence d'une couverture observationnelle imparfaite sur la recherche de périodicités dans des séries temporelles telles que celles présentées au paragraphe 1.1 peut être décrite dans l'espace de Fourier. Il est bien connu que la restriction de l'observation à un nombre fini de données

3. Notons que pour ce faire les auteurs ont recours à la fusion de ces données avec un second jeu de données issu d'observations effectuées au cours de l'année 2000.

échantillonnées provoque, au niveau du spectre de Fourier, la *convolution* du spectre théorique par la réponse en fréquence de la fenêtre d'observation. De manière formelle, considérons une fonction $y(t)$ représentant l'évolution d'une grandeur physique au cours du temps, et $Y(f)$ sa transformée de Fourier :

$$Y(f) = \mathcal{TF}\{y(t)\} \triangleq \int_{t=-\infty}^{+\infty} y(t)\, e^{-j2\pi ft} dt.$$

Par la suite, le terme de *spectre* sera employé pour désigner la transformée de Fourier $Y(f)$ de la fonction $y(t)$. La grandeur y est mesurée à des instants $t_{n,\, n=1\ldots N}$. En considérant les observations parfaites, c'est-à-dire en l'absence de bruit, les données disponibles correspondent alors au signal échantillonné :

$$y_e(t) = y(t) \times w_e(t), \text{ avec } w_e(t) \triangleq \sum_{n=1}^{N} \delta(t - t_n)$$

où $\delta(t)$ est la fonction de Dirac[4]. Le terme $w_e(t)$ sera par la suite désigné comme la *fenêtre d'observation*. Le spectre de Fourier des données peut alors s'écrire comme la convolution :

$$Y_e(f) = Y(f) \star W_e(f), \text{ avec } W_e(f) = \mathcal{TF}\{w_e(t)\} = \sum_{n=1}^{N} e^{j2\pi f t_n}. \tag{1.1}$$

$Y_e(f)$ sera appelé le *spectre observé* ou *spectre des données* et la transformée de Fourier $W_e(f)$ de la fenêtre d'observation $w_e(t)$ sera référencée comme la *fenêtre spectrale*.

Dans le cas fréquent en traitement du signal où l'échantillonnage est effectué de manière régulière à la période T_e – soit $t_n = n\,T_e, n \in \mathbb{Z}$ –, la fenêtre spectrale s'écrit :

$$W_e^{\text{reg}}(f) = e^{-j\pi f(N+1)T_e} \frac{\sin \pi f N T_e}{\sin \pi f T_e}$$

et est périodique, de période $1/T_e$. Deux conséquences majeures sont alors à noter :

- le spectre observé $Y_e(f)$ est également périodique, de période $1/T_e$. De ce fait, aucune composante spectrale ne peut être retrouvée sans ambiguïté à partir des données au-delà de la limite de Nyquist $f_{\text{Nyquist}} \triangleq 1/2T_e$: c'est une version du théorème de Shannon ;
- le motif périodique de la fenêtre spectrale a une allure proche de celle d'une fonction sinus cardinal. En particulier, cette fenêtre a un lobe principal de largeur $2/NT_e$ et une atténuation du premier lobe secondaire par rapport au lobe principal de l'ordre de 13 dB [Kunt, 1984].

Dans le cas de schémas d'échantillonnage irrégulier, de telles propriétés ne sont plus valables. Nous étudions donc le problème de la définition d'une fréquence de Nyquist et les conséquences, sur le spectre observé, d'une fenêtre spectrale plus complexe.

1.3.1 Définition d'une fréquence limite de Nyquist

Lorsque les données sont échantillonnées régulièrement à la fréquence $f_e = 1/T_e$, la périodicité de la fenêtre spectrale, et donc celle du spectre observé, rend ambiguë la détection de

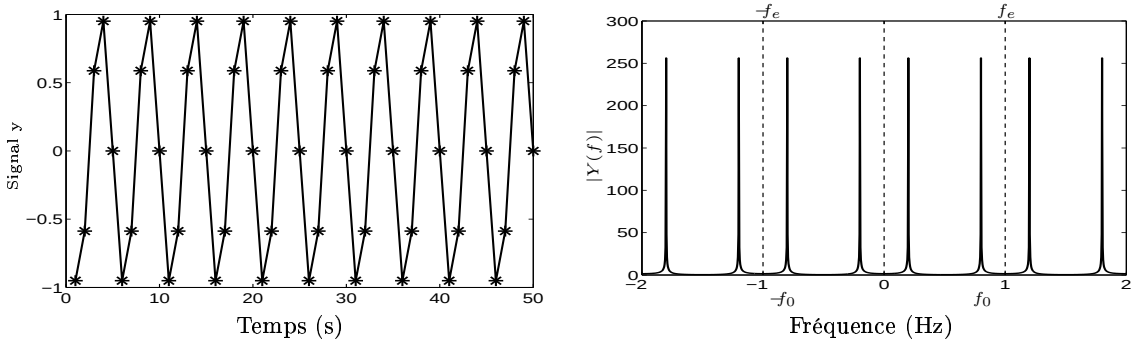
4. La « fonction » δ de Dirac est proprement définie au sens des distributions, comme la dérivée de la fonction échelon valant 0 pour $t < 0$ et 1 pour $t \geq 0$.

fréquences supérieures à $f_{\text{Nyquist}} = f_e/2$: c'est le phénomène de recouvrement, ou *aliasing*. Pour un échantillonnage irrégulier, la fenêtre spectrale $W_e(f)$ perd sa périodicité et la notion de recouvrement n'a plus lieu d'être. En ce sens, le caractère irrégulier est un avantage, puisqu'il autorise la recherche de fréquences sur une plage de valeurs beaucoup plus large en éliminant les alias, ou au moins en les repoussant à des fréquences beaucoup plus éloignées. Pour un nombre donné de mesures, le cas où elles sont régulièrement espacées est même celui qui produit la plus petite plage de fréquences reconstructibles [Eyer et Bartholdi, 1999].

Prenons par exemple le cas d'un signal sinusoïdal à la fréquence f_0. Son spectre théorique est constitué de deux raies aux fréquences $\pm f_0$:

$$y(t) = \cos 2\pi f_0 t \quad \longleftrightarrow \quad Y(f) = \frac{1}{2}(\delta(f - f_0) + \delta(f + f_0)).$$

Supposons ce signal échantillonné à une fréquence $f_e < 2f_0$, en dessous de la fréquence minimale imposée par le théorème de Shannon. Le spectre observé a l'allure représentée sur la figure 1.5. Il fait apparaître plusieurs pics de même hauteur dans l'intervalle de fréquences $[0, 2f_e]$, en $f_e - f_0$, f_0, $2f_e - f_0$ et $f_e + f_0$, et l'unique pic de fréquence inférieure à $f_{\text{Nyquist}} = f_e/2$ se trouve en $f_e - f_0$ et est un alias de la fréquence négative $-f_0$.



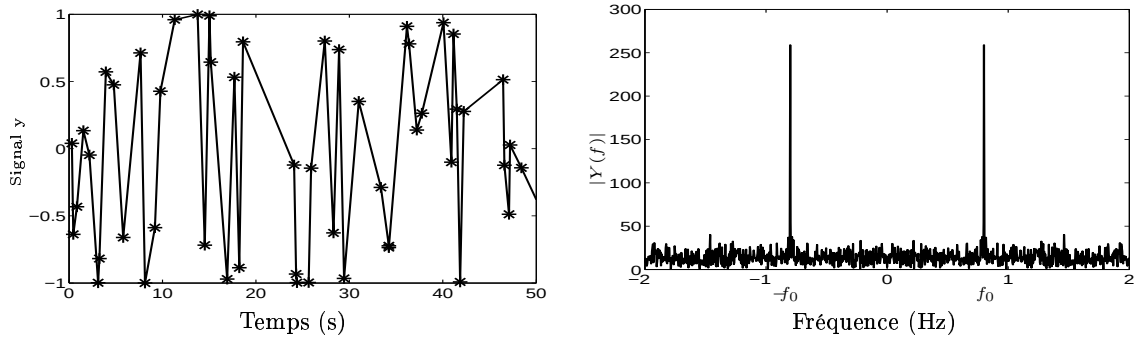
Une partie du signal — Spectre de Fourier

FIG. 1.5 – *À gauche, signal sinusoïdal échantillonné à $f_e = 1$ (512 points) et de fréquence au-delà de la limite de Nyquist ($f_0 = 0.8$). À droite, spectre de Fourier correspondant.*

Considérons à présent la même sinusoïde échantillonnée en le même nombre de points, mais qui sont cette fois répartis aléatoirement de manière uniforme sur un intervalle de la même durée : un seul pic est maintenant présent dans le spectre de Fourier, à la fréquence f_0, comme le montre la figure 1.6. Notons néanmoins que le spectre comporte de plus fortes perturbations sur l'ensemble des fréquences que dans le cas régulier ; nous y reviendrons de manière détaillée au paragraphe 1.3.2. Le caractère irrégulier de l'échantillonnage, perçu par ailleurs comme une contrainte, permet donc d'étendre la plage de fréquences exempte d'alias, rendant ainsi possible la détection de très courtes périodes d'oscillation [Eyer et Bartholdi, 1999]. Comment définir alors une plage de fréquences $[0, f_{\max}]$ exempte d'alias pour des instants d'échantillonnage $\{t_n\}_{n=1\ldots N}$ irrégulièrement espacés ?

L'origine des alias étant la périodicité de la fenêtre spectrale, Bretthorst [1988] recherche la plus petite période de cette fenêtre. Il définit ainsi un *temps critique* τ, éventuellement très petit, comme le plus grand diviseur de tous les instants d'échantillonnage[5] : on a alors $t_n = i_n\tau, i_n \in \mathbb{N}$,

5. Quitte à changer l'origine des temps, nous supposons ici l'échantillonnage tel que $t_1 = 0$.



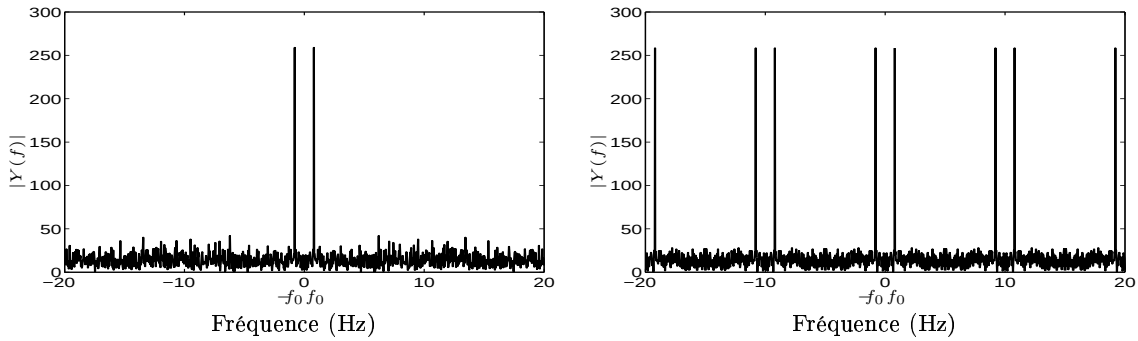
Une partie du signal — Spectre de Fourier

FIG. 1.6 – *À gauche, signal sinusoïdal avec un échantillonnage irrégulier de distribution uniforme (512 points) et de fréquence* $f_0 = 0.8$. *Spectre de Fourier correspondant.*

ce qui revient à considérer que les données proviennent d'un schéma d'échantillonnage régulier, de période τ, et pour lequel la plupart des données sont manquantes. La période de la fenêtre spectrale est alors de $1/\tau$ – par définition de τ, il ne peut y avoir de période plus courte – et les auteurs proposent de définir la fréquence de Nyquist comme :

$$f_{\text{Nyquist}} \stackrel{\Delta}{=} 1/2\tau, \text{ où } \tau \text{ est la plus grande valeur telle que } \forall n,\ t_n \text{ s'écrit } t_n = i_n\tau, n \in \mathbb{N}.$$

Lorsque l'échantillonnage est totalement irrégulier, rien n'impose aux instants t_n d'être multiples d'une même durée τ ; la valeur du temps critique est alors de l'ordre de la précision du codage des instants d'échantillonnage. Par exemple, si les t_n sont définis (en secondes) avec quatre décimales, on aura τ de l'ordre de 10^{-4} sec, soit une fréquence de Nyquist de 5 000 Hz ; le passage à une précision de 10 décimales correspondra, au moins sur le plan théorique, à une fréquence limite de 5 GHz! Considérons à nouveau la sinusoïde de l'exemple de la figure 1.6 (les instants d'échantillonnage étant représentés avec 16 décimales), pour laquelle les instants t_n ont été arrondis à 10^{-1} près : on a alors $\tau = 0.1$ sec. La figure 1.7 compare les deux spectres sur une bande de fréquences de 0 à 20 Hz. Dans le premier cas, aucun alias n'est présent alors que pour le second, l'effet de la périodisation à $1/\tau = 10$ Hz apparaît.



t_n avec 16 décimales — t_n arrondis à 1 décimale

FIG. 1.7 – *Spectres de Fourier d'une sinusoïde échantillonnée irrégulièrement, où les instants d'échantillonnage* t_n *en secondes sont codés avec une précision différente.*

Dans le cadre astrophysique qui nous intéresse, la précision des instants t_n sera généralement

suffisante pour éviter tout phénomène d'aliasing. Néanmoins, la recherche de fréquences jusqu'à la limite de Nyquist au sens défini ci-dessus sera généralement inutile : d'une part, des oscillations à haute fréquence n'auront pas forcément de sens au niveau de la physique sous-jacente, et d'autre part la couverture d'un large domaine fréquentiel risque d'augmenter de manière déraisonnée la complexité de mise en œuvre des algorithmes d'analyse spectrale. En pratique, il sera donc plus raisonnable de fixer la fréquence maximale de recherche en accord avec une certaine connaisance physique *a priori*, en s'assurant que la fenêtre spectrale ne présente pas de périodicité dans la bande considérée. Pour terminer ce paragraphe, notons encore quelques remarques au sujet du choix de la fréquence maximale.

1. Cette définition du temps critique τ sous-entend également que l'ajout, à des données régulièrement espacées ($t_n = nT_e$), d'une seule mesure à un instant t' non multiple de T_e suffit à augmenter la fréquence maximale sans alias. Bretthorst [2001] analyse ce point et montre que le spectre obtenu présente des répliques aux fréquences multiples de $1/T_e$, mais dont l'amplitude est d'autant plus atténuée par rapport au motif principal que la proportion de points « irrégulièrement espacés » augmente.
2. D'autres définitions plus « pessimistes » de la fréquence de Nyquist pour l'échantillonnage irrégulier sont parfois rencontrées. En particulier, Lenz et Breger [2005] considèrent le temps critique $\overline{\tau}$ comme la durée moyenne entre deux instants consécutifs $\delta t_n = t_n - t_{n-1}$[6]. Bretthorst [2001] envisage encore de définir le temps critique comme la plus petite durée entre deux instants consécutifs, ce qui débouche sur une plus grande fréquence de Nyquist que dans le cas précédent. L'auteur montre néanmoins que, en pratique, ce choix (et donc *a fortiori* celui de $\overline{\tau}$) sous-estime encore largement la fréquence maximale prévenant l'apparition d'alias.
3. Le cas d'un échantillonnage régulier (à la fréquence T_e) mais avec des *données trouées*, introduit au paragraphe 1.2.1, peut également s'interpréter de la manière ci-dessus en considérant $\tau = T_e$. En termes d'aliasing, on obtient donc des conclusions similaires au cas de l'échantillonnage régulier, c'est-à-dire une limitation de la plage fréquentielle utile à l'intervalle $[0,1/2T_e]$. L'influence des trous se caractérise alors par une forme différente du motif périodique de la fenêtre spectrale. Ce point est l'objet du paragraphe suivant.

1.3.2 Une fenêtre spectrale complexe

Outre la perte de la périodicité de la fenêtre spectrale, le passage à un échantillonnage non régulier confère par ailleurs à celle-ci une forme plus complexe. Nous distinguons ici les effets produits par les deux sources d'irrégularités, à savoir l'absence périodique de données au cours de la période d'observation et l'espacement irrégulier des instants d'échantillonnage.

i) Conséquences de l'absence périodique de données

La présence de trous dans les données provoque l'apparition de lobes secondaires, qui sont d'autant plus marqués si les trous sont périodiques. En effet, un échantillonnage troué peut être vu comme un fenêtrage des données par une fenêtre indicatrice continue $w_i(t)$ à valeurs binaires, telle que :

6. Plus exactement, les auteurs proposent de considérer l'espacement moyen entre deux instants d'échantilonnage sans prendre en compte les différences δt_n élevées correspondant aux indices des « trous » éventuels.

- $w_i(t) = 1$ lorsque t est dans une période d'observation (par exemple : la nuit),
- $w_i(t) = 0$ pour les zones « trouées » (par exemple : le jour).

La fenêtre d'observation peut alors s'écrire :

$$w_e(t) = \sum_{n=1}^{N} \delta(t - t_n) \times w_i(t) \tag{1.2}$$

où la fenêtre $w_i(t)$ reflète les effets des phénomènes occultant la visibilité de l'objet étudié, comme l'alternance du jour et de la nuit ou les non visibilités saisonnières étudiées au paragraphe 1.2.1. Ces phénomènes étant périodiques, la fenêtre le sera également. Dans le domaine fréquentiel, la relation (1.2) s'écrit :

$$W_e(f) = \sum_{n=1}^{N} e^{j2\pi f t_n} \star W_i(f)$$

et l'influence des trous sur la fenêtre spectrale se traduit sous la forme d'une *convolution* par le motif $W_i(f)$. À titre d'exemple considérons, pour le cas de l'alternance jour/nuit, des périodes d'observation de $T_0 = 8$ heures espacées toutes les $T = 24$ heures. La fenêtre associée $w_i(t)$ a l'allure représentée sur la figure 1.8, et se décompose en une somme de fenêtres rectangulaires :

$$w_i(t) = \sum_{k=0}^{K} \mathbb{1}_{[0,\,T_0]}(t - kT)$$

où $\mathbb{1}_{[0,\,T_0]}(t)$ est la fonction porte valant 1 pour $t \in [0, T_0]$ et 0 ailleurs, de transformée de Fourier $\Gamma_{T_0}(f) = \Gamma^0 \operatorname{sinc}(fT_0)$ avec $\Gamma^0 = T_0\, e^{-j\pi fT}$ et $\operatorname{sinc}(u) \triangleq \sin(\pi u)/\pi u$. Le motif spectral correspondant à $w_i(t)$ s'écrit alors :

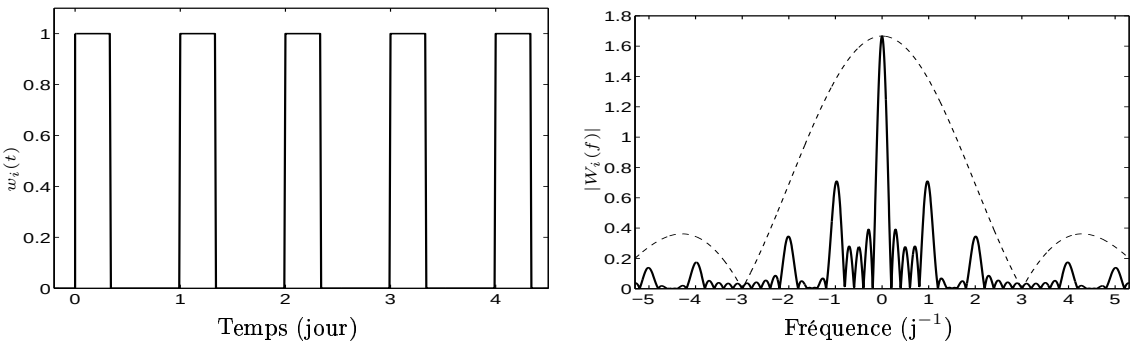
$$\begin{aligned} W_i(f) &= \sum_{k=0}^{K} \Gamma^0 \operatorname{sinc}(fT_0)\, e^{-j2\pi fkT} \\ &= e^{-j\pi(K-1)fT}\, \Gamma^0 \operatorname{sinc}(fT_0)\, \frac{\sin\left(\pi f(K+1)T\right)}{\sin(\pi fT)} \end{aligned}$$

et est constitué d'un motif périodique, le dernier terme, de période $1/T$, apodisé par un sinus cardinal de pseudo-période $1/T_0 > 1/T$. Le module $|W_i(f)|$ est représenté sur la figure 1.8, qui fait apparaître des lobes secondaires assez élevés pour des fréquences multiples de $1\ \mathrm{j}^{-1}$. Remarquons au passage que l'apodisation par $\Gamma_{T_0}(f)$, de pseudo-période $1/T_0$, provoque la disparition du lobe en $1/T_0 = 3\ \mathrm{j}^{-1}$.

En conclusion, la présence de trous périodiques dans le schéma d'échantillonnage génère des lobes secondaires dans le motif $W_i(f)$, à des fréquences multiples de la fréquence des trous. Ces lobes seront d'autant plus élevés que la couverture observationnelle est trouée, *i.e.*, le rapport T_0/T est faible. La fenêtre spectrale $W_e(f)$, qui s'écrit comme la convolution du terme $\sum_{n=1}^{N} e^{j2\pi f t_n}$ par le motif spectral $W_i(f)$, présentera donc également des lobes secondaires aux fréquences multiples de la fréquence des trous.

ii) Effet de l'échantillonnage irrégulier

Nous avons vu au début du paragraphe 1.3 que la régularité de l'échantillonnage se traduit par une certaine régularité de la fenêtre spectrale, permettant de quantifier son effet (allure de sinus



Fenêtre indicatrice — Motif spectral

FIG. 1.8 – *Fenêtre indicatrice (à gauche) et motif spectral associé (à droite) pour des observations périodiques de 8 heures par jour et sur une durée de 5 jours. Les pointillés représentent le terme* $|\Gamma_{T_0}(f)|$.

cardinal, largeur du lobe principal, atténuation des lobes secondaires). Dans le cas de données irrégulièrement espacées, l'expression de $W_e(f) = \sum_{n=1}^{N} e^{j2\pi f t_n}$ ne revêtant pas de forme plus simple, peu de propriétés exactes peuvent être établies, si ce n'est la conservation de la symétrie hermitienne ($W_e(-f) = W_e(f)^*$) et le maximum de $|W_e(f)|$ en 0.

Nous pouvons néanmoins effectuer quelques considérations qualitatives en écrivant les instants $\{t_n\}_{n=1\ldots N}$, répartis de manière irrégulière sur l'intervalle $[0, T]$, sous la forme $t_n = n\overline{t} + \epsilon_n$, où $\overline{t} = T/N$ est une sorte de « période moyenne » et $\{\epsilon_n\}_n$ représente l'écart du schéma d'échantillonnage à un schéma régulier de période $\overline{t}$: dans le cas d'un échantillonnage régulier, $\overline{t}$ est la période d'échantillonnage et $\epsilon_n = 0$. Si les irrégularités ne sont pas trop marquées (en particulier, nous supposons ici que l'échantillonnage ne comporte pas de gros trous), les valeurs de ϵ_n peuvent être supposées de l'ordre de grandeur de $\overline{t}$. La fenêtre spectrale se décompose alors en :

$$W_e(f) = \sum_{n=1}^{N} e^{j\overbrace{2\pi f \epsilon_n}^{\phi_n^\epsilon}} e^{j\overbrace{2\pi f n\overline{t}}^{\overline{\phi}_n}}$$

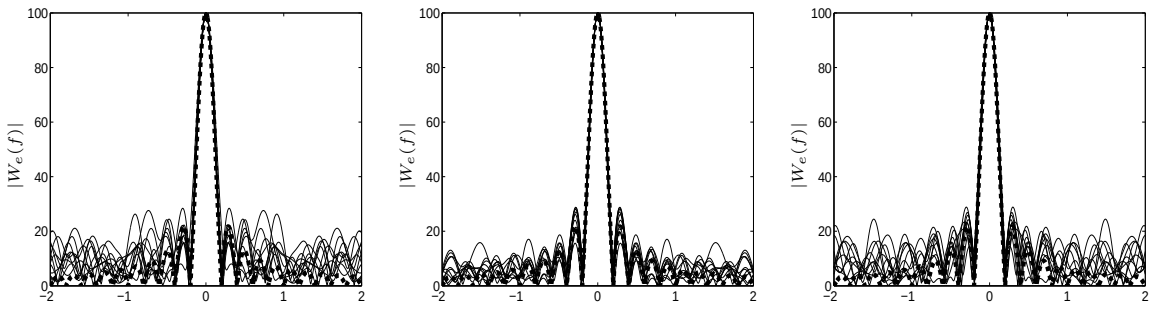
et les perturbations dues à l'irrégularité de l'échantillonnage s'expriment à travers le terme de pondération $e^{j\phi_n^\epsilon}$. On peut alors montrer que ce terme est de peu d'influence sur $W_e(f)$ pour les fréquences proches de 0. En effet, considérons $f \in [0, 1/(N\overline{t})]$, intervalle qui correspond au premier lobe de W_e dans le cas de l'échantillonnage régulier. On a :

$$f = \frac{\alpha}{N\overline{t}},\ 0 \le \alpha \le 1 : \left\{ \begin{array}{lll} \overline{\phi}_n = & 2\pi\alpha\frac{n}{N} & \text{varie de 0 à } 2\pi\alpha \text{ pour n=1 ...N} \\ \phi_n^\epsilon = & 2\pi\alpha\frac{\epsilon_n}{N\overline{t}} \sim \frac{2\pi\alpha}{N} & \text{varie de 0 à } \frac{2\pi\alpha}{N} \text{ pour n=1 ...N} \end{array} \right.$$

Par conséquent, dans la mesure où le nombre de données N est suffisamment grand et les irrégularités de l'échantillonnage « régulières » (permettant d'avoir $\epsilon_n \sim \overline{t}$), le rapport des angles $\phi_n^\epsilon/\overline{\phi}_n$ est de l'ordre de $1/n$. On s'attend donc à ce que la pondération par $e^{j\phi_n^\epsilon}$ soit d'influence plutôt faible sur la fenêtre spectrale W_e, et donc que celle-ci présente un comportement à basse fréquence similaire à celui d'une fenêtre correspondant à un échantillonnage régulier de période $\overline{t}$. En particulier, les mêmes caractéristiques sont attendues en ce qui concerne la forme du lobe principal. Ce n'est par contre plus le cas lorsque l'on s'éloigne vers des fréquences plus élevées, où pour toute valeur de n, l'angle ϕ_n^ϵ prend des valeurs qui ne sont plus négligeables devant $\overline{\phi}_n$, et aucune propriété de décroissance de $|W_e(f)|$ ne peut être établie de manière similaire au cas régulier,

pour lequel $|We(f)|$ décroît en $1/f$. Les simulations présentées sur la figure 1.9 confirment cette conjecture, en présentant l'allure de la fenêtre spectrale pour différentes réalisations d'échantillonnage irrégulier :

- t_n répartis uniformément sur $[0, T]$,
- $\delta t_n = t_n - t_{n-1}$ répartis uniformément sur $[0, \overline{t}]$,
- t_n réalisation d'un processus de Poisson [7] de paramètre $h = \overline{t}^{-1}$.



(a) t_n uniformes (b) δt_n uniformes (c) t_n Poissoniens

FIG. 1.9 – *Fenêtres spectrales correspondant à 10 réalisations de différents types d'échantillonnage irrégulier (trait plein), avec* $N = 100$ *points couvrant une période de* $T = 5$ *jours. En pointillé, fenêtre spectrale pour un échantillonnage régulier à la période équivalente* $\overline{t} = 0.05$ *jour : son lobe secondaire est alors de largeur* $2/N\overline{t} = 2/T = 0.4$ j^{-1}.

Dans chaque cas, le lobe principal a une largeur approximative de $2/T$, et l'atténuation du premier lobe secondaire est comparable à celle du cas régulier, à savoir 13 dB. Pour les plus hautes fréquences, la fenêtre présente un comportement plus imprévisible, où notamment l'atténuation des lobes secondaires semble rester à un niveau constant lorsque f augmente alors que, dans le cas de l'échantillonnage régulier, celle-ci décroît en $1/F$. Notons que la thèse de doctorat de Martin [1998], qui considère des schémas d'échantillonnage *aléatoires*, étudie de manière probabiliste la forme de la fenêtre spectrale en fonction de la distribution de probabilité des instants t_n. Des conclusions similaires sont obtenues, notamment en ce qui concerne l'atténuation moins nette et tendant vers un niveau constant lorsque l'on s'éloigne de $f = 0$, dans le cas de différences δt_n suivant une loi uniforme (figure (b)) ou pour un échantillonnage poissonien (figure (c)).

iii) Conclusion

Lorsque s'ajoutent des trous périodiques à un schéma d'échantillonnage irrégulier, il faut inclure à la forme de la fenêtre spectrale analysée ci-dessus l'effet convolutif par le motif $W_i(f)$ étudié au paragraphe précédent. La fenêtre résultante voit donc se superposer les deux effets, à savoir :

- l'apparition de lobes secondaires aux fréquences multiples de celle des trous,
- par rapport au cas régulier, des oscillations similaires autour de ces lobes, et une atténuation moins marquée sur l'ensemble des fréquences.

7. Un tel processus correspond à $\delta t_n = -\overline{t} \ln u_n$, où u_n suit une répartition uniforme sur [0,1].

La figure 1.10 résume l'effet, sur la fenêtre spectrale, des trous dus à l'alternance du jour et de la nuit (5 nuits d'observations de 8 heures chacune) et de l'échantillonnage irrégulier. Dans cet exemple, les instants d'échantillonnage irréguliers sont choisis uniformément répartis dans la zone d'observation. Remarquons encore que, dans le cas de l'échantillonnage à la fois irrégulier et troué de la figure 1.10 (d), le lobe principal de la fenêtre est de largeur similaire à celle du lobe principal pour l'échantillonnage régulier troué (figure 1.10 (b)), à savoir $2/T = 0.2$ j^{-1}. Ainsi, même si l'échantillonnage irrégulier présente des trous, le lobe principal de la fenêtre spectrale a encore une largeur de l'ordre de $2/T$, où T est la durée totale couverte par les observations.

1.3.3 Conséquences sur le périodogramme

Considérons une fonction $y(t)$, de transformée de Fourier $Y(f)$. Les données échantillonnées s'écrivent $\{y(t_n)\}_{n=1...N}$ et ont un spectre de Fourier noté $Y_e(f)$. L'outil le plus naturel pour l'analyse spectrale est sans doute le périodogramme, défini comme :

$$\begin{aligned} P_Y(f) &= \frac{1}{N}\left|\sum_{n=1}^{N} y(t_n)e^{-j2\pi f t_n}\right|^2 \\ &= \frac{1}{N}|Y_e(f)|^2. \end{aligned} \tag{1.3}$$

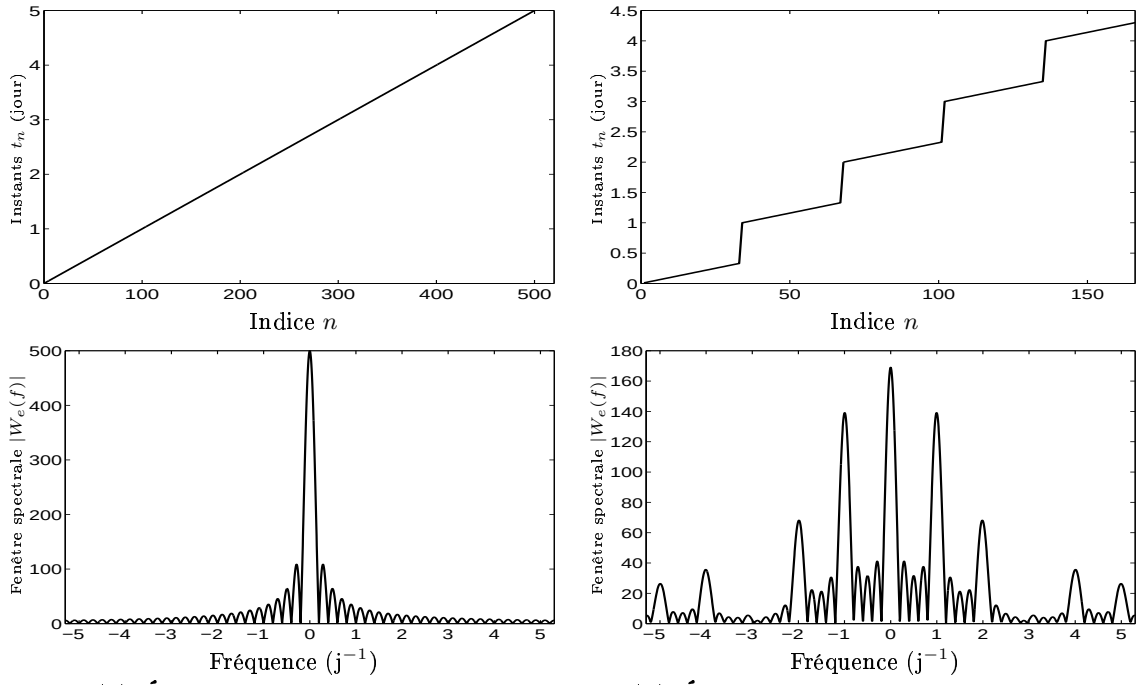
Pour des données régulièrement échantillonnées, on peut montrer [Kunt, 1984] que $P_Y(f)$ constitue un estimateur de la densité spectrale de puissance $S_y(f) = |Y(f)|^2$ qui est :

- biaisé : $\begin{aligned}\mathbb{E}\{P_Y(f)\} &= S_y(f) \star W_{\text{rect}}(f) \\ &\neq S_y(f)\end{aligned}$
  où $W_{\text{rect}}(f)$ est la transformée de Fourier de la fenêtre triangulaire ;
- asymptotiquement non biaisé : $\lim_{N\to+\infty} \mathbb{E}\{P_Y(f)\} = S_y(f)$;
- non consistant : $\lim_{N\to+\infty} \text{var}\{P_Y(f)\} \neq 0$.

Le biais et la non consistance de cet estimateur en font donc en théorie un outil particulièrement délicat dans le cadre général de l'analyse spectrale. Néanmoins, une justification de son utilisation est fournie dans le cadre statistique pour des données contenant une seule fréquence ; nous y reviendrons au chapitre suivant.

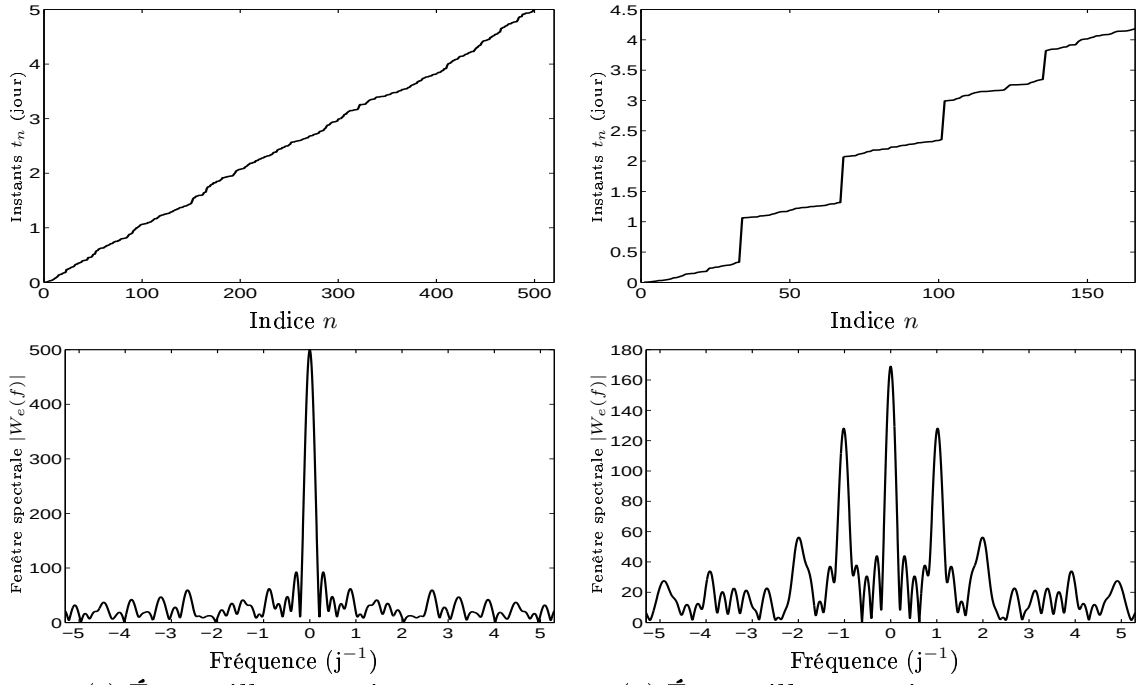
L'influence d'une couverture temporelle irrégulière et trouée se traduit sur le spectre observé sous la forme de la convolution du spectre théorique par la fenêtre $W_e(f)$ étudiée ci-dessus : $Y_e(f) = Y(f) \star W_e(f)$. Deux caractéristiques essentielles sont alors à noter.

– Nous avons vu au paragraphe précédent que le lobe principal de la fenêtre W_e présente des caractéristiques similaires à celles obtenues pour un échantillonnage régulier, à savoir une largeur d'environ $2/T$ et un lobe principal de forme similaire à un sinus cardinal. De ce fait, la *résolution* du périodogramme, *i.e.,* le pouvoir de séparation entre deux fréquences proches, sera similaire à celle obtenue pour un échantillonnage régulier, c'est-à-dire de l'ordre de $1/T$, où T est la durée des observations. En effet, pour des instants d'échantillonnage régulièrement espacés $t_n = nT_e$, la résolution du périodogramme est de l'ordre de $1/NT_e$ [Kunt, 1984]. Ce point est particulièrement intéressant : il signifie, pour l'échantillonnage irrégulier, que c'est la durée totale couverte par les observations qui conditionne



(a) **Échantillonnage régulier**
$N = 500$ points

(b) **Échantillonnage régulier troué**
$N = 166$ points



(c) **Échantillonnage irrégulier**
$N = 500$ points

(d) **Échantillonnage irrégulier troué**
$N = 166$ points

FIG. 1.10 – *Allure de la fenêtre spectrale $|W_e(f)|$ selon les différents types d'irrégularités de l'échantillonnage t_n.*

la résolution du périodogramme. De ce point de vue, il sera donc préférable d'avoir un grand étalement des données au cours du temps, ce qui est une situation courante pour des séries temporelles astronomiques.

– Lors de la présence de raies spectrales multiples, le périodogramme des données présente des pics qui sont à la fois dus aux raies et aux lobes secondaires de la fenêtre W_e, provoqués principalement par les trous de l'échantillonnage. Par conséquent, l'interprétation du périodogramme en tant qu'estimateur de spectres de raies peut devenir extrêmement délicate.

Afin d'illustrer ce second point, considérons le cas d'un signal sinusoïdal, échantillonné selon le schéma de la figure 1.10 (d), présentant à la fois des trous et un espacement irrégulier des mesures dans la zone d'observation. Les données ainsi que leur périodogramme sont représentés en figure 1.11 : le maximum du périodogramme correspond à la fréquence du signal, même si le motif convolutif de la fenêtre spectrale fait apparaître des lobes secondaires, espacés de $1\,\mathrm{j}^{-1}$ par rapport au pic principal. Par contre, dans le cas d'un signal à plusieurs fréquences, le périodogramme

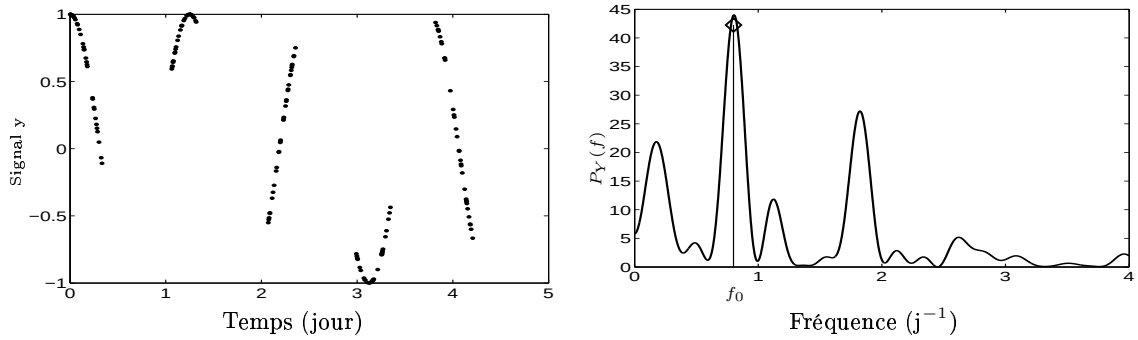


FIG. 1.11 – *À gauche, signal sinusoïdal de fréquence* $f_0 = 0.8$ *observé pendant* 5 *nuits et à échantillonnage irrégulier. À droite, son périodogramme (–) et le spectre théorique (◇).*

peut vite devenir illisible. Pour le même échantillonnage, un signal composé de deux sinusoïdes de fréquences f_0 et $f_1 = f_0 + 2\,\mathrm{j}^{-1}$ et d'amplitudes égales est représenté figure 1.12, ainsi que son périodogramme : la superposition des motifs convolués autour des raies en f_0 et f_1 fait apparaître un troisième pic à la fréquence $(f_0 + f_1)/2$, d'amplitude supérieure à celles associées aux deux raies. Ce cas pathologique[8] illustre l'inadéquation de l'utilisation du périodogramme pour la détection de fréquences multiples lorsque l'échantillonnage présente des trous périodiques.

Pour les problèmes envisagés dans ce manuscrit, le périodogramme est donc un outil à utiliser avec précaution et la forme complexe de la fenêtre spectrale, notamment la présence de forts lobes secondaires, compromet l'usage d'un quelconque fenêtrage des données pour adoucir l'effet de ces lobes, dont l'amplitude est essentiellement fixée par la proportion de trous dans la couverture temporelle. Dans le cas de signaux à échantillonnage irrégulier, la définition du périodogramme (1.3) a été adaptée par Lomb [1976], puis reprise et complétée par Scargle [1982] : il s'agit du périodogramme de Lomb-Scargle, qui présente des propriétés statistiques plus complètes que le premier. S'agissant d'un outil très couramment utilisé en astrophysique – les auteurs mentionnés l'ont initialement introduit pour l'étude de séries temporelles astronomiques – nous

8. L'espacement fréquentiel entre les deux raies est ici volontairement choisi de la manière la plus défavorable, *i.e.*, pour lequel les lobes secondaires dus aux raies en f_0 et f_1 coincident en $\frac{f_0+f_1}{2}$.

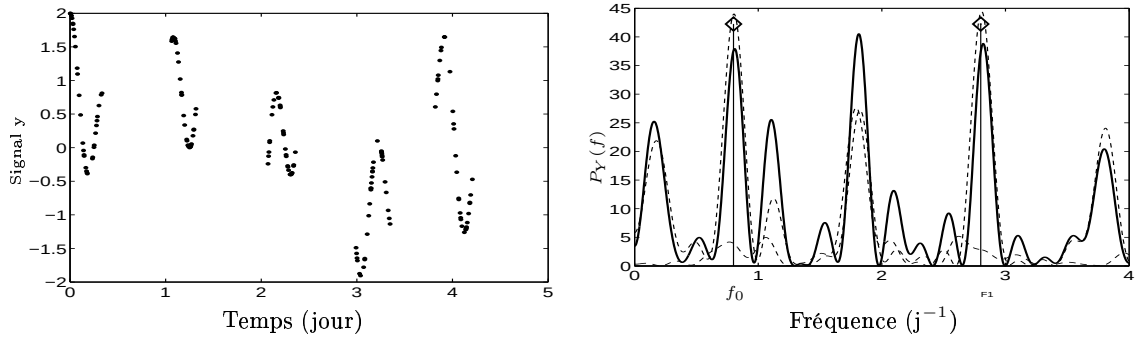


FIG. 1.12 – *À gauche, signal composé de deux sinusoïdes de fréquences* $f_0 = 0.8$ *et* $f_1 = 2.8$*, observé pendant* 5 *nuits et à échantillonnage irrégulier. À droite, son périodogramme (trait plein), le spectre théorique (*$\diamond$*) et les périodogrammes associés aux sinusoïdes en* f_0 *et* f_1 *(pointillés)*

l'étudierons en détail au début du chapitre suivant. Pour l'analyse de séries temporelles présentant des oscillations à fréquences multiples, l'absence de cadre théorique génère naturellement des réticences vis-à-vis de l'utilisation du périodogramme, renforcées par le caractère « troué » de l'échantillonnage. L'interprétation du périodogramme est alors compromise par la forme de la fenêtre spectrale et peut déboucher sur la détection de faux pics et la non-détection de vraies raies.

Chapitre 2

Les outils classiques et leurs limites

Nous considérons le problème de l'estimation de spectres de raies à partir de mesures irrégulièrement échantillonnées. Nous supposerons tout au long de ce manuscrit que les spectres recherchés sont *stationnaires*, c'est-à-dire que leur contenu fréquentiel ne varie pas durant toute la période d'observation. Ce cadre peut sembler à première vue restrictif lorsque l'on s'intéresse à des variations portant sur des observations couvrant de longues périodes. Dans certains cas, des glissements de fréquences au fil du temps ou des excitations temporaires de modes d'oscillation sont même spécifiquement recherchés. L'hypothèse de stationnarité est néanmoins fondamentale au niveau des méthodes d'analyse des séries temporelles : d'une part, le formalisme de la modélisation correspondante s'en trouve fortement simplifié ; d'autre part, les procédures d'estimation qui en résultent présentent des propriétés plus robustes, *a fortiori* lorsque les données sont peu nombreuses et la couverture temporelle pauvre. L'extension à des méthodes envisageant la prise en compte de non stationnarités dans le cadre de l'*analyse temps-fréquence* sera évoquée dans les perspectives de poursuite de ce travail de thèse. Nous nous attacherons par la suite à plusieurs notions essentielles pour décrire une méthode d'analyse spectrale, que nous précisons dès maintenant :

- la *résolution* d'un estimateur représente l'écart fréquentiel minimal qui doit séparer deux raies proches pour être détectées. Ainsi, nous parlerons d'analyse à *haute résolution*, lorsque l'estimateur est capable de détecter deux raies spectrales plus proches que la limite obtenue par le périodogramme, dont le pouvoir de séparation fréquentielle est de l'ordre de l'inverse de la durée des observations [Kunt, 1984] ;
- le terme de *précision* est associé à l'erreur commise sur les valeurs estimées des paramètres spectraux. La précision sur l'estimation des fréquences des raies en sera le point essentiel, sans pour autant négliger la précision d'estimation des amplitudes associées ;
- le *niveau de confiance* est une notion fondamentale associée à la recherche de fréquences dans des séries temporelles, qui concerne la validité à accorder à une éventuelle détection : avec quelle certitude peut-on affirmer, à partir d'une série d'observations, qu'une raie est présente à une fréquence donnée ? L'obtention de statistiques cherchant à décrire la validité de l'estimation est donc essentielle, et nous lui accorderons à ce titre une place importante au fil des descriptions suivantes.

Nous notons par la suite $\{t_n\}_{n=1...N}$ les N instants d'échantillonnage et $\{y_n\}_{n=1...N}$ les mesures associées. La modélisation des données par une somme de sinusoïdes est l'approche la plus

directe, traduisant exactement les variations temporelles recherchées dans les données :

$$\begin{aligned} y_n &= \sum_{k=1}^{K} a_k \cos(2\pi f_k t_n) + b_k \sin(2\pi f_k t_n) + \epsilon_n, \quad n = 1 \dots N \\ &\triangleq \mathcal{M}_K(\boldsymbol{\theta}^K, t_n) + \epsilon_n, \quad n = 1 \dots N \end{aligned} \tag{2.1}$$

où $\boldsymbol{\theta}^K = [a_k, b_k, f_k]_{k=1\dots K}$ regroupe les paramètres inconnus et le terme ϵ_n rend compte des perturbations, comme par exemple le bruit associé à chaque mesure ou les erreurs de modélisation. Ces perturbations supposées additives sont en pratique inconnues, mais pourront être décrites de manière statistique sous la forme d'une distribution de probabilité : $\epsilon_n \sim p(\epsilon_n)$.

L'identification des paramètres $\boldsymbol{\theta}^K$ est un problème complexe. En considérant les perturbations ϵ_n indépendantes et identiquement distribuées (*i.i.d.*) centrées et de statistique gaussienne, soit $\epsilon_n \sim \mathcal{N}(0, \sigma_\epsilon^2)$, la *fonction de vraisemblance* s'écrit :

$$\mathcal{L}(y_1, \dots, y_N; \boldsymbol{\theta}^K) = \frac{1}{(2\pi)^{N/2} \sigma_\epsilon^N} \exp\Big\{ -\frac{1}{2\sigma_\epsilon^2} \big(y_n - \mathcal{M}_K(\boldsymbol{\theta}^K, t_n)\big)^2 \Big\}$$

et l'estimation de $\boldsymbol{\theta}^K$ au sens du *maximum de vraisemblance* (MV) revient à minimiser le terme des moindres carrés correspondant :

$$\begin{aligned} \widehat{\boldsymbol{\theta}}^K_{\text{MV}} &= \arg\min_{\theta_K} \; -\log \mathcal{L}(y_1, \dots, y_N; \boldsymbol{\theta}^K) \\ &= \arg\min_{\theta_K} \; \sum_n \Big(y_n - \sum_{k=1}^{K} a_k \cos(2\pi f_k t_n) + b_k \sin(2\pi f_k t_n) \Big)^2. \end{aligned} \tag{2.2}$$

Dans le cas d'un modèle sinusoïdal, c'est-à-dire $K = 1$, nous rappelons au § 2.1 que l'estimation de $\widehat{\boldsymbol{\theta}}^1_{\text{MV}}$ au sens du maximum de vraisemblance revient, pour des données échantillonnées régulièrement, à considérer le maximum de leur périodogramme :

$$P_S(f) \triangleq \frac{1}{N} \left| \sum_{n=1}^{N} y_n e^{-j2\pi f t_n} \right|^2. \tag{2.3}$$

L'extension au cas de l'échantillonnage irrégulier fait apparaître les lacunes du périodogramme (2.3), appelé encore périodogramme *de Schuster* : une généralisation plus appropriée débouche alors sur le périodogramme *de Lomb-Scargle*, et de manière plus rigoureuse sur le périodogramme *à date compensée*, que nous étudions également au § 2.1.

L'estimation des paramètres d'un modèle à fréquences multiples est un problème bien plus délicat. D'une part, pour un ordre K fixé, la non linéarité du modèle (2.1) en les paramètres f_k rend leur estimation très délicate. Dans le cas de données régulièrement espacées, Stoica *et al.* [1989] montrent que le critère quadratique (2.2) est multimodal, ce qui rend sa minimisation nécessairement sous-optimale. La contrainte d'un échantillonnage irrégulier – en particulier le caractère troué des données – rend encore plus délicate l'estimation de $\boldsymbol{\theta}^K$ au sens du maximum de vraisemblance, en ce sens qu'elle accentue le caractère multimodal du critère à minimiser. L'analyse de l'influence de l'échantillonnage sur le périodogramme, présentée au chapitre précédent, a en effet montré que les pics secondaires du périodogramme dûs aux trous de l'échantillonnage s'interprètent comme des maxima locaux de la vraisemblance du modèle sinusoïdal $\mathcal{M}_1(\boldsymbol{\theta}^1)$; l'extension au cas multi-sinusoïdal laisse alors prévoir une multimodalité accrue du critère (2.2) lorsque les

données présentent des trous périodiques. D'autre part, le nombre K de sinusoïdes, c'est-à-dire le nombre de raies spectrales, est en pratique inconnu. Des critères de sélection de l'ordre de type *Akaike Information Criterion* [Akaike, 1974] ou *Minimum Description Length* [Rissanen, 1978] peuvent alors être envisagés, ou plus spécifiquement dans le cas d'un modèle multi-sinusoïdal le critère proposé par Djurić [1996]. Dans chaque cas, le principe consiste à choisir l'ordre K minimisant un critère pénalisé :

$$\widehat{K} = \arg\min_{K} -\log \mathcal{L}(y_1, \ldots, y_N; \widehat{\boldsymbol{\theta}}^K_{\text{MV}}) + p_K$$

où le premier terme correspond à la valeur minimale du terme des moindres carrés (2.2) atteinte en $\widehat{\boldsymbol{\theta}}^K_{\text{MV}}$. L'expression de la pénalisation p_K diffère alors en fonction du critère de sélection de l'ordre. Ce genre de critère suppose néanmoins implicitement l'obtention, pour différents ordres candidats K, de l'estimateur du maximum de vraisemblance $\widehat{\boldsymbol{\theta}}^K_{\text{MV}}$ correspondant, tâche qui, nous venons de le voir, semble d'autant plus ardue que l'échantillonnage est irrégulier. Une alternative proposée par Andrieu et Doucet [1999] envisage de baser l'estimation conjointe de l'ordre du modèle K et de ses paramètres $\boldsymbol{\theta}^K$ à partir de l'échantillonnage stochastique de la distribution de probabilité *a posteriori* $p(K,\boldsymbol{\theta}^K|\boldsymbol{y})$ en considérant une loi *a priori* Poisson-Gaussienne sur $(K,\boldsymbol{\theta}^K)$. Le recours à des méthodes de Monte-Carlo par chaînes de Markov (MCMC) à sauts réversibles permet ainsi de contourner les problèmes spécifiques à une approche de type optimisation. Cette approche a été étudiée dans le cas de données à échantillonnage irrégulier lors du stage de DEA de S. Bourguignon [2002], et a fourni des résultats encourageants ; cependant, l'exploitation de la chaîne produite s'est révélée très délicate : en particulier, la variation de la dimension du modèle au fil des échantillons générés peut perturber l'indexation des fréquences le long de la chaîne et déboucher alors sur une estimation insatisfaisante.

La déconvolution du spectre de Fourier par des méthodes itératives de type CLEAN a également pour objectif l'identification des paramètres d'un modèle de type $\mathcal{M}_K$, en extrayant un par un les pics du spectre de Fourier et la contribution associée de la fenêtre spectrale. L'estimation de l'ordre K est alors reliée à la définition d'un test d'arrêt de la procédure itérative. S'agissant d'un outil fréquemment rencontré pour le traitement de données astrophysiques, nous lui consacrons une étude détaillée au § 2.2, où nous mettrons également en évidence ses limites. Nous décrivons ensuite brièvement au § 2.3 des techniques propres à l'estimation d'une seule période et développées pour l'étude de systèmes multiples. Spécifiquement adaptés pour des données peu nombreuses et très éparses, ces méthodes « intuitives » ont fourni des résultats satisfaisants sur ce genre de problème mais manquent de propriétés théoriques solides.

L'utilisation d'autres modèles paramétriques de type autorégressifs (AR), dont l'efficacité pour l'analyse spectrale *à haute résolution* est incontestable dans le cas de données régulièrement espacées [Kay, 1988], ne s'adapte pas facilement au cas de l'échantillonnage irrégulier. En effet, la structure même d'un modèle aléatoire AR, reliant la valeur du signal à un instant donné à ses valeurs aux instants précédents, appelle naturellement un espacement constant des données : l'estimation de la fonction d'autocorrélation, par exemple, s'appuie naturellement sur un échantillonnage régulier. Néanmoins, différents travaux sur cette voie [Martin, 1999, Rivoira *et al.*, 2002] ont proposé la généralisation de concepts proprement définis pour l'échantillonnage régulier (autocorrélation, énergie d'erreur de prédiction), bien qu'au prix de la perte de certaines propriétés analytiques. Si elles semblent appropriées pour des données irrégulièrement échantillonnées et peu nombreuses, ces approches considèrent surtout des schémas d'échantillonnage *aléatoires* et les propriétés des estimateurs proposés s'appuient sur la distribution de probabilité de ces schémas. À ce titre, cette démarche nous paraît plus adaptée au cas d'irrégularités

« régulières » de l'échantillonnage, c'est-à-dire ne présentant pas de gros trous. Les spécificités des schémas d'échantillonnage rencontrés dans les séries temporelles astrophysiques nous rendent donc peu enclins à l'utilisation de telles méthodes.

Pour clore ce chapitre, nous présentons au § 2.4 l'approche adoptée dans la suite de ce manuscrit. Afin de s'affranchir des contraintes propres à l'estimation de modèles paramétriques, nous abordons l'analyse spectrale sous la forme d'un *problème inverse* linéaire – le spectre étant discrétisé sur une grille arbitrairement fine – auquel le manque de données disponibles et la nécessité d'une analyse à haute résolution fréquentielle confèrent un caractère sous-déterminé. Sa *régularisation* est alors envisagée en prenant en compte la nature *a priori parcimonieuse* de l'objet à reconstruire, puisque nous nous intéressons exclusivement à la recherche de spectres de raies. Une telle méthodologie a déjà été employée pour l'estimation de spectres de raies à partir de données régulièrement échantillonnées (nous ferons notamment référence aux travaux de Dublanchet [1996], Sacchi *et al.* [1998], Ciuciu *et al.* [2001]) et a permis de lever la traditionnelle incompatibilité entre l'utilisation d'un modèle *non paramétrique* et l'obtention d'un estimateur spectral à *haute résolution*, c'est-à-dire dont le pouvoir de séparation entre deux raies proches dépasse celui du périodogramme.

2.1 Les périodogrammes

2.1.1 Le périodogramme de Lomb-Scargle

Pour des données $\{t_n,y_n\}_{n=1\ldots N}$ irrégulièrement échantillonnées, considérons le cas d'un modèle sinusoïdal :

$$y_n = a\cos(2\pi f t_n) + b\sin(2\pi f t_n) + \epsilon_n,\ n = 1\ldots N \tag{2.4}$$

et le problème de l'estimation de ses paramètres $\boldsymbol{\theta} = [a,b,f]$. Sous forme matricielle, on peut encore écrire :

$$\boldsymbol{y} = \mathbf{R}(f)\,\boldsymbol{s} + \boldsymbol{\epsilon}$$

en notant $\boldsymbol{y} = [y_1,\ldots,y_n]^{\mathrm{t}}$, $\boldsymbol{\epsilon} = [\epsilon_1,\ldots,\epsilon_n]^{\mathrm{t}}$, $\boldsymbol{s} = [a,b]^{\mathrm{t}}$ et $\mathbf{R}(f)$ la matrice $N \times 2$ formée par les N lignes $[\cos(2\pi f t_n),\ \sin(2\pi f t_n)]_{n=1\ldots N}$. En supposant les perturbations ϵ_n indépendantes et distribuées selon une même distribution de probabilité p_ϵ, la *fonction de vraisemblance* du modèle (2.4) est définie comme :

$$\mathcal{L}(\boldsymbol{y};f,a,b) = \Pi_{n=1}^{N} p_\epsilon\left(y_n - a\cos(2\pi f t_n) - b\sin(2\pi f t_n)\right)$$

et sous l'hypothèse supplémentaire de perturbations gaussiennes centrées et de variance connue σ_ϵ^2, la vraisemblance prend la forme :

$$\mathcal{L}(\boldsymbol{y};f,\boldsymbol{s}) = \frac{1}{(\sigma_\epsilon\sqrt{2\pi})^N}\exp\{-\frac{1}{2\sigma_\epsilon^2}||\boldsymbol{y} - \mathbf{R}(f)\boldsymbol{s}||^2\}. \tag{2.5}$$

Les paramètres $\boldsymbol{\theta}$ maximisant cette vraisemblance sont alors ceux minimisant le terme de moindres carrés $\mathcal{I}(f,\boldsymbol{s}) = ||\boldsymbol{y} - \mathbf{R}(f)\boldsymbol{s}||^2$. Pour une fréquence f donnée, la valeur de $\boldsymbol{s}$ minimisant $\mathcal{I}(f,\boldsymbol{s})$ est classiquement :

$$\begin{aligned}\widehat{\boldsymbol{s}}(f) &= \left(\mathbf{R}(f)^{\mathrm{t}}\mathbf{R}(f)\right)^{-1}\mathbf{R}(f)^{\mathrm{t}}\boldsymbol{y} \\ &= \begin{bmatrix} CC(f) & CS(f) \\ CS(f) & SS(f) \end{bmatrix}^{-1} \begin{bmatrix} \sum_n y_n\cos(2\pi f t_n) \\ \sum_n y_n\sin(2\pi f t_n) \end{bmatrix}\end{aligned} \tag{2.6}$$

où $CC(f) = \sum_n \cos^2(2\pi f t_n)$, $CS(f) = \sum_n \cos(2\pi f t_n)\sin(2\pi f t_n)$ et $SS(f) = \sum_n \sin^2(2\pi f t_n)$. L'injection de la relation (2.6) dans l'expression de $\mathcal{I}$ permet ainsi d'obtenir un terme ne dépendant que du paramètre f :

$$\begin{aligned} \mathcal{I}(f,\widehat{\boldsymbol{s}}(f)) &= ||\boldsymbol{y} - \mathbf{R}(f)\widehat{\boldsymbol{s}}(f)||^2 \\ &= ||\boldsymbol{y}||^2 - \widehat{\boldsymbol{s}}(f)^\dagger \mathbf{R}(f)^\dagger \boldsymbol{y} \end{aligned}$$

dont la minimisation en f est équivalente à la maximisation de $J(f) \triangleq \widehat{\boldsymbol{s}}(f)^\dagger \mathbf{R}(f)^\dagger \boldsymbol{y}$. Afin de simplifier l'écriture de ce terme, Lomb [1976] propose de modifier le modèle sinusoïdal initial en introduisant un décalage τ :

$$y_n = a\cos 2\pi f(t_n - \tau) + b\sin 2\pi f(t_n - \tau) + \epsilon_n$$

et de choisir τ, pour une fréquence f fixée, de manière à annuler les termes extra-diagonaux $CS(f)$ de la matrice $\mathbf{R}(f)^\dagger\mathbf{R}(f)$ ainsi modifiée, soit τ tel que $\sum_n \cos\left(2\pi f(t_n - \tau)\right)\sin\left(2\pi f(t_n - \tau)\right) = 0$. Cet artifice permet alors d'écrire sous forme simplifiée :

$$\frac{1}{2}J(f) = \frac{1}{2}\left\{ \frac{\left[\sum_n y_n \cos 2\pi f(t_n - \tau)\right]^2}{\sum_n \cos^2 2\pi f(t_n - \tau)} + \frac{\left[\sum_n y_n \sin 2\pi f(t_n - \tau)\right]^2}{\sum_n \sin^2 2\pi f(t_n - \tau)} \right\} \triangleq P_{\text{LS}}(f) \tag{2.7}$$

$$\text{en choisissant } \tau = \frac{1}{4\pi f}\tan^{-1}\frac{\sum_n \sin 4\pi f t_n}{\sum_n \cos 4\pi f t_n}$$

qui est la définition du *périodogramme de Lomb-Scargle* (PLS), où le facteur 1/2 permet d'identifier cette définition à celle du périodogramme classique (2.3) dans le cas d'un échantillonnage régulier. La valeur $\widehat{f}$ maximisant $J(f)$ est donc la fréquence correspondant au maximum de la vraisemblance du modèle (2.4), et les amplitudes associées $\widehat{\boldsymbol{s}}(\widehat{f})$ se déduisent *a posteriori* de la relation (2.6).

Bien sûr, l'écriture initiale du critère $\mathcal{I}(f,\widehat{\boldsymbol{s}}(f))$ permet également d'estimer les paramètres du modèle (2.4) au sens du maximum de vraisemblance. L'artifice consistant à introduire un décalage τ a pour objectif une écriture plus simple du périodogramme sous la forme de la somme de deux carrés, ce qui est interprété de manière *statistique* par Lomb [1976], puis repris par Scargle [1982] ; nous y reviendrons au § 2.1.3. Nous considérons néanmoins que l'interprétation *géométrique* explicitée au prochain paragraphe permet de mieux décrire le périodogramme en tant que mesure de l'adéquation entre les données et un modèle sinusoïdal.

Interprétation géométrique

L'analyse de Fourier s'interprète en termes de projections dans des espaces fonctionnels en munissant l'ensemble $\mathcal{E}_t$ des signaux réels de taille N échantillonnés aux instants $\{t_n\}_{n=1\ldots N}$ du produit scalaire naturel :

$$\langle y|z\rangle \triangleq \sum_{n=1}^{N} y_n z_n \quad \text{et de la norme associée } \|y\| = \sqrt{\langle y|y\rangle},$$

où y_n et z_n représentent les valeurs respectives des signaux y et z aux instants t_n. Cette approche permet d'obtenir une interprétation supplémentaire des différents périodogrammes et met en évidence les limites du périodogramme de Lomb-Scargle, avantageusement remplacé par le périodogramme à date compensée.

Lorsque l'échantillonnage est régulier ($t_n = nT_e, n = 1 \dots N$), il est bien connu que l'ensemble des N fonctions :

$$\sqrt{\frac{1}{N}}\left\{1, \left\{\sqrt{2}\cos 2\pi f_k t\right\}_{k=1\dots N/2}, \left\{\sqrt{2}\sin 2\pi f_k t\right\}_{k=1\dots N/2-1}\right\} \text{ pour les fréquences } f_k = \frac{k}{NT_e}$$

forme une base orthonormale de $\mathcal{E}_t$ et la décomposition d'un signal y sur cette base s'identifie aux parties réelle et imaginaire de sa transformée de Fourier discrète (TFD), qui est le spectre de Fourier évalué aux fréquences f_k :

$$\begin{aligned} \forall k = 0 \dots N/2-1, \quad Y(f_k) &= \sum_{n=1}^{N} y_n e^{-j2\pi f_k t_n} = \sum_{n=1}^{N} y_n e^{-j2\pi \frac{kn}{N}} \\ &= \langle y|\cos 2\pi f_k t\rangle - j\,\langle y|\sin 2\pi f_k t\rangle\,. \end{aligned}$$

L'adéquation entre un modèle sinusoïdal de fréquence f_k et les données y peut alors être mesurée par la norme de la projection de y sur les vecteurs de base $B_{\text{REG}} = (e_1(t),e_2(t)) = \left(\sqrt{2/N}\cos 2\pi f_k t, \sqrt{2/N}\sin 2\pi f_k t\right)$. Ces vecteurs étant orthonormaux, la projection s'écrit :

$$\begin{aligned} \overline{y} \triangleq P^{\perp}_{B_{\text{REG}}}\{y\} \quad \text{avec} \quad \overline{y}_n &= \langle y|e_1\rangle\, e_1(t_n) + \langle y|e_2\rangle\, e_2(t_n) \\ &= 2\,\langle y|\cos 2\pi f_k t\rangle \cos 2\pi f_k t_n + 2\,\langle y|\sin 2\pi f_k t\rangle \sin 2\pi f_k t_n \end{aligned}$$

et sa norme vaut :

$$\begin{aligned} Q_{\text{REG}}(f_k) &\triangleq \langle \overline{y}|\overline{y}\rangle \\ &= \frac{2}{N}\,\langle y|\cos 2\pi f_k t\rangle^2 + \frac{2}{N}\,\langle y|\sin 2\pi f_k t\rangle^2 \\ &= \frac{2}{N}|Y(f_k)|^2 \\ &= 2P_S(f_k) \end{aligned}$$

où $P_{\text{S}}\,(f_k) = \frac{1}{2}\left(\langle y|e_1\rangle^2 + \langle y|e_2\rangle^2\right)$ est le périodogramme de Schuster défini par la relation (2.3). Ce dernier est donc une mesure de la distance des données à un modèle sinusoïdal de fréquence f_k. Remarquons que cette interprétation, basée sur l'orthogonalité des vecteurs $\cos 2\pi f_k t$ et $\sin 2\pi f_k t$, n'est valable que pour les fréquences f_k vérifiant $f_k = \frac{k}{NT_e}$.

Lorsque l'échantillonnage est irrégulier[1], le terme $Q_{\text{S}}\,(f) = \frac{2}{N}P_S(f)$ dérivé du périodogramme de Schuster ne peut désormais plus être interprété en termes de projection de y sur la famille B_{REG} car ses éléments ne sont plus orthonormaux : $\langle \cos 2\pi ft|\sin 2\pi ft\rangle \neq 0$ et $\langle \cos 2\pi ft|\cos 2\pi ft\rangle$, $\langle \sin 2\pi ft|\sin 2\pi ft\rangle \neq N/2$. Néanmoins, pour une fréquence f donnée, il est possible de les orthonormaliser en considérant par exemple la famille :

$$B_{\text{LS}} = \big(h_1(t),h_2(t)\big) = \left(\frac{\cos 2\pi f(t-\tau)}{\|\cos 2\pi f(t-\tau)\|}, \frac{\sin 2\pi f(t-\tau)}{\|\sin 2\pi f(t-\tau)\|}\right)$$

où τ a été défini par la relation (2.7) : l'annulation des termes extra-diagonaux $CS(f)$ dans l'expression (2.7) revient exactement à orthogonaliser les contributions en cosinus et en sinus

1. Ces considérations sont également valables dans le cas d'un échantillonnage régulier de période T_e et une fréquence f ne s'écrivant pas sous la forme $f_k = \frac{k}{NT_e}$.

d'une même fréquence. Ainsi, par construction, B_{LS} est une famille orthonormale et la projection $\overline{y}$ des données sur B_{LS} s'écrit :

$$\begin{aligned}\overline{y} \triangleq P^{\perp}_{B_{\text{LS}}}\{y\} \quad \text{avec} \quad \overline{y}_n &= \langle y|h_1\rangle\, h_1(t_n) + \langle y|h_2\rangle\, h_2(t_n) \\ &= \langle y|\cos 2\pi f_k(t-\tau)\rangle \frac{\cos 2\pi f(t_n-\tau)}{\|\cos 2\pi f(t-\tau)\|^2} \quad \dots \\ &\quad \dots \quad + \langle y|\sin 2\pi f_k(t_n-\tau)\rangle \frac{\sin 2\pi f(t-\tau)}{\|\sin 2\pi f(t-\tau)\|^2}\end{aligned}$$

dont la norme vaut :

$$\begin{aligned}Q_{\text{LS}}(f) &\triangleq \langle\overline{y}|\overline{y}\rangle \\ &= \frac{\langle y|\cos 2\pi f(t-\tau)\rangle^2}{\|\cos 2\pi f(t-\tau)\|^2} + \frac{\langle y|\sin 2\pi f(t-\tau)\rangle^2}{\|\sin 2\pi f(t-\tau)\|^2} \\ &= 2P_{\text{LS}}(f).\end{aligned}$$

À f fixé, la norme de la projection de y sur la famille B_{LS} est donc proportionnelle au périodogramme de Lomb-Scargle :

$$P_{\text{LS}}(f) = \frac{1}{2}\left(\langle y|h_1\rangle^2 + \langle y|h_2\rangle^2\right).$$

Cette interprétation permet donc d'associer à l'amplitude du PLS un sens géométrique dont le périodogramme de Schuster est dépourvu en échantillonnage irrégulier : celui de la distance des données à un modèle sinusoïdal de fréquence f. Ainsi, pour un signal sinusoïdal échantillonné irrégulièrement, d'amplitude unité, de fréquence f_0 et ayant le déphasage τ associé par l'équation (2.7), on aura $Q_{\text{LS}}(f_0) = 1$ et $P_{\text{LS}}(f_0) = 1/2$, alors qu'une propriété similaire ne peut pas être établie à partir du périodogramme de Schuster.

2.1.2 La méconnue DCDFT et le périodogramme à date compensée

Selon cette même interprétation géométrique, cependant, la non orthogonalité des fonctions cosinus et sinus à la fonction constante 1 pour un échantillonnage irrégulier n'est pas prise en compte dans le périodogramme de Lomb-Scargle. Ferraz-Mello [1981] puis Foster [1995] montrent ainsi que cette caractéristique est à l'origine d'effets indésirables, aussi bien dans le périodogramme de Schuster que dans le PLS : pour des données sinusoïdales, la position du maximum du périodogramme peut en être affectée et l'estimation de son amplitude est altérée par la persistance d'une composante continue dans les données. En effet, le recentrage des données par rapport à leur valeur moyenne n'est pas une solution suffisante pour un échantillonnage irrégulier car la répartition, *a priori* non uniforme, des instants d'échantillonnage par rapport aux variations de la fonction échantillonnée biaise l'estimation de la moyenne. Par exemple, une plus grande concentration de mesures à des instants pour lesquels le signal est d'amplitude élevée donnera une valeur moyenne des données surestimant la véritable moyenne. Autrement dit, ce n'est pas parce que les données sont de moyenne nulle, *i.e.*, $\sum_{n=1}^{N} y_n = 0$, qu'il ne subsiste pas de composante continue dans le signal. L'exemple de la figure 2.1 illustre cette situation. Le signal présenté est sinusoïdal et l'échantillonnage comporte des trous périodiques, dont la période est proche de celle du signal ; de ce fait, la valeur moyenne des données est largement positive alors que le signal est à moyenne théorique nulle.

Afin de pallier cette insuffisance, Ferraz-Mello [1981] considère un modèle sinusoïdal prenant en compte une composante continue additionnelle. Il envisage alors la projection des données y,

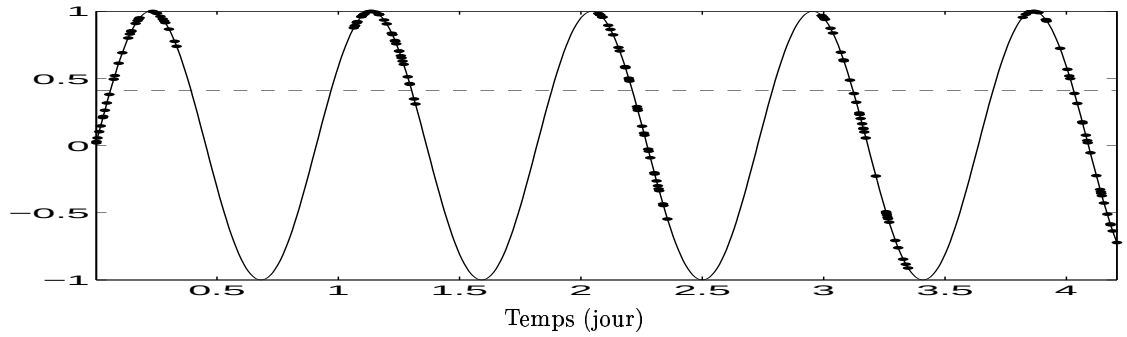


FIG. 2.1 – *Effet d'un échantillonnage troué (de fréquence 1 j^{-1}) sur l'estimation de la moyenne. Le signal est une sinusoïde de fréquence $f_0 = 1.1\ j^{-1}$. Données disponibles (·), signal complet (–) et valeur moyenne des données (- -).*

pour une fréquence f fixée, sur le sous-espace engendré par les trois vecteurs $(1, \cos 2\pi ft, \sin 2\pi ft)$. Ceux-ci ne formant pas une famille orthonormale, il propose de lui appliquer une procédure d'orthonormalisation de Gram-Schmidt, obtenant ainsi une famille $B_{\text{DC}} = \big(k_0(t),k_1(t),k_2(t)\big)$ orthonormale **pour la fréquence f et pour le schéma d'échantillonnage donné**, où $k_0(t) = \frac{1}{\sqrt{N}}$ et les expressions correspondantes de k_1 et k_2 sont détaillées dans [Ferraz-Mello, 1981]. La projection des données sur B_{DC} s'écrit alors :

$$\overline{y} \triangleq P^{\perp}_{B_{\text{DC}}}\{y\} \quad \text{avec} \quad \overline{y}_n = \langle y|k_0\rangle\, k_0(t_n) + \langle y|k_1\rangle\, k_1(t_n) + \langle y|k_2\rangle\, k_2(t_n)$$

et, pour des données auxquelles on a soustrait leur valeur moyenne, on a $\langle y|k_0\rangle = 0$. La norme de la projection vaut donc :

$$\begin{aligned} Q_{\text{DC}}(f) &\triangleq \langle \overline{y}|\overline{y}\rangle \\ &= \langle y|k_1\rangle^2 + \langle y|k_2\rangle^2 \\ &= \frac{2}{N}|Y_{\text{DC}}(f)|^2 = 2P_{\text{DC}}(f) \end{aligned} \tag{2.8}$$

où la DCDFT (*Date-Compensated Discrete Fourier Transform*) $Y_{\text{DC}}(f)$ est définie par Ferraz-Mello [1981] comme :

$$Y_{\text{DC}}(f) \triangleq \sqrt{\frac{N}{2}}\big(\,\langle y|k_1\rangle + j\,\langle y|k_2\rangle\,\big)$$

et le *périodogramme à date compensée* (PDC) est défini par :

$$P_{\text{DC}}(f) \triangleq \frac{1}{2}\left(\langle y|k_1\rangle^2 + \langle y|k_2\rangle^2\right). \tag{2.9}$$

L'estimation par le périodogramme à date compensée possède donc des propriétés plus rigoureuses que le périodogramme de Lomb-Scargle et est à ce titre injustement méconnue. De la même manière que le maximum du PLS peut être vu comme l'estimateur du maximum de vraisemblance de la fréquence d'un modèle sinusoïdal, le maximum du PDC correspond à l'estimation au sens du maximum de vraisemblance de la fréquence d'un modèle de la forme « sinusoïde + constante ». La figure 2.2 compare la valeur des trois périodogrammes $P_{\text{S}}(f)$, $P_{\text{LS}}(f)$ et $P_{\text{DC}}(f)$

– le premier, rappelons-le, ne s'interprète plus en termes de projections si l'échantillonnage est irrégulier – correspondant au signal introduit sur la figure 2.1. Ces trois grandeurs sont également représentées dans le cas où les données sont en même nombre, mais cette fois pour un échantillonnage aléatoire uniforme sur la même durée d'observation. Parmi les trois, $P_{\rm DC}(f)$ est la seule fonction dont le maximum est situé en f_0 dans les deux cas. Remarquons encore que les propriétés caractérisant ces trois périodogrammes sont intrinsèquement liées à la recherche d'**une seule** sinusoïde dans les données. L'expérience montre néanmoins que, même si seul le périodogramme de Schuster est issu d'un produit de convolution, et si seul le PDC s'interprète proprement en termes de projections, les trois périodogrammes présentent des comportements similaires vis-à-vis de modèles multi-sinusoïdaux, avec les conséquences décrites au § 1.3.3 du premier chapitre : la convolution des vraies raies avec une fenêtre spectrale pouvant comporter des lobes secondaires très marqués (si l'échantillonnage présente des trous périodiques) peut déboucher sur la détection de faux pics. L'utilisation des périodogrammes – plus exactement du spectre de Fourier et de la DCDFT – sera cependant à la base des algorithmes de déconvolution itérative de type CLEAN que nous étudierons au § 2.2.

À propos de l'amplitude des périodogrammes, notons encore que pour un signal sinusoïdal de fréquence f_0, l'orthogonalité de la projection impose dans le cas du PDC l'égalité $||y||^2 = \langle y|k_1\rangle^2 + \langle y|k_2\rangle^2$ et donc $P_{\rm DC}(f_0) = \frac{1}{2}||y||^2 = \frac{1}{2}\sum y_n^2$, ce qui n'est pas une relation exacte pour $P_{\rm S}\,(f_0)$ et $P_{\rm LS}(f_0)$. Pour un échantillonnage présentant des trous, cependant, la valeur de $\frac{1}{N}||y||^2$ ne correspond pas à la puissance théorique de la sinusoïde à cause du biais introduit par la persistance de la composante continue, comme nous le voyons sur la figure 2.2 à gauche : la puissance théorique du signal sinusoïdal vaut $1/2$, mais on a $||y||^2 \neq N/2$. Lorsque l'échantillonnage est régulier, ou tout du moins si les irrégularités ne perturbent pas l'estimation de la moyenne du signal, comme dans la figure 2.2 à droite, le terme $\frac{1}{N}||y||^2$ se rapproche de la puissance théorique du signal, qui vaut $1/2$ dans cet exemple, soit $P_{\rm DC}(f_0) \simeq N/4$. En conclusion, l'irrégularité de l'échantillonnage, et notamment son caractère troué, peut également venir perturber l'interprétation de l'amplitude du PDC.

2.1.3 Interprétation statistique

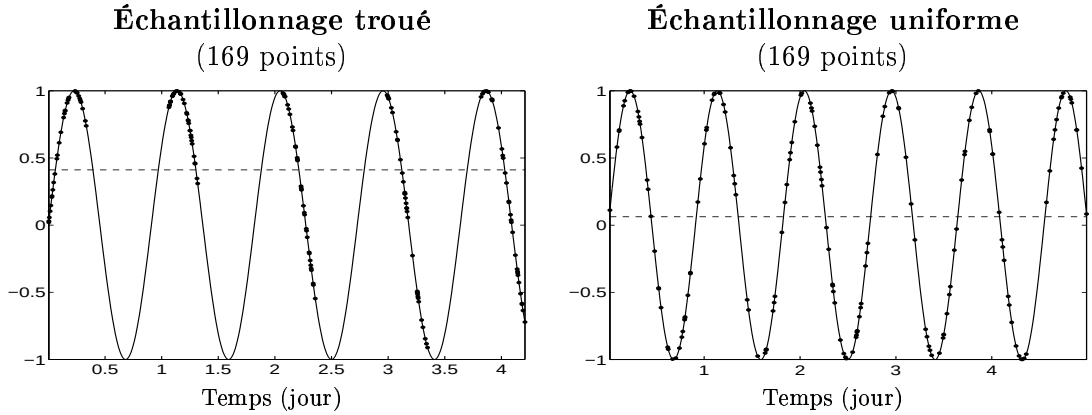
Nous supposerons par la suite que les données sont toujours centrées, c'est-à-dire telles que $\sum_n y_n = 0$. Puisque les périodogrammes de Schuster, de Lomb-Scargle et à date compensée sont par nature adaptés à la recherche d'une seule fréquence dans les données, un test statistique immédiat à considérer est celui de l'hypothèse :

$$\mathcal{H}_0 : \textit{aucune sinusoïde n'est présente dans les données,}$$

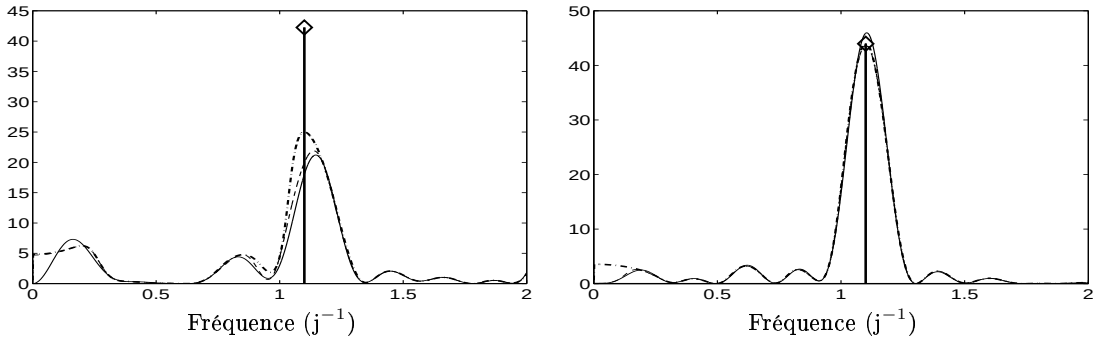
permettant d'associer un niveau de confiance à un pic d'amplitude dans le périodogramme. L'approche du § 2.1.1 décrivant les périodogrammes en termes de projections orthogonales fournit un cadre adéquat pour l'utilisation d'un tel test. Elle permet en effet, pour une fréquence f fixée, d'écrire le périodogramme $P(f)$ sous la forme de la somme de deux carrés :

$$P(f) = \frac{1}{2}\{C^2(f) + S^2(f)\}$$

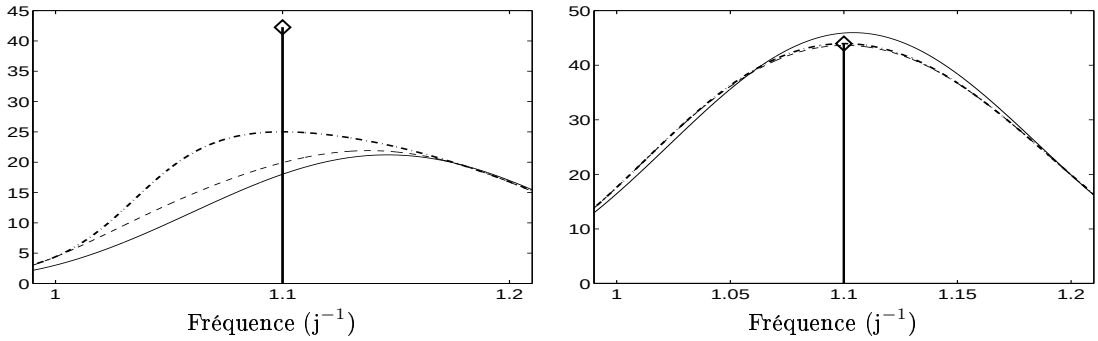
où $C(f)$ et $S(f)$ sont les coefficients de projection des données sur les deux vecteurs d'une famille orthonormale, notée de manière générique $B = \big(\ell_1(t),\ell_2(t)\big)$. Sous l'hypothèse $\mathcal{H}_0$, les données contiennent une unique composante continue μ, *a priori* inconnue (conformément aux considérations du § 2.1.2, l'estimation de μ est biaisée par l'irrégularité de l'échantillonnage) et



Données disponibles (·), signal théorique sinusoïdal (–) et valeur moyenne des données (- -).



Fréquences de 0 à 2 j^{-1}.



Zoom autour de la fréquence théorique $f_0 = 1.1$ j^{-1}.

Périodogramme de Schuster (–), de Lomb-Scargle (- -) et à date compensée (· -). Le ◇ représente le point (f_0,$N/4$).

FIG. 2.2 – *Importance de la prise en compte de la composante continue. À gauche, cas d'un échantillonnage « pathologique » présentant des trous de fréquence* 1 j^{-1} *pour une sinusoïde de fréquence* 1.1 j^{-1}. *À droite, échantillonnage irrégulier uniforme. Dans les deux cas,* $N = 169$ *points.*

on a alors :

$$\mathcal{H}_0 \ : y_n = \mu + \epsilon_n$$

où $\{\epsilon_n\}_n$ est un bruit blanc gaussien centré et de puissance σ_ϵ^2 ; les coefficients de projection sur B sont alors également gaussiens avec :

$$C(f) = \langle y|\ell_1\rangle = \mu\,\langle 1|\ell_1\rangle + \langle \epsilon|\ell_1\rangle \qquad \text{et} \qquad S(f) = \langle y|\ell_2\rangle = \mu\,\langle 1|\ell_2\rangle + \langle \epsilon|\ell_2\rangle$$

$$\Rightarrow \left\{ \begin{array}{rcl} \mathbb{E}[C(f)] & = & \mu\,\langle 1|\ell_1\rangle \\ \text{var } C(f) & = & \sigma_\epsilon^2\,\|\ell_1\|^2 = \sigma_\epsilon^2 \end{array} \right. \qquad \Rightarrow \left\{ \begin{array}{rcl} \mathbb{E}[S(f)] & = & \mu\,\langle 1|\ell_2\rangle \\ \text{var } S(f) & = & \sigma_\epsilon^2\,\langle \ell_2|1\rangle = \sigma_\epsilon^2. \end{array} \right.$$

Dans le cas du périodogramme à date compensée, la famille de projection – les vecteurs notés k_1 et k_2 au § 2.1.2 – est orthonormale à la fonction constante 1. Les coefficients de projection correspondants $C_{\text{DC}}(f)$ et $S_{\text{DC}}(f)$ suivent donc également des lois gaussiennes $\mathcal{N}(0,\sigma_\epsilon^2)$ et on a la propriété suivante :

Sous l'hypothèse $\mathcal{H}_0$, $Z_{DC} = \frac{2}{\sigma_\epsilon^2} P_{DC}(f)$ *suit une loi du* χ^2 *à deux degrés de liberté, c'est-à-dire une loi exponentielle :* $p_Z(z) = \frac{1}{2\sigma_\epsilon^2} \exp -z/2\sigma_\epsilon^2$.

Cette propriété est à la base du test statistique permettant d'évaluer le niveau de confiance accordé à une éventuelle détection, *i.e.,* du test contredisant l'hypothèse $\mathcal{H}_0$.

Pour le périodogramme de Lomb-Scargle, les vecteurs de B_{LS} ne sont pas orthogonaux à 1. Par conséquent, $\mathbb{E}[C_{\text{LS}}(f)] \neq \mathbb{E}[S_{\text{LS}}(f)] \neq 0$ et $\frac{2}{\sigma_\epsilon^2} P_{\text{LS}}(f)$ ne suit plus exactement une loi de type χ^2. Les travaux précités de G. Foster soulignent cette inexactitude et la pratique fréquente visant à considérer $\frac{2}{\sigma_\epsilon^2} P_{\text{LS}}(f) \sim \chi_2^2$, qui devient une relation exacte uniquement lorsque la constante μ est nulle. Signalons encore une fois que ce problème n'est pas lié au centrage des données (nous supposons ici $\langle y|1\rangle = 0$), mais à la non orthogonalité de la famille de projection à la fonction constante 1, due uniquement à l'échantillonnage : avec les notations du § 2.1.1, on a $\langle h_1|1\rangle \propto \sum_n \cos 2\pi f(t_n - \tau) \neq 0$ et $\langle h_2|1\rangle \propto \sum_n \sin 2\pi f(t_n - \tau) \neq 0$.

L'interprétation statistique de l'amplitude du périodogramme dans le cas de l'échantillonnage irrégulier n'est donc rigoureusement valide que pour le périodogramme à date compensée, ce qui montre une nouvelle fois son intérêt par rapport aux périodogrammes de Schuster et de Lomb-Scargle. Néanmoins, il convient d'associer à cette interprétation toutes les réserves liées à un test d'hypothèse de type $\mathcal{H}_0$ et de ne pas se méprendre sur sa signification. La proposition : « le test est significatif à 95% » est parfois énoncée sous la forme : « la présence d'une raie est détectée avec un niveau de confiance de 95% », ce qui détourne le test de son véritable sens qui est : « sous l'hypothèse $\mathcal{H}_0$, il y a 5% de chances que la valeur de Z calculée dépasse la limite donnée par la table du χ_2^2 », ou encore « au vu des données, en affirmant la détection d'une raie, on se trompe *statistiquement* moins de 5% des fois ». Foster [1996] émet encore d'autre réserves limitant la portée d'un test de type χ_2^2, portant notamment sur l'aspect *paramétrique* de la projection : ce test, normalement appliqué à *une* fréquence f choisie *au hasard*, est ici appliqué pour la fréquence maximisant le périodogramme, qui n'est pas choisie au hasard! Le test du χ^2 n'est donc pas assez sélectif pour justifier la détection d'une raie au vu d'un pic dans le périodogramme. Ce genre de raffinement génère donc des ambiguïtés au niveau de l'interprétation statistique des périodogrammes. Ainsi, même si le maximum du périodogramme – en particulier dans sa version à date compensée – permet de localiser la position la plus probable de la fréquence correspondant à un modèle sinusoïdal, l'interprétation de son amplitude en termes de niveau de confiance reste délicate. Signalons enfin que, bien que ce test ne soit théoriquement valable que dans le cas où une unique sinusoïde est présente dans les données, il est le plus souvent utilisé pour valider la détection de plusieurs raies dans le spectre.

2.1.4 Prise en compte de perturbations non *i.i.d.*

Nous avons jusqu'à présent considéré un modèle sinusoïdal bruité, pour lequel les perturbations ϵ_n étaient supposées additives, indépendantes et identiquement distribuées selon une loi gaussienne centrée $\mathcal{N}(0,\sigma_\epsilon^2)$. La définition des périodogrammes précédents est cependant adaptable à des perturbations non *i.i.d.*, qui est un cas fréquemment rencontré pour des séries temporelles astronomiques à cause de la variabilité temporelle des conditions d'observation (voir le § 1.2.2). Sous le formalisme du maximum de vraisemblance introduit au § 2.1.1, nous supposons pouvoir décrire sous forme probabiliste l'ensemble des perturbations $\boldsymbol{\epsilon} = [\epsilon_1,\ldots,\epsilon_n]^{\mathrm{t}}$ par une loi normale *vectorielle* $\epsilon \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$ où $\boldsymbol{\Sigma}$ est une matrice définie positive supposée connue, définissant la covariance des perturbations : $\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\dagger]$. La vraisemblance (2.5) prend alors la forme plus générale :

$$\mathcal{L}(\boldsymbol{y}; f,\boldsymbol{s}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}\|\boldsymbol{y} - \mathbf{R}(f)\boldsymbol{s}\|_{\boldsymbol{\Sigma}^{-1}}\right\}$$

$$\text{où } ||\mathbf{u}||_{\boldsymbol{\Sigma}^{-1}} \triangleq \boldsymbol{u}^\dagger\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \tag{2.10}$$

et la démarche réalisée au § 2.1.1 définissant le périodogramme de Lomb-Scargle reste valable en remplaçant la norme euclidienne $||\boldsymbol{u}||$ par[2] $||\mathbf{u}||_{\boldsymbol{\Sigma}^{-1}}$. De manière analogue, l'interprétation des périodogrammes en termes de projections, réalisée au § 2.1.1, peut être étendue en remplaçant le produit scalaire $\langle y|z\rangle$ par la version pondérée :

$$\langle y|z\rangle_{\boldsymbol{\Sigma}^{-1}} \triangleq [y_1,\ldots,y_n]\ \boldsymbol{\Sigma}^{-1}\ [z_1,\ldots,z_n]^{\mathrm{t}}.$$

La construction de familles orthonormales et de projections orthogonales par rapport à ce produit scalaire s'effectue alors de manière similaire et prend à présent en compte une description plus générale des perturbations $\boldsymbol{\epsilon}$. Un cas particulier classique est celui de perturbations *indépendantes* mais de rapports signal sur bruit différents, rencontré par exemple lors d'observations sur de longues durées ou pour les campagnes d'observations multi-sites (voir le § 1.2.2). Dans ce cas, on aura $\epsilon_n \sim \mathcal{N}(0,\sigma_n^2)$ et $\boldsymbol{\Sigma}$ diagonale d'éléments σ_n^2, où σ_n correspond au niveau de bruit associé à la $n^{\text{ème}}$ mesure. Le produit scalaire correspondant s'écrit alors sous la forme pondérée : $\langle y|z\rangle_{\boldsymbol{\Sigma}^{-1}} = \sum_{n=1}^{N} \frac{y_n z_n}{\sigma_n^2}$, et l'expression généralisée du périodogramme de Lomb-Scargle des données $\{t_n,y_n\}$ s'écrit :

$$P_{\text{LS}}(f) \triangleq \frac{1}{2}\left\{\frac{\left[\sum_n y_n \cos 2\pi f(t_n-\tau)/\sigma_n\right]^2}{\sum_n \cos^2 2\pi f(t_n-\tau)/\sigma_n^2} + \frac{\left[\sum_n y_n \sin 2\pi f(t_n-\tau)/\sigma_n\right]^2}{\sum_n \sin^2 2\pi f(t_n-\tau)/\sigma_n^2}\right\}$$

$$\text{en choisissant } \tau = \frac{1}{4\pi f}\tan^{-1}\frac{\sum_n \sin 4\pi f t_n/\sigma_n}{\sum_n \cos 4\pi f t_n/\sigma_n}.$$

La DCDFT et le périodogramme à date compensée se construisent de manière similaire en considérant le produit scalaire pondéré dans les équations du processus d'orthonormalisation.

2.2 Déconvolution itérative : méthodes CLEAN

Les algorithmes de type CLEAN abordent l'identification des paramètres du modèle (2.1) sous l'angle de la *déconvolution* du spectre de Fourier des données. L'origine de cette approche remonte

2. La matrice $\boldsymbol{\Sigma}^{-1}$ étant supposée définie positive, $||\mathbf{u}||_{\boldsymbol{\Sigma}^{-1}}$ définit bien une norme sur l'espace vectoriel des signaux échantillonnés aux instants t_n.

aux travaux de Högbom [1974] en imagerie radio-astronomique, où l'on cherche à reconstruire un *objet*, supposé constitué de plusieurs sources de lumière ponctuelles, à partir de son *image* qui est une version convoluée de ces sources par la réponse impulsionnelle de l'instrument d'observation. L'application de cette démarche pour l'analyse spectrale de séries temporelles, ainsi que l'introduction de certains raffinements, a été formalisée par Roberts *et al.* [1987], même si le schéma général apparaît déjà chez Ferraz-Mello [1981].

2.2.1 Principe général

L'influence d'un échantillonnage irrégulier au niveau de la fenêtre spectrale a été décrite au § 1.3.3. Si $Y_e(f)$ est le spectre de Fourier des données, on a :

$$Y_e(f) = Y(f) \star W_e(f), \text{ avec } W_e(f) = \mathcal{TF}\{w_e(t)\} = \sum_{n=1}^{N} e^{j2\pi f t_n}. \tag{2.11}$$

En particulier, nous avons vu que la présence de trous périodiques dans les données provoque de forts lobes secondaires dans la fenêtre $W_e(f)$ et que la superposition du motif de $W_e(f)$ autour de chaque raie spectrale génère de nombreux pics indésirables dans le spectre de Fourier observé. Le principe des algorithmes CLEAN est d'extraire de manière itérative les pics du spectre de Fourier, en « nettoyant » à chaque itération ce dernier de la contribution de la fenêtre spectrale associée à chaque pic. La base de cette méthode est alors dans le schéma suivant, où $R^k(f)$ désigne le spectre *résiduel* après k itérations de l'algorithme :

1. $k = 0$: initialiser $R^0(f) = Y_e(f)$
2. Itération k : identifier la fréquence f_k du maximum de $|R^{k-1}(f)|$ comme une des fréquences du modèle multi-sinusoïdal (2.1) et les amplitudes (a_k,b_k) associées.
3. Retirer la contribution spectrale correspondante : en notant $\rho_k\delta(f - f_k) + \rho_k^*\delta(f + f_k)$ le spectre théorique de la sinusoïde identifiée[3], la contribution sur le spectre observé de cette composante vaut $\rho_k W_e(f - f_k) + \rho_k^* W_e(f + f_k)$ et le spectre résiduel après l'itération k s'écrit :
$$R^k(f) = R^{k-1}(f) - \left[\rho_k W_e(f - f_k) + \rho_k^* W_e(f + f_k)\right].$$
4. Retour en 2 jusqu'à ce qu'un critère d'arrêt à définir soit vérifié.

De manière idéale, chaque étape correspond à l'identification d'une sinusoïde du modèle (2.1) ; Schwartz [1978] montre ainsi, dans un contexte très général, que cette procédure débouche sur *un* minimum du critère des moindres carrés de type (2.2) associé à un modèle multi-sinusoïdal d'ordre K *fixé*. Néanmoins, nous avons vu au début de ce chapitre qu'un tel critère est fortement multimodal et que sa minimisation pouvait rester bloquée vers un minimum local. La procédure itérative décrite ci-dessus peut donc être vue comme une minimisation du critère (2.2) privilégiant la recherche de solutions *a priori* satisfaisantes, en ce sens qu'elle est spécifiquement conçue pour éviter les minima locaux correspondant aux faux pics du spectre de Fourier initial : ceux-ci sont itérativement retirés en même temps que la raie spectrale qui en est la cause.

3. soit $\rho_k = (a_k - jb_k)/2$

2.2.2 Nécessité d'une stabilisation

Un principe inhérent à la démarche de CLEAN exposée ci-dessus repose sur l'hypothèse d'une identification correcte, à chaque itération, des paramètres de la sinusoïde correspondant au maximum du module du spectre de Fourier $|R^{k-1}(f)|$. Comme nous l'avons vu précédemment, ce point est en lui-même problématique. En particulier, dans le cas d'un modèle multi-sinusoïdal, nous avons montré au premier chapitre (§ 1.3.3) que l'interaction entre les différentes raies spectrales perturbe l'estimation, par le spectre de Fourier, des paramètres de la raie de plus forte amplitude qui doit correspondre, en principe, au maximum de $|R^{k-1}(f)|$. Ainsi, le maximum de $|Y_e(f)|$ peut ne pas correspondre à la position exacte de la fréquence d'une sinusoïde : la figure 1.12 en est un exemple flagrant. Dans ce cas, l'utilisation telle quelle de l'algorithme précédent soustraira au spectre de Fourier observé une composante inexistante (ou au moins inexacte), produisant irréversiblement la propagation d'erreurs dans le modèle estimé. Par ailleurs, nous avons vu au § 2.1.2 que l'estimation de l'*amplitude* d'une sinusoïde à partir de la valeur du maximum du spectre de Fourier (ou du périodogramme de Schuster) était également perturbée par l'irrégularité de l'échantillonnage.

Introduction d'un *clean factor*

Afin de stabiliser la procédure et d'éviter la propagation en chaîne d'erreurs de modélisation, Roberts *et al.* [1987] ont proposé, à l'étape 3 de l'algorithme précédent, de ne retirer qu'une fraction g de la contribution de la fréquence f_k estimée à l'étape 2. Le paramètre g, appelé *clean factor*, doit vérifier $0 < g < 2$ pour assurer la stabilité de l'algorithme, mais est en pratique fixé entre 0.1 et 1. Cette astuce permet de corriger l'erreur éventuelle effectuée sur l'estimation de la fréquence f_k et de l'amplitude correspondante ρ_k au fil des itérations suivantes : plus g est faible, plus la partie retirée du spectre de Fourier est faible, permettant d'affiner l'estimation de la fréquence d'une même raie au fil des itérations. *A contrario*, une telle procédure nécessite un plus grand nombre d'itérations pour « nettoyer » le spectre observé de la contribution produite par une seule raie (et estimer les paramètres de celle-ci).

En pratique, les résultats de l'algorithme sont présentés sous la forme de la convolution des différentes composantes sinusoïdales extraites par le *clean beam*, qui correspond à une approximation gaussienne du lobe principal de la fenêtre spectrale, formant alors le *spectre nettoyé*. Cette opération a pour effet de lisser le spectre obtenu et de masquer les décalages induits par l'introduction du *clean factor* sur l'estimation de la fréquence d'une même raie. En effet, les composantes fréquentielles lui correspondant, extraites en plusieurs itérations, ne se trouvent pas toujours exactement à la même position. Néanmoins, une telle convolution présente l'inconvénient de limiter la résolution du spectre *nettoyé* à la largeur du lobe central de la fenêtre spectrale, *i.e.,* à la résolution du périodogramme[4], alors qu'en pratique une telle méthode peut s'avérer capable de détecter deux raies plus proches que cette limite. Il sera donc également important de représenter de manière « brute » les différentes composantes extraites. Enfin, afin de contrôler la procédure, et notamment la validité de la condition d'arrêt, le spectre résiduel final y est généralement superposé.

4. Nous avons vu au cours du premier chapitre que cette largeur était de l'ordre de $2/T$, où T est la durée totale couverte par les observations.

Choix d'une condition d'arrêt

La définition d'un critère d'arrêt de la procédure itérative CLEAN est un de ses points délicats. Choisir d'arrêter l'extraction de composantes à l'itération K revient à considérer que l'on a extrait toutes les raies contenues dans le spectre des données, et que le spectre résiduel $R^K(f)$ n'est dû qu'au bruit et à l'effet convolutif de la fenêtre spectrale. Nous retrouvons ici le genre de tests statistiques du § 2.1.3, cherchant à déterminer si un spectre donné est uniquement dû aux perturbations ou s'il traduit la présence d'une raie, et les ambiguïtés interprétatives qui en résultent. Une solution *ad hoc* consiste à comparer, à chaque itération, le maximum du spectre résiduel à un seuil μ en deçà duquel tout pic dans le spectre de Fourier est considéré comme dû au bruit. De nombreux travaux donnent d'ailleurs à ce seuil une interprétation en termes de *niveau de confiance* qu'il convient ici d'éclaircir. L'approche empirique de Breger *et al.* [1993] propose ainsi de ne retenir le pic maximal à l'itération k que si son amplitude $|R^{k-1}(f_k)|$ est supérieure à 4 fois le *niveau de bruit*, qui correspond à l'amplitude moyenne du spectre résiduel *blanchi*, c'est-à-dire après extraction de la composante sinusoïdale correspondante en f_k :

$$\text{un pic de } R^{k-1}(f) \text{ en } f_k \text{ est significatif si } \frac{|R^{k-1}(f_k)|}{\overline{R^k(f)}} \geq 4.$$

Le complément apporté par Kuschnig *et al.* [1997] compare, *pour un échantillonnage donné* et un grand nombre de réalisations d'un bruit blanc centré $\boldsymbol{\epsilon}$, l'amplitude maximale du spectre de Fourier de $\boldsymbol{\epsilon}$ à sa valeur moyenne. Les auteurs fondent alors un critère de décision sur l'observation statistique que 99.9% des réalisations (respectivement, 99%) donnent lieu à des rapports de puissance inférieurs à 4 (respectivement, 3.6). Un tel critère permet ainsi d'affirmer qu'un pic dans le spectre résiduel ayant une puissance supérieure à 4 fois la puissance moyenne du spectre *blanchi* correspond à la détection d'une raie avec un niveau de confiance de 99.9% :

$$\text{un pic de } R^{k-1}(f) \text{ en } f_k \text{ est significatif à 99.9\% (resp. 99\%) si } \frac{|R^{k-1}(f_k)|^2}{\overline{|R^k(f)|^2}} \geq 4 \text{ (resp. 3.6).}$$

Il ne faut cependant pas oublier le sens bien spécifique que prend ici le terme de niveau de confiance qui, fondé sur des considérations plutôt empiriques, n'est pas à proprement parler un indicateur de l'incertitude associée à la détection éventuelle d'une raie. En particulier, nous avons vu au long du § 2.1 que l'interprétation de l'amplitude du spectre de Fourier et des différents périodogrammes est un point délicat si l'échantillonnage est irrégulier. Par ailleurs, l'estimation du niveau de bruit moyen dans le spectre de Fourier est un problème délicat et l'approche ci-dessus suppose, de manière analogue à la méthodologie de CLEAN, que le spectre résiduel après extraction de la composante testée en f_k (*blanchi*) est celui d'un bruit blanc (voir le § 2.2.4 ci-dessous pour un exposé des limites de cette hypothèse).

Exemple de résultat

La figure 2.3 présente l'estimation obtenue par CLEAN sur un signal composé de cinq sinusoïdes échantillonnées irrégulièrement et « suffisamment » espacées, auquel est ajouté un bruit blanc gaussien tel que le rapport signal sur bruit est de 10 dB où, pour un signal composé de sinusoïdes d'amplitudes (a_k,b_k), le rapport signal sur bruit est défini par :

$$RSB_{\text{dB}} = 10 \log \frac{\sum_k a_k^2 + b_k^2}{2\sigma_\epsilon^2} \tag{2.12}$$

où σ_ϵ^2 est la variance du bruit. Ce même signal relativement simple sera repris tout au long de ce manuscrit pour illustrer le fonctionnement des différentes méthodes d'estimation et les algorithmes associés ; nous le décrirons plus en détail au § 2.4 à la fin de ce chapitre. Dans ces simulations, l'algorithme est arrêté lorsque le maximum du spectre résiduel ne dépasse pas un certain pourcentage μ du maximum du spectre de Fourier des données, avec $\mu = 5\%$ et $\mu = 10\%$. La convolution des raies théoriques par la fenêtre spectrale rend le spectre de Fourier des données difficilement interprétable directement et l'application de CLEAN permet de retrouver les cinq raies. mais nécessite un réglage adéquat des deux paramètres que sont le *clean factor* et la condition d'arrêt. En effet, l'application de CLEAN sans *clean factor* provoque la propagation d'erreurs qui se traduit par l'apparition de faux pics. L'introduction d'un facteur $g = 0.2$ permet de contrôler ce problème : les faux pics sont désormais noyés dans le spectre résiduel. Néanmoins, en poussant trop loin l'algorithme, de nombreux artefacts apparaissent le long de l'axe fréquentiel.

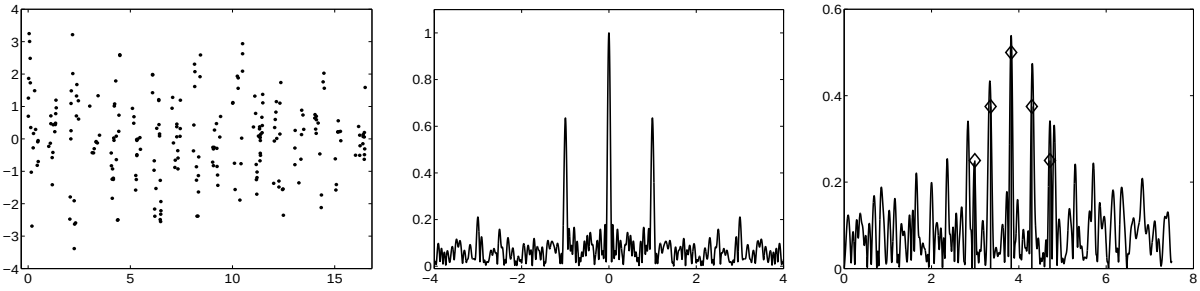
2.2.3 Extensions

Notons encore la possibilité d'inclure diverses variations agrémentant le schéma général de l'algorithme CLEAN. Ainsi, la version CLEANEST proposée par Foster [1995] aborde le même problème d'identification itérative des paramètres d'un modèle multi-sinusoïdal, mais en termes de projections sur des familles de fonctions sinusoïdales dont la dimension augmente au fil des itérations. Par analogie avec l'interprétation géométrique des différents périodogrammes du § 2.1.1, cette approche conduit à l'usage de la DCDFT plutôt que du spectre de Fourier, et ne s'interprète donc plus exactement en termes de déconvolution, mais repose sur un principe analogue : si à l'itération k correspond un modèle multi-sinusoïdal de k fréquences[5] $(f_1, \ldots, f_k)$, l'itération suivante recherchera un modèle à $k+1$ fréquences $(f'_1, \ldots, f'_k, f'_{k+1})$, en autorisant des petites variations des fréquences estimées à l'itération k ($f'_i \simeq f_{i\,,\,i=1\ldots k}$) au moyen d'une procédure d'*optimisation locale*. Par rapport à l'algorithme CLEAN initial, la remise en cause, à chaque itération, des estimées de l'itération précédente permet ainsi d'affiner l'estimation des raies spectrales. Une méthode similaire est mise en œuvre dans le programme de recherche de périodes PERIOD04 [Lenz et Breger, 2005] qui sera utilisé par la suite, notamment en application aux données réelles du chapitre 5.

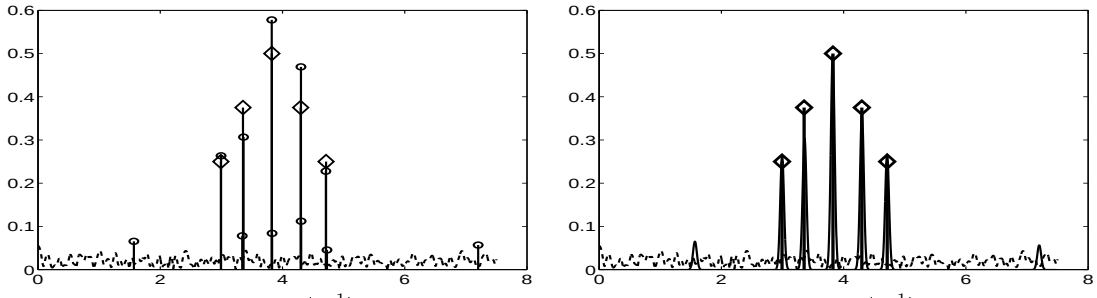
2.2.4 Conclusion

La méthodologie des algorithmes de type CLEAN procure des outils dont l'efficacité s'est révélée incontestable pour la recherche de périodes dans des données à échantillonnage irrégulier, et sont à juste titre un des outils favoris des astrophysiciens pour l'analyse de séries temporelles. L'évaluation des spectres de Fourier et résiduels, possible en tout point de l'axe fréquentiel, permet en particulier d'envisager une très grande précision sur l'estimation des fréquences. L'obtention de résultats satisfaisants reste néanmoins conditionnée par un réglage souvent *ad hoc* des paramètres de contrôle g et μ : la procédure doit être *supervisée* par l'utilisateur. Par ailleurs, l'absence d'une totale rigueur théorique (ainsi que la sous-optimalité de la procédure d'optimisation dans ses raffinements de type CLEANEST) peut compromettre son usage. En particulier, l'utilisation du maximum du spectre de Fourier (ou de la DCDFT) pour localiser une raie n'est théoriquement valable que dans le cas d'un modèle mono-sinusoïdal (nous l'avons montré au long de la partie 2.1) ; cette procédure recherche pourtant, après extraction de la première raie f_1,

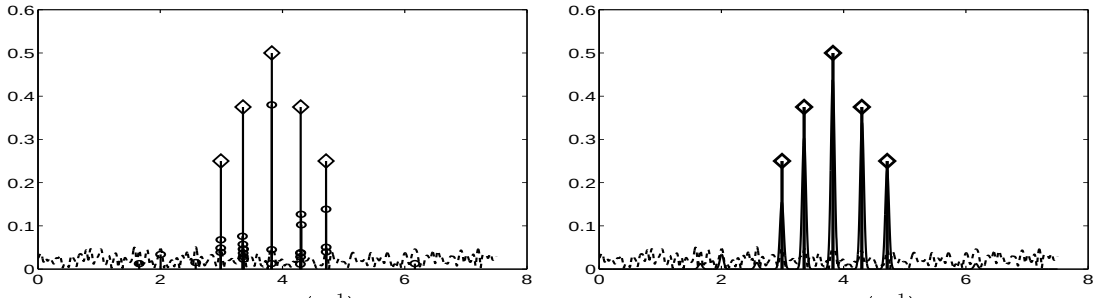
5. Pour simplifier l'explication, nous supposons ici que l'algorithme n'utilise pas la subtilité du *clean factor*. Néanmoins, la démarche reste identique dans le cas où un gain $g < 1$ est introduit.



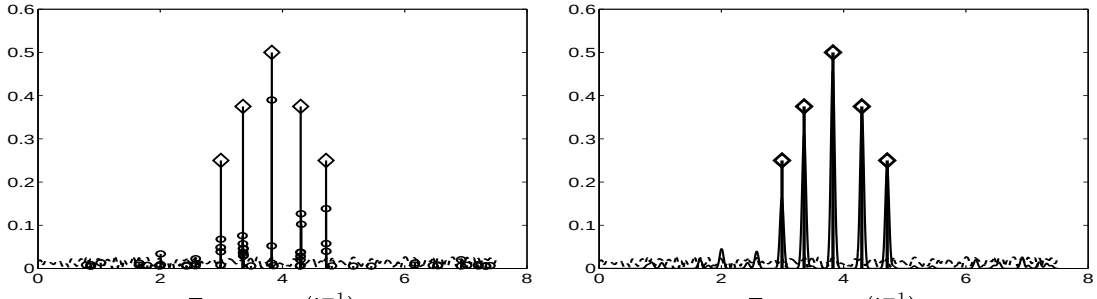
De gauche à droite, signal composé de cinq sinusoïdes à échantillonnage irrégulier et troué, fenêtre spectrale correspondante $|W_e(f)|$ et spectre de Fourier des données $|Y(f)|$.



$g = 1$, $\mu = 10\%$.



$g = 0.2$, $\mu = 10\%$.



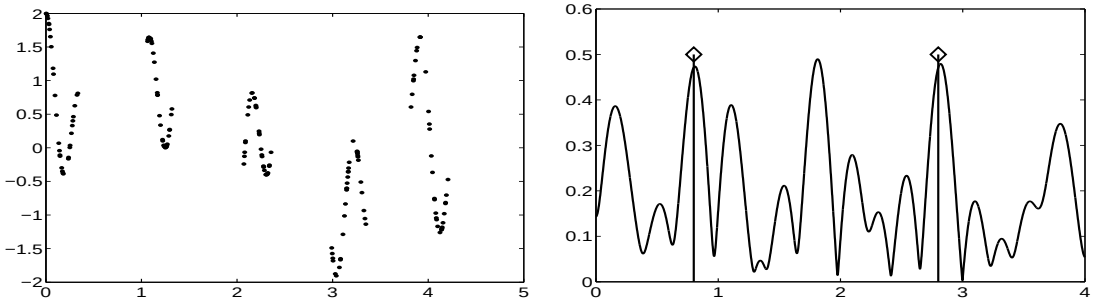
$g = 0.2$, $\mu = 5\%$

Estimation produite par CLEAN : à gauche, les composantes extraites ($\circ$) et à droite, leur convolution *a posteriori* par le *clean beam*. Les pointillés représentent le spectre résiduel $|R^k(f)|$ et les $\diamond$ les raies théoriques.

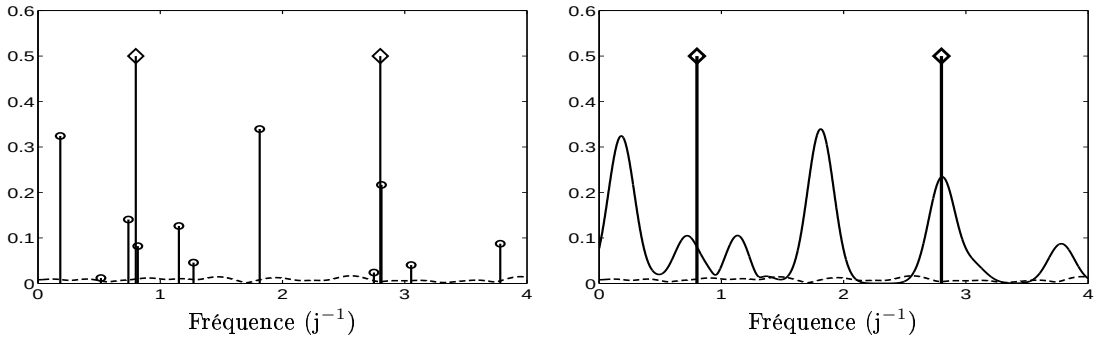
FIG. 2.3 – *Résultats de l'algorithme* CLEAN *sur un signal multi-sinusoïdal, pour plusieurs valeurs du facteur g et de la condition d'arrêt μ : l'algorithme est arrêté lorsque le maximum du spectre résiduel est inférieur à μ% de l'amplitude maximale du spectre de Fourier des données.*

une nouvelle raie f_2 dans le spectre résiduel R^1, alors que l'estimation de la première fait implicitement l'hypothèse que R^1 ne contient rien d'autre que du bruit. Lorsque les raies sont bien séparées en fréquence, cette approximation se révèle satisfaisante en pratique ; cependant, dans le cas d'espacements critiques des raies, la solution fournie par CLEAN peut se révéler délicate à interpréter. C'est le cas sur l'exemple « pathologique » introduit au § 1.3.3, figure 1.12, que nous avons repris en figure 2.4 : le maximum du spectre de Fourier (mais aussi de la DCDFT) est dû à la superposition de deux lobes secondaires et ne correspond aucunement à une raie présente dans les données, débouchant sur une mauvaise initialisation de CLEAN. Les résultats de la figure 2.4 illustrent ce point, mais montrent également qu'une utilisation adéquate du *clean factor* g permet de compenser la mauvaise initialisation et de retrouver les deux véritables raies. Néanmoins, même dans le cas où g est réglé de manière favorable, la solution présentera nécessairement une composante inexacte correspondant à l'initialisation défectueuse de l'algorithme.

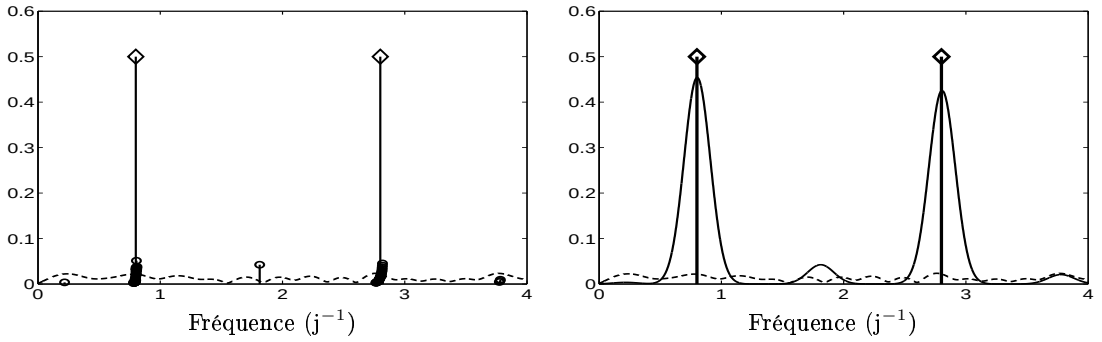
Par ailleurs, l'utilisation d'une telle démarche sur des données perturbées par un spectre continu à basse fréquence, telles les données présentées au § 1.2.2, semble difficilement réalisable puisque le spectre de Fourier observé est dominé par les composantes à basse fréquence ; en pratique, l'application brute de CLEAN sur ce type de données a produit des résultats aberrants, sous la forme conjointe de faux pics le long de l'axe fréquentiel et de non détections de vraies raies. Cette limitation est d'ordre plus général et est intrinsèque à la *modélisation* des données : bien que l'on s'intéresse à la recherche de spectres de raies, la nature inhomogène des perturbations rend une modélisation paramétrique sous la forme d'une somme de sinusoïdes noyées dans un bruit blanc gaussien trop contraignante et peu adaptable à des données plus complexes.



Signal composé de deux sinusoïdes (à gauche) et module du spectre de Fourier associé (à droite). Les $\diamond$ représentent les positions des vraies raies spectrales.



$g = 0.8$: l'erreur due à l'initialisation se propage et fournit une solution inacceptable.



$g = 0.1$: la procédure est stabilisée par une valeur plus faible du *clean factor*.

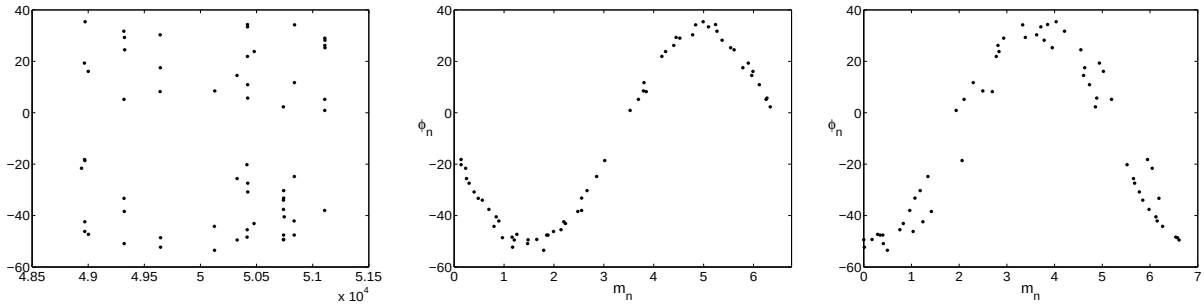
Estimation produite par CLEAN : à gauche, les composantes extraites ($\circ$) et à droite, leur convolution *a posteriori* par le *clean beam*. Les pointillés représentent le spectre résiduel et les $\diamond$ les raies théoriques.

FIG. 2.4 – *Résultats de l'algorithme* CLEAN *sur le signal « pathologique » de la figure 1.12, pour deux valeurs différentes du clean factor. L'algorithme est arrêté lorsque le maximum du spectre résiduel est inférieur à* $\mu = 5\%$ *de l'amplitude maximale du spectre de Fourier des données.*

2.3 Problèmes à une période : techniques spécifiques

Nous mentionnons ici l'existence de méthodes spécifiquement adaptées au problème de l'estimation d'*une* périodicité à partir de séries temporelles irrégulièrement échantillonnées. Cette situation est notamment rencontrée lors de l'étude de systèmes d'étoiles doubles, pour laquelle les données recueillies, souvent peu nombreuses, couvrent une période temporelle très large. De par le faible nombre de mesures disponibles (parfois de l'ordre de la dizaine de points), l'utilisation du spectre de Fourier ou des différents périodogrammes semble compromise, d'autant plus que les variations périodiques à l'étude ne sont que rarement sinusoïdales, comme nous l'avons vu au § 1.1.2.

L'idée à la base de ce type de méthodes est de pallier la pauvreté de la couverture temporelle en ramenant les données dans un intervalle de temps où elles seront plus lisibles. Puisque l'on recherche des mouvements périodiques, le principe consiste alors à essayer plusieurs valeurs de test pour une période candidate et à analyser le *diagramme de phase* correspondant, qui est une représentation des données ramenées dans un intervalle de temps correspondant à cette période : en considérant les données $(t_n,y_n)_{n=1...N}$ et la période test T, le diagramme de phase correspondant est la représentation des points (t'_n,y_n), où $t'_n = t_n \bmod T$. Ainsi, lorsque T correspond à la vraie période, les points du diagramme de phase correspondant sont répartis selon le motif périodisé, présentant une courbe de variations « douces ». Un exemple est présenté sur la figure 2.5, appliqué aux variations de vitesse radiale du système multiple HD 7119 [Carquillat *et al.*, 2002] : le signal représenté correspond au premier corps du système et comprend 60 mesures s'étalant sur une période de 2000 jours. La période principale du système[6] a été identifiée à $T_0 = 6.761$ jours.



Courbe de vitesse radiale de HD 7119 — Diagramme de phase $T = T_0$ — Diagramme de phase $T = T_0 + 2.10^{-3}$

FIG. 2.5 – *Exemples de diagrammes de phase de la courbe de vitesse radiale de HD 7119, pour deux périodes test : une faible variation de la période test par rapport à sa valeur estimée provoque un diagramme de phase beaucoup moins régulier.*

Différentes méthodes ont ainsi été proposées, cherchant à maximiser, en le paramètre T, un critère de douceur sur le diagramme de phase correspondant, noté par la suite $(m_n,\phi_n)_{n=1...N}$ où m_n représentent les instants t'_n introduits ci-dessus et réordonnées dans l'ordre croissant,

6. Carquillat *et al.* [2002] établissent qu'il s'agit en réalité d'un système triple et la courbe de la figure 2.5 présente donc deux périodicités, la seconde de période bien plus longue (1700 jours) et d'amplitude bien plus faible.

et ϕ_n les *phases* correspondantes, *i.e.,* ϕ_n est l'amplitude de la mesure prise à l'instant représenté par m_n. Les méthodes de type *Phase Dispersion Minimization* [Lafler et Kinman, 1965] visent à minimiser les différences entre des observations de phases adjacentes, soit à minimiser un critère du type :

$$\Theta_{\text{PDM}} = \sum_n (\phi_n - \phi_{n+1})^2$$

dépendant implicitement de la période de test T. La méthode *String-Length* proposée par Dworetsky [1983] minimise la longueur de la courbe formée par les points du diagramme de phase :

$$\Theta_{\text{SL}} = \sum_n \sqrt{(m_n - m_{n+1})^2 + (\phi_n - \phi_{n+1})^2}$$

en ayant préalablement effectué un changement d'échelle afin d'avoir les mêmes amplitudes de variation en abscisse et en ordonnée.

Étant spécifiquement développées pour des données pauvres et périodiques, ces méthodes peuvent s'avérer plus efficaces que les différents périodogrammes étudiés précédemment, *a fortiori* lorsque le motif périodique est fortement non-sinusoïdal. Néanmoins, la difficulté de dériver des propriétés théoriques satisfaisantes en restreignent quelque peu la portée.

2.4 Méthodologie adoptée : un problème inverse mal posé

Nous avons vu dans ce chapitre que l'identification directe d'un modèle paramétrique multi-sinusoïdal est un problème délicat : la vraisemblance correspondante est multimodale et l'interaction des différentes composantes spectrales peut déboucher sur la détection de fausses fréquences et ce phénomène est aggravé par la présence de trous périodiques dans l'échantillonnage des données. De plus, la modélisation multi-sinusoïdale peut s'avérer trop contraignante lors de la présence dans les séries temporelles d'un continuum à basses fréquences.

Afin de s'affranchir de ces difficultés, les différentes méthodes développées au cours de cette thèse considèrent l'estimation d'un nombre fini, arbitrairement grand, de composantes spectrales discrétisées sur une grille fréquentielle, et parmi lesquelles seules quelques unes prendront des valeurs non nulles. Nous considérons ainsi un modèle composé d'un nombre P de sinusoïdes de fréquences *fixées* entre 0 et une fréquence maximale $f_{\max}$, sur la grille $\mathcal{G} = \{\frac{k}{P} f_{\max}\}_{k=0\ldots P}$. Pour des données irrégulièrement échantillonnées, nous avons vu au § 1.3.1 que la fréquence maximale correspondant à une plage fréquentielle exempte d'alias était poussée à des valeurs bien plus grandes que dans le cas de l'échantillonnage régulier : de ce fait, le paramètre $f_{\max}$ sera choisi en fonction du contenu spectral *a priori* recherché dans les observations et non au sens d'une limite imposée par le théorème de Shannon. Le problème devient alors celui de l'estimation des amplitudes spectrales correspondantes, que l'on peut écrire sous la forme :

$$y_n = \sum_{k=0}^{P} a_k \cos(2\pi f_k t_n) + b_k \sin(2\pi f_k t_n) + \epsilon_n, \;\; n = 1 \ldots N \quad \text{avec } f_k \triangleq \frac{k}{P} f_{\max}$$

où, comme précédemment, y_n représente la mesure effectuée à l'instant t_n et le terme ϵ_n rend compte des perturbations dues au bruit de mesure ou encore à une erreur de modélisation. En notant $\boldsymbol{s}_k = [a_k, b_k]$ et en formant le vecteur à $(2P+2)$ inconnues réelles $\boldsymbol{s} = [\boldsymbol{s}_0, \ldots, \boldsymbol{s}_P]^T$, on peut encore écrire sous forme matricielle :

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{s} + \boldsymbol{\epsilon} \tag{2.13}$$

où $\boldsymbol{y} = [y_1, \ldots, y_N]^{\mathrm{t}}$ regroupe les données, $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_N]^{\mathrm{t}}$ les perturbations et la matrice $\mathbf{H}$ de taille $N \times (2P+2)$ est constituée des éléments $\mathbf{H}_{n,2k} = \cos(2\pi f_k t_n)$ et $\mathbf{H}_{n,2k+1} = \sin(2\pi f_k t_n)$ pour $k = 0 \ldots P$ et $n = 1 \ldots N$.

À la différence du modèle multi-sinusoïdal (2.1) où les inconnues sont le nombre K de sinusoïdes, les K fréquences et les $2K$ amplitudes correspondantes, cette modélisation considère des fréquences fixées et un ordre P connu. Le problème est alors celui de l'estimation des seules amplitudes spectrales $\{a_k, b_k\}_{k=0\ldots P}$ à partir des données $\{y_n\}_{n=1\ldots N}$: le modèle est maintenant *linéaire*, mais le nombre de paramètres à estimer est beaucoup plus grand : $2P+2 \gg 3K+1$. L'intérêt de la formulation du problème inverse (2.13) réside essentiellement dans la simplicité du modèle, puisque la relation *linéaire* entre les inconnues $\boldsymbol{s}$ et les données $\boldsymbol{y}$ ouvre le problème au champ d'application de nombreuses méthodes d'estimation. Par ailleurs, cette modélisation, moins stricte que celle fournie par une approche paramétrique, permet d'envisager un comportement plus robuste des estimateurs spectraux, par exemple face à des perturbations à basses fréquences comme celles présentées au § 1.2.2. Le prix à payer est évidemment en conséquence, et le gain en simplicité se fait en contrepartie de l'apparition d'indéterminations : l'objectif d'une analyse à haute résolution impose l'utilisation d'une grille fréquentielle $\mathcal{G}$ la plus fine possible, c'est-à-dire un nombre de paramètres (a_k,b_k) arbitrairement grand. De fait, le nombre de composantes spectrales à estimer sera supérieur au nombre de données : $2P+2 \gg N$. Le problème *exact* $\boldsymbol{y} = \mathbf{H}\boldsymbol{s}$ admet alors une infinité de solutions : le problème est dit sous-déterminé. Il s'agit alors de *régulariser* ce problème, c'est-à-dire de contraindre la solution $\boldsymbol{s}$ à vérifier certaines propriétés supplémentaires permettant de la définir de manière unique. Dans le cadre qui est le nôtre, nous recherchons exclusivement des spectres de raies : le vecteur d'amplitudes spectrales $\boldsymbol{s}$ doit donc revêtir un caractère *parcimonieux* ou *lacunaire*, c'est-à-dire être nul en dehors d'un nombre restreint de positions fréquentielles. La régularisation consiste alors ici en la prise en compte de cette propriété structurelle *a priori*, et peut être envisagée de différentes façons : c'est l'objet des deux chapitres suivants.

Notons encore que la formulation du problème d'analyse spectrale sous la forme du problème inverse (2.13) peut être adaptée au problème de la recherche d'un mouvement périodique non sinusoïdal. En effet, un signal périodique de fréquence f_0 peut être décomposé en série de Fourier, comme une somme (en théorie infinie) de sinusoïdes de fréquences multiples de f_0. L'écriture d'un modèle de type (2.13), pour lequel les fréquences discrétisées le sont cette fois sur une grille $\mathcal{G} = \{k.f_0\}_{k=0\ldots P}$, vise alors à la reconstruction du développement en série de Fourier du signal en question, tronqué aux P premières harmoniques. Le problème est alors celui de l'estimation des amplitudes associées et, bien entendu, de la fréquence fondamentale f_0 : ce n'est donc plus un problème linéaire. Néanmoins, la non-linéarité du modèle étant réduite à un seul paramètre, il est tout à fait envisageable de procéder à une telles estimation. Ce problème a d'ailleurs déjà été abordé dans ce cadre pour des applications musicales (voir par exemple [Godsill et Davy, 2002]).

Nous signalons pour terminer ce chapitre que cette approche du problème d'analyse spectrale est plus souvent rencontrée sous la forme de la restauration des amplitudes *complexes* x_k associées aux fréquences f_k :

$$\begin{aligned} y_n &= \sum_{k=-P}^{P} x_k \exp\left(j2\pi f_k t_n\right) + \epsilon_n, \;\; n = 1 \ldots N \\ \Leftrightarrow \quad \boldsymbol{y} &= \mathbf{W}\boldsymbol{x} + \epsilon \end{aligned} \tag{2.14}$$

où $\mathbf{W} = \{e^{j2\pi f_k t_n}\}_{k=-P\ldots P,\, n=1\ldots N}$ et $\boldsymbol{x} = [x_{-P}, \ldots, x_P]^{\mathrm{t}}$ contient les amplitudes spectrales à

estimer. Lorsque les données $\boldsymbol{y}$ sont à valeurs réelles, le vecteur d'amplitudes spectrales $\boldsymbol{x}$ est bien entendu à symétrie hermitienne : $x_{-k} = x_k^*$, les perturbations $\boldsymbol{\epsilon}$ sont également à valeurs réelles et les formulations (2.13) à $2P+2$ inconnues réelles et (2.14) à $P+1$ inconnues complexes – en imposant la contrainte d'hermiticité – sont équivalentes avec $\boldsymbol{s}_k = [a_k,b_k] = 2\,[\Re(x_k),\, -\Im(x_k)]$. Dans le cas d'un échantillonnage régulier sur N points ($t_n = nT_e$) et si la fréquence $f_{\max}$ est choisie à la limite imposée par le théorème de Shannon ($f_{\max} = 1/2T_e$), l'opérateur $\mathbf{W}$ correspondant à la reconstruction du même nombre N de composantes fréquentielles est carré et s'identifie à l'opérateur de transformée de Fourier discrète inverse : $\mathbf{W} = \{\exp j2\pi\frac{kn}{N}\}_{k,n}$. La structure de l'opérateur $\mathbf{W}$ dans le cas de l'échantillonnage irrégulier sera étudiée en détail au chapitre 3. Par la suite, nous utiliserons alternativement les deux formulations réelle et complexe. Plus précisément, le chapitre 3 consacré à la pénalisation fera usage de l'écriture des sinusoïdes sous la forme d'exponentielles complexes car nous verrons que les fonctions de pénalisation portent sur le *module* des amplitudes spectrales $|x_k| = \frac{1}{2}\sqrt{a_k^2 + b_k^2}$. La modélisation probabiliste Bernoulli-Gaussienne mise en place au chapitre 4 se révélera quant à elle plus appropriée à l'utilisation de variables réelles bidimensionnelles $\boldsymbol{s}_k = [a_k, b_k]$ afin de pouvoir expliciter des distributions de probabilité de paramètres réels.

Utilisation d'un signal test

Afin d'illustrer les différentes méthodes mises en œuvre dans la suite de ce manuscrit, nous les appliquerons à chaque fois au même signal synthétique que nous présentons ici. Il s'agit de de cinq sinusoïdes de fréquences $\nu_1, \ldots, \nu_5$ allant de 0.1 à 0.16 j^{-1} ;

$$y(t_n) = \sum_{j=1}^{5} \alpha_j \cos(2\pi\nu_j t_n) + \beta_j \sin(2\pi\nu_j t_n) + \epsilon_n,$$

noyées dans un bruit blanc gaussien $\epsilon \sim \mathcal{N}(0,\sigma_\epsilon^2)$ avec un rapport signal sur bruit de 10 dB, celui-ci étant défini par : $\mathrm{RSB}_{\mathrm{dB}} = 10\log_{10}\frac{P_y}{\sigma_\epsilon^2}$, avec $P_y = \frac{1}{2}\sum_{j=1}^{5}\alpha_j^2 + \beta_j^2$. Afin d'analyser la robustesse des procédures d'estimation face aux erreurs de modélisation, les fréquences sont choisies, sauf précision contraire, en dehors de la grille de reconstruction $\mathcal{G} = \frac{k}{P}f_{\max}$, où $f_{\max}$ sera par la suite fixé à 0.25 j^{-1}. L'échantillonnage couvre une durée de 15 jours et est à la fois irrégulier et troué, où les trous sont de période journalière et occultent environ la moitié de la période d'observation. Le nombre d'échantillons disponibles est de $N = 250$. Ce signal est représenté sur la figure 2.6 à gauche. En raison d'un fort pic secondaire dans la fenêtre spectrale (figure centrale), le spectre de Fourier des données présente de nombreux faux pics (figure de droite).

Notons que des résultats d'une qualité satisfaisante peuvent être obtenus sur ce problème par un algorithme de déconvolution de type CLEAN correctement supervisé. L'application à ce signal d'un algorithme CLEAN a d'ailleurs été présentée au § 2.2 : on pourra revoir la figure de la page 41. L'utilisation d'un tel signal servira donc plutôt, au cours des chapitres suivants, à illustrer les caractéristiques générales des estimateurs qui seront introduits ainsi que le comportement des différents algorithmes proposés.

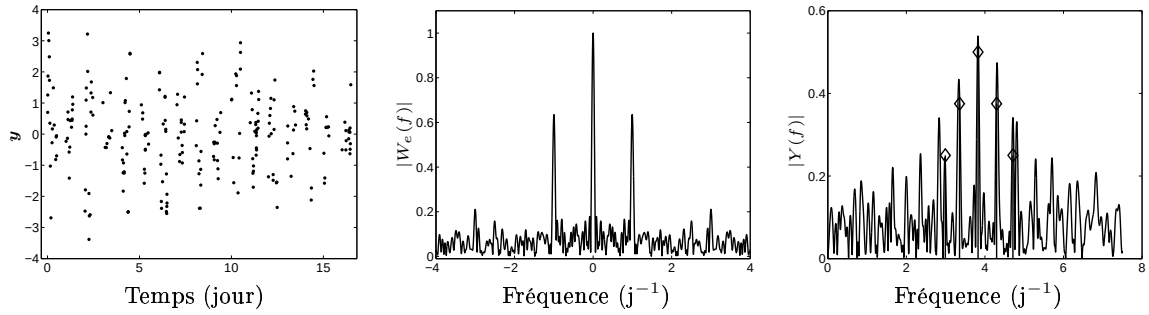


FIG. 2.6 – *De gauche à droite, signal composé de cinq sinusoïdes à échantillonnage irrégulier et troué, fenêtre spectrale correspondante $|W_e(f)|$ et spectre de Fourier des données $|Y(f)|$. Les $\diamond$ représentent les raies spectrales théoriques.*

Chapitre 3

Approche par pénalisation

Nous nous plaçons dans le cadre mis en place à la fin du chapitre précédent : la reconstruction d'un spectre de raies à partir de données irrégulièrement échantillonnées $\{t_n,y_n\}_{n=1...N}$ est abordée sous la forme de l'estimation d'un vecteur d'amplitudes spectrales correspondant aux fréquences discrétisées sur la grille $\mathcal{G} = \{f_k\}_{k=0...P} = \{\frac{k}{P} f_{\max}\}_{k=0...P}$. Nous considérons donc un modèle linéaire, qui s'écrit sous la forme matricielle :

$$\boldsymbol{y} = \mathbf{W}\boldsymbol{x} + \epsilon \tag{3.1}$$

où $\mathbf{W} = \{e^{j2\pi \frac{k}{P} f_{\max} t_n}\}_{k=-P...P,\, n=1...N}$, $\boldsymbol{y} = [y_1, \ldots, y_N]^{\mathrm{t}}$ regroupe les données, $\boldsymbol{x} = [x_{-P}, \ldots, x_P]^{\mathrm{t}}$ contient les amplitudes spectrales à estimer et $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_n]^{\mathrm{t}}$ les perturbations. Bien entendu, les données $\boldsymbol{y}$ étant à valeurs réelles, le spectre à reconstruire $\boldsymbol{x}$ possède la propriété de symétrie hermitienne : $x_{-k} = x_k^*$. Pour simplifier les notations cependant, nous nous plaçons pour l'instant sous la formulation générale ci-dessus à $2P+1$ inconnues complexes, les données $\boldsymbol{y}$ et les perturbations $\boldsymbol{\epsilon}$ étant à valeurs *a priori* complexes. La prise en compte de la symétrie hermitienne de $\boldsymbol{x}$ sera étudiée à la fin de ce chapitre.

Dans la perspective d'une analyse à haute résolution, nous considérons ici le cas sous-déterminé ($2P+1 > N$), pour lequel le problème exact $\boldsymbol{y} = \mathbf{W}\boldsymbol{x}$ admet une infinité de spectres $\boldsymbol{x}$ modélisant parfaitement les données. Une approche simple et immédiate consiste dans ce cas à considérer la solution du problème exact de norme minimale, dite encore inverse généralisée ou inverse à droite, définie de manière unique comme :

$$\begin{aligned} \widehat{\boldsymbol{x}}_{\mathrm{IG}} &\stackrel{\Delta}{=} \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_2 \text{ sous la contrainte } \boldsymbol{y} = \mathbf{W}\boldsymbol{x} \\ &= \mathbf{W}^\dagger (\mathbf{W}\mathbf{W}^\dagger)^{-1} \boldsymbol{y}. \end{aligned}$$

Nous verrons que l'inversion généralisée débouche sur un estimateur proche du spectre de Fourier des données et donc, conformément à l'étude menée au premier chapitre, ne fournit pas une solution acceptable en termes de spectres de raies.

Une manière plus fine d'imposer l'unicité de la solution a trait au domaine de la régularisation, au sens décrit par Tikhonov et Arsenin [1977], qui a connu un succès massif au cours des années 1990 [Demoment, 1989, Idier, 2001a] : l'approche dite *pénalisée* de la régularisation définit l'estimateur $\widehat{\boldsymbol{x}}$ comme le minimiseur d'un critère composite :

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} J(\boldsymbol{x}), \quad J(\boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|^2 + \lambda R(\boldsymbol{x}) \tag{3.2}$$

où la notation $\|\boldsymbol{u}\|$ correspond à la norme ℓ^2 du vecteur $\boldsymbol{u}$. On ne recherche donc plus une solution correspondant parfaitement aux données, mais un compromis relativement fidèle aux données d'une part (le terme $\|\boldsymbol{y} - \mathbf{W}\widehat{\boldsymbol{x}}\|^2$ doit être de faible valeur), mais aussi tel que le terme de pénalisation $R(\widehat{\boldsymbol{x}})$ soit faible. L'unicité d'un tel estimateur est alors garantie sous certaines conditions de régularité imposées à la fonction $R : \mathbb{C}^{2P+1} \to \mathbb{R}_+$; ce sera le cas en particulier de toutes les fonctions strictement convexes, qui assurent l'unicité du minimiseur du critère (3.2) ainsi que sa continuité par rapport aux données [Bouman et Sauer, 1993] et confèrent de ce fait au problème régularisé un caractère *bien posé* [Tikhonov et Arsenin, 1977].

Il s'agit donc de définir une fonction de pénalisation R traduisant des propriétés *a priori* attendues sur le spectre $\boldsymbol{x}$. Dans notre cas, la pénalisation doit refléter le fait que l'on recherche un nombre *faible* de sinusoïdes sur la grille $\mathcal{G}$, c'est-à-dire encourager la présence de zéros dans le vecteur d'amplitudes spectrales : $R(\boldsymbol{x})$ doit prendre des valeurs plus faibles pour des solutions $\boldsymbol{x}$ parcimonieuses. Le choix du paramètre λ permet alors de régler le compromis entre la fidélité aux données et la confiance en l'information *a priori* exprimée par la pénalisation. Le cas $\lambda = 0$ considère ainsi une fidélité parfaite aux données ; à l'inverse, le cas $\lambda \to +\infty$ ne prend en compte que l'information *a priori* apportée par la pénalisation sans tenir compte des données : la solution régularisée est alors identiquement nulle. Dans le cadre bayésien et pour des perturbations $\boldsymbol{\epsilon}$ centrées, indépendantes et identiquement distribuées selon une loi gaussienne *circulaire* [1] $\mathcal{N}_C(0,2\sigma_\epsilon^2)$, le terme de moindres carrés $\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|^2$ correspond, à un facteur additif près, à l'anti log-vraisemblance du modèle (3.1) et $\widehat{\boldsymbol{x}}$ est l'estimateur du maximum *a posteriori* correspondant à une loi *a priori* de la forme $p(\boldsymbol{x}) \propto \exp -\frac{\lambda}{\sigma_\epsilon^2} R(\boldsymbol{x})$ [Idier, 2001a].

Si la formulation (3.2) pour l'analyse spectrale apparaît chez Cabrera et Parks [1991], son application à l'estimation de spectres de raies remonte aux travaux de Sacchi *et al.* [1998], qui ont été étendus par Ciuciu *et al.* [2001] pour la prise en compte de formes spectrales plus complètes. Ces deux contributions proposent des fonctions de pénalisation adaptées à la recherche de spectres de raies ainsi que des algorithmes d'optimisation spécifiques, dont l'efficacité repose essentiellement sur le caractère régulier de l'échantillonnage des données [2]. Lors de son stage de DEA au Laboratoire d'Astrophysique de l'Observatoire Midi-Pyrénées, L. Jahan [2001] a montré leur possible adaptation au cas de l'échantillonnage régulier à données manquantes. L'objectif principal de ce chapitre est la proposition d'algorithmes performants pour le calcul de la solution régularisée, exploitant dans la mesure du possible les propriétés des opérateurs mis en jeu dans le cas général de l'échantillonnage irrégulier, ainsi que les spécificités des fonctions de pénalisation employées.

Nous étudions tout d'abord au § 3.1 les propriétés structurelles des opérateurs mis en jeu – l'opérateur du problème $\mathbf{W}$ ainsi que les matrices $\mathbf{W}\mathbf{W}^\dagger$ et $\mathbf{W}^\dagger\mathbf{W}$, dont l'importance apparaîtra par la suite – en fonction de la forme de l'échantillonnage. Dans le cas de mesures régulièrement espacées, ces propriétés sont à la base de l'interprétation de la pénalisation *quadratique* en termes de fenêtrage des données [Giovannelli et Idier, 2001]. Nous montrons au § 3.2 l'inadéquation d'une telle pénalisation pour la recherche de raies spectrales quel que soit le schéma d'échantillonnage et étudions deux types de pénalisation mieux adaptés à notre problème, chacun d'eux débouchant sur des propriétés différentes au niveau des algorithmes d'optimisation correspondants. Dans un premier temps, nous considérons une pénalisation de type $\ell^2\ell^1$ comme la *branche hy-*

1. Pour des perturbations ϵ_n à valeurs *a priori* complexes, la loi gaussienne circulaire $\mathcal{N}_C(0,2\sigma_\epsilon^2)$ sur ϵ_n correspond à des lois gaussiennes $\mathcal{N}(0,\sigma_\epsilon^2)$ sur les parties réelle et imaginaire de ϵ_n, supposées indépendantes : $p(\epsilon_n) = \frac{1}{2\pi\sigma_\epsilon^2} \exp -\frac{1}{2\sigma_\epsilon^2}|\epsilon_n|^2$.

2. Les premiers auteurs envisagent l'extension au cas de l'échantillonnage irrégulier : nous y reviendrons au § 3.4.

perbolique, proposée avec succès par Ciuciu *et al.* [2001] pour l'estimation de spectres de raies à partir de données régulièrement échantillonnées. Nous proposons ensuite l'emploi de la pénalisation par la *norme* ℓ^1, qui présente de solides propriétés théoriques en termes de parcimonie apportée sur la solution [Fuchs, 2004] mais n'a à notre connaissance jamais été associée à la formulation (3.1) du problème d'analyse spectrale. Des simulations montrent alors la nature très proche des estimateurs obtenus par ces deux pénalisations.

Le § 3.3 est consacré au choix des hyperparamètres intervenant dans l'écriture du critère pénalisé (3.2). Alors que le réglage d'un paramètre supplémentaire dans le cas de la pénalisation hyperbolique ne pose pas de problème particulier, la sélection d'une valeur satisfaisante du paramètre de régularisation λ est un point plus sensible dans le but d'obtenir une estimation *non supervisée*. Si ce paramètre peut être interprété en termes de rapport signal sur bruit lorsque les observations peuvent être accompagnées d'une estimation de ce dernier, nous décrivons l'utilisation d'une méthode empirique basée sur la « courbe en L » [Hansen, 1992] qui permet de fixer approximativement une valeur adéquate de λ ; nous montrons cependant sur des simulations la sensibilité de l'estimation au réglage de λ pour les pénalisations étudiées, qui ne peut par conséquent pas s'affranchir totalement du contrôle de l'utilisateur.

Plusieurs algorithmes d'optimisation sont proposés au § 3.4 dans le cas de la pénalisation hyperbolique, pour laquelle le critère à minimiser est strictement convexe. Nous montrons en particulier que les spécificités des opérateurs établies au § 3.1 peuvent être prises en compte par des algorithmes de type gradient, de manière différente en fonction de l'échantillonnage. La pénalisation ℓ^1 produit quant à elle un critère non continûment différentiable, pour lequel une stratégie de type gradient n'est plus envisageable. La norme ℓ^1 présente cependant d'autres caractéristiques intéressantes sur le plan algorithmique, que nous décrivons et exploitons au § 3.5, qui constitue notre principale contribution méthodologique pour cette approche. Nous utilisons notamment une version à paramètres complexes de l'algorithme IRLS (*Iterative Reweighted Least Squares*) proposée pour des variables réelles par Fuchs [2005] dans le cas de la pénalisation ℓ^1 et dont la mise en œuvre peut être réalisée à moindre coût. La convergence de cet algorithme n'a pas encore été démontrée à ce jour ; aussi, nous établissons un test complémentaire permettant de vérifier *a posteriori* l'optimalité de la solution obtenue. Nous proposons également un algorithme de type Gauss-Seidel effectuant des minimisations successives variable par variable, pour lequel chaque étape de minimisation est explicite et immédiate pour une pénalisation ℓ^1, permettant ici encore d'obtenir un coût calculatoire réduit.

Alors que l'ensemble des travaux précités considèrent la formulation du modèle (3.1) à $2P+1$ variables complexes, les observations $\boldsymbol{y}$ sont à valeurs réelles et le spectre $\boldsymbol{x}$ est par conséquent à symétrie hermitienne. Nous montrons au § 3.6 que la possibilité de prise en compte de cette symétrie, permettant de réduire de moitié le nombre d'inconnues et donc de diminuer le coût calculatoire de l'optimisation, dépend de la procédure utilisée, et sera notamment facilement exploitable par les algorithmes de type gradient étudiés au § 3.4.

Tout au long de ce chapitre, nous illustrons nos propos par des applications au signal multisinusoïdal introduit à la fin du chapitre précédent en figure 2.6. Ce signal est également utilisé pour établir la comparaison des performances des différents algorithmes proposés, que nous présentons finalement au § 3.7.

3.1 Analyse de l'opérateur W

Nous étudions en premier lieu la structure de la matrice $\mathbf{W}$ en fonction de la forme de l'échantillonnage. Nous avons détaillé au § 1.2.1 deux sortes d'irrégularités des instants de mesure :

- les « trous » dans les données, dus essentiellement aux effets de la rotation de la terre (alternance jour/nuit) et de sa révolution autour du soleil (alternance des saisons),
- l'échantillonnage irrégulier, où la durée séparant deux observations successives est totalement arbitraire.

Nous traiterons de manière séparée le cas particulier d'un échantillonnage *régulier troué*, qui confère aux matrices mises en jeu des propriétés structurelles similaires au cas de l'échantillonnage régulier « complet », l'absence de données ayant un effet équivalent de celui de la mise à zéro des valeurs manquantes. Nous nous intéressons également à l'influence du type d'échantillonnage (régulier complet et à données manquantes, échantillonnage irrégulier) sur les propriétés des matrices $\mathbf{W}\mathbf{W}^\dagger$ et $\mathbf{W}^\dagger\mathbf{W}$, qui nous seront utiles par la suite. En effet, l'inversion généralisée, qui sera étudiée au § 3.1.4, fait intervenir l'opérateur $\mathbf{W}\mathbf{W}^\dagger$. La structure de $\mathbf{W}^\dagger\mathbf{W}$, quant à elle, permettra d'étudier la régularisation quadratique au § 3.2.1 et sera déterminante lors de la mise en œuvre algorithmique de la procédure d'optimisation du critère (3.2).

3.1.1 Échantillonnage régulier

Dans le cas de données échantillonnées régulièrement à la période T_e et si la fréquence $f_{\max}$ est choisie à la limite de Nyquist $1/2T_e$, on a $\mathbf{W} = \{e^{j\pi\frac{kn}{P}}\}_{n=1\ldots N,\ k=-P\ldots P}$. Les première et dernière ligne de $\mathbf{W}$ étant ici identiques (elles correspondent à la contribution de la fréquence $f_{\max} = 1/2T_e$), on ne considère que ses $2P$ dernières lignes : $\mathcal{G} = \{\frac{k}{P} f_{\max}\}_{k=-P+1\ldots P}$. Les propriétés suivantes ont déjà été établies par Giovannelli et Idier [2001] ; nous les rappelons ici afin de voir dans quelle mesure elles pourront être généralisées au cas de l'échantillonnage irrégulier. Le modèle (3.1) s'écrit :

$$y(t_n) = \sum_{k=-P+1}^{P} x_k e^{j\pi\frac{kn}{P}} + \epsilon_n$$

et, dans le cas « classique » $2P = N$, $\mathbf{W}$ est la matrice (carrée) de Fourier :

$$\mathbf{W} = \mathbf{F}_{2P} \quad \triangleq \quad \{e^{j\pi\frac{kn}{P}}\}_{n=1\ldots 2P,\, k=-P+1\ldots P}.$$

On a alors $\mathbf{F}_{2P}\mathbf{F}_{2P}{}^\dagger = \mathbf{F}_{2P}{}^\dagger\mathbf{F}_{2P} = 2P\mathbf{I}_{2P}$ et l'inversion directe du problème ci-dessus correspond à la Transformée de Fourier Discrète (TFD) des données $\boldsymbol{y}$. Dans le cas dit *à haute résolution* où $2P \gg N$, on peut écrire $\mathbf{W} = \{e^{j\pi\frac{kn}{P}}\}_{k=-P+1\ldots P,\, n=1\ldots N}$ sous la forme $\mathbf{W} = \mathbf{Z}_{N,2P} = \mathbf{I}_{N,2P}\mathbf{F}_{2P}$, où $\mathbf{I}_{N,2P}$ est la matrice identité tronquée de taille $N \times 2P$. Il est alors facile de voir que :

- $\mathbf{Z}_{N,2P}\mathbf{Z}_{N,2P}{}^\dagger = 2P\,\mathbf{I}_N$,
- $\mathbf{Z}_{N,2P}{}^\dagger\mathbf{Z}_{N,2P} = \mathbf{F}_{2P}{}^\dagger\,\mathbf{D}_{\mathrm{z}}\,\mathbf{F}_{2P}$ est *circulante*,

où $\mathbf{D}_{\mathrm{z}}$ est la matrice diagonale $2P \times 2P$ constituée de N fois l'unité et $2P - N$ fois zéro.

3.1.2 Échantillonnage régulier à données manquantes

Pour un échantillonnage régulier comportant des trous, les instants d'échantillonnage peuvent se mettre sous la forme $t_n = i_n T_e$, $i_n \in \mathbb{N}$. En supposant à nouveau le paramètre $f_{\max}$ à la limite imposée par le théorème de Shannon $f_{\max} = 1/2T_e$, l'opérateur s'écrit de manière similaire au cas précédent sous la forme :

$$\begin{aligned} \mathbf{W} = \mathbf{M}_{N,2P} &\triangleq \{e^{j\pi \frac{k i_n}{P}}\}_{n=1\ldots N,\, k=-P+1\ldots P} \\ &= \mathbf{L}_{N,2P}\mathbf{F}_{2P}, \end{aligned}$$

où $\mathbf{L}_{N,2P}$ correspond à la matrice identité $\mathbf{I}_{2P}$ dont seules les N lignes d'indices i_n sont conservées[3]. Dans ce cas, on montre de manière similaire au cas non troué que :

- $\mathbf{M}_{N,2P}\mathbf{M}_{N,2P}{}^{\dagger} = 2P\,\mathbf{I}_N$,
- $\mathbf{M}_{N,2P}{}^{\dagger}\mathbf{M}_{N,2P} = \mathbf{F}_{2P}{}^{\dagger}\,\mathbf{D}_{\mathrm{M}}\,\mathbf{F}_{2P}$ est *circulante*,

où $\mathbf{D}_{\mathrm{M}}$ désigne ici la matrice diagonale de taille $2P \times 2P$ constituée de 1 aux indices i_n et de 0 ailleurs.

3.1.3 Échantillonnage irrégulier

Dans le cas plus général d'un échantillonnage totalement irrégulier, il est tout d'abord impossible de définir une fréquence maximale en fonction d'une limite de Nyquist : la plage fréquentielle exempte du phénomène de recouvrement spectral est dans ce cas bien plus large et le choix de $f_{\max}$ doit être fait en fonction de connaissances physiques *a priori* sur l'objet étudié. Le § 1.3.1 a été consacré à l'étude de ce point. Par ailleurs l'opérateur $\mathbf{W}$, noté dans ce cas $\mathbf{W} = \mathbf{U}_{N,2P+1} = \{e^{j2\pi \frac{k}{P} f_{\max} t_n}\}_{n=1\ldots N,\, k=-P\ldots P}$ ne peut plus être écrit au moyen de matrices de Fourier et l'on s'attend à ce que les matrices $\mathbf{U}_{N,2P+1}\mathbf{U}_{N,2P+1}^{\dagger}$ et $\mathbf{U}_{N,2P+1}^{\dagger}\mathbf{U}_{N,2P+1}$ perdent les propriétés des cas précédents. À cet effet, on montre que :

$$\begin{aligned} \{\mathbf{U}_{N,2P+1}\mathbf{U}_{N,2P+1}^{\dagger}\}_{m,n=1\ldots N} &= \frac{\sin\left(\pi(t_m - t_n)(2P+1)\frac{f_{\max}}{P}\right)}{\sin\left(\pi(t_m - t_n)\frac{f_{\max}}{P}\right)} \\ \text{et} \quad \{\mathbf{U}_{N,2P+1}^{\dagger}\mathbf{U}_{N,2P+1}\}_{k,\ell=-P\ldots P} &= \sum_n \exp\left(-2j\pi(k-\ell)t_n\frac{f_{\max}}{P}\right). \end{aligned}$$

Par conséquent, $\mathbf{U}_{N,2P+1}\mathbf{U}_{N,2P+1}^{\dagger}$ n'est pas proportionnel à l'identité et $\mathbf{U}_{N,2P+1}^{\dagger}\mathbf{U}_{N,2P+1}$ est encore une matrice de Toeplitz mais n'est plus circulante[4] Plus précisément, il vient des deux relations ci-dessus que ces matrices sont respectivement proportionnelle à l'identité et circulante si et seulement si les instants de mesure sont du type $t_n = i_n T_e$ et $f_{\max} = \frac{2P}{2P+1}\frac{1}{2T_e}$, c'est-à-dire exactement si l'échantillonnage est régulier et éventuellement à données manquantes[5].

3. Nous considérons ici le cas où $i_N < 2P$. Le cas d'un échantillonnage pour lequel $i_N > 2P$, ayant lieu par exemple lorsque l'échantillonnage présente de gros trous, donne lieu aux mêmes conclusions. Il suffit pour cela de considérer les représentants modulo $2P$ des indices i_n dans l'intervalle $[O,2P-1]$.

4. Toute matrice de Toeplitz étant asymptotiquement circulante lorsque sa dimension tend vers l'infini [Gray, 1977], une *approximation* de $\mathbf{U}_{N,2P+1}^{\dagger}\mathbf{U}_{N,2P+1}$ par une matrice circulante pourrait être envisagée ; nous n'avons néanmoins pas poursuivi dans cette voie.

5. Remarquons que dans ce cas, on n'a pas exactement $f_{\max}$ à la fréquence de Nyquist $1/2T_e$ alors que nous venons de montrer que cette valeur de $f_{\max} = f_{\text{Nyquist}}$ conditionnait les propriétés étudiées des deux matrices $\mathbf{W}^{\dagger}\mathbf{W}$ et $\mathbf{W}\mathbf{W}^{\dagger}$. Cette légère différence s'explique par le fait que les opérateurs diffèrent légèrement dans les deux cas. En échantillonnage irrégulier, les deux lignes $k = -P$ et $k = P$ sont différentes (et donc conservées) alors que dans le cas précédent, ces deux lignes étant identiques, on ne garde que la contribution en $k = P$.

3.1.4 Conséquences sur l'inversion généralisée

La structure de l'opérateur $\mathbf{W}\mathbf{W}^\dagger$ joue un rôle essentiel dans l'écriture de la solution inverse généralisée du problème (3.1) :

$$\widehat{\boldsymbol{x}}_{\text{IG}} \triangleq \mathbf{W}^\dagger(\mathbf{W}\mathbf{W}^\dagger)^{-1}\boldsymbol{y}.$$

Dans le cas de l'échantillonnage régulier, Giovannelli et Idier [2001] ont montré que l'inversion généralisée fournit un estimateur proportionnel à la TFD des données, éventuellement complétée par bourrage de zéros dans le cas de la haute résolution ($2P > N$). Il est facile de voir que cette propriété s'étend au cas d'un échantillonnage à données manquantes : puisque dans ce cas $\mathbf{M}_{N,2P}\mathbf{M}_{N,2P}^\dagger \propto \mathbf{I}_N$, l'inverse à droite $\widehat{\boldsymbol{x}}_{\text{IG}}$ est proportionnel à $\mathbf{M}_{N,2P}^\dagger\boldsymbol{y}$, qui correspond à la TFD des données $\boldsymbol{y}$ complétées par des zéros aux points où les données sont manquantes, et par bourrage de zéros pour atteindre les $2P$ points :

$$\mathbf{M}_{N,2P}^\dagger\boldsymbol{y} = \mathbf{F}_{2P}\,\widetilde{\boldsymbol{y}}, \quad \text{où } \widetilde{\boldsymbol{y}}[i_n] = \boldsymbol{y}[n] \text{ pour } n = 1\dots N \text{ et } \widetilde{\boldsymbol{y}}[n] = 0 \text{ ailleurs.}$$

La TFD de $\widetilde{\boldsymbol{y}}$ peut être vue comme un échantillonnage en fréquence, sur la grille $\mathcal{G}$, du spectre recherché convolué par la réponse en fréquence de la fenêtre d'observation, dite encore *fenêtre spectrale*. Comme nous l'avons vu au cours du premier chapitre, du fait des irrégularités dans l'échantillonnage, la fenêtre spectrale n'a plus l'allure d'un simple sinus cardinal et peut posséder de forts lobes secondaires indésirables, d'autant plus marqués si les trous dans l'échantillonnage sont périodiques. L'inversion généralisée ne fournit donc pas une solution intéressante. Le cas de l'échantillonnage totalement irrégulier, qui enlève aux matrices mises en jeu les propriétés des cas précédents, ne permet pas de généraliser l'interprétation de $\widehat{\boldsymbol{x}}_{\text{IG}}$ en termes de TFD. Néanmoins, il semble illusoire d'en attendre un comportement satisfaisant pour l'estimation de spectres de raies ; l'exemple de la figure 3.1, présenté au prochain paragraphe, confirmera cette intuition.

3.2 Différents types de pénalisation

3.2.1 Régularisation quadratique

Dans la littérature consacrée à la régularisation de type Tikhonov, la pénalisation quadratique occupe une place primordiale. Historiquement, d'une part, les travaux fondateurs de Tikhonov et Arsenin [1977] – bien qu'ils envisagent la régularisation dans un cadre théorique pour des objets *continus* – considèrent exclusivement des pénalisations de type quadratique opérant sur la fonction à restaurer ou sur ses dérivées. D'autre part, le cas quadratique est également intéressant sur le plan de la simplicité de la solution qui en résulte. Lorsque le problème initial est linéaire en les paramètres, une fonction quadratique présente notamment l'avantage de fournir une solution explicite, dont le calcul se ramène à la résolution d'un système linéaire, et qui est elle-même en relation linéaire vis-à-vis des données. Sous le formalisme bayésien, le minimiseur du critère (3.2) associé à une pénalisation quadratique de la forme $R_\Pi(\boldsymbol{x}) = \boldsymbol{x}^\dagger\Pi\boldsymbol{x}$ s'interprète comme l'estimateur du maximum *a posteriori* (MAP) correspondant à :

- des perturbations $\boldsymbol{\epsilon}$ *i.i.d.* selon une loi gaussienne circulaire $\mathcal{N}_C(0,2\sigma_\epsilon^2)$, permettant d'écrire la vraisemblance du modèle (3.1) sous la forme :

$$\mathcal{L}(\boldsymbol{y};\boldsymbol{x},\sigma_\epsilon) = \frac{1}{(2\pi\sigma_\epsilon^2)^N} e^{-\frac{\|\boldsymbol{y}-\mathbf{W}\boldsymbol{x}\|^2}{2\sigma_\epsilon^2}},$$

– une loi *a priori* gaussienne sur les amplitudes spectrales : $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0},\Sigma_x)$ où $\Sigma_x = \dfrac{\sigma_\epsilon^2}{2\lambda}\Pi^{-1}$. Le cas particulier de la pénalisation $R_{\mathbf{I}}(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$ revient alors à considérer les amplitudes x_k *i.i.d.* selon une loi gaussienne circulaire centrée $\mathcal{N}_C(0,\sigma_x^2)$ avec $\sigma_x^2 = \dfrac{\sigma_\epsilon^2}{2\lambda}$:

$$\forall k, \quad p(x_k) = \frac{1}{2\pi\sigma_x^2} e^{-\frac{|x_k|^2}{2\sigma_x^2}}.$$

Pour l'analyse spectrale à haute résolution de signaux régulièrement échantillonnés, Giovannelli et Idier [2001] ont étudié la régularisation quadratique et démontré les propriétés suivantes, en utilisant les notations du § 3.1 (en particulier, nous écrivons $\mathbf{W} = \mathbf{Z}_{N,2P}$) :

P.1 La solution régularisée pour la fonction quadratique $R_{\mathbf{I}}(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$ s'écrit :

$$\widehat{\boldsymbol{x}}_2^{\mathbf{I}} = (\mathbf{Z}_{N,2P}{}^\dagger \mathbf{Z}_{N,2P} + \lambda \mathbf{I}_{2P})^{-1} \mathbf{Z}_{N,2P}{}^\dagger \boldsymbol{y} = \frac{1}{2P+\lambda} \mathbf{Z}_{N,2P}{}^\dagger \widetilde{\boldsymbol{y}}$$

où $\widetilde{\boldsymbol{y}}[n] = \boldsymbol{y}[n]$ pour $n = 1 \dots N$ et $\widetilde{\boldsymbol{y}}[n] = 0$ pour $n = N+1 \dots 2P$. Cette solution est proportionnelle à la TFD des données complétées par bourrage de zéros, le paramètre de régularisation ne contrôlant que le coefficient de proportionnalité.

P.2 La généralisation au cas d'une fonction de pénalisation quadratique du type $R_\Pi(\boldsymbol{x}) = \boldsymbol{x}^\dagger \Pi \boldsymbol{x}$ est également étudiée, où Π est une matrice *circulante*. Cette hypothèse additionnelle permet de diagonaliser Π dans l'espace de Fourier : $\Pi = \mathbf{F}_{2P}{}^\dagger \, \mathbf{D}_\Pi \, \mathbf{F}_{2P}$, où $\mathbf{D}_\Pi$ est une matrice diagonale d'éléments notés $d_\Pi[k]$. La solution régularisée se met alors sous la forme :

$$\widehat{\boldsymbol{x}}_2^{\Pi} = (\mathbf{Z}_{N,2P}{}^\dagger \mathbf{Z}_{N,2P} + \lambda \Pi)^{-1} \mathbf{Z}_{N,2P}{}^\dagger \boldsymbol{y} = \mathbf{F}_{2P}{}^\dagger \breve{\boldsymbol{y}}$$

où $\breve{\boldsymbol{y}}[n] = \dfrac{2P}{2P + \lambda \boldsymbol{d}_\Pi[n]} \widetilde{\boldsymbol{y}}[n]$ pour $n = 1 \dots 2P$. $\widehat{\boldsymbol{x}}_2^{\Pi}$ correspond donc à la TFD des données fenêtrées et complétées par bourrage de zéros, la structure de la matrice $\mathbf{D}_\Pi$ contrôlant la forme du fenêtrage.

Le cas d'un échantillonnage régulier à données manquantes permet d'étendre sans difficulté ces deux propriétés, qui reposent sur la structure circulante de la matrice $\mathbf{W}^\dagger\mathbf{W}$. Ainsi, la solution $\widehat{\boldsymbol{x}}_2^{\mathbf{I}}$ est également proportionnelle à la TFD des données complétées par des zéros :

$$\widehat{\boldsymbol{x}}_2^{\mathbf{I}} = \frac{1}{2P+\lambda} \mathbf{M}_{N,2P}{}^\dagger \widetilde{\boldsymbol{y}}$$

qui, une fois de plus, n'est pas une solution satisfaisante. De ce fait, la régularisation quadratique par une fonction $R_{\mathbf{I}}(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$ n'est pas intéressante dans le cas de données manquantes. Pour une pénalisation de type $R_\Pi(\boldsymbol{x}) = \boldsymbol{x}^\dagger \Pi \boldsymbol{x}$, la propriété P.2 est interprétée dans Giovannelli et Idier [2001] en termes de fenêtrage des données, qui permet de contrôler l'influence de la fenêtre d'observation sur la TFD. Néanmoins, la forme plus complexe de celle-ci dans le cas de données manquantes rend l'utilisation du fenêtrage des données plus que délicate.

Les propriétés précédentes ne peuvent pas être généralisées au cas de l'échantillonnage irrégulier : nous avons vu au § 3.1.3 que, la matrice $\mathbf{W} = \mathbf{U}_{N,2P+1}$ ne pouvant plus s'écrire à partir d'une matrice de Fourier, les opérateurs $\mathbf{U}_{N,2P+1}^\dagger \mathbf{U}_{N,2P+1}$ et $\mathbf{U}_{N,2P+1} \mathbf{U}_{N,2P+1}^\dagger$ perdaient leurs propriétés structurelles. Par conséquent, la solution $\widehat{\boldsymbol{x}}_2^{\mathbf{I}}$ n'est plus proportionnelle à $\mathbf{U}_{N,2P+1}^\dagger \boldsymbol{y}$,

qui représente le spectre de Fourier des données échantillonné sur la grille $\mathcal{G}$. Malgré l'absence de conclusion théorique, le résultat obtenu dans le cas de données manquantes rend toutefois dissuasive l'utilisation d'une telle régularisation pour notre problème, quel que soit le type d'échantillonnage.

Exemple sur un signal simple

En application sur le signal test échantillonné irrégulièrement introduit à la fin du chapitre précédent (voir la figure 2.6 au § 2.4), la figure 3.1 à gauche présente la solution régularisée $\widehat{\boldsymbol{x}}_2$ pour une valeur de $\lambda = 0.1$. S'il n'est pas exactement proportionnel au spectre de Fourier des données, l'estimateur $\widehat{\boldsymbol{x}}_2^{\mathbf{I}}$ présente globalement le même comportement et est à ce titre inacceptable. La figure de droite représente la solution inverse généralisée correspondant au même signal. En réalité, le calcul de $\widehat{\boldsymbol{x}}_{\mathrm{IG}}$ n'est pas réalisable directement à cause du trop mauvais conditionnement de l'opérateur $\mathbf{U}_{N,2P+1}\mathbf{U}_{N,2P+1}^{\dagger}$; cette figure représente donc son approximation par $\widehat{\boldsymbol{x}}_2^{\mathbf{I}}$ pour une valeur faible de $\lambda = 10^{-4}$, dont l'allure est également inadéquate pour l'estimation de raies spectrales.

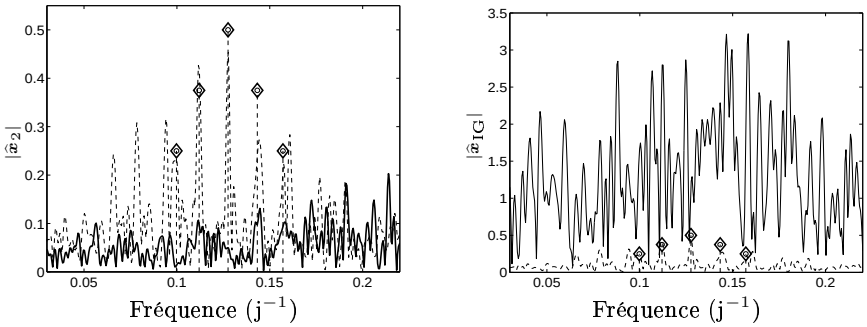


FIG. 3.1 – *Estimation pour une pénalisation quadratique $R_2(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$. En trait plein, à gauche, module $|\widehat{\boldsymbol{x}}_2^{\mathbf{I}}|$ pour $\lambda = 0.1$ et, à droite, approximation de $|\widehat{\boldsymbol{x}}_{IG}|$. Le module du spectre de Fourier des données $|\mathbf{U}_{N,2P+1}^{\dagger}\,\boldsymbol{y}|$ est tracé en pointillé et les raies théoriques sont représentées par $\diamond$.*

3.2.2 Régularisations non quadratiques

Le cas quadratique se révélant inapproprié pour l'estimation de spectres de raies, nous étudions à présent le choix d'une pénalisation mieux adaptée à notre problème. En l'absence d'information *a priori* sur le contenu du spectre recherché – si ce n'est sa nature parcimonieuse – nous nous plaçons ici dans le cas de pénalisations *indépendantes* et *identiques* de toutes les composantes spectrales : la probabilité d'avoir une raie est la même en tout point de la grille $\mathcal{G}$. Par ailleurs, l'information *a priori* de parcimonie porte uniquement sur les *modules* des amplitudes complexes x_k. Dans le formalisme bayésien, cette double hypothèse revient à considérer la loi *a priori* sur le spectre $\boldsymbol{x}$ sous la forme produit : $p(\boldsymbol{x}) = \prod_{k=-P}^{P} p(x_k)$, où $p(x_k)$ s'écrit comme une fonction du seul module $|x_k|$. La fonction de pénalisation peut alors se décomposer sous la forme séparable :

$$R(\boldsymbol{x}) = \sum_{k=-P}^{P} \phi(x_k), \text{ avec } \phi(x_k) = \varphi(|x_k|)$$

Nous avons vu précédemment que la pénalisation quadratique, c'est-à-dire une distribution *a priori* gaussienne des amplitudes $\boldsymbol{x}$, débouchait sur des solutions inacceptables car proches du spectre de Fourier des données. Il s'agit donc d'obtenir des distributions plus parcimonieuses des amplitudes que le cas gaussien, et beaucoup de fonctions φ ont été proposées à ce titre. Nous n'en étudions ici que trois, qui sont représentatives des différents types de choix possibles.

Régularisation par *a priori* de Cauchy

Historiquement, la première formulation du problème d'estimation de raies spectrales sous la forme linéaire (3.1) est due à Sacchi *et al.* [1998], qui proposent une distribution *a priori* de Cauchy sur les modules des amplitudes spectrales $|x_k|$, supposés indépendants et identiquement distribués selon :

$$p_C(\,|x_k|\,;\sigma_x) \propto \frac{1}{1+\frac{|x_k|^2}{2\sigma_x^2}}.$$

De par sa longue queue, une telle distribution favorise particulièrement des solutions parcimonieuses : en autorisant des valeurs d'amplitudes élevées, l'énergie du signal peut ainsi être concentrée en un faible nombre de raies spectrales. Cette distribution correspond, à un facteur additif constant près, à la fonction de pénalisation :

$$\phi_{\mathrm{C}}(x_k) = \ln(1+\frac{|x_k|^2}{2\sigma_x^2}).$$

Néanmoins, cette forte adéquation de la pénalisation à la structure recherchée s'obtient au prix de la perte d'une propriété essentielle sur le plan algorithmique : la convexité. En effet, la fonction ϕ n'étant pas convexe, le critère pénalisé qui lui est associé :

$$J_{\mathrm{C}}(\boldsymbol{x}) \triangleq \frac{1}{2}\left\|\boldsymbol{y}-\mathbf{W}\boldsymbol{x}\right\|^2 + \lambda\sum_k \ln(1+\frac{|x_k|^2}{2\sigma_x^2})$$

est également non convexe et l'existence éventuelle de minima locaux rend l'optimisation délicate. Pour notre problème, en particulier, nous avons vu au § 1.3 que la présence de trous périodiques dans l'échantillonnage peut provoquer l'apparition de faux pics dans le spectre de Fourier, qui peuvent être interprétés comme des maxima locaux de la fonction de vraisemblance[6]. En d'autres termes, les perturbations introduites par un échantillonnage non régulier favorisent l'existence de minima locaux dans le critère $J_{\mathrm{C}}(\boldsymbol{x})$ et augmentent ainsi le risque de converger vers un minimum local, comprenant éventuellement de faux pics. Notons également que, lors de simulations, un réglage des hyperparamètres (λ,σ_x^2) permettant d'obtenir une solution satisfaisante s'est avéré assez délicat. Par la suite, nous considérons donc uniquement des fonctions de pénalisation convexes.

Pénalisation hyperbolique

Il s'agit donc de rechercher une fonction de pénalisation ϕ de croissance plus lente que la fonction quadratique, mais qui reste pourvue de la propriété de convexité. C'est par exemple le cas des fonctions ℓ^p : $\phi_p(u) = |u|^p$ avec $1 < p < 2$ utilisées en reconstruction d'image [Bouman

6. En toute rigueur, il s'agit de la vraisemblance des données pour un modèle mono-sinusoïdal, qui n'est pas le modèle considéré ici ; cette analogie nous rend toutefois peu enclins à l'utilisation d'un critère pouvant présenter des extrema locaux.

et Sauer, 1993] et en déconvolution à caractère impulsionnel [Yarlagadda *et al.*, 1985], ou des fonctions de type $\ell^2\ell^1$, c'est-à-dire quadratiques à l'origine et linéaires à l'infini. Dans ce même cadre d'estimation de spectres de raies (pour des données régulièrement échantillonnées), Ciuciu *et al.* [2001] utilisent ainsi la branche d'hyperbole :

$$\phi_{\text{hyp}}(x_k) = \sqrt{s^2 + |x_k|^2}.$$

En termes de parcimonie, cette pénalisation correspond à un *a priori* moins fort que celui représenté par la fonction précédente ϕ_{C}, mais reste strictement convexe. Un comportement affine à l'infini est même le cas limite de fonction convexe qui reste croissante. La stricte convexité de ϕ_{hyp} assure alors l'existence d'un unique minimum global du critère pénalisé correspondant :

$$J_{\text{hyp}}(\boldsymbol{x}) \triangleq \frac{1}{2}\left\|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\right\|^2 + \lambda \sum_k \sqrt{s^2 + |x_k|^2}$$

qui est alors également strictement convexe. De plus, la fonction ϕ_{hyp} étant continûment différentiable sur $\mathbb{R}^+$, la minimisation de J_{hyp} pourra être envisagée par un algorithme d'optimisation classique de type descente de gradient : nous reviendrons sur la mise en œuvre algorithmique au § 3.4. Notons enfin que pour une faible valeur de s, la fonction de pénalisation correspondante peut être vue comme une approximation, conservant la propriété essentielle de *stricte* convexité, de la norme ℓ^1 qui, comme nous le verrons au paragraphe suivant, présente des propriétés théoriques intéressantes en termes de parcimonie.

Pénalisation par la norme ℓ^1

Depuis les années 1990, beaucoup d'attention a été portée au lien entre l'information de parcimonie et l'usage de la norme ℓ^1 : $\|\boldsymbol{x}\|_1 = \sum_k |x_k|$. Ces travaux s'inscrivent dans une thématique très suivie aujourd'hui (elle a par exemple donné lieu à une session spéciale à la conférence ICASSP'05 : *Advances in Sparse Signal Representation*) : il s'agit de la représentation d'un signal à partir d'un dictionnaire *surcomplet*, constitué d'un large éventail de fonctions de nature analytique différente, pouvant par exemple inclure les fonctions sinusoïdales, plusieurs types d'ondelettes, des signaux « chirp », etc. Un tel dictionnaire est alors souvent *redondant*, c'est-à-dire que certains de ses éléments peuvent s'exprimer en fonction des autres. Il ne s'agit donc pas exactement de décomposer le signal sur une base au sens propre du terme, mais plutôt de rechercher la représentation la plus parcimonieuse, c'est-à-dire utilisant le plus petit nombre possible d'éléments du dictionnaire. La méthodologie des algorithmes de décomposition temps-fréquence de type *Matching Pursuit* [Mallat et Zhang, 1993] s'inscrit par exemple dans ce cadre. À partir d'un dictionnaire de représentations temps-fréquence élémentaires, un signal est décomposé en une série d'*atomes* où, à chaque itération, l'atome sélectionné correspond à la meilleure projection du signal résiduel – le signal initial auquel ont été retirés les premiers atomes – parmi l'ensemble des éléments du dictionnaire. Dans le cas d'un dictionnaire limité aux fonctions sinusoïdales, cette démarche est à rapprocher des algorithmes CLEAN décrits au deuxième chapitre (bien que sans *clean factor*, voir le § 2.2) et souffre des limites liées à l'aspect *itératif* de la méthode, soit essentiellement la propagation irréversible d'erreurs et le choix d'un critère d'arrêt satisfaisant, fixant le nombre de composantes de la représentation

Ce problème a récemment été abordé comme un problème d'optimisation *globale* [Chen *et al.*, 1999, Fuchs, 2004], pouvant s'écrire sous la forme :

$$\text{(P0)} \quad : \quad \widehat{\boldsymbol{x}}_{\text{P0}} \triangleq \arg\min \|\boldsymbol{x}\|_0 \quad \text{s.c.} \quad \boldsymbol{y} = \mathbf{W}\boldsymbol{x},$$

où la norme $\|\boldsymbol{x}\|_0$ est le nombre de composantes non nulles de $\boldsymbol{x}$ et, dans un cadre très général, les colonnes de $\mathbf{W}$ représentent chaque élément du dictionnaire. Les valeurs non nulles de $\boldsymbol{x}$ correspondent alors aux coefficients de projection du signal sur chacun des éléments sélectionnés. Le problème (P0) est un problème d'optimisation de type *combinatoire* dans un espace de dimension très élevé et n'est donc en pratique pas solvable de manière optimale. Donoho et Logan [1992] ont montré, sous certaines conditions, l'équivalence entre la recherche de la solution la plus parcimonieuse (de norme ℓ^0 minimale) et celle de la solution de norme ℓ^1 minimale :

$$(\mathrm{P0}) \Leftrightarrow (\mathrm{P1}) \ : \quad \widehat{\boldsymbol{x}}_{\mathrm{P1}} \triangleq \arg\min \|\boldsymbol{x}\|_1 \quad \text{s.c.} \quad \boldsymbol{y} = \mathbf{W}\boldsymbol{x},$$

le problème (P1) pouvant être traité par des méthodes d'optimisation plus maîtrisables. C'est la méthodologie retenue par les approches de type *Basis Pursuit* [Chen *et al.*, 1999] pour ce même problème de décomposition parcimonieuse à partir de dictionnaires surcomplets. Les travaux de J.-J. Fuchs établissent en particulier le lien entre la résolution d'un problème de type (P1) et la minimisation du critère :

$$\widehat{\boldsymbol{x}}_1 = \arg\min J_1(\boldsymbol{x}), \quad J_1(\boldsymbol{x}) \triangleq \frac{1}{2}\|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{x}\|_1.$$

Dans le cas de paramètres à valeurs réelles, Fuchs [1997] montre que, pour des valeurs suffisamment faibles du paramètre de régularisation λ, ces deux problèmes sont équivalents, dans le sens où ils débouchent sur des estimés ayant leurs composantes non nulles localisées aux mêmes endroits, et de surcroît de même signe. Fuchs [2004] établit ensuite des conditions suffisantes liant le minimiseur $\widehat{\boldsymbol{x}}_1$ du critère J_1, l'opérateur $\mathbf{W}$ et le paramètre de régularisation λ, pour lesquelles le minimiseur $\widehat{\boldsymbol{x}}_1$ est unique et correspond à *la* solution la plus parcimonieuse du problème exact : $\widehat{\boldsymbol{x}}_1 = \widehat{\boldsymbol{x}}_{\mathrm{P1}}$.

Ces propriétés, bien que très intéressantes sur le plan théorique pour établir le lien entre la pénalisation ℓ^1 et une solution parcimonieuse du critère régularisé, ne sont cependant pas directement applicables au problème que nous envisageons ici : en pratique, les données observationnelles sont entachées de bruit et nous ne recherchons donc pas une solution les modélisant parfaitement, qui n'est pas forcément satisfaisante. En particulier, il pourra être nécessaire de choisir librement le paramètre de régularisation alors que les conditions établies[7] dans [Fuchs, 2004] imposent une limite maximale à la valeur de λ. Notons encore que, dans le cas d'un modèle bruité, la minimisation de J_1 pour la recherche d'une décomposition parcimonieuse est également l'approche adoptée par [Chen *et al.*, 1999], sous le terme de *Basis Pursuit Denoising*. Ainsi, même si le cadre théorique décrit ci-dessus n'est pas exactement celui de notre problème (pour l'analyse spectrale, par ailleurs, les inconnues sont *complexes*), la pénalisation ℓ^1 reste intéressante dans notre cas, puisque la fonction $\phi_1(x) = |x|$ est convexe et de croissance plus lente qu'une fonction quadratique. C'est d'ailleurs, parmi les fonctions ℓ^p, le cas limite assurant une croissance plus lente que la pénalisation quadratique, soit $p < 2$, tout en restant convexe, ce qui n'est vrai que si $p \geq 1$. Nous verrons par la suite lors de simulations que cette pénalisation débouche effectivement sur des solutions rigoureusement parcimonieuses, c'est-à-dire exactement nulles sauf en un nombre restreint de positions.

Un point important pour le problème de l'analyse spectrale est l'adaptation de la pénalisation par la norme ℓ^1 pour des variables *à valeurs complexes*, qui n'est pas sans poser de difficultés.

– Au niveau de la modélisation, d'une part, les propriétés théoriques établissant le lien de la

7. Ces conditions imposent également un nombre maximal de composantes non nulles, lié à la forme de l'opérateur $\mathbf{W}$. Dans le cas de l'analyse spectrale à échantillonnage irrégulier, et pour les simulations que nous présenterons au paragraphe 3.2.3, ces conditions ne sont pas respectées.

norme ℓ^1 avec une représentation parcimonieuse considèrent spécifiquement des paramètres à valeurs réelles. Les travaux précités s'appuient en effet sur l'utilisation de la *relation d'ordre* et du *signe* des paramètres, notions propres à l'écriture de la norme ℓ^1 dans $\mathbb{R}$ à partir de la valeur absolue : pour $u \in \mathbb{R}$, $|u| = u$ si $u \geq 0$ et $-u$ si $u < 0$. Pour des paramètres à valeurs complexes, la norme ℓ^1 est définie à partir du module complexe et ne possède plus ce lien à des notions d'ordre et de signe. En pratique, nous verrons par la suite que cette pénalisation débouche effectivement sur des solutions à valeurs complexes rigoureusement parcimonieuses, c'est-à-dire exactement nulles en module sauf en un nombre restreint de positions fréquentielles.

– Mais surtout, sur le plan algorithmique, la décomposition précédente de la valeur absolue permet d'écrire, pour des paramètres réels, la minimisation de J_1 sous la forme d'un problème quadratique sous contrainte, ouvrant alors de nombreuses possibilités algorithmiques. Ce n'est plus le cas pour des paramètres complexes. L'analogie entre $\mathbb{C}$ et $\mathbb{R}^2$ permet de ne plus considérer de variables complexes en utilisant la formulation à paramètres réels (2.13) introduite à la fin du chapitre précédent, qui considère les inconnues $\boldsymbol{s}_k = [a_k, b_k] \in \mathbb{R}^2$ au lieu de $x_k \in \mathbb{C}$, avec $x_k = (a_k - j\, b_k)/2$. La pénalisation par la norme ℓ^1 du vecteur d'amplitudes $\boldsymbol{s}$ conduit alors à des pénalisations *indépendantes* des parties réelle et imaginaire d'une même amplitude complexe x_k :

$$\|\boldsymbol{s}\|_1 = \sum_k |a_k| + |b_k| \neq \sum_k |a_k + j\, b_k| = 2\, \|\boldsymbol{x}\|_1$$

Par conséquent, la pénalisation par $\|\boldsymbol{s}\|_1$ n'est pas exactement celle de la norme ℓ^1 pour des variables complexes et est moins satisfaisante en termes de modélisation *a priori*, puisqu'elle considère indépendamment les parties réelle et imaginaire de l'amplitude associée à une même raie. Nous remarquons que c'est pourtant l'approche adoptée par Chen et Donoho [1998] qui, dans l'esprit des méthodes de *Basis Pursuit* évoquées plus haut, cherchent la décomposition la plus parcimonieuse d'un signal sur un dictionnaire de fonctions cosinus et sinus.

Exemple sur un signal simple

Nous pouvons illustrer l'effet insatisfaisant d'une pénalisation indépendante de $\Re(x_k)$ et $\Im(x_k)$ sur notre exemple habituel. La figure 3.2 représente les estimateurs obtenus selon les deux types de pénalisation : la figure de gauche correspond à une pénalisation de la forme $R_1^{\mathbb{R}}(\boldsymbol{x}) = \sum_k |\Re(x_k)| + |\Im(x_k)|$ et la figure de droite à une pénalisation sur les modules complexes : $R_1^{\mathbb{C}}(\boldsymbol{x}) = \sum_k |x_k|$. Les méthodes mises en œuvre pour obtenir de tels estimateurs seront détaillées au § 3.5. Sur cet exemple, à gauche, la pénalisation indépendante des parties réelle et imaginaire de $\boldsymbol{x}$ produit des vecteurs $\Re(\boldsymbol{x})$ et $\Im(\boldsymbol{x})$ qui prennent leurs valeurs non nulles en des positions légèrement décalées : le module $|\boldsymbol{x}|$ correspondant présente donc parfois deux pics pour la localisation d'une seule fréquence (voir le zoom autour de la fréquence f_4). À l'inverse, sur la figure de droite, la pénalisation portant sur le module contraint les vecteurs $\Re(\boldsymbol{x})$ et $\Im(\boldsymbol{x})$ à être non nuls simultanément. Par conséquent le spectre $|\boldsymbol{x}|$ présente dans ce cas une allure plus parcimonieuse et bien plus satisfaisante en termes de localisation fréquentielle.

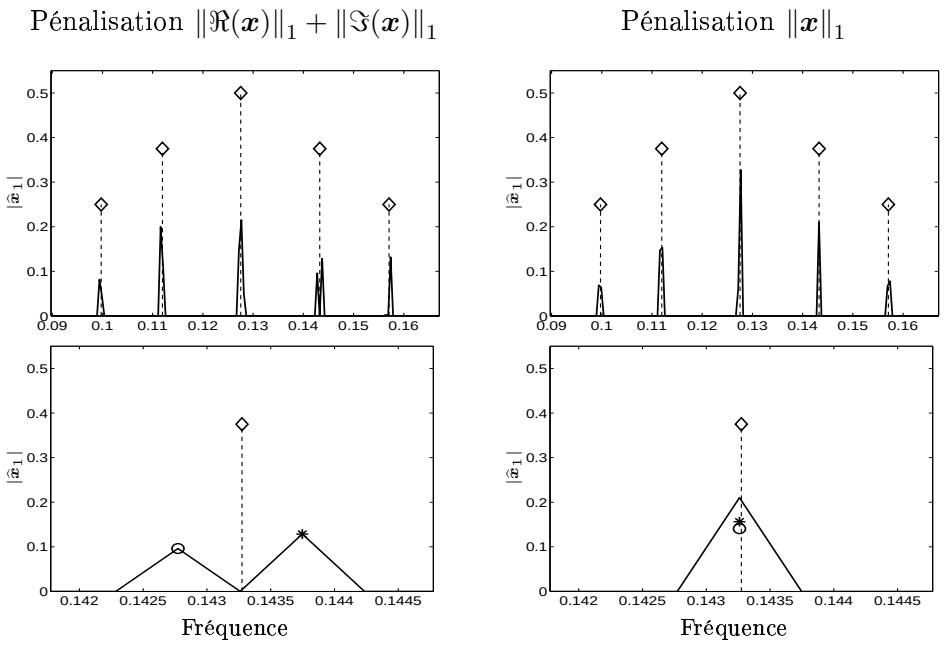


FIG. 3.2 – *Illustration de la différence entre* $|\Re(x_k)| + |\Im(x_k)|$ *et* $|x_k|$ *pour la pénalisation par la norme* ℓ^1. *En haut, spectres estimés en module (en trait plein) pour les deux formes de pénalisation et raies théoriques* ($\diamond$). *En bas, zoom autour de la fréquence* f_4 *: les* $\circ$ *et* $\star$ *représentent respectivement les valeurs non nulles de* $|\Re(\widehat{\boldsymbol{x}})|$ *et* $|\Im(\widehat{\boldsymbol{x}})|$ *et le trait plein correspond au module associé* $|\widehat{\boldsymbol{x}}|$. *La raie théorique à la fréquence* f_4 *est représentée par* $\diamond$.

Si la pénalisation ℓ^1 opérant sur les modules complexes a déjà été appliquée ([Davies et Daudet, 2003] constitue à notre connaissance le seul exemple à ce sujet), elle n'a jusqu'ici jamais été envisagée pour le problème de l'analyse spectrale de la forme (3.1) pour lequel une approximation présentant de plus fortes propriétés analytiques lui est souvent préférée, comme par exemple la branche d'hyperbole décrite au paragraphe précédent. En effet, la norme ℓ^1 souffre du manque de deux caractéristiques essentielles. Premièrement, la pénalisation n'est pas *strictement* convexe et, dans le cas sous-déterminé dans lequel nous nous plaçons, la partie quadratique $\frac{1}{2}\|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|^2$, qui est convexe, ne l'est pas au sens strict. La convexité de J_1 n'est donc pas stricte non plus et l'unicité de son minimiseur ne peut pas être garantie : la simple convexité permet uniquement d'affirmer que ses minima sont globaux, c'est-à-dire que l'ensemble de ses minima (qui est convexe) donne la même valeur du critère [Bertsekas, 1995, Annexe B]. Ensuite, le critère J_1 n'est pas différentiable en tout vecteur ayant un élément nul, ce qui rend délicate, voire impossible, l'application de schémas algorithmiques classiques visant à sa minimisation. À l'inverse, nous verrons au § 3.5 que la norme ℓ^1 possède des propriétés très intéressantes à exploiter en termes algorithmiques et qui constituent, en plus de la parcimonie, une forte motivation à l'usage de la pénalisation ℓ^1 pour notre problème d'analyse spectrale.

3.2.3 Bilan

La figure 3.3 ci-après propose une comparaison du comportement des différentes fonctions de pénalisation présentées, ainsi que des densités de probabilité correspondantes. Vu qu'elles sont toutes du type $R(\boldsymbol{x}) = \sum_k \phi(x_k)$ avec $\phi(x_k) = \varphi(|x_k|)$, nous traçons les fonctions $\varphi(u)$ de la va-

riable réelle $u \geq 0$. De même, nous traçons les distributions de probabilité $p(u) = k_\varphi \exp -\lambda\varphi(u)$ pour $\lambda = 1$ et k_φ tel que $\int_{x=0}^{+\infty} p(u)\, du = 1$. La distribution *a priori* de Cauchy est celle présentant la plus longue queue, et donc la mieux adaptée pour des objets parcimonieux, mais la pénalisation correspondante n'est pas convexe. À l'inverse, une pénalisation quadratique pénalise trop fortement les amplitudes élevées : sa distribution, gaussienne, contraint plutôt les amplitudes à prendre des valeurs plus faibles. On comprend alors aisément que, pour la restauration d'un spectre d'énergie donnée, une pénalisation quadratique favorisera les solutions présentant de faibles amplitudes tout au long de la grille fréquentielle de reconstruction, alors qu'une fonction pénalisant moins les grandes valeurs permettra la reconstruction d'un spectre parcimonieux, concentrant l'énergie dans un nombre limité de plus grandes amplitudes. Les pénalisations ℓ^1 et hyperbolique ont des représentations très proches et présentent des propriétés intermédiaires, à savoir un écrasement moins marqué des grandes amplitudes par rapport au cas gaussien, tout en maintenant la convexité de la pénalisation (la convexité au sens *strict* étant obtenue uniquement dans le cas de la pénalisation hyperbolique).

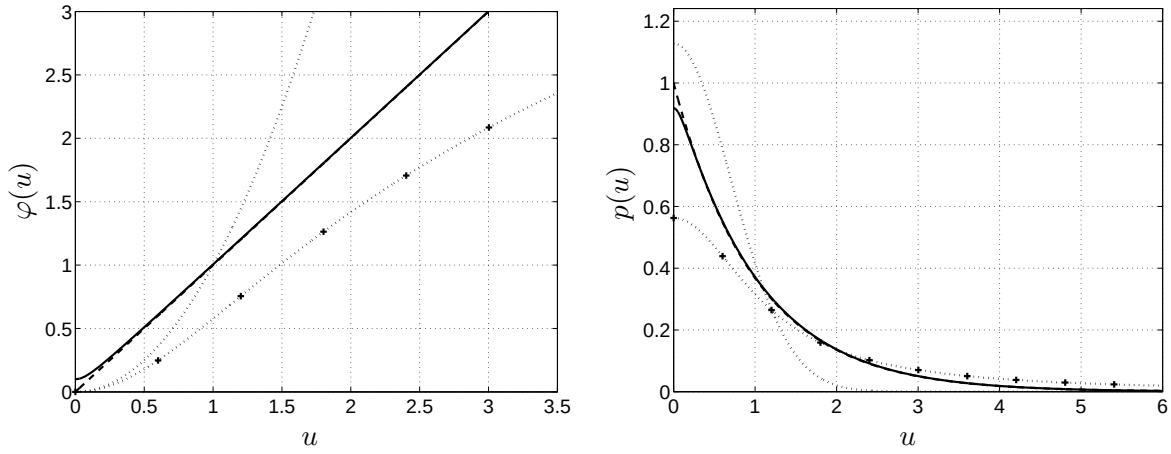


FIG. 3.3 – *Fonctions de pénalisation (à gauche) et distributions de probabilité associées (à droite) : pénalisations quadratique (...), ℓ^1 (- -), log-Cauchy avec $\sigma_x = 0.8$ (+) et hyperbolique avec $s = 10^{-1}$ (-).*

Notons que la définition des fonctions ϕ_{C} et ϕ_{hyp} fait intervenir un paramètre supplémentaire, à savoir respectivement, σ_x et s. Ce paramètre réglant la forme de la courbe de $\phi_{\mathrm{C}}(u)$ (respectivement $\phi_{\mathrm{hyp}}(u)$) peut être interprété comme un paramètre de *seuillage* des amplitudes. En effet, ces fonctions ayant un comportement quadratique (donc *plat*) lorsque u tend vers 0, elles pénalisent sensiblement de la même manière les amplitudes correspondant à cette zone, dont l'étendue est déterminée par la valeur de σ_x (respectivement de s).

Exemple sur un signal simple

La figure 3.4 illustre le comportement typique des estimateurs sur le spectre de raies de notre signal habituel, obtenu pour les trois pénalisations étudiées ci-dessus. Un paramètre de régularisation de $\lambda = 50$ a été choisi pour $\widehat{\boldsymbol{x}}_1$ et $\widehat{\boldsymbol{x}}_{\mathrm{hyp}}$, et pour ce dernier le paramètre s a été fixé à 10^{-3} (nous reviendrons sur ces choix au paragraphe suivant) ; un réglage *ad hoc* de λ et σ_x a été utilisé pour $\widehat{\boldsymbol{x}}_{\mathrm{C}}$. La grille fréquentielle comporte ici $P = 510$ fréquences positives (rappelons que le signal comporte $N = 250$ points) et le paramètre $f_{\max} = 0.25$; nous ne représentons ici

que la zone fréquentielle correspondant aux fréquences positives du signal. Dans chaque cas, une solution parcimonieuse satisfaisante est estimée (figures de gauche) et les pics dans les valeurs de $|\widehat{\boldsymbol{x}}|$ réalisent un encadrement des fréquences théoriques sur la grille de reconstruction (figures de droite). L'agrandissement effectué sur les figures de la colonne centrale montrent également que la pénalisation par la norme ℓ^1 fournit une solution *exactement* parcimonieuse, c'est-à-dire exactement nulle en dehors des positions des raies contenues dans le signal. Ce n'est pas le cas pour les pénalisations log-Cauchy et hyperbolique, pour lesquelles les amplitudes ne sont jamais rigoureusement nulles, et qui présente quelques artefacts nécessitant une opération postérieure de seuillage. Notons encore que, pour les pénalisation hyperbolique et ℓ^1 que nous proposons d'utiliser, les amplitudes correspondant aux raies spectrales sont systématiquement sous-estimées. Ce phénomène, inhérent au principe même de la régularisation, est d'ailleurs moins marqué pour la pénalisation log-Cauchy, puisque celle-ci pénalise moins fortement les amplitudes élevées (voir la figure 3.3). Cependant, il ne doit pas être considéré comme un problème en soi. L'étape la plus délicate dans le problème d'estimation de spectres de raies est bien entendu celle de la *détection* des fréquences. Une fois que celles-ci ont été correctement localisées, il est alors possible de réestimer *a posteriori* les amplitudes du modèle multi-sinusoïdal correspondant par moindres carrés, le problème étant alors linéaire et largement surdéterminé (peu de paramètres inconnus devant le nombre de données N). Ici aussi, signalons l'avantage de la pénalisation ℓ^1 qui, à la différence de la pénalisation hyperbolique, ne nécessite pas la définition d'un seuil de détection. La réestimation des amplitudes correspondant aux cinq raies détectées par ces trois estimateurs, c'est-à-dire aux positions des pics dans les valeurs de $|\widehat{\boldsymbol{x}}|$, est également représentée sur les figures de la colonne de droite : une estimation satisfaisante de l'ensemble des paramètres spectraux est ainsi finalement obtenue.

3.3 Sélection des hyperparamètres

Nous avons étudié au cours du paragraphe précédent différents critères pénalisés, dont la définition fait intervenir des hyperparamètres : le paramètre s pour la pénalisation hyperbolique et, dans tous les cas, le paramètre de régularisation λ, dont la valeur n'est pas d'un intérêt direct, mais qu'il convient de régler de façon adéquate.

3.3.1 Choix de s pour la pénalisation hyperbolique

Dans le cas de la pénalisation hyperbolique, le paramètre s est introduit de façon à obtenir une pénalisation approximant la norme ℓ^1, et doit pour cela être de faible valeur : plus s est petit, plus le caractère parcimonieux de l'estimateur sera accentué et nous avons vu au § 3.2.2 que la valeur de s contrôle le niveau de seuil des amplitudes en dessous duquel elles sont considérées comme nulles. Cependant, plus s est petit, plus le critère correspondant s'approche de la non-différentiabilité en tout vecteur ayant une composante nulle, point critique de la pénalisation ℓ^1. Une valeur trop faible de s risque donc de ralentir fortement la vitesse de convergence de la procédure d'optimisation. Ciuciu *et al.* [2001] envisagent la décroissance progressive du paramètre s au cours de l'optimisation, mais préfèrent fixer sa valeur dans le cas de la restauration de spectres de raies. Nous avons constaté en pratique que l'estimateur est peu sensible au réglage de s à partir du moment où ce paramètre est assez faible et, pour des amplitudes spectrales de l'ordre de l'unité, une valeur de s de l'ordre de 10^{-3} est un compromis satisfaisant, permettant d'obtenir des solutions parcimonieuses pour un coût calculatoire raisonnable. Sur l'ensemble des

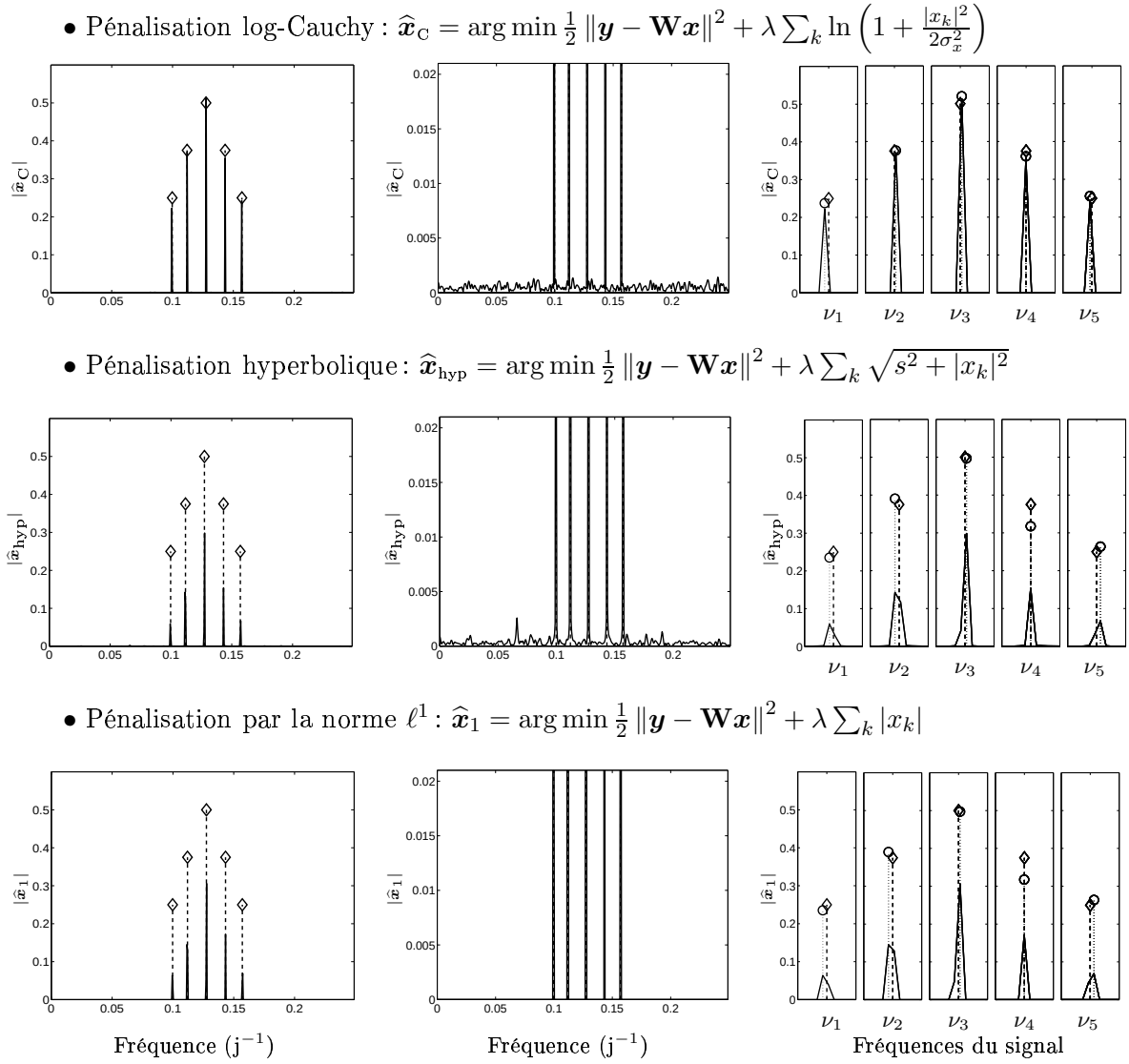


FIG. 3.4 – *Estimation pour des pénalisations non quadratiques. À gauche, estimé en module sur l'ensemble de la grille fréquentielle. Au centre, zoom en amplitude pour les faibles valeurs. À droite, zoom en fréquence autour de chaque raie. Les traits pleins représentent le module du spectre estimé, les ◇ les raies théoriques et, sur les figures de droite, les ○ correspondent à la réestimation a posteriori des amplitudes localisées par les pics de* $|\widehat{\boldsymbol{x}}|$.

résultats présentés dans ce chapitre, ce paramètre sera donc fixé à 10^{-3} sauf indication contraire.

3.3.2 Choix du paramètre de régularisation

Pour cette approche pénalisée, le choix du coefficient de régularisation λ est un point primordial dans la perspective d'obtenir une méthode *non supervisée*, c'est-à-dire permettant un réglage automatique de ce paramètre. La valeur de λ règle le compromis entre les deux sources d'information que sont la fidélité du spectre à reconstruire envers les données, exprimée par le terme quadratique $Q(\boldsymbol{x})$, et l'information *a priori* de parcimonie apportée par le terme de pénalisation $R(\boldsymbol{x})$. Ainsi le cas $\lambda \to 0$ (pas de pénalisation) correspond à celui de données « complètes », c'est-à-dire contenant suffisamment d'information pour estimer le spectre. Nous avons vu que, dans le cas sous-déterminé qui nous intéresse, le minimum du critère correspondant est nul et atteint pour une infinité de solutions. À l'inverse, lorsque $\lambda \to +\infty$, aucune information n'est apportée par les données $\boldsymbol{y}$ et seule l'information *a priori* de parcimonie est prise en compte ; le critère est alors minimal pour une solution identiquement nulle. Entre ces deux extrêmes, il s'agit donc de chercher une valeur, ou une plage de valeurs, du paramètre λ pour lesquelles le spectre $\widehat{\boldsymbol{x}}_\lambda$ est de nature parcimonieuse : dans le cas de la pénalisation ℓ^1, Fuchs [2004] montre ainsi que, pour des valeurs suffisamment faibles de λ, tous les minimiseurs $\widehat{\boldsymbol{x}}_\lambda$ ont leurs composantes non nulles localisées aux mêmes endroits[8]. Rappelons qu'il ne s'agit pas ici de déterminer *la* valeur optimale de λ – dans un sens qu'il faudrait d'ailleurs définir – mais plutôt un ordre de grandeur approximatif débouchant sur un estimateur satisfaisant.

Lorsque les données peuvent être accompagnées d'une estimation du niveau de bruit – en fonction des conditions observationnelles, des caractéristiques de l'instrument utilisé, etc. –, le paramètre λ peut être interprété en termes de *rapport signal sur bruit* (RSB). En effet, dans le cadre bayésien, le minimiseur du critère (3.2) correspond, en utilisant les notations du § 3.2.3, à l'estimateur du maximum *a posteriori* pour des perturbations $\boldsymbol{\epsilon}$ *i.i.d.* gaussiennes circulaires et une loi *a priori* $p(x_k) = k_\varphi \exp\left(-\lambda\varphi(|x_k|)/\sigma_\epsilon^2\right)$ sur chaque amplitude spectrale. En notant $y_n = \boldsymbol{l}_n\boldsymbol{x} + \epsilon_n$ où $\boldsymbol{l}_n$ est le $n^{\text{ème}}$ vecteur ligne de $\mathbf{W}$, la puissance du signal vaut :

$$\begin{aligned} \mathbb{E}[\,|\boldsymbol{l}_n\boldsymbol{x}|^2\,] &= \boldsymbol{l}_n\,\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\dagger]\,\boldsymbol{l}_n^\dagger = (2P+1)\,\sigma_x^2 \\ \text{où}\quad \sigma_x^2 &= \mathbb{E}[\,|x_k|^2\,] = k_\varphi \int_u u^2 \exp\left(-\frac{\lambda}{\sigma_\epsilon^2}\,\varphi(u)\right) du. \end{aligned}$$

et le rapport signal sur bruit s'écrit :

$$\mathrm{RSB}_{\mathrm{dB}} = 10\,\log_{10}(2P+1)\frac{\sigma_x^2}{\sigma_\epsilon^2},$$

où σ_x^2 est une fonction de $\lambda/\sigma_\epsilon^2$. Par exemple :

- Pour une pénalisation quadratique, c'est-à-dire pour des amplitudes supposées *i.i.d.* selon une loi gaussienne circulaire $x_k \sim \mathcal{N}_C(0, \sigma_x^2)$, nous avons vu au § 3.2.1 que $\sigma_x^2 = \frac{\sigma_\epsilon^2}{2\lambda}$ et donc $\mathrm{RSB}_{\mathrm{dB}} = 10\log_{10}\frac{2P+1}{2\lambda}$. La connaissance du RSB, ou conjointement de σ_x^2 et σ_ϵ^2, permet alors de fixer la valeur de λ.
- La pénalisation par la norme ℓ^1 correspond à une distribution *a priori* de Laplace des amplitudes spectrales : $p(x_k) \propto \exp\left(-\lambda|x_k|/\sigma_\epsilon^2\right)$, pour laquelle nous obtenons $\sigma_x^2 = \frac{2\sigma_\epsilon^4}{\lambda^2}$ et donc $\mathrm{RSB}_{\mathrm{dB}} = 10\log_{10}(2P+1)\,\frac{2\sigma_\epsilon^2}{\lambda^2}$. Ainsi, la connaissance conjointe du RSB et du niveau

8. Notons que cette propriété est établie pour des paramètres *réels* et surtout pour un modèle *non bruité*.

de bruit σ_ϵ^2 permet de fixer en conséquence la valeur de λ. Dans le cas de la pénalisation hyperbolique, qui peut être considérée comme une approximation de la pénalisation ℓ^1 pour une faible valeur de s, il est naturel de choisir λ de manière similaire.

Lorsque ces caractéristiques (RSB, σ_ϵ^2) ne sont pas disponibles, la sélection d'une valeur adéquate de λ reste un problème ouvert. Plusieurs approches rencontrées dans la littérature reposent sur le même schéma : un calcul préalable des estimés $\widehat{\boldsymbol{x}}_\lambda$ pour différentes valeurs de λ, puis le choix parmi ceux-ci de la meilleure solution au sens d'un critère à définir. Ainsi Sacchi *et al.* [1998] exploitent la statistique du bruit qui, toujours dans le cas *i.i.d.* centré gaussien, permet d'affirmer que, lorsque $\widehat{\boldsymbol{x}}_\lambda$ est la *vraie* solution parcimonieuse au problème $\boldsymbol{y} = \mathbf{W}\boldsymbol{x} + \epsilon$, le terme $\Psi(\widehat{\boldsymbol{x}}_\lambda) = (1/2\sigma_\epsilon^2) \left\| \boldsymbol{y} - \mathbf{W}\widehat{\boldsymbol{x}}_\lambda \right\|^2$ suit une statistique de type χ^2 à N degrés de liberté. Dans le cas où le niveau de bruit σ_ϵ est connu, ils proposent ainsi de choisir, parmi les solutions $\widehat{\boldsymbol{x}}_\lambda$, celle(s) débouchant sur une valeur du chi-carré $\Psi(\widehat{\boldsymbol{x}}_\lambda)$ inférieure à $N + 3.3\sqrt{N}$ [9]. Néanmoins, un telle approche peut conduire à une sur-régularisation de la solution [Idier, 2001a, chapitre 2]. En effet, l'estimée $\widehat{\boldsymbol{x}}_\lambda$ diffère nécessairement de la vraie solution, notamment par la localisation inexacte des raies sur une grille de valeurs discrètes, et l'erreur résiduelle $\boldsymbol{y} - \mathbf{W}\widehat{\boldsymbol{x}}_\lambda$ ne suit alors aucune distribution connue, mais sera probablement de niveau supérieur au seul niveau du bruit.

Dans le cas particulier de la recherche d'une solution parcimonieuse, une approche alternative consiste à sélectionner, parmi les solutions $\widehat{\boldsymbol{x}}_\lambda$ éventuellement seuillées pour éliminer les faibles amplitudes, celle satisfaisant au mieux un critère de sélection de l'ordre de type *Akaike Information Criterion* ou *Minimum Description Length*, que nous avons déjà évoqués au début du deuxième chapitre de ce manuscrit. Cette approche est proposée pour la pénalisation ℓ^1 par Fuchs [1999], qui relie la valeur de λ au nombre de composantes non nulles de $\widehat{\boldsymbol{x}}_\lambda$. L'interprétation d'un tel critère souffre néanmoins des mêmes limitations que celles évoquées au chapitre 2 : elle suppose en particulier que $\widehat{\boldsymbol{x}}_\lambda$ *est* la meilleure solution pour un modèle multi-sinusoïdal à un nombre de raies donné. Pour les raisons évoquées ci-dessus, cette hypothèse est ici encore optimiste.

Notons que l'interprétation bayésienne évoquée précédemment ouvre ce problème à l'application de nombreuses techniques statistiquement plus rigoureuses visant à *estimer*, simultanément aux paramètres spectraux d'intérêt $\boldsymbol{x}$, une valeur optimale du paramètre de régularisation [Idier, 2001a, chapitre 3]. Cependant, leur mise en œuvre peut se révéler d'un coût élevé et sans rapport avec l'enjeu qui est ici uniquement l'obtention d'un ordre de grandeur de λ fournissant une solution acceptable. Nous proposons donc d'utiliser un critère empirique pour le choix du paramètre λ, basé sur la « courbe en L », introduite par Hansen [1992]. Dans le cas d'une fonction de pénalisation R quadratique, ce dernier étudie l'allure de la courbe $\big(Q(\widehat{\boldsymbol{x}}_\lambda),R(\widehat{\boldsymbol{x}}_\lambda)\big)$ paramétrée en λ et tracée en échelle logarithmique, et montre que, pour une pénalisation quadratique, cette courbe a une forme de « L ». Un exemple est présenté sur la figure 3.5 : la partie verticale correspond à des valeurs trop faibles de λ pour lesquelles l'estimateur est peu sensible à la régularisation et la partie horizontale correspond à des valeurs de λ élevées, où l'estimateur ne prend plus en compte le terme d'attache aux données. L'auteur propose alors de choisir le paramètre λ correspondant au « coude » de la courbe. Bien qu'aucune étude théorique ne justifie l'emploi d'un tel critère dans le cas d'une pénalisation non quadratique, l'utilisation empirique d'une représentation similaire s'est révélée satisfaisante : nous utiliserons la courbe $\big(Q(\widehat{\boldsymbol{x}}_\lambda),R(\widehat{\boldsymbol{x}}_\lambda)\big)$ paramétrée en λ et tracée en échelle *linéaire*, ou la courbe $\big(Q(\widehat{\boldsymbol{x}}_\lambda),R(\widehat{\boldsymbol{x}}_\lambda)^2\big)$, afin de représenter sur les deux axes des grandeurs homogènes à une énergie. Lors des simulations effectuées, de telles courbes ont toujours présenté une forme de « L », bien que moins marquée que celle de la figure 3.5 pour

9. Si $\Psi \sim \chi^2_N$, on a $\mathbb{E}(\Psi) = N$ et la valeur maximale acceptable pour un niveau de confiance à 99% est d'environ $N + 3.3\sqrt{N}$.

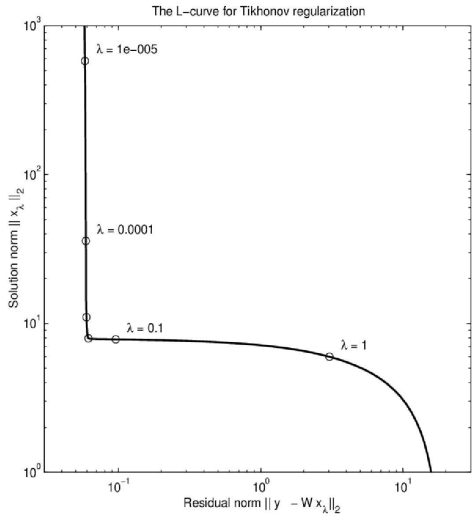


FIG. 3.5 – *Exemple de courbe en L dans le cas d'une pénalisation quadratique [Hansen, 1992].*

une pénalisation quadratique.

Notons encore que, l'estimateur $\widehat{\boldsymbol{x}}_\lambda$ étant calculé pour une large gamme de valeurs de λ, il est également possible de sélectionner le spectre présentant visuellement le meilleur rendu, comme le font Ciuciu *et al.* [2001]. Néanmoins, l'utilisation d'un critère comme la courbe en L confère à la procédure d'estimation un caractère un peu plus « automatisé ».

Exemple sur un signal simple

Reprenant notre exemple habituel, nous traçons sur la figure 3.6 les courbes en « L » correspondant aux pénalisations hyperbolique avec $s = 10^{-3}$ et par la norme ℓ^1. Nous représentons également en module les spectres estimés en fonction de λ. Pour ces deux pénalisations non quadratiques, les courbes $\big(Q(\widehat{\boldsymbol{x}}_\lambda),R(\widehat{\boldsymbol{x}}_\lambda)\big)$ sont d'allures très proches et présentent un « L » bien moins marqué que dans le cas quadratique. Notons que le tracé de $\big(Q(\widehat{\boldsymbol{x}}_\lambda),R(\widehat{\boldsymbol{x}}_\lambda)^2\big)$, plus « judicieux » au niveau de l'homogénéité des grandeurs représentées, permet en oudre d'obtenir un coude plus marqué.

Si la valeur prise dans ce coude (soit $\log\lambda \simeq 1.5$) fournit une solution satisfaisante, les figures 3.6 en bas montrent que la plage de valeurs fournissant un estimateur acceptable est relativement étroite, soit ici $1 \leq \log\lambda \leq 1.7$: l'estimation sur cet exemple est donc assez sensible au réglage de λ. Si la courbe en « L » permet de localiser approximativement la plage de valeurs de λ acceptables, la procédure d'estimation qui résulte de son usage ne s'affranchit donc pas totalement de la supervision de l'utilisateur.

Rappelons également que sur ces simulations un rapport signal sur bruit de 10 dB et un niveau de bruit correspondant de $\sigma_\epsilon = 0.36$ ont été utilisés [10]. Conformément à l'interprétation du début de ce paragraphe, la valeur de λ correspondante est alors de $\lambda = \sqrt{2\,(2P+1)}\sigma_\epsilon\, 10^{-\frac{\mathrm{RSB_{dB}}}{20}}$, soit $\log_{10}\lambda \simeq 1.2$, qui se trouve dans le coude du « L ».

10. Le spectre de ce signal n'a cependant pas été généré comme la réalisation d'un processus aléatoire selon une loi de Laplace et les amplitudes des cinq raies spectrales ont été manuellement fixées ; le RSB a alors été défini de la manière suivante : $\mathrm{RSB_{dB}} = 10\log_{10}\frac{P_y}{\sigma_\epsilon^2}$, avec $P_y = \frac{1}{2}\sum_{j=1}^{5}\rho_j^2$, où ρ_j est le module associé à la $j^{\text{ème}}$ sinusoïde du modèle.

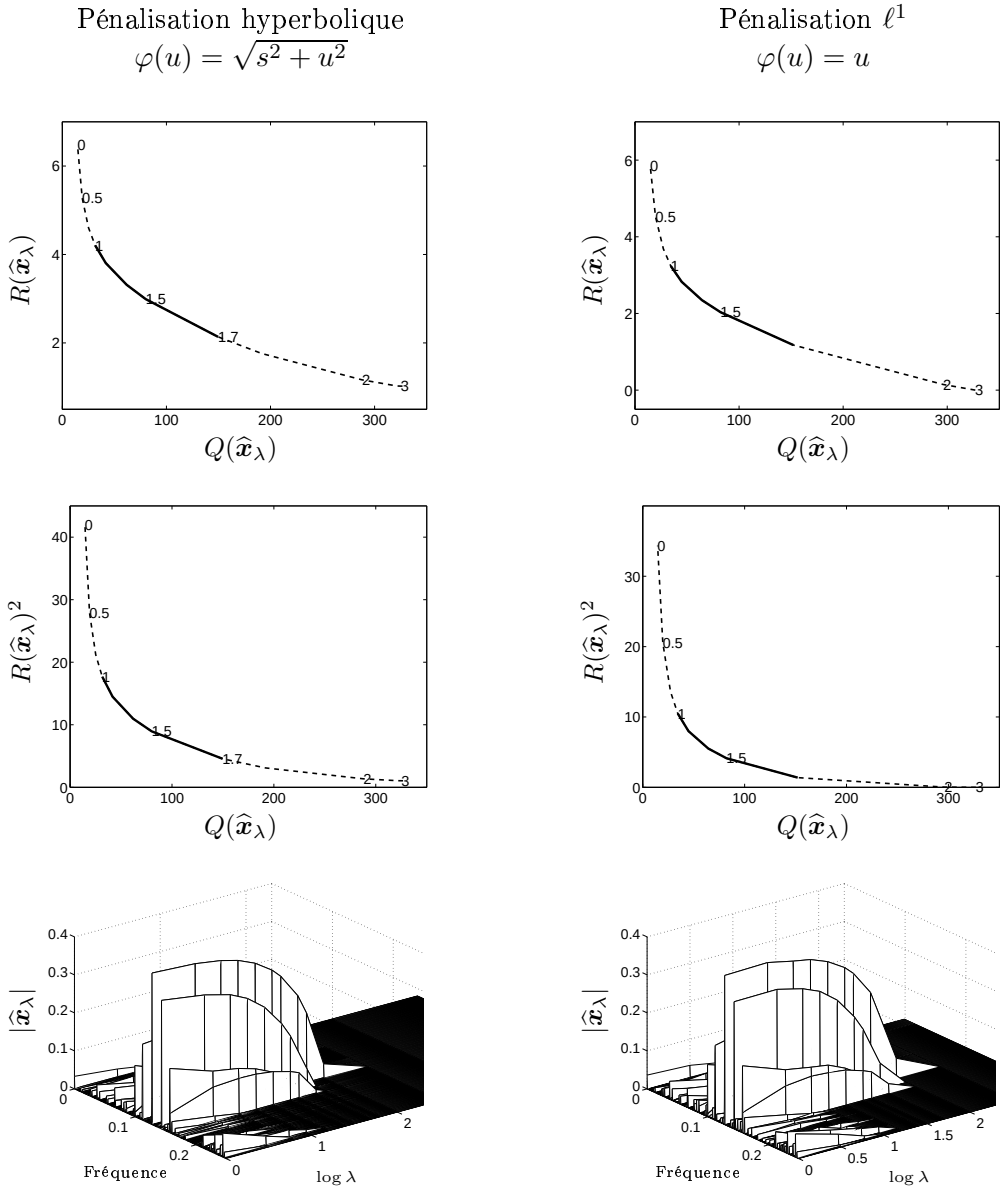


FIG. 3.6 – *En haut, courbes* $\big(Q(\widehat{\boldsymbol{x}}_\lambda),R(\widehat{\boldsymbol{x}}_\lambda)\big)$ *paramétrées en* $\log_{10}\lambda$ *pour les trois types de pénalisation* φ*, où* $R(\boldsymbol{x}) = \sum_k \varphi(|x_k|)$*. Au centre, courbes* $\big(Q(\widehat{\boldsymbol{x}}_\lambda),R(\widehat{\boldsymbol{x}}_\lambda)^2\big)$ *pour les pénalisations hyperbolique et* ℓ^1*. Le trait plein représente la zone fournissant une estimation acceptable. En bas, estimés* $|\widehat{\boldsymbol{x}}_\lambda|$ *en fonction de* λ*.*

3.4 Algorithmes d'optimisation pour une pénalisation hyperbolique

Nous reprenons ici le formalisme introduit pour l'estimation de spectres de raies par Sacchi *et al.* [1998] et repris par Ciuciu *et al.* [2001], qui consiste à minimiser un critère de type « moindres carrés » pénalisé par une fonction de régularisation. Alors que les premiers utilisent la fonction non convexe $\phi_{\mathrm{C}}(x_k) = \ln(1 + |x_k|^2/2\sigma_x^2)$, nous utilisons ici la branche hyperbolique, strictement convexe, utilisée par les seconds : $\phi_{\mathrm{hyp}}(x_k) = \sqrt{s^2 + |x_k|^2}$ qui présente l'ensemble des propriétés analytiques garantissant existence et unicité du minimum $\widehat{\boldsymbol{x}}_{\mathrm{hyp}}$ du critère asssocié :

$$J_{\mathrm{hyp}}(\boldsymbol{x}) = \frac{1}{2}\left\|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\right\|^2 + \lambda R_{\mathrm{hyp}}(\boldsymbol{x}), \quad R_{\mathrm{hyp}}(\boldsymbol{x}) = \sum_k \sqrt{s^2 + |x_k|^2}. \tag{3.3}$$

La pénalisation envisagée étant strictement convexe et continûment différentiable, la minimisation de J_{hyp} est alors équivalente à l'annulation du gradient [11] :

$$\overrightarrow{\nabla J_{\mathrm{hyp}}}(\boldsymbol{x}) = -\mathbf{W}^\dagger(\boldsymbol{y} - \mathbf{W}\boldsymbol{x}) + \lambda\overrightarrow{\nabla R}_{\mathrm{hyp}}(\boldsymbol{x}), \quad \overrightarrow{\nabla R}_{\mathrm{hyp}}(\boldsymbol{x})\big|_k = \phi'(x_k) = \frac{x_k}{\sqrt{s^2 + |x_k|^2}}.$$

Nous envisageons différentes solutions algorithmiques visant à la minimisation du critère (3.3) et en particulier les conséquences sur ces solutions du caractère irrégulier de l'échantillonnage. En effet, les procédures d'optimisation rencontrées chez Sacchi *et al.* et Ciuciu *et al.* tirent parti de la régularité de l'échantillonnage et de la similarité résultante des matrices mises en jeu avec des opérateurs de Fourier : dans le cas de l'échantillonnage régulier, éventuellement à données manquantes, les produits $\mathbf{W}\cdot$ s'identifient à des transformées de Fourier discrètes (TFD) inverses et sont donc calculables rapidement par des algorithmes FFT. De même, les produits $\mathbf{W}^\dagger\cdot$ sont des opérations de TFD. Par contre, pour un échantillonnage totalement irrégulier, $\mathbf{W}$ ne présente aucune structure particulière exploitable et le coût du calcul associé aux produits $\mathbf{W}\cdot$ devient rapidement prohibitif lorsque la dimension de l'espace des paramètres augmente. La préservation de la structure de Toeplitz de la matrice $\mathbf{W}^\dagger\mathbf{W}$ pourra toutefois être mise à profit pour résoudre à moindre coût certains systèmes linéaires ou calculer des produits de type $\mathbf{W}^\dagger\mathbf{W}\cdot$.

3.4.1 Gradients Conjugués

Une approche classique pour la minimisation du critère (3.3) consiste à utiliser un algorithme de descente de type gradients conjugués (GC). Celle-ci peut sembler à première vue inappropriée pour la restauration de solutions parcimonieuses dans le sens où, pour un algorithme de GC, les composantes de basses fréquences de la solution tendent à converger plus rapidement que celles de hautes fréquences [12] (voir par exemple les résultats expérimentaux de Sauer et Bouman [1993]).

Cependant, la mise en œuvre d'un algorithme de GC permet d'exploiter les différentes structures spécifiques étudiées au § 3.1. Ainsi pour un échantillonnage régulier, éventuellement troué,

11. La dérivation d'une fonction $f : \mathbb{C} \to \mathbb{R}$ s'écrit $f'(u) = 2\dfrac{\partial g(u,u^*)}{\partial u^*}$, où $f(u) = g(u,u^*)$ et en considérant indépendamment u et u^* [Brandwood, 1983].

12. Notons bien que les termes de basses et hautes *fréquences* employés ici ne font pas référence à la distribution spectrale de l'énergie du signal $\boldsymbol{y}$ le long de la grille fréquentielle $\mathcal{G}$, mais est à entendre au sens de la *forme* du vecteur $\boldsymbol{x}$: les basses fréquences correspondent à des variations douces des composantes x_k lorsque k varie de $-P$ à P, alors que les hautes fréquences correspondent à des variations rapides, ce qui est le cas lorsque l'on recherche une solution parcimonieuse.

l'écriture de $\mathbf{W}$ à partir d'une matrice de Fourier permet de calculer le critère J_{hyp} au moyen d'un algorithme FFT sur $2P$ points (à condition, bien sûr, que $2P$ soit une puissance de 2), ainsi que son gradient au moyen d'un double algorithme FFT sur $2P$ points. Pour un échantillonnage totalement irrégulier, l'absence de propriétés structurelles intéressantes de $\mathbf{W}$ alourdit les coûts de calcul des produits de type $\mathbf{W}\cdot$. Par contre, l'exploitation de la structure de Toeplitz de la matrice $\mathbf{W}^\dagger\mathbf{W}$ permet de réduire la complexité de calcul du critère J_{hyp} ainsi que de son gradient. En effet, d'une part, le terme quadratique de J_{hyp} peut s'écrire sous la forme :

$$\|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|^2 = \boldsymbol{y}^\dagger\boldsymbol{y} - 2\Re(\boldsymbol{x}^\dagger\boldsymbol{u}) + \boldsymbol{x}^\dagger\mathbf{W}^\dagger\mathbf{W}\boldsymbol{x} \text{ où } \boldsymbol{u} = \mathbf{W}^\dagger\boldsymbol{y}$$

et, d'autre part, la complexité du calcul du gradient $\overrightarrow{\nabla J}_{\text{hyp}}$ réside essentiellement dans le calcul du produit $\mathbf{W}^\dagger\mathbf{W}\boldsymbol{x}$. Or, en enveloppant la matrice $\mathbf{W}^\dagger\mathbf{W}$ dans une matrice *circulante* deux fois plus grande, les produits $\mathbf{W}^\dagger\mathbf{W}\cdot$ sont calculables par un double algorithme FFT sur $4P+2$ points (voir par exemple [Vio *et al.*, 2000]). Ainsi, quelle que soit la nature de l'échantillonnage, critère et gradient sont calculables par des algorithmes rapides. Le tableau 3.1 résume l'ordre de grandeur des coûts de calcul associés aux critères et gradients pour un échantillonnage régulier et irrégulier, en rappelant que le coût de calcul d'un algorithme FFT sur L points est de l'ordre de $\frac{L}{2}\log_2 L$ multipilications complexes.

	calcul de J_{hyp}	calcul de $\overrightarrow{\nabla J_{\text{hyp}}}$
Échantillonnage régulier	$P\log_2(2P)$	$2P\log_2(2P)$
Échantillonnage irrégulier	$4P\log_2(4P)$	$4P\log_2(4P)$

TAB. 3.1 – *Comparaison des coûts de calcul du critère J_{hyp} et de son gradient en fonction de la nature de l'échantillonnage.*

Par la suite, nous envisagerons deux types d'algorithmes de type GC. Le premier consiste en une version « classique » des gradients conjugués (la mise en œuvre utilisera ici la version de Polak et Ribière [Press *et al.*, 1992]), alors que le second s'affranchit d'une recherche exacte du pas optimal de descente. Sun et Zhang [2001] proposent ainsi un algorithme de GC basé uniquement sur des calculs successifs du gradient : pour chaque direction de descente, l'étape de minimisation monodimensionnelle déterminant la valeur optimale du pas est remplacée par une valeur sous-optimale, obtenue uniquement à partir du gradient. L'avantage d'une telle procédure réside alors essentiellement dans sa plus grande simplicité, tout en disposant d'une garantie de convergence [Sun et Zhang, 2001].

3.4.2 L'algorithme IRLS

L'algorithme IRLS (*Iterative Reweighted Least-Squares*) a été initialement introduit pour la minimisation de critères de la forme $\mathcal{K}(\boldsymbol{x}) = \sum_{n=1}^{N}\phi(y_n - \boldsymbol{w}_n\boldsymbol{x})$, où $\boldsymbol{w}_n$ est le $n^{\text{ème}}$ vecteur ligne de $\mathbf{W}$, et dans un contexte de déconvolution ℓ^p : $\phi(u) = |u|^p$ ou $\mathcal{K}(\boldsymbol{x}) = \|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|_P$ avec $1 \leq p \leq 2$ [Byrd et Payne, 1979, Yarlagadda *et al.*, 1985, les seconds auteurs faisant référence au rapport interne des premiers, qui semble être introuvable]. Son application à la minimisation d'un critère pénalisé de type (3.2) a été proposée, dans le cadre de l'analyse spectrale, par Sacchi *et al.* [1998] pour la régularisation par *a priori* de Cauchy et par Ciuciu *et al.* [2001] pour la pénalisation hyperbolique que nous étudions ici. L'algorithme est basé sur l'introduction d'une quantité annexe $q_k(x_k) = x_k/\phi'_{\text{hyp}}(x_k)$, qui permet d'écrire l'annulation du gradient $\overrightarrow{\nabla J}_{\text{hyp}}$ sous la forme équivalente de l'équation implicite :

$$(\mathbf{W}^\dagger\mathbf{W} + \lambda\mathbf{Q}_{\boldsymbol{x}}^{-1})\,\boldsymbol{x} = \mathbf{W}^\dagger\boldsymbol{y},$$

où $\mathbf{Q}_{\boldsymbol{x}}$ est la matrice diagonale d'éléments $q_k(x_k)$. Notons que cette écriture correspond, en omettant la dépendance en $\boldsymbol{x}$ de la matrice $\mathbf{Q}_{\boldsymbol{x}}$, à la recherche de la solution régularisée pour une pénalisation quadratique de type $R(\boldsymbol{x}) = \boldsymbol{x}^\dagger \mathbf{Q}^{-1} \boldsymbol{x}$ (voir le § 3.2.1). L'algorithme IRLS consiste alors en un schéma algorithmique itératif de type « recherche de point fixe », basé sur l'alternance des étapes :

i) actualisation de $\mathbf{Q}^{(t)} = \mathrm{diag}(q_k^{(t)})$ avec $q_k^{(t)} = \dfrac{x_k^{(t)}}{\phi'_{\mathrm{hyp}}\left(x_k^{(t)}\right)} = \sqrt{s^2 + \left|x_k^{(t)}\right|^2}$;

ii) résolution du système linéaire correspondant : $\boldsymbol{x}^{(t+1)} = \left(\mathbf{W}^\dagger \mathbf{W} + \lambda \mathbf{Q}^{(t)^{-1}}\right)^{-1} \mathbf{W}^\dagger \boldsymbol{y}$;

iii) $t \leftarrow t+1$ et retour en i) jusqu'à convergence.

Plusieurs jeux de conditions sur la fonction de pénalisation ϕ ont été établis, assurant la convergence d'un tel algorithme vers le minimum d'un critère convexe [voir notamment Yarlagadda *et al.*, 1985, Bube et Langan, 1997, Idier, 2001b][13]. Dans le cas de la branche d'hyperbole $\phi = \phi_{\mathrm{hyp}}$, les conditions proposées par Bube et Langan [1997] sont vérifiées ; l'approche proposée par Idier [2001b] interprétant l'IRLS comme un algorithme de minimisation d'un critère semi-quadratique permet également d'établir cette convergence.

Au niveau calculatoire, le coût essentiel est celui de la résolution du système linéaire (ii). Sacchi *et al.* [1998] proposent à cette fin d'utiliser l'égalité :

$$\left(\mathbf{W}^\dagger \mathbf{W} + \lambda \mathbf{Q}^{(t)^{-1}}\right)^{-1} \mathbf{W}^\dagger \boldsymbol{y} = \mathbf{Q}^{(t)}\, \mathbf{W}^\dagger \left(\lambda \mathbf{I}_N + \mathbf{W}\, \mathbf{Q}^{(t)}\, \mathbf{W}^\dagger\right)^{-1} \boldsymbol{y}$$

où l'écriture du terme de droite correspond à l'inversion d'un système de taille $N \times N$ inférieure, dans le cas d'une analyse à haute résolution, à la taille de $(2P+1) \times (2P+1)$ intervenant dans la première formulation. De manière similaire à l'étude menée au § 3.1, il est facile de voir que la matrice à inverser dans l'expression de droite a une structure de *Toeplitz* si et seulement si l'échantillonnage est régulier. Dans ce cas, son inversion peut être réalisée à moindre coût par une procédure récursive de type Levinson. Dans le cas d'un échantillonnage régulier troué ou de l'échantillonnage irrégulier, la matrice à inverser n'est plus de Toeplitz. Sacchi *et al.* suggèrent alors de recourir à une décomposition de Cholesky pour réaliser l'inversion, mais le coût de calcul associé à chaque itération de l'algorithme IRLS s'en trouve fortement alourdi. Notons encore que la matrice $\mathbf{W}^\dagger \mathbf{W} + \lambda \mathbf{Q}^{(t)^{-1}}$ intervenant dans la formulation initiale à gauche n'est pas de Toeplitz à cause de l'influence du terme diagonal $\mathbf{Q}^{(t)^{-1}}$.

3.4.3 Algorithme de type Gauss-Seidel

Une alternative à la complexité calculatoire, liée aux produits matriciels et aux résolutions de systèmes linéaires de dimension élevée, consiste à envisager uniquement des minimisations *scalaires* successivement par rapport à chacun des paramètres x_k. Cette démarche a notamment été envisagée en traitement d'images, où chaque pixel est mis à jour de manière individuelle pour un coût calculatoire réduit [Sauer et Bouman, 1993, Brette et Idier, 1996], et s'est avérée particulièrement adaptée à la restauration de solutions présentant des « variations rapides », au

13. Notons que chez les deux premiers, le problème envisagé est celui de la minimisation d'un critère du type $\mathcal{K}(\boldsymbol{x}) = \sum_{n=1}^{N} \phi(y_n - \boldsymbol{w}_n \boldsymbol{x})$ qui ne débouche pas exactement sur la même formulation de l'IRLS. Néanmoins, les conditions de convergence portant sur la fonction ϕ sont les mêmes [Ciuciu et Idier, 2002], où dans notre cas ϕ est la pénalisation opérant sur chaque variable x_k.

sens où nous l'avons entendu au § 3.4.1. En effet, contrairement à une approche de type gradient, la mise à jour de chaque paramètre de manière indépendante favorise la restauration de « hautes fréquences » dans la forme du spectre $\boldsymbol{x}$, c'est-à-dire l'existence de raies.

Minimisation explicite

L'intérêt d'un tel algorithme réside dans la possibilité de réaliser facilement chaque minimisation scalaire. Dans le cas de la pénalisation hyperbolique, la minimisation du critère J_{hyp} en la variable complexe x_k peut être obtenue analytiquement et nécessite un coût calculatoire réduit. Plus précisément, nous montrons en annexe A.3 que la résolution de :

$$\mathring{x_k} = \arg\min_{x_k} J_{\text{hyp}}(\boldsymbol{x}) \Leftrightarrow \mathring{x_k} = \arg\min_{x_k} f(x_k), \;\; f(x_k) = \frac{1}{2} \left\| \boldsymbol{e}_k - \boldsymbol{w}_k x_k \right\|^2 + \lambda \sqrt{s^2 + |x_k|^2}$$

nécessite pour tout calcul la recherche des racines d'un polynôme de degré quatre, pour laquelle les solutions sont explicites [Abramowitz et Stegun, 1964, pages 17–18], puis le calcul de la valeur de f en ces quatre racines. L'algorithme GS_{hyp}(algorithme de Gauss-Seidel pour une pénalisation hyperbolique) consiste alors en le schéma itératif suivant :

1. Itération t : pour $k = -P \dots P$, réaliser les opérations :
   i) calcul de $\boldsymbol{e}_k^{(t)} = \boldsymbol{y} - \sum_{\ell<k} \boldsymbol{w}_\ell \, x_\ell^{(t)} - \sum_{\ell>k} \boldsymbol{w}_\ell \, x_\ell^{(t-1)}$;
   ii) minimisation en x_k de $f(x_k) = \frac{1}{2} \left\| \boldsymbol{e}_k^{(t)} - \boldsymbol{w}_k x_k \right\|^2 + \lambda \sqrt{s^2 + |x_k|^2}$ selon le schéma explicité à l'annexe A.3 ;
   iii) mise à jour de $x_k^{(t)}$;
2. $t \leftarrow t + 1$ et retour en 1 jusqu'à convergence.

Cette procédure ne requiert donc aucun calcul en dimension élevée et n'exploite par conséquent aucune structure particulière liée à la nature de l'échantillonnage des données, mais la mise à jour sous forme séquentielle à l'étape i) et le calcul des racines d'un polynôme d'ordre quatre pour l'étape ii) nécessitent un coût limité à chaque itération. Notons que la fonction f à minimiser à l'étape ii) est strictement convexe, ce qui garantit l'existence et l'unicité de son minimiseur. De même, la stricte convexité du critère J_{hyp} assure la convergence de la procédure globale de Gauss-Seidel vers son unique minimiseur [Bertsekas, 1995, chapitre 2, proposition 2.7.1].

Remarquons ici que le choix d'une autre fonction de type «$\ell^2\ell^1$» que la branche d'hyperbole pourrait s'avérer plus judicieux pour ce type d'algorithme. En effet, nous montrons à l'annexe A.3 que la minimisation de la fonction $f(x_k)$ introduite ci-dessus revient à résoudre une équation de la forme $a\rho - b + \lambda\varphi'(\rho)$ où $\rho = |x_k|$ et φ est la fonction de pénalisation opérant sur le module de chaque composante. Ainsi pour la fonction de Huber $\varphi_H(\rho) = \begin{cases} \frac{1}{2}\rho^2 \text{ pour } \rho < h \\ h\rho - \frac{1}{2}\rho^2 \text{ pour } \rho \geq h \end{cases}$, la minimisation de f est quasiment immédiate puisqu'elle revient à rechercher les racines de deux polynômes du premier degré. La pénalisation par la norme ℓ^1, soit $\varphi_1(\rho) = \rho$ donc $\varphi_1'(\rho) = 1$, est à ce titre optimale : nous y reviendrons au § 3.5.4.

Minimisation semi-quadratique

Une option alternative pour la minimisation de l'étape 1ii) est d'utiliser un algorithme semblable à l'IRLS, mais ici pour des variables scalaires : une démarche similaire est adoptée

dans [Brette et Idier, 1996] avec une pénalisation similaire de la forme $\ell^2\ell^1$, sur des variables réelles et pour un problème de déconvolution d'images. En utilisant les notations introduites au paragraphe précédent, la minimisation en la seule variable x_k du critère J_{hyp} revient en effet à résoudre l'équation implicite :

$$\left(\boldsymbol{w}_k^\dagger \boldsymbol{w}_k + \frac{\lambda}{q_k(x_k)}\right) x_k = \boldsymbol{w}_k^\dagger \boldsymbol{e}_k \text{ avec } q_k(x_k) = \sqrt{s^2 + |x_k|^2},$$

où $\boldsymbol{w}_k$ représente la colonne d'indice k de la matrice $\mathbf{W}$ et $\boldsymbol{e}_k = \boldsymbol{y} - \sum_{\ell \neq k} \boldsymbol{w}_\ell x_\ell$. L'étape de minimisation 1ii) peut alors être remplacée par le schéma itératif suivant :

ii_a) initialisation $x_*^{(0)} = x_k^{(t-1)}$;

ii_b) calcul de $q_*^{(u)} = \dfrac{x_*^{(u)}}{\phi'_{\text{hyp}}\left(x_*^{(u)}\right)} = \sqrt{s^2 + \left|x_*^{(u)}\right|^2}$;

ii_c) mise à jour de $x_*^{(u+1)} = \dfrac{\boldsymbol{w}_k^\dagger \boldsymbol{e}_k^{(t)}}{\boldsymbol{w}_k^\dagger \boldsymbol{w}_k + \lambda / q_*^{(u)}}$ où : $\left\{ \begin{array}{l} \boldsymbol{w}_k^\dagger \boldsymbol{w}_k = N \\ \boldsymbol{w}_k^\dagger \boldsymbol{e}_k^{(t)} \text{ a été calculé à l'étape 1i)} \end{array} \right.$;

ii_d) $u \leftarrow u+1$ et retour en ii_b) jusqu'à convergence, puis mise à jour de $x_k^{(t)} = x_*^{(u)}$;

Si elle semble plus lourde à mettre en œuvre, cette deuxième forme s'est révélée moins coûteuse en calcul que la minimisation explicite par les formules de l'annexe A.3. En effet, il résulte que seulement quelques itérations ii_b)-ii_c) sont nécessaires en pratique, qui sont très peu coûteuses en temps de calcul, alors que le calcul des quatre racines d'un polynôme d'ordre 4, puis l'évaluation de la fonction scalaire f en ces valeurs nécessitent plus de temps (un gain de temps cpu d'un facteur environ 2 est ainsi obtenu). Remarquons, ici encore, que la stricte convexité de la fonction f assure la convergence de la minimisation scalaire ii_a)-ii_d) par un algorithme de type IRLS.

3.5 Algorithmes d'optimisation pour une pénalisation ℓ^1

La pénalisation par la norme ℓ^1, c'est-à-dire $\phi_1(x) = |x|$, bien que théoriquement satisfaisante en termes de parcimonie, rend la minimisation du critère pénalisé :

$$J_1(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|^2 + \lambda \|\boldsymbol{x}\|_1 \tag{3.4}$$

relativement complexe à cause de la non différentiabilité de $\|\boldsymbol{x}\|_1$ en tout vecteur ayant un élément nul. En particulier, toute méthode d'optimisation basée sur l'utilisation du gradient :

$$\overrightarrow{\nabla J_1}(\boldsymbol{x}) = -\mathbf{W}^\dagger(\boldsymbol{y} - \mathbf{W}\boldsymbol{x}) + \lambda \operatorname{sgn}(\boldsymbol{x})$$

est délicate, puisque celui-ci n'est pas défini lorsqu'une composante x_k s'annule. Rappelons par ailleurs que le critère J_1 n'est pas strictement convexe et que l'unicité du minimiseur ne peut être garantie : la convexité simple permet uniquement d'affirmer que ses minima donnent la même valeur du critère et qu'ils forment un ensemble convexe.

Nous étudions maintenant diverses solutions algorithmiques visant à la minimisation de J_1. Alors que la première concerne la programmation quadratique, pour laquelle nous montrons sa délicate extension au cas de paramètres à valeurs complexes, les deux approches suivantes peuvent être vues comme des généralisations des algorithmes étudiés précédemment au cas d'une

pénalisation non strictement convexe et non différentiable en 0. Plus précisément, nous étudions l'adaptation des algorithmes IRLS et de Gauss-Seidel au cas de la pénalisation ℓ^1, dont nous verrons que le principal handicap réside dans la limite infinie de $\phi_1'(x)/x$ lorsque $x \to 0$. Alors que le premier permet une mise en œuvre très intéressante en termes de temps d'exécution, il souffre de lacunes théoriques au niveau des propriétés de convergence ; nous établirons toutefois une condition nécessaire et suffisante vérifiée par tout minimiseur du critère (3.4), permettant de vérifier *a posteriori* la validité de la solution. L'algorithme de Gauss-Seidel étudié par la suite présente quant à lui des propriétés plus solides au niveau de la convergence.

3.5.1 Programmation quadratique

Une première approche à la minimisation d'un critère comme J_1 consiste à réécrire le problème d'optimisation sous la forme de la minimisation d'un problème linéaire sous contrainte [Fuchs, 2004]. En effet, pour des paramètres x_k *réels*, il est possible d'écrire $\|\boldsymbol{x}\|_1$ sous forme linéaire en séparant $x_k^+ = \max(x_k,0) \geq 0$ et $x_k^- = \max(-x_k,0) \geq 0$. On a alors $\boldsymbol{x} = \boldsymbol{x}^+ - \boldsymbol{x}^-$ et $\|\boldsymbol{x}\|_1 = \boldsymbol{x}^+ + \boldsymbol{x}^-$. La minimisation du critère (3.4) revient encore à minimiser en $[\boldsymbol{x}^+,\boldsymbol{x}^-]$ la forme quadratique :

$$\frac{1}{2}\begin{pmatrix} \boldsymbol{x}^+ \\ \boldsymbol{x}^- \end{pmatrix}^{\mathrm{t}} \begin{pmatrix} \mathbf{W}^{\mathrm{t}}\mathbf{W} & -\mathbf{W}^{\mathrm{t}}\mathbf{W} \\ -\mathbf{W}^{\mathrm{t}}\mathbf{W} & \mathbf{W}^{\mathrm{t}}\mathbf{W} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^+ \\ \boldsymbol{x}^- \end{pmatrix} + \begin{pmatrix} \lambda - \boldsymbol{y}^{\mathrm{t}}\mathbf{W}, & \lambda + \boldsymbol{y}^{\mathrm{t}}\mathbf{W} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^+ \\ \boldsymbol{x}^- \end{pmatrix}$$

sous la contrainte $\boldsymbol{x}^+,\boldsymbol{x}^- \geq 0$. Une telle minimisation est alors réalisable, en utilisant par exemple des algorithmes de programmation quadratique [Bertsekas, 1995, chapitres 2 et 3].

Cependant, l'adaptation d'une telle démarche pour des paramètres x_k *complexes* est délicate, puisqu'elle nécessite un reparamétrage du problème afin de se ramener à une écriture linéaire de la norme ℓ^1 et à des contraintes de positivité : comme nous l'avons expliqué au § 3.2.2, nous sommes alors amenés à considérer la formulation à paramètres réels (2.13) introduite à la fin du deuxième chapitre et par conséquent une pénalisation indépendante des partie réelles et imaginaires. En utilisant les inconnues $\boldsymbol{s}_k = [a_k, b_k] \in \mathbb{R}^2$ au lieu de $x_k \in \mathbb{C}$, avec $x_k = (a_k - j\, b_k)/2$, et l'opérateur $\mathbf{H}$ de taille $N \times (2P+2)$ tel que $\mathbf{H}_{n,2k} = \cos(2\pi f_k t_n)$ et $\mathbf{H}_{n,2k+1} = \sin(2\pi f_k t_n)$, l'écriture du critère sous une forme quadratique contraint alors à minimiser :

$$\widetilde{J}_1(\boldsymbol{s}^+,\boldsymbol{s}^-) = \frac{1}{2}\begin{pmatrix} \boldsymbol{s}^+ \\ \boldsymbol{s}^- \end{pmatrix}^{\mathrm{t}} \begin{pmatrix} \mathbf{H}^{\mathrm{t}}\mathbf{H} & -\mathbf{H}^{\mathrm{t}}\mathbf{H} \\ -\mathbf{H}^{\mathrm{t}}\mathbf{H} & \mathbf{H}^{\mathrm{t}}\mathbf{H} \end{pmatrix} \begin{pmatrix} \boldsymbol{s}^+ \\ \boldsymbol{s}^- \end{pmatrix} + \begin{pmatrix} \lambda - \boldsymbol{y}^{\mathrm{t}}\mathbf{H} & \lambda + \boldsymbol{y}^{\mathrm{t}}\mathbf{H} \end{pmatrix} \begin{pmatrix} \boldsymbol{s}^+ \\ \boldsymbol{s}^- \end{pmatrix}$$

sous la contrainte $\boldsymbol{s}^+ \geq 0$, $\boldsymbol{s}^- \geq 0$, avec $\boldsymbol{s}^+ = \max(\boldsymbol{s},0)$, $\boldsymbol{s}^- = \max(-\boldsymbol{s},0)$. La minimisation de $\widetilde{J}_1$ n'est cependant pas équivalente à celle du critère (3.4), puisque la pénalisation résultante porte sur la norme ℓ^1 des paramètres $\boldsymbol{s}$, c'est-à-dire les valeurs absolues des parties réelles et imaginaires des amplitudes complexes $\boldsymbol{x}$, plutôt que sur les modules complexes. Cette approximation, bien que nécessaire pour la réécriture du problème de minimisation sous une forme quadratique, est cependant bien moins satisfaisante en termes de modélisation *a priori*, comme nous l'avons illustré au § 3.2.2. Nous envisageons donc d'autres solutions algorithmiques permettant de prendre en compte une véritable pénalisation par la norme $\|x\|_1 = \sum_k |x_k|$.

3.5.2 Algorithme IRLS

L'utilisation d'un algorithme de type IRLS, que nous avons décrit au § 3.4.1, a déjà été envisagée pour la minimisation d'un critère pénalisé par la norme ℓ^1 [Figueiredo, 2003, Davies

et Daudet, 2003]. À notre connaissance, cependant, aucune démonstration de la convergence de l'IRLS vers un minimiseur du critère (3.4) n'a été publiée jusqu'à présent. De manière plus générale, pour une pénalisation par la norme ℓ^p avec $1 \leq p < 2$, les conditions de convergence proposées par Yarlagadda *et al.* [1985] et celles obtenues par l'interprétation semi-quadratique de Idier [2001b] sont mises en échec par la limite infinie de $\frac{\phi'(x)}{x}$ lorsque x tend vers 0 ; les conditions établies par Bube et Langan [1997] ne sont pas satisfaites non plus [14] à cause de la limite infinie de $\phi''(x)$ en 0. Il semble toutefois que les travaux en cours de Fuchs [2005] aient établi une preuve de convergence, encore non publiée à ce jour, dans le cas où les inconnues sont réelles. Ceci peut paraître de faible intérêt au niveau des propriétés de l'estimateur puisque l'on peut en principe approcher la fonction $\phi_1(x) = |x|$ par la branche d'hyperbole $\phi_{\text{hyp}}(x) = \sqrt{s^2 + |x|^2}$ avec une valeur suffisamment faible de s. Cependant, nous avons vu au § 3.2.3 que la pénalisation ℓ^1 fournissait une estimation rigoureusement parcimonieuse, c'est-à-dire exactement nulle en dehors d'un faible nombre de points. De plus, au niveau de la mise en œuvre algorithmique, l'association d'une pénalisation de type ℓ^1 à l'IRLS permet de bénéficier d'une propriété intéressante non vérifiée par ϕ_{hyp} que nous détaillons ici. L'absence de propriétés de convergence pourra par ailleurs être partiellement compensée : nous proposons au § 3.5.3 un test simple, valable pour des variables complexes, qui permet de vérifier *a posteriori* l'optimalité de la solution obtenue. Ainsi, toute convergence éventuelle de l'algorithme vers un point fixe autre qu'un des minima (globaux) du critère J_1 peut être décelée. Notons d'ailleurs que la solution obtenue par un tel algorithme a en pratique toujours satisfait cette condition, ce qui renforce l'intérêt de cette approche.

Nous avons vu au § 3.4.1 que l'algorithme IRLS repose sur la résolution de l'équation implicite (EI) suivante, traduisant l'annulation du gradient :

$$\text{EI(a):}\ \left(\mathbf{W}^\dagger\mathbf{W} + \lambda\mathbf{Q}_{\boldsymbol{x}}^{-1}\right)\boldsymbol{x} = \mathbf{W}^\dagger\boldsymbol{y} \quad\Leftrightarrow\quad \text{EI(b):}\ \boldsymbol{x} = \mathbf{Q}_{\boldsymbol{x}}\,\mathbf{W}^\dagger\Big(\lambda\mathbf{I}_N + \mathbf{W}\,\mathbf{Q}_{\boldsymbol{x}}\,\mathbf{W}^\dagger\Big)^{-1}\boldsymbol{y}$$

où $\mathbf{Q}_{\boldsymbol{x}} = \text{diag}\big(q_k(x_k)\big)$ avec $q_k(x_k) = \dfrac{x_k}{\phi'(x_k)}$, et la seconde formulation est celle proposée par Sacchi *et al.* [1998], permettant de réduire la dimension du système à inverser. Dans le cas de la pénalisation ℓ^1, on a $\phi_1'(x_k) = x_k/|x_k|$ et donc $q_{k_1}(x_k) = |x_k|$. Par conséquent, la matrice $\mathbf{Q}_{\boldsymbol{x}}^{-1}$ n'existe que si aucune composante de $\boldsymbol{x}$ ne s'annule (remarquons que ce problème n'apparaît pas avec la fonction ϕ_{hyp} pour laquelle les $q_k(x_k)$ ne s'annulent jamais). C'est pourquoi la mise en œuvre dans [Davies et Daudet, 2003, Fuchs, 2005] s'appuie sur la résolution de l'équation implicite EI(b), qui est définie pour tout vecteur complexe $\boldsymbol{x}$. Notons toutefois qu'elle ne correspond pas exactement à l'annulation du gradient, celui-ci n'étant pas défini lorsque $x_k = 0$, et que cette contrainte limitant l'analogie avec l'algorithme IRLS initial est exactement la même que celle qui invalide l'application des propriétés de convergence précitées, à savoir l'annulation de $q_{k_1}(x_k) = |x_k|$ lorsque x tend vers 0.

Malgré des lacunes au niveau des propriétés de convergence, que nous pallierons en partie avec le test du § 3.5.3, la résolution de l'équation implicite EI(b) par un algorithme IRLS présente un grand intérêt par la possibilité de réduire la dimension des matrices à inverser dans l'étape de mise à jour :

$$\boldsymbol{x}^{(t+1)} = \mathbf{Q}^{(t)}\,\mathbf{W}^\dagger\Big(\lambda\mathbf{I}_N + \mathbf{W}\,\mathbf{Q}^{(t)}\,\mathbf{W}^\dagger\Big)^{-1}\boldsymbol{y}.$$

En effet, si $\mathbf{Q}^{(t)}$ a ses éléments diagonaux $q_k^{(t)}$ nuls pour un certain nombre d'indices $\mathcal{I}$, alors $\boldsymbol{x}^{(t+1)}$ sera nul pour ces mêmes indices. Le calcul de $\boldsymbol{x}^{(t+1)}$ pour ces indices est de ce fait inutile, ce qui

14. Rappelons que les hypothèses proposées par les deux derniers papiers assurent par contre la convergence de l'IRLS pour la pénalisation hyperbolique.

réduit la dimension du problème. Un cas particulièrement intéressant est celui d'une fonction de pénalisation telle que l'on ait également : $x_k^{(t)} = 0 \Rightarrow q_k^{(t)} = 0$ puisqu'alors la nullité des composantes d'indices $\mathcal{I}$ se propage au fil des itérations : si $x_k^{(t)} = 0$, alors $x_k^{(u)} = 0$ pour $u > t$. Remarquons que cette condition s'écrit encore :

$$\frac{x}{\phi'(x)} \to 0 \text{ pour } x \to 0,$$

qui est contradictoire avec les hypothèses de convergence de l'algorithme IRLS proposées par Yarlagadda *et al.* [1985] et Idier [2001b]. Cette condition est satisfaite pour $\phi(x) = |x|^p$ avec $p < 2$, et donc en particulier pour la pénalisation ℓ^1 ; elle n'est par contre pas vérifiée pour la branche d'hyperbole ϕ_{hyp} précédemment utilisée comme approximation strictement convexe et dérivable de la norme ℓ^1.

Soient $\mathcal{I}^{(t)}$ les indices des éléments nuls de $\boldsymbol{x}^{(t)}$ et $-\mathcal{I}^{(t)}$ les indices des éléments non nuls. En notant $\mathbf{W}_{-\mathcal{I}}$ (respectivement $\mathbf{Q}_{-\mathcal{I}}$) la matrice $\mathbf{W}$ (respectivement $\mathbf{Q}$) dans laquelle on a supprimé les colonnes (respectivement colonnes et lignes) correspondant aux indices $\mathcal{I}^{(t)}$, l'étape de mise à jour de l'algorithme IRLS peut par conséquent être réalisée sous la forme :

$$\boldsymbol{x}_{-\mathcal{I}}^{(t+1)} = \left(\mathbf{W}_{-\mathcal{I}}{}^{\dagger}\mathbf{W}_{-\mathcal{I}} + \lambda\mathbf{Q}_{-\mathcal{I}}^{-1}\right)^{-1} \mathbf{W}_{-\mathcal{I}}^{\dagger}\,\boldsymbol{y}$$

qui correspond exactement à l'écriture initiale, de la forme EI(a), de l'étape de mise à jour dans l'IRLS en ne considérant que les éléments pour lesquels $\boldsymbol{x}^{(t)}$ est non nul : cette fois, la matrice $\mathbf{Q}_{-\mathcal{I}}^{-1}$ est correctement définie. La mise à jour pour les indices $\mathcal{I}$ est bien évidemment $\boldsymbol{x}_{\mathcal{I}}^{(t+1)} = \mathbf{0}$. Notons qu'une technique similaire est employée par Figueiredo [2003] pour la minimisation d'un critère de type (3.4) par un algorithme EM (*Expectation-Maximization*), pour lequel l'étape de maximisation s'avère être identique à l'étape de mise à jour de l'IRLS.

Une mise en œuvre efficace de l'algorithme IRLS s'effectue finalement en deux étapes distinctes suivant que $N_{-\mathcal{I}}^{(t)}$, le nombre d'éléments non nuls de $\boldsymbol{x}^{(t)}$ à l'itération t, est inférieur ou supérieur au nombre de données N :

1. Tant que $N_{-\mathcal{I}}^{(t)} > N$: **résolution d'un système de N équations à N inconnues** par l'alternance des deux étapes :
   i) $\mathbf{Q}^{(t)} = \operatorname{diag}\left(\{|x_k^{(t)}|\}_k\right)$,
   ii) $\boldsymbol{x}^{(t+1)} = \mathbf{Q}^{(t)}\,\mathbf{W}\left(\lambda\mathbf{I} + \mathbf{W}\,\mathbf{Q}^{(t)}\,\mathbf{W}^{\dagger}\right)^{-1}\boldsymbol{y}$.
2. Dès que $N_{-\mathcal{I}}^{(t)} \leq N$: **résolution d'un système de $N_{-\mathcal{I}}^{(t)}$ équations à $N_{-\mathcal{I}}^{(t)}$ inconnues** par l'alternance des deux étapes :
   i) $\mathbf{Q}_{-\mathcal{I}} = \operatorname{diag}(\{|x_k^{(t)}|\}_{k\notin\mathcal{I}})$;
   ii) $\boldsymbol{x}_{\mathcal{I}}^{(t+1)} = \mathbf{0}$ et $\boldsymbol{x}_{-\mathcal{I}}^{(t+1)} = \left(\mathbf{W}_{-\mathcal{I}}{}^{\dagger}\mathbf{W}_{-\mathcal{I}} + \lambda\mathbf{Q}_{-\mathcal{I}}^{-1}\right)^{-1}\mathbf{W}_{-\mathcal{I}}{}^{\dagger}\,\boldsymbol{y}$.

   Notons que le produit $\mathbf{W}_{-\mathcal{I}}{}^{\dagger}\boldsymbol{y}$ ne nécessite pas de calcul (ce sont des éléments de $\mathbf{W}^{\dagger}\,\boldsymbol{y}$, calculable une fois pour toutes) et que la matrice $\mathbf{W}_{-\mathcal{I}}{}^{\dagger}\mathbf{W}_{-\mathcal{I}}$ n'est calculée qu'à chaque changement du nombre d'éléments non nuls de $\boldsymbol{x}^{(t+1)}$.

Puisque la pénalisation ℓ^1 est théoriquement adaptée à l'estimation de solutions parcimonieuses (voir le § 3.2.2), la décroissance au fil des itérations du nombre $N_{-\mathcal{I}}^{(t)}$ de composantes non nulles est assurée et chaque itération aura un coût d'autant plus faible que ce nombre est réduit à partir du moment où $N_{-\mathcal{I}}^{(t)} \leq N$: les itérations seront donc de plus en plus rapides. Notons que,

en pratique, il convient d'initialiser l'algorithme avec un spectre $\boldsymbol{x}^{(0)}$ non nul en tout point afin de ne pas introduire artificiellement de « faux zéros » [Fuchs, 2005]. Par ailleurs, les étapes de mise à jour 1.ii) et 2.ii) débouchent sur des valeurs numériques des amplitudes qui ne sont jamais rigoureusement nulles ; afin de pouvoir bénéficier de la propriété précédente de réduction de dimension, il sera donc nécessaire d'effectuer un *seuillage* des amplitudes considérées comme nulles à chaque itération.

3.5.3 Un test d'optimalité

L'utilisation de l'algorithme IRLS avec une pénalisation ℓ^1 est donc très intéressante en termes de coût de calcul. Notons encore une fois que, si une preuve de convergence de cette procédure semble être établie dans le cas de paramètres réels, ce n'est pour l'instant pas le cas de l'extension à des variables complexes. Néanmoins, nous pouvons définir une condition nécessaire et suffisante (CNS) remplie par tout minimiseur du critère J_1, qui permet de vérifier postérieurement que la solution obtenue par l'algorithme ci-dessus minimise effectivement le critère (3.4). En effet, dans le cas de variables réelles, il est montré dans [Alliney et Ruzinsky, 1994] que :

$$\boldsymbol{x} \text{ est un minimum de } J_1 \text{ si et seulement si } \begin{cases} (i) & \forall k \text{ tq. } |x_k| = 0 : \quad |r_k| \leq \lambda, \\ (ii) & \forall k \text{ tq. } |x_k| \neq 0 : \quad r_k + \lambda \operatorname{sign}(x_k) = 0, \end{cases}$$

où $\boldsymbol{r} = \mathbf{W}^{\mathrm{t}}(\mathbf{W}\boldsymbol{x} - \boldsymbol{y})$. Nous avons généralisé ce résultat au cas de variables complexes :

$$\boldsymbol{x} \text{ est un minimum de } J_1 \text{ si et seulement si } \begin{cases} (i) & \forall k \text{ tq. } |x_k| = 0 : \quad |r_k| \leq \lambda, \\ (ii) & \forall k \text{ tq. } |x_k| \neq 0 : \quad r_k + \lambda \dfrac{x_k}{|x_k|} = 0, \end{cases}$$

où $\boldsymbol{r} = \mathbf{W}^{\dagger}(\mathbf{W}\boldsymbol{x} - \boldsymbol{y})$. D'un point de vue pratique, il est donc immédiat de vérifier *a posteriori* si la solution obtenue en sortie de l'IRLS vérifie ces conditions, c'est-à-dire est bien un minimum du critère J_1. La démonstration établissant ce test est renvoyée en annexe A.2). Notons que, au vu d'une telle CNS, l'information de parcimonie du minimiseur du critère J_1 apparaît clairement par la séparation des conditions (i) et (ii) selon que chaque composante x_k est nulle ou non.

Exemple sur un signal simple

L'illustration pratique de ce test est représentée sur la figure 3.7 pour le signal habituel composé de cinq raies spectrales [15]. Conformément aux considérations du paragraphe précédent, nous avons dû fixer un seuil en-dessous duquel toute valeur numérique est considérée comme nulle par l'algorithme IRLS ; ce seuil a été fixé ici à 10^{-10}. La solution $\widehat{\boldsymbol{x}}_1$ obtenue par l'algorithme IRLS est rigoureusement parcimonieuse, dans le sens où ses seules composantes non nulles en module sont localisées autour des fréquences théoriques (figures du haut). La figure en bas à gauche représente la partie (i) du test, où tous les résidus en module $|r_k|$ sont bien inférieurs au paramètre de régularisation λ pour l'ensemble des composantes nulles de $\boldsymbol{x}$. La partie (ii) du test est représentée sur la figure en bas à droite : les valeurs de $r_k + \lambda \frac{x_k}{|x_k|}$ ne sont pas rigoureusement nulles mais inférieures à 10^{-10}, c'est-à-dire au seuil de nullité fixé plus haut. De ce fait, nous pouvons conclure que la solution $\widehat{\boldsymbol{x}}_1$ est bien un minimiseur du critère J_1.

15. Nous ne représentons sur cette figure que la partie de $\boldsymbol{r}$ correspondant aux fréquences positives. En effet, si la solution $\widehat{\boldsymbol{x}}_1$ est à symétrie hermitienne – nous verrons au § 3.6.1 qu'en pratique $\widehat{\boldsymbol{x}}_1$ l'est toujours – alors il est facile de voir que $\boldsymbol{r} = \mathbf{W}^{\dagger}(\mathbf{W}\boldsymbol{x} - \boldsymbol{y})$ l'est également lorsque les données $\boldsymbol{y}$ sont à valeurs réelles.

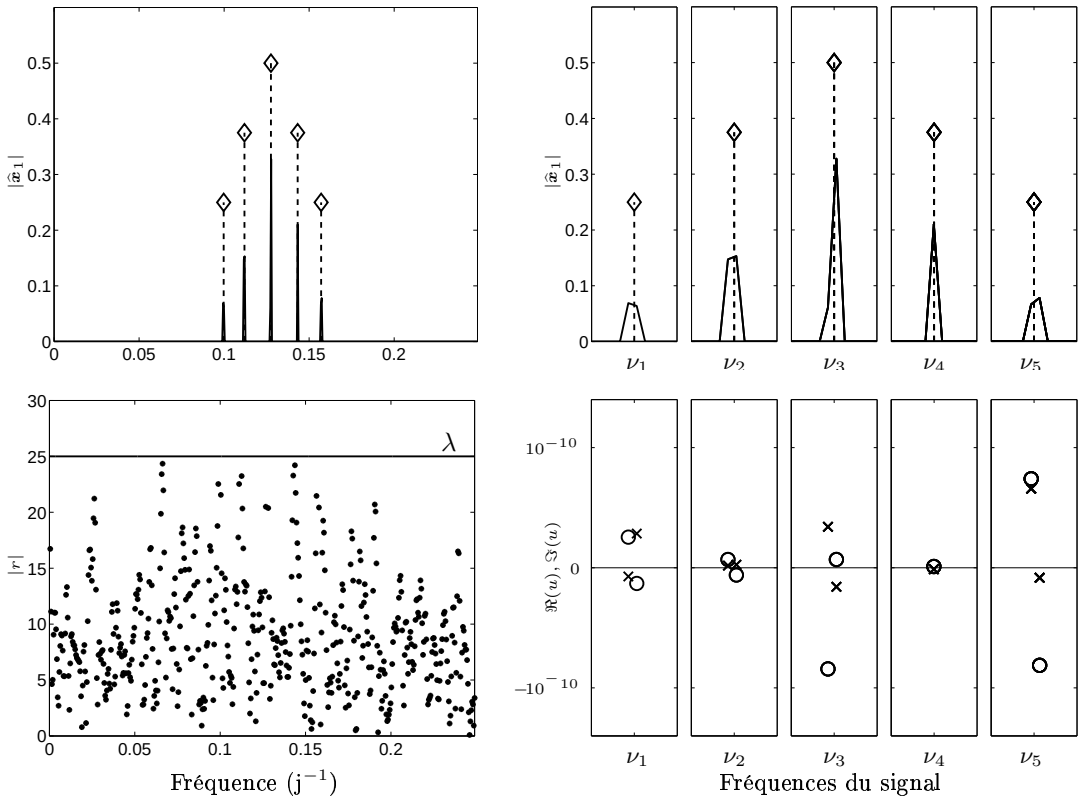


FIG. 3.7 – *Illustration du test d'optimalité d'un minimiseur du critère J_1. En haut, solution obtenue par l'algorithme IRLS. Les traits pleins représentent le module du spectre estimé, et les $\diamond$ les raies théoriques. En bas, représentation du test : à gauche sont représentés $|r_k|$ pour les composantes nulles de $\boldsymbol{x}$ ainsi que le seuil de valeur λ ; à droite, représentation des parties réelle ($\circ$) et imaginaire ($\times$) de $u_k = r_k + \lambda x_k/|x_k|$ pour les composantes non nulles de $\boldsymbol{x}$.*

3.5.4 Algorithme de type Gauss-Seidel

Un algorithme de descente de type Gauss-Seidel permet une mise en œuvre très efficace pour une pénalisation ℓ^1 car chaque minimisation scalaire de J_1 en la variable complexe x_k a une solution analytique explicite et immédiate. En notant $\boldsymbol{w}_k$ la colonne d'indice k de la matrice $\mathbf{W}$ et $\boldsymbol{e}_k = \boldsymbol{y} - \sum_{\ell \neq k} \boldsymbol{w}_\ell x_\ell$, nous avons en effet montré que :

$$\begin{aligned} \mathring{x}_k = \arg\min_{x_k} J_1(\boldsymbol{x}) \quad &\Leftrightarrow \quad \mathring{x}_k = \arg\min_{x_k} \frac{1}{2} \|\boldsymbol{e}_k - \boldsymbol{w}_k x_k\|^2 + \lambda |x_k| \\ &\Leftrightarrow \quad \begin{cases} \text{si } |\boldsymbol{w}_k^\dagger \boldsymbol{e}_k| \leq \lambda \ : \ \mathring{x}_k = 0, \\ \text{si } |\boldsymbol{w}_k^\dagger \boldsymbol{e}_k| > \lambda \ : \ \arg \mathring{x}_k = \arg \boldsymbol{w}_k^\dagger \boldsymbol{e}_k \text{ et } |\mathring{x}_k| = \dfrac{1}{N} \left(|\boldsymbol{w}_k^\dagger \boldsymbol{e}_k| - \lambda \right). \end{cases} \end{aligned} \tag{3.5}$$

La démonstration correspondante est établie à l'annexe A.3. Ce résultat est très intéressant puisque les seules opérations nécessaires pour chaque minimisation scalaire sont l'actualisation de $\boldsymbol{e}_k$ et le calcul de $\boldsymbol{w}_k^\dagger \boldsymbol{e}_k$. Chaque minimisation est donc très simple, notamment par rapport au

cas de la pénalisation hyperbolique[16] où la minimisation, également explicite, nécessite le calcul des racines d'un polynôme de degré quatre. L'algorithme GS_1 (algorithme de Gauss-Seidel pour une pénalisation ℓ^1) consiste alors en le schéma itératif suivant :

1. Itération t : pour $k = -P \dots P$, réaliser les opérations :
   i) calcul de $\boldsymbol{e}_k^{(t)} = \boldsymbol{y} - \sum_{\ell<k} \boldsymbol{w}_\ell \, x_\ell^{(t)} - \sum_{\ell>k} \boldsymbol{w}_\ell \, x_\ell^{(t-1)}$;
   ii) si $|\boldsymbol{w}_k^\dagger \boldsymbol{e}_k| < \lambda$: prendre $x_k^{(t)} = 0$; sinon prendre $x_k^{(t)} = \dfrac{1}{N}\Big(|\boldsymbol{w}_k^\dagger \boldsymbol{e}_k| - \lambda\Big)$;
2. $t \leftarrow t+1$ et retour en 1 jusqu'à convergence.

Ce schéma peut être rapproché de l'algorithme proposé pour des variables réelles par Alliney et Ruzinsky [1994] qui, plutôt que d'effectuer la *minimisation* scalaire de J_1 en x_k, réalisent la mise à jour de chaque composante $x_k^{(t)}$ de manière sous-optimale : $x_k^{(t+1)} = x_k^{(t)} + \beta_k$, où β_k est déterminé explicitement de façon à diminuer la valeur du critère à chaque itération. Les auteurs prouvent alors la convergence de l'algorithme résultant vers un minimum, nécessairement global, de J_1. Par conséquent, la procédure décrite ci-dessus, réalisant à chaque itération la plus grande diminution du critère, converge également. Remarquons au passage que la démarche employée par ces auteurs semble passer à côté de la nature *explicite* de la minimisation scalaire (3.5), dont l'expression est encore plus simple dans le cas considéré de variables réelles ; la sous-optimalité de chaque mise à jour de $x_k^{(t)}$ risque alors d'augmenter inutilement le nombre d'itérations nécessaires à la convergence de l'algorithme. L'introduction d'un facteur de *relaxation* à l'étape de mise à jour 1.ii) est une démarche qui, tout en s'écartant d'une minimisation exacte, vise à accélérer la convergence de l'algorithme ; nous y reviendrons au § 3.7.

3.6 Prise en compte de la symétrie hermitienne

3.6.1 Des estimateurs « souvent » identiques

Les séries temporelles étudiées étant à valeurs *réelles*, le spectre $\boldsymbol{x} = [x_{-K}, \dots, x_0, \dots, x_K]$ que l'on cherche à reconstruire possède bien entendu la propriété de symétrie hermitienne : $x_{-k} = x_k^*$. Cependant, nous avons vu au § 3.2.1 que le terme de moindres carrés $\|\boldsymbol{y} - \mathbf{W}\boldsymbol{x}\|^2$ correspond, d'un point de vue statistique, à l'anti log-vraisemblance du modèle $\boldsymbol{y} = \mathbf{W}\boldsymbol{x} + \epsilon$ en considérant des perturbations $\boldsymbol{\epsilon}$ *i.i.d.* gaussiennes complexes *circulaires*, c'est-à-dire :

$$p(\boldsymbol{\epsilon}) \propto \frac{1}{(2\pi\sigma_\epsilon^2)^N} \exp\left(-\frac{\sum_{n=1}^N |\epsilon_n|^2}{2\sigma_\epsilon^2}\right)$$

En d'autres termes, on n'impose pas aux perturbations $\boldsymbol{\epsilon}$ d'être à valeurs réelles ou, ce qui revient au même, au spectre $\boldsymbol{x}$ d'être à symétrie hermitienne. Plus précisément, cette formulation autorise la même erreur de reconstruction, de puissance σ_ϵ^2, sur les parties réelle et imaginaire de $\boldsymbol{y} - \mathbf{W}\boldsymbol{x}$, alors que les données *sont* à valeurs réelles. C'est d'ailleurs l'approche adoptée dans les travaux précités, qui utilisent cette même formulation pour des raisons de simplicité du modèle [Sacchi *et al.*, 1998, Giovannelli et Idier, 2001, Ciuciu *et al.*, 2001]. Les derniers

16. Ce gain en simplicité résulte d'une plus grande simplicité de l'expression analytique de la fonction φ', où la pénalisation de chaque amplitude s'écrit $\phi(x) = \varphi(|x|)$: dans le cas ℓ^1, on a $\varphi_1(u) = u$ et $\varphi_1'(u) = 1$ alors que pour la pénalisation hyperbolique, on a $\varphi_{\mathrm{hyp}}(u) = \sqrt{s^2+u^2}$ et $\varphi_{\mathrm{hyp}}'(u) = \frac{u}{\sqrt{s^2+u^2}}$.

auteurs montrent cependant que, dans le cas de données à valeurs réelles et d'une pénalisation strictement convexe, l'unique minimiseur du critère pénalisé est à symétrie hermitienne : c'est le cas de l'estimateur $\widehat{\boldsymbol{x}}_{\text{hyp}}$.

Dans le cas de la pénalisation par la norme ℓ^1, une telle propriété ne peut être établie à cause de la non stricte convexité du critère résultant. Toutefois, il est facile de voir que l'étape de mise à jour de l'algorithme IRLS propage la symétrie hermitienne : l'initialisation par une valeur à symétrie hermitienne assure ainsi la préservation de cette symétrie en sortie de l'algorithme. Cette propriété ne s'applique pas à un algorithme de type Gauss-Seidel, qui pourtant est assuré de converger vers un minimum (global) du critère pénalisé associé ; en pratique, cependant, les solutions obtenues ont toujours possédé la propriété de symétrie hermitienne.

3.6.2 Un modèle à deux fois moins d'inconnues

Si elle ne modifie pas l'estimateur obtenu, la prise en compte d'une contrainte de symétrie hermitienne sur le spectre à reconstruire permet de réduire de moitié le nombre d'inconnues du problème. En ne conservant que les paramètres de fréquences positives, nous pouvons considérer le modèle :

$$\begin{aligned} \boldsymbol{y} &= 2\,\Re(\mathbf{W}_+\boldsymbol{x}_+) + \boldsymbol{\epsilon} \\ &= \mathbf{W}_+\boldsymbol{x}_+ + \mathbf{W}_+^*\boldsymbol{x}_+^* + \boldsymbol{\epsilon} \end{aligned} \tag{3.6}$$

avec cette fois $\boldsymbol{x}_+ = [x_0, \ldots, x_P]^{\text{t}}$, $\mathbf{W}_+ = \{e^{j2\pi\frac{k}{P}f_{\max}t_n}\}_{k=0\ldots P,\, n=1\ldots N}$ opérant sur les fréquences positives et $\mathbf{W}_+^* = \{e^{j2\pi\frac{k}{P}f_{\max}t_n}\}_{k=-P\ldots 0,\, n=1\ldots N}$ sur les fréquences négatives. Nous introduisons alors le critère :

$$\begin{aligned} J_+(\boldsymbol{x}_+) &= \frac{1}{2}\left\|\boldsymbol{y} - 2\Re(\mathbf{W}_+\boldsymbol{x}_+)\right\|^2 + \lambda R(\boldsymbol{x}_+) \\ &= \frac{1}{2}\left\|\boldsymbol{y} - (\mathbf{W}_+\boldsymbol{x}_+ + \mathbf{W}_+^*\boldsymbol{x}_+^*)\right\|^2 + \lambda R(\boldsymbol{x}_+) \end{aligned} \tag{3.7}$$

dont les seuls paramètres sont les amplitudes $\boldsymbol{x}_+$ correspondant aux fréquences positives $\{\frac{k}{P}f_{\max}\}_{k=0\ldots P}$. Ainsi, la partie quadratique du critère J_+ correspond à l'anti log-vraisemblance du modèle (3.6) pour des perturbations $\boldsymbol{\epsilon}$ gaussiennes qui sont maintenant à valeurs *réelles*, et le critère à minimiser possède deux fois moins d'inconnues. Lorsque le gradient associé est défini – c'est le cas pour la pénalisation hyperbolique – celui-ci s'écrit alors (le principe du calcul correspondant est explicité à l'annexe A.1) :

$$\begin{aligned} \overrightarrow{\nabla J}_+(\boldsymbol{x}_+) &= -\mathbf{W}_+^\dagger\big(\boldsymbol{y} - 2\Re(\mathbf{W}_+\boldsymbol{x}_+)\big) + \lambda\overrightarrow{\nabla R}(\boldsymbol{x}) \\ &= -\mathbf{W}_+^\dagger\big(\boldsymbol{y} - (\mathbf{W}_+\boldsymbol{x}_+ + \mathbf{W}_+^*\boldsymbol{x}_+^*)\big) + \lambda\overrightarrow{\nabla R}(\boldsymbol{x}). \end{aligned}$$

Notons que pour un spectre $\boldsymbol{x}$ à symétrie hermitienne, le terme de pénalisation $R(\boldsymbol{x}_+)$ du critère (3.7) aura une valeur deux fois plus faible que celle de la pénalisation $R(\boldsymbol{x})$ du critère initial (3.2) : $\phi(x_k) = \phi(x_k^*)$ donc $\phi(x_k) = \frac{1}{2}\big(\phi(x_k) + \phi(x_{-k})\big)$. Aussi, il convient d'adapter correctement le paramètre de régularisation en le choisissant deux fois plus grand pour $J(\boldsymbol{x}_+)$ que pour $J(\boldsymbol{x})$.

3.6.3 Intérêt pour les algorithmes de gradients conjugués

Nous étudions ici l'apport de la réduction du nombre d'inconnues au niveau des différentes mises en œuvre proposées aux § 3.4 et 3.5. Dans le cas de l'algorithme IRLS, cette réduction n'est pas intéressante. En effet, cet algorithme est basé sur l'écriture de l'annulation du gradient sous la forme d'une équation implicite du type $\mathbf{A}_{\boldsymbol{u}}\boldsymbol{u} = \boldsymbol{b}$. Dans le cas du gradient $\overrightarrow{\nabla J}_+$, celle-ci s'écrit $\mathbf{A}^1_{\boldsymbol{x}}\boldsymbol{x}_+ + \mathbf{A}^2_{\boldsymbol{x}}\boldsymbol{x}^*_+ = \boldsymbol{b}$ et nécessite donc un reparamétrage en $\boldsymbol{u} = [\boldsymbol{x}_+,\boldsymbol{x}^*_+]^{\mathrm{t}}$, c'est-à-dire un retour au vecteur $\boldsymbol{x}$ à l'ordonnancement près. La réduction du nombre d'inconnues n'entraîne donc pas de réduction de la complexité calculatoire. Pour les algorithmes de type Gauss-Seidel, nous n'avons pas réussi non plus à prendre en compte de manière simple la symétrie hermitienne.

Par contre, la réduction à un nombre moitié d'inconnues est intéressante en termes de coûts calculatoires pour la mise en œuvre des algorithmes de gradients conjugués, étudiés précédemment dans le cas de la pénalisation hyperbolique, qui exploitent la structure des opérateurs mis en jeu. De manière analogue aux propriétés établies au § 3.1, il est facile de voir que, quelque soit le type d'échantillonnage :

- $\mathbf{T}_+ \triangleq \mathbf{W}_+^\dagger \mathbf{W}_+ = \{\sum_{n=1}^N e^{-j2\pi\frac{k-\ell}{P}f_{\max}t_n}\}_{k,\ell=0\ldots P}$ est une matrice de Toeplitz.
- $\mathbf{H}_+ \triangleq \mathbf{W}_+^\dagger \mathbf{W}_+^* = \{\sum_{n=1}^N e^{-j2\pi\frac{k+\ell}{P}f_{\max}t_n}\}_{k,\ell=0\ldots P}$ est une matrice de Hankel, *i.e.* constante sur les antidiagonales.

Les produits $\mathbf{W}_+^\dagger \mathbf{W}_+ \cdot$ et $\mathbf{W}_+^\dagger \mathbf{W}_+^* \cdot$ peuvent donc se calculer par des algorithmes FFT sur $2P+2$ points, permettant ici encore le calcul du critère J_+ ainsi que celui du gradient $\overrightarrow{\nabla J}_+$ à moindre coût. Notons que par rapport au critère initial à $2P+1$ inconnues, pour lequel critère et gradient se calculent par un double algorithme FFT sur $4P+2$ points, la réduction du nombre d'inconnues permet cette fois de calculer critère et gradient par deux doubles algorithmes FFT sur $2P+2$ points.

Le tableau 3.2 résume les coûts de calcul correspondant, pour un échantillonnage irrégulier, aux différents critères et gradients introduits pour une approche algorithmique de type Gradients Conjugués. Le calcul d'un produit de type $\boldsymbol{v} = \mathbf{A}\boldsymbol{u}$, où $\mathbf{A}$ est une matrice circulante de taille $L \times L$ et $\boldsymbol{u}$ un vecteur de L composantes, s'effectue de la façon suivante :

i) calcul de la TFD $\widehat{\boldsymbol{u}}$ de $\boldsymbol{u}$: $\frac{L}{2}\log_2 L$ multiplications complexes ,

ii) calcul du produit $\widehat{\boldsymbol{v}} = \widehat{\boldsymbol{u}} \times \widehat{\boldsymbol{a}}$: L multiplications complexes ,

iii) calcul de la TFD inverse $\boldsymbol{v}$ de $\widehat{\boldsymbol{v}}$: $\frac{L}{2}\log_2 L$ multiplications complexes ,

où $\widehat{\boldsymbol{a}}$ est la TFD, sur P points, de la première ligne de $\mathbf{A}$ et, afin de tirer parti des algorithmes FFT, L est supposé être une puissance de 2. Ce calcul requiert donc un nombre de multiplications de l'ordre de $L\log_2 L$.

Nous constatons que la prise en compte d'un modèle hermitien permet donc de réduire de moitié le coût de calcul du critère initial J, mais qu'elle réduit plus légèrement le coût associé à son gradient : la différence n'est en effet que de $4P\log_2(4P) - 4P\log_2(2P) = 4P$ multiplications complexes. De même, les coûts de calcul du critère et du gradient modifiés sont réduits d'environ $4P$ et $8P$ multiplications complexes, respectivement. Les simulations du § 3.7 montreront cependant un gain de temps de calcul non négligeable entre les versions non hermitienne et hermitienne des algorithmes de Gradients Conjugués.

	quantité à calculer	coût associé	exemple $N = 250$, $P = 500$
Critère initial J	$\mathbf{W}\boldsymbol{x}$	$2NP$	2.5×10^5
Gradient $\overrightarrow{\nabla J}$	$\mathbf{T}\boldsymbol{x}$	$4P\log_2(4P)$	2.2×10^4
Critère initial J_+	$\mathbf{W}_+\boldsymbol{x}_+$	NP	1.2×10^5
Gradient $\overrightarrow{\nabla J}_+$	$\mathbf{T}_+\boldsymbol{x}_+ + \mathbf{H}_+\boldsymbol{x}_+^*$	$4P\log_2(2P)$	2×10^4

TAB. 3.2 – *Comparaison des coûts de calcul associés aux différents critères et gradients pour un échantillonnage irrégulier dans le cas de modèles non hermitien et hermitien. La matrice* $\mathbf{T} \triangleq \mathbf{W}^\dagger\mathbf{W}$ *est ici une matrice de Toeplitz et les matrices* $\mathbf{T}_+ \triangleq \mathbf{W}_+^\dagger\mathbf{W}_+$ *et* $\mathbf{H}_+ \triangleq \mathbf{W}_+^\dagger\mathbf{W}_+^*$ *sont respectivement des matrices de Toeplitz et de Hankel.*

3.7 Performances algorithmiques

Au long de ce chapitre, nous avons illustré sur un signal relativement simple quelques caractéristiques générales concernant l'estimation de spectres de raies par minimisation d'un critère pénalisé. Des comportements similaires en termes de parcimonie ont été mis en évidence pour les estimateurs $\widehat{\boldsymbol{x}}_{\text{hyp}}$, $\widehat{\boldsymbol{x}}_1$. La recherche de solutions algorithmiques efficaces constitue notre principale contribution dans ce chapitre ; nous évaluons donc à présent les performances en termes de vitesse d'exécution des différents algorithmes introduits aux paragraphes 3.4 et 3.5. Toutes les optimisations sont effectuées sur le signal habituel de la figure 2.6, en utilisant un paramètre de régularisation de $\lambda = 25$, soit $\log_{10}\lambda \simeq 1.4$: nous avons vu au § 3.3.2 que cette valeur est satisfaisante pour l'obtention d'une solution parcimonieuse.

3.7.1 Pénalisation hyperbolique

Nous avons envisagé au § 3.4 plusieurs options algorithmiques pour la minimisation du critère pénalisé par la fonction $R_{\text{hyp}}(\boldsymbol{x}) = \sum_k \sqrt{s^2 + |x_k|^2}$, dont nous résumons les principales caractéristiques pour le problème d'analyse spectrale des signaux à échantillonnage irrégulier :

- un algorithme de Gradients Conjugués (GC), où la structure de Toeplitz de la matrice $\mathbf{W}^\dagger\mathbf{W}$ est exploitée pour le calcul du gradient au moyen d'algorithmes FFT. Dans la version mise en œuvre, la direction de descente est calculée selon la formule de Polak et Ribière [Press *et al.*, 1992] et la minimisation scalaire déterminant le pas de descente est effectuée par approximations paraboliques successives. La version prenant en compte la symétrie hermitienne est également proposée, suivant les considérations du § 3.6, et sera notée GC_+ ;
- un algorithme de Gradients Conjugués sans recherche exacte du pas de descente (GCSRE) s'affranchissant du calcul du critère et basé uniquement sur des calculs du gradient ;
- un algorithme IRLS (*Iterative Reweighted Least-Squares*), qui a donné des résultats satisfaisants pour l'estimation de raies spectrales en échantillonnage régulier [Ciuciu *et al.*, 2001], mais qui n'exploite aucune structure particulière des opérateurs dans le cas traité ici ;
- un algorithme de descente de type Gauss-Seidel (GS), où chaque minimisation scalaire est réalisée par un algorithme de type IRLS scalaire. Nous avons vu au § 3.4.3 que cette minimisation peut être réalisée explicitement, mais en pratique le coût calculatoire associé est plus élevé. Nous envisageons également une version sur-relaxée notée GSOR (OR

pour *Over-Relaxation*) par l'introduction d'un coefficient de relaxation ω. En utilisant les notations du § 3.4.1, chaque mise à jour scalaire est effectuée par :

$$\begin{aligned} x_k^{(t+1)} &= x'_k \quad \text{si } f(x'_k) < f(x_k^{(t)}), \text{ avec } x'_k = \omega\, \mathring{x}_k + (1-\omega) x_k^{(t)}, \\ &= x_k^{(t)} \quad \text{sinon,} \end{aligned}$$

où $\mathring{x}_k$ est la valeur (scalaire) minimisant la fonction $f(x_k)$ et $\omega \in]1,2[$, qui permet d'accélérer la convergence de l'algorithme GS [Press *et al.*, 1992]. Un réglage de $\omega = 1.5$ a empiriquement fourni la convergence la plus rapide ;

La valeur du paramètre s a été fixée à 10^{-3} et tous les algorithmes utilisent la même condition d'arrêt :

$$\left\| \boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)} \right\| < 10^{-6} \text{ et } \left\| \overrightarrow{\nabla J}_{\text{hyp}}{}^{(t)} \right\| < 10^{-4}.$$

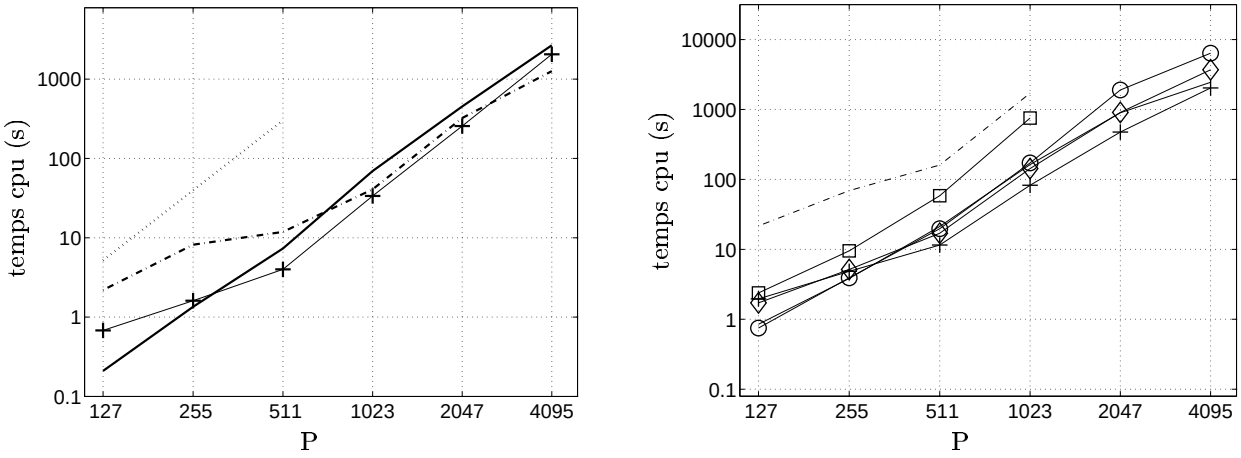
La mise en œuvre de l'ensemble de ces algorithmes a été réalisée sous *Matlab* et les coûts calculatoires présentés en figure 3.8 ont été obtenus avec un processeur cadencé à 1.4 GHz doté de 1 Go de mémoire vive. Les valeurs du paramètre P ont été choisies sous la forme $P = 2^K - 1, K \in \mathbb{N}$ de sorte que $2P+1$ soit une puissance de deux ; un tel choix favorise ainsi les différents algorithmes de type gradient qui utilisent des algorithmes FFT sur $2P+1$ et $4P+2$ points.

Les performances de ces différents algorithmes sont illustrées sur la figure 3.8, où les algorithmes IRLS et GCSRE montrent des performances relativement mauvaises [17]. Pour une grille de $P = 1023$ fréquences discrétisées, l'algorithme GSOR est le plus efficace (figure 3.8 en bas à gauche) ; les algorithmes GS et de type gradients conjugués (GC et GC_+) sont sensiblement plus longs mais restent ici plus performants que l'algorithme de Gauss-Seidel sans sur-relaxation. Pour un problème de dimension supérieure, les algorithmes de type Gauss-Seidel, c'est-à-dire GS et GSOR, deviennent largement plus rapides, comme nous pouvons le voir sur la figure 3.8 en bas à droite : l'introduction d'un facteur de sur-relaxation de $\omega = 1.5$ permet en particulier de réduire de manière significative (de moitié pour $P = 1023$, d'environ 20% pour P=4095) le coût calculatoire de ce type d'algorithme. Remarquons encore que la prise en compte de la symétrie hermitienne pour les algorithmes de gradients conjugués (GC_+) permet de gagner environ un facteur deux en vitesse d'exécution par rapport à la version GC.

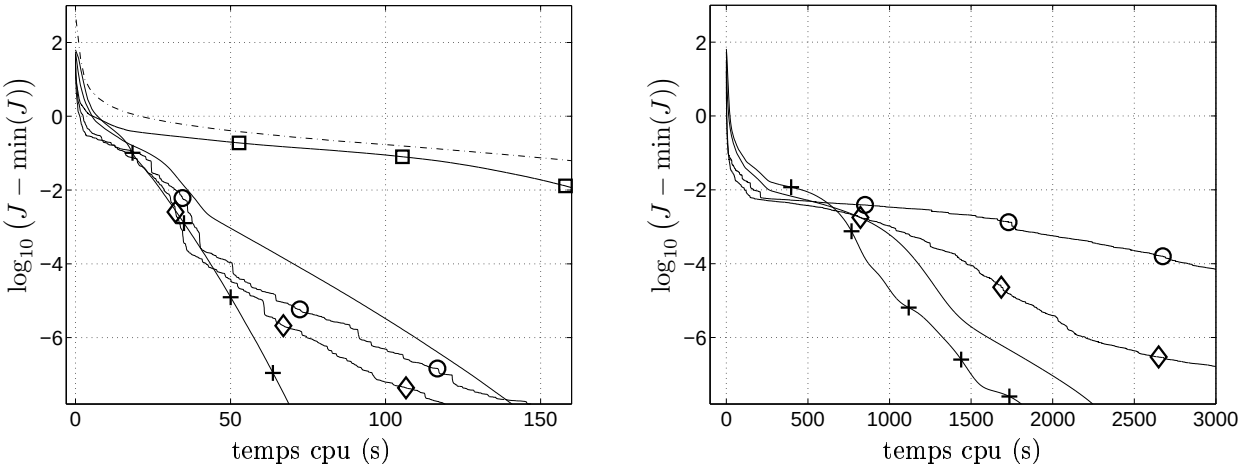
Parmi les algorithmes étudiés, l'algorithme GSOR est donc finalement le plus efficace pour la minimisation du critère à pénalisation hyperbolique, pour peu que la résolution fréquentielle requise soit suffisamment élevée.

Notons enfin que si nous avons ici suivi Ciuciu *et al.* [2001] dans l'utilisation de la branche d'hyperbole $\big(\phi(x_k) = \varphi(|x_k|) = \sqrt{s^2 + |x_k|^2}\big)$, le choix d'une autre fonction de type «$\ell^2\ell^1$» peut s'avérer plus judicieux pour une procédure d'optimisation de type Gauss-Seidel : nous montrons à l'annexe A.3 que la complexité analytique de chaque minimisation scalaire est déterminée par l'expression de la fonction φ'. Dans le cas de la branche d'hyperbole, la présence d'une racine carrée au dénominateur de φ' rend cette minimisation équivalente à la recherche des racines d'un polynôme de degré quatre, qui s'est avérée plus coûteuse qu'une minimisation itérative de type IRLS (voir le § 3.4.3). L'utilisation alternative, par exemple, de la fonction de Huber définie à ce même paragraphe et dont la dérivée est affine par morceaux, permet de réaliser chaque minimisation scalaire en résolvant deux équations du premier degré, ce qui est immédiat. Le coût

17. La version de GCSRE mise en œuvre ici ne prend pas en compte la symétrie hermitienne ; cependant la convergence plus lente de GCSRE par rapport à GC laisse penser qu'une version à symétrie hermitienne sera à son tour plus lente que GC_+.



Évolution du coût de l'optimisation en fonction de la dimension P.
À gauche, échelle de temps linéaire ; à droite, échelle de temps logarithmique.



Évolution des critères en fonction du temps pour $P = 1023$ (à gauche) et $P = 4095$ (à droite).

FIG. 3.8 – *Comparaison des performances des différents algorithmes pour une pénalisation hyperbolique : GC (∘), GC_+ (∘), GCSRE (⋄), IRLS (−·), GS (−) et GSOR (+).*

calculatoire de chaque itération de GS peut donc encore être réduit. Nous n'allons cependant pas plus loin dans cette voie car nous verrons au paragraphe suivant que le choix d'une pénalisation par la norme ℓ^1, pour laquelle le coût de ces minimisations scalaires est le plus faible $\big(\varphi'(\rho) = 1\big)$, surpasse celui d'une pénalisation $\ell^2\ell^1$ en termes de rapidité algorithmique.

3.7.2 Pénalisation ℓ^1

Nous comparons ici les algorithmes d'optimisation présentés au § 3.5 pour la minimisation du critère pénalisé par la norme ℓ^1, à savoir :

- un algorithme de programmation quadratique (PQ). Alors que les autres algorithmes ont été implantés sous *Matlab*, celui-ci fait appel aux programmes de la bibliothèque Mosek [18] écrits en langage C. Nous avons vu au § 3.5.1 que la mise en œuvre de cet algorithme, basée sur la réécriture du problème initial sous la forme d'un problème quadratique sous contrainte, prend en compte une pénalisation indépendante des parties réelle et imaginaire des amplitudes spectrales, qui est moins satisfaisante en termes d'estimation qu'une pénalisation opérant sur les modules complexes. Nous l'utilisons quand même ici afin de comparer son coût calculatoire aux autres algorithmes et discuter ainsi de leur éventuelle application pour des problèmes à variables réelles ;
- un algorithme IRLS dans sa version prenant en compte la réduction de dimension au fil des itérations ;
- un algorithme de descente de type Gauss-Seidel (GS), où chaque minimisation scalaire est obtenue de manière explicite et pour un coût très réduit. Comme dans le cas précédent, nous considérons également une version sur-relaxée (GSOR) avec un paramètre de relaxation $\omega = 1.5$.

Ces algorithmes n'exploitent aucune structure matricielle particulière et sont de ce fait applicables de la même manière pour tout type d'échantillonnage. Nous proposons d'utiliser le test d'optimalité introduit au § 3.5.3 comme critère d'arrêt, soit :

$$\boldsymbol{x}^{(t)} \text{ minimise } J_1 \Leftrightarrow \begin{cases} \forall k \text{ tq. } |x_k^{(t)}| = 0 : & |r_k^{(t)}| \leq \lambda & (i) \\ \forall k \text{ tq. } |x_k^{(t)}| \neq 0 : & r_k^{(t)} + \lambda \dfrac{x_k^{(t)}}{|x_k^{(t)}|} = 0 & (ii) \end{cases}$$

avec $\boldsymbol{r}^{(t)} = \mathbf{W}^\dagger(\mathbf{W}\boldsymbol{x}^{(t)} - \boldsymbol{y})$, excepté pour l'algorithme de programmation quadratique dont nous ne contrôlons pas le critère de convergence. La mise en œuvre de l'algorithme IRLS nécessitant la définition d'un seuil ν en-dessous duquel les amplitudes sont considérées comme numériquement nulles, les relations (i) et (ii) sont également considérées avec une tolérance numérique de ν, soit respectivement $|r_k^{(t)}| < \lambda + \nu$ et $\left| r_k^{(t)} + \lambda \dfrac{x_k^{(t)}}{|x_k^{(t)}|} \right| < \nu$. Nous notons en particulier que la relation (ii) correspond à l'annulation du gradient du critère J_1 en les points où celui-ci est défini, c'est-à-dire en les composantes non nulles du vecteur $\boldsymbol{x}^{(t)}$. Pour les valeurs de ν qui seront employées par la suite (soit $\nu = 10^{-6}$ et $\nu = 10^{-10}$), ce test d'arrêt est plus sévère que le critère d'arrêt sur le gradient employé au paragraphe précédent pour la pénalisation hyperbolique $\big(||\overrightarrow{\nabla J}_{\text{hyp}}|| < 10^{-4}\big)$, pour lequel, en pratique, le gradient présente encore des composantes supérieures à 10^{-5}. Signalons également que sur l'ensemble des optimisations réalisées l'algorithme IRLS a toujours

18. Logiciel disponible sur le site `http://www.mosek.com`.

convergé au sens de ce test malgré le manque, à ce jour, de propriétés établissant formellement sa convergence.

La figure 3.9 présente les coûts associés aux différents algorithmes en fonction de la taille du problème, pour deux valeurs du seuil $\nu = 10^{-6}$ et $\nu = 10^{-10}$. Hormis le fait qu'il ne débouche pas sur une solution complètement satisfaisante dans le cas de variables complexes, l'algorithme de programmation quadratique, bien qu'utilisant le langage C alors que les deux autres algorithmes sont mis en œuvre sous *Matlab*, se révèle bien plus coûteux ; nous ne lui faisons donc plus référence par la suite[19]. Les valeurs des coûts représentés sur la figure 3.9 montrent tout d'abord que la minimisation du critère pénalisé par la norme ℓ^1 peut être envisagée pour un coût plus faible que celle du critère J_{hyp}. Par exemple, pour une valeur de $P = 1023$, l'algorithme GSOR met ici environ 33 s contre 75 s pour la pénalisation hyperbolique, alors que le critère d'arrêt est plus sévère dans le premier cas. Nous notons que la version sur-relaxée permet de réduire approximativement de moitié le coût de l'algorithme GS (à partir de $P = 511$) et que GSOR est également plus efficace que l'IRLS jusqu'à une certaine taille de l'espace des paramètres, au-delà de laquelle cette relation s'inverse. Au vu de la figure 3.9, cette inversion des coûts calculatoires a lieu entre $P = 2047$ et $P = 4095$. Notons également que le choix d'un seuil ν plus sévère augmente « raisonnablement » le coût calculatoire associé : quel que soit l'algorithme, pour P de l'ordre du millier, la durée requise par l'optimisation est à peine doublée entre $\nu = 10^{-6}$ et $\nu = 10^{-10}$. Les solutions obtenues sont bien sûr identiques, à la valeur de ce seuil près : leurs composantes non nulles sont en particulier localisées aux mêmes positions. Par la suite, les résultats présentés ont été obtenus avec $\nu = 10^{-6}$.

L'évolution du critère en fonction du temps, représentée en figure 3.10 pour les algorithmes IRLS et GSOR, permet de mieux en appréhender le fonctionnement : alors que pour GSOR la diminution du critère est régulière au fil du temps, les premières itérations de l'IRLS sont pénalisées par la dimension élevée des systèmes linéaires à inverser[20]. Ce point est d'ailleurs d'autant plus pénalisant que le nombre de données N, fixant la complexité de la première phase de l'IRLS, est élevé. À l'inverse, dès que le nombre de composantes non nulles devient suffisamment faible, cet algorithme devient plus efficace que GSOR dans le sens où la décroissance du critère est plus rapide. Par conséquent, pour un paramètre de discrétisation P élevé, la rapidité de la seconde phase de l'IRLS l'emporte sur la lourdeur des premières itérations (dont la complexité est fixée par le nombre de données N et non par P) et l'IRLS devient globalement plus efficace que GSOR, ce dernier nécessitant un balayage de l'ensemble de la grille fréquentielle (de taille $2P+1$) à chaque itération. Les figures 3.10 en bas représentent, pour ces deux algorithmes, l'évolution correspondante du nombre $N_{-\mathcal{I}}^{(t)}$ de composantes non nulles, qui est décroissante dans les deux cas : alors que c'est une conséquence de la structure algorithmique de l'IRLS, cette remarque est plutôt à caractère empirique dans le cas de GSOR et propre à cet exemple : lors de l'application à des données plus complexes (notamment celles du chapitre 5), il est ainsi arrivé que le nombre $N_{-\mathcal{I}}^{(t)}$ varie de manière non monotone au fil des itérations. La valeur finale $N_{-\mathcal{I}}^{\infty}$ est rapidement atteinte par l'algorithme GSOR, pour lequel la majeure partie du temps de calcul s'effectue avec $N_{-\mathcal{I}}^{(t)} = N_{-\mathcal{I}}^{\infty}$. À l'inverse, la plupart du temps nécessaire à la convergence de l'IRLS est consacrée à la réduction de $N_{-\mathcal{I}}^{(t)}$, et chaque « accélération » de l'IRLS correspond à une diminution du nombre de composantes non nulles. Lorsque celui-ci atteint sa valeur finale, la convergence de l'IRLS est

19. Rappelons cependant que nous ne contrôlons pas le test d'arrêt de cet algorithme.

20. Il est en effet nécessaire d'initialiser l'IRLS avec une solution sans composante nulle (ici choisie constamment égale à l'unité), alors que l'algorithme GS peut être initialisé avec une solution identiquement nulle, plus proche de la solution parcimonieuse recherchée.

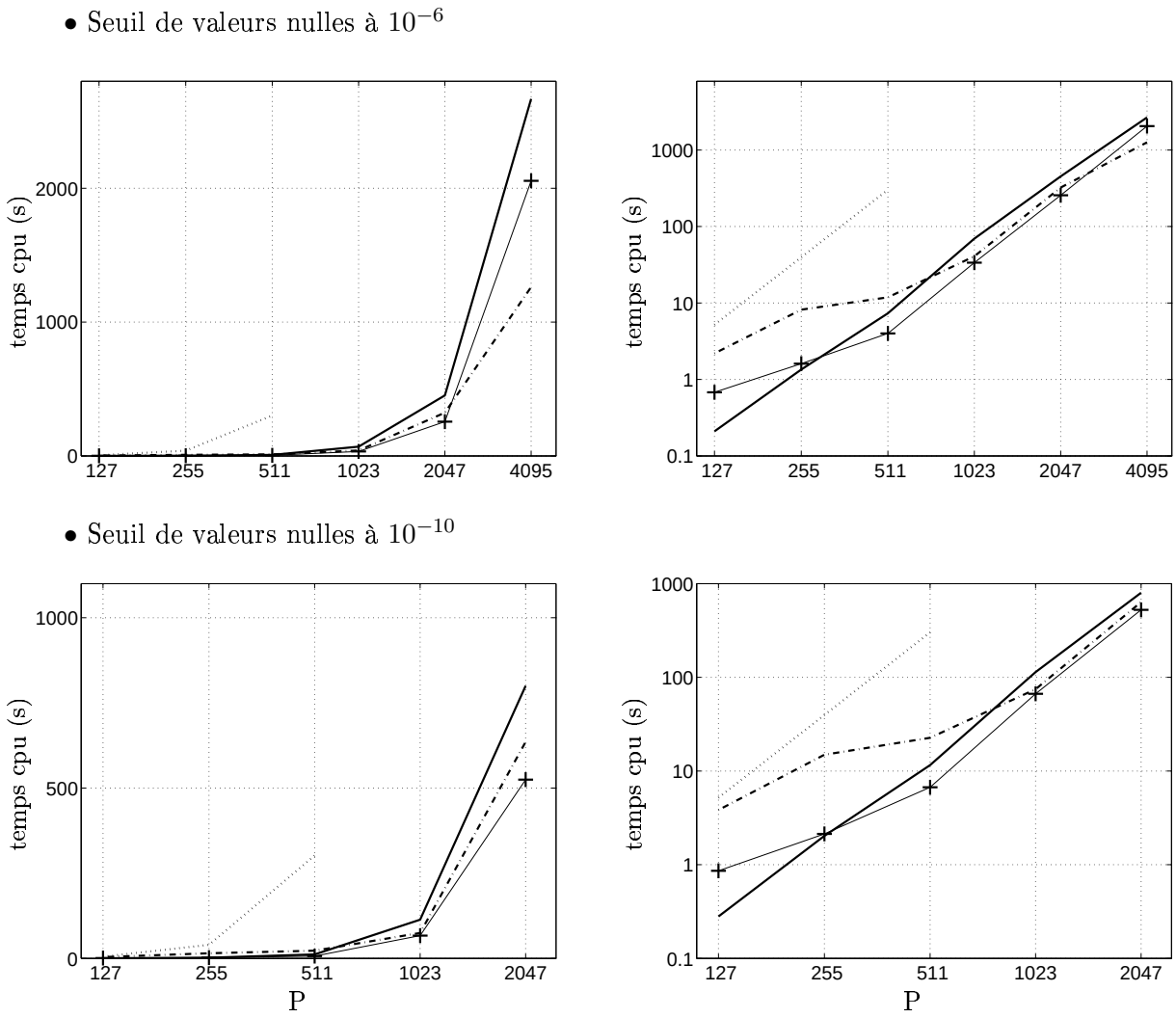


FIG. 3.9 – *Pénalisation ℓ^1 : temps cpu en fonction de la dimension P pour les algorithmes de programmation quadratique (···), IRLS (−·), GS (–) et GSOR (+) pour deux valeurs du seuil de valeurs nulles. À gauche, échelle de temps linéaire ; à droite, échelle de temps logarithmique.*

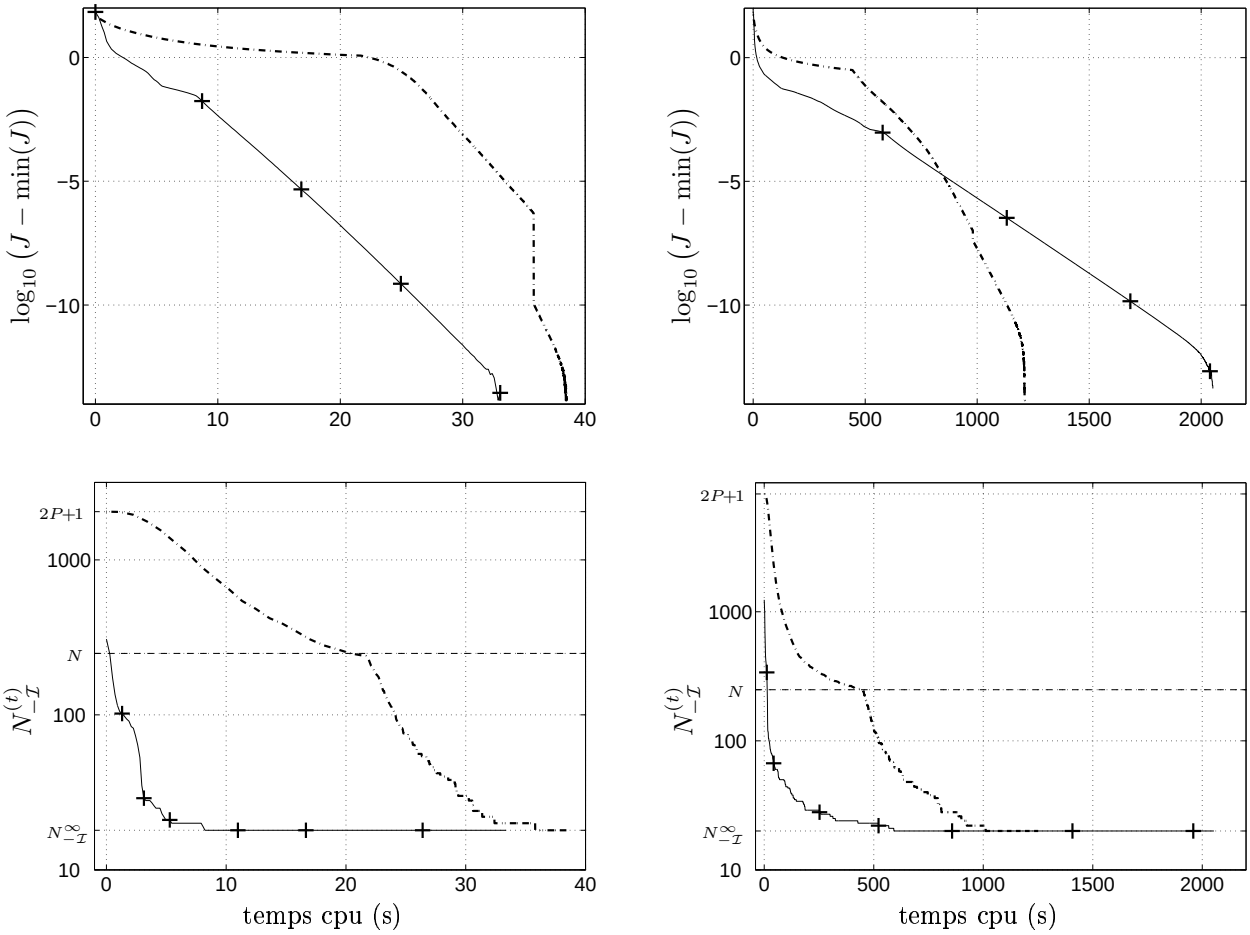


FIG. 3.10 – *Pénalisation ℓ^1 : comportement des algorithmes IRLS (−·) et GSOR (+) pour $P = 1023$ (à gauche) et $P = 4095$ (à droite). En haut, évolution du critère en fonction du temps. En bas, évolution correspondante du nombre $N_{-\mathcal{I}}$ de composantes non nulles.*

alors très rapide.

Les algorithmes IRLS et GSOR possèdent donc des caractéristiques complémentaires en fonction de la taille du problème considéré. Le tableau 3.3 résume les temps de calcul pour chacune des deux phases de l'optimisation décrites ci-dessus. Notons enfin qu'une stratégie hybride vi-

P	Temps cpu 1ère phase $N_{-\mathcal{I}}^{(t)} > N_{-\mathcal{I}}^{\infty}$		Temps cpu 2ème phase $N_{-\mathcal{I}}^{(t)} = N_{-\mathcal{I}}^{\infty}$		Temps cpu total	
	IRLS	GSOR	IRLS	GSOR	IRLS	GSOR
127	2.19	**0.47**	**0**	0.21	2.19	**0.68**
255	4.59	**0.87**	3.58	**0.73**	8.17	**1.6**
511	11.66	**1.62**	**0.17**	2.37	11.83	**3.99**
1023	38.05	**8.55**	**2.7**	25.15	40.75	**33.7**
2047	157.51	**72.43**	**165.69**	183.94	323.2	**256.37**
4095	1021.55	**594.64**	**240.12**	1460.93	**1261.67**	2055.57

TAB. 3.3 – *Pénalisation ℓ^1 : temps cpu en secondes des deux étapes des algorithmes IRLS et GSOR en fonction de la dimension P.*

sant à réduire dans un premier temps le nombre de composantes non nulles par GSOR, puis à appliquer une procédure IRLS, semble difficile à maîtriser. En effet, le premier algorithme ne bénéficie pas de la propriété de propagation des composantes nulles de l'IRLS : les indices des composantes non nulles peuvent théoriquement varier au fil des itérations. Par conséquent, une localisation erronée des composantes non nulles à l'initialisation de l'IRLS aura des conséquences irréversibles sur l'estimation produite.

3.8 Conclusions

Nous avons envisagé dans ce chapitre l'estimation de spectres de raies par la minimisation d'un critère pénalisé. La pénalisation quadratique se révélant inappropriée pour l'estimation de spectres de raies quelle que soit la forme de l'échantillonnage, nous avons utilisé la pénalisation hyperbolique adoptée par Ciuciu *et al.* [2001] en échantillonnage régulier et proposé alternativement l'emploi de la norme ℓ^1 opérant sur les modules complexes qui, à notre connaissance, n'a jamais été utilisé sous cette formulation à variables complexes du problème d'analyse spectrale. Bien que les nombreux travaux théoriques récents sur la norme ℓ^1 considèrent des variables réelles, nous avons montré que la pénalisation sur le module s'avère plus satisfaisante en pratique qu'une pénalisation opérant indépendamment sur les parties réelle et imaginaire, qui est pourtant l'approche adoptée par Chen et Donoho [1998].

Une structure rigoureusement parcimonieuse est ainsi obtenue pour l'estimateur $\widehat{\boldsymbol{x}}_1$, pour lequel la présence de composantes exactement nulles transparaît dans le détail des procédures d'optimisation associées. Ce n'est pas le cas de l'estimateur $\widehat{\boldsymbol{x}}_{\text{hyp}}$ pour lequel un seuillage *a posteriori* des amplitudes est nécessaire, ce qui ne pose généralement pas de problème particulier, et les deux estimateurs présentent finalement des allures similaires. Si la détection des fréquences est correctement effectuée, les amplitudes associées sont systématiquement sous-estimées, ce qui est inhérent à la régularisation par pénalisation ; cependant, une réestimation postérieure des amplitudes du modèle multi-sinusoïdal identifié par l'étape de détection précédente est immédiate par moindres carrés et fournit une estimation satisfaisante.

En l'absence d'informations concernant le niveau du bruit associé aux observations, le réglage non supervisé du paramètre de régularisation λ est un point qui reste délicat. Si la méthode de la « courbe en L » permet de déterminer un ordre de grandeur cohérent, l'estimation produite est assez sensible aux variations de λ. Une approche plus empirique consiste, à partir de la plage de valeurs fournie par la courbe en L, à détecter la valeur minimale de λ ne produisant pas visuellement de sous-régularisation (dans ce cas, de nombreux artéfacts de faible amplitude sont disséminés le long de la grille fréquentielle) : nous reviendrons sur cette approche lors du traitement de données réelles au chapitre 5.

Dans le cas de la pénalisation hyperbolique, où le critère à minimiser est strictement convexe et continûment différentiable, nous avons proposé une approximation permettant d'exploiter la structure de Toeplitz de la matrice $\mathbf{W}^\dagger\mathbf{W}$ lorsque l'échantillonnage est irrégulier. Le critère approché ainsi que son gradient étant calculables à moindre coût par des algorithmes FFT, la mise en œuvre de l'optimisation par un algorithme de Gradients Conjugués a permis d'obtenir une réduction de coût considérable par rapport à l'algorithme IRLS employé par Sacchi *et al.* [1998]. De plus, la prise en compte de la symétrie hermitienne du spectre à reconstruire a permis de réduire de moitié le temps de calcul d'une telle optimisation. Cependant, un schéma algorithmique de type GSOR – une version sur-relaxée d'une méthode de type Gauss-Seidel effectuant des minimisations itératives variable par variable – s'est révélé plus efficace, notamment lorsque le nombre de paramètres inconnus est élevé. Notons que, à la différence des résultats de simulation publiés dans [Bourguignon *et al.*, 2005], le critère d'arrêt que nous avons employé ici est plus sévère (la norme du gradient devant notamment être inférieure à 10^{-4}) et pénalise les algorithmes de GC qui, s'ils convergent assez rapidement vers une solution assez proche du minimiseur du critère, se sont révélés particulièrement lents pour obtenir une solution plus précise.

Pour la minimisation du critère pénalisé par la norme ℓ^1, l'approche par programmation quadratique n'est pas satisfaisante aussi bien sur le plan de la qualité de l'estimation – elle conduit à une pénalisation indépendante des parties réelle et imaginaire des amplitudes complexes, et nous avons vu que l'estimateur résultant pouvait en être affecté – qu'au niveau de la vitesse d'exécution. Nous avons proposé une mise en œuvre particulièrement efficace de l'algorithme IRLS pour une telle pénalisation portant sur les modules complexes et, afin de pallier les lacunes théoriques causées par l'absence de preuves de convergence, avons établi une condition permettant de vérifier *a posteriori* et de manière immédiate que la solution obtenue minimise le critère, rendant ainsi valide la procédure algorithmique. En pratique, ce test a toujours été vérifié sur l'ensemble des simulations effectuées. Nous avons également proposé un algorithme de type GSOR où, pour une pénalisation ℓ^1, chaque étape de minimisation scalaire nécessite un coût de calcul très réduit. Sur le plan de la mise en œuvre, nous notons encore que l'algorithme GSOR est de loin le plus simple, générant naturellement des composantes *exactement* nulles (et donc une solution parcimonieuse) alors que pour l'IRLS un seuillage des valeurs numériques faibles est nécessaire. Les procédures IRLS et GSOR ont montré en simulation des comportements complémentaires en termes de vitesse d'exécution. Si les premières itérations de l'IRLS sont pénalisées par la haute dimension des systèmes linéaires à inverser, liée au nombre de données, la convergence de celui-ci est plus rapide que celle de GSOR lorsque le nombre de paramètres non nuls dans l'itération en cours devient faible. La préférence vers l'une des deux solutions est donc un compromis en fonction du nombre de données, dont l'augmentation pénalise l'IRLS, et du nombre de paramètres, dont l'augmentation pénalise GSOR.

La pénalisation ℓ^1 est donc finalement sur ce problème plus avantageuse que la pénalisation hyperbolique. D'une part, elle représente une meilleure information de parcimonie, puisqu'elle

débouche sur une solution comportant essentiellement des valeurs *exactement* nulles (aux erreurs numériques près dans le cas de l'IRLS), et permet de s'affranchir d'une étape de post-traitement supplémentaire de seuillage des amplitudes faibles. Bien que, sur les simulations présentées, ce seuillage ne pose aucun problème interprétatif, ce point n'est cependant pas sans importance puisqu'il facilite la détection du « seuil de sous-régularisation » pouvant être associé au choix du paramètre de régularisation λ. D'autre part, et surtout, la comparaison des performances algorithmiques développées pour ces deux pénalisations donne un net avantage à la pénalisation ℓ^1. En particulier, la nature immédiate de l'optimisation à chaque itération de GSOR présente pour la norme ℓ^1 un coût optimal, qui ne peut qu'augmenter lors de l'emploi de toute fonction de pénalisation de type $\ell^2\ell^1$. Les temps de calcul « raisonnables » des procédures d'optimisation ont ainsi permis d'envisager une analyse à très haute résolution fréquentielle. Nous notons cependant que, sur les simulations effectuées, les raies détectées par $\widehat{\boldsymbol{x}}_1$ pour une grille $\mathcal{G}$ de $P = 4095$ fréquences positives présentent parfois un léger décalage (d'une ou deux positions fréquentielles) par rapport à la meilleure approximation dans $\mathcal{G}$ des fréquences théoriques.

Chapitre 4

Approche par modélisation Bernoulli-Gaussienne

Nous abordons dans ce chapitre une deuxième approche pour la recherche d'une solution parcimonieuse au problème d'analyse spectrale modélisé sous la forme linéaire introduite à la fin du chapitre 2 :

$$\begin{aligned} y(t_n) &= \sum_{k=0}^{P} a_k \cos 2\pi f_k t_n + b_k \sin 2\pi f_k t_n + \epsilon_n, \quad n = 1 \dots N \\ \Leftrightarrow \quad \boldsymbol{y} &= \mathbf{H}\boldsymbol{s} + \boldsymbol{\epsilon} \end{aligned} \tag{4.1}$$

où $\boldsymbol{s}_k = [a_k, b_k]$, $\boldsymbol{s}$ est le vecteur à $2P+2$ inconnues réelles $\boldsymbol{s} = [\boldsymbol{s}_0, \dots, \boldsymbol{s}_P]^T$ et $\mathbf{H}_{n,2k} = \cos(2\pi f_k t_n)$, $\mathbf{H}_{n,2k+1} = \sin(2\pi f_k t_n)$ pour $k = 0 \dots P$ et $n = 1 \dots N$. Rappelons que, comme nous l'avons proposé à la fin du chapitre 2, nous préférons utiliser ici l'écriture à variables *réelles* du modèle linéaire pour l'analyse spectrale alors qu'au chapitre précédent la formulation à variables complexes s'avère plus judicieuse afin de traduire plus simplement l'information de parcimonie par une fonction de pénalisation portant sur les *modules* $|x_k| = \frac{1}{2}\sqrt{a_k^2 + b_k^2}$. L'avantage de l'écriture (4.1) apparaîtra clairement lors de l'écriture des distributions de probabilité *a posteriori* des paramètres spectraux. La régularisation par pénalisation étudiée au chapitre précédent a été traitée dans un cadre *déterministe*, même si une interprétation bayésienne a été évoquée. Nous envisageons ici la prise en compte de l'information de parcimonie de manière *probabiliste*, sous la forme d'un modèle *a priori* Bernoulli-Gaussien (B-G) sur le vecteur d'amplitudes spectrales $\boldsymbol{s}$. Cette modélisation est fortement adaptée aux objets parcimonieux puisque la structure même du modèle impose la nullité de la plupart des composantes spectrales, la proportion de valeurs nulles étant réglée par la valeur du paramètre de Bernoulli.

La modélisation Bernoulli-Gaussienne a été initialement introduite pour traiter des problèmes de déconvolution impulsionnelle en géophysique, plus précisément dans le domaine de la sismique-réflexion : un train d'impulsions modélise les discontinuités rencontrées à l'interface de deux couches homogènes lors de l'exploration du sous-sol terrestre [Kormylo et Mendel, 1982]. Depuis, son champ d'application a été largement étendu, allant de l'échographie (voir par exemple [Champagnat, 1993]) aux télécommunications (par exemple [Fathi *et al.*, 2005]). Dans le cadre applicatif de l'estimation de spectres de raies, les travaux de Duvaut et Dublanchet [1995] et la thèse de doctorat de Dublanchet [1996] constituent à notre connaissance les seuls travaux étudiant la prise en compte d'une modélisation des amplitudes spectrales sous la forme Bernoulli-Gaussienne. Le

formalisme d'un tel modèle, ainsi que le cadre de l'estimation bayésienne dans lequel nous nous plaçons, sont introduits au § 4.1.

La plupart des travaux en déconvolution impulsionnelle conjuguent la modélisation B-G avec une estimation au sens du maximum *a posteriori* (MAP) de ses paramètres ; c'est également la démarche retenue par Dublanchet [1996] en application à l'analyse spectrale. Cette approche, que nous décrivons au § 4.2, présente néanmoins certains désagréments, notamment par la sous-optimalité de la procédure d'optimisation résultante et sa difficile adaptation en vue d'une estimation non supervisée. Une approche alternative introduite par Cheng *et al.* [1996] en déconvolution sismique aborde l'estimation des paramètres B-G au sens de la moyenne *a posteriori* (MP), mise en œuvre au moyen de techniques d'échantillonnage stochastique de type Monte-Carlo par chaînes de Markov (MCMC). Nous justifions au § 4.3 le choix de cet estimateur pour le problème de l'analyse spectrale, qui permet d'envisager de manière élégante une estimation totalement bayésienne non supervisée. En outre, l'échantillonnage de la distribution *a posteriori* d'intérêt permet d'envisager une procédure d'estimation particulièrement attrayante dans le cas de l'estimation de spectres de raies. À la différence de Cheng *et al.* qui estiment directement les amplitudes au sens de la moyenne *a posteriori*, nous proposons en particulier l'estimation de la moyenne *a posteriori* de la séquence de Bernoulli : l'estimé résultant, qui n'est alors plus une séquence binaire, peut être interprété en termes de probabilités associées à la détection éventuelle de raies spectrales, donnant ainsi un *niveau de confiance* sur les résultats de la procédure.

La richesse informationnelle obtenue par une telle démarche a néanmoins un prix, et la mise en œuvre de l'échantillonnage stochastique peut s'avérer délicate et coûteuse en temps de calcul. Le § 4.4 étudie l'échantillonneur de Gibbs introduit par Cheng *et al.* et que nous avons généralisé au cas de variables complexes. Nous mettons en évidence des lenteurs de convergence en application au problème d'analyse spectrale considéré, qui sont dues à l'inertie intrinsèque de l'algorithme de Gibbs classique, et proposons alors deux solutions permettant d'obtenir un schéma d'échantillonnage plus efficace. Les résultats du § 4.5 révèlent l'apport de cette approche pour l'estimation de raies spectrales. En particulier, nous montrons l'efficacité des algorithmes proposés et étudions l'influence des distributions *a priori* des hyperparamètres afin d'envisager une procédure non supervisée. Des résultats satisfaisants fournis par cette méthodologie pour la restauration de spectres de raies sont ensuite mis en évidence sur des données simulées.

Nous dégageons également certaines limites associées à cet estimateur, dont les caractéristiques de précision et de résolution sont restreintes par la discrétisation des fréquences sur une grille fixe. Dublanchet [1996] a proposé une extension du modèle B-G prenant en compte des *décalages* fréquentiels, associé à une estimation par une approche de type optimisation, mais qui n'a pas donné entière satisfaction. En effet, en plus de la sous-optimalité de la procédure de détection que nous avons évoquée plus haut, l'estimation des décalages au sens du MAP est à son tour également délicate. Nous reprenons cette modélisation au § 4.6 dans le cadre de l'estimation au sens de la moyenne *a posteriori* et proposons d'utiliser un algorithme MCMC hybride, permettant d'envisager l'estimation conjointe de l'ensemble des paramètres et de dépasser les limites associées à la mise en œuvre d'une approche par MAP. Comme dans le cas du modèle Bernoulli-Gaussien classique, la procédure que nous proposons donne en outre accès à des informations déterminantes pour la recherche de raies spectrales, à savoir des niveaux de confiance sur chacun des paramètres estimés : présence d'une raie, amplitude et fréquence associées, tout en étant non supervisée. L'application de cette nouvelle démarche et le gain apporté par rapport au modèle B-G classique sont finalement présentés au § 4.7 sur les données simulées utilisées au chapitre précédent.

Remarquons que, à la différence de l'approche par pénalisation où nous avons essayé de prendre en compte les spécificités de l'échantillonnage irrégulier au niveau des algorithmes proposés, le travail effectué ici s'applique indifféremment à toute forme d'échantillonnage. Nous verrons en effet lors de la description algorithmique du § 4.4 que les propriétés structurelles de l'opérateur $\mathbf{H}$ ne sont pas exploitables pour une procédure d'échantillonnage de Gibbs et n'influent donc pas sur la complexité algorithmique. Il n'en reste pas moins qu'un espacement irrégulier des observations confère au problème d'analyse spectrale des caractéristiques différentes : en particulier, le choix arbitraire de la fréquence maximale est un « avantage » sur le cas régulier (voir le § 1.3.1) mais, à l'inverse, une plus grande dégradation du spectre de Fourier, notamment pour des données trouées, rend plus complexe le problème de déconvolution associé.

4.1 Description du modèle

Nous décrivons en premier lieu le modèle probabiliste Bernoulli-Gaussien qui, bien qu'initialement développé pour décrire des séquences à valeurs réelles scalaires, s'adapte sans complication majeure au cas de paramètres complexes ou, de manière équivalente, réels bidimensionnels. La formulation ci-dessous est alors équivalente à celle proposée dans la thèse de Dublanchet [1996]. Nous introduisons par la suite le formalisme bayésien sous lequel nous étudierons l'estimation, naturellement adapté à une description probabiliste, puisqu'il permet de fusionner, sous la forme de la distribution de probabilité *a posteriori*, l'information *a priori* de structure Bernoulli–Gaussienne des inconnues et la *vraisemblance* des observations. Enfin, nous nous intéresserons au rôle des hyperparamètres du modèle qui, s'ils peuvent avoir une signification physique, doivent également pouvoir être réglés en l'absence totale d'information préalable : on parlera alors d'une procédure *non supervisée*.

4.1.1 Une structure hiérarchique ...

Alors que la plupart des travaux en déconvolution Bernoulli-Gaussienne considèrent des inconnues à valeurs réelles scalaires, le problème d'analyse spectrale mis sous la forme du modèle linéaire (4.1) fait intervenir des paramètres bidimensionnels : à chaque point de la grille fréquentielle $\mathcal{G} = \{\frac{k}{P} f_{\max}\}_{k=-P\dots P}$ correspond un vecteur d'amplitudes $[a_k, b_k]$, sur lesquelles on désire faire porter conjointement l'information de parcimonie. Nous définissons donc ici le modèle probabiliste Bernoulli-Gaussien (B-G) comme un processus aléatoire composite $(\boldsymbol{q},\boldsymbol{s})$, où :

- $\boldsymbol{q} = [q_0, \dots q_P]^{\mathrm{t}}$ est une séquence blanche de Bernoulli de paramètre $\lambda \in [0,1]$ conditionnant la présence d'une raie le long de la grille fréquentielle $\{f_k\}_{k=0\dots P}$:

$$\begin{cases} \Pr(q_k = 1) = \lambda, \text{ probabilité } \textit{a priori} \text{ d'existence d'une raie à la fréquence } f_k, \\ \Pr(q_k = 0) = 1 - \lambda. \end{cases} \tag{4.2}$$

  Les fréquences contenues dans le signal sont ainsi supposées indépendantes et réparties le long de la grille $\mathcal{G}$ avec la même probabilité d'existence en tout point. Une valeur faible de λ contrôlera alors le degré de parcimonie de la solution.
- Conditionnellement à la localisation apportée par la séquence $\boldsymbol{q}$, les amplitudes correspondantes $(a_k|q_k = 1)$ et $(b_k|q_k = 1)$ sont supposées *i.i.d.* gaussiennes centrées et de même variance σ^2. D'une manière formelle, nous pouvons écrire $(a_k|q_k), (b_k|q_k) \sim \mathcal{N}(0,q_k\sigma^2)$, soit encore sous forme bidimensionnelle :

$$p(\boldsymbol{s}_k|q_k) = g_2(\boldsymbol{s}_k,q_k\sigma^2\mathbf{I}_2), \tag{4.3}$$

où $g_m(\boldsymbol{u},\Sigma)$ est la distribution gaussienne multivariée de dimension m :

$$g_m(\boldsymbol{u},\Sigma) = \frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\boldsymbol{u}^T\Sigma^{-1}\boldsymbol{u}}$$

Lorsque $q_k = 0$, l'écriture (4.3) correspond à une distribution impropre gaussienne de variance nulle, c'est-à-dire à une distribution de Dirac bidimensionnelle $\delta_2(\boldsymbol{s}_k)$. Afin d'éviter de manipuler de telles distributions pour mener à terme les différents calculs qui seront nécessaires par la suite, nous remplacerons si besoin est la distribution δ_2 par une distribution de probabilité gaussienne $\mathcal{N}(0, \sigma_0^2\mathbf{I}_2)$ pour laquelle on fera tendre la variance σ_0^2 vers 0.
Notons que cette distribution *a priori* des amplitudes (a_k,b_k) correspond à une loi gaussienne *circulaire* sur la variable complexe $x_k = \frac{1}{2}(a_k - jb_k)$ du modèle (2.14), c'est-à-dire que sa phase suit une loi uniforme et son module une loi de Rayleigh de paramètre $\sigma/2$ (voir l'annexe B.1) :

$$\begin{cases} p(\arg x_k | q_k = 1) &= \dfrac{1}{2\pi}\ \mathbb{1}_{[0,\,2\pi]}(\arg x_k) \\ p(|x_k|\ \big|\ q_k = 1) &= \dfrac{|x_k|}{(\sigma/2)^2}\ \exp\left(-\dfrac{|x_k|^2}{2(\sigma/2)^2}\right) \end{cases}$$

Cette distribution est de moyenne $\frac{\sigma}{2}\sqrt{\frac{\pi}{2}}$ et prend son maximum en $\sigma/2$; une représentation est donnée en figure 4.1 pour $\sigma = 1$.

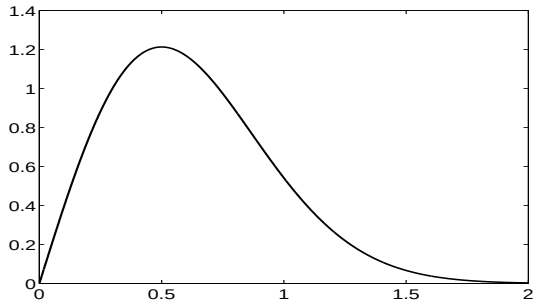


FIG. 4.1 – *Distribution a priori sur le module des amplitudes spectrales $|x_k|$ conditionnellement à l'existence d'une raie en f_k : loi de Rayleigh de paramètre $\sigma/2 = 1/2$, maximale en $1/2$ et de moyenne $\frac{1}{2}\sqrt{\frac{\pi}{2}} \simeq 0.63$.*

Selon cette description du modèle B-G $(\boldsymbol{q},\boldsymbol{s})$, les amplitudes spectrales $\boldsymbol{s}$ forment donc un processus blanc non gaussien, alors que la variable $\boldsymbol{s}|\boldsymbol{q}$ est gaussienne, centrée et de covariance $\sigma^2\mathrm{diag}(\underline{\underline{\boldsymbol{q}}})$, où $\underline{\underline{\boldsymbol{q}}}$ représente la séquence $\boldsymbol{q}$ « doublée » : $\underline{\underline{q}}_{2k} = \underline{\underline{q}}_{2k+1} = q_k$ pour $k = 0 \ldots P$, et $\mathrm{diag}(\underline{\underline{\boldsymbol{q}}})$ est la matrice diagonale d'éléments $\{\underline{\underline{q}}_k\}_{k=0\ldots 2P+1}$ à valeurs 0 ou 1. L'information apportée par une telle modélisation peut alors s'écrire sous la forme condensée de la distribution de probabilité *a priori*, qui admet la structure hiérarchique suivante, dans laquelle N_e représente le nombre de composantes non nulles dans la séquence $\boldsymbol{q}$:

$$\begin{aligned} p(\boldsymbol{q},\boldsymbol{s}) &= p(\boldsymbol{s}|\boldsymbol{q})\ \Pr(\boldsymbol{q}) \\ &= g_{P+1}(\boldsymbol{s},\boldsymbol{\Sigma}_{s|q})\ \lambda^{N_e}(1-\lambda)^{P+1-N_e} \quad \text{où} \quad \boldsymbol{\Sigma}_{s|q} \triangleq \sigma^2\mathbf{D}_{\underline{\underline{q}}} = \sigma^2\mathrm{diag}(\underline{\underline{\boldsymbol{q}}}) \end{aligned}$$

où la notation formelle $g_{P+1}(\boldsymbol{s},\boldsymbol{\Sigma}_{s|q})$ est impropre puisqu'elle fait intervenir une matrice de covariance diagonale[1] :

$$\boldsymbol{\Sigma}_{s|q} = \text{diag } (\mathbb{E}[\boldsymbol{s}_k \boldsymbol{s}_k^{\text{t}} | \boldsymbol{q}_k])_k = \text{diag}(q_k \sigma^2 \mathbf{I}_2)_k$$

comportant des éléments diagonaux nuls correspondant aux zéros de $\boldsymbol{q}$. Elle correspond alors rigoureusement à la distribution de probabilité :

$$g_{P+1}(\boldsymbol{s},\boldsymbol{\Sigma}_{s|q}) = \prod_{k/q_k=1} g_2(\boldsymbol{s}_k,\sigma^2\mathbf{I}_2) \prod_{k/q_k=0} \delta_2(\boldsymbol{s}_k). \tag{4.4}$$

Notons enfin que pour la fréquence nulle en $k = 0$, une loi *a priori* légèrement différente doit être considérée puisque l'on a alors $b_0 = -2\Im(x_0) = 0$. En choisissant conformément aux définitions précédentes $(a_0|q_0 = 1) \sim \mathcal{N}(0,\sigma^2)$, la relation (4.3) s'écrit donc de manière rigoureuse pour $k = 0$:

$$p(\boldsymbol{s}_0|q_0 = 1) = g_1(a_0,\sigma^2)\, \delta(b_0)$$

Par souci de clarté, les modifications correspondantes ne seront pas reportées dans l'exposé qui suit, mais la prise en compte des spécificités résultantes sera bien entendu détaillée lorsque cela s'avérera nécessaire.

4.1.2 ... adaptée à l'estimation bayésienne

Le cadre bayésien se prête naturellement à l'estimation des paramètres B-G $(\boldsymbol{q},\boldsymbol{s})$. En effet, toutes les informations concernant le modèle linéaire (4.1) peuvent être décrites de manière statistique, avec :

- la modélisation *a priori* Bernoulli-Gaussienne dont l'information est portée sous la forme de la distribution de probabilité *a priori* $p(\boldsymbol{q},\boldsymbol{s})$ décrite au paragraphe précédent ;
- une description statistique des perturbations ϵ, permettant de définir la *vraisemblance* des observations. En supposant comme au chapitre 3 les perturbations $\boldsymbol{\epsilon}$ additives, *i.i.d.* gaussiennes centrées et de variance σ_ϵ^2, la vraisemblance s'écrit :

$$\mathcal{L}(\boldsymbol{y};\boldsymbol{q},\boldsymbol{s},\sigma_\epsilon^2) = p(\boldsymbol{y}|\boldsymbol{q},\boldsymbol{s},\sigma_\epsilon^2) = g_N(\boldsymbol{y} - \mathbf{H}\boldsymbol{s}, \sigma_\epsilon^2\mathbf{I}_N).$$

La règle de Bayes permet alors de fusionner l'information *a priori* et celle contenue dans les données, sous la forme de la distribution de probabilité *a posteriori* des paramètres B-G :

$$\begin{aligned} p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y},\boldsymbol{\theta}) &\propto p(\boldsymbol{y}|\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta})\, p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{\theta}) \\ &\propto g_N(\boldsymbol{y} - \mathbf{H}\boldsymbol{s}, \sigma_\epsilon^2\mathbf{I}_N)\ g_{P+1}(\boldsymbol{s},\boldsymbol{\Sigma}_{s|q})\ \lambda^{N_e}(1-\lambda)^{P+1-N_e} \end{aligned} \tag{4.5}$$

où $\boldsymbol{\theta}$ regroupe les *hyperparamètres* du problème, c'est-à-dire les paramètres réglant le comportement des différentes distributions introduites : $\boldsymbol{\theta} = (\lambda,\sigma_\epsilon^2,\sigma^2)$. La théorie bayésienne de l'estimation laisse alors le champ libre à l'usage de différentes sortes d'estimateurs basés sur l'exploitation de la distribution *a posteriori* $p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y},\boldsymbol{\theta})$. Le choix d'une stratégie adaptée à notre problème fera l'objet des paragraphes 4.2 et 4.3.

Remarquons que le cas plus général de perturbations non *i.i.d.* – mais de statistique connue – peut être pris en compte en modifiant en conséquence le terme de vraisemblance. En particulier,

1. La notation $\text{diag}(\boldsymbol{a}_k)_k$ correspond à la matrice diagonale d'*éléments* $\boldsymbol{a}_k$ lorsque ceux-ci sont scalaires, et par extension à la matrice bloc-diagonale formée par les *blocs* $\boldsymbol{a}_k$ lorsque ceux-ci sont des matrices.

le cas de perturbations indépendantes et gaussiennes mais de variances différentes connues $\epsilon_n \sim \mathcal{N}(0,\sigma^2_{\epsilon,n})$, situation fréquente pour des observations astronomiques, sera traité en remplaçant le premier terme dans l'expression ci-dessus par la vraisemblance correspondante $g_N(\boldsymbol{y} - \mathbf{H}\boldsymbol{s}, \boldsymbol{\Sigma}_\epsilon)$, où $\boldsymbol{\Sigma}_\epsilon$ est la matrice diagonale d'éléments $\sigma^2_{\epsilon,n}$.

Notons que le papier fondateur pour la déconvolution B-G de Kormylo et Mendel [1982], ainsi que les travaux postérieurs de l'équipe de Mendel (par exemple [Chi et Mendel, 1984, Goutsias et Mendel, 1986]) utilisent une définition légèrement différente d'un processus B-G. Pour des paramètres *réels*, ils considèrent $\boldsymbol{s}$ comme un processus produit $s_k = q_k w_k$, où w_k est un processus blanc gaussien centré de variance σ^2 et *indépendant* de q_k, et examinent le problème de l'estimation de $(\boldsymbol{q},\boldsymbol{w})$. Goussard [1989] montre que la maximisation de la loi *a posteriori* jointe en $(\boldsymbol{q},\boldsymbol{w})$ ou en $(\boldsymbol{q},\boldsymbol{s})$ débouche sur des solutions différentes selon la formulation adoptée, et justifie le choix du modèle $(\boldsymbol{q},\boldsymbol{s})$ qui revêt une structure plus naturelle que le modèle produit $(\boldsymbol{q},\boldsymbol{w})$. Cependant, l'ensemble de ces travaux privilégie une estimation basée sur la loi *marginale* $p(\boldsymbol{q}|\boldsymbol{y},\boldsymbol{\theta})$, qui revêt la même expression indépendamment de la définition sous-jacente du modèle B-G et les deux formulations sont dans ce cas équivalentes.

4.1.3 Loi jointe ou loi marginale?

Un premier élément à envisager dans l'optique de l'estimation bayésienne des paramètres Bernoulli-Gaussiens concerne le choix même de la loi à la base de la procédure d'estimation. En effet, sous l'hypothèse de gaussianité des amplitudes $\boldsymbol{s}$, leur marginalisation dans la loi (4.5) s'obtient sans difficulté. Se pose alors la question de savoir laquelle, de la loi jointe ou de la loi marginale, est la mieux adaptée.

Afin d'alléger les écritures, nous omettons pour l'instant la dépendance en $\boldsymbol{\theta}$ des différentes distributions. Nous rappelons tout d'abord l'expression exacte de la loi jointe *a posteriori*, dérivée des relations (4.4) et (4.5) :

$$p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y}) = g_N(\boldsymbol{y} - \mathbf{H}\boldsymbol{s}, \sigma^2_\epsilon \mathbf{I}_N) \prod_{k/q_k=1} g_2(\boldsymbol{s}_k,\sigma^2\mathbf{I}_2) \prod_{k/q_k=0} \delta_2(\boldsymbol{s}_k)\ \lambda^{N_e}(1-\lambda)^{P+1-N_e} \tag{4.6}$$

où N_e est le nombre de valeurs non nulles dans la séquence de Bernoulli $\boldsymbol{q}$. Sous les hypothèses de gaussianité de $\boldsymbol{\epsilon}$ et des amplitudes $\boldsymbol{s}$, la linéarité du modèle (4.1) permet d'affirmer que la variable conditionnelle $\boldsymbol{y}|\boldsymbol{q}$ est également gaussienne, avec :

$$\begin{aligned} \mathbb{E}[\boldsymbol{y}|\boldsymbol{q}] &= \mathbf{H}\mathbb{E}[\boldsymbol{s}] = \mathbf{0} \\ \mathbb{E}[\boldsymbol{y}\boldsymbol{y}^{\mathrm{t}}|\boldsymbol{q}] &= \sigma^2_\epsilon \mathbf{I}_N + \mathbf{H}\boldsymbol{\Sigma}_{s|q}\mathbf{H}^{\mathrm{t}} \ \triangleq\ \boldsymbol{\Sigma}_{y|q} \end{aligned}$$

où, rappelons-le, $\boldsymbol{\Sigma}_{s|q}$ est la covariance *a priori* : $\boldsymbol{\Sigma}_{s|q} = \sigma^2 \mathrm{diag}(\underline{\underline{\boldsymbol{q}}})$. L'expression de la loi *a posteriori* marginale $\Pr(\boldsymbol{q}|\boldsymbol{y})$ s'obtient alors par application de la règle de Bayes :

$$\begin{aligned} \Pr(\boldsymbol{q}|\boldsymbol{y}) &= p(\boldsymbol{y}|\boldsymbol{q})\ \Pr(\boldsymbol{q}) \\ &= g_N(\boldsymbol{y},\boldsymbol{\Sigma}_{y|q})\ \lambda^{N_e}(1-\lambda)^{P+1-N_e} \\ &\text{où}\quad \boldsymbol{\Sigma}_{y|q} = \sigma^2_\epsilon\, \mathbf{I}_N + \sigma^2\, \mathbf{H}\, \mathrm{diag}(\underline{\underline{\boldsymbol{q}}})\, \mathbf{H}^{\mathrm{t}}. \end{aligned} \tag{4.7}$$

Deux stratégies différentes d'estimation des paramètres B-G peuvent alors être menées :

- l'estimation conjointe des paramètres $(\boldsymbol{q},\boldsymbol{s})$ à partir de la loi jointe (4.6) ;

– la détection préalable de la séquence de Bernoulli $\widehat{\boldsymbol{q}}$ à partir de la loi marginale (4.7), puis l'estimation des amplitudes conditionnellement à la connaissance de $\widehat{\boldsymbol{q}}$. Cette deuxième étape peut être effectuée indistinctement, sous l'hypothèse gaussienne de $\boldsymbol{s}|\boldsymbol{q}$, au sens du maximum *a posteriori* ou de la moyenne *a posteriori*. Le vecteur d'amplitudes $\widehat{\boldsymbol{s}}$ maximisant la loi $p(\boldsymbol{s}|\widehat{\boldsymbol{q}},\boldsymbol{y})$ s'obtient alors immédiatement par [Idier *et al.*, 2004] :

$$\widehat{\boldsymbol{s}} = \mathbb{E}[\boldsymbol{s} \,|\, \widehat{\boldsymbol{q}}, \boldsymbol{y}] = \boldsymbol{\Sigma}_{s|\widehat{q}}\, \mathbf{H}^{\mathrm{t}}\, \boldsymbol{\Sigma}_{y|\widehat{q}}{}^{-1}\, \boldsymbol{y}. \tag{4.8}$$

De par la complexité des deux distributions (4.6) et (4.7), notamment leur aspect multidimensionnel, une étude comparative théorique de leurs comportements semble difficile et ne permet pas de privilégier de manière générale une des deux stratégies. Quelques éléments importants retiendront cependant notre attention à ce sujet :

– La covariance *a posteriori* $\boldsymbol{\Sigma}_{y|q}$ étant définie positive, la loi marginale (4.7) est proprement définie. Ce n'est pas le cas de la loi jointe (4.6), dont l'utilisation nécessite la manipulation de distributions de Dirac.
– La marginalisation des amplitudes complique la dépendance de la distribution en les hyperparamètres $(\sigma_\epsilon^2,\sigma^2)$, qui est alors « cachée » dans l'expression de $\boldsymbol{\Sigma}_{y|q} = \sigma_\epsilon^2\mathbf{I}_N + \boldsymbol{\sigma}^2\mathbf{H}\underline{\underline{\mathbf{D}}}_{\boldsymbol{q}}\mathbf{H}^{\mathrm{t}}$. De ce fait, les lois *a posteriori* de σ_ϵ^2 et σ^2 extraites de la loi marginale perdent la simplicité analytique qui caractérise les lois obtenues à partir de la loi jointe (nous verrons au § 4.3.4 que ces lois sont de type inverse-Gamma). Le choix d'une paramétrisation alternative en $(\sigma_\epsilon^2, \delta^2 = \sigma^2/\sigma_\epsilon^2)$, proposé par Andrieu et Doucet [1999] pour une modélisation *a priori* Poisson-Gaussienne, permet d'obtenir une dépendance plus explicite de la loi (4.7) en σ_ϵ^2, qui est alors également de type inverse-Gamma, réduisant ce désagrément au seul cas du paramètre δ^2.
– Nous verrons aux paragraphes 4.2 et 4.3 que le choix de la distribution à la base de la procédure d'estimation est également en lien avec la définition même de l'estimateur : une approche de type maximum *a posteriori* (MAP) fournira des résultats plus satisfaisants en se fondant sur la loi marginale $\Pr(\boldsymbol{q}|\boldsymbol{y})$, alors que l'utilisation de la loi jointe $p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y})$ s'avérera plus efficace pour le calcul de l'estimateur de la moyenne *a posteriori* (MP).

Par souci de clarté de l'exposé, nous noterons par la suite de manière générale $\boldsymbol{u}$ l'ensemble des paramètres B-G considérés : dans le cas de la distribution jointe, $\boldsymbol{u}$ désignera l'ensemble $(\boldsymbol{q},\boldsymbol{s})$ alors que, pour la loi marginale, $\boldsymbol{u}$ fera uniquement référence à la séquence de Bernoulli $\boldsymbol{q}$.

4.1.4 Quid des hyperparamètres?

L'obtention d'un réglage satisfaisant des hyperparamètres $\boldsymbol{\theta} = (\lambda,\sigma_\epsilon^2,\sigma^2)$ est une étape essentielle pour la procédure d'estimation. En effet, la valeur du paramètre de Bernoulli λ conditionne le nombre *a priori* recherché d'oscillations sinusoïdales dans les données, puisqu'il correspond à la probabilité d'existence d'une raie en chaque point de la grille fréquentielle. Pour une grille $\mathcal{G} = \{\frac{k}{P} f_{\max}\}_{k=0\ldots P}$ de $P+1$ points, on aura donc *a priori* en moyenne $(P+1)\lambda$ fréquences d'oscillation. Les écarts-type σ_ϵ^2 et σ^2 sont quant à eux reliés respectivement au niveau de bruit perturbant les données et à l'amplitude moyenne *a priori* attendue des sinusoïdes, caractérisée par la distribution $p\big(|x_k|\big)$ représentée en figure 4.1.

Selon la nature des observations, certaines informations préalables concernant les hyperparamètres peuvent être disponibles. Par exemple, le type de l'objet étudié peut renseigner sur le nombre approximatif de sinusoïdes recherchées et on peut disposer dans certains cas d'une

estimation du niveau de bruit σ_ϵ en fonction des conditions observationnelles (instrument utilisé, qualité du site d'observation et conditions météorologiques). Il est encore possible d'obtenir un ordre de grandeur du rapport signal sur bruit associé au modèle B-G, dont la valeur est reliée à celle des trois hyperparamètres : si P_s et P_ϵ représentent les puissances respectives du signal et du bruit, on a $P_\epsilon = \mathbb{E}[\epsilon_n^2] = \sigma_\epsilon^2$ et :

$$\begin{aligned} P_s &= \mathbb{E}[\big(\sum_{k=0}^{P} a_k \cos 2\pi f_k t_n + b_k \sin 2\pi f_k t_n\big)^2] \\ &= \mathbb{E}[a_k^2] \sum_k \cos^2 2\pi f_k t_n + \mathbb{E}[b_k]^2 \sum_k \sin^2 2\pi f_k t_n \\ &= 2\,(P+1)\,\lambda\,\sigma^2. \end{aligned}$$

Le rapport signal sur bruit (RSB) s'exprime alors sous la forme :

$$\mathrm{RSB} = \frac{P_s}{P_\epsilon} = 2\,(P+1)\,\frac{\lambda\,\sigma^2}{\sigma_\epsilon^2}.$$

La connaissance de certaines caractéristiques observationnelles peut donc aider à fixer les valeurs des hyperparamètres $\boldsymbol{\theta}$. Néanmoins, la perspective de leur réglage de manière automatique représente un enjeu majeur afin, d'une part, de pallier le manque éventuel d'information sur $\boldsymbol{\theta}$, et d'autre part d'assouplir la procédure d'estimation qui pourrait se révéler trop sensible à une valeur fixée de manière incorrecte d'un ou de plusieurs hyperparamètres. Lorsqu'elle s'affranchit de tout réglage préalable des hyperparamètres, la procédure d'estimation est alors dite *non supervisée*. Afin de desservir cet objectif, l'approche la plus intuitive consiste à *estimer* les paramètres $\boldsymbol{\theta}$ à partir des données en maximisant la vraisemblance marginale :

$$\begin{aligned} p(\boldsymbol{y}|\boldsymbol{\theta}) &= \int p(\boldsymbol{y},\boldsymbol{q},\boldsymbol{s}|\boldsymbol{\theta})\, d(\boldsymbol{q},\boldsymbol{s}) \\ &= \sum_{\boldsymbol{q}} \Pr(\boldsymbol{q}|\boldsymbol{\theta}) \int p(\boldsymbol{y}|\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta})\, p(\boldsymbol{s}|\boldsymbol{q},\boldsymbol{\theta}) d\boldsymbol{s} \end{aligned}$$

Le nombre de séquences $\boldsymbol{q}$ possibles étant de 2^{P+1}, le calcul de $p(\boldsymbol{y}|\boldsymbol{\theta})$, sans parler de sa maximisation en $\boldsymbol{\theta}$, est cependant irréalisable en pratique. Des solutions alternatives doivent alors être proposées, qui dépendront de la stratégie d'estimation employée [Idier, 2001a, chapitre 8]. La maximisation en $\boldsymbol{\theta}$ de la vraisemblance marginale peut être abordée de manière indirecte par un algorithme SEM (*Stochastic Expectation-Maximization*). L'estimation des hyperparamètres peut encore être envisagée par la maximisation conjointe en les paramètres B-G et $\boldsymbol{\theta}$ de la *vraisemblance généralisée* $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{y}|\boldsymbol{\theta})$. Ces solutions, étant par nature liées à une approche de type maximisation, seront détaillées au § 4.2. Une procédure alternative totalement bayésienne permet, par une description probabiliste des hyperparamètres $\boldsymbol{\theta}$, de considérer la loi jointe *a posteriori* $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y})$ et l'estimation conjointe de l'ensemble des paramètres *aléatoires* $(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta})$. Nous verrons au § 4.3 que l'exploitation de cette distribution prend tout son intérêt lorsqu'elle est associée à une estimation au sens de la moyenne *a posteriori*.

4.2 Estimation au sens du maximum *a posteriori* (MAP)

La stratégie d'estimation la plus fréquemment rencontrée dans un cadre bayésien est certainement l'estimation au sens du maximum *a posteriori* (MAP). Historiquement, la plupart

des travaux en déconvolution Bernoulli-Gaussienne abordent ainsi le problème d'estimation sous l'angle de la maximisation de la distribution *a posteriori* des paramètres B-G :

$$\mathring{\boldsymbol{u}}_{\mathrm{MAP}} \triangleq \arg\max_{\boldsymbol{u}} p(\boldsymbol{u}|\boldsymbol{y})$$

où $\boldsymbol{u}$ représente alternativement $\boldsymbol{u} = (\boldsymbol{q},\boldsymbol{s})$ (loi jointe) ou $\boldsymbol{u} = \boldsymbol{q}$ (loi marginale). Afin d'alléger les notations, la dépendance des diverses distributions considérées en les hyperparamètres $\boldsymbol{\theta}$ ne sera pas explicitée pour l'instant. Ces paramètres seront toutefois réintroduits lorsque nous envisagerons le problème de l'estimation en mode non supervisé au § 4.2.2.

Pour une séquence de Bernoulli $\boldsymbol{q}$ donnée, la maximisation de la loi jointe en les amplitudes $\boldsymbol{s}$ correspond à l'estimation MAP de la variable conditionnelle $(\boldsymbol{s}|\boldsymbol{q},\boldsymbol{y})$, qui prend la forme explicite (4.8). Par conséquent, la maximisation de $p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y})$ peut encore s'écrire essentiellement comme un problème d'optimisation en les seuls paramètres $\boldsymbol{q}$, les amplitudes $\boldsymbol{s}$ correspondantes se déduisant *a posteriori* par la relation (4.8). Ainsi, que l'on cherche à maximiser la loi jointe (4.6) ou la loi marginale (4.7), le point délicat de la procédure d'estimation réside dans la maximisation d'un critère en les variables de Bernoulli, que nous noterons sous la forme générique $J(\boldsymbol{q})$. Notons que la maximisation en $\boldsymbol{q}$ des deux critères ne débouche pas *a priori* sur des séquences de Bernoulli identiques ; cependant, Kormylo et Mendel [1982] montrent que ceux-ci présentent des expressions proches : seule l'intervention d'un déterminant matriciel les différencie. La mise en œuvre de leur optimisation est donc de complexité comparable, qui est celle d'un problème d'ordre *combinatoire*, pour lequel 2^{P+1} séquences $\boldsymbol{q}$ sont possibles. Le nombre d'inconnues P étant élevé pour satisfaire les exigences d'une analyse à haute résolution – typiquement de l'ordre du millier –, il est inconcevable d'envisager une exploration exhaustive de l'espace des solutions.

4.2.1 Un algorithme performant : SMLR

De nombreux travaux en déconvolution B-G dérivent de l'algorithme SMLR (*Single Most Likely Replacement*), introduit par le papier fondateur de Kormylo et Mendel [1982], visant à l'optimisation du critère $J(\boldsymbol{q})$. Le principe consiste, pour une itération correspondant à la séquence $\boldsymbol{q}_t$, à rechercher un maximum du critère parmi un ensemble restreint de séquences $\boldsymbol{q}_r$ *voisines*, et à itérer le processus jusqu'à une séquence au voisinage de laquelle aucune amélioration n'est obtenue. L'intérêt de cette démarche réside alors dans l'obtention de formules numériquement peu coûteuses pour calculer la valeur de $J(\boldsymbol{q}_r|\boldsymbol{y})$ à partir de $J(\boldsymbol{q}_t|\boldsymbol{y})$: les auteurs choisissent ainsi des voisinages pour lesquelles deux séquences ne diffèrent qu'en un point, produisant des variations de J calculables à moindre coût. Différentes variantes ont ainsi été proposées sur ce schéma en enrichissant la structure du voisinage parcouru [Chi et Mendel, 1984, Kaaresen, 1997], débouchant sur une exploration plus complète de l'espace des séquences de Bernoulli au prix d'une complexité plus élevée de chaque itération. Depuis la contribution de Kormylo et Mendel, des algorithmes performants ont été développés pour le calcul de $J(\boldsymbol{q}_r)$ à partir de $J(\boldsymbol{q}_t)$. Ainsi, l'introduction de variables auxiliaires adéquates dans [Goussard *et al.*, 1990] permet d'utiliser un algorithme de type moindres carrés récursifs, et l'exploitation de la structure creuse de la séquence de Bernoulli proposée par Champagnat *et al.* [1996] permet de réduire la dimension de l'inversion matricielle intervenant dans ces calculs. Une telle démarche permet donc d'aboutir à un maximum – *a priori* local – du critère J, c'est-à-dire de la loi marginale ou, après estimation des amplitudes, de la loi jointe, pour un coût calculatoire raisonnable : ainsi, des travaux récents en communications numériques envisagent un traitement en temps réel par l'implantation d'une procédure de type SMLR à bord de récepteurs de téléphonie mobile dans les années à venir [Fathi *et al.*, 2005].

Notons que la mise en œuvre de l'algorithme SMLR et de ses variantes est en général préférablement associée à la maximisation de la loi marginale $Pr(\boldsymbol{q}|\boldsymbol{y})$ [Kormylo et Mendel, 1982, Goutsias et Mendel, 1986, Champagnat *et al.*, 1996], sous l'argument empirique que la maximisation de la loi jointe produit des résultats surestimant le nombre de valeurs non nulles dans la séquence de Bernoulli $\boldsymbol{q}$. La thèse de doctorat de Goussard [1989] présente à ce sujet des simulations mettant en évidence un plus grand nombre de fausses détections lors de la maximisation de la loi jointe, alors que des valeurs correctes des hyperparamètres ont été employées. L'auteur propose de pallier ce problème avec un réglage *ad hoc* des hyperparamètres, solution qui permet de retrouver une estimation satisfaisante à partir de la loi jointe. Néanmoins, une meilleure adéquation des hyperparamètres pour la loi marginale, ainsi que la similarité des problèmes d'optimisation associés à chaque cas, incite à préférer la maximisation de la loi marginale. C'est d'ailleurs l'approche envisagée par Dublanchet [1996] pour ce même problème d'analyse spectrale, où l'estimation est réalisée selon la démarche séquentielle suivante :

i) $\boldsymbol{q}_{\text{MMAP}}$ est l'estimateur du maximum *a posteriori* marginal[2] :

$$\boldsymbol{q}_{\text{MMAP}} = \arg\max_{\boldsymbol{q}} \Pr(\boldsymbol{q}|\boldsymbol{y}),$$

ii) $\boldsymbol{s}_{\text{MAP}}$ est l'estimateur du maximum *a posteriori* de $\boldsymbol{s}|\boldsymbol{q}$:

$$\boldsymbol{s}_{\text{MAP}} = \arg\max_{\boldsymbol{s}} p(\boldsymbol{s} \,|\, \boldsymbol{y}, \boldsymbol{q} = \boldsymbol{q}_{\text{MMAP}}).$$

Il faut néanmoins signaler qu'une telle méthodologie est nécessairement sous-optimale : le caractère non exhaustif de l'exploration de l'ensemble des séquences $\boldsymbol{q}$ ne peut garantir l'obtention de la séquence maximisant le critère J sur l'ensemble des 2^{P+1} séquences candidates. Des exemples de simulations pour lesquelles l'algorithme SMLR reste bloqué dans des maxima locaux sont présentés dans [Goussard, 1989]. Dans le cas de l'analyse spectrale, Dublanchet [1996] observe ainsi, pour une discrétisation fine de la grille fréquentielle – c'est-à-dire une valeur élevée de P – le dédoublement des raies spectrales ; nous verrons au § 4.4.1 que ce problème peut également correspondre à un blocage de la procédure d'optimisation au voisinage d'un minimum local de la distribution *a posteriori*[3].

4.2.2 Estimation délicate des hyperparamètres

Une seconde limite concernant les algorithmes d'optimisation de type SMLR réside en ce qu'ils considèrent le jeu d'hyperparamètres connu, et leur structure n'est pas directement adaptable à une estimation simultanée de $\boldsymbol{\theta}$ et des paramètres B-G : la maximisation en $\boldsymbol{u} = \boldsymbol{q}$ (respectivement en $\boldsymbol{u} = (\boldsymbol{q},\boldsymbol{s})$) de la distribution *a posteriori* $\Pr(\boldsymbol{q}|\boldsymbol{y},\boldsymbol{\theta})$ (respectivement $p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y},\boldsymbol{\theta})$) s'effectue pour une valeur donnée de $\boldsymbol{\theta}$. Sur le plan informationnel, nous considérons ici que les hyperparamètres sont *fixés*, même s'ils sont de valeur inconnue : toutes les distributions sont définies conditionnellement à la valeur de $\boldsymbol{\theta}$. L'approche considérant les hyperparamètres comme des variables aléatoires permet d'envisager des distributions jointes en les paramètres B-G et $\boldsymbol{\theta}$ (« le passage de $\boldsymbol{\theta}$ dans le terme de gauche des distributions conditionnelles ») : cette démarche *totalement bayésienne* sera envisagée au § 4.3.

Une première approche adaptée à une stratégie de type MAP consiste en la maximisation

2. Plus exactement, une approximation $\widehat{\boldsymbol{q}}_{\text{MMAP}}$ obtenue par l'algorithme SMLR, qui ne présente pas de garantie d'obtention du maximum global.

3. Nous faisons ici référence à des dédoublements de type $(q_{k_0-1}, q_{k_0}, q_{k_0+1}) = (1, 0, 1)$ alors que la véritable raie est localisée en f_{k_0}.

de la *vraisemblance généralisée*, définie comme la densité de probabilité conjointe $p(\boldsymbol{y},\boldsymbol{u}|\boldsymbol{\theta})$ de toutes les quantités aléatoires connaissant les hyperparamètres. Un tel estimateur a été étudié pour la loi jointe en $\boldsymbol{u} = (\boldsymbol{q},\boldsymbol{s})$ par [Gassiat *et al.*, 1992] et pour la loi marginale en $\boldsymbol{u} = \boldsymbol{q}$ par [Champagnat et Idier, 1995], dans le cas d'un modèle sans convolution, c'est-à-dire de type $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\epsilon}$, où $\boldsymbol{x}$ est la réalisation d'un processus Bernoulli-Gaussien. Les données y_n étant alors statistiquement indépendantes, cette simplification permet d'établir des propriétés théoriques sur les estimateurs résultants. Dans le cas de la loi jointe, Gassiat *et al.* [1992] ont montré l'inefficacité de l'estimateur résultant, dont l'existence même n'est pas garantie. À l'inverse, l'estimation de $\boldsymbol{\theta}$ par maximisation de la vraisemblance généralisée marginale $p(\boldsymbol{y},\boldsymbol{q}|\boldsymbol{\theta})$ présente des propriétés plus solides [Champagnat et Idier, 1995] et s'avère en particulier intéressante lorsque le nombre de données est faible. Notons par contre que si, dans le cadre de ces travaux, l'estimation des hyperparamètres est obtenue de manière simple – en particulier, la maximisation en $\boldsymbol{\theta}$ de $p(\boldsymbol{y},\boldsymbol{u}|\boldsymbol{\theta})$ ne dépend pas des paramètres Bernoulli-Gaussiens $\boldsymbol{u}$ – ce n'est plus le cas pour un modèle convolutif comme le nôtre. L'estimateur peut cependant être obtenu (de manière sous-optimale) par l'alternance itérée d'une étape de maximisation en les paramètres B-G à $\boldsymbol{\theta}$ fixé avec une étape d'optimisation en les hyperparamètres à $\boldsymbol{u}$ fixé. Alors que la première étape est celle qui a été décrite plus haut – réalisable par exemple par un algorithme SMLR – l'étape de maximisation en $\boldsymbol{\theta}$ à $\boldsymbol{u}$ fixé conduit à une valeur explicite immédiate de $\lambda = N_e/(P+1)$, où N_e est le nombre de valeurs non nulles dans la séquence de Bernoulli. Dans le cas de la loi jointe, les valeurs de $(\sigma_\epsilon^2,\sigma^2)$ s'obtiennent également de manière explicite. Nous avons vu par contre au § 4.1.3 que la dépendance en $(\sigma_\epsilon^2,\sigma^2)$ de la loi marginale prend une forme plus complexe ; la maximisation correspondante se ramène cependant à un problème d'optimisation scalaire qui peut être résolu pour un coût calculatoire limité [Champagnat *et al.*, 1996].

Une seconde approche [Lavielle, 1995, Champagnat *et al.*, 1996] consiste à considérer les variables B-G $\boldsymbol{u}$ comme des variables auxiliaires dans l'expression de la vraisemblance généralisée $p(\boldsymbol{y},\boldsymbol{u}|\boldsymbol{\theta})$ et d'appliquer un algorithme de type EM (*Expectation-Maximization*), permettant d'obtenir un jeu d'hyperparamètres maximisant la *vraisemblance marginale* $p(\boldsymbol{y}|\boldsymbol{\theta})$. La procédure se décompose alors également en l'alternance répétée de deux étapes. L'étape (E) correspond au calcul d'une espérance en les variables $\boldsymbol{u}$ qui est incalculable en pratique, mais qui peut être approximée de manière stochastique en générant des variables aléatoires distribuées selon la distribution d'intérêt . L'étape (M) de maximisation par rapport aux hyperparamètres $\boldsymbol{\theta}$ à $\boldsymbol{u}$ fixé est identique à celle décrite pour l'approche par MVG. Cette approche dite SEM (pour *Stochastic Expectation-Maximization*) provoque l'augmentation à chaque étape de la vraisemblance marginale $p(\boldsymbol{y}|\boldsymbol{\theta})$, sans garantir pour autant la convergence en $\boldsymbol{\theta}$ vers son maximum global. Notons que le choix de la loi jointe en $\boldsymbol{u} = (\boldsymbol{q},\boldsymbol{s})$ proposé par Lavielle [1995] permet une mise en œuvre de l'étape (E) à un coût plus faible que chez Champagnat *et al.* [1996] qui utilisent la loi marginale en $\boldsymbol{u} = \boldsymbol{q}$: nous montrerons en effet au § 4.3 que l'échantillonnage de la loi *jointe* est plus efficace que celui de la loi marginale. De plus, nous avons vu plus haut que l'étape de maximisation est immédiate dans le cas de la loi jointe alors qu'elle nécessite la mise en œuvre d'un algorithme d'optimisation mono-dimensionnelle pour la loi marginale. Notons que ces deux options ont fourni chez leurs auteurs des résultats satisfaisants pour l'estimation des hyperparamètres d'un modèle B-G, bien qu'au prix d'un accroissement considérable du coût calculatoire.

L'objectif visant à estimer simultanément les paramètres B-G et les hyperparamètres alourdit donc la démarche algorithmique associée à une approche de type MAP et pénalise l'utilisation d'un algorithme SMLR dont un avantage essentiel est la rapidité d'exécution. Nous verrons au § 4.3 que la mise en œuvre de méthodes d'échantillonnage stochastique, utilisée dans l'algorithme SEM pour la seule estimation des hyperparamètres, permet d'envisager de manière unifiée

l'estimation *conjointe* de l'ensemble des paramètres et hyperparamètres.

4.2.3 Optimisations alternatives

Parmi les solutions alternatives pour une approche de type MAP, nous mentionnons l'existence de méthodes stochastiques visant à pallier les limitations de l'algorithme SMLR évoquées aux deux paragraphes précédents. Lavielle [1993] propose ainsi une procédure de recuit simulé, qui permet théoriquement d'éviter le piégeage algorithmique dans des minima locaux du critère à optimiser mais qui reste également supervisée. Doucet et Duvaut [1997] proposent encore de choisir l'estimateur du maximum *a posteriori* marginal parmi des échantillons $\boldsymbol{q}^{(t)}$ distribués selon la loi $\Pr(\boldsymbol{q}|\boldsymbol{y})$: la mise en œuvre d'un algorithme d'échantillonnage stochastique permet également d'échapper aux problèmes de minima locaux et rend en outre l'extension au cas non supervisé réalisable pour un coût calculatoire additionnel négligeable. Cette méthodologie se rapproche de la démarche que nous proposons d'adopter et que nous décrivons au paragraphe suivant. Pour notre problème, néanmoins, il nous semblera plus judicieux de conjuguer des méthodes d'échantillonnage stochastique avec une estimation de la moyenne *a posteriori* de la séquence de Bernoulli.

4.3 Estimation au sens de la moyenne *a posteriori* (MP)

Toujours dans le contexte bayésien, l'estimation des paramètres Bernoulli-Gaussiens au sens de la moyenne *a posteriori* (MP) représente une alternative aux estimateurs de type MAP ou MMAP étudiés au paragraphe précédent. En reprenant les notations précédentes où $\boldsymbol{u}$ désigne l'ensemble des paramètres $(\boldsymbol{q},\boldsymbol{s})$ ou la seule séquence $\boldsymbol{q}$ selon la loi *a posteriori* considérée, la moyenne *a posteriori* est définie comme :

$$\boldsymbol{u}_{\text{MP}} \triangleq \mathbb{E}[\boldsymbol{u}|\boldsymbol{y}] = \int \boldsymbol{u}\, p(\boldsymbol{u}|\boldsymbol{y})\, d\boldsymbol{u} \quad \begin{array}{ll} = \displaystyle\sum_{\boldsymbol{q}} \boldsymbol{q} \Pr(\boldsymbol{q}|\boldsymbol{y}) \int \boldsymbol{s}\, p(\boldsymbol{s}|\boldsymbol{q},\boldsymbol{y})\, d\boldsymbol{s} & \text{pour } \boldsymbol{u} = (\boldsymbol{q},\boldsymbol{s}) \\ = \displaystyle\sum_{\boldsymbol{q}} \boldsymbol{q} \Pr(\boldsymbol{q}|\boldsymbol{y}) & \text{pour } \boldsymbol{u} = \boldsymbol{q}. \end{array} \tag{4.9}$$

Pour simplifier les expressions à venir, la dépendance des différentes distributions en les hyperparamètres n'est pas explicitée pour l'instant. On ne cherche donc plus les paramètres « les plus probables au vu des données », mais ceux minimisant l'erreur en moyenne quadratique [Tourneret, 2002] :

$$\boldsymbol{u}_{\text{MP}} = \arg\min_{\boldsymbol{u}} \mathbb{E}[\|\boldsymbol{u} - \boldsymbol{u}_{\text{th}}\|^2 |\boldsymbol{y}], \text{ où } \boldsymbol{u}_{\text{th}} \text{ est la véritable valeur des paramètres.}$$

L'estimation au sens de la moyenne *a posteriori*, alternative à une stratégie de type MAP, présente l'avantage de s'affranchir en théorie des problèmes spécifiques à la mise en œuvre d'une méthode d'optimisation, et en particulier du blocage de l'algorithme dans un maximum local de la distribution considérée. Pour des raisons déjà évoquées au paragraphe précédent, un tel estimateur n'est cependant pas calculable directement : si l'intégration des amplitudes dans le cas de la loi jointe est explicite (la variable aléatoire $\boldsymbol{s}|\boldsymbol{q}$ est gaussienne), le nombre de séquences $\boldsymbol{q}$ sur lequel s'effectue la moyenne est, dans les deux cas, de 2^{P+1} pour $P+1$ paramètres. Le recours à des méthodes d'échantillonnage stochastique est une solution visant à obtenir des échantillons $\boldsymbol{u}^{(t)}$

distribués selon la loi *a posteriori* $p(\boldsymbol{u}|\boldsymbol{y})$ et dont la moyenne réalise une approximation de l'espérance $\mathbb{E}[\boldsymbol{u}|\boldsymbol{y}]$.

Un reproche souvent effectué à un tel estimateur est son comportement envers des distributions fortement multimodales, puisque l'effet de moyennage entre deux modes peut déboucher sur un estimateur tombant dans une région de faible probabilité. Cet argument est effectivement valable lorsque les modes secondaires se voient affectés d'une densité de probabilité de valeur élevée et nous avons justement vu au premier chapitre que la vraisemblance d'un modèle multisinusoïdal est multimodale, laissant prévoir l'apparition de tels désagréments sur l'estimation de la séquence $\boldsymbol{q}$ localisant les raies spectrales. Ce comportement pathologique d'alternance entre deux modes, néanmoins, peut être décelé par une exploration plus complète de la distribution de probabilité échantillonnée, rendue possible par l'analyse des échantillons $\boldsymbol{u}^{(t)}$: nous en verrons un exemple sur l'estimation de $\boldsymbol{q}$ au § 4.5.1 consacré aux résultats de simulations. Notons encore que la procédure d'estimation des amplitudes $\boldsymbol{s}_k$, qui sera détaillée plus loin, sera réalisée quant à elle de façon à éviter le moyennage entre les deux modes $q_k = 0$ et $q_k = 1$.

4.3.1 La puissance de l'échantillonnage stochastique

Les techniques d'échantillonnage stochastique, et en particulier les méthodes de Monte-Carlo par chaînes de Markov, sont des outils très puissants, dont l'exploitation a connu un essor croissant depuis l'avènement de moyens informatiques performants [Robert, 1996]. Elles permettent de réaliser des intégrations numériques complexes telles que (4.9) en produisant des échantillons $\{\boldsymbol{u}^{(t)}\}_{t=1\ldots T}$ distribués selon la loi d'intérêt $p(\boldsymbol{u}|\boldsymbol{y})$: la moyenne *a posteriori* $\mathbb{E}[\boldsymbol{u}|\boldsymbol{y}]$ est alors estimée par :

$$\widehat{\boldsymbol{u}}_{\mathrm{MP}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{u}^{(t)}, \quad \text{avec } \boldsymbol{u}^{(t)} \sim p(\boldsymbol{u}|\boldsymbol{y}). \tag{4.10}$$

Outre le calcul de l'estimateur de la moyenne *a posteriori*, l'obtention d'un grand nombre d'échantillons distribués selon la loi $p(\boldsymbol{u}|\boldsymbol{y})$ ouvre l'accès à une quantité énorme d'informations, permettant d'analyser la forme de la loi échantillonnée. Ainsi, pour chaque paramètre scalaire u_k, l'histogramme des échantillons $\{u_k^{(t)}\}_{t=1\ldots T}$ est une approximation de la distribution *a posteriori* $p(u_k|\boldsymbol{y})$. En particulier, nous pouvons estimer la *variance* associée à l'estimateur $\widehat{u}_{k_{\mathrm{MP}}}$:

$$\mathrm{var}\,\widehat{u}_{k_{\mathrm{MP}}} = \mathbb{E}[\,(u_k - \widehat{u}_{k_{\mathrm{MP}}})^2 \,|\, \boldsymbol{y}] \quad \text{par} \quad \widehat{\mathrm{var}}\,\widehat{u}_{k_{\mathrm{MP}}} = \frac{1}{T-1}\sum_{t=1}^{T} (u_k^{(t)} - \widehat{u}_{k_{\mathrm{MP}}})^2. \tag{4.11}$$

L'estimation de la variance est un point fondamental pour le problème d'analyse spectrale considéré, puisqu'il s'interprète en termes de *niveau de confiance* associé à l'estimation, c'est-à-dire de fiabilité du résultat.

Dans le cas d'une stratégie de type MAP, notons qu'une évaluation postérieure de la qualité de l'estimation peut également être envisagée en échantillonnant la loi *a posteriori* autour de la solution $\widehat{\boldsymbol{u}}_{\mathrm{MAP}}$: cette démarche est par exemple utilisée, dans un tout autre domaine astrophysique par Kneib *et al.* [2003] pour déterminer les paramètres d'ellipticité de galaxies en leur associant des barres d'erreur. Cependant, l'utilisation d'un algorithme MCMC visant à l'estimation simultanée des paramètres et des variances correspondantes permet ici l'obtention de l'ensemble des informations par une démarche plus élégante et théoriquement solide.

Nous remarquons encore que, pour les paramètres de Bernoulli $\boldsymbol{q}$ à valeurs 0 ou 1, l'estimation de la variance n'apporte aucune information puisque l'on a toujours $\left(q_k^{(t)}\right)^2 = q_k^{(t)}\ \forall k, t$. Plus

précisément, si $\widehat{q}_{k_{\mathrm{MP}}} \in [0,1]$ est l'estimation de la probabilité que $q_k = 1$ au vu des données $\boldsymbol{y}$, on a :

$$\begin{aligned} \widehat{q}_{k_{\mathrm{MP}}} &= \frac{T_k}{T} \text{ où } T_k = \mathrm{Card}\{t \,|\, q_k^{(t)} = 1\} \\ \text{et} \quad \widehat{\mathrm{var}}\, \widehat{q}_{k_{\mathrm{MP}}} &= \frac{T_k}{T-1}(1 - \frac{T_k}{T})^2 = \frac{T}{T-1}(\widehat{q}_{k_{\mathrm{MP}}} - \widehat{q}_{k_{\mathrm{MP}}}^2). \end{aligned}$$

Par conséquent $\widehat{\mathrm{var}}\, \widehat{\boldsymbol{q}}_{\mathrm{MP}}$, comme tous les moments d'ordre supérieur de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$, représente la même information que $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$: l'incertitude associée à l'estimation de la séquence de Bernoulli $\boldsymbol{q} \in \{0;1\}^{2P+1}$ est déjà entièrement contenue dans la valeur de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$, qui n'est plus à valeurs binaires. Par ailleurs, nous proposerons plus loin une estimation *conditionnelle* des amplitudes $\boldsymbol{s}$, qui sont alors de loi *a posteriori* gaussienne. La variance associée à l'estimateur revêt dans ce cas une forme analytique explicite, rendant une fois de plus l'estimation des variances à partir des échantillons $\boldsymbol{u}^{(t)}$ superflue. Néanmoins, la relation (4.11) permet d'associer une variance à l'estimation des hyperparamètres, et sera surtout utile lors de l'extension au modèle Bernoulli-Gaussien avec décalages fréquentiels qui sera introduit au § 4.6.

4.3.2 Avantages techniques de la loi jointe

Nous avons vu lors du § 4.2 que l'estimation des paramètres B-G au sens du maximum *a posteriori* présente des caractéristiques calculatoires similaires lorsqu'elle est associée à la loi jointe $p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y})$ ou à la loi marginale $Pr(\boldsymbol{q}|\boldsymbol{y})$. Les estimateurs obtenus dans les deux cas sont toutefois différents et l'optimisation par un algorithme de type SMLR fournit empiriquement des résultats plus satisfaisants avec la loi marginale. Qu'en est-il pour l'approche MP, visant à réaliser l'échantillonnage stochastique d'une loi *a posteriori* ? Nous détaillons notre réponse en les quatre points suivants.

i) Les paramètres d'intérêt sont les valeurs de $\boldsymbol{q}$. L'estimation de la séquence de Bernoulli est en effet le point le plus délicat et, sous les hypothèses gaussiennes réalisées sur la loi des perturbations $\boldsymbol{\epsilon}$ et les amplitudes $\boldsymbol{s}|\boldsymbol{q}$, les amplitudes peuvent être estimées *a posteriori* selon la relation (4.8). L'importance essentielle de l'estimation de $\boldsymbol{q}$ est *a fortiori* renforcée lorsque son estimation est réalisée au sens de la MP par des méthodes d'échantillonnage stochastique : les échantillons $\{\boldsymbol{q}^{(t)}\}$ permettent alors d'extraire de nombreuses informations au niveau de l'interprétation des résultats.

ii) Les estimateurs $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ obtenus à partir de $p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y})$ et $\mathrm{Pr}(\boldsymbol{q}|\boldsymbol{y})$ sont identiques, ce qui n'est pas le cas pour une approche de type MAP. Cette propriété s'obtient immédiatement en remarquant que si $(\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)}) \sim p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y})$, alors $\boldsymbol{q}^{(t)} \sim \mathrm{Pr}(\boldsymbol{q}|\boldsymbol{y})$. Le choix entre lois jointe et marginale sera donc essentiellement lié à la complexité associée à l'échantillonnage de chaque loi.

iii) La dépendance en les hyperparamètres est plus simple pour la loi jointe. Nous avons effectivement vu au S 4.1.3 que l'étape de marginalisation des amplitudes masque la relation des variances $(\sigma_\epsilon^2,\sigma^2)$ dans la formulation de la loi marginale. Cet argument pénalise l'utilisation de la loi marginale dans le but d'aborder l'estimation non supervisée dans un cadre totalement bayésien, comme nous l'envisagerons au § 4.3.4.

iv) Mais surtout, l'échantillonnage de la loi jointe *a posteriori* est plus efficace. Bien que le nombre de paramètres à échantillonner soit plus élevé – il faut ajouter l'échantillonnage des $2P+2$ amplitudes $\boldsymbol{s}$ dans le cas de la loi jointe – l'échantillonnage conjoint en $(\boldsymbol{q},\boldsymbol{s})$ est plus attractif sur le plan calculatoire. Cette proposition est à rapprocher de l'argumentaire de Davy et Idier [2004] affirmant, dans un contexte plus général, que l'échantillonnage de la loi jointe peut prévaloir sur celui de la loi marginale en fonction de la structure du problème considéré : même si l'estimation basée sur la loi marginale, du fait d'une quantité réduite de paramètres, nécessite un nombre plus faible d'échantillons, l'obtention d'échantillons distribués selon la loi jointe peut se révéler bien plus efficace, si bien qu'au final la procédure d'estimation se montre plus avantageuse en exploitant la structure de la seconde. Afin de valider ce dernier point, nous notons ci-après les expressions des lois $p(q_k,\boldsymbol{s}_k|\boldsymbol{q}_{-k},\boldsymbol{s}_{-k},\boldsymbol{y})$ et $\Pr(q_k|\boldsymbol{q}_{-k},\boldsymbol{y})$ qui sont à la base de l'*échantillonnage de Gibbs* des lois *a posteriori* jointe et marginale, respectivement. La notation employée $\boldsymbol{u}_{-k}$ désigne le vecteur formé par toutes les composantes de $\boldsymbol{u}$ sauf la $k^{\text{ème}}$. La justification de l'usage d'un algorithme d'échantillonnage de Gibbs, ainsi que les détails de sa mise en œuvre, feront l'objet du § 4.4.1. Dans les expressions suivantes, $\mathbf{H}_k$ représente la matrice formée par les deux colonnes d'indices $2k$ et $2k+1$ de $\mathbf{H}$, c'est-à-dire la matrice $N \times 2$ formée par les N lignes $[\cos(2\pi f_k t_n),\, \sin(2\pi f_k t_n)]_{n=1\ldots N}$.

Lois conditionnelles tirées de la distribution a posteriori jointe : l'expression de la distribution $p(q_k,\boldsymbol{s}_k|\boldsymbol{q}_{-k},\boldsymbol{s}_{-k},\boldsymbol{y})$ s'obtient à partir de la loi jointe (4.6), en remplaçant les distributions de Dirac $\delta_2(\boldsymbol{s}_k)$ par des lois gaussiennes $\mathcal{N}(0,\sigma_0^2\mathbf{I}_2)$ dont on fait tendre la variance vers 0 : le détail des calculs peut être trouvé à l'annexe B.2, et constituent une adaptation de ceux menés par Cheng *et al.* [1996] au cas de paramètres $\boldsymbol{s}_k$ bidimensionnels. On montre que les variables $(q_k,\boldsymbol{s}_k|\boldsymbol{q}_{-k},\boldsymbol{s}_{-k},\boldsymbol{y})$ ont également une structure Bernoulli-Gaussienne :

- $\Pr(q_k=1|\boldsymbol{q}_{-k},\boldsymbol{s}_{-k},\boldsymbol{y}) = \dfrac{\widetilde{\lambda}_k}{\widetilde{\lambda}_k+(1-\lambda)}$ avec $\widetilde{\lambda}_k = \dfrac{\lambda}{\sigma^2}\,|\boldsymbol{\Gamma}_k|^{1/2}\,\exp\{\dfrac{1}{2}\boldsymbol{v}_k^{\mathrm{t}}\boldsymbol{\Gamma}_k^{-1}\boldsymbol{v}_k\}$

  $$\text{où}\ \left\{\begin{array}{rcl}\boldsymbol{\Gamma}_k^{-1} &=& \frac{1}{\sigma^2}\,\mathbf{I}_2 + \frac{1}{\sigma_\epsilon^2}\,\mathbf{H}_k^{\mathrm{t}}\,\mathbf{H}_k \\ \boldsymbol{v}_k &=& \frac{1}{\sigma_\epsilon^2}\,\boldsymbol{\Gamma}_k\,\mathbf{H}_k^{\mathrm{t}}\,\boldsymbol{e}_k,\ \text{avec}\ \boldsymbol{e}_k = \boldsymbol{y} - \sum_{j\neq k}\mathbf{H}_j\boldsymbol{s}_j\end{array}\right.$$

- $p(\boldsymbol{s}_k|q_k=1,\boldsymbol{q}_{-k},\boldsymbol{s}_{-k},\boldsymbol{y}) = g_2(\boldsymbol{s}_k-\boldsymbol{v}_k,\boldsymbol{\Gamma}_k)$ et $p(\boldsymbol{s}_k|q_k=0,\boldsymbol{q}_{-k},\boldsymbol{s}_{-k},\boldsymbol{y}) = \delta_2(\boldsymbol{s}_k)$.

L'échantillonnage de chaque couple $(\boldsymbol{q}_k,\boldsymbol{s}_k)$ correspond donc à l'échantillonnage d'une variable aléatoire de Bernoulli et éventuellement celui d'un vecteur aléatoire gaussien de dimension 2. Il requiert donc un effort calculatoire limité – la matrice $\boldsymbol{\Gamma}_k$ est de dimension 2 – qui consiste essentiellement en :

- $2N$ multiplications scalaires et autant d'additions pour le produit $\mathbf{H}_k^{\mathrm{t}}\boldsymbol{e}_k$,
- $2N$ additions scalaires pour la mise à jour séquentielle du terme $\boldsymbol{e}_k$.

Lois conditionnelles tirées de la distribution a posteriori marginale : la loi marginale $\Pr(q_k|\boldsymbol{q}_{-k},\boldsymbol{y})$ se déduit à partir de l'expression (4.7) et l'écriture donnée ci-dessous est sa forme de complexité minimale, utilisant des réductions analogues à celles proposées par Kormylo et Mendel [1982] pour l'algorithme SMLR, bien que l'aspect bidimensionnel des variables $\boldsymbol{s}_k$ com-

plique quelque peu leur formulation. Les calculs correspondants sont explicités à l'annexe B.2.

$$\Pr(q_k = 1|\boldsymbol{q}_{-k},\boldsymbol{y}) = \frac{\breve{\lambda}_k}{\breve{\lambda}_k + (1-\lambda)} \text{ avec } \breve{\lambda}_k = \lambda\,|\boldsymbol{\Lambda}_k|^{1/2}\,\exp\{\frac{1}{2}\sigma^2 \boldsymbol{w}_k^{\mathrm{t}}\boldsymbol{\Lambda}_k^{-1}\boldsymbol{w}_k\}$$

$$\text{où } \left\{ \begin{array}{rcl} \boldsymbol{\Sigma}_{\boldsymbol{y}|\boldsymbol{q}^1_{-k}} & = & \sigma_\epsilon^2\,\mathbf{I}_N + \sigma^2\,\mathbf{H}\,\mathrm{diag}(\underline{\underline{\boldsymbol{q}}}^1_{-k})\,\mathbf{H}^{\mathrm{t}} \\ \boldsymbol{w}_k & = & \mathbf{H}_k^{\mathrm{t}}\,\boldsymbol{\Sigma}^{-1}_{\boldsymbol{y}|\boldsymbol{q}^1_{-k}}\,\boldsymbol{y} \\ \boldsymbol{\Lambda}_k & = & \mathbf{I}_2 - \sigma^2\,\mathbf{H}_k^{\mathrm{t}}\,\boldsymbol{\Sigma}^{-1}_{\boldsymbol{y}|\boldsymbol{q}^1_{-k}}\,\mathbf{H}_k \end{array} \right.$$

où $\boldsymbol{q}^1_{-k}$ correspond à la séquence de Bernoulli valant 1 en k et $\boldsymbol{q}_{-k}$ ailleurs et, comme précédemment, la notation $\underline{\underline{\boldsymbol{q}}}^1_{-k}$ correspond à la séquence $\boldsymbol{q}^1_{-k}$ « doublée » (voir en page 96). La marginalisation des amplitudes se manifeste sous la forme de la covariance *a posteriori* $\boldsymbol{\Sigma}_{\boldsymbol{y}|\boldsymbol{q}^1_{-k}}$, dont l'expression complexe pénalise considérablement le coût de calcul des probabilités *a posteriori* $\Pr(q_k|\boldsymbol{q}_{-k},\boldsymbol{y})$. Chaque évaluation de cette quantité nécessite essentiellement :

- le calcul de la matrice $\boldsymbol{\Sigma}^{-1}_{\boldsymbol{y}|\boldsymbol{q}^1_{-k}}$ de dimension N et le calcul de son déterminant.
- plusieurs produits matrice - vecteur en dimension N pour le calcul de $\boldsymbol{w}_k$ et Λ_k, soit un total d'environ $5N$ multiplications et autant d'additions.

Notons que l'évaluation de la matrice $\boldsymbol{\Sigma}_{\boldsymbol{y}|\boldsymbol{q}^1_{-k}}$ n'est pas nécessaire, et que le calcul de son inverse peut se ramener à l'inversion d'une matrice de taille $2N_e$ en exploitant la structure creuse de la séquence $\boldsymbol{q}^1_{-k}$ [Champagnat *et al.*, 1996], où $2N_e$ est le nombre de valeurs non nulles dans $\underline{\underline{\boldsymbol{q}}}^1_{-k}$. De fait, l'échantillonnage de $\Pr(q_k|\boldsymbol{q}_{-k},\boldsymbol{y})$ pour $k = 0 \ldots P$, présente des aspects similaires au balayage de séquences voisines de l'algorithme d'optimisation SMLR (voir le § 4.2.1), et les diverses astuces associées à cette procédure peuvent être employées, permettant de réduire le coût de calcul et le stockage en mémoire. Néanmoins, l'échantillonnage de la loi marginale restera toujours de complexité supérieure à celle de l'échantillonnage de la distribution jointe $p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y})$. En dehors du champ de l'analyse spectrale, notons encore que l'échantillonnage de la loi marginale peut se révéler avantageux pour des problèmes de déconvolution : lorsque la réponse impulsionnelle du filtre correspondant – éventuellement inconnue – est d'ordre réduit M, le schéma d'échantillonnage proposé par Cappé *et al.* [1999] exploite la structure creuse de l'opérateur associé $\mathbf{H}$ pour n'inverser que des matrices de taille M.

Enfin, un dernier argument renforçant le choix de la seconde concerne les hyperparamètres, dont les lois *a posteriori* prennent des formes plus simples lorsqu'elles sont conjuguées à la distribution jointe. Le problème d'estimation peut alors être abordé dans un cadre totalement bayésien : ce point sera abordé au § 4.3.4.

4.3.3 L'intérêt de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$

Limitations de l'estimation jointe $(\widehat{\boldsymbol{q}},\widehat{\boldsymbol{s}})_{\mathrm{MP}}$

À partir des échantillons $\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)} \sim p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y})$, la stratégie d'estimation directe de $(\boldsymbol{q},\boldsymbol{s})$ conjointement au sens de la moyenne *a posteriori* est une solution élégante et rigoureuse qui fournit des estimateurs aux caractéristiques bien connues, avec en particulier l'accès à leur variance. L'obtention de tels estimateurs à partir des échantillons de la loi *a posteriori* est de plus immédiate : c'est une simple moyenne. C'est la solution retenue par Cheng *et al.* [1996] pour un problème de déconvolution myope, qui s'intéresse directement à l'estimation des amplitudes $\widehat{\boldsymbol{s}}_{\mathrm{MP}}$. Ce choix,

s'il est solide au niveau théorique, peut néanmoins se révéler problématique puisqu'il effectue, pour chaque vecteur (bidimensionnel) d'amplitude $\boldsymbol{s}_k$ un moyennage entre les deux modes $q_k = 0$ et $q_k = 1$ de la distribution *a posteriori* :

$$\mathbb{E}[\boldsymbol{s}_k|\boldsymbol{y}] = \Pr(q_k = 0|\boldsymbol{y})\,\mathbb{E}[\boldsymbol{s}_k|q_k = 0, \boldsymbol{y}] + \Pr(q_k = 1|\boldsymbol{y})\,\mathbb{E}[\boldsymbol{s}_k|q_k = 1, \boldsymbol{y}],$$

mélangeant ainsi des amplitudes nulles et des amplitudes de loi *a posteriori* $p(\boldsymbol{s}_k|q_k = 1,\boldsymbol{y})$ gaussienne. Si la distribution *a posteriori* marginale $\Pr(q_k|\boldsymbol{y})$ n'est pas assez concentrée en 1 ou en 0, les amplitudes $\boldsymbol{s}_{k_{\mathrm{MP}}}$ correspondant à la présence d'une raie à la fréquence f_k seront donc sous-estimées. À l'inverse, le caractère parcimonieux de $\boldsymbol{s}_{k_{\mathrm{MP}}}$ ne sera plus rigoureusement vérifié, puisque l'on n'a jamais exactement $Pr(q_k = 1|\boldsymbol{y}) = 0$.

Estimation séquentielle

Une démarche séquentielle semble alors plus adaptée, visant à réaliser en premier lieu la localisation des raies spectrales, c'est-à-dire la *détection* de la séquence de Bernoulli, puis postérieurement l'*estimation* des amplitudes correspondant aux fréquences sélectionnées. C'est ce que proposent Doucet et Duvaut [1997] avec un schéma en deux étapes :

1. l'estimation préalable de la séquence de Bernoulli au sens du maximum *a posteriori* marginal en sélectionnant l'échantillon $\boldsymbol{q}^{(t)}$ qui maximise la distribution $\Pr(\boldsymbol{q}|\boldsymbol{y})$;
2. l'estimation des amplitudes maximisant $p(\boldsymbol{s}|\boldsymbol{q}, \boldsymbol{y})$, donnée par l'expression (4.8).

Notons que la première étape nécessite l'évaluation supplémentaire de la loi $\Pr(\boldsymbol{q}^{(t)}|\boldsymbol{y})$ en chaque échantillon $\boldsymbol{q}^{(t)}$, et que la maximisation par l'échantillonnage de l'ensemble de cette loi ne semble pas optimale en termes de coût calculatoire. En explorant théoriquement l'ensemble de la loi, elle permet néanmoins de s'affranchir des problèmes spécifiques des approches « orientées optimisation », comme l'existence de minima locaux du critère à optimiser pénalisant des approches itératives de type SMLR[4]. Mais surtout, cette procédure souffre, en ce qui concerne l'application envisagée à la détection de raies spectrales, du reproche commun à tous les estimateurs de type MAP, qui estiment $\boldsymbol{q}$ par une véritable séquence de Bernoulli. En effet, l'obtention de $\widehat{q}_{k_{\mathrm{MP}}}$ permet d'associer le résultat à un degré de confiance accordé sur la détection éventuelle d'une raie spectrale :

> *$\widehat{q}_{k_{MP}}$ est une estimation de la probabilité d'existence d'une sinusoïde à la fréquence f_k au vu des observations $\boldsymbol{y}$.*

La structure du modèle Bernoulli-Gaussien, couplée à une estimation au sens de la moyenne *a posteriori* (4.10), permet donc d'obtenir avec $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ une estimation de la probabilité d'existence d'une raie en tout point de la grille fréquentielle. Un intérêt majeur de l'estimation au sens de la moyenne *a posteriori* par rapport au maximum *a posteriori* apparaît ici : le MAP, que ce soit pour la loi jointe $p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y})$ ou la loi marginale $\Pr(\boldsymbol{q}|\boldsymbol{y})$, fournit un estimateur $\widehat{\boldsymbol{q}}_{\mathrm{MAP}}$ qui est lui-même une séquence de Bernoulli, à valeurs 0 ou 1. $\widehat{\boldsymbol{q}}_{\mathrm{MAP}}$ est certes la séquence « la plus probable », mais l'estimation n'apporte aucune information au sujet de la cohérence du résultat. En d'autres termes, l'échantillonnage de la distribution *a posteriori* $p(\boldsymbol{q}, \boldsymbol{s}|\boldsymbol{y})$ permet d'accéder à une quantité d'informations pertinentes dont la richesse n'est pas suffisamment exploitée par la seule extraction du maximum. Cette interprétation est cependant un point clé de la procédure

4. Nous verrons toutefois au § 4.4.1 que l'algorithme MCMC peut être perturbé par l'attraction de la chaîne vers des modes locaux, ralentissant sa convergence vers la distribution échantillonnée.

d'estimation que nous proposons dans le contexte de la détection de raies spectrales, dont le schéma est décrit au tableau 4.1.

1. l'estimation de $\mathbb{E}[\boldsymbol{q}|\boldsymbol{y}]$ par la moyenne $\widehat{\boldsymbol{q}}_{\text{MP}} = \frac{1}{T}\sum_t \boldsymbol{q}^{(t)}$ ainsi que celle des hyperparamètres $\widehat{\boldsymbol{\theta}}_{\text{MP}} = \frac{1}{T}\sum_t \boldsymbol{\theta}^{(t)}$;
2. l'interprétation du résultat en termes de *probabilités de détection*, et la sélection des fréquences significatives par *seuillage* des valeurs de $\widehat{\boldsymbol{q}}_{\text{MP}}$, produisant une séquence de Bernoulli $\widehat{\boldsymbol{q}}_\alpha$: $\widehat{\boldsymbol{q}}_{\alpha_k} = 1$ si $\widehat{q}_{\text{MP}k} \geq \alpha$ et 0 sinon. Le niveau de seuil α est laissé au libre choix de l'expert ;
3. l'estimation des amplitudes correspondantes, au sens du maximum (ou de manière équivalente de la moyenne) de la loi gaussienne $p(\boldsymbol{s}|\boldsymbol{q} = \widehat{\boldsymbol{q}}_\alpha,\boldsymbol{y})$, obtenu par l'équation (4.8), rappelée ici :

$$\widehat{\boldsymbol{s}}_\alpha = \mathbb{E}[\boldsymbol{s} \,|\, \widehat{\boldsymbol{q}}_\alpha, \boldsymbol{y}] = \boldsymbol{\Sigma}_{s|\widehat{q}_\alpha}\, \mathbf{H}^{\text{t}}\, \boldsymbol{\Sigma}_{y|\widehat{q}_\alpha}{}^{-1}\, \boldsymbol{y}$$

où $\boldsymbol{\Sigma}_{s|\widehat{q}_\alpha} = \sigma^2 \text{diag}(\underline{\widehat{\boldsymbol{q}}}_\alpha)$ et $\boldsymbol{\Sigma}_{y|\widehat{q}_\alpha} = \sigma_\epsilon^2 \mathbf{I}_N + \mathbf{H}\,\boldsymbol{\Sigma}_{s|\widehat{q}_\alpha}\,\mathbf{H}^{\text{t}}$ et en utilisant les valeurs estimées à l'étape 1 pour les hyperparamètres σ^2 et σ_ϵ^2. L'expression analytique de la matrice de covariance associée est alors donnée par [Idier *et al.*, 2004] :

$$\begin{aligned} \text{Cov}\,[\boldsymbol{s} \,|\, \widehat{q}_\alpha,\boldsymbol{y}] &= \mathbb{E}[(\boldsymbol{s} - \widehat{\boldsymbol{s}}_\alpha)(\boldsymbol{s} - \widehat{\boldsymbol{s}}_\alpha)^{\text{t}} |\, \widehat{\boldsymbol{q}}_\alpha, \boldsymbol{y}] \\ &= \boldsymbol{\Sigma}_{s|\widehat{q}_\alpha} - \boldsymbol{\Sigma}_{s|\widehat{q}_\alpha}\,\mathbf{H}^{\text{t}}\,\boldsymbol{\Sigma}_{y|\widehat{q}_\alpha}^{-1}\,\mathbf{H}\,\boldsymbol{\Sigma}_{s|\widehat{q}_\alpha}. \end{aligned}$$

En particulier, les éléments diagonaux de cette matrice sont les variances $\text{var}\,\widehat{a}_k$, $\text{var}\,\widehat{b}_k$ où $\widehat{\boldsymbol{s}}_k = [\widehat{a}_k, \widehat{b}_k]$.

TAB. 4.1 – *Procédure d'estimation de* $(\boldsymbol{q},\boldsymbol{s})$ *à partir de l'échantillonnage de la loi a posteriori* $p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y})$.

En ce qui concerne l'estimation des amplitudes, notons encore que les paramètres d'intérêt sont essentiellement les modules $\rho_k = \sqrt{a_k^2 + b_k^2}$ associés à chaque sinusoïde, que nous pouvons également estimer au sens de la MP à partir des échantillons $a_k^{(t)}, b_k^{(t)}$, ainsi que la variance associée. Nous préférons cependant utiliser les expressions théoriques fournies par Goodman [1985] où, si $\widehat{a}_k \sim \mathcal{N}(m_a,v)$ et $\widehat{b}_k \sim \mathcal{N}(m_b,v)$, la moyenne et la variance de ρ_k ont les expressions analytiques données à l'annexe B.1. Ce choix, bien qu'approximatif puisqu'il suppose que les variables *a posteriori* $a_k|\boldsymbol{y}$ et $b_k|\boldsymbol{y}$ sont indépendantes et de même variance, prendra tout son sens lors de l'estimation des paramètres du modèle Bernoulli-Gaussien avec décalages en fréquence au § 4.6 pour lequel nous ne pourrons plus estimer le module à partir des échantillons $a_k^{(t)}, b_k^{(t)}$. Nous n'avons pas trouvé ni pu établir d'expression analytique exacte des moments de $\widehat{\rho}$ dans le cas général d'un couple $\boldsymbol{s}_k = [\widehat{a}_k, \widehat{b}_k]$ gaussien et de matrice de covariance non proportionnelle à l'identité. En pratique cependant, nous avons toujours obtenu des valeurs $\text{Cov}\,[\widehat{\boldsymbol{s}}_k|\boldsymbol{y}] \simeq v\,\mathbf{I}_2$.

4.3.4 Échantillonnage des hyperparamètres

Nous avons considéré tout au long de ce paragraphe une dépendance muette des différentes lois de probabilité étudiées envers les hyperparamètres $\boldsymbol{\theta} = (\lambda,\sigma_\epsilon^2,\sigma^2)$. En particulier, la distribution

sur laquelle nous proposons de baser l'estimation MP s'écrit sous forme complète :

$$p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{y},\boldsymbol{\theta}) \propto p(\boldsymbol{y}|\boldsymbol{q},\boldsymbol{s},\sigma_\epsilon^2)\, p(\boldsymbol{s}|\boldsymbol{y},\boldsymbol{q},\sigma^2)\, \Pr(\boldsymbol{q}|\boldsymbol{y},\lambda).$$

Le paradigme bayésien dans lequel nous nous plaçons est un cadre idéal pour considérer, au moins sur le plan théorique, une estimation *non supervisée* totalement bayésienne en se munissant d'une description probabiliste des hyperparamètres sous la forme d'une distribution $p(\boldsymbol{\theta})$. Il est alors possible de réaliser l'estimation à partir de la loi *a posteriori* conjointe en $(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta})$, qui s'écrit par application de la règle de Bayes :

$$p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta})\, p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\mathcal{I}} \propto p(\boldsymbol{y}|\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta})\, p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}) \tag{4.12}$$

où $\mathcal{I}$ est la constante de normalisation intégrant le terme de droite sur l'ensemble des paramètres aléatoires $\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}$. L'utilisation de distributions $p(\boldsymbol{\theta})$ *non informatives*, c'est-à-dire ne nécessitant aucune connaissance *a priori* sur les valeurs de $\boldsymbol{\theta}$, permet d'envisager une estimation non supervisée.

La *maximisation conjointe* de la distribution (4.12) en les paramètres et hyperparamètres est théoriquement possible, mais bute sur les mêmes problèmes que l'estimation au sens du maximum de vraisemblance généralisée décrite au § 4.2.2 (les deux approches coïncident d'ailleurs lorsque les lois *a priori* $p(\boldsymbol{\theta})$ sont choisies uniformes) : la maximisation est en elle-même une tâche délicate, et les propriétés de l'estimateur résultant sont mal connues [Idier, 2001a, chapitre 8]. Une stratégie de type moyenne *a posteriori* permet par contre d'inclure les hyperparamètres dans la procédure d'estimation en se dotant de lois *a priori* adéquates garantissant la simplicité de l'échantillonnage de la loi jointe $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y})$. L'estimation de l'ensemble des inconnues au sens de la MP est alors envisageable, la moyenne :

$$\mathbb{E}[\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y}] = \int_{\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}} (\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta})\, p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta})$$

étant estimée par $(\widehat{\boldsymbol{q}},\widehat{\boldsymbol{s}},\widehat{\boldsymbol{\theta}})_{\mathrm{MP}} = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T}(\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)},\boldsymbol{\theta}^{(t)})$, où $(\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)},\boldsymbol{\theta}^{(t)}) \sim p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y})$. Cette approche totalement bayésienne se révèle attrayante, offrant un cadre homogène pour lequel les paramètres B-G et les hyperparamètres sont considérés sur le même plan, aussi bien par la nature conjointe de la loi que par la méthode d'estimation. Une réserve est cependant à émettre au sujet de l'estimation des amplitudes $\boldsymbol{s}$: comme nous l'avons vu plus haut, il sera en effet préférable d'estimer $\boldsymbol{s}$ postérieurement à l'estimation de $\boldsymbol{q}$ afin d'éviter les effets de moyennage entre les deux modes $q_k = 0$ et $q_k = 1$. La formule d'estimation conjointe ci-dessus sera donc en réalité utilisée pour les paramètres $\boldsymbol{q}$ et $\boldsymbol{\theta}$.

Nous examinons donc l'utilisation de lois *a priori non informatives* sur les hyperparamètres $\boldsymbol{\theta}$, c'est-à-dire ne traduisant pas ou peu de contraintes sur les valeurs de $\boldsymbol{\theta}$ et, dans la mesure du possible, ne nécessitant pas elles-mêmes le réglage de paramètres supplémentaires. Une fois de plus, notons que l'objectif de cette démarche est de stabiliser la procédure d'estimation plutôt que d'influer sur son résultat. Nous optons pour une structure réglant indépendamment les trois hyperparamètres :

$$p(\boldsymbol{\theta}) = p(\lambda)\, p(\sigma_\epsilon^2)\, p(\sigma^2),$$

bien qu'il soit également possible d'envisager une paramétrisation différente. L'utilisation d'un jeu d'hyperparamètres équivalent $(\lambda,\sigma_\epsilon^2,\delta^2)$ où $\delta^2 = \sigma^2/\sigma_\epsilon^2$ utilisé par Andrieu et Doucet [1999] permet par exemple de réaliser l'inférence sur δ^2 plutôt que sur σ^2, le paramètre δ^2 ayant une

interprétation physique plus « parlante » en termes de rapport signal sur bruit [5]. Néanmoins, la structure proposée nous semble la plus appropriée, pour son caractère naturel par rapport à la modélisation et la simplicité analytique des lois *a posteriori* qui en résultent.

Une loi *a priori* uniforme est-elle vraiment non informative?

Le choix de lois uniformes peut être perçu comme non informatif, dans le sens où celles-ci ne privilégient aucune valeur *a priori* de l'hyperparamètre. Ces distributions sont souvent impropres, c'est-à-dire qu'elles ne sont pas normalisables : ce sera le cas de tout paramètre prenant ses valeurs dans un domaine non borné, comme ici les variances σ_ϵ^2 et σ^2. Néanmoins, l'utilisation de lois *a priori* impropres est justifiable sous la condition que la loi jointe en $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y})$ soit propre [Robert, 1996], c'est-à-dire lorsque la constante de normalisation $\mathcal{I}$ intervenant dans la relation (4.12) converge, où :

$$\mathcal{I} = \sum_{\boldsymbol{q}} \int p(\boldsymbol{y}|\boldsymbol{q},\boldsymbol{s},\sigma_\epsilon^2)\, p(\boldsymbol{s}|\boldsymbol{q},\sigma^2)\ \Pr(\boldsymbol{q}|\lambda)\, p(\lambda)\, p(\sigma_\epsilon^2)\, p(\sigma^2)\, d\boldsymbol{s}\, d\lambda\, d\sigma_\epsilon^2\, d\sigma^2.$$

Un reproche plus dommageable aux lois uniformes en termes informationnels concerne leur forte dépendance envers la paramétrisation choisie. Ainsi, le choix d'une loi $p(\sigma^2)$ uniforme sur l'axe réel positif, impropre, correspond à une loi en σ, également impropre, de la forme $p(\sigma) = \sigma$, favorisant *a priori* les amplitudes à l'infini ! À l'inverse, le choix de $\sigma \sim \mathcal{U}_{[0,+\infty[}$ revient à considérer que la variance σ^2 est distribuée selon une loi (impropre) de la forme $p(\sigma^2) \propto 1/\sqrt{\sigma}$. Une loi non informative en σ^2 peut donc être très informative en σ! Il est donc illusoire de rechercher sous cette forme une loi totalement non informative puisque l'on disposera toujours d'une interprétation permettant de lui associer une information *a priori*. Les motivations poussant au choix d'un *a priori* impropre correspondent donc plus à des arguments de nature technique (simplicité de la loi *a posteriori* résultante et absence de réglage de paramètres supplémentaires) qu'à un choix naturel pour l'inférence statistique.

La loi *a priori* de Jeffreys

Lors de la sélection d'une loi *a priori* en l'absence d'information préalable, une approche bayésienne traditionnelle est celle utilisant la loi non informative de Jeffreys, qui s'écrit comme la racine de l'information de Fisher du modèle [Bernardo et Smith, 1994] :

$$p_{\text{F}}(\boldsymbol{\theta}) \propto \sqrt{I(\boldsymbol{\theta})}, \text{ où } I(\boldsymbol{\theta}|\boldsymbol{y}) = -\mathbb{E}[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log p(\boldsymbol{q},\boldsymbol{s}|\boldsymbol{\theta})].$$

Par rapport à l'*a priori* uniforme, cette distribution présente l'avantage d'*invariance* selon la paramétrisation choisie parmi la famille des puissances $\{\theta^m\}_{m>0}$ représentant un paramètre θ. En reprenant l'exemple de la variance σ^2 d'une loi gaussienne, l'*a priori* selon Jeffreys revient alors à considérer $p(\sigma) \propto 1/\sigma$, ou de manière équivalente $p(\sigma^m) \propto 1/\sigma^m$: ainsi la loi *a priori* contient la même information selon que l'on considère σ ou σ^m. Notons que dans ce cas, l'*a priori* de Jeffreys s'identifie avec la définition de la loi non informative *de référence*, issue de la théorie de l'information [Bernardo et Smith, 1994]. Remarquons encore que cette distribution

5. Les motivations poussant les auteurs à utiliser une telle paramétrisation ont cependant plus à voir avec un aspect technique lié à la marginalisation des amplitudes dans la loi *a posteriori* : nous avons étudié ce point au § 4.1.3.

peut également être impropre (c'est le cas pour la variance d'une loi gaussienne), ce qui n'est pas un problème dans la mesure où la loi *a posteriori* associée est intégrable.

Lois *a priori* conjuguées

Une autre option, dont la justification est plutôt d'ordre technique, consiste à utiliser une loi *a priori conjuguée*, c'est-à-dire d'une forme telle que la loi *a posteriori* correspondante appartienne à la même famille de fonctions, l'étape d'intégration des données n'intervenant qu'au niveau de la valeur des paramètres régissant cette loi. Lorsque les familles considérées sont de nature simple, cette propriété se révèle en particulier de grande importance pour l'utilisation de méthodes d'échantillonnage stochastique, garantissant ainsi l'obtention de distributions *a posteriori* selon lesquelles il sera possible de générer facilement des échantillons. Dans un problème de déconvolution aveugle B-G utilisant ce même formalisme, Cheng *et al.* [1996] proposent ainsi d'utiliser des lois *a priori* conjuguées de type Beta pour λ : $\lambda \sim \mathcal{B}e(a,b)$ et inverse-chi deux pour les variances des lois gaussiennes : $\sigma_\epsilon^2 \sim \chi^{-2}(\nu,\eta)$ et $\sigma^2 \sim \chi^{-2}(\mu,\xi)$, nécessitant le réglage de six paramètres. Doucet et Duvaut [1997] utilisent quant à eux des lois Beta sur λ et inverse-Gamma sur σ_ϵ^2 et σ^2, également conjuguées avec la loi *a posteriori* (4.6). Il est ainsi théoriquement possible de rajouter un nombre indéterminé de niveaux d'inférence statistique, où les *hyper-hyperparamètres* réglant les distributions $p(\boldsymbol{\theta})$ doivent être fixés, voire probabilisés à leur tour puis estimés conjointement dans le cadre bayésien. En choisissant pour chaque niveau des distributions *a priori* de plus en plus vagues, la procédure autorise ainsi des variations peu contraintes des hyperparamètres, satisfaisant l'objectif non supervisé (Richardson et Green [1997], Andrieu et Doucet [1999] remontent ainsi à trois niveaux d'inférence). Néanmoins, par souci de simplicité, nous privilégions la recherche de lois $p(\boldsymbol{\theta})$ de description plus directe et s'affranchissant dans la mesure du possible du réglage de paramètres supplémentaires. Nous verrons que, en application à notre problème, un second niveau d'inférence sur $p(\boldsymbol{\theta})$ s'avérera suffisant pour obtenir une estimation non supervisée satisfaisante.

Remarquons que dans le cas où l'on disposerait d'une quelconque information préalable sur tout ou partie de $\boldsymbol{\theta}$, sa prise en compte peut être effectuée de manière probabiliste en contraignant plus ou moins les distributions correspondantes au moyen du réglage des hyper-hyperparamètres, ce qui reste une démarche plus souple que de fixer de manière « définitive » de la valeur d'un hyperparamètre, en la rendant plus permissive vis-à-vis d'éventuelles erreurs.

Choix adoptés

Nous détaillons ci-dessous les types de distributions *a priori* retenues pour les trois hyperparamètres. Une étude comparative de leur influence sur la qualité de l'estimation résultante sera présentée au § 4.5.

Choix de $p(\lambda)$: La dépendance de la loi *a posteriori* (4.6) en λ est de la forme $\lambda^{N_e}(1-\lambda)^{P+1-N_e}$, où N_e représente le nombre de raies dans le modèle. Une loi *a priori* de type $\mathcal{B}e$ est alors conjuguée à (4.6) et il est possible de contraindre λ à prendre des valeurs faibles (puisque nous recherchons une solution parcimonieuse, on aura $\lambda \ll 1$). Cependant, nous préférons éviter de régler des paramètres supplémentaires et choisissons une loi *a priori* uniforme sur $[0,1]$, choix qui s'avérera satisfaisant en pratique :

$$\lambda \sim \mathcal{U}_{[0,1]} \ : \ p(\lambda) = \mathbb{1}_{[0,1]}(\lambda).$$

Choix de $p(\sigma_\epsilon^2)$: Plusieurs stratégies sont envisageables pour choisir une loi $p(\sigma_\epsilon^2)$ non informative. La possibilité de prise en compte d'une distribution *a priori* uniforme de σ_ϵ^2 sur $[0, +\infty[$ est reliée aux considérations évoquées plus haut dans ce même paragraphe : la loi jointe sera proprement définie si l'intégrale $\int \sigma_\epsilon^{-N} \exp -\frac{\|\boldsymbol{y}-\mathbf{H}\boldsymbol{s}\|^2}{2\sigma_\epsilon^2} d\sigma_\epsilon^2$ converge, ce qui est garanti à partir du moment où $N/2 > 1$, soit un nombre de données $N \geq 3$, ce qui sera toujours le cas en pratique. Par ailleurs, le terme sous l'intégrale précédente décrit la dépendance de la loi *a posteriori* (4.6) en σ_ϵ^2, qui prend la forme d'une loi inverse-Gamma ($\mathcal{IG}$). Une loi *a priori* conjuguée de type $\sigma_\epsilon^2 \sim \mathcal{IG}(\alpha,\beta)$ débouchera alors sur une loi *a posteriori* $p(\sigma_\epsilon^2|\boldsymbol{y},\boldsymbol{q},\boldsymbol{s},\lambda,\sigma^2)$ également $\mathcal{IG}$. Il s'agira alors de choisir un *a priori vague*, c'est-à-dire des paramètres α et β autorisant des valeurs de σ_ϵ^2 sur une grande partie de l'axe réel positif. Notons encore que l'*a priori* de Jeffreys $p(\sigma_\epsilon^2) \propto 1/\sigma_\epsilon^2$ est impropre, mais débouche sur une loi *a posteriori* propre, également de type inverse-Gamma ($p(\sigma_\epsilon^2) \propto 1/\sigma_\epsilon^2$ peut être considéré comme une loi conjuguée $\mathcal{IG}(1,0)$). Nous étudierons donc les différentes lois *a posteriori* suivantes :

$$\begin{cases} \sigma_\epsilon^2 \sim \mathcal{U}_{[0,+\infty[} \ : \ p(\sigma_\epsilon^2) = \mathbb{1}_{[0,+\infty[}(\sigma_\epsilon^2), \\ \sigma_\epsilon^2 \sim \mathcal{IG}(\alpha,\beta) \text{ où } \alpha \text{ et } \beta \text{ seront à régler,} \\ \sigma_\epsilon \text{ suivant l'} a\ priori \text{ de Jeffreys } : \ p(\sigma_\epsilon^2) = 1/\sigma_\epsilon^2. \end{cases}$$

Choix de $p(\sigma^2)$: La dépendance en σ^2 de la loi *a posteriori* (4.6) peut s'écrire sous la forme :

$$\prod_{k/q_k=1} g_2(\boldsymbol{s}_k,\sigma^2\mathbf{I}_2) = \sigma^{-2N_e} \exp -\frac{\sum_{k/q_k=1} ||\boldsymbol{s}_k||^2}{2\boldsymbol{\sigma}^2},$$

où N_e représente le nombre de raies dans le modèle. La validité de l'utilisation d'une loi *a priori* uniforme impropre sur σ^2 est alors conditionnée par l'intégrabilité du terme ci-dessus en σ^2, qui est obtenue lorsque le nombre de raies $N_e > 1$, ce qui n'est pas vrai pour toutes les séquences $\boldsymbol{q}$. Par conséquent, l'intégrale $\mathcal{I}$ explicitée au début de ce paragraphe diverge, et il semble donc plus prudent de ne pas utiliser une telle distribution. De manière analogue à la discussion menée plus haut pour le paramètre σ_ϵ^2 – ces paramètres correspondent tous deux à des variances de lois normales et leur influence dans l'expression de (4.6) revêt des formes similaires – le choix d'un *a priori* conjugué se fera dans la famille des lois $\mathcal{IG}(\mu,\eta)$; le choix de $\mu \geq 1$ assure alors une définition propre de la loi *a posteriori* d'intérêt. En particulier, la loi de Jeffreys $p(\sigma^2) \propto 1/\sigma^2$, qui correspond à une loi impropre $\mathcal{IG}(1,0)$, est également envisageable. Nous étudierons donc les différentes lois *a posteriori* suivantes :

$$\begin{cases} \sigma^2 \sim \mathcal{IG}(\mu,\eta) \quad \text{où } \mu \text{ et } \eta \text{ seront à régler,} \\ \sigma \text{ suivant l'} a\ priori \text{ de Jeffreys } : \ p(\sigma^2) \propto 1/\sigma^2. \end{cases}$$

4.4 Mise en œuvre de l'estimation MP

Nous nous intéressons maintenant à l'échantillonnage de la loi jointe *a posteriori* $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y})$, dont nous détaillons l'expression :

$$\begin{aligned} p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y}) \ \propto \ & \lambda^{N_e}(1-\lambda)^{P+1-N_e} \ \frac{1}{\sigma_\epsilon^N} \exp -\frac{1}{2\sigma_\epsilon^2} \|\boldsymbol{y} - \mathbf{H}\boldsymbol{s}\|^2 \\ & \times \frac{1}{\sigma^{2N_e}} \exp -\frac{1}{2\sigma^2} \|\boldsymbol{s}\|^2 \prod_{k/q_k=0} \delta_2(\boldsymbol{s}_k) \\ & \times p(\lambda) \ p(\sigma_\epsilon^2) \ p(\sigma^2). \end{aligned} \tag{4.13}$$

De par la complexité de sa forme analytique, il est évidemment impossible de simuler directement des variables aléatoires distribuées selon cette distribution. L'algorithme d'*échantillonnage de Gibbs* est une méthode MCMC reposant sur l'utilisation des densités de probabilité conditionnelles, et se révèle particulièrement bien adapté à la structure ci-dessus, pour laquelle les lois conditionnelles de chaque paramètre connaissant les autres variables et les données revêtent des expressions relativement simples, selon lesquelles il est « facile » de générer des échantillons.

4.4.1 L'échantillonnage de Gibbs classique

Cheng *et al.* [1996] puis Doucet et Duvaut [1997] ont proposé, pour des applications de déconvolution, un algorithme d'échantillonnage de Gibbs des paramètres B-G et des hyperparamètres, que nous avons généralisé au problème d'analyse spectrale faisant intervenir des variables $\boldsymbol{s}_k$ bidimensionnelles. Les calculs menant aux expressions des différentes lois conditionnelles sont détaillés dans l'annexe B.2 et le détail de la procédure est présenté ci-dessous, où il est fait usage des notations $\boldsymbol{u}_{i\to j}^{(t)} \triangleq (u_i^{(t)}, u_{i+1}^{(t)}, \ldots, u_j^{(t)})$ et $\boldsymbol{u}_{k,\ell}^{(t,t-1)} \triangleq (u_{1\to k}^{(t)}, u_{\ell\to P}^{(t-1)})$. Le schéma algorithmique décrit au tableau 4.2 génère des échantillons $(\boldsymbol{q}^{(t)}, \boldsymbol{s}^{(t)}, \boldsymbol{\theta}^{(t)})_{t=1\ldots T}$ asymptotiquement distribués selon la distribution *a posteriori* (4.13) ; une preuve de convergence de la procédure générale d'échantillonnage de Gibbs dans le cas général peut être trouvée dans [Robert, 1996], ainsi que les expressions analytiques des différentes lois « classiques » intervenant dans l'algorithme.

Remarquons que les différentes expressions à calculer à chaque itération ne font pas intervenir de calculs en dimension élevée et le balayage itératif des paramètres B-G $(q_k, \boldsymbol{s}_k)$ pour $k = 0 \ldots P$ permet l'actualisation à moindre coût des quantités $\boldsymbol{e}_k$; de ce fait, le coût principal associé à chaque itération est celui des P produits $\mathbf{H}_k^\dagger \boldsymbol{e}_k$. Quelle que soit la forme de l'échantillonnage temporel des données, le coût résultant d'une itération de l'algorithme de Gibbs est donc le même. Nous retrouvons ici une caractéristique commune avec les algorithmes de descente coordonnée par coordonnée étudiés aux paragraphes 3.4.3 et 3.5.4 : le traitement *individuel* de chaque paramètre spectral ne permet pas de tirer parti des propriétés structurelles liées à l'échantillonnage régulier comme les calculs de produits matrice-vecteur par algorithmes FFT, qui n'interviennent que lorsque le vecteur d'amplitudes spectrales est considéré de manière *globale*.

o) Initialisation $t=0$: $\left(\boldsymbol{q}^{(0)},\boldsymbol{s}^{(0)}\right)=(\mathbf{0},\mathbf{0})$ et $\boldsymbol{\theta}^{(0)}=(0.1\,;1\,;1)$.
Pour les itérations $t=1\ldots T$, effectuer les tirages aléatoires suivants :

i) Pour $k=0$, échantillonner les paramètres B-G :

$$\boxed{q_0^{(t)},a_0^{(t)} \sim p\left(q_0,a_0 \,|\, \boldsymbol{q}_{1\to P}^{(t-1)},\boldsymbol{s}_{1\to P}^{(t-1)},\boldsymbol{\theta}^{(t-1)},\boldsymbol{y}\right)}$$

avec $p(q_0,a_0|reste)=\lambda_0\,g(a_0-v_0)+(1-\lambda_0)\,\delta(a_0)$, où :

$$\left\{\begin{array}{rcl} \lambda_0 & = & \dfrac{\widetilde{\lambda}_0}{\widetilde{\lambda}_0+(1-\lambda)},\widetilde{\lambda}_0=\lambda\,\dfrac{\sigma_0}{\sigma}\,\exp\left(\dfrac{v_0^2}{2\gamma_0^2}\right) \\ \gamma_0^2 & = & \sigma^2\sigma_\epsilon^2/(\sigma_\epsilon^2+N\sigma^2) \\ v_0 & = & \frac{\gamma_0^2}{\sigma_\epsilon^2}\sum_n \boldsymbol{e}_{0n}, \text{ avec } \boldsymbol{e}_0=\boldsymbol{y}-\sum_{j=1}^{P+1}\mathbf{H}_j\boldsymbol{s}_j \end{array}\right.$$

et former $\boldsymbol{s}_0^{(t)}=(a_0^{(t)},0)$.

ii) Pour $k=1\ldots P$, échantillonner les paramètres B-G :

$$\boxed{q_k^{(t)},\boldsymbol{s}_k^{(t)} \sim p\left(q_k,\boldsymbol{s}_k \,|\, \boldsymbol{q}_{k-1,k+1}^{(t,t-1)},\boldsymbol{s}_{k-1,k+1}^{(t,t-1)},\boldsymbol{\theta}^{(t-1)},\boldsymbol{y}\right)}$$

avec $p(q_k,\boldsymbol{s}_k|reste)=\lambda_k\,g_2(\boldsymbol{s}_k-\boldsymbol{v}_k)+(1-\lambda_k)\,\delta_2(\boldsymbol{s}_k)$, où :

$$\left\{\begin{array}{rcl} \lambda_k & = & \dfrac{\widetilde{\lambda}_k}{\widetilde{\lambda}_k+(1-\lambda)},\widetilde{\lambda}_k=\dfrac{\lambda}{\sigma^2}\,|\boldsymbol{\Gamma}_k|^{1/2}\,\exp\left(\dfrac{1}{2}\boldsymbol{v}_k^{\mathrm{t}}\boldsymbol{\Gamma}_k^{-1}\boldsymbol{v}_k\right) \\ \boldsymbol{\Gamma}_k^{-1} & = & \frac{1}{\sigma^2}\mathbf{I}_2+\frac{1}{\sigma_\epsilon^2}\mathbf{H}_k^{\mathrm{t}}\mathbf{H}_k \\ \boldsymbol{v}_k & = & \frac{1}{\sigma_\epsilon^2}\boldsymbol{\Gamma}_k\mathbf{H}_k^{\mathrm{t}}\boldsymbol{e}_k, \text{ avec } \boldsymbol{e}_k=\boldsymbol{y}-\sum_{j\neq k}\mathbf{H}_j\boldsymbol{s}_j. \end{array}\right.$$

iii) Échantillonner $\boxed{\lambda^{(t)} \sim p\left(\lambda \,|\, \boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)},\sigma_\epsilon^{2(t-1)},\sigma^{2(t-1)},\boldsymbol{y}\right)}$

avec $\lambda|reste \sim \mathcal{B}e(N_e,P+1-N_e)$.

iv) Échantillonner $\boxed{\sigma_\epsilon^{2(t)} \sim p\left(\sigma_\epsilon^2 \,|\, \boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)},\lambda^{(t)},\sigma^{2(t-1)},\boldsymbol{y}\right)}$

avec $\sigma_\epsilon^2|reste \sim \mathcal{IG}(N/2+\alpha,\|\boldsymbol{y}-\mathbf{H}\boldsymbol{s}\|^2/2+\beta)$

en considérant une loi *a priori* $\sigma_\epsilon^2 \sim \mathcal{IG}(\alpha,\beta)$. Dans le cas d'un *a priori* de Jeffreys, on aura $\alpha=1,\beta=0$ et le cas d'un *a priori* uniforme (impropre) correspond à $\alpha=-1,\beta=0$.

v) Échantillonner $\boxed{\sigma^{2(t)} \sim p\left(\sigma^2 \,|\, \boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)},\lambda^{(t)},\sigma_\epsilon^{2(t)},\boldsymbol{y}\right)}$

$$\text{avec } \sigma^2|reste \sim \left\{\begin{array}{l} \mathcal{IG}(N_e+\mu,\|\boldsymbol{s}\|^2/2+\eta) \text{ si } q_0=0, \\ \mathcal{IG}(N_e-1/2+\mu,\|\boldsymbol{s}\|^2/2+\eta) \text{ si } q_0\neq 0 \end{array}\right.$$

en considérant une loi *a priori* $\sigma^2 \sim \mathcal{IG}(\mu,\eta)$. Pour un *a priori* de Jeffreys, on aura $\mu=1,\eta=0$. La spécificité du cas $q_0=0$ provient de la conséquente nullité de b_0 dans le modèle B-G.

TAB. 4.2 – *Algorithme d'échantillonnage de Gibbs générant* $\left(\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)},\boldsymbol{\theta}^{(t)}\right) \sim p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}\,|\,\boldsymbol{y})$.

Mise en évidence de lenteurs de convergence

La mise en œuvre du schéma d'échantillonnage précédent sur le problème d'analyse spectrale révèle des lenteurs de convergence, où la chaîne $(\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)})$ reste bloquée dans des configurations telles que celle décrite sur la figure 4.2 : si, à une itération donnée, l'échantillonneur produit des échantillons B-G $(\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)})$ modélisant une raie de fréquence f_{k_0} par deux raies de fréquences voisines f_{k_0-1} et f_{k_0+1}, la chaîne de Markov produite par l'algorithme du tableau 4.2 va difficilement sortir de cette configuration. Ce comportement pathologique est lié à la structure même de l'échantillonnage de Gibbs [Robert, 1996, page 205] qui ne modifie qu'une seule composante à la fois, ce qui limite les déplacements « rapides » de la chaîne entre deux configurations différentes, d'autant plus lorsque celles-ci correspondent à des modes locaux de la distribution d'intérêt. L'expression de la loi à partir de laquelle sont générées les composantes B-G permet de mieux comprendre ce phénomène par l'observation suivante, en rappelant que la variable de Bernoulli q_k conditionne l'existence d'une raie de fréquence $f_k = \frac{k}{P} f_{\max}$:

> *Conditionnellement à la connaissance des données et de l'ensemble des composantes B-G aux fréquences $f_{j\neq k}$, la probabilité d'avoir une raie en f_k est d'autant plus élevée que l'est le terme $\check{\boldsymbol{e}}_k = \mathbf{H}_k^{\mathrm{t}}(\boldsymbol{y} - \sum_{j\neq k} \mathbf{H}_j \boldsymbol{s}_j)$, qui correspond au spectre de Fourier*[6] *pris en f_k de la partie des données non modélisée par ces composantes B-G.*

De ce fait, tout changement d'une **unique** composante autour des indices k_0 dans la configuration de type (1-0-1) de la figure 4.2 à gauche se voit affecter d'une probabilité faible :

- La probabilité $\Pr(q_{k_0-1} = 1|\textit{reste})$ est dans ce cas élevée, le reste des composantes B-G (et notamment la raie en f_{k_0+1}) ne suffisant pas à modéliser la fréquence d'indice k_0. De même, $\Pr(q_{k_0+1} = 1|\textit{reste})$ est également élevée.
- À l'inverse, $\Pr(q_{k_0} = 1|\textit{reste})$ sera faible puisque les composantes existantes à l'itération courante en f_{k_0-1} et f_{k_0+1} suffisent à modéliser approximativement la contribution de la fréquence théorique en f_{k_0}. On a en effet la relation :

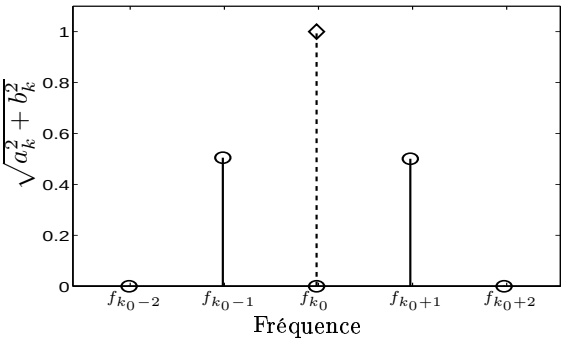
$$\begin{aligned}\exp\big(j2\pi \underbrace{\tfrac{k_0-1}{P} f_{\max}}_{f_{k_0-1}} t_n\big) + \exp\big(j2\pi \underbrace{\tfrac{k_0+1}{P} f_{\max}}_{f_{k_0+1}} t_n\big) &= 2\cos\big(2\pi \frac{1}{P} f_{\max} t_n\big) \exp\big(j2\pi \underbrace{\tfrac{k_0}{P} f_{\max}}_{f_{k_0}} t_n\big) \\ &\simeq 2\exp\big(j2\pi f_{k_0} t_n\big)\end{aligned}$$

  où cette approximation est d'autant plus valable que P est grand.

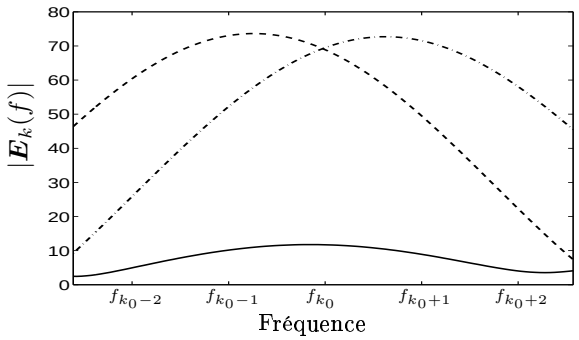
La figure 4.2 à droite illustre ce propos en représentant les spectres de Fourier $\boldsymbol{E}_k(f)$ des signaux $\boldsymbol{e}_k$ (le vecteur bidimensionnel $\check{\boldsymbol{e}}_k$ contient les parties réelle et imaginaire de $\boldsymbol{E}_k(f_k)$) : les spectres de Fourier $\boldsymbol{E}_{k_0-1}(f)$ et $\boldsymbol{E}_{k_0+1}(f)$ présentent des pics aux fréquences f_{k_0-1} et f_{k_0+1} respectivement, traduisant l'interdépendance des deux raies modélisées par la chaîne courante. Dans cet exemple, les probabilités correspondantes $\Pr(q_{k_0-1} = 1|\textit{reste})$ et $\Pr(q_{k_0+1} = 1|\textit{reste})$ sont toutes deux pratiquement égales à 1. À l'inverse, le spectre de Fourier des données auxquelles on a retiré les contributions en f_{k_0-1} et f_{k_0+1} apportées par la chaîne courante présente un comportement très peu marqué en f_{k_0} ; la probabilité correspondante $\Pr(q_{k_0} = 1|\textit{reste})$ vaut ici 2.10^{-3}.

Ce désagrément est à rapprocher des problèmes rencontrés par les techniques d'optimisation, pouvant être piégées dans des modes locaux de la distribution à maximiser : l'algorithme SMLR,

6. Plus précisément, on a par définition $\check{\boldsymbol{y}}_k \triangleq \mathbf{H}_k^{\mathrm{t}}\boldsymbol{y} = [\sum_n y(t_n)\cos(2\pi f_k t_n), \sum_n y(t_n)\sin(2\pi f_k t_n)]$ regroupe les parties réelle et imaginaires du spectre de Fourier $Y(f)$ du signal $\boldsymbol{y}$ pris en la fréquence f_k.



Configuration critique (1-0-1)
$(q_{k_0-1}, q_{k_0}, q_{k_0+1}) = (1, 0, 1)$



Modules de $\boldsymbol{E}_{k_0-1}(f)$, $\boldsymbol{E}_{k_0}(f)$ et $\boldsymbol{E}_{k_0+1}(f)$

FIG. 4.2 – *À gauche, blocage de l'échantillonnage de Gibbs : les* $\circ$ *représentent la chaîne courante et le* $\diamond$ *la raie en* f_{k_0}*. La hauteur des* $\circ$ *est donnée par l'amplitude associée à* $\boldsymbol{s}_k = (a_k, b_k)$*, soit* $\sqrt{a_k^2+b_k^2}$*. À droite, représentation en module des spectres de Fourier* $\boldsymbol{E}_{k_0-1}(f)$ *(- -),* $\boldsymbol{E}_{k_0}(f)$ *(–) et* $\boldsymbol{E}_{k_0+1}(f)$ *(·–)*

dans sa formulation initiale où les voisinages ne diffèrent que d'une seule composante, se retrouvera également bloqué s'il se trouve à une itération donnée dans une configuration autour de f_{k_0} de type (1-0-1) puisque toute configuration qui en diffère par au plus une composante donne lieu à une valeur plus faible de la densité de probabilité *a posteriori*. Les conséquences sur une approche par échantillonnage stochastique sont cependant moins graves, puisque la distribution des échantillons selon la loi d'intérêt est *asymptotiquement* garantie : dans l'exemple ci-dessus, la probabilité $\Pr(q_{k_0=1}|reste)$ n'est pas exactement nulle et l'algorithme de Gibbs est assuré d'aboutir après un nombre suffisant d'itérations à une configuration de type $(q_{k_0-1}, q_{k_0}, q_{k_0+1}) = (0, 1, 0)$, et de ne revenir que *très peu probablement* dans une configuration de la forme (1-0-1), qui est un mode local mais de probabilité bien plus faible que la véritable configuration. Néanmoins, il semble primordial d'envisager des solutions palliant cette lenteur qui pénalise de manière considérable une procédure dont la mise en œuvre est déjà intrinsèquement lourde : c'est l'objet du prochain paragraphe. Notons pour finir que ce genre de configuration problématique n'a lieu d'être que lorsque l'on considère une grille fréquentielle suffisamment fine, *i.e.,* des valeurs élevées de P : l'approximation explicitée ci-dessus d'une raie fréquentielle en f_{k_0} par deux raies d'amplitude moitié en f_{k_0-1} et f_{k_0+1} est d'autant plus valable que la résolution fréquentielle est élevée.

4.4.2 Accélérations de l'algorithme

Au cours de ce travail de thèse, nous avons envisagé deux solutions afin d'éviter le blocage de l'échantillonneur de Gibbs dans de telles configurations critiques. Afin d'alléger les différentes expressions des lois de probabilité considérées, leur dépendance en les hyperparamètres n'est pas détaillée par la suite : les problèmes mentionnés apparaissent à l'intérieur de l'échantillonneur de Gibbs dans les distributions conditionnelles de type $p(q_k, \boldsymbol{s}_k | \boldsymbol{q}_{-k}, \boldsymbol{s}_{-k}, \boldsymbol{y}, \boldsymbol{\theta})$, pour lesquelles $\boldsymbol{\theta}$ ne joue aucun rôle déterminant.

Encourager des transitions depuis les configurations critiques

Nous avons étudié une première approche consistant à favoriser, au cours de l'algorithme d'échantillonnage stochastique, des transitions vers des configurations ne présentant pas ce type de « trous ». Une solution idéale à cet effet serait l'introduction dans le schéma du tableau 4.2 d'une étape de Hastings-Metropolis proposant le passage d'une configuration $(q_{k_0-1}, q_{k_0}, q_{k_0+1}) = (1,0,1)$ à une configuration $(q_{k_0-1}, q_{k_0}, q_{k_0+1}) = (0,1,0)$ et réciproquement, afin de s'assurer de la réversibilité de la chaîne ainsi produite. En décrivant les transitions entre deux configurations sous la forme probabiliste d'une *loi instrumentale* $h_{AB}(\boldsymbol{q}_B,\boldsymbol{s}_B|\boldsymbol{q}_A,\boldsymbol{s}_A)$, une telle étape fonctionne de la façon suivante [Robert, 1996] :

i) étant donné un échantillon courant $(\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)})$ dans une configuration (1-0-1), générer un candidat dans la configuration B selon :

$$\boldsymbol{q}',\boldsymbol{s}' \sim h_{AB}(\boldsymbol{q}_B,\boldsymbol{s}_B|\boldsymbol{q}_A,\boldsymbol{s}_A)$$

ii) $\begin{cases} \text{accepter les nouvelles valeurs } (\boldsymbol{q}^{(t+1)},\boldsymbol{s}^{(t+1)}) = (\boldsymbol{q}',\boldsymbol{s}') \text{ avec une probabilité } \rho_{\text{HM}}, \\ \text{rester en } (\boldsymbol{q}^{(t+1)},\boldsymbol{s}^{(t+1)}) = (\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)}) \text{ avec une probabilité } 1-\rho_{\text{HM}}, \end{cases}$

$$\text{où } \rho_{\text{HM}} = \frac{p(\boldsymbol{q}',\boldsymbol{s}'|\boldsymbol{y})}{p(\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)}|\boldsymbol{y})} \, \frac{h_{BA}(\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)}|\boldsymbol{q}',\boldsymbol{s}')}{h_{AB}(\boldsymbol{q}',\boldsymbol{s}'|\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)})}$$

En revenant à l'exemple pris au paragraphe précédent, si la configuration courante est de type $(q^{(t)}_{k_0-1}, q^{(t)}_{k_0}, q^{(t)}_{k_0+1}) = (1,0,1)$ pour une raie effectivement localisée en f_{k_0}, la proposition d'un candidat présentant une séquence $\boldsymbol{q}'$ de type $(q'_{k_0-1}, q'_{k_0}, q'_{k_0+1}) = (0,1,0)$ et $q'_k = q^{(t)}_k$ ailleurs aura une forte probabilité d'être acceptée, puisque ce candidat donne lieu à une plus forte densité de probabilité *a posteriori* : le premier terme de ρ sera supérieur à 1. Cependant, la nature composite du couple $(\boldsymbol{q},\boldsymbol{s})$ associant une variable discrète et une variable continue rend délicate la définition d'une loi instrumentale h, mélange de probabilités discrètes et de lois continues.

Dans une approche de la déconvolution B-G basée sur des algorithmes d'optimisation itératifs de type SMLR, Chi et Mendel [1984] remarquent des perturbations similaires à celles identifiées au paragraphe précédent et proposent d'inclure, parmi les séquences voisines explorées, celles différant de la séquence courante en deux indices consécutifs. D'une manière analogue, une solution dans le cas de l'échantillonnage stochastique est d'inclure une nouvelle étape d'échantillonnage dans le schéma de l'algorithme de Gibbs du tableau 4.2, favorisant le décalage d'un indice de la position d'une raie. Nous proposons donc le schéma d'échantillonnage modifié suivant :

Échantillonner $\boldsymbol{q}^{(t)}, \boldsymbol{s}^{(t)}, \boldsymbol{\theta}^{(t)}$ selon le schéma de l'algorithme du tableau 4.2 et, avec une probabilité $\eta^{(t)}$, proposer la modification suivante :

i) sélectionner un indice k_1 de la chaîne courante tel que $q_{k_1}^{(t)} = 1$;

ii) avec une probabilité $1/2$, prendre $\delta k_1 = 1$, sinon prendre $\delta k_1 = -1$ et former la séquence de Bernoulli $\boldsymbol{q}'$ telle que :

$$q'_{k_1} = 0,\ q'_{k_1+\delta k_1} = 1 \text{ et } q'_k = q_k^{(t)} \text{ ailleurs ;}$$

iii) estimer les amplitudes correspondantes $\boldsymbol{s}'$ au sens du maximum *a posteriori* :

$$\boldsymbol{s}' = \arg\max_{\boldsymbol{s}} p(\boldsymbol{s}'|\boldsymbol{y},\boldsymbol{q}'), \text{ dont l'expression est donnée par (4.8) ;}$$

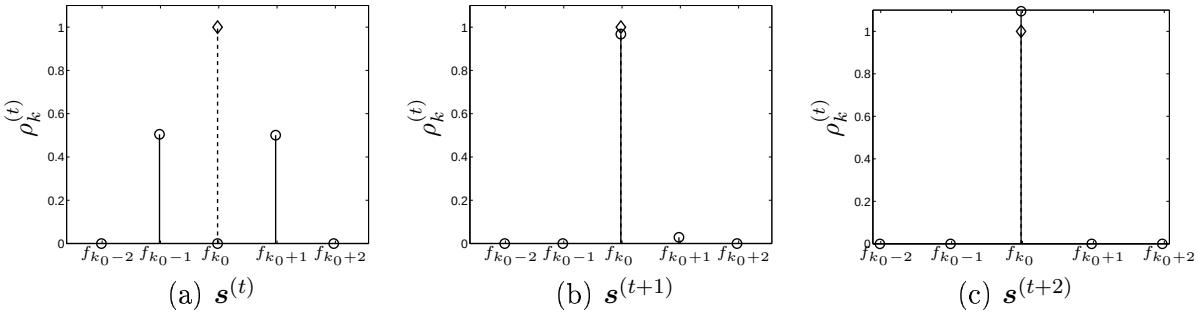
iv) accepter $(\boldsymbol{q}^{(t+1)},\boldsymbol{s}^{(t+1)}) = (\boldsymbol{q}',\boldsymbol{s}')$ si le candidat augmente la loi *a posteriori*,

$$\text{c'est-à-dire si } \mu = \frac{p(\boldsymbol{q}',\boldsymbol{s}'|\boldsymbol{y})}{p(\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)}|\boldsymbol{y})} > 1$$

et rester en $(\boldsymbol{q}^{(t+1)},\boldsymbol{s}^{(t+1)}) = (\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)})$ sinon.

TAB. 4.3 – *Algorithme modifié pour l'échantillonnage de* $p(\boldsymbol{q}, \boldsymbol{s}, \boldsymbol{\theta}|\boldsymbol{y})$.

À titre d'exemple, la figure 4.3 détaille le comportement de cette procédure sur la configuration (1-0-1) précédente, témoin du blocage de l'échantillonneur de Gibbs. La figure (a) montre la configuration initiale critique : $(q_{k_0-1}^{(t)}, q_{k_0}^{(t)}, q_{k_0+1}^{(t)}) = (1, 0, 1)$ alors que la raie théorique se trouve en f_{k_0}. En sélectionnant par exemple $k_1 = k_0 - 1$ et $\delta k_1 = 1$, la séquence de Bernoulli proposée $\boldsymbol{q}'$ vaut $(q'_{k_0-1}, q'_{k_0}, q'_{k_0+1}) = (0, 1, 1)$ et l'estimation au sens du MAP des amplitudes correspondant à $\boldsymbol{q}'$ est représentée sur la figure (b) : pour un modèle courant avec deux sinusoïdes de fréquences f_{k_0} et f_{k_0+1}, l'amplitude en f_{k_0} se voit affecter une valeur élevée, alors que l'amplitude associée à f_{k_0+1} est faible. Le rapport des lois *a posteriori* vaut dans cet exemple $\mu \simeq 24$, et la transition $(\boldsymbol{q}^{(t+1)},\boldsymbol{s}^{(t+1)}) = (\boldsymbol{q}',\boldsymbol{s}')$ est acceptée. Á la suite d'une itération supplémentaire de l'échantillonneur de Gibbs, la figure (c) montre que la composante en f_{k_0+1} a même complètement disparu : on a à présent $(q_{k_0-1}^{(t+2)}, q_{k_0}^{(t+2)}, q_{k_0+1}^{(t+2)}) = (0, 1, 0)$.



(a) $\boldsymbol{s}^{(t)}$ (b) $\boldsymbol{s}^{(t+1)}$ (c) $\boldsymbol{s}^{(t+2)}$

FIG. 4.3 – *Comportement de la procédure du tableau 4.3 sur la situation de blocage de l'échantillonneur de Gibbs classique de la figure 4.2. Les* $\diamond$ *localisent la raie théorique en* f_{k_0} *et les* $\circ$ *le spectre correspond à la chaîne courante à divers moments de la procédure. La hauteur des* $\circ$ *est donnée par l'amplitude associée à* $\boldsymbol{s}_k = (a_k, b_k)$, *soit* $\rho_k = \sqrt{a_k^2 + b_k^2}$. (t) *: configuration critique,* $(t+1)$ *: sortie de l'étape (iv) et* $(t+2)$ *: après une itération supplémentaire de l'échantillonnage de Gibbs.*

Il faut noter que cette étape n'est pas à proprement parler un pas de Hastings-Metropolis, puisque les amplitudes sont proposées de manière *déterministe* et au sens du MAP, favorisant ainsi le taux d'acceptation à l'étape iv) de l'algorithme modifié : μ n'est pas une probabilité de la forme ρ_{HM} et la chaîne produite n'est alors pas stationnaire. Par conséquent, cette étape ne peut être réalisée tout au long de la procédure d'échantillonnage mais doit plus être interprétée comme une étape de « chauffage », évitant une initialisation critique de l'échantillonneur de Gibbs. En pratique, son application sera proposée, à chaque itération t, avec une probabilité $\eta^{(t)}$ telle que $\eta^{(t)} \to 0$ lorsque $t \to +\infty$.

Échantillonnage par blocs

Vu que les configurations problématiques mettent en jeu trois indices consécutifs de la séquence de Bernoulli, nous proposons une seconde solution mettant en œuvre l'échantillonnage de Gibbs de la loi $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y})$ non plus par la simulation des lois conditionnelles $p(q_k,\boldsymbol{s}_k|reste)$, mais en groupant les paramètres par blocs de trois indices consécutifs, en échantillonnant conjointement :

$$\boldsymbol{q}_{k\to k+2},\boldsymbol{s}_{k\to k+2} \sim p(\boldsymbol{q}_{k\to k+2},\boldsymbol{s}_{k\to k+2}|\boldsymbol{q}_{-(k\to k+2)},\boldsymbol{s}_{-(k\to k+2)},\boldsymbol{y},\boldsymbol{\theta})$$

où $\boldsymbol{q}_{-(k\to k+2)},\boldsymbol{s}_{-(k\to k+2)}$ représente l'ensemble des échantillons $\boldsymbol{q},\boldsymbol{s}$ excepté ceux d'indices k à $k+2$. Cette approche permet d'envisager de manière plus élégante la transition directe d'un échantillon « critique » vers une configuration correcte : en reprenant notre exemple phare de la figure 4.2 et les notations correspondantes, il est naturel de penser que l'échantillonnage conjoint de $(\boldsymbol{q}_{k_0-1\to k_0+1},\boldsymbol{s}_{k_0-1\to k_0+1})$ favorisera une séquence de Bernoulli de la forme (0,1,0) puisque l'échantillonnage en k_0 n'est plus conditionné par l'existence de raies aux points voisins $k_0 \pm 1$.

La loi jointe présente toutefois une expression analytique plus complexe que pour le cas « scalaire » précédent, et la mise en œuvre de l'échantillonnage correspondant nécessite un coût calculatoire plus élevé. En effet, d'une manière générale, l'échantillonnage conjoint de m paramètres B-G nécessite le calcul des 2^m probabilités correspondant à chaque séquence de Bernoulli possible, au lieu de m probabilités pour un échantillonnage séparé. Dans le cas d'un échantillonnage de Gibbs par blocs de trois composantes, la loi *a posteriori* s'écrit sous la forme :

$$p(\boldsymbol{q}_{k\to k+2},\boldsymbol{s}_{k\to k+2}|reste) = \sum_{(u,v,w)\in\{0;1\}^3} \lambda_{ku,v,w}\; g_6(\boldsymbol{s}_{k\to k+2} - \boldsymbol{v}_{ku,v,w}, \boldsymbol{\Gamma}_{ku,v,w}) \qquad (4.14)$$

où $\lambda_{ku,v,w}$ est la probabilité *a posteriori*, connaissant les données et le reste des paramètres B-G, que les variables de Bernoulli $\boldsymbol{q}_{k\to k+2}$ soient dans la configuration (u,v,w) et $g_6(\boldsymbol{s}_{k\to k+2}-\boldsymbol{v}_{ku,v,w},\boldsymbol{\Gamma}_{ku,v,w})$ est la distribution *a posteriori* gaussienne correspondante des amplitudes : $\boldsymbol{\Gamma}_{ku,v,w}$ est une matrice de taille 6×6, diagonale par blocs 2×2, d'expressions similaires à la matrice $\boldsymbol{\Gamma}_k$ intervenant dans le cas de l'échantillonnage classique de l'algorithme du tableau 4.2. Notons que cette distribution gaussienne est à considérer au sens des distributions et inclut des blocs identiquement nuls correspondant aux zéros de la configuration (u,v,w). Afin de ne pas pénaliser la lisibilité de ce manuscrit, les expressions de $\lambda_{u,v,w}$ et $\boldsymbol{\Gamma}_{u,v,w}$ sont renvoyées à l'annexe B.2. L'échantillonnage par blocs, plus coûteux, ne sera effectué en pratique qu'en alternance avec l'échantillonnage de Gibbs classique. Par ailleurs, afin de mieux couvrir d'éventuelles configurations problématiques, la procédure devra également éviter de toujours échantillonner les mêmes blocs 3×3. L'algorithme résultant est donc au final le suivant :

i) échantillonner $\boldsymbol{q}^{(t)},\boldsymbol{s}^{(t)},\boldsymbol{\theta}^{(t)}$ selon le schéma de l'algorithme du tableau 4.2 ;
ii) toutes les T_3 itérations, choisir de manière équiprobable $\delta k = 0,1$ ou 2 puis pour $k = 0\ldots[\frac{P+1}{3}]$, échantillonner $\boldsymbol{q}^{(t)}_{3k+\delta_k\to 3k+\delta_k+2}, \boldsymbol{s}^{(t)}_{3k+\delta_k\to 3k+\delta_k+2}$ selon la loi (4.14).

TAB. 4.4 – *Algorithme d'échantillonnage de $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y})$ incluant des étapes d'échantillonnage de Gibbs par blocs.*

En reprenant l'exemple de la configuration critique précédente, la figure 4.4 compare les probabilités $\lambda_{u,v,w}$ associées aux blocs de trois composantes comprenant k_0. Dans chaque cas, la

probabilité est quasiment égale à 1 pour la configuration localisant la raie en k_0, alors que les autres ne dépassent pas 10^{-3}.

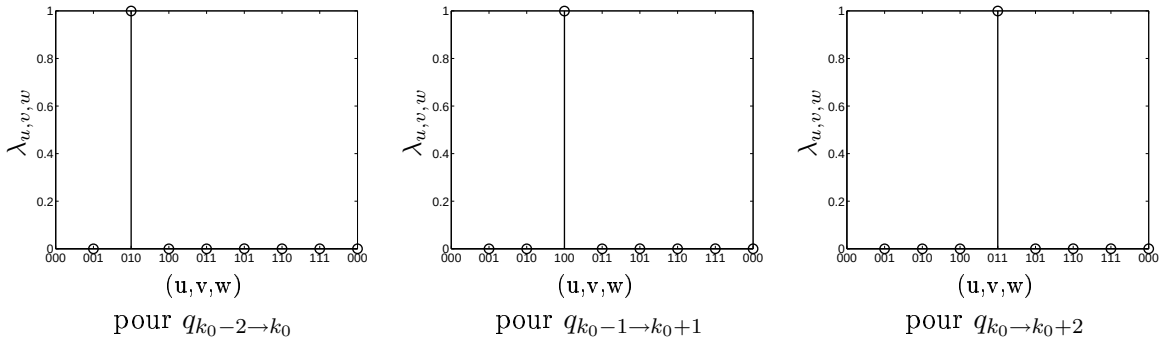


FIG. 4.4 – *Probabilités a posteriori $\lambda_{u,v,w}$ pour les trois configurations impliquant k_0, correspondant à une étape d'échantillonnage par blocs associée à la configuration critique de la figure 4.2.*

4.5 Résultats d'estimation sur le modèle B-G

Afin d'illustrer la qualité de l'estimation et le comportement des algorithmes d'échantillonnage stochastique, nous appliquons la procédure d'estimation par modélisation Bernoulli-Gaussienne sur le signal multi-sinusoïdal à la base des simulations du chapitre 3 et présenté sur la figure 2.6, composé de cinq sinusoïdes $\nu_1, \ldots, \nu_5$ bruitées à 10 dB. Des échantillons $(\boldsymbol{q}^{(t)}, \boldsymbol{s}^{(t)}, \boldsymbol{\theta}^{(t)})$ sont générés selon les trois algorithmes étudiés, à savoir :

- le schéma d'échantillonnage classique du tableau 4.2, noté G_0 ;
- le schéma favorisant les décalages des positions des raies dans $\boldsymbol{q}^{(t)}$ selon l'algorithme du tableau 4.3. Nous notons cette version G_μ en référence à la probabilité $\mu^{(t)}$ à laquelle ces transitions sont proposées ;
- l'algorithme du tableau 4.4 proposant l'échantillonnage par blocs de trois composantes B-G, que nous désignerons par G_3.

La détermination d'un test de contrôle assurant une exploration suffisante de la distribution de probabilité échantillonnée est une problématique qui reste très ouverte [Robert, 1996, chapitre 6]. L'ensemble des résulats présentés dans ce paragraphe fait usage de la règle suivante : l'échantillonnage est arrêté lorsque, à une itération donnée, l'ensemble des paramètres estimés $\widehat{\boldsymbol{q}}^T, \widehat{\boldsymbol{s}}^T, \widehat{\boldsymbol{\theta}}^T$ ne varie pratiquement plus, soit plus précisément lorsque :

$$||\widehat{\boldsymbol{u}}^T - \widehat{\boldsymbol{u}}^{T-T_0}||_\infty < \alpha \quad \text{avec } \alpha = 10^{-3} \quad \text{et} \quad \widehat{\boldsymbol{u}}^T \triangleq \frac{2}{T} \sum_{t=T/2+1}^{T} \boldsymbol{u}^{(t)}, \tag{4.15}$$

où $\widehat{\boldsymbol{u}}^T$ regroupe l'ensemble des paramètres estimés $(\widehat{\boldsymbol{q}}^T, \widehat{\boldsymbol{s}}^T, \widehat{\boldsymbol{\theta}}^T)$ à l'itération T. L'utilisation de la norme $\|\cdot\|_\infty$ signifie que les algorithmes sont arrêtés lorsque l'estimé d'aucun paramètre *scalaire* parmi $\boldsymbol{q}$, $\boldsymbol{s}$ et $\boldsymbol{\theta}$ ne varie de plus de 10^{-3} entre les itérations $T - T_0$ et T. Par la suite, T_0 est fixé à 10 itérations. Le rejet, à une itération T, de la moitié des échantillons permet de s'affranchir des effets dûs à l'initialisation de la chaîne, les premiers échantillons étant en pratique très éloignés des valeurs finales. Ainsi, même si une telle procédure rejette un nombre croissant d'échantillons

lorsque T augmente, le calcul de la moyenne sur la seconde moitié de la chaîne produite vise à ne considérer que sa partie stationnaire, tout en évitant de régler un nombre fixe d'itérations de « chauffage ». Notons que la définition (4.15) permet une estimation *récursive* des différents paramètres et ne nécessite que le stockage en mémoire de blocs de taille T_0.

4.5.1 Comportement algorithmique

Fréquences $\nu_1, \ldots, \nu_5 \in \mathcal{G}$

Nous illustrons dans un premier temps le comportement des différents algorithmes dans le cas favorable où les cinq fréquences ν_j sont sur la grille de reconstruction, c'est-à-dire s'écrivent sous la forme $\nu_j = \frac{k_j}{P} f_{\max}$. La figure 4.5 présente un exemple d'estimés obtenus après $T = 2500$ itérations de chaque algorithme, qui est le nombre typique d'échantillons nécessaires à la satisfaction du critère (4.15) pour les algorithmes G_μ et G_3. Tous les échantillons produits par G_0

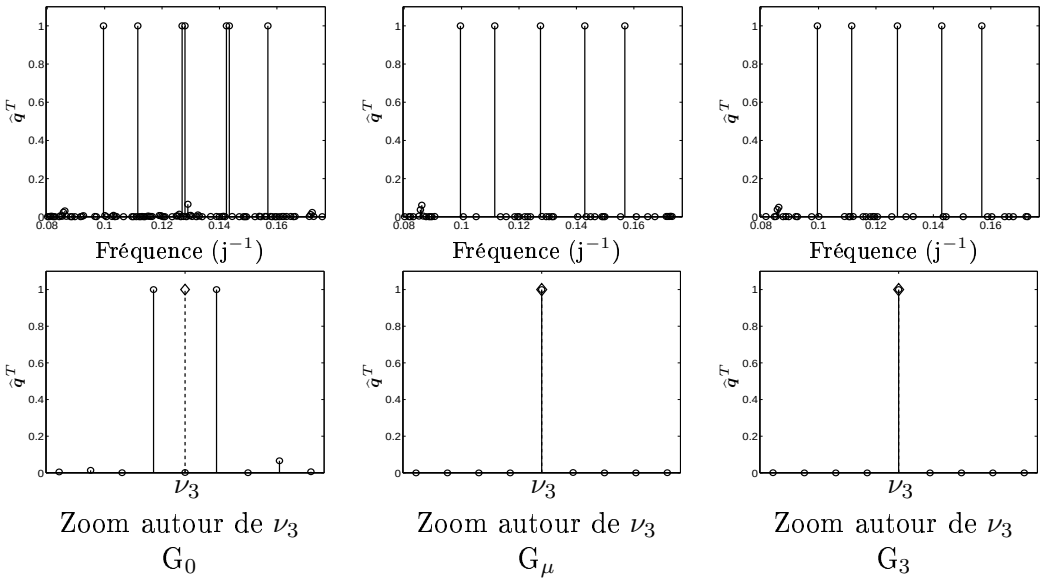
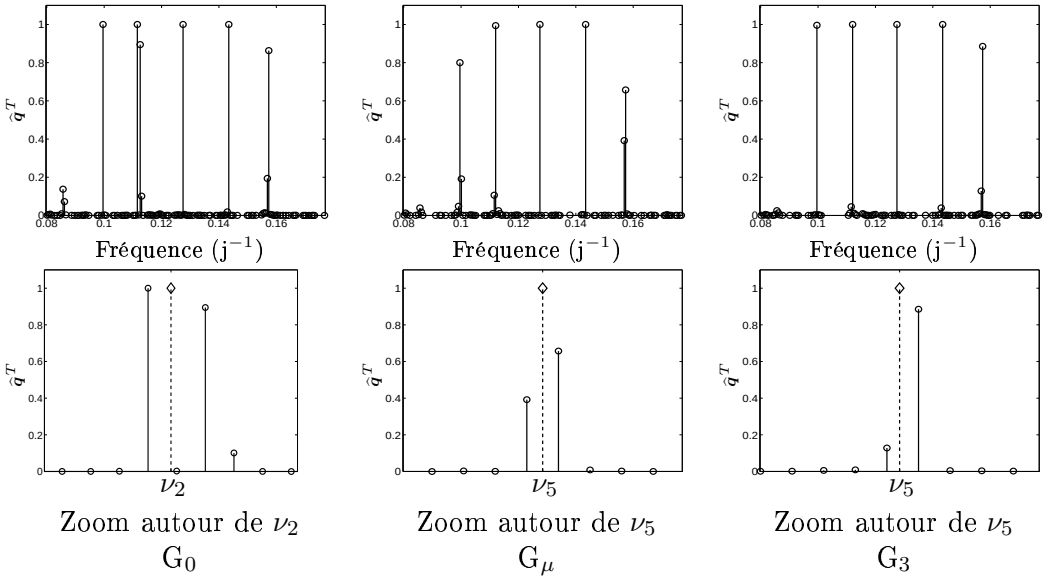


FIG. 4.5 – *Pour des fréquences $\nu_j \in \mathcal{G}$, probabilités $\boldsymbol{q}^T$ estimées ($\circ$) après $T = 2500$ itérations des trois algorithmes. Par souci de lisibilité, seules les probabilités non nulles sont représentées. Les $\diamond$ localisent les raies spectrales théoriques.*

présentent une configuration critique de la forme (1-0-1) autour de ν_3 et ν_5, donnant lieu à des probabilités estimées $\widehat{\boldsymbol{q}}^T$ de cette forme. Notons que l'algorithme est ici arrêté après un nombre fixe d'itérations, mais que la persistance de la chaîne $\boldsymbol{q}^{(t)}$ dans cette configuration peut provoquer en pratique la satisfaction du critère de convergence (4.15) pour un estimé semblable à celui de la figure de gauche. Les algorithmes G_μ et G_3 permettent de résoudre ce problème. Dans le cas de G_μ, la décroissance de la probabilité $\mu^{(t)}$ est choisie sous forme exponentielle, avec $\mu^{(0)} = 1$ et $\mu^{(T/2)} = 10^{-2}$: de cette façon l'estimation, prenant en compte les $T/2$ derniers éléments de la chaîne, n'est que très faiblement perturbée par l'étape de chauffage décrite dans l'algorithme du tableau 4.3 qui, rappelons-le, n'est pas une étape de Hastings-Metropolis et, par un trop fort taux d'acceptation, produit une chaîne non stationnaire.

Fréquences $\nu_1, \ldots, \nu_5 \notin \mathcal{G}$

Nous illustrons ici l'effet d'une discrétisation inexacte des fréquences sur la grille $\mathcal{G}$. Les estimés $\widehat{\boldsymbol{q}}^T$ obtenus après $T = 4000$ itérations, qui est le nombre typique d'échantillons nécessaires à la satisfaction du critère (4.15) pour les algorithmes G_μ et G_3, sont présentés pour chaque algorithme sur la figure 4.6. Comme dans le cas précédent, l'algorithme de Gibbs classique reste bloqué dans des modes locaux correspondant à des configurations de la séquence de Bernoulli de la forme (1-0-1), desquels les deux alternatives G_μ et G_3 permettent de sortir. Néanmoins, l'impossibilité de modéliser exactement les raies théoriques provoque l'alternance, au cours de la chaîne $\boldsymbol{q}^{(t)}$, de localisations des raies théoriques aux positions discrétisées de part et d'autre des fréquences ν_j. L'estimation de $\boldsymbol{q}$ au sens de la moyenne *a posteriori* réalise alors le moyennage entre les indices adjacents $k_j - 1$ et k_j tels que $f_{k_j-1} \leq \nu_j \leq f_{k_j}$ où $f_k = \frac{k}{P} f_{\max}$, et complique l'interprétation des résultats. Notons que cet effet d'alternance est propre à l'estimation MP et n'a évidemment pas lieu pour une approche MAP, qui « choisira » entre l'une des deux localisations en f_{k_j-1} ou f_{k_j}.



Zoom autour de ν_2 — G_0 | Zoom autour de ν_5 — G_μ | Zoom autour de ν_5 — G_3

FIG. 4.6 – *Pour des fréquences $\nu_j \notin \mathcal{G}$, probabilités $\boldsymbol{q}^T$ estimées ($\circ$) après $T = 4000$ itérations des trois algorithmes. Par souci de lisibilité, seules les probabilités non nulles sont représentées. Les $\diamond$ localisent les raies spectrales théoriques.*

Il ne faut cependant pas confondre les deux effets de déboublement d'une raie présentés sur la figure 4.6 à gauche d'une part (algorithme G_0) et sur les deux figures de droite d'autre part (algorithmes G_μ et G_3). Dans le premier cas, la séquence $\boldsymbol{q}^T$ est faussée par le blocage de l'algorithme d'échantillonnage dans un mode local alors que pour le second, le dédoublement est dû à l'impossibilité de modéliser de façon exacte une raie en dehors de la grille de reconstruction. Néanmoins, l'interprétation postérieure des résultats de l'estimation permet de résoudre partiellement ce problème : dans l'exemple ci-dessus, pour les deux estimés produits par G_μ et G_3, la somme des probabilités estimées de part et d'autre de ν_j vaut $\widehat{q}_{k_j-1} + \widehat{q}_{k_j} \simeq 1$. Cette observation permet de conclure, au vu des figures pour G_μ et G_3, à l'existence d'une *unique* composante fréquentielle située entre f_{k_j-1} et f_{k_j} alors qu'une interprétation du résultat de la figure pour G_0 peut laisser croire à la présence effective de deux raies proches en f_{k_j-1} et f_{k_j+1} (sur la figure 4.6

à gauche, on a $\widehat{q}_{k_j-1}$ et $\widehat{q}_{k_j+1}$ toutes deux proches de 1).

4.5.2 Influence des hyperparamètres

Nous étudions ici la qualité d'estimation des hyperparamètres $\boldsymbol{\theta} = (\lambda,\sigma_\epsilon^2,\sigma^2)$ et son éventuelle dépendance envers les distributions *a priori* $p(\boldsymbol{\theta})$ choisies. Nous utilisons le même signal que précédemment, où les fréquences $\nu_{j,\,i=1\ldots5}$ sont choisies sur la grille de reconstruction, constituée ici de $P+1 = 501$ fréquences. Le choix de fréquences ν_j en dehors de la grille a débouché en simulation sur des résultats similaires. Les valeurs « attendues » des hyperparamètres correspondent alors à :

- $\lambda \simeq 5/(P+1) \simeq 0.01$, nombre « moyen » de raies ;
- $\sigma_\epsilon^2 = \sigma_{\epsilon\,\mathrm{th}}^2 = 0.131$, qui est la variance du bruit gaussien additif ajouté aux données pour avoir un rapport signal sur bruit de 10 dB, en rappelant la définition utilisée du RSB pour un signal multi-sinusoïdal de la forme $y(t) = \sum_{k=1}^{K} \alpha_k \cos(2\pi\phi_k t) + \beta_k \sin(2\pi\phi_k t) + \epsilon(t)$:

$$\mathrm{RSB}_{\mathrm{dB}} = 10\,\log_{10}\,\frac{\sum_{k=1}^{K}\alpha_k^2+\beta_k^2}{2\,\sigma_\epsilon^2}\;;$$

- σ^2 est la variance associée aux amplitudes (α_k,β_k). Dans ce signal de test, les amplitudes ont été fixées, mais nous pouvons considérer : $\sigma^2 \simeq \frac{1}{2K}\sum_{k=1}^{K}\alpha_k^2+\beta_k^2$, soit $\sigma^2 = 0.263$ dans cet exemple.

Notons que nous ne considérons pas ici le problème de l'estimation des hyperparamètres en tant que tel : cette étude (caractérisation du biais et de la variance associé à l'estimation de θ) demanderait la réalisation d'un grand nombre de spectres *réellement* issus d'un processus aléatoire B-G réglé par les paramètres $\boldsymbol{\theta}$. Ce n'est pas notre objectif ici, qui est plutôt l'étude de l'influence du choix de la loi *a priori* sur la distribution *a posteriori* $p(\boldsymbol{\theta}|\boldsymbol{y})$ et sur l'estimation des paramètres B-G.

L'échantillonnage de la loi jointe $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y})$ est réalisé pour les différentes lois *a priori* envisagées sur les hyperparamètres. Les distributions *a posteriori* $p(\boldsymbol{\theta}|\boldsymbol{y})$ estimées à partir des échantillons $\boldsymbol{\theta}^{(t)}$ sont représentées sur la figure 4.7, après convergence de l'algorithme G_3 au sens du critère (4.15). Nous présentons ici les distributions obtenues pour une seule réalisation du bruit et une même initialisation du générateur aléatoire, mais dont la forme est représentative des résultats obtenus de manière plus générale. Dans chaque cas, le nombre d'itérations nécessaires est de l'ordre de 2 500. La loi *a priori* $p(\boldsymbol{\theta})$ étant choisie sous la forme séparable $p(\boldsymbol{\theta}) = p(\lambda)\,p(\sigma_\epsilon^2)\,p(\sigma^2)$, nous étudions l'influence de chaque loi *a priori* de manière indépendante sur la distribution *a posteriori* du paramètre correspondant.

Pour le paramètre de Bernoulli, seule une distribution $p(\lambda)$ uniforme sur $[0,1]$ a été envisagée. Même si aucune information *a priori* ne privilégie une faible valeur de λ (et donc une forme parcimonieuse du spectre $\boldsymbol{s}$ à reconstruire), la distribution *a posteriori* $p(\lambda|\boldsymbol{y})$ présente une allure assez « piquée » autour de 0.01 (figure a)). En réalité, l'estimation au sens de la MP de λ débouche sur une valeur de l'ordre de :

$$\widehat{\lambda}_{\mathrm{MP}} = 0.013 \pm 5.10^{-3},$$

c'est-à-dire légèrement supérieure au nombre de raies moyen.

La distribution *a posteriori* $p(\sigma_\epsilon^2|\boldsymbol{y})$ est représentée pour les différentes lois *a priori* décrites au § 4.3.4, à savoir des lois impropres uniformes et de Jeffreys, et une loi conjuguée

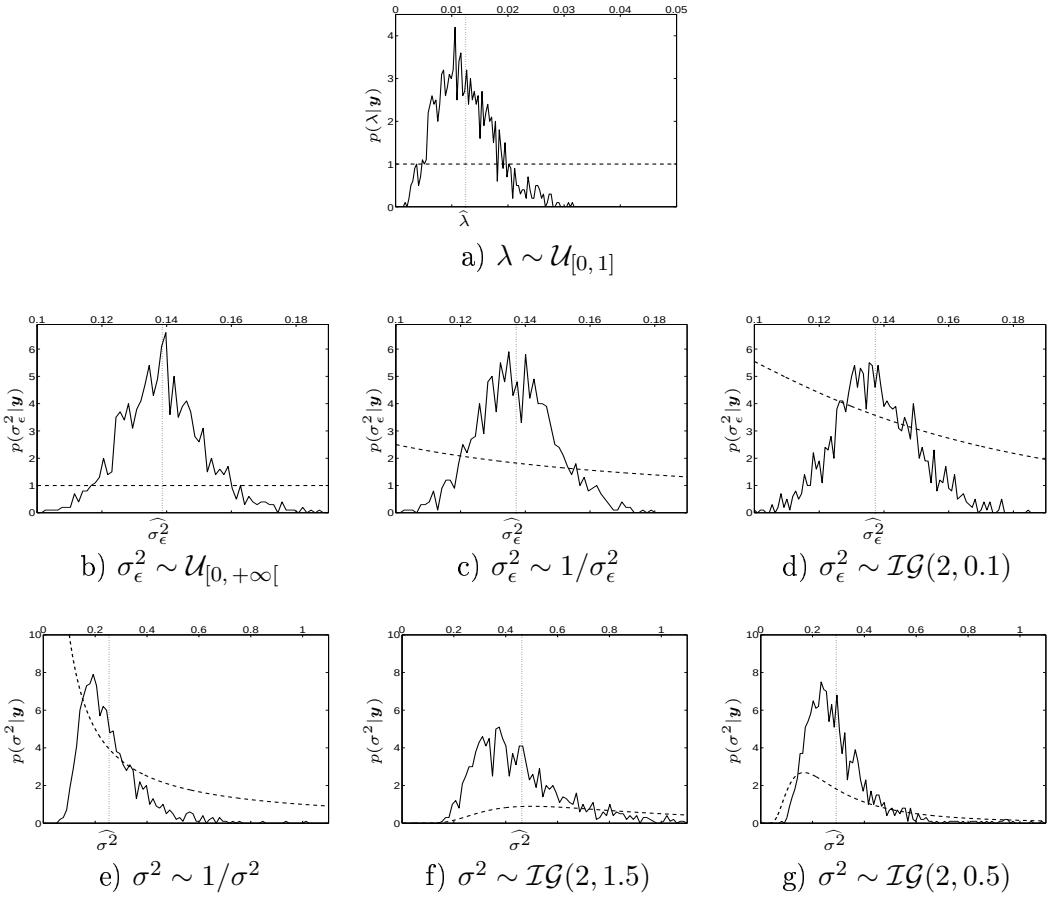


FIG. 4.7 – *Distributions a posteriori des hyperparamètres estimées par les histogrammes de* $\boldsymbol{\theta}^{(t)}$ *(–) en fonction de la loi a priori adoptée (- -). Les traits verticaux correspondent à la position de la moyenne a posteriori estimée.*

$\mathcal{IG}(\alpha = 2, \beta = 0.1)$, dont les paramètres ont été choisis de manière à avoir $\mathbb{E}[\sigma_\epsilon^2] = \frac{\beta}{\alpha - 1} = 0.1$ (c'est-à-dire proche de la valeur théorique) et var $\sigma_\epsilon^2 = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} = +\infty$. Quelle que soit la loi $p(\sigma_\epsilon^2)$ envisagée, les distributions *a posteriori* obtenues sont très semblables (figures b), c) et d)). En particulier, l'estimation de σ_ϵ^2 au sens de la MP fournit (ici pour un *a priori* de Jeffreys) une valeur de :

$$\widehat{\sigma_\epsilon^2}_{\text{MP}} = 0.138 \pm 0.013,$$

qui est à comparer à la valeur théorique utilisée pour générer le signal : $\sigma_{\epsilon\,\text{th}}^2 = 0.1313$.

Enfin, nous considérons l'influence sur $p(\sigma^2|\boldsymbol{y})$ des lois *a priori* $p(\sigma^2)$ de Jeffreys et $\mathcal{IG}(\mu = 2, \eta = 1.5)$. Les paramètres μ et η ont été choisis de sorte que $\mathbb{E}[\sigma^2] = \frac{1}{N}\boldsymbol{y}^{\text{t}}\boldsymbol{y}$, puissance moyenne du signal observé, et var $\sigma^2 = \frac{\eta^2}{(\mu-1)^2(\mu-2)} = +\infty$, ne nécessitant ainsi aucun réglage arbitraire. Alors que la loi de Jeffreys procure une estimation satisfaisante (figure e)) :

$$p(\sigma^2) \propto 1/\sigma^2 \;\Rightarrow\; \widehat{\sigma^2}_{\text{MP}} = 0.251 \pm 0.121,$$

la loi $\mathcal{IG}(\mu = 2, \eta = 1.5)$ débouche sur une distribution *a posteriori* centrée sur des valeurs trop élevées (figure f)) :

$$\sigma^2 \sim \mathcal{IG}(2, 1.5) \;\Rightarrow\; \widehat{\sigma^2}_{\text{MP}} = 0.462 \pm 0.202.$$

En effet, la loi $\mathcal{IG}$ présente une allure très écrasée pour σ^2 faible. Néanmoins, dans ces simulations, la surestimation de σ^2 n'a pas entraîné de dégradation notable sur l'estimation des paramètres B-G. Notons encore qu'un réglage *ad hoc* des paramètres μ et η permet de retrouver une estimation plus cohérente de σ^2 (figure g)) :

$$\sigma^2 \sim \mathcal{IG}(2, 0.5) \;\Rightarrow\; \widehat{\sigma^2}_{\text{MP}} = 0.291 \pm 0.128.$$

Sur ces simulations, le choix de l'*a priori* en λ et σ_ϵ^2 n'a donc que très peu d'influence sur l'estimation de ces paramètres, et *a fortiori* sur l'allure du spectre estimé. En particulier, le réglage *ad hoc* des paramètres d'une loi $\mathcal{IG}$ n'apporte pas d'amélioration notable. Par contre, l'usage d'une loi *a priori* inverse Gamma sur la variance σ^2 montre une sensibilité assez forte au réglage des paramètres correspondants, et l'allure « écrasée » de la loi $\mathcal{IG}$ en 0 peut considérablement pénaliser des faibles valeurs de σ^2. Par conséquent, l'emploi pour ce problème de lois conjuguées inverse-Gamma est plutôt une source de complications. Nous préconisons donc le choix d'une loi uniforme sur $[0, 1]$ pour le paramètre de Bernoulli et, pour les variances l'usage de lois impropres de Jeffreys, ou encore impropre uniforme dans le cas de σ_ϵ^2.

4.5.3 Bilan de l'estimation

Nous consignons ici les résultats produits par l'estimation des paramètres B-G décrite au tableau 4.1 sur le signal multi-sinusoïdal habituel, pour lequel les fréquences sont choisies en dehors de la grille $\mathcal{G}$ composée ici de P+1 = 501 fréquences. L'échantillonnage de la loi *a posteriori* $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y})$ est mis en œuvre dans sa version G_3 où une étape d'échantillonnage par blocs de trois composantes B-G est réalisée toutes les $T_3 = 10$ itérations. Des lois *a priori* uniformes ont été choisies sur λ et σ_ϵ^2 ainsi qu'une loi *a priori* de Jeffreys sur σ^2. Pour la réalisation de l'algorithme d'échantillonnage dont nous présentons ici les résultats, un total de $T = 3800$ itérations sont nécessaires[7] à la satisfaction du critère d'arrêt (4.15), qui demandent un temps cpu

7. Le nombre d'itérations peut en effet varier en fonction de l'initialisation du générateur aléatoire, mais reste de cet ordre de grandeur.

d'environ 600 s avec un processeur de 1.4 GHz pourvu de 1Go de mémoire vive, les algorithmes ayant été mis en œuvre sous *Matlab*.

La moyenne *a posteriori* de la séquence de Bernoulli est représentée en figure 4.8. La probabilité *a posteriori* estimée d'existence d'une raie est très faible en dehors des localisation des cinq raies théoriques : la figure du centre montre que la probabilité maximale de « fausse alarme» est d'environ 0.05. L'agrandissement de la figure de droite montre que, pour les fréquences ν_1 à ν_4, une probabilité égale à 1 est estimée en les points de la grille fréquentielle les plus proche des valeurs théoriques. Dans le cas de ν_5, cette probabilité se répartit sur les localisations de part et d'autre de ν_5, comme nous l'avons déjà vu au § 4.5.1.

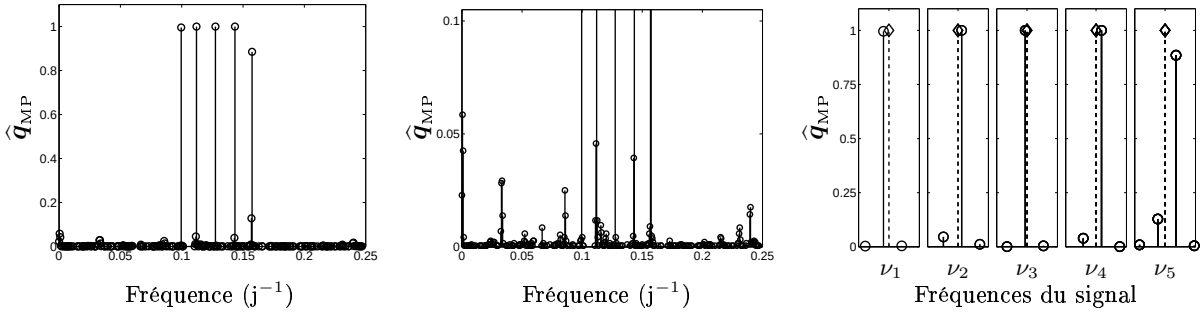


FIG. 4.8 – *Estimation $\widehat{\boldsymbol{q}}_{MP}$ pour un modèle $B-G$ avec décalages sur un signal multi-sinusoïdal. À gauche, sur l'ensemble de la grille fréquentielle. Au centre, zoom en amplitude pour les faibles valeurs. À droite, zoom en fréquence autour de chaque raie. Les $\diamond$ représentent la position des raies théoriques.*

Au vu de la figure 4.8, nous ne conservons que les positions fréquentielles donnant lieu à une valeur de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ supérieure à $\alpha = 0.5$. Sur cet exemple, tout seuil α choisi entre 0.2 et 0.8 donne lieu à la même séquence seuillée $\widehat{\boldsymbol{q}}_{\alpha}$ et permet en particulier de ne conserver que la position la plus probable parmi les deux fréquences localisant approximativement ν_5. La seconde phase d'estimation des amplitudes correspondantes est réalisée conditionnellement à la connaissance des positions fréquentielles au sens du MAP (ou de manière identique de la MP) selon l'expression analytique du tableau 4.1 et l'estimation finale résultante est représentée sur la figure 4.9. En notant $\widehat{\boldsymbol{s}}_k = [\widehat{a}_k, \widehat{b}_k]$ les paramètres d'amplitude estimés pour chaque raie, nous représentons la moyenne de $\rho_k \stackrel{\Delta}{=} \sqrt{a_k^2 + b_k^2}$ et les barres d'erreur sur la figure 4.9 représentent les intervalles $[-\sigma_{\rho_k}, \sigma_{\rho_k}]$ où $\sigma^2_{\rho_k} = \operatorname{var} \widehat{\rho}_k$, comme nous l'avons expliqué au § 4.3.3 ; les expressions analytiques de la moyenne et de la variance en fonction de celles de a_k et b_k sont données à l'annexe B.1.

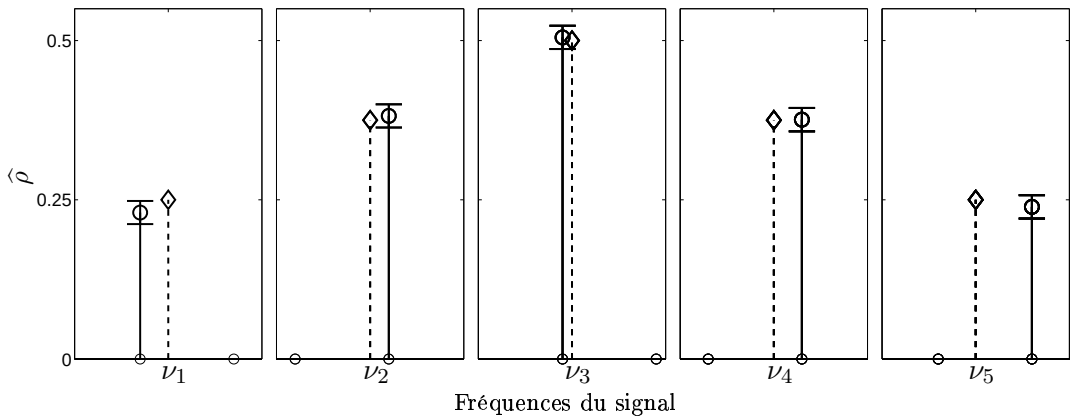


FIG. 4.9 – *Module du spectre estimé pour un modèle Bernoulli-Gaussien sur une grille de* $P+1 = 501$ *fréquences : agrandissement autour de chaque fréquence détectée. En trait plein, raie estimée et barres d'erreur associées à plus ou moins l'écart-type. Les* $\diamond$ *représentent les raies théoriques. La largeur de chaque fenêtre est de* $2\frac{f_{max}}{P} = 10^{-3}\ j^{-1}$

4.6 Prise en compte de décalages fréquentiels

Nous avons montré au long de ce chapitre que la modélisation Bernoulli-Gaussienne des amplitudes spectrales permet de restaurer une solution parcimonieuse pour le problème d'analyse spectrale modélisé par la relation linéaire (4.1). Une démarche d'estimation séquentielle basée sur la moyenne *a posteriori* de la séquence de Bernoulli d'une part, puis l'estimation en un second temps des amplitudes associées, ont notamment permis de traiter de manière satisfaisante le problème de *détection* de raies spectrales, en associant un niveau de confiance à chaque détection éventuelle.

Néanmoins, les simulations du § 4.5 ont mis en évidence une limitation fortement pénalisante, provoquée par la contrainte de discrétisation des fréquences sur la grille fixée $\mathcal{G} = \{\frac{k}{P} f_{\max}\}_{k=0\ldots P}$. Ainsi, l'estimateur proposé jusqu'ici ne peut détecter deux raies que si elles sont séparées d'un minimum de deux positions sur la grille, puisque l'inexactitude de la discrétisation fréquentielle peut générer des valeurs de $\boldsymbol{q}$ non nulles de part et d'autre de chaque fréquence théorique. Ce dédoublement des raies, s'il est partiellement corrigible par une interprétation postérieure des résultats, conserve toutefois un aspect déplaisant et limite la *résolution* de la méthode qui est alors de $2f_{\max}/P$. La *précision* de l'estimation est quant à elle intrinsèquement fixée par le pas de la grille, soit $f_{\max}/P$. Cette limitation est d'ailleurs commune à toutes les approches formalisant l'analyse spectrale sous la forme non paramétrique du modèle linéaire (4.1) ; c'est donc également le cas pour l'approche pénalisée étudiée au chapitre 3.

L'amélioration de ces deux caractéristiques (résolution et précision) ne peut alors se faire qu'au prix de la hausse du nombre d'inconnues fixé par la valeur de P, qui ne peut être augmenté indéfiniment : outre les problèmes de blocage algorithmique liés à une discrétisation trop fine, pour lesquels nous avons proposé des solutions alternatives, le coût calculatoire requis par les algorithmes MCMC mis en œuvre devient rapidement prohibitif lorsque le nombre de composantes spectrales augmente. En pratique, un nombre P de paramètres de l'ordre du millier nous semble être une borne supérieure pour assurer un temps d'exécution raisonnable des algorithmes précédents (sur ces simulations et au sens du critère d'arrêt (4.15), la procédure d'échantillonnage

pour $P = 1000$ paramètres B-G nécessite un temps de calcul d'environ 2200 s).

Afin de s'affranchir des problèmes inhérents à une discrétisation fixée des fréquences du modèle (4.1), la thèse de Dublanchet [1996] propose la prise en compte d'un modèle Bernoulli-Gaussien plus complet, autorisant des décalages fréquentiels par rapport aux positions de la grille :

$$f'_0 = 0 \;\text{ et }\; \forall k = 1 \dots P, \quad f'_k = \frac{k}{P} f_{\max} - df_k, \;\text{ avec } df_k \in \Delta f \triangleq \left[0, \frac{f_{\max}}{P}\right],$$
$$i.e., \quad f'_k \in \mathcal{I}_k \triangleq \left[\frac{k-1}{P} f_{\max}, \frac{k}{P} f_{\max}\right].$$

Le problème correspondant s'écrit à présent sous la forme :

$$\boldsymbol{y} = \mathbf{H}_{\mathrm{df}} \boldsymbol{s} + \boldsymbol{\epsilon} \tag{4.16}$$

où $\mathbf{H}_{\mathrm{df}}$ représente la matrice d'éléments $\mathbf{H}_{\mathrm{df}_{n,2k}} = \cos(2\pi f'_k t_n) = \cos\left(2\pi(\frac{k}{P} f_{\max} - df_k) t_n\right)$ et $\mathbf{H}_{\mathrm{df}_{n,2k+1}} = \sin(2\pi f'_k t_n) = \sin\left(2\pi(\frac{k}{P} f_{\max} - df_k) t_n\right)$ pour $k = 0 \dots P$ et $n = 1 \dots N$. Les paramètres $f'_{k,\,k=1\dots P}$ évoluant à présent dans un domaine *continu*, cette modélisation permet d'envisager une précision en fréquence théoriquement illimitée de l'estimation des raies spectrales et la valeur de P fixe alors uniquement sa résolution à $f_{\max}/P$: une seule fréquence peut être restaurée dans chaque intervalle $\mathcal{I}_k$. Notons que sous cette formulation, le modèle spectral conserve une composante de fréquence nulle f'_0, qui permet de prendre en compte l'inexactitude du centrage des données, causée par l'irrégularité de l'échantillonnage (voir le § 2.1.2). Afin d'utiliser des notations générales, nous noterons $\mathcal{I}_0 = \{0\}$; les spécificités liées à l'estimation de a_0, l'amplitude spectrale associée, ont déjà été décrites au long de ce chapitre et ne seront pas détaillées par la suite.

4.6.1 Description du modèle

En conservant une structure sous-jacente Bernoulli-Gaussienne sur les amplitudes spectrales sous la forme composite ($\boldsymbol{q}$,$\boldsymbol{s}$), nous définissons le modèle Bernoulli-Gaussien avec décalages, noté par la suite B-G$_{\boldsymbol{df}}$, comme un processus aléatoire composite ($\boldsymbol{q}$,$\boldsymbol{s}$,$\boldsymbol{df}$) où :

- $\boldsymbol{q} = [q_0, \dots q_P]^{\mathrm{t}}$ est une séquence blanche de Bernoulli de paramètre $\lambda \in [0,1]$ conditionnant la présence d'une raie spectrale dans chaque intervalle $\mathcal{I}_{k,\,k=0\dots P}$:

  $$\begin{cases} \Pr(q_k = 1) = \lambda, \text{ probabilité } \textit{a priori} \text{ d'existence d'une raie de fréquence } f'_k \in \mathcal{I}_k, \\ \Pr(q_k = 0) = 1 - \lambda. \end{cases}$$

- Les décalages $\boldsymbol{df}$ sont supposés indépendants et uniformément répartis dans l'intervalle $\Delta f = [0, \frac{f_{\max}}{P}]$:

  $$\text{pour } k = 1 \dots P, \quad p(df_k) = \frac{P}{f_{\max}} \mathbb{1}_{\Delta f}(df_k).$$

  Sur le plan informationnel, la variable df_k ne prend un sens que conditionnellement à l'existence d'une raie dans l'intervalle $\mathcal{I}_k$, c'est-à-dire lorsque $q_k = 1$. Néanmoins, cette modélisation supposant l'*indépendance* des deux processus $\boldsymbol{q}$ et $\boldsymbol{df}$ permet de définir df_k $\forall k = 0 \dots P$, et simplifie alors la description du modèle. Une utilisation appropriée des variables df_k sera réalisée à l'étape d'estimation, que nous envisagerons au § 4.6.2.

– Conditionnellement à la localisation apportée par la séquence $\boldsymbol{q}$, les amplitudes $(a_k|\, q_k = 1)$ et $(b_k|\, q_k = 1)$ correspondant à la raie spectrale présente dans l'intervalle $\mathcal{I}_k$ sont supposées *i.i.d.* gaussiennes centrées et de même variance σ^2. Comme pour le modèle initial B-G$_{\mathbf{0}}$, nous pouvons écrire $(a_k|q_k), (b_k|q_k) \sim \mathcal{N}(0,q_k\sigma^2)$, soit encore sous la forme bidimensionnelle :

$$p(\boldsymbol{s}_k|q_k) = g_2(\boldsymbol{s}_k,q_k\sigma^2\mathbf{I}_2),$$

Notons que cette définition des amplitudes suppose l'*indépendance* entre $\boldsymbol{s}$ et $\boldsymbol{df}$ plutôt qu'une définition des amplitudes conditionnée par $\boldsymbol{df}$. On a ici : « sachant $q_k = 1$, $\boldsymbol{s}_k$ représente l'amplitude de la raie présente *dans l'intervalle* $\mathcal{I}_k$ » et non pas : « sachant $q_k = 1$ *et* df_k, $\boldsymbol{s}_k$ représente l'amplitude de la raie présente *à la fréquence* $f'_k = \frac{k}{P} f_{max} - df_k$ ». Bien que la seconde définition soit de prime abord plus intuitive, le choix de la première est effectué pour des raisons algorithmiques qui seront détaillées au § 4.6.3.

Le problème de l'estimation de l'ensemble des paramètres du modèle B-G$_{\boldsymbol{df}}$ peut alors être envisagé de manière similaire au cas du modèle B-G$_{\mathbf{0}}$. Sous le paradigme Bayésien et en supposant les perturbations $\boldsymbol{\epsilon}$ additives, *i.i.d.* et gaussiennes de variance σ_ϵ^2, la loi *a posteriori* s'écrit dorénavant :

$$\begin{aligned} p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{df}|\boldsymbol{y},\boldsymbol{\theta}) &\propto p(\boldsymbol{y}|\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta})\, p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{df}|\boldsymbol{\theta}) \\ &\propto \lambda^{N_e}(1-\lambda)^{P+1-N_e}\ g_N(\boldsymbol{y} - \mathbf{H}_{\mathrm{df}}\,\boldsymbol{s}, \sigma_\epsilon^2\mathbf{I}_N)\ g_{P+1}(\boldsymbol{s},\boldsymbol{\Sigma}_{s|q}) \\ &\quad \times \prod_{k=1}^{P} \mathbb{1}_{\Delta f}\,(df_k) \end{aligned} \tag{4.17}$$

où, comme précédemment, N_e est le nombre de composantes non nulles de $\boldsymbol{q}$, $\boldsymbol{\Sigma}_{s|q} = \sigma^2\mathrm{diag}(\underline{\boldsymbol{q}})$ et $\boldsymbol{\theta}$ regroupe les trois hyperparamètres $\boldsymbol{\theta} = [\lambda,\sigma_\epsilon^2,\sigma^2]$.

4.6.2 Principe de l'estimation

Cet enrichissement du modèle spectral se fait cependant au prix d'une considérable augmentation en complexité associée à l'estimation de $(\boldsymbol{q},\boldsymbol{s},\boldsymbol{df})$, de par la dépendance non linéaire de la définition (4.16) en les variables additionnelles $\boldsymbol{df}$. Pour une stratégie de type maximum *a posteriori*, nous avons vu dans le cas du modèle B-G$_{\mathbf{0}}$ que l'estimation de la séquence $\widehat{\boldsymbol{q}}_{\mathrm{MAP}}$ constitue en soi un problème d'optimisation combinatoire rendu délicat par la dimension de l'espace des paramètres. La non-linéarité du modèle (4.16) en $\boldsymbol{df}$ complique alors toute procédure d'optimisation conjointe et conduit naturellement Dublanchet [1996] à considérer, par analogie avec le modèle B-G$_{\mathbf{0}}$, une démarche séquentielle visant à estimer, dans un premier temps, la séquence de Bernoulli au sens du MAP, puis à procéder à l'estimation MAP des décalages fréquentiels et des amplitudes associées. Néanmoins, outre les problèmes liés à l'estimation MAP de $\boldsymbol{q}$ (voir le § 4.2), une telle approche n'est pas totalement satisfaisante. D'une part, elle considère l'estimation de la séquence $\boldsymbol{q}$ correspondant à un modèle à fréquences *fixées* sur la grille $\mathcal{G}$, *i.e.*, $\boldsymbol{df} = \mathbf{0}$, et de ce fait reste soumise aux limitations évoquées au paragraphe précédent concernant le dédoublement des raies, alors que l'introduction des variables $\boldsymbol{df}$ a en partie pour objectif d'échapper à ce problème causé par l'inexactitude de la discrétisation. D'autre part, la seconde étape d'estimation des décalages df_k correspondant aux localisations approximatives fournies par $\widehat{q}_{k_{\mathrm{MAP}}} = 1$ est à son tour délicate. Le recours à des procédures d'optimisation locales (voir le § 2.2.3) semble cependant plus satisfaisant que la solution proposée par Dublanchet [1996], basée sur une linéarisation au

premier ordre en $\boldsymbol{df}$ du modèle (4.16) : l'approximation résultante est théoriquement différente selon que la fréquence associée s'approche de la fréquence nulle ($k = 0$) ou de la fréquence maximale ($k = P$).

À l'inverse, le cadre présenté au long de ce chapitre, visant à réaliser l'estimation au sens de la moyenne *a posteriori* par la mise en œuvre d'algorithmes d'échantillonnage stochastique, peut être étendu à l'estimation de l'ensemble des paramètres B-G$_{\boldsymbol{df}}$ à partir de l'échantillonnage de la loi (4.17). Une telle approche conjointe permet alors de dépasser les limites propres à une approche de type optimisation, que son nécessaire caractère séquentiel rend sous-optimale : l'estimation des décalages ne remet pas en cause la localisation fournie par la séquence $\boldsymbol{q}$. Pour une procédure non supervisée, nous envisagerons plus exactement l'échantillonnage de la loi $p(\boldsymbol{q}, \boldsymbol{s}, \boldsymbol{df}, \boldsymbol{\theta} \mid \boldsymbol{y})$. Les hyperparamètres jouant des rôles similaires dans les modèles B-G$_{\mathbf{0}}$ et B-G$_{\boldsymbol{df}}$, les considérations du § 4.3.4 à leur sujet restent valables, et l'inclusion de $\boldsymbol{\theta}$ à l'algorithme d'échantillonnage s'effectuera ici de la même manière. Pour cette raison, la dépendance en $\boldsymbol{\theta}$ des différentes distributions ne sera pas explicitée ici excepté lorsqu'elle s'avèrera nécessaire.

En supposant l'obtention d'échantillons $(\boldsymbol{q}^{(t)}, \boldsymbol{s}^{(t)}, \boldsymbol{df}^{(t)})$ distribués selon $p(\boldsymbol{q}, \boldsymbol{s}, \boldsymbol{df} \mid \boldsymbol{y})$ – nous proposerons au § 4.6.3 une solution algorithmique à ce sujet – nous pouvons envisager une procédure d'estimation selon le schéma suivant :

1. l'estimation de $\mathbb{E}[\boldsymbol{q}|\boldsymbol{y}]$ par la moyenne $\frac{1}{T}\sum_t \boldsymbol{q}^{(t)}$ ainsi que celle des hyperparamètres $\widehat{\boldsymbol{\theta}}_{\text{MP}} = \frac{1}{T}\sum_t \boldsymbol{\theta}^{(t)}$;
2. l'interprétation du résultat en termes de *probabilités de détection*, et la sélection des fréquences significatives par *seuillage* des valeurs de $\widehat{\boldsymbol{q}}_{\text{MP}}$, produisant une séquence de Bernoulli $\widehat{\boldsymbol{q}}_\alpha$: $\widehat{\boldsymbol{q}}_{\alpha_k} = 1$ si $\widehat{q}_{\text{MP}\,k} \geq \alpha$ et 0 sinon. Le niveau de seuil α est laissé au libre choix de l'expert ;
3. l'estimation des décalages en fréquence associés au sens de la moyenne *a posteriori* conditionnelle de $(\boldsymbol{df} \,|\, \boldsymbol{y}, \boldsymbol{q} = \widehat{\boldsymbol{q}}_\alpha)$:

$$\text{pour } k \text{ tel que } \widehat{\boldsymbol{q}}_{\alpha_k} = 1,\ \widehat{df_k} = \frac{1}{T_\alpha} \sum_{t\,|\boldsymbol{q}^{(t)} = \widehat{\boldsymbol{q}}_\alpha} df_k^{(t)} \ \text{ où } T_\alpha = \text{Card}\{t \,|\, \boldsymbol{q}^{(t)} = \widehat{\boldsymbol{q}}_\alpha\}$$

et l'estimation de la variance associée :

$$\text{var } \widehat{df_k} \simeq \frac{1}{T_\alpha - 1} \sum_{t\,|\boldsymbol{q}^{(t)} = \widehat{\boldsymbol{q}}_\alpha} \left(df_k^{(t)} - \widehat{df_k}\right)^2 ;$$

4. l'estimation des amplitudes correspondantes, au sens du maximum (ou de manière équivalente de la moyenne) de la loi gaussienne $p(\boldsymbol{s} \,|\, \boldsymbol{q} = \widehat{\boldsymbol{q}}_\alpha, \widehat{\boldsymbol{df}}, \boldsymbol{y})$, fournie par l'expression analytique [Idier *et al.*, 2004] :

$$\widehat{\boldsymbol{s}}_\alpha = \mathbb{E}[\boldsymbol{s} \,|\, \widehat{\boldsymbol{q}}_\alpha, \widehat{\boldsymbol{df}}, \boldsymbol{y}] = \boldsymbol{\Sigma}_{s|\widehat{q}_\alpha} \mathbf{H}_{\widehat{\text{df}}}{}^{\text{t}}\, \boldsymbol{\Sigma}_{y|\widehat{q}_\alpha,\widehat{df}}{}^{-1}\, \boldsymbol{y}$$

$$\text{où } \boldsymbol{\Sigma}_{s|\widehat{q}_\alpha} = \sigma^2 \text{diag}(\underline{\underline{\widehat{\boldsymbol{q}}}}_\alpha) \text{ et } \boldsymbol{\Sigma}_{y|\widehat{q}_\alpha,\widehat{df}} = \sigma_\epsilon^2 \mathbf{I}_N + \mathbf{H}_{\widehat{\text{df}}}\, \boldsymbol{\Sigma}_{s|\widehat{q}_\alpha} \mathbf{H}_{\widehat{\text{df}}}{}^{\text{t}}$$

en utilisant les valeurs estimées à l'étape 1 pour les hyperparamètres σ^2 et σ_ϵ^2. Nous calculons également la matrice de covariance associée :

$$\begin{aligned} \text{Cov}\,[\,\boldsymbol{s} \,|\, \widehat{\boldsymbol{q}}_\alpha, \widehat{\boldsymbol{df}}, \boldsymbol{y}] &= \mathbb{E}[(\boldsymbol{s} - \widehat{\boldsymbol{s}}_\alpha)(\boldsymbol{s} - \widehat{\boldsymbol{s}}_\alpha)^{\text{t}} | \, \widehat{\boldsymbol{q}}_\alpha, \widehat{\boldsymbol{df}}, \boldsymbol{y}] \\ &= \boldsymbol{\Sigma}_{s|\widehat{q}_\alpha} - \boldsymbol{\Sigma}_{s|\widehat{q}_\alpha} \mathbf{H}_{\text{df}}^{\text{t}}\, \boldsymbol{\Sigma}_{y|\widehat{q}_\alpha,\widehat{df}}^{-1}\, \mathbf{H}_{\widehat{\text{df}}}\, \boldsymbol{\Sigma}_{s|\widehat{q}_\alpha}. \end{aligned}$$

En particulier, les éléments diagonaux de cette matrice sont les variances $\text{var}\,\widehat{a}_k, \text{var}\,\widehat{b}_k$ où $\widehat{\boldsymbol{s}}_k = [\widehat{a}_k, \widehat{b}_k]$.

TAB. 4.5 – *Procédure d'estimation de* $(\boldsymbol{q},\boldsymbol{s},\boldsymbol{df})$ *à partir de l'échantillonnage de la loi a posteriori* $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{df}\,|\,\boldsymbol{y})$.

Nous signalons au sujet de l'estimation des amplitudes que, les paramètres d'intérêt étant surtout les modules $\rho_k = \sqrt{a_k^2 + b_k^2}$, l'estimation de leur moyenne et variance sera réalisée à partir de celles des paramètres a_k et b_k, calculées à l'étape 3 de la procédure du tableau 4.5 : les expressions analytiques effectuant ce lien sont données à l'annexe B.1. Remarquons encore que, contrairement au cas du modèle B-$\text{G}_\mathbf{0}$, nous ne pouvons plus ici estimer la moyenne des paramètres d'amplitude a_k, b_k ou ρ_k au sens de la MP à partir des échantillons $\boldsymbol{s}^{(t)}$, puisque ceux-ci correspondent, pour chaque itération t, à une fréquence différente $f_k'^{(t)} = \frac{k}{P} f_{\max} - df_k^{(t)}$. Les échantillons en amplitude $\boldsymbol{s}^{(t)}$ n'interviennent donc pas dans l'étape d'estimation et, comme pour le modèle B-$\text{G}_\mathbf{0}$, les motivations poussant à l'échantillonnage selon la loi jointe (4.17) plutôt qu'à celui de la loi marginalisée en amplitude sont d'ordre technique : l'échantillonnage des

variables $p(q_k,\boldsymbol{s}_k|\boldsymbol{y},\boldsymbol{q}_{-k},\boldsymbol{x}_{-k},\boldsymbol{df})$ est bien moins coûteux que l'échantillonnage de $\Pr(q_k|\boldsymbol{y},\boldsymbol{q}_{-k},\boldsymbol{df})$.

L'estimateur produit par une telle méthode permet alors d'atteindre la précision des méthodes à fréquences *continues*, habituellement propres aux modèles paramétriques, tout en conservant les avantages fournis par la structure Bernoulli-Gaussienne, à savoir :

- l'obtention, conjointement aux estimés, de niveaux de confiance associés. La séquence $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ s'interprète comme la probabilité de détection d'une raie dans chaque intervalle $\mathcal{I}_k$, et l'accès aux variances associées aux estimateurs $\widehat{\boldsymbol{df}}$ et $\widehat{\boldsymbol{s}}$ permettent d'en mesurer l'incertitude ;
- l'estimation du nombre de raies, qui est ici un sous-produit de l'analyse de la chaîne $\boldsymbol{q}^{(t)}$ alors que la détermination de l'ordre d'un modèle paramétrique engendre en général des complications supplémentaires. Andrieu et Doucet [1999] proposent l'estimation conjointe de l'ordre et des paramètres d'un modèle spectral de type Poisson-Gaussien au moyen d'un algorithme MCMC et permettant également d'associer une interprétation statistique à l'estimation de l'ordre. Nous avons étudié l'estimateur proposé pour lequel, les paramètres fréquentiels évoluant sur un domaine continu, de hauts niveaux de précision et de résolution peuvent être obtenus ; cependant, le recours à des algorithmes à sauts réversibles pour pallier l'ignorance de l'ordre alourdit la procédure d'échantillonnage, mais surtout rend bien plus délicate l'exploitation des résultats [Bourguignon, 2002] ;
- une plus grande robustesse face aux erreurs de modélisation par rapport aux méthodes paramétriques. En particulier, la structure du modèle B-G$_{\boldsymbol{df}}$ impose la reconstruction d'un nombre fixe de raies éventuelles ne pouvant se chevaucher : il y a au plus une raie par intervalle $\mathcal{I}_k$. Une telle structure est alors moins sensible aux fréquentes perturbations des données par un spectre continu de basses fréquences que nous avons déjà évoquées à plusieurs reprises. La taille des intervalles $\mathcal{I}_k$, déterminée par le paramètre P, permet alors de régler le compromis entre cette robustesse et la résolution de la méthode, qui reste structurellement limitée à $f_{\max}/P$.

4.6.3 Mise en œuvre : algorithme hybride

Si la modélisation (4.16) permet en théorie de dépasser les limites intrinsèques à la modélisation Bernoulli-Gaussienne classique, l'introduction des variables supplémentaires $\boldsymbol{df}$ pénalise la mise en œuvre d'une procédure d'échantillonnage de la loi *a posteriori* correspondante. En effet, l'inclusion dans l'algorithme de Gibbs classique d'une étape d'échantillonnage des décalages df_k n'est pas immédiate : les lois conditionnelles $p(df_k \,|\, \boldsymbol{df}_{-k}, \boldsymbol{q}, \boldsymbol{s}, \boldsymbol{y})$ ne sont pas simulables directement. En notant $\mathbf{H}_{\mathrm{df}_k}$ la matrice $N \times 2$ formée par les deux colonnes d'indices $2k$ et $2k+1$ de $\mathbf{H}_{\mathrm{df}}$ et $\boldsymbol{e}_k = \boldsymbol{y} - \sum_{j\neq k} \mathbf{H}_{\mathrm{df}_j}\boldsymbol{s}_j$, on peut écrire :

$$\begin{aligned} p(df_k \,|\, \boldsymbol{df}_{-k}, \boldsymbol{q}, \boldsymbol{s}, \boldsymbol{y}) &\propto \exp\left(-\frac{\|\boldsymbol{y} - \mathbf{H}_{\mathrm{df}}\,\boldsymbol{s}\|^2}{2\sigma_\epsilon^2}\right) \times \mathbb{1}_{\Delta f}\,(df_k) \\ &\propto \exp\left(-\frac{\|\boldsymbol{e}_k - \mathbf{H}_{\mathrm{df}_k}\boldsymbol{s}_k\|^2}{2\sigma_\epsilon^2}\right) \times \mathbb{1}_{\Delta f}\,(df_k) \end{aligned} \tag{4.18}$$

Remarquons que lorsque $q_k = 0$, on a $\boldsymbol{s}_k = 0$ et $p(df_k \,|\, \boldsymbol{df}_{-k}, q_k = 0, \boldsymbol{q}_{-k}, \boldsymbol{s}, \boldsymbol{y})$ est la distribution uniforme sur l'intervalle Δf. Par contre, lorsque $q_k = 1$ – et c'est le seul cas qui nous intéresse –, il est inenvisageable de générer directement un échantillon df_k selon la loi (4.18). Nous proposons alors de recourir à une version *hybride* de l'échantillonneur de Gibbs [Robert, 1996], où la

simulation selon (4.18) est remplacée par une étape d'algorithme de Hastings-Metropolis (HM) ayant cette même loi pour cible. À l'itération t de l'échantillonnage, cette étape consiste à :

1. générer un candidat φ selon une loi instrumentale $\Phi(\varphi \,|\, df_k^{(t-1)})$;
2. prendre
   - $df_k^{(t)} = \varphi$ avec la probabilité $\rho_{\text{HM}}(df_k^{(t-1)}, \varphi)$
   - $df_k^{(t)} = df_k^{(t-1)}$ avec la probabilité $1 - \rho_{\text{HM}}(df_k^{(t-1)}, \varphi)$

   avec $\rho_{\text{HM}}(df_k^{(t-1)}, \varphi) = \min\left\{1, \dfrac{p(\varphi \,|\, \boldsymbol{df}_{-k}, \boldsymbol{q}, \boldsymbol{s}, \boldsymbol{y})}{p(df_k^{(t-1)} \,|\, \boldsymbol{df}_{-k}, \boldsymbol{q}, \boldsymbol{s}, \boldsymbol{y})} \dfrac{\Phi(df_k^{(t-1)} \,|\, \varphi)}{\Phi(\varphi \,|\, df_k^{(t-1)})}\right\}$

Même si l'étape HM ci-dessus génère asymptotiquement des échantillons distribués selon la loi d'intérêt quelle que soit la loi instrumentale – à condition que celle-ci puisse explorer l'intégralité du support Δf de (4.18) – il s'agit de choisir une loi Φ ne provoquant pas trop de rejets à l'étape 2 afin de ne pas pénaliser la vitesse de convergence de l'algorithme d'échantillonnage. En nous inspirant des choix effectués par Andrieu et Doucet [1999], nous proposons d'utiliser alternativement deux lois instrumentales Φ_1 et Φ_2, où $\Phi_1(\varphi)$ est une loi uniforme sur Δf et $\Phi_2(\varphi \,|\, df_k^{(t-1)})$ est une marche aléatoire gaussienne centrée sur l'échantillon courant $df_k^{(t-1)}$ et de faible variance, favorisant une exploration locale de la distribution cible. Afin d'assurer l'appartenance de chaque échantillon $df_k^{(t)}$ à l'intervalle Δf, il est nécessaire de restreindre le support de la marche aléatoire, de support infini, à cet intervalle. Φ_2 est donc de manière exacte une distribution gaussienne *tronquée*. La simulation répétitive selon une loi gaussienne jusqu'à l'obtention d'un candidat appartenant à Δf correspond à un algorithme d'*acceptation-rejet* [Robert, 1996] fournissant des variables distribuées selon Φ_2 et la faible valeur de l'écart-type de la gaussienne (que nous choisissons ici à un dixième de la taille de Δf) assure un fort taux d'acceptation. Notons que, ces deux lois étant symétriques, *i.e.*, $\Phi(\varphi \,|\, df_k^{(t-1)}) = \Phi(df_k^{(t-1)} \,|\, \varphi)$, la probabilité $\rho_{\text{HM}}(df_k^{(t-1)}, \varphi)$ se voit simplifiée du rapport instrumental et correspond uniquement au rapport des lois *a posteriori* (4.18) prises en φ et en $df_k^{(t-1)}$; en particulier, tout candidat φ augmentant la valeur de cette loi sera accepté.

Nous envisageons donc une procédure d'échantillonnage de Gibbs, couplée avec une étape de Hastings-Metropolis, visant à échantillonner la loi conjointe $p(\boldsymbol{q}, \boldsymbol{s}, \boldsymbol{df}, \boldsymbol{\theta} \,|\, \boldsymbol{y})$. Comme nous l'avons déjà signalé, l'inclusion dans l'échantillonnage de Gibbs des hyperparamètres s'effectue de manière similaire au cas B-G_0 que nous avons étudié au §4.4.1. L'algorithme proposé est décrit au tableau 4.6, où il est fait usage des notations $\boldsymbol{u}_{i\to j}^{(t)} = (u_i^{(t)}, u_{i+1}^{(t)}, \ldots, u_j^{(t)})$ et $\boldsymbol{u}_{k,\ell}^{(t,t-1)} = (u_{1\to k}^{(t)}, u_{\ell\to P}^{(t-1)})$.

o. Initialisation $t = 0$: $\left(\boldsymbol{q}^{(0)}, \boldsymbol{s}^{(0)}\, \boldsymbol{df}^{(0)}\right) = (\mathbf{0},\mathbf{0},\mathbf{0})$ et $\boldsymbol{\theta}^{(0)} = (0.1\,;1\,;1)$.
Pour les itérations $t = 1 \ldots T$, effectuer les tirages aléatoires suivants :

i. Pour $k = 0 \ldots P$:

a) échantillonner les paramètres B-G :

$$\boxed{q_k^{(t)}, \boldsymbol{s}_k^{(t)} \sim p\left(q_k, \boldsymbol{s}_k \,|\, \boldsymbol{q}_{k-1,k+1}^{(t,t-1)}, \boldsymbol{s}_{k-1,k+1}^{(t,t-1)}, \boldsymbol{df}_{k-1,k}^{(t,t-1)}, \boldsymbol{\theta}^{(t-1)}, \boldsymbol{y}\right)}$$

Cette loi revêt une expression similaire à celle explicitée pour le modèle B-G$_\mathbf{0}$ à l'algorithme du tableau 4.2 en considérant $\mathbf{H}_{\mathrm{df}}$ au lieu de $\mathbf{H}$, avec $\boldsymbol{df} = \boldsymbol{df}_{k-1,k}^{(t,t-1)}$.

b) pour $k \geq 1$, réaliser pour $df_k^{(t)}$ une étape de Hastings-Metropolis ayant pour loi cible :

$$\boxed{df_k^{(t)} \sim p\left(df_k \,|\, \boldsymbol{q}_{k-1,k+1}^{(t,t-1)}, q_k^{(t)}, \boldsymbol{s}_{k-1,k+1}^{(t,t-1)}, \boldsymbol{s}_k^{(t)}, \boldsymbol{\theta}^{(t-1)}, \boldsymbol{y}\right)}$$

en utilisant, avec une probabilité $1/2$, la loi instrumentale uniforme $\Phi_1(\varphi) = \mathbb{1}_{\Delta f}(\varphi)$ et, avec une probabilité $1/2$, la marche aléatoire tronquée $\Phi_2(\varphi \,|\, df_k^{(t-1)}) \propto g(\varphi - df_k^{(t-1)}, \sigma_\Phi^2)\mathbb{1}_{\Delta f}(\varphi)$. L'écart-type de Φ_2 sera fixé par la suite à $\sigma_\Phi = f_{\max}/10P$.

ii. Échantillonner les hyperparamètres :

$$\boxed{\begin{array}{rcl} \lambda^{(t)} & \sim & p\left(\lambda \,|\, \boldsymbol{q}^{(t)}, \boldsymbol{s}^{(t)}, \boldsymbol{df}^{(t)}, \sigma_\epsilon^{2(t-1)}, \sigma^{2(t-1)}, \boldsymbol{y}\right) \\ \sigma_\epsilon^{2(t)} & \sim & p\left(\sigma_\epsilon^2 \,|\, \boldsymbol{q}^{(t)}, \boldsymbol{s}^{(t)}, \boldsymbol{df}^{(t)}, \lambda^{(t)}, \sigma^{2(t-1)}, \boldsymbol{y}\right) \\ \sigma^{2(t)} & \sim & p\left(\sigma^2 \,|\, \boldsymbol{q}^{(t)}, \boldsymbol{s}^{(t)}, \boldsymbol{df}^{(t)}, \lambda^{(t)}, \sigma_\epsilon^{2(t)}, \boldsymbol{y}\right) \end{array}}$$

Ces trois lois conditionnelles revêtent des expressions similaires à celles explicitées à l'algorithme du tableau 4.2 en considérant $\mathbf{H}_{\mathrm{df}}$ au lieu de $\mathbf{H}$, avec $\boldsymbol{df} = \boldsymbol{df}^{(t)}$.

TAB. 4.6 – *Algorithme hybride d'échantillonnage de* $p(\boldsymbol{q}, \boldsymbol{s}, \boldsymbol{df}, \boldsymbol{\theta} \,|\, \boldsymbol{y})$.

Plusieurs remarques relatives à cette procédure sont à effectuer :

- Le cas particulier $k = 0$ n'est pas détaillé ici mais donne lieu aux mêmes spécificités que l'échantillonneur de Gibbs classique pour l'échantillonnage des paramètres $(q_0, \boldsymbol{s}_0)$.
- L'étape **i.**b) d'échantillonnage de la loi *a posteriori* de $df_k^{(t)}$ prend un sens uniquement lorsque $q_k^{(t)} \neq 0$. Dans le cas contraire, cette loi est uniforme et son échantillonnage n'est d'aucun intérêt pour l'estimation des décalages, qui sera réalisée en moyennant uniquement sur les échantillons correspondant à la séquence de Bernoulli $\boldsymbol{q}^{(t)} = \widehat{\boldsymbol{q}}_\alpha$, c'est-à-dire des échantillons pour lesquels $q_k^{(t)} = 1$ (voir le § 4.6.2). La réalisation de cet échantillonnage « artificiel » a pour unique but de garantir la définition des $df_k^{(t)}$ $\forall\, k,t$ et de permettre ainsi un usage correct de la marche aléatoire Φ_2 dont la définition nécessite l'existence, à l'itération t, des $df_k^{(t-1)}$.
- La génération d'un décalage en fréquence df_k selon la loi conditionnelle fixant l'amplitude $\boldsymbol{s}_k$ qui lui est associée peut paraître inappropriée à la hiérarchie du modèle : la démarche inverse semble en effet plus naturelle, visant à l'échantillonnage des amplitudes conditionnellement

à la connaissance de la fréquence *exacte* correspondante $f'_k = \frac{k}{P} f_{\max} - df_k$. Il faut noter ici que l'introduction des variables d'amplitude a pour objectif l'échantillonnage à moindre coût des paramètres B-G : nous avons vu au §4.3.2 que les lois conditionnelles $p(q_k, \boldsymbol{s}_k \mid reste)$ sont simulables bien plus facilement que les lois marginales $\Pr(q_k \mid reste)$. Ce choix impose donc un schéma algorithmique du type de celui décrit au tableau 4.6 et correspond, en termes de modélisation, à supposer l'*indépendance* entre $\boldsymbol{s}$ et $\boldsymbol{df}$. Le risque majeur associé à cette hypothèse est alors celui d'un trop fort taux de rejet à l'étape i.(b), provoqué par l'inadéquation, par rapport aux données, entre la fréquence déterminée par df_k et l'amplitude $\boldsymbol{s}_k$ qui lui est imposée. Néanmoins, si P est suffisamment grand, il est raisonnable de considérer que l'amplitude modélisant correctement une raie dans l'intervalle $\mathcal{I}_k$ est à peu près stable sur la longueur de cet intervalle.

4.7 Résultats d'estimation sur le modèle BG avec prise en compte de décalages

Nous illustrons les caractéristiques de l'estimation décrite au paragraphe précédent sur le signal multi-sinusoïdal habituel (celui de la figure 2.6). Nous considérons dans un premier temps une grille de $P + 1 = 501$ fréquences, à laquelle n'appartiennent pas les fréquences recherchées. L'échantillonnage de la loi *a posteriori* $p(\boldsymbol{q}, \boldsymbol{df}, \boldsymbol{\theta} | \boldsymbol{y})$ est mis en œuvre selon le schéma hybride décrit par l'algorithme du tableau 4.6, et est arrêté lorsque la moyenne *a posteriori* (MP) estimée de l'ensemble des paramètres $(\boldsymbol{q}, \boldsymbol{\theta})$ ne varie pratiquement plus, de manière similaire au critère (4.15) utilisé pour l'échantillonnage des paramètres B-$\mathrm{G}_{\mathbf{0}}$.

La figure 4.10 montre l'estimation de la moyenne *a posteriori* de la séquence de Bernoulli après convergence de l'algorithme, qui prend des valeurs faibles en la plupart des fréquences de la grille. La figure de droite montre que les cinq seules valeurs élevées de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$, qui sont égales à 1, localisent parfaitement les intervalles auxquels appartiennent les fréquences théoriques. En effet, au sens du modèle B-$\mathrm{G}_{\boldsymbol{df}}$, la variable de Bernoulli q_k, positionnée à la fréquence $\frac{k}{P} f_{\max}$, conditionne l'existence d'une raie dans l'intervalle $\mathcal{I}_k = [\frac{k-1}{P} f_{\max}, \frac{k}{P} f_{\max}]$.

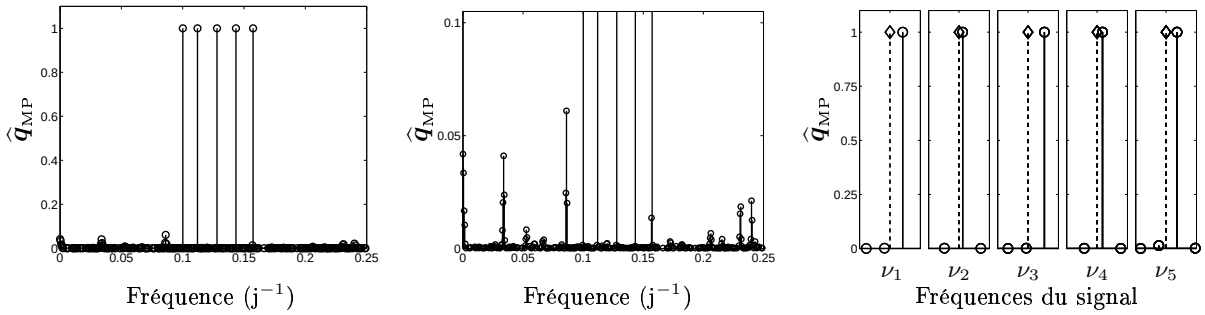


FIG. 4.10 – *Estimation $\widehat{\boldsymbol{q}}_{MP}$ pour un modèle $B-G$ avec décalages sur un signal multi-sinusoïdal. À gauche, sur l'ensemble de la grille fréquentielle. Au centre, zoom en amplitude pour les faibles valeurs. À droite, zoom en fréquence autour de chaque raie. Les $\diamond$ représentent la position fréquentielle des raies théoriques.*

La deuxième phase d'estimation des décalages en fréquence au sens de la MP conditionnellement à la localisation apportée par les cinq valeurs élevées de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$, puis celle des amplitudes

correspondantes, est réalisée selon la procédure du tableau 4.5. Les résultats sont présentés sur la figure 4.11, ainsi que les barres d'erreur que nous pouvons associer à l'estimation. En utilisant les notations du § 4.6, les fréquences sont localisées en $\widehat{f'_k} = \frac{k}{P} f_{\max} - \widehat{df_k}$ et les barres d'erreur horizontales représentent les intervalles $[-\sigma_{\mathrm{df}_k}, \sigma_{\mathrm{df}_k}]$ où $\sigma^2_{\mathrm{df}_k} = \widehat{\mathrm{var}}\, \widehat{f'_k} = \widehat{\mathrm{var}}\, \widehat{df_k}$, qui est estimée à partir des échantillons $df_k^{(t)}$. L'estimation des paramètres d'amplitude est réalisée conditionnellement à la connaissance des positions fréquentielles au sens du MAP (ou de manière identique de la MP) selon l'expression analytique du tableau 4.5, ainsi que celle des variances associées. En notant $\widehat{\boldsymbol{s}}_k = [\widehat{a}_k, \widehat{b}_k]$ les paramètres d'amplitude estimés pour chaque raie, nous représentons la moyenne *a posteriori* de $\rho_k = \sqrt{a_k^2 + b_k^2}$ estimée à partir de $\widehat{a}_k, \widehat{b}_k$ et les barres d'erreur verticales sur la figure 4.11 représentent les intervalles $[-\sigma_{\rho_k}, \sigma_{\rho_k}]$ où $\sigma^2_{\rho_k}$ est la variance associée à l'estimation de ρ_k : les expressions analytiques correspondantes sont détaillées à l'annexe B.1.

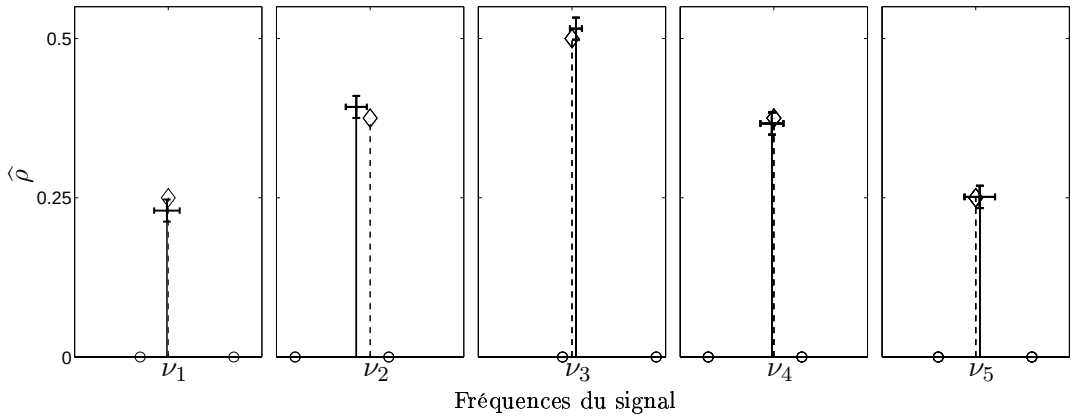
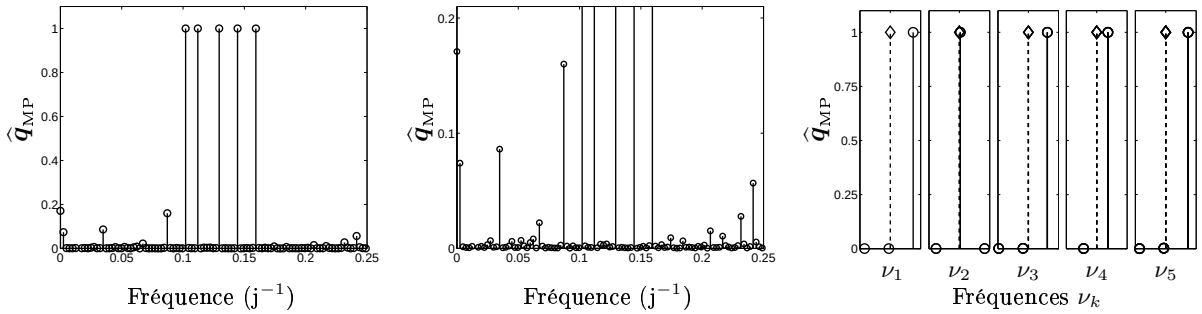


FIG. 4.11 – *Module du spectre estimé pour un modèle Bernoulli-Gaussien avec décalages en fréquence pour* $P = 500$ *: agrandissement autour de chaque fréquence détectée. En trait plein, raie estimée et barres d'erreur associées à plus ou moins l'écart-type. Les* $\diamond$ *représentent les raies théoriques et les* $\circ$ *les coordonnées de la grille* $\{\frac{k}{P} f_{max}\}_{k=0\ldots P}$*. La largeur de chaque fenêtre est de* $2\frac{f_{max}}{P} = 10^{-3}$ j^{-1}.

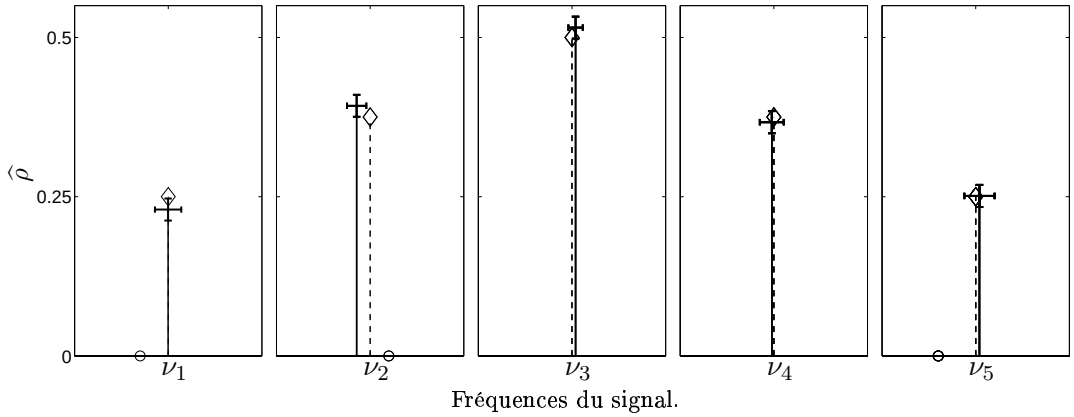
Au vu de cette figure, l'extension prenant en compte des variables *continues* de décalages fréquentiels permet donc d'augmenter considérablement la précision de l'estimation des fréquences par rapport au modèle Bernoulli-Gaussien classique. Sur cet exemple, l'erreur maximale commise sur l'estimation des fréquences est de 7.3×10^{-5} j^{-1} pour une erreur moyenne de 2.7×10^{-5} j^{-1}, alors que la précision inhérente à la modélisation B-G$_{\mathbf{0}}$ est celle du pas de la grille qui vaut ici $\frac{f_{\max}}{P} = 5 \times 10^{-4}$ j^{-1}. De plus, si l'amélioration des performances de l'estimation pour B-G$_{\mathbf{0}}$ ne peut s'effectuer qu'au prix de l'augmentation de la taille P de la grille, cette dernière ne limite que la résolution de B-G$_{\boldsymbol{df}}$ et il n'est donc pas nécessaire d'employer une grille arbitrairement fine pour améliorer sa précision. La diminution du nombre de paramètres, fixé par la dimension P, permet alors de limiter le coût calculatoire de la procédure d'échantillonnage stochastique tout en conservant une qualité satisfaisante d'estimation. Sur cet exemple, la réduction à $P+1 = 101$ fréquences discrétisées a ainsi permis de réduire le temps de calcul d'un facteur d'environ trois[8], celui-ci étant de l'ordre de 300 s pour $P = 100$ (les mises en œuvre ont été réalisées sur un ordinateur PC avec un processeur cadencé à 1.4 GHz et doté de 1 Go de mémoire vive). Sur un tel

8. Notons que le nombre d'itérations nécessaires de la procédure d'échantillonnage reste du même ordre dans les deux cas.

signal pour lequel les raies sont bien séparées en fréquence, l'estimation obtenue, représentée en figure 4.12, est de qualité tout à fait comparable à celle de la figure 4.11. Nous notons toutefois que les probabilités *a posteriori* de « fausse alarme », représentées par les valeurs de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ en dehors des cinq composantes égales à 1, sont dans ce deuxième cas plus élevées même si elles restent faibles : au vu de la figure 4.12 en haut, la plus grande probabilité correspondant à une « fausse alarme » est ici de 0.17 contre 0.06 dans le premier cas.



Première phase : estimation de la moyenne a posteriori $\widehat{\boldsymbol{q}}_{MP}$. À gauche, sur l'ensemble de la grille fréquentielle. Au centre, zoom en amplitude pour les faibles valeurs. À droite, zoom en fréquence autour de chaque raie. Les $\diamond$ représentent la position fréquentielle des raies théoriques.



Seconde phase : estimation des fréquences et des amplitudes associées. En trait plein, raie estimée en module et barres d'erreur associées à plus ou moins l'écart-type. Les $\diamond$ représentent les raies théoriques et les $\circ$ les coordonnées de la grille $\{\frac{k}{P} f_{max}\}_{k=0\ldots P}$. La largeur de chaque fenêtre est la même que sur la figure 4.11, soit 10^{-3} j^{-1}.

FIG. 4.12 – *Estimation obtenue pour un modèle Bernoulli-Gaussien avec décalages en fréquence pour $P = 100$.*

4.8 Conclusion

Nous avons étudié dans ce chapitre la prise en compte d'un modèle probabiliste de type Bernoulli-Gaussien sur les amplitudes spectrales, qui est naturellement adapté à l'estimation d'objets parcimonieux. Si Duvaut et Dublanchet [1995], Dublanchet [1996] ont adopté pour ce problème une approche par maximum *a posteriori*, nous avons suivi les travaux de Cheng *et al.* [1996] en déconvolution impulsionnelle proposant l'estimation des paramètres B-G au sens de

la moyenne *a posteriori* au moyen d'algoritmes MCMC, qui permet d'envisager une procédure totalement non supervisée. À la différence de ceux-ci, nous avons proposé d'estimer la moyenne *a posteriori* de la séquence de Bernoulli : l'estimée $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ obtenue n'est alors plus à valeurs binaires, mais permet une interprétation en termes de probabilité de détection d'une raie spectrale en chaque fréquence de la grille de reconstruction $\mathcal{G}$, information qui n'est pas accessible par une stratégie de type MAP. La sélection des fréquences significatives, et l'estimation postérieure des amplitudes associées, peut alors être effectuée par un seuillage adéquat en termes probabilistes. Notons dès à présent que si, sur les simulations présentées au cours de ce chapitre, l'estimation permet de localiser chaque raie spectrale en lui associant une probabilité proche de un, l'interprétation de ces probabilités sera plus délicate lors de l'application à des données plus complexes que nous présenterons au chapitre 5.

Sur l'ensemble des simulations réalisées ici, les différents choix de lois *a priori* envisagés sur les hyperparamètres se sont révélés sans grande influence sur leur estimation *a posteriori*, et par conséquent sur l'estimation des paramètres B-G. En particulier le choix de lois *a priori* impropres de Jeffreys sur les deux paramètres de variance, ou encore uniforme pour la variance du bruit, ont fourni des résultats satisfaisants et ne requièrent aucun réglage de l'utilisateur.

Au niveau algorithmique, nous avons d'abord étendu l'échantillonneur de Gibbs proposé par Cheng *et al.* [1996] au cas de variables bidimensionnelles associées à une même variable de Bernoulli. Sa mise en œuvre pour notre problème a révélé des lenteurs dues au piégeage des échantillons au voisinage d'un mode local de la distribution *a posteriori* et pouvant déboucher, au sens du critère d'arrêt de l'échantillonnage que nous avons utilisé, sur une estimation présentant un dédoublement d'une raie spectrale en des localisations incorrectes de part et d'autre de sa véritable valeur. Nous avons proposé deux alternatives : la première favorise la sortie de tels modes locaux au cours des premières itérations de « chauffage » de la procédure ; la seconde propose l'échantillonnage des paramètres B-G par blocs de trois composantes adjacentes. Des résultats satisfaisants ont été obtenus dans les deux cas et, si la seconde solution est plus onéreuse sur le plan calculatoire, elle est cependant plus élégante que la première qui nécessite le réglage supplémentaire du nombre d'itérations de chauffage.

Si l'estimation $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ permet sur les simulations présentées de localiser correctement les raies spectrales, nous notons cependant des complications causées par l'inexactitude de la discrétisation des fréquences sur la grille $\mathcal{G}$. L'alternance, au cours de l'échantillonnage, de valeurs non nulles des variables de Bernoulli en chacune des deux fréquences de la grille encadrant une fréquence théorique, débouche sur des valeurs adjacentes élevées de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$, ce phénomène étant d'autant plus marqué que les fréquences théoriques sont éloignées de la grille. L'interprétation postérieure de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ – les deux valeurs concernées étant alors de somme approximativement égale à 1 – permet cependant de déceler un tel effet d'alternance et de le prendre en compte pour réaliser correctement l'estimation finale des paramètres spectraux.

Les performances de résolution et de précision associées à cette modélisation restent toutefois intrinsèquement limitées à $2f_{\max}/P$, où $f_{\max}/P$ est le pas de discrétisation de la grille $\mathcal{G}$, comme c'est le cas de l'approche par pénalisation étudiée au chapitre 3. Si pour cette dernière une analyse à très haute résolution s'est avérée possible en choisissant P élevé, la contrainte calculatoire associée aux algorithmes d'échantillonnage stochastique rend difficilement envisageable leur utilisation dans des dimensions comparables : les 33 secondes nécessaires à la minimisation du critère pénalisé pour une grille composée d'un millier de fréquences positives sont à comparer à un temps d'environ 2200 secondes nécessaire pour l'échantillonnage des paramètres B-G. Nous avons alors repris l'extension proposée par Dublanchet [1996] à un modèle Bernoulli-Gaussien

autorisant des décalages en fréquence par rapport à la grille $\mathcal{G}$, mais que nous avons à nouveau envisagée dans le cadre de l'estimation MP : outre l'interprétation toujours valable de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ en termes probabilistes, cette approche nous a également permis d'estimer les décalages ainsi que leur variance de manière unifiée à partir de l'échantillonnage de la loi conjointe de l'ensemble des paramètres inconnus. Nous avons pour cela proposé un algorithme d'échantillonnage hybride qui a montré un comportement satisfaisant en simulation, résolvant notamment les problèmes d'alternance entre deux fréquences adjacentes liés à la nature discrète des fréquences du modèle B-G initial.

Cette extension, pour laquelle le pas de discrétisation $f_{\max}/P$ limite uniquement la résolution de l'estimateur, a donc permis d'améliorer considérablement la précision fréquentielle de la modélisation Bernoulli-Gaussienne tout en conservant un coût calculatoire raisonnable : pour une grille de $P+1 = 501$ fréquences positives, les résultats présentés plus haut, dont la précision dépasse celle obtenue pour la modélisation B-G initiale sur une grille deux fois plus fine, le temps calculatoire est de l'ordre du millier de secondes.

Nous verrons cependant en application aux données complexes du chapitre 5 que lorsque l'échantillonnage des fréquences associées aux décalages présente une trop forte variabilité, la localisation fréquentielle peut évoluer au fil des itérations dans plusieurs intervalles $\mathcal{I}_k$, ce qui pénalise considérablement la procédure d'estimation proposée.

Chapitre 5

Application au traitement de données réelles

Ce dernier chapitre est consacré à la mise en application des méthodes d'estimation que nous avons proposées sur un exemple de séries temporelles issues d'observations astrophysiques. Les données de la figure 5.1 ont déjà été présentées au premier chapitre de ce manuscrit : il s'agit de la courbe de vitesse radiale de l'étoile de Herbig HD 104237 obtenue à la suite d'observations réalisées au cours du mois d'avril 1999 à l'Observatoire Astronomique Sud-Africain (SAAO, *South African Astronomical Observatory*) couvrant une durée de cinq nuits [Böhm *et al.*, 2004]. Plus précisément, ces observations correspondent aux variations, en fonction du temps, de la vitesse radiale de la première composante d'un système multiple. Les mouvements orbitaux dus à la multiplicité du système provoquent des variations de la vitesse radiale de période relativement longue, auxquelles la nature pulsante du premier corps ajoute des oscillations de fréquences plus élevées et d'amplitudes plus faibles : l'allure générale de la courbe de la figure 5.1 à gauche est donc celle de la composante orbitale. Afin d'étudier la variabilité à « haute » fréquence, Böhm *et al.* [2004] ont proposé de corriger cette courbe de vitesse radiale en identifiant approximativement la contribution due au mouvement orbital. La composante orbitale extraite et la courbe de vitesse radiale corrigée sont présentées sur cette même figure, respectivement au centre et à droite.

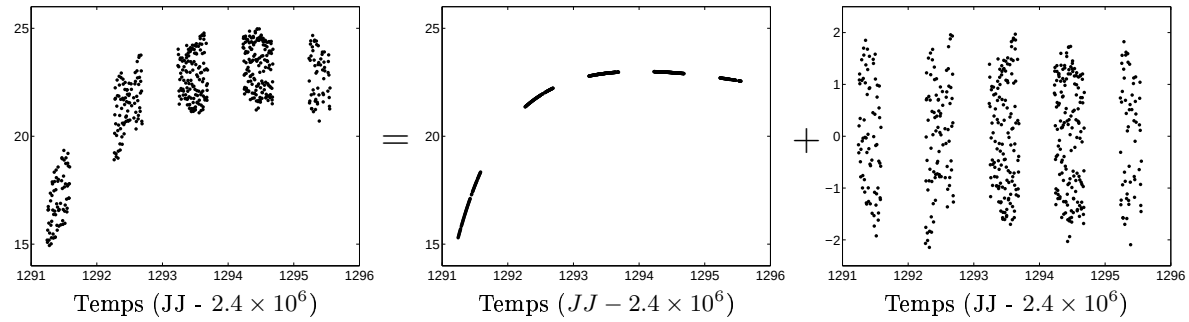


FIG. 5.1 – *Courbe de vitesse radiale de HD 104237. À gauche, données brutes ; au centre, approximation des variations à basse fréquence dues au mouvement orbital ; à droite, données corrigées correspondant aux données brutes desquelles ont été retranchées les variations à basse fréquence. L'échelle de l'axe des temps est en jour Julien (JJ) moins* 2.4×10^6 *jours.*

La recherche d'oscillations à partir de la série temporelle corrigée est alors plus aisée. En particulier, l'utilisation d'outils de type CLEAN est rendue possible : les auteurs précités utilisent ainsi le logiciel PERIOD04 [Lenz et Breger, 2005] que nous avons déjà évoqué au § 2.2 et identifient huit fréquences d'oscillation entre 28.5 et 35.6 j^{-1}.

Il nous faut cependant émettre quelques réserves concernant la généralisation d'une telle démarche :

- D'une part, l'identification du mouvement orbital n'est bien entendu pas exacte et sa soustraction aux données initiales introduit nécessairement des erreurs sur les hautes fréquences : il ne s'agit pas d'une opération de *filtrage* à proprement parler. Notons surtout qu'une telle correction n'est pas toujours possible et nécessite en particulier de disposer d'observations couvrant une plus grande période temporelle : sur l'exemple de la figure 5.1, l'approximation de la contribution due au mouvement orbital a été réalisée en fusionnant les données d'août 1999 avec celles provenant d'une seconde campagne réalisée au cours de l'année 2000.
- L'utilisation de méthodes de type CLEAN, qui permettent généralement d'obtenir des résultats satisfaisants en termes de localisation fréquentielle, reste par ailleurs soumise aux limitations que nous avons soulevées au § 2.2 : la nécessité de contrôler le déroulement itératif de la procédure et la définition d'une condition d'arrêt sont deux points essentiels, le second étant relié à la notion de seuil significatif de détection d'une fréquence. Parmi les huit fréquences détectées à partir de ce jeu de données, Böhm *et al.* considèrent ainsi que trois le sont avec un *niveau de confiance* supérieur à 99.9%, trois avec un niveau de 99%, les deux autres l'étant avec un niveau plus faible, où le terme de niveau de confiance est à considérer avec les précautions que nous avons évoquées au § 2.2.2.

Le jeu de données brutes présenté constitue donc un test particulièrement sévère pour l'évaluation des méthodes d'estimation spectrale que nous avons proposées. En particulier les modèles utilisés dans ces méthodes, qui considèrent une somme de sinusoïdes dans un bruit blanc additif, ne correspondent pas aux données perturbées par un spectre continu de basses fréquences. De plus, aucune des méthodes classiques d'analyse spectrale appliquées à ces données ne produit de résultat intéressant. Nous rechercherons donc les variations à haute fréquence directement à partir des données brutes de la figure 5.1, et pourrons alors comparer l'estimation à celle obtenue par Böhm *et al.* [2004] à partir des données corrigées.

Nous effectuons tout d'abord au § 5.1 une brève analyse de ces données brutes pour lesquelles le spectre de Fourier, s'il ne permet pas de déterminer de manière précise les périodes d'oscillation recherchées, permet néanmoins de localiser grossièrement la plage fréquentielle à l'intérieur de laquelle nous pourrons ensuite les rechercher. Nous fixons également la valeur de la fréquence maximale du modèle en accord avec l'échantillonnage des données.

La courbe de vitesse radiale de la figure 5.1 ne résulte pas d'observations directes mais a été obtenue après différentes étapes de prétraitement [Böhm *et al.*, 2004] qui, si elles sont nécessaires à l'obtention d'un jeu de données exploitables, sont inévitablement à la source d'erreurs additionnelles[1]. Nous simulons donc dans un premier temps un signal présentant le même échantillonnage et le même comportement à basse fréquence que les données brutes, ainsi que les quatre composantes sinusoïdales de plus forte amplitude issues de l'estimation obtenue sur les données corrigées et publiée par les auteurs précités. Nous présentons au § 5.2 les résultats de

1. Notons que ces étapes de prétraitement touchent à la fois les valeurs des amplitudes et celles des instants d'échantillonnage.

l'estimation obtenue sur ces données simulées par les différentes méthodes proposées au cours de ce manuscrit, à savoir l'approche par pénalisation, l'estimation au sens de la moyenne *a posteriori* des paramètres d'un modèle Bernoulli-Gaussien B-G_0 et celle des paramètres du modèle Bernoulli-Gaussien avec décalages en fréquence B-G_{df}.

Le § 5.3 est finalement dédié à l'application de ces trois méthodes au traitement des données brutes, où nous comparons l'estimation des raies spectrales avec celle obtenue par Böhm *et al.* à partir des données corrigées. Nous appliquerons également nos méthodes aux données corrigées afin d'étudier l'influence d'une telle correction. Nous rappelons que tous les temps de calcul mentionnés ont été obtenus par des mises en œuvre sous *Matlab* avec un processeur cadencé à 1.4 GHz et doté de 1 Go de mémoire vive.

5.1 Présentation des données

Nous avons déjà représenté l'échantillonnage de la série temporelle de la figure 5.1 au premier chapitre de ce manuscrit, figure 1.1. Cet échantillonnage comporte $N = 514$ points et est à la fois irrégulier et troué par l'alternance du jour et de la nuit, où les observations couvrent environ 45% de chaque tranche de 24 heures. La fenêtre spectrale correspondante, représentée sur la figure 5.2, exhibe par conséquent un fort lobe secondaire à la fréquence de 1 j^{-1}. Notons que l'espacement irrégulier des instants d'échantillonnage permet d'obtenir une zone fréquentielle exempte d'alias d'environ 125 j^{-1} puisque la première « pseudo-périodicité » de la fenêtre spectrale apparaît vers 250 j^{-1}.

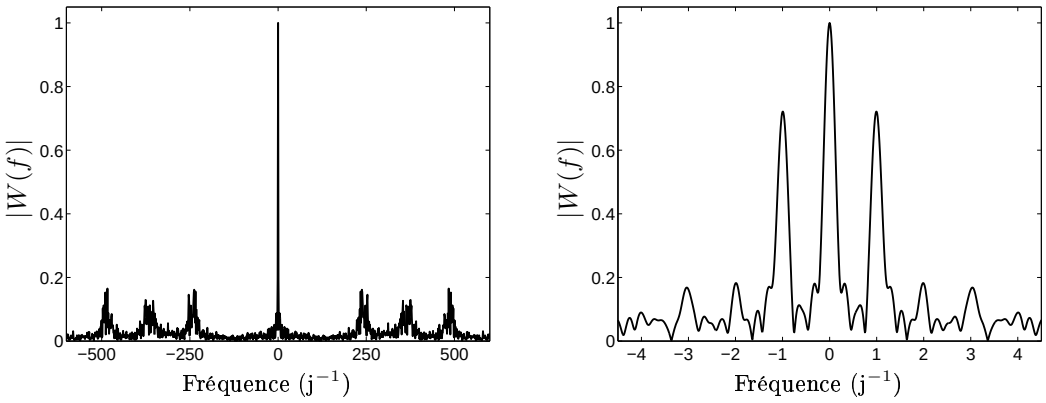


FIG. 5.2 – *Fenêtre spectrale correspondant à l'échantillonnage des données de la figure 5.1. À gauche, entre -500 et 500 j^{-1} ; à droite, entre -5 et 5 j^{-1}.*

Le spectre de Fourier des données centrées est représenté sur la figure 5.3 et est naturellement dominé par les basses fréquences correspondant au mouvement orbital. Cependant, sa représentation entre 0 et 125 j^{-1}, sur la figure de gauche, montre la présence de composantes spectrales autour de 30 j^{-1}. L'agrandissement de cette région sur la figure de droite permet d'identifier un pic dans le spectre de Fourier à une fréquence approximative de 33.3 j^{-1}, le reste de cette plage fréquentielle étant essentiellement marqué par les répliques de ce pic à ± 1 j^{-1} causées par l'alternance jour/nuit dans l'échantillonnage. Nous notons encore que le spectre de Fourier des données corrigées, également représenté sur la figure 5.3 à droite, présente pour les plus hautes fréquences une allure proche du spectre des données brutes. Les deux courbes sont cependant

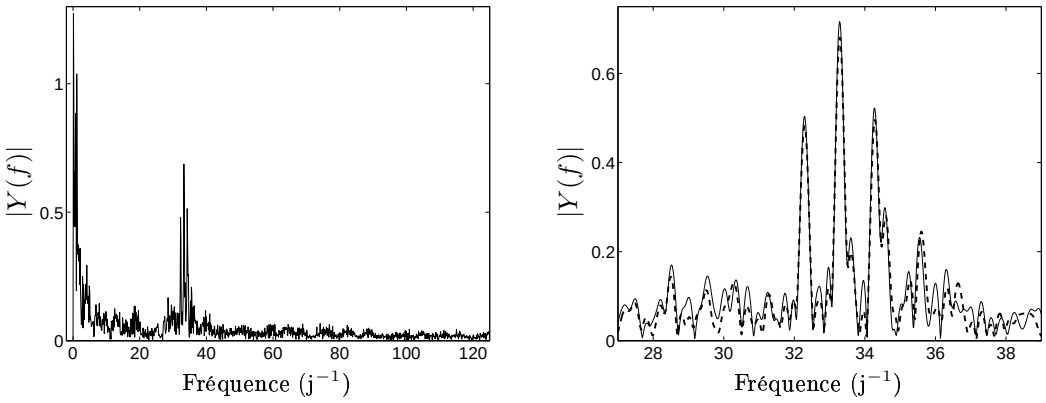


FIG. 5.3 – *Spectre de Fourier des données brutes de la figure 5.1 (en trait plein). À gauche, entre 0 et 125 j^{-1} ; à droite, agrandissement autour de 33 j^{-1}. En pointillé, spectre de Fourier des données corrigées.*

différentes ; en particulier, le pic secondaire aux alentours de 35.6 j^{-1} qui sera par la suite identifié comme représentatif d'une fréquence d'oscillation, est localisé avec une différence de l'ordre de 0.05 j^{-1} par les deux spectres de Fourier.

À la vue de cette figure, même si la plage fréquentielle libre d'alias s'étend jusqu'à 125 j^{-1}, nous fixerons donc par la suite la fréquence maximale à $f_{\max} = 40$ j^{-1} afin de ne pas alourdir inutilement la charge de calcul. Signalons cependant que nous avons mené des essais avec une valeur supérieure de $f_{\max}$, mais aucune composante n'a jamais été détectée au-delà de 40 j^{-1}.

5.2 Comparaison des méthodes sur des données simulées

Nous proposons dans un premier temps d'évaluer les méthodes proposées aux chapitres 3 et 4 sur des données artificielles construites sur la forme du signal de la figure 5.1. Plus précisément, nous générons un signal échantillonné en les mêmes instants que les véritables observations sous la forme :

$$y(t_n) = \mathrm{BF}(t_n) + \sum_{k=1}^{K} a_k \sin\left(2\pi\nu_k t_n + \phi_k\right) + \epsilon_n, \quad \text{où :}$$

- BF est la composante orbitale de basses fréquences telle qu'elle a été extraite par Böhm *et al.* [2004], représentée sur la figure 5.1 au centre ;
- les paramètres des $K = 4$ sinusoïdes correspondent aux quatre raies spectrales de plus forte amplitude détectées par ces auteurs à partir des données corrigées. Les fréquences ν_k sont choisies en dehors de la grille d'estimation $\mathcal{G} = \left\{\frac{k}{P} f_{\max}\right\}_{k=0\ldots P}$. Nous avons volontairement relevé l'amplitude de la plus faible des quatre raies afin d'obtenir une meilleure lisibilité des résultats ;
- un bruit blanc additif ϵ centré gaussien, de puissance σ_ϵ^2 telle que le RSB soit de 10 dB, où nous définissons le RSB par rapport à l'amplitude des sinusoïdes :

$$\mathrm{RSB}_{\mathrm{dB}} = 10 \log_{10}\left(\frac{\sum_k a_k^2}{2\sigma_\epsilon^2}\right).$$

La représentation temporelle de ce signal est visuellement très proche de celle du signal réel et n'est pas tracée ici. Le spectre de Fourier correspondant est représenté sur la figure 5.4 et présente une allure également très proche du spectre de Fourier des observations, à savoir une prédominance des basses fréquences et la détection d'un seul pic significatif vers 33.3 j^{-1}.

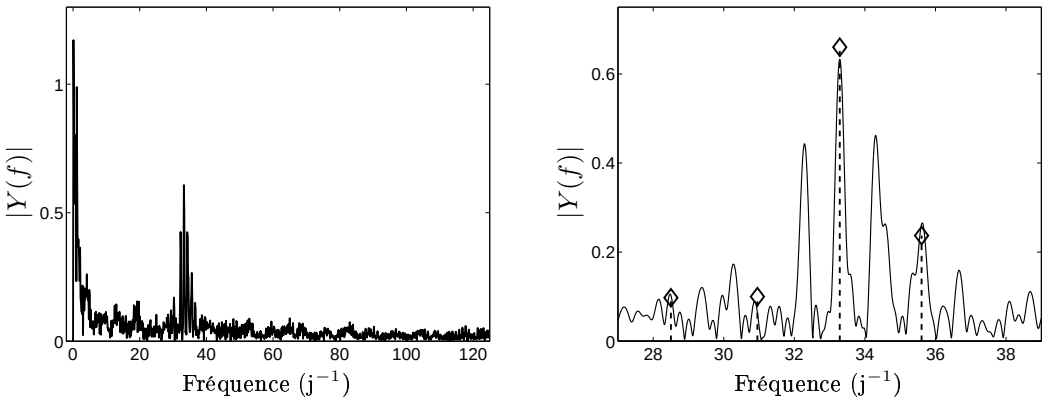


FIG. 5.4 – *Spectre de Fourier des données simulées. À gauche, entre 0 et 125 j^{-1} ; à droite, agrandissement autour de 33 j^{-1}. Les $\diamond$ représentent les raies théoriques, de fréquences ν_k et d'amplitudes $a_k/2$.*

Nous présentons aux paragraphes suivants l'application à ce signal des différentes procédures d'estimation développées au cours de cette thèse, à savoir la minimisation d'un critère pénalisé par la norme ℓ^1 ou par une fonction hyperbolique et l'estimation au sens de la moyenne *a posteriori* des paramètres des modèles Bernoulli-Gaussien classique et avec décalages en fréquence.

5.2.1 Estimation par pénalisation

Nous étudions ici l'estimation spectrale obtenue, pour une grille de $P = 1000$ fréquences positives, par la minimisation des critères pénalisés par la norme ℓ^1 et par son approximation hyperbolique :

$$\begin{array}{rcl} \widehat{\boldsymbol{x}}_1 & = & \arg\min J_1(\boldsymbol{x}), \qquad J_1(\boldsymbol{x}) \;=\; \frac{1}{2}\left\|\boldsymbol{y}-\mathbf{W}\boldsymbol{x}\right\|^2 + \lambda \sum_k |x_k| \\ \text{et } \widehat{\boldsymbol{x}}_{\mathrm{hyp}} & = & \arg\min J_{\mathrm{hyp}}(\boldsymbol{x}), \quad J_{\mathrm{hyp}}(\boldsymbol{x}) \;=\; \frac{1}{2}\left\|\boldsymbol{y}-\mathbf{W}\boldsymbol{x}\right\|^2 + \lambda \sum_k \sqrt{s^2+|x_k|^2}. \end{array}$$

Pour une valeur du paramètre de régularisation λ comprise entre 20 et 45, les spectres estimés $\widehat{\boldsymbol{x}}_1$ et $\widehat{\boldsymbol{x}}_{\mathrm{hyp}}$ permettent d'identifier les quatre raies et des valeurs en dehors de cet intervalle ne fournissent pas une estimation satisfaisante : la figure 5.5 montre l'estimation de $\widehat{\boldsymbol{x}}_1$ obtenue pour des solutions sur- et sous-régularisées, avec $\lambda = 55$ et $\lambda = 5$, respectivement. Dans le premier cas, seules les deux raies de plus fortes amplitudes sont détectées alors que pour le second des composantes non nulles et de faible amplitude sont disséminées tout au long de la grille fréquentielle. La courbe $\big(Q(\widehat{\boldsymbol{x}}_\lambda),R(\widehat{\boldsymbol{x}}_\lambda)\big)$, qui a permis dans le cas de simulations simples d'obtenir une valeur de λ satisfaisante, montre que la plage de valeurs acceptables pour λ est ici plus restreinte. Une approche plus heuristique pour sélectionner une solution satisfaisante consiste alors à choisir visuellement la valeur la plus faible de λ qui fournit une solution ne présentant pas de faibles pics répartis tout au long de la grille fréquentielle comme c'est le cas pour la solutions sous-régularisée de la figure 5.5 à droite. Nous reviendrons sur cet point lors de l'application sur les données réelles au § 5.3.

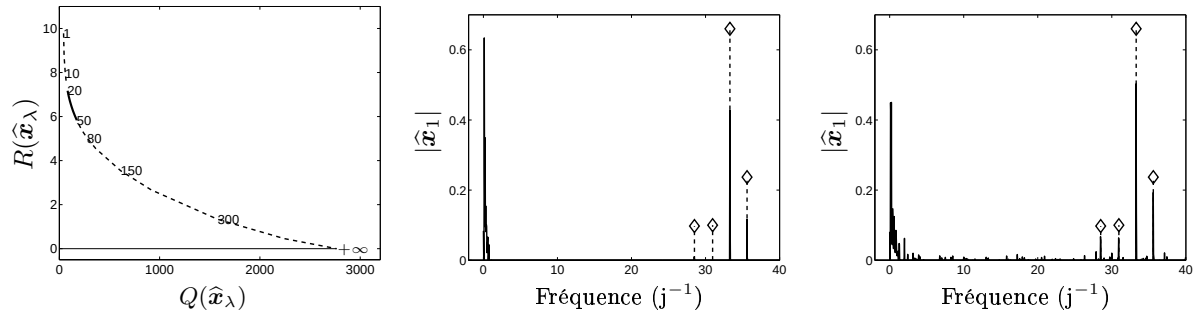


FIG. 5.5 – *Sensibilité de l'estimation par pénalisation ℓ^1 à la valeur du paramètre de régularisation. À gauche, courbe $(Q(\widehat{\boldsymbol{x}}_\lambda),R(\widehat{\boldsymbol{x}}_\lambda))$ paramétrée en λ. Le trait plein représente la zone fournissant une estimation acceptable. Au centre, sur-régularisation ($\lambda = 55$) et à droite, sous-régularisation ($\lambda = 5$). Les traits pleins représentent le module du spectre estimé et les $\diamond$ les raies théoriques.*

L'estimation obtenue pour $\lambda = 25$ est présentée sur la figure 5.6. Dans le cas de la pénalisation ℓ^1, les seules composantes non nulles de $\widehat{\boldsymbol{x}}_1$ sont localisées aux basses fréquences et au niveau de chaque raie. Pour la pénalisation hyperbolique, bien que la forme de $\widehat{\boldsymbol{x}}_{\text{hyp}}$ ne soit pas rigoureusement parcimonieuse, l'allure suffisamment « piquée » de la solution permet également de retrouver les quatre raies contenues dans le signal simulé. Nous remarquons néanmoins un léger décalage sur l'estimation de la fréquence ν_2 qui est décalée d'un point (soit $f_{\max}/P = 0.04$ j^{-1}) par rapport à la meilleure approximation de ν_2 sur la grille fréquentielle du modèle.

Comme dans les simulations présentées au cours du chapitre 3 pour un signal multi-sinusoïdal, nous notons également une sous-estimation des amplitudes. Nous avons alors vu qu'il était possible de réestimer *a posteriori* les amplitudes une fois que les fréquences ont été localisées. Ici encore, malgré les perturbations à basse fréquence (BF), une étape de réestimation postérieure permet d'estimer correctement les amplitudes par les étapes suivantes :

i) ne conserver que les composantes d'amplitudes non nulles de $\widehat{\boldsymbol{x}}_1$, ou les composantes supérieures à un certain seuil pour $\widehat{\boldsymbol{x}}_{\text{hyp}}$;

ii) dans la zone spectrale qui nous intéresse, lorsqu'une raie est estimée par deux amplitudes spectrales adjacentes non nulles[2], ne conserver que la plus grande des deux ;

iii) former le modèle multi-sinusoïdal correspondant : $\boldsymbol{y} = \mathbf{W}_{-\mathcal{I}}\boldsymbol{x}_{-\mathcal{I}} + \boldsymbol{\epsilon}$ où $\mathbf{W}_{-\mathcal{I}}$ (respectivement $\boldsymbol{x}_{-\mathcal{I}}$) est composé des colonnes de $\mathbf{W}$ (respectivement des composantes de $\boldsymbol{x}$) correspondant aux indices conservés. Si ceux-ci sont en nombre restreint devant le nombre de données (sur le signal traité ici, ce modèle est composé de seulement 32 composantes fréquentielles, soit le double d'inconnues complexes pour $N = 514$ données), le problème linéaire d'estimation des amplitudes correspondantes est surdéterminé. Toutefois, la présence de plusieurs composantes fréquentielles voisines aux basses fréquences rend les colonnes de $\mathbf{W}_{-\mathcal{I}}$ proches de la colinéarité et l'inversion directe par moindres carrés : $\widehat{\boldsymbol{x}}_{-\mathcal{I}} = (\mathbf{W}_{-\mathcal{I}}^\dagger\mathbf{W}_{-\mathcal{I}})^{-1}\mathbf{W}_{-\mathcal{I}}^\dagger\,\boldsymbol{y}$ n'est pas numériquement stable. Celle-ci peut être à son tour régularisée sous la forme :

$$\widehat{\boldsymbol{x}}_{-\mathcal{I}} = (\mathbf{W}_{-\mathcal{I}}^\dagger\mathbf{W}_{-\mathcal{I}} + \alpha\mathbf{I})^{-1}\mathbf{W}_{-\mathcal{I}}^\dagger\,\boldsymbol{y},$$

où α prend une valeur faible mais suffisante pour assurer une inversion numérique correcte de la matrice $\mathbf{W}_{-\mathcal{I}}^\dagger\mathbf{W}_{-\mathcal{I}} + \alpha\mathbf{I}$. Sur ce signal, la réestimation des amplitudes est représentée

2. C'est par exemple le cas des fréquences ν_1, ν_3 et ν_4 sur l'exemple de la figure 5.6.

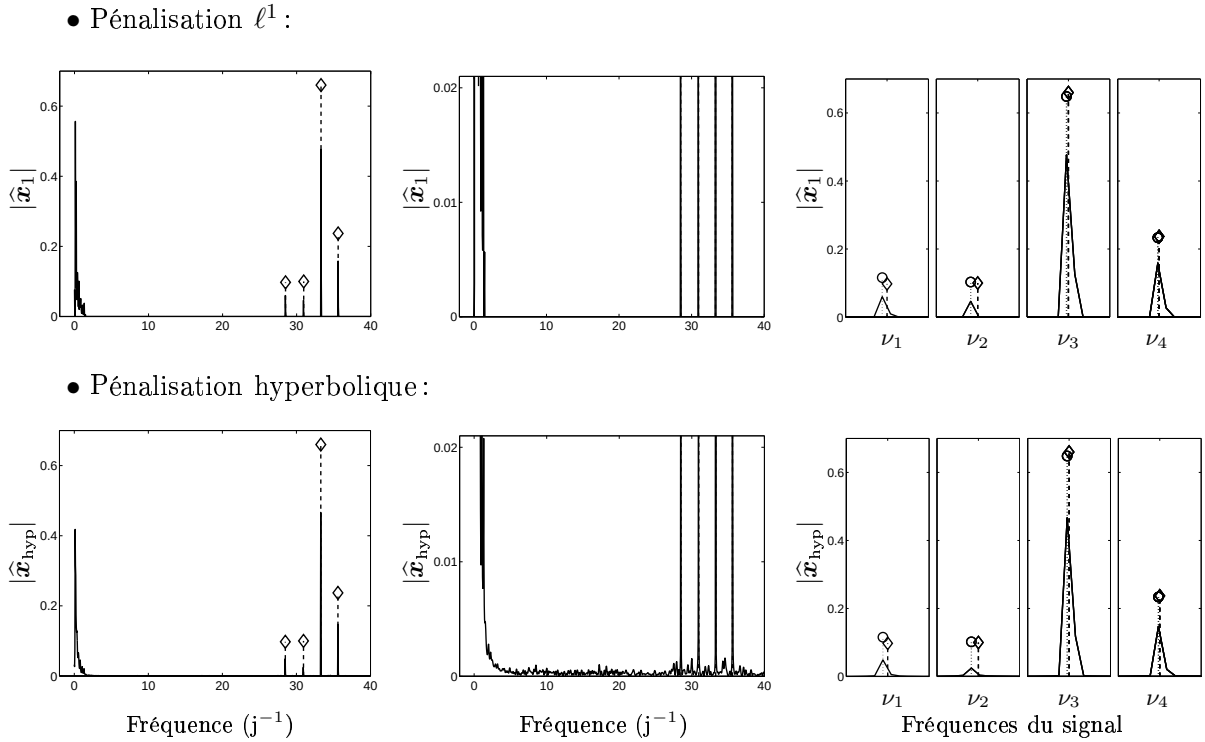


FIG. 5.6 – *Résultats d'estimation par minimisation d'un critère pénalisé. À gauche, estimé en module sur l'ensemble de la grille fréquentielle de $P + 1 = 1001$ points. Au centre, zoom en amplitude pour les faibles valeurs. À droite, zoom en fréquence autour de chaque raie. Les traits pleins représentent le module du spectre estimé, les $\diamond$ les raies théoriques et, sur les figures de droite, les $\circ$ correspondent à la réestimation a posteriori par moindres carrés régularisés des amplitudes localisées par les pics de $|\widehat{\boldsymbol{x}}|$.*

sur la figure 5.6 à droite pour une valeur de $\alpha = 10^{-3}$: les valeurs obtenues de cette façon sont tout à fait acceptables.

À propos de la mise en œuvre

Au niveau de la mise en œuvre algorithmique, il nous faut signaler que le temps nécessaire à l'optimisation des critères pénalisés J_1 et J_{hyp} est largement plus élevé pour ce type de données par rapport au signal plus simple utilisé au long du chapitre 3. Les algorithmes de type Gauss-Seidel, par nature plus adaptés à la recherche de solutions présentant des variations rapides, sont à ce titre les plus touchés et convergent ici plus lentement du fait de la forme *continue* de la partie à basse fréquence du spectre due au mouvement orbital.

En application à ce genre de données, nous proposons donc d'utiliser un critère d'arrêt moins sévère que celui utilisé auparavant pour lequel la norme du gradient en la solution devait être inférieure à 10^{-4} pour le critère J_{hyp} et la valeur du gradient de J_1 devait être inférieure à 10^{-6}, voire 10^{-10}, en tout point où celui-ci est défini, c'est-à-dire en les composantes non nulles de la solution. Les estimateurs $\widehat{\boldsymbol{x}}_1$ et $\widehat{\boldsymbol{x}}_{\text{hyp}}$ ont été en réalité obtenus par des algorithmes de type GSOR (Gauss-Seidel avec sur-relaxation) arrêtés à l'itération t telle que :

$$\frac{\left\|\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}\right\|}{\left\|\boldsymbol{x}^{(t)}\right\|} < 10^{-4} \text{ et } \frac{|J(\boldsymbol{x}^{(t)}) - J(\boldsymbol{x}^{(t-1)})|}{J(\boldsymbol{x}^{(t)})} < 10^{-8} \tag{5.1}$$

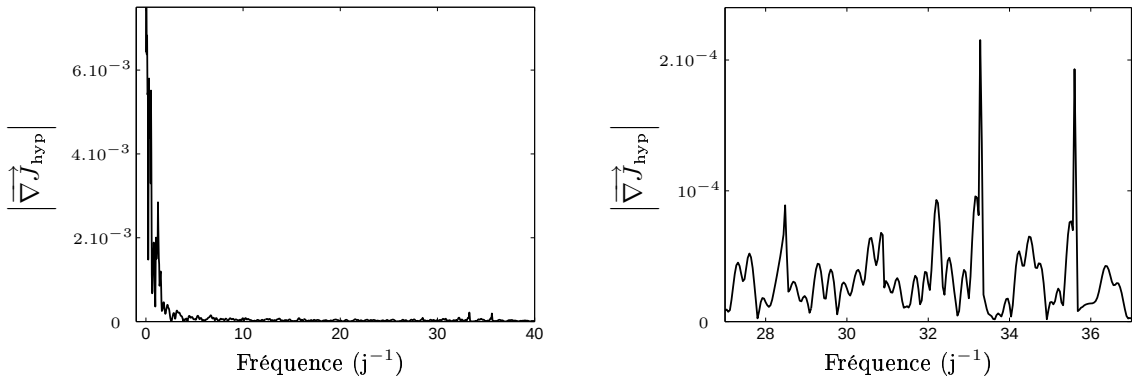
La figure 5.7 en haut montre l'allure du gradient du critère J_{hyp} en la solution $\widehat{\boldsymbol{x}}_{\text{hyp}}$. Lorsque le test d'arrêt (5.1) est vérifié, la norme du gradient vaut encore $\left\|\overrightarrow{\nabla J}_{\text{hyp}}\right\| \simeq 5.10^{-2}$ mais dans la plage fréquentielle qui nous intéresse, celui-ci prend des valeurs plus faibles de l'ordre de 10^{-4}. Nous représentons également sur la figure 5.7 en bas les caractéristiques, en la solution $\widehat{\boldsymbol{x}}_1$, du test d'optimalité vérifié par un minimiseur du critère J_1, au sens duquel on doit avoir :

$$\begin{cases} (i) & \forall k \text{ tq. } |x_k| = 0 : \quad |r_k| \leq \lambda, \\ (ii) & \forall k \text{ tq. } |x_k| \neq 0 : \quad r_k + \lambda \dfrac{x_k}{|x_k|} = 0, \end{cases}$$

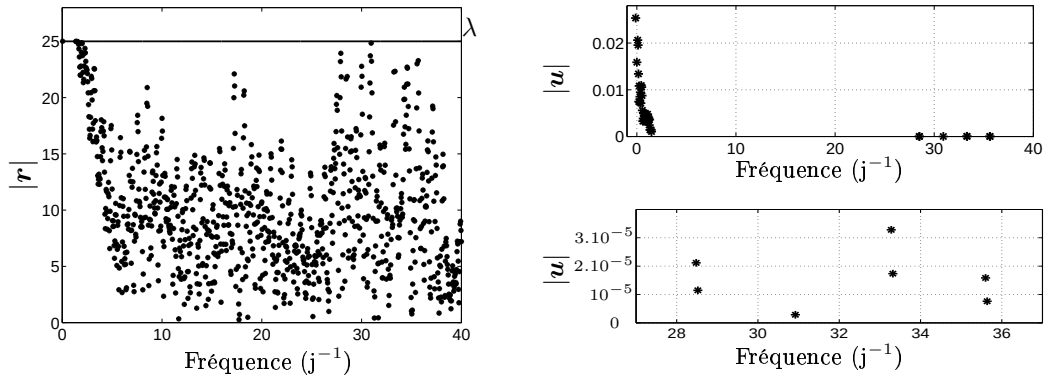
où $\boldsymbol{r}^{(t)} = \mathbf{W}^\dagger(\mathbf{W}\boldsymbol{x}^{(t)} - \boldsymbol{y})$. À la vue de cette figure, la condition (i) est satisfaite excepté pour une composante nulle de $\boldsymbol{x}^{(t)}$ à basse fréquence[3]. La condition (ii) est quant à elle loin d'être satisfaite puisqu'en les composantes non nulles à basse fréquence on a $r_k^{(t)} + \lambda x_k^{(t)}/|x_k^{(t)}|$ de l'ordre de 10^{-2}. Cependant, nous voyons sur la figure 5.7 en bas à droite que pour la zone fréquentielle à laquelle appartiennent les fréquences $\nu_1, \ldots, \nu_4$, les valeurs théoriquement nulles au sens de la condition (ii) sont de l'ordre de 10^{-5}. Bien que moins rigoureux en termes d'optimalité, le critère d'arrêt (5.1) reste donc suffisamment sévère sur la partie de l'axe fréquentiel pour laquelle la solution est parcimonieuse et a permis en pratique d'obtenir des estimateurs satisfaisants tout en maintenant des coûts calculatoires raisonnables : pour $P = 1000$, le calcul de $\widehat{\boldsymbol{x}}_{\text{hyp}}$ et $\widehat{\boldsymbol{x}}_1$ par les algorithmes GSOR prend environ 390 s et 166 s respectivement[4], alors que le calcul de $\widehat{\boldsymbol{x}}_1$ au sens du test d'arrêt initial nécessite environ 10 heures.

3. Cette figure ne représente que la moitié des composantes r_k correspondant aux fréquences positives. Le vecteur $\boldsymbol{r}^{(t)}$ étant à symétrie hermitienne lorsque $\boldsymbol{x}^{(t)}$ l'est, il y a en réalité deux composantes pour lesquelles la condition (i) n'est pas satisfaite.

4. Ces valeurs sont à comparer aux temps respectifs établis au § 3.7 de 75 s et 33 s nécessaires à la convergence des algorithmes GSOR pour un problème de taille similaire, mais sur un signal multi-sinusoïdal sans perturbations à basse fréquence et pour un test d'arrêt plus strict.



Gradient du critère J_{hyp} en la solution $\widehat{\boldsymbol{x}}_{hyp}$. La figure de gauche couvre l'ensemble de la grille fréquentielle et celle de droite est un agrandissement de la zone fréquentielle à laquelle appartiennent les raies du signal.



Test d'optimalité du minimiseur du critère J_1 en la solution $\widehat{\boldsymbol{x}}_1$. À gauche, condition (i) : tracé de $|r_k|$ pour les composantes nulles de $\boldsymbol{x}$. Le trait plein représente le seuil de valeur λ. À droite, condition (ii) : valeurs en module de $u_k = r_k + \lambda x_k/|x_k|$ pour les composantes non nulles de $\boldsymbol{x}$. La figure en haut couvre l'ensemble de la grille fréquentielle et celle du bas est un agrandissement autour de chaque raie du signal.

FIG. 5.7 – *Évaluation des solutions $\widehat{\boldsymbol{x}}_{hyp}$ et $\widehat{\boldsymbol{x}}_1$ obtenues par un algorithme GSOR au sens du critère d'arrêt (5.1).*

5.2.2 Estimation par modélisation B-G

Nous montrons ici les résultats obtenus sur ce même signal en modélisant son spectre discrétisé sous la forme d'un processus Bernoulli-Gaussien. Conformément à la stratégie proposée au § 4.3.3, les paramètres B-G sont estimés de manière séquentielle : l'estimation de la moyenne *a posteriori* de la séquence de Bernoulli est d'abord réalisée. Nous désignerons par la suite $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ l'estimation obtenue, tout en rappelant que celle-ci n'est pas une séquence de Bernoulli : $\widehat{q}_{\mathrm{MP}_k} \in [0,\, 1]$ mais en général $\widehat{q}_{\mathrm{MP}_k} \neq 0$, $\widehat{q}_{\mathrm{MP}_k} \neq 1$. Cette estimation est ensuite seuillée pour ne conserver que les fréquences significatives et l'estimation des amplitudes correspondantes est réalisée au sens du maximum *a posteriori* conditionnellement à la localisation des raies apportée par l'étape précédente, qui prend une forme analytique explicite en raison des hypothèses gaussiennes sur les distributions du bruit et des amplitudes conditionnelles $\boldsymbol{s}|\boldsymbol{q}$.

En reprenant les notations du chapitre 4, l'échantillonnage stochastique de la loi *a posteriori* $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y})$ est réalisé ici selon la version « par blocs » de l'échantillonnage de Gibbs, notée G_3 et décrite par l'algorithme 4.4 à la page 122. L'estimation est totalement non supervisée et nous utilisons des lois *a priori* de Jeffreys sur les hyperparamètres σ^2 et σ_ϵ^2 et une loi *a priori* uniforme sur le paramètre de Bernoulli λ. La grille fréquentielle $\mathcal{G} = \{f_k\}_{k=0\ldots P} = \{\frac{k}{P} f_{\max}\}_{k=0\ldots P}$ est la même qu'au paragraphe précédent, c'est-à-dire $f_{\max} = 40\ \mathrm{j}^{-1}$ et $P = 1000$. Les fréquences ν_j du signal simulé étant choisies en dehors de la grille de reconstruction, nous noterons $k_{j,\,j=1\ldots4}$ les indices tels que $f_{k_j} < \nu_j < f_{k_j+1}$.

Tout comme nous avons dû adapter le seuil de convergence pour les algorithmes d'optimisation au paragraphe précédent, il s'est également avéré nécessaire pour l'application à ces données d'adapter le critère (4.15) arrêtant la procédure d'échantillonnage. Avec les notations de (4.15), les résultats présentés ci-dessous ont été obtenus pour le test d'arrêt :

$$||\widehat{\boldsymbol{u}}^T - \widehat{\boldsymbol{u}}^{T-T_0}||_\infty < \alpha \quad \text{avec } \alpha = 10^{-2.5} \quad \text{et} \quad \widehat{\boldsymbol{u}}^T = \frac{2}{T} \sum_{t=T/2+1}^{T} \boldsymbol{u}^{(t)},$$

au lieu de la valeur de $\alpha = 10^{-3}$ utilisée auparavant, et nécessitent déjà un temps de calcul très élevé, supérieur à l'heure. Le nombre d'itérations nécessaires sur cet exemple est de l'ordre de $T = 13000$. L'estimation $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ est représentée sur la figure 5.8 et prend des valeurs très faibles, voire nulles, en la plupart des points de la grille $\mathcal{G}$: seules quelques composantes aux basses fréquences et autour de chaque raie contenue dans le signal prennent des valeurs significatives. Rappelons que $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ peut s'interpréter comme la probabilité d'existence d'une raie pour chaque fréquence f_k de la grille. Un seuillage approprié permet alors de ne sélectionner que les raies les plus probables. En raison de l'inexactitude de la discrétisation fréquentielle, l'estimation de $\boldsymbol{q}$ est sensible au même effet, étudié au § 4.5.1, de moyennage des échantillons modélisant chaque raie par son approximation de part et d'autre de la grille $\mathcal{G}$. Il en résulte pour les fréquences ν_1 et ν_2 des probabilités $\widehat{q}_{\mathrm{MP}_k}$ non nulles de part et d'autre de la grille, mais *dont la somme vaut environ 1.* La figure 5.9 à gauche montre par exemple pour la fréquence ν_2 l'évolution des échantillons $q_{k_2}^{(t)}$ et $q_{k_2+1}^{(t)}$, où nous voyons la « complémentarité » des deux localisations : en général, si à une itération donnée la raie n'est pas détectée en f_{k_2}, elle l'est en f_{k_2+1} et vice-versa. L'interprétation n'est dans ce cas pas problématique, puisqu'elle traduit l'existence d'une raie dont la fréquence est encadrée par les deux composantes non nulles adjacentes de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$, à savoir (f_{k_1}, f_{k_1+1}) pour ν_1 et (f_{k_2}, f_{k_2+2}) pour ν_2. Sur cet exemple, un seuillage adéquat consiste alors à ne conserver que les indices donnant une probabilité *a posteriori* supérieure à 1/2, qui permet de ne garder, pour ν_1 et ν_2, qu'une seule de ces deux composantes : la plus « probable ». Par contre, le comportement

de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ autour de la fréquence ν_3 est plus ambigu et la figure 5.8 à droite laisse croire à l'existence de deux raies : si une raie est estimée à la fréquence f_{k_3} avec une probabilité de 1, les valeurs non nulles de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ aux indices k_3-3, k_3-2 et k_3-1, d'ailleurs complémentaires puisque leur somme vaut approximativement 1, laissent penser à l'existence d'une seconde raie de fréquence comprise entre f_{k_3-2} et f_{k_3}. La figure 5.9 à droite met en évidence la complémentarité des localisations en f_{k_3-3}, f_{k_3-2} et f_{k_3-1}.

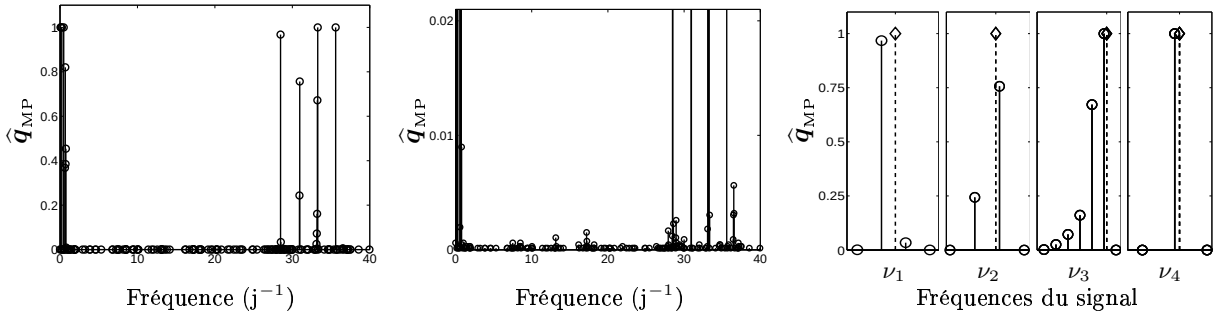


FIG. 5.8 – *Moyenne estimée $\widehat{\boldsymbol{q}}_{MP}$ pour une modélisation B-G sur une grille de $P+1=1001$ points. À gauche, sur l'ensemble de la grille fréquentielle. Au centre, zoom en amplitude pour les faibles valeurs. À droite, zoom en fréquence autour de chaque raie. Les $\diamond$ représentent les positions fréquentielles des raies théoriques.*

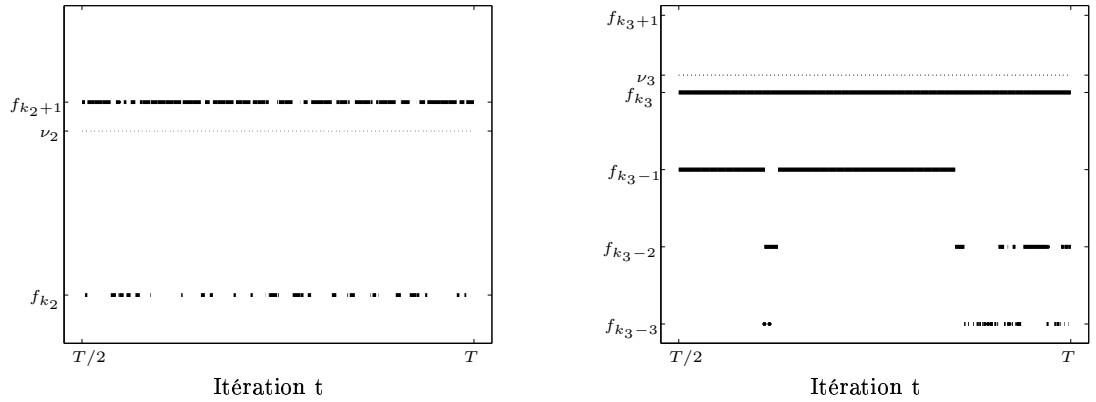


FIG. 5.9 – *Échantillons de Bernoulli $q_{k_j}^{(t)}$ au fil des itérations. À gauche, pour les fréquences autour de ν_2 ; à droite, pour les fréquences autour de ν_3. Le trait plein marque les valeurs $q_{k_j}^{(t)}=1$.*

La figure 5.10 montre l'estimation MAP des amplitudes après seuillage de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ – nous n'avons conservé ici que les fréquences pour lesquelles la probabilité *a posteriori* d'existence d'une raie est supérieure à 1/2, ce qui permet de conserver également la plupart des composantes à basse fréquence. Les barres d'erreur à $\pm\sigma_k$ sont également représentées, où σ_k est l'écart-type associé à l'amplitude de la raie à la fréquence f_k, dont l'expression analytique est également explicite (voir le § 4.3.3). Des valeurs satisfaisantes des paramètres spectraux sont finalement obtenues pour les raies en ν_1, ν_2 et ν_4, et les variances associées à l'estimation des amplitudes sont faibles. En raison d'une étape de détection légèrement erronée, l'estimation pour la fréquence ν_3 est moins satisfaisante puisqu'elle débouche sur deux amplitudes spectrales élevées aux fréquences f_{k_3-1} et f_{k_3}. Cependant, la variance associée à ces amplitudes est anormalement élevée comparée à celle

des trois autres raies spectrales, laissant entrevoir un comportement « suspect » de l'estimation pour cette fréquence.

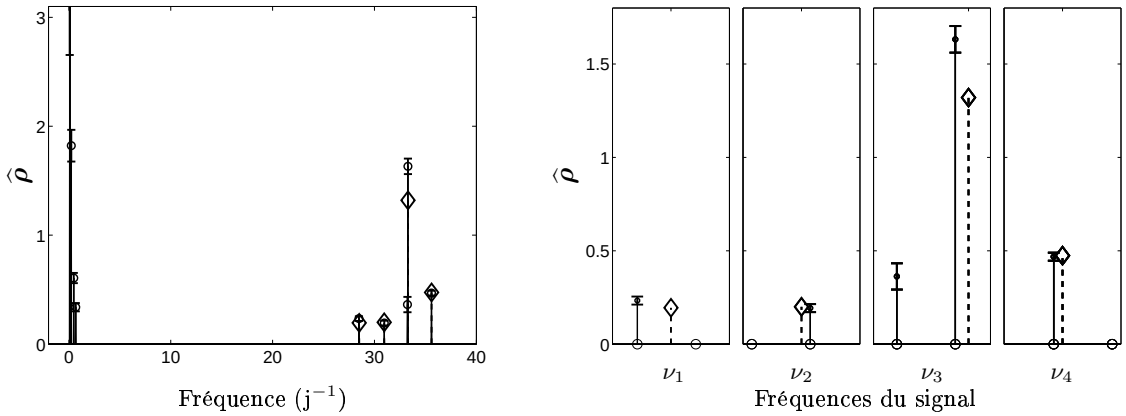


FIG. 5.10 – *Estimation au sens du maximum a posteriori des amplitudes pour les fréquences localisées par* $\widehat{\boldsymbol{q}}_{MP} > 0.5$. *À gauche, sur l'ensemble de la grille fréquentielle de* $P + 1 = 1001$ *points : à droite, zoom en fréquence autour de chaque raie. Les barres d'erreur correspondent à* $\pm\sigma_k$, *où* σ_k *est l'écart-type associé, et les* $\diamond$ *représentent les raies théoriques. Chaque fenêtre en fréquence sur la figure de droite couvre une plage de* $2\frac{f_{max}}{P} = 8 \times 10^{-2}\ j^{-1}$.

5.2.3 Estimation par modélisation B-G avec décalages en fréquence

L'estimation des paramètres d'un modèle Bernoulli-Gaussien prenant en compte des décalages fréquentiels est réalisée au moyen de l'échantillonnage de la loi jointe *a posteriori* $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{df}|\boldsymbol{y})$, selon le schéma décrit par l'algorithme du tableau 4.6, en utilisant les mêmes lois *a priori* sur les hyperparamètres que dans le cas précédent. Les paramètres sont ici discrétisés sur une grille de $P + 1 = 501$ fréquences positives et l'arrêt de d'échantillonnage est déterminé au sens du critère (4.15), pour lequel $T \simeq 27000$ itérations et un temps de calcul d'environ 4300 s sont nécessaires pour la réalisation présentée plus bas[5].

L'estimation $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ est représentée sur la figure 5.11 : comme pour le modèle B-G$_\mathbf{0}$, $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ prend des valeurs faibles excepté à basse fréquence et autour de chaque raie du signal. En rappelant que, pour ce modèle, $\widehat{q}_{\mathrm{MP}_k}$ est une estimation de la probabilité *a posteriori* qu'une raie existe dans l'intervalle de fréquences $\mathcal{I}_k = [f_{k-1}, f_k] = \left[\frac{k-1}{P} f_{\max}, \frac{k}{P} f_{\max}\right]$, nous notons ici $\mathcal{I}_{k_j,\, j=1\ldots4}$ les intervalles auxquels appartiennent les fréquences respectives $\nu_{j,\, j=1\ldots4}$. Les intervalles $\mathcal{I}_{k_j,\, j=1,2,3}$ sont correctement localisés puisque $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ prend des valeurs égales à 1 en chaque indice $k_{j,\, j=1,2,3}$, c'est-à-dire en la première valeur de la grille fréquentielle supérieure à ν_j. Pour la fréquence ν_3, une seconde fausse composante est détectée dans l'intervalle fréquentiel voisin $\mathcal{I}_{k_3-1}$, mais avec une probabilité de l'ordre de 0.14, ce qui reste faible. Pour la fréquence ν_4, l'estimation présente un moyennage de la localisation fréquentielle entre les intervalles $\mathcal{I}_{k_4-1}$ et $\mathcal{I}_{k_4}$: on a $q_{k_4-1} + q_{k_4} \simeq 1$, où ces deux probabilités prennent des valeurs significatives.

Si dans le cas du modèle B-G$_\mathbf{0}$ l'effet de moyennage entre deux positions adjacentes est

5. Le nombre de fréquences discrétisées est ici deux fois plus faible qu'au paragraphe précédent pour la modélisation B-G ; contrairement à ce cas, l'utilisation du critère d'arrêt initial (4.15) est ici envisageable pour un temps de calcul plus « raisonnable ».

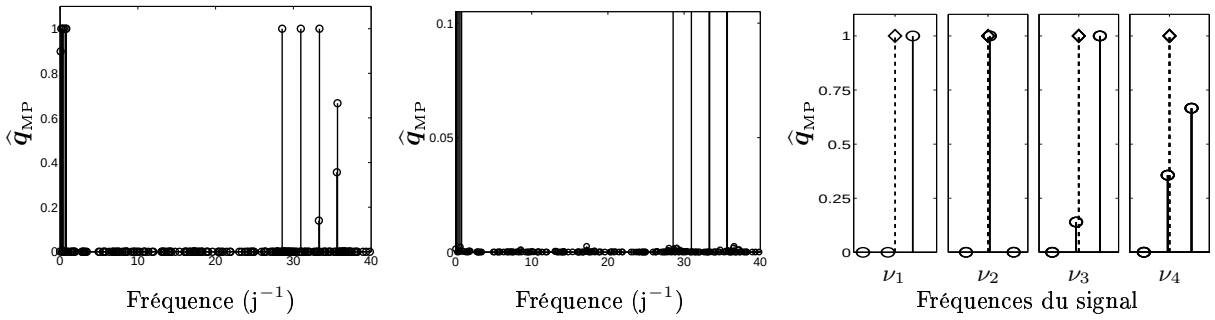


FIG. 5.11 – *Moyenne estimée $\widehat{\boldsymbol{q}}_{MP}$ pour une modélisation B-G$_{\boldsymbol{df}}$ sur une grille de $P+1=501$ points. À gauche, sur l'ensemble de la grille fréquentielle. Au centre, zoom en amplitude pour les faibles valeurs. À droite, zoom en fréquence autour de chaque raie. Les $\diamond$ représentent les positions fréquentielles des raies théoriques.*

une conséquence logique de la discrétisation inexacte de l'axe fréquentiel, son apparition dans un modèle à fréquences continues traduit ici plutôt un problème *algorithmique*. En effet, une analyse plus détaillée des échantillons générés, présentée sur la figure 5.12 à droite, montre que la stationnarité de la chaîne n'a pas été atteinte sur l'ensemble des échantillons pris en compte pour l'estimation : la première partie des échantillons conservés comprend la raie dans l'intervalle $\mathcal{I}_{k_4-1}$ et, pour la seconde partie, la raie s'est « décalée » dans l'intervalle $\mathcal{I}_{k_4}$. Vu le nombre d'échantillons produits – sur cet exemple, l'estimation est réalisée à partir de $T/2 = 13500$ échantillons dont environ les 5000 premiers sont « bloqués » avec une raie dans l'intervalle $\mathcal{I}_{k_4-1}$ –, la variation résultante de la moyenne estimée à chaque itération devient suffisamment faible pour que le test d'arrêt (4.15) soit satisfait. Nous observons également sur la figure 5.12 à gauche que la détection « parasite » d'une fréquence dans l'intervalle $\mathcal{I}_{k_3-1}$ n'est présente que sur les 2000 premiers échantillons considérés.

Le critère d'arrêt (4.15) n'est donc pas suffisant dans ce cas, et une condition plus sévère amènerait à la diminution des valeurs de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ en k_3-1 et k_4-1. Nous avons cependant préféré conserver cet exemple afin d'illustrer la difficulté du choix d'une condition d'arrêt qui fait partie des principales limitations de cette approche : la génération, sur cet exemple, de $T = 27000$ échantillons n'est pas suffisante alors qu'elle requiert déjà un temps de calcul supérieur à une heure. Remarquons que ce problème de blocage de l'échantillonnage dans une configuration erronée n'est pas sans rappeler les lenteurs algorithmiques de l'échantillonnage de Gibbs mises en évidence au § 4.4.1 pour le modèle B-G$_{\mathbf{0}}$. À ce jour, nous n'avons toutefois pas réussi à accélérer l'algorithme pour l'échantillonnage des paramètres de B-G$_{\boldsymbol{df}}$ comme nous l'avons fait dans le cas de B-G$_{\mathbf{0}}$.

Notons enfin que, en fonction de l'initialisation du générateur aléatoire, l'estimation obtenue en arrêtant l'échantillonnage au sens du test (4.15) a produit des résultats parfois légèrement différents. Cependant, ce cas nous semble représentatif de la plupart des réalisations effectuées et, si la stationnarité n'a pas été atteinte, l'utilisation de la moyenne $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ ainsi produite permet cependant d'obtenir une bonne qualité d'estimation. Ainsi sur cet exemple le seuillage de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ ne conservant que les valeurs supérieures à 0.5 permet de localiser parfaitement les quatre intervalles $\mathcal{I}_k$. Les résultats de la seconde phase d'estimation des décalages fréquentiels et des amplitudes sont représentés sur la figure 5.13. L'estimation des décalages au sens de la MP, telle que nous l'avons décrite au tableau 4.5, procure une précision fréquentielle très satisfaisante et

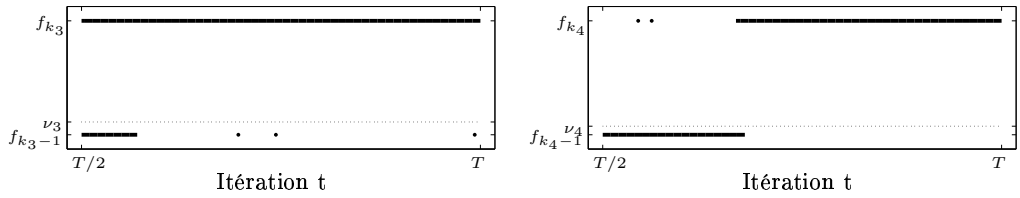


FIG. 5.12 – *Échantillons de Bernoulli* $q_{k_j}^{(t)}$ *au fil des itérations. À gauche, pour les fréquences autour de* ν_3 *; à droite, pour les fréquences autour de* ν_4*. Le trait plein marque les valeurs* $q_{k_j}^{(t)} = 1$.

l'estimation postérieure des amplitudes correspondantes est également de très bonne qualité.

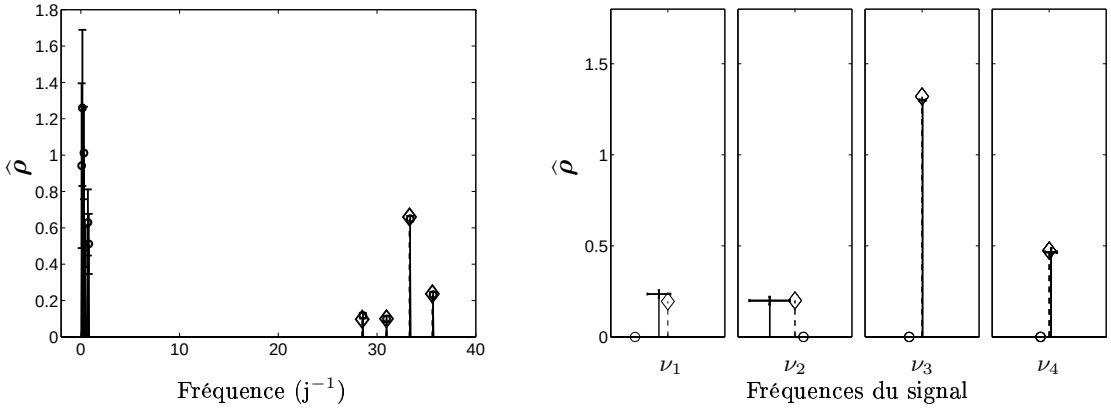


FIG. 5.13 – *Estimation au sens du maximum a posteriori des amplitudes pour les fréquences localisées par* $\widehat{\boldsymbol{q}}_{MP} > 0.5$*. À gauche, sur l'ensemble de la grille fréquentielle : à droite, zoom en fréquence autour de chaque raie. Les barres d'erreur correspondent à* $\pm\sigma_k$*, où* σ_k *est l'écart-type associé, et les* $\diamond$ *représentent les raies théoriques. Chaque fenêtre en fréquence sur la figure de droite couvre une plage de la même largeur qu'à la figure 5.10, soit* $8 \times 10^{-2}\ j^{-1}$.

5.2.4 Bilan

Le tableau 5.1 compare les différents résultats obtenus en termes d'estimation des fréquences des quatre raies contenues dans le signal simulé, qui sont correctement détectées par les trois méthodes mises en œuvre : $\Delta\nu_j$ représente l'erreur d'estimation en valeur absolue sur la fréquence ν_j et $\overline{\Delta\nu}$ leur valeur moyenne. Le coût relativement faible de l'optimisation du critère pénalisé par la norme ℓ^1 devant celui des algorithmes MCMC nous a amenés à calculer l'estimateur résultant dans un cas de *très haute résolution*, avec $P + 1 = 5001$ fréquences positives, c'est-à-dire un vecteur de 10001 inconnues complexes. Si le coût calculatoire est bien plus élevé que pour la grille initiale avec $P = 1000$, il reste cependant inférieur à celui de l'estimation des paramètres des modèles B-$\mathrm{G}_{\mathbf{0}}$ et B-$\mathrm{G}_{\boldsymbol{df}}$. Le résultat de l'estimation n'est pas représenté graphiquement ici mais améliore sensiblement la précision fréquentielle ; notons toutefois que la valeur de ν_2 n'est pas estimée par sa meilleure approximation sur la grille, comme c'était déjà le cas pour une grille de 1001 fréquences.

La modélisation Bernoulli-Gaussienne B-$\mathrm{G}_{\mathbf{0}}$ atteint sur cet exemple de meilleures perfor-

mances que la pénalisation ℓ^1 pour une grille de même taille, mais au prix d'un coût calculatoire beaucoup plus élevé ; l'augmentation du nombre de paramètres pour l'approche pénalisée permet alors d'obtenir des résultats légèrement meilleurs que B-G$_{\mathbf{0}}$ pour un coût de calcul qui reste plus faible. L'extension à la modélisation B-G$_{\boldsymbol{df}}$ permet de s'affranchir de la limite de précision des deux approches précédentes, imposée par le pas de la grille $f_{\max}/P$, et présente l'estimation fréquentielle la plus satisfaisante. Celle-ci s'obtient cependant au prix d'un accroissement non négligeable (de plus de 50%) du coût calculatoire par rapport à l'approche pénalisée, ainsi que du caractère plus délicat du choix d'un critère d'arrêt de l'algorithme d'échantillonnage stochastique. Notons enfin que l'approche par pénalisation mise en œuvre sur une grille encore plus fine de $P+1 = 10001$ fréquences positives n'a pas donné d'estimation plus précise que pour $P = 5000$ sur cet exemple et que le coût de calculatoire correspondant était supérieur à 9000 s, rendant globalement la modélisation B-G$_{\boldsymbol{df}}$ la plus efficace ici.

Signalons enfin que les modélisations B-G et B-G$_{\boldsymbol{df}}$ permettent d'apporter pour chaque fréquence une information supplémentaire en termes de probabilité. Si sur ces données simulées la détection des quatre raies, après une éventuelle fusion des composantes adjacentes, est estimée à 1 (les raies sont présentes sur l'intégralité des échantillons), ce ne sera plus le cas lors de l'application aux données réelles du § 5.3, où cette interprétation supplémentaire sera plus utile.

	$\Delta\nu_1$ (j^{-1})	$\Delta\nu_2$ (j^{-1})	$\Delta\nu_3$ (j^{-1})	$\Delta\nu_4$ (j^{-1})	$\overline{\Delta\nu}$ (j^{-1})	Temps cpu (s)
Pénalisation ℓ^1						
$P = 1000$: $\frac{f_{\max}}{P} = 0.04$ j^{-1}	0.023	0.034	0.009	0.006	0.018	168
$P = 5000$: $\frac{f_{\max}}{P} = 0.008$ j^{-1}	0.007	0.026	0.001	0.002	0.009	2638
Modélisation B-G$_{\mathbf{0}}$						
$P = 1000$: $\frac{f_{\max}}{P} = 0.04$ j^{-1}	0.023	0.006	0.009	0.006	0.011	4076
Modélisation B-G$_{\boldsymbol{df}}$						
$P = 500$	0.006	0.018	0.0004	0.001	0.006	4258

TAB. 5.1 – *Erreurs sur l'estimation des fréquences pour les différentes approches développées et coûts calculatoires associés.*

5.3 Application aux données réelles

5.3.1 Estimation par pénalisation

À la vue des résultats similaires obtenus sur des données simulées pour les pénalisation ℓ^1 et hyperbolique, nous considérons ici uniquement le cas de la première, qui présente l'avantage de fournir un estimateur rigoureusement parcimonieux. La taille de la grille utilisée est ici de $P = 5000$ fréquences positives et l'optimisation est réalisée par un algorithme GSOR arrêté au sens du critère (5.1), ce qui nécessite un temps de calcul de l'ordre de 40 minutes. En notant $\widehat{\boldsymbol{x}}_\lambda$ le minimiseur du critère pénalisé pour une valeur donnée de λ, le tracé de la courbe « en L » $(Q(\widehat{\boldsymbol{x}}_\lambda),R(\widehat{\boldsymbol{x}}_\lambda))$ paramétrée en λ est représenté sur la figure 5.14 et l'estimation obtenue pour des valeurs du paramètre de régularisation dans le « coude » de cette courbe est décrite par les figures 5.15 et 5.16.

Les valeurs de λ supérieures à 10 donnent lieu à un spectre estimé présentant des composantes

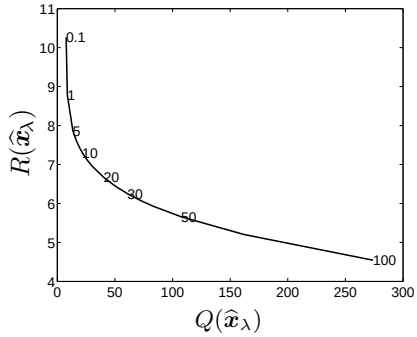


FIG. 5.14 – *Courbe* $(Q(\widehat{\boldsymbol{x}}_\lambda),R(\widehat{\boldsymbol{x}}_\lambda))$ *paramétrée en* λ *pour l'estimation par pénalisation* ℓ^1 *sur des données réelles.*

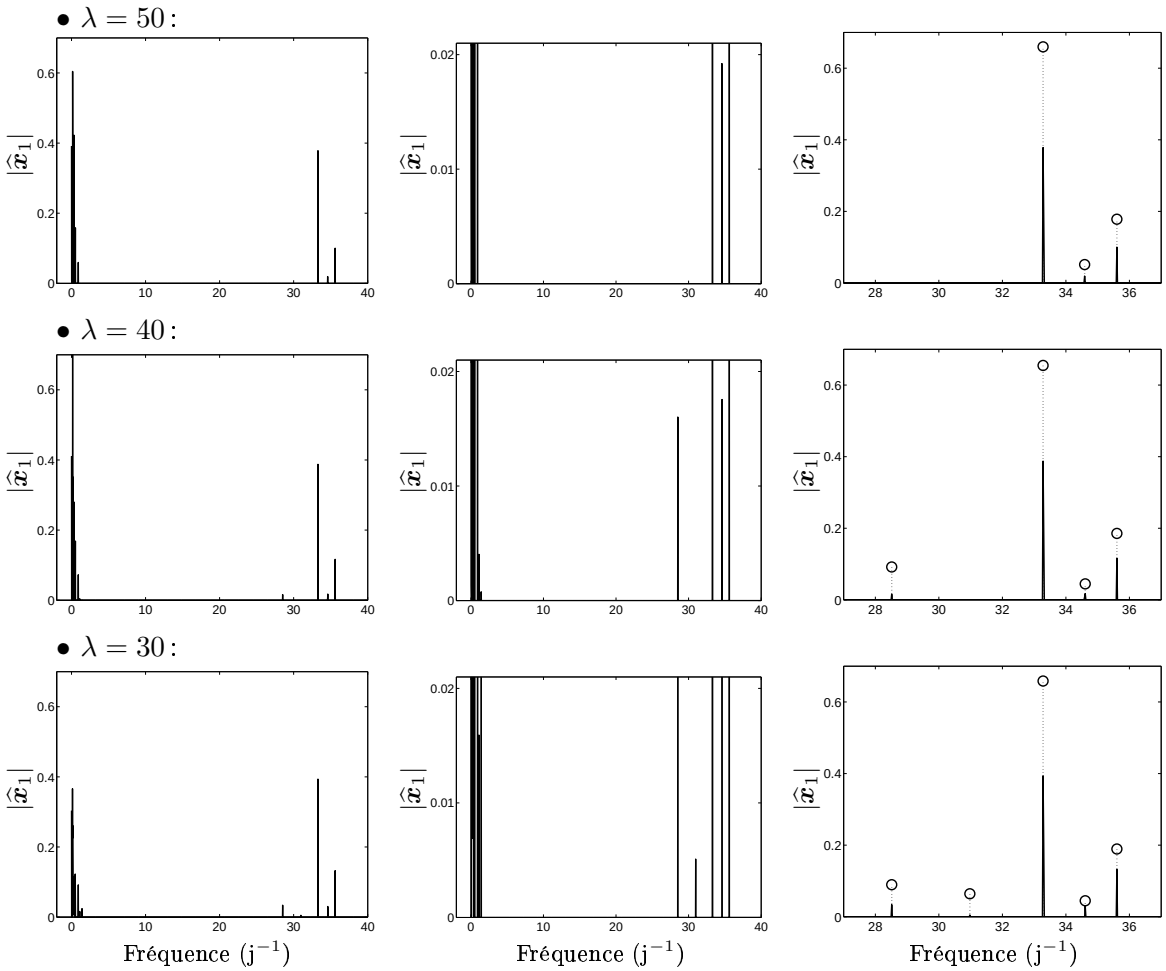


FIG. 5.15 – *Résultats d'estimation par minimisation d'un critère pénalisé pour* $\lambda = 50$ *à* $\lambda = 30$. *À gauche, estimé en module sur l'ensemble de la grille fréquentielle de* $P + 1 = 5001$ *points. Au centre, zoom en amplitude pour les faibles valeurs. À droite, zoom en fréquence sur l'intervalle* $\mathcal{I} = [27\ j^{-1}, 37\ j^{-1}]$. *Les traits pleins représentent le module du spectre estimé et les* $\circ$ *correspondent à la réestimation a posteriori par moindres carrés régularisés des amplitudes localisées par les pics de* $|\widehat{\boldsymbol{x}}_1|$.

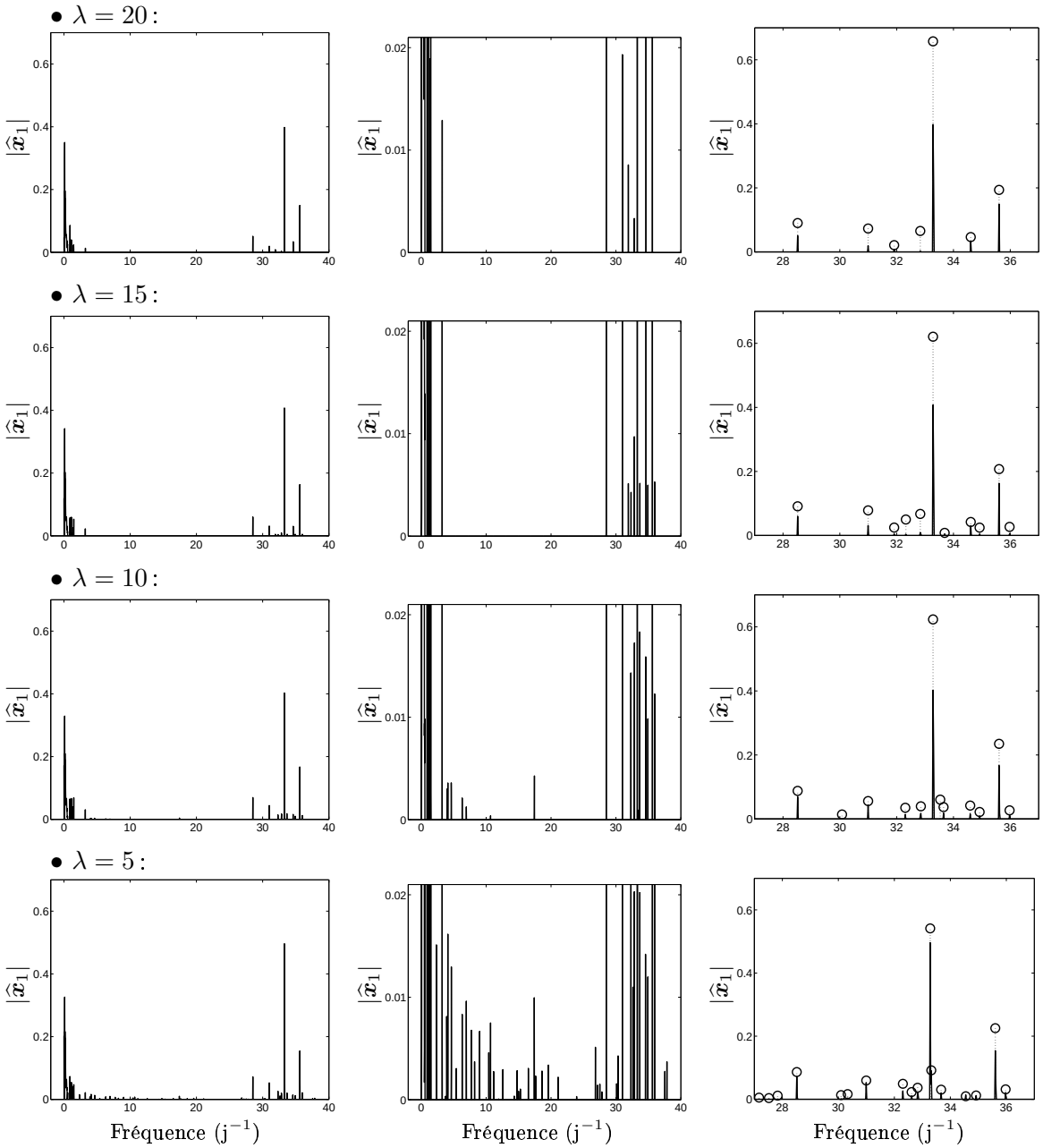


FIG. 5.16 – *Résultats d'estimation par minimisation d'un critère pénalisé pour* $\lambda = 20$ *à* $\lambda = 5$. *À gauche, estimé en module sur l'ensemble de la grille fréquentielle de* $P + 1 = 5001$ *points. Au centre, zoom en amplitude pour les faibles valeurs. À droite, zoom en fréquence sur l'intervalle* $\mathcal{I} = [27\ j^{-1}, 37\ j^{-1}]$. *Les traits pleins représentent le module du spectre estimé et les* $\circ$ *correspondent à la réestimation a posteriori par moindres carrés régularisés des amplitudes localisées par les pics de* $|\widehat{\boldsymbol{x}}_1|$.

à basse fréquence ainsi que des raies dans l'intervalle $\mathcal{I} = [27\ \mathrm{j}^{-1}, 37\ \mathrm{j}^{-1}]$, tout en étant rigoureusement nul dans la plage de fréquences intermédiaire. Pour $\lambda = 50$, trois raies sont détectées dans l'intervalle $\mathcal{I}$, et le nombre de détections augmente au fur et à mesure que le paramètre de régularisation diminue. Pour $\lambda \leq 5$, l'estimateur présente des pics répartis tout au long de la grille fréquentielle et est clairement sous-régularisé.

Si elle permet de sélectionner l'ordre de grandeur du paramètre λ, la courbe $(Q(\widehat{\boldsymbol{x}}_\lambda), R(\widehat{\boldsymbol{x}}_\lambda))$ n'est donc pas suffisante pour définir convenablement l'estimation, puisque le nombre de raies détectées dans l'intervalle $\mathcal{I}$ varie pour une plage relativement étroite de valeurs de λ. Une limitation de l'approche par pénalisation apparaît ici, où le choix entre les différents estimés s'avère assez délicat. Une méthode empirique visant à rechercher la valeur minimale de λ ne produisant pas visuellement de sous-régularisation semble ici complémentaire et s'est révélée satisfaisante en simulations : lorsque la solution est sous-régularisée, des faux pics apparaissent *tout au long* de l'axe fréquentiel alors que, en accord avec la physique de l'objet étudié, nous recherchons des oscillations *a priori* localisées dans l'intervalle $\mathcal{I}$. Les représentations des figures 5.15 et 5.16 montrent ainsi que pour $\lambda \geq 15$ le spectre estimé est nul en dehors de $\mathcal{I}$ et des basses fréquences, ce qui n'est plus vrai pour une valeur inférieure du paramètre de régularisation.

Nous représentons également sur les figures 5.15 et 5.16 à droite la réestimation *a posteriori* des amplitudes par moindres carrés régularisés comme il a été expliqué au § 5.2.1. L'analyse de ces amplitudes permet alors d'affiner l'estimation : puisque les raies spectrales qui différencient $\widehat{\boldsymbol{x}}_{15}$ de $\widehat{\boldsymbol{x}}_{20}$ sont d'amplitudes plus faibles, une attitude plus conservatrice nous incite à ne retenir que les six raies principales de $\widehat{\boldsymbol{x}}_{20}$. Un agrandissement de l'estimation finalement obtenue dans l'intervalle $[27\ \mathrm{j}^{-1}, 37\ \mathrm{j}^{-1}]$ est présenté sur la figure 5.17, où la hauteur des raies correspond à l'amplitude de chaque sinusoïde détectée, soit $\widehat{\boldsymbol{\rho}} = 2|\widehat{\boldsymbol{x}}_1|$.

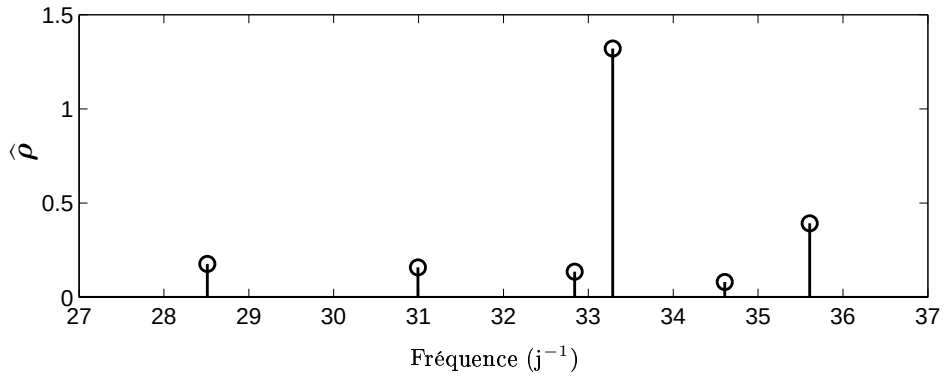


FIG. 5.17 – *Pénalisation ℓ^1 : réestimation par moindres carrés régularisés des amplitudes associées aux fréquences détectées, représentées dans l'intervalle* $[27\ j^{-1}, 37\ j^{-1}]$.

5.3.2 Estimation par modélisation B-G

Nous avons vu aux simulations du § 5.2 que la mise en œuvre des algorithmes stochastiques comportait un point critique en la définition d'une condition d'arrêt. Les résultats présentés dans ce paragraphe, ainsi que ceux du paragraphe suivant pour le modèle B-G$_{\boldsymbol{df}}$, ont été obtenus par l'algoritme G$_3$ du tableau 4.4, en fixant le nombre d'itérations à $T = 100000$. Le temps de calcul nécessaire à la génération d'un tel nombre d'échantillons est bien sûr très élevé – de

l'ordre de la dizaine d'heures – mais permet de s'affranchir du choix d'une condition d'arrêt de l'échantillonnage qui, comme nous l'avons vu sur les simulations du § 5.2, peut s'avérer problématique. L'estimation de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ est représentée en figure 5.18 pour une grille de $P+1 = 1001$ fréquences positives. Hormis la localisation de composantes de basses fréquences, nous pouvons

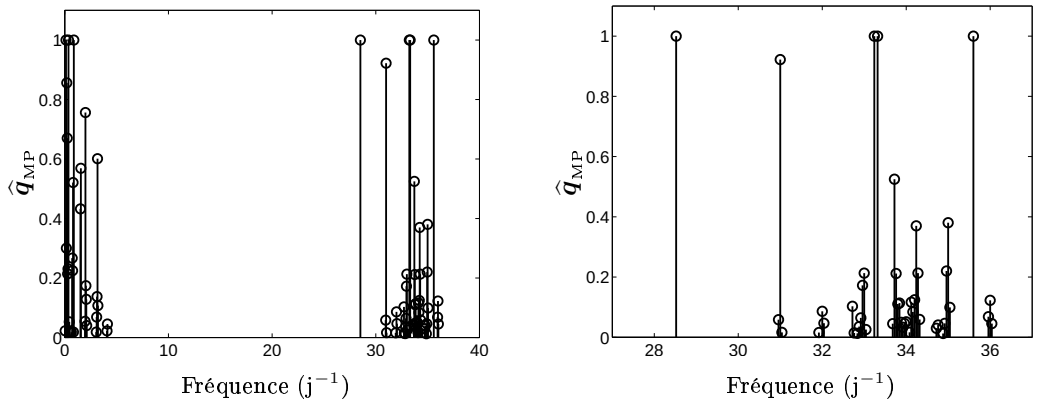


FIG. 5.18 – *Modèle B-G_0 : moyenne estimée $\widehat{\boldsymbol{q}}_{MP}$. À gauche, sur l'ensemble de la grille fréquentielle de $P+1 = 1001$ points. À droite, entre 27 et 37 j^{-1}. Afin d'améliorer la lisibilité, seules les valeurs supérieures à 10^{-2} ont été représentées.*

en extraire les informations suivantes dans la plage fréquentielle de 27 à 37 j^{-1} qui nous intéresse :

- Quatre raies sont détectées avec une probabilité estimée à 1, où deux d'entre elles sont localisées à des fréquences très proches. Plus précisément, $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ prend une valeur de 1 pour les indices correspondant aux fréquences 28.52, 33.24, 33.32 et 35.6 j^{-1}. La résolution de la grille fréquentielle étant de $f_{\max}/P = 0.04$ j^{-1}, les deux fréquences vers 33 j^{-1} ne sont séparées que par une valeur quasiment nulle de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$.
- L'obtention de valeurs adjacentes de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ non nulles, inférieures à 1 et de somme quasiment égale à 1 vers la fréquence $f_{k_1} = 31$ j^{-1} traduit la présence d'une raie dont la localisation alterne entre les fréquences voisines f_{k_1-1}, f_{k_1} et f_{k_1+1}. L'agrandissement de la figure 5.19 montre ainsi qu'une raie en f_{k_1} est présente sur 92.2% des échantillons et que, lorsque $q_{k_1}^{(t)} = 0$, on a en général $q_{k_1-1}^{(t)} = 1$ ou $q_{k_1+1}^{(t)} = 1$, de sorte qu'une raie est présente en l'une de ces trois fréquences sur 99.6% des échantillons.
- Des comportements similaires sont également à noter autour des fréquences $f_{k_2} = 33.72$ j^{-1}, $f_{k_3} = 34.24$ j^{-1} et $f_{k_4} = 35$ j^{-1}, représentés respectivement sur la figure 5.20 en haut, au centre et en bas. Leur interprétation est cependant plus délicate que dans le cas de f_{k_1}. Pour f_{k_2}, par exemple, neuf valeurs non nulles adjacentes de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ sont obtenues entre 33.5 et 34 j^{-1} et la figure 5.20 en haut à droite, représentant l'évolution des échantillons de Bernoulli correspondants au fil des itérations, montre leur complémentarité : si la raie la plus fréquente en $f_{k_2} = 33.72$ j^{-1} n'est présente que sur 55% des échantillons, nous remarquons qu'une raie est présente parmi les fréquences $\{f_k\}_{k=k_2-1,\ldots,k_2+7}$ sur environ 94% des échantillons. Dans le deuxième cas, une raie est présente parmi les fréquences $\{f_k\}_{k=k_3-1,\ldots,k_3+2}$ sur environ 98% des échantillons alors que pour le troisième, une raie est présente parmi les fréquences $\{f_k\}_{k=k_4-2,\ldots,k_4+1}$ sur seulement 76% des échantillons.

L'analyse ci-dessus permet donc d'affirmer la détection de quatre raies spectrales principales avec une probabilité de 1 ainsi que celle d'une raie de fréquence approximative f_{k_1} avec une pro-

• Fréquences entre 30.9 et 31.1 j^{-1} ($f_{k_1} = 31\ \mathrm{j}^{-1}$) :

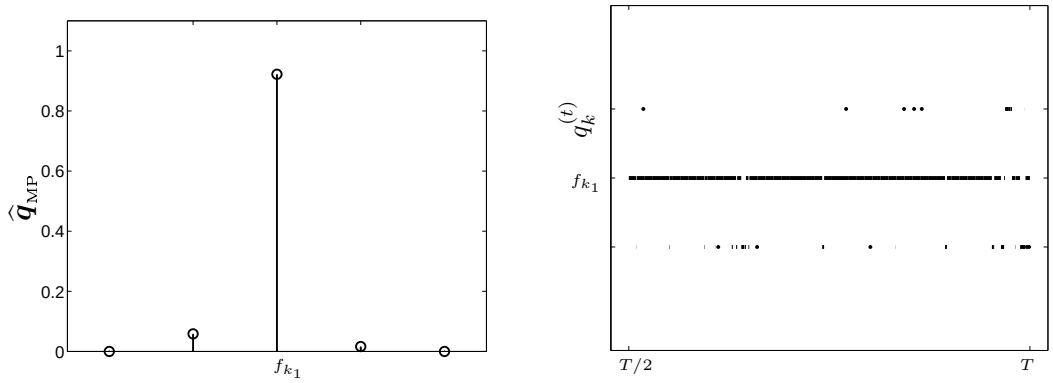
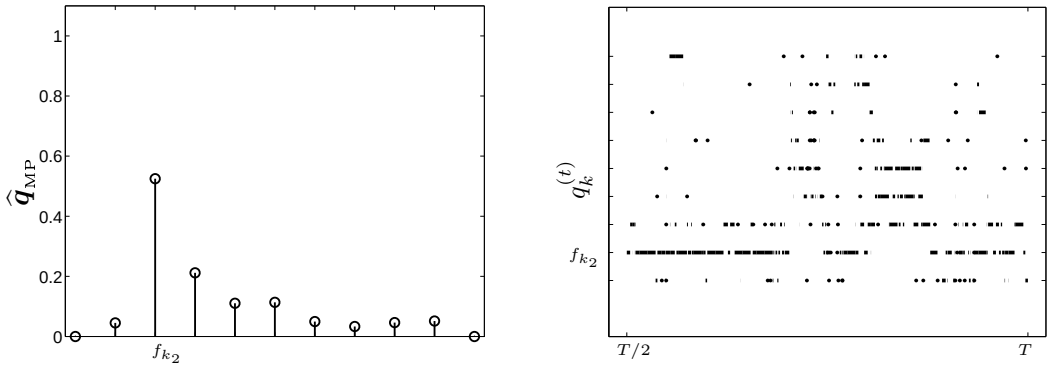


FIG. 5.19 – *Modèle B-G_0 : mise en évidence de la variabilité des échantillons $\boldsymbol{q}^{(t)}$ autour de la fréquence f_{k_1}. À gauche, agrandissement de l'estimation $\widehat{\boldsymbol{q}}_{MP}$ dans la plage fréquentielle concernée. À droite, évolution des échantillons de Bernoulli $q_k^{(t)}$ correspondants au fil des itérations.*

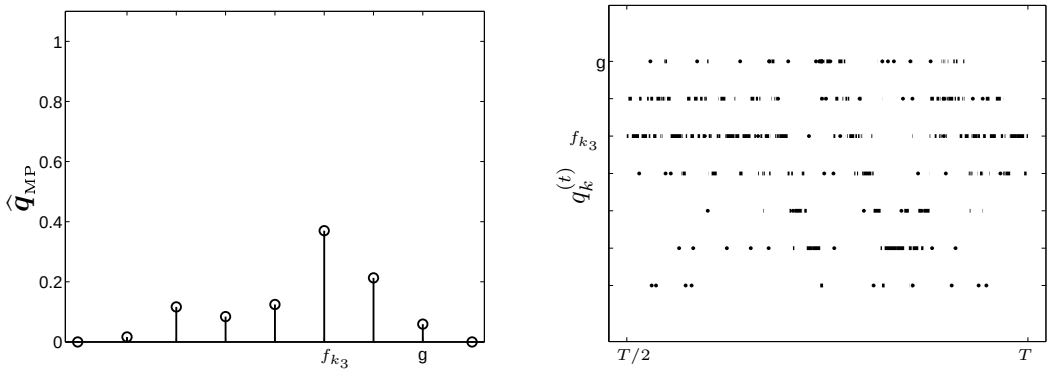
babilité très proche de 1 et de deux raies de fréquences approximatives f_{k_2} et f_{k_3} auxquelles nous pouvons également associer de fortes probabilités d'existence. Cependant, l'allure de la chaîne $\boldsymbol{q}^{(t)}$ produite aux environs de $f_{k_i,\, i=1,2,3}$ rend délicate une estimation précise de leur fréquence. La décision la plus « logique » consiste alors à considérer que ces fréquences sont effectivement localisées en $f_{k_i,\, i=1,2,3}$ et calculer les paramètres d'amplitude correspondants : plutôt que d'effectuer un seuillage de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ comme nous l'avons fait jusqu'à présent, les observations précédentes permettent de sélectionner les positions fréquentielles de manière plus fine, tout en leur associant une probabilité d'existence.

En ne conservant que les sept fréquences ainsi sélectionnées et les composantes à basse fréquence, dont nous ne conservons que la partie pour laquelle $\widehat{\boldsymbol{q}}_{\mathrm{MP}} > 0.5$, l'estimation MAP des amplitudes correspondantes est représentée sur la figure 5.21. Les modules des amplitudes associées à chaque fréquence ainsi détectée sont représentés ainsi que les barres d'erreur à plus ou moins l'écart-type, comme nous l'avons expliqué au § 4.3.3.

- Fréquences entre 33.63 et 34.05 j^{-1} ($f_{k_2} = 33.72\ \mathrm{j}^{-1}$) :



- Fréquences entre 34.03 et 34.37 j^{-1} ($f_{k_3} = 34.24\ \mathrm{j}^{-1}$) :



- Fréquences entre 34.83 et 35.09 j^{-1} ($f_{k_4} = 35\ \mathrm{j}^{-1}$) :

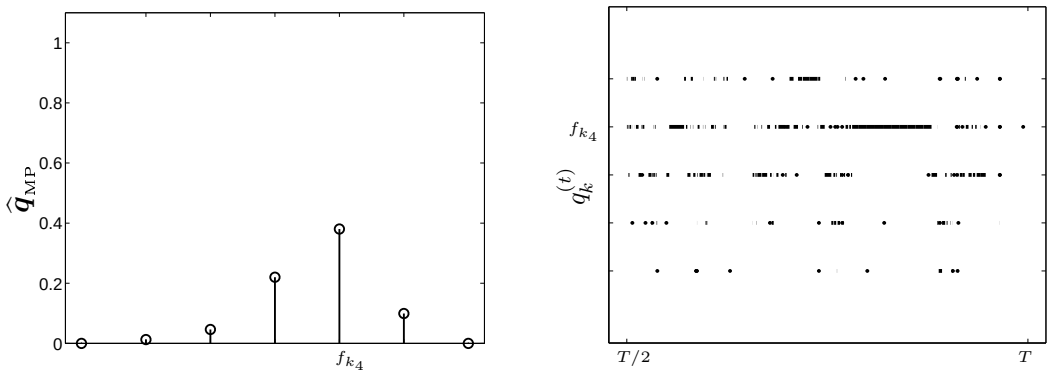


FIG. 5.20 – *Modèle B-G_0 : mise en évidence de la variabilité des échantillons $\boldsymbol{q}^{(t)}$ autour des fréquences f_{k_2}, f_{k_3} et f_{k_4}. À gauche, agrandissement de l'estimation $\widehat{\boldsymbol{q}}_{MP}$ dans chaque plage fréquentielle. À droite, évolution des échantillons de Bernoulli $q_k^{(t)}$ correspondants au fil des itérations.*

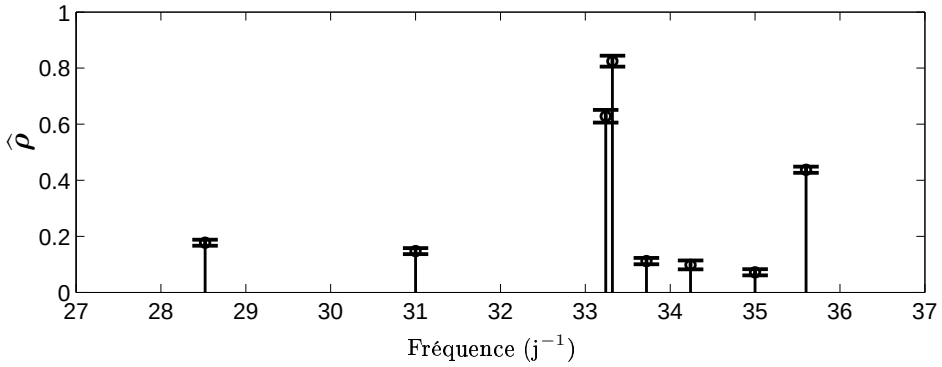


FIG. 5.21 – *Modèle B-G_0 : estimation au sens du maximum a posteriori des amplitudes associées aux fréquences sélectionnées et barres d'erreur à plus ou moins l'écart-type, représentées dans l'intervalle* $[27\ j^{-1}, 37\ j^{-1}]$.

5.3.3 Estimation par modélisation B-G avec décalages en fréquence

L'estimation de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ obtenue après $T = 100000$ itérations de l'algorithme d'échantillonnage du tableau 4.6 – obtenue pour un coût calculatoire d'environ cinq heures – est représentée sur la figure 5.22. De manière similaire au paragraphe précédent, une interprétation détaillée permet

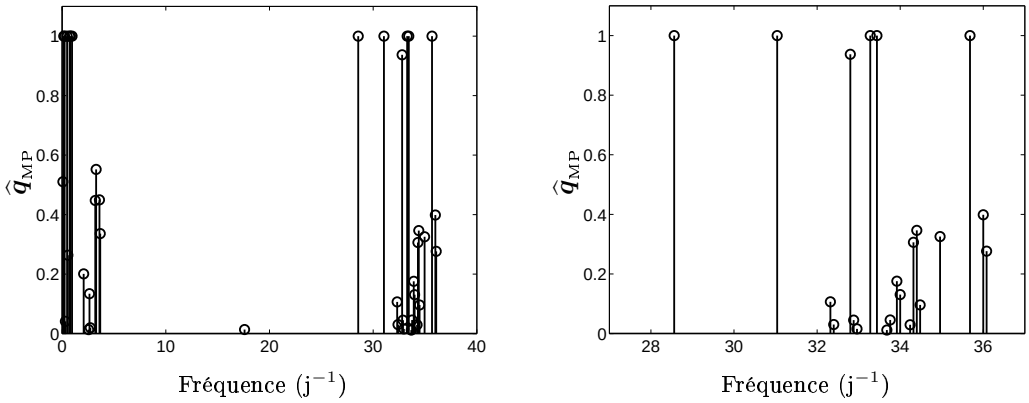


FIG. 5.22 – *Modèle B-$G_{\mathbf{df}}$: moyenne estimée $\widehat{\boldsymbol{q}}_{MP}$. À gauche, sur l'ensemble de la grille fréquentielle de $P = 501$ points. À droite, entre 27 et 37 j^{-1}. Afin d'améliorer la lisibilité, seules les valeurs supérieures à 10^{-2} ont été représentées.*

d'en extraire des informations intéressantes. En rappelant que $\widehat{\boldsymbol{q}}_{\mathrm{MP}_k}$ est une estimation de la probabilité *a posteriori* d'existence d'une raie dans l'intervalle $\mathcal{I}_k = [f_{k-1}, f_k] = \left[\frac{k-1}{P} f_{\max}, \frac{k}{P} f_{\max}\right]$, la figure 5.22 permet d'effectuer l'analyse suivante, hormis la localisation de composantes de basses fréquences.

- Cinq fréquences sont détectées avec une probabilité estimée à 1 ;
- Des valeurs adjacentes de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ non nulles et inférieures à 1 sont obtenues, traduisant la présence d'une raie dont la localisation alterne au cours des itérations entre les différents intervalles concernés. La figure 5.23 en haut à gauche montre ainsi plusieurs valeurs non nulles de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ autour de $f_{k_1} = 32.8$ j^{-1}, en notant k_1 l'indice de la grille dans

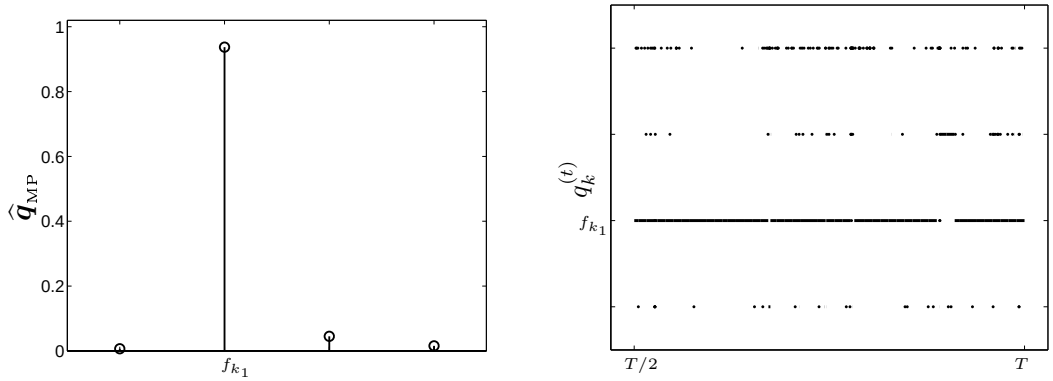
cette zone pour lequel la probabilité $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ est la plus élevée. On a en effet $\widehat{q}_{\mathrm{MP}_{k_1}} \simeq 94\%$, mais également $\widehat{q}_{\mathrm{MP}_{k_1+1}} \simeq 5\%$ et $\widehat{q}_{\mathrm{MP}_{k_1+2}} \simeq 2\%$. La figure 5.23 en haut à droite met en évidence la « complémentarité » de ces valeurs adjacentes, dans le sens où lorsque la raie n'est pas présente dans l'intervalle $\mathcal{I}_{k_1}$, elle l'est généralement dans un des intervalles voisins $\mathcal{I}_{k_1+1}$ ou $\mathcal{I}_{k_1+2}$: sur cet exemple, nous observons ainsi qu'une raie est présente dans $U_{k=k_1}^{k1+2}\mathcal{I}_k = [f_{k_1-1}, f_{k_1+2}]$ sur 98% des échantillons. De manière analogue, les figures 5.23 au centre et en bas représentent les probabilités $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ estimées autour des fréquences $f_{k_2} = 34.4$ j^{-1} et $f_{k_3} = 36$ j^{-1} respectivement, ainsi que l'évolution des échantillons $q_k^{(t)}$ correspondant aux différentes composantes adjacentes. Nous observons ainsi qu'une raie est présente dans l'intervalle $[f_{k_2-3}, f_{k_2+1}]$ sur 78% des échantillons, ainsi qu'une raie dans $[f_{k_3-1}, f_{k_3+1}]$ sur 67% des échantillons. Le comportement de la chaîne autour des fréquences f_{k_j} traduit donc la présence d'une unique raie, à laquelle nous pouvons associer une probabilité *a posteriori* d'existence.

– Des comportements similaires sont également obtenus autour des fréquence 32.32 j^{-1} et 34 j^{-1}, mais pour lesquels les valeurs cumulées des probabilités *a posteriori* associées ne dépassent pas 40 % ; par conséquent, nous ne les prendrons pas en compte pour la suite de l'estimation.

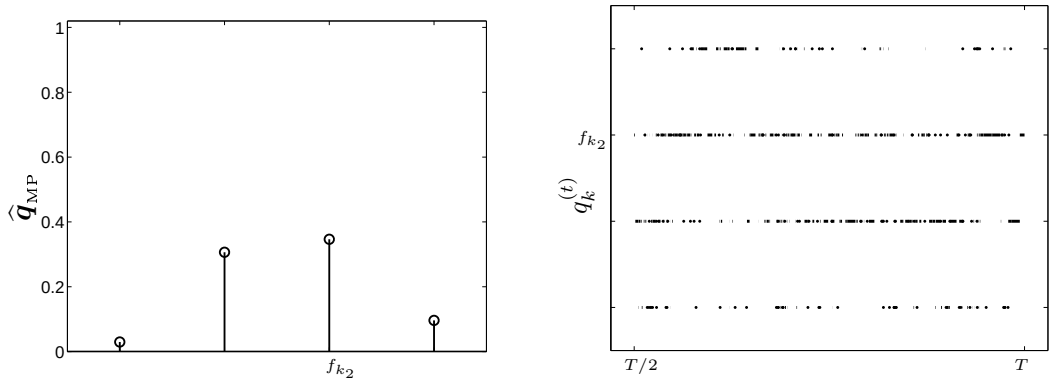
Cette interprétation postérieure délicate, mais pourtant nécessaire, des échantillons générés reflète un point problématique du modèle B-G$_{\boldsymbol{df}}$: sur cet exemple, l'évolution des fréquences détectées au fil des itérations, c'est-à-dire des valeurs $f_k - df_k^{(t)}$ pour $q_k^{(t)} = 1$, présente une variabilité d'amplitude supérieure au pas de discrétisation de la grille fréquentielle[6]. Si l'étape de détection peut être effectuée *a posteriori* en fusionnant des localisations adjacentes, la deuxième phase d'estimation des décalages associés s'en trouve également compliquée, et la stratégie consistant à suivre l'évolution, au fil des itérations, d'une même composante fréquentielle ainsi identifiée, est problématique. En effet, les représentations de la colonne de droite de la figure 5.23 montrent que les échantillons $q_k^{(t)}$ correspondant à deux positions adjacentes peuvent également prendre des valeurs non nulles simultanément. Nous proposons ici pour l'estimation des fréquences de considérer uniquement les intervalles donnant lieu à la plus grande probabilité *a posteriori*, c'est-à-dire les intervalles $\mathcal{I}_{k_j}$, mais ne pouvons pas alors de cette façon calculer de variances associées. L'estimation finale des fréquences et des amplitudes est représentée sur la figure 5.24.

6. La solution consistant à employer des intervalles $\mathcal{I}_k$ plus larges, c'est-à-dire une valeur plus faible de P, a également été testée pour une grille de $P+1 = 251$ fréquences positives, où le même comportement a été obtenu.

- Fréquences entre 32.7 et 33 j^{-1} ($f_{k_1} = 32.8\ \mathrm{j}^{-1}$) :



- Fréquences entre 34.2 et 34.5 j^{-1} ($f_{k_2} = 34.4\ \mathrm{j}^{-1}$) :



- Fréquences entre 35.9 et 36.1 j^{-1} ($f_{k_3} = 36\ \mathrm{j}^{-1}$) :

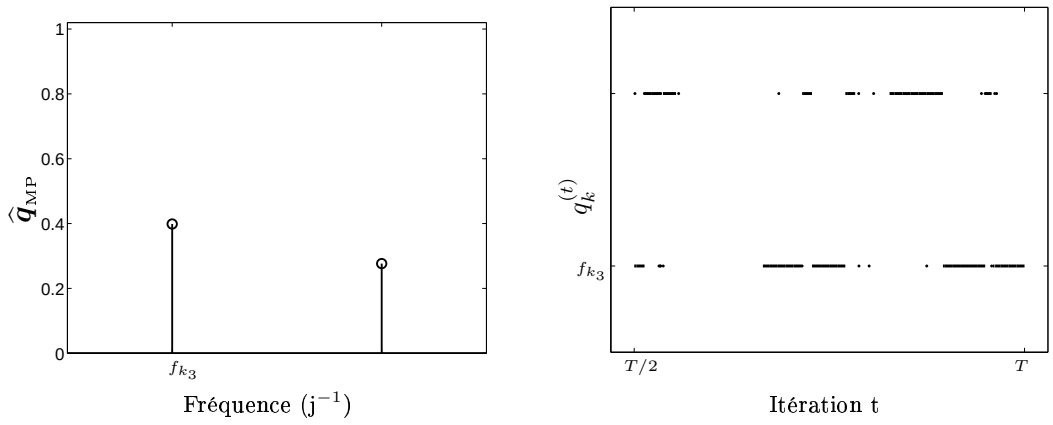


FIG. 5.23 – *Modèle B-G$_{\mathbf{df}}$: Mise en évidence de la variabilité des échantillons $\boldsymbol{q}^{(t)}$ autour des fréquences f_{k_1}, f_{k_2} et f_{k_3}. À gauche, agrandissement de l'estimation $\widehat{\boldsymbol{q}}_{MP}$ dans chaque plage fréquentielle. À droite, évolution des échantillons de Bernoulli $q_k^{(t)}$ correspondants au fil des itérations.*

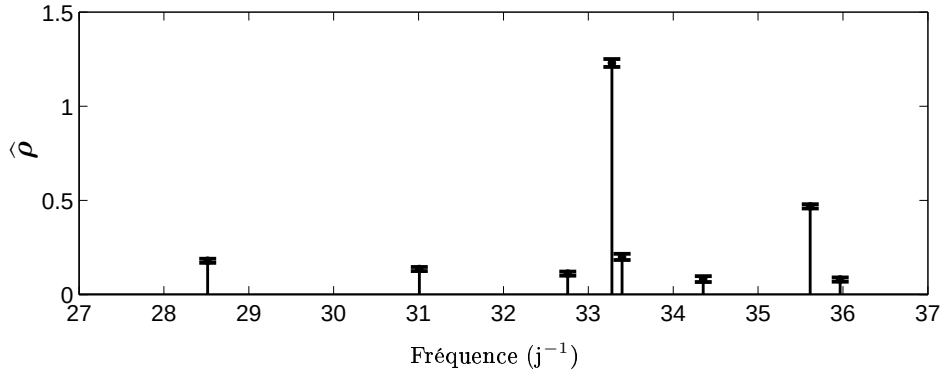


FIG. 5.24 – *Modèle B-G$_{df}$: estimation au sens du maximum a posteriori des amplitudes associées aux fréquences sélectionnées et barres d'erreur en amplitude à plus ou moins l'écart-type, représentées dans l'intervalle* $[27\ j^{-1}, 37\ j^{-1}]$.

5.3.4 Bilan

Le tableau 5.2 compare les résultats d'estimation des raies dans l'intervalle fréquentiel $[27\ \text{j}^{-1}, 37\ \text{j}^{-1}]$ obtenus sur les données brutes, ainsi que les résultats obtenus par Böhm *et al.* [2004] à partir des données corrigées.

	Pénalisation ℓ^1 $\frac{f_{\max}}{P} = 0.008$ j^{-1}		Modélisation B-G$_0$ $\frac{f_{\max}}{P} = 0.04$ j^{-1}			Modélisation B-G$_{df}$			Böhm *et al.* [2004] précision 0.04 j^{-1}		
	Fréq. (j^{-1})	Amp. (km.s^{-1})	Fréq. (j^{-1})	Amp. (km.s^{-1})	Proba. (%)	Fréq. (j^{-1})	Amp. (km.s^{-1})	Proba. (%)	Fréq. (j^{-1})	Amp. (km.s^{-1})	Conf. ($\geq$ %)
ν_1	33.288	1.321	33.32	0.83	100	33.278	1.230	100	33.289	1.320	99.9
			33.24	0.63	100	33.398	0.200	100			
ν_2	35.608	0.393	35.6	0.44	100	35.614	0.468	100	35.606	0.474	99.9
ν_3	28.512	0.178	28.52	0.18	100	28.516	0.180	100	28.503	0.195	99.9
ν_4	30.992	0.158	31	0.15	99.6	31.009	0.136	100	30.954	0.139	99
ν_5	32.840	0.136	33.72	0.11	94	32.76	0.112	98	32.616	0.105	99
ν_6	34.608	0.081	34.24	0.10	98	34.35	0.082	78	33.862	0.099	99
			35	0.07	76	35.97	0.080	67			

TAB. 5.2 – *Bilan comparatif de l'estimation obtenue par différentes méthodes sur les données réelles. Les méthodes que nous avons proposées sont ici appliquées aux données brutes alors que l'estimation proposée par Böhm et al. [2004] a été effectuée sur les données corrigées.*

Nous notons tout d'abord que, si la raie principale à une fréquence ν_1 d'approximativement 33.28 j^{-1} est détectée par l'ensemble des méthodes (rappelons qu'elle apparaît également dans le spectre de Fourier des données brutes sur la figure 5.3), les estimations pour les deux modélisations de type B-G présentent au lieu d'une seule raie deux composantes fréquentielles proches, avec une probabilité estimée à 1. Un agrandissement de la plage fréquentielle concernée montre que la moyenne $\widehat{\boldsymbol{q}}_{\text{MP}}$ estimée dans cette zone présente dans les deux cas une configuration de la forme $(\widehat{q}_{\text{MP}k_1-1}, \widehat{q}_{\text{MP}k_1}, \widehat{q}_{\text{MP}k_1+1}) = (1,0,1)$, où k_1 est l'indice de la fréquence sur la grille $f_{k_1} = \frac{k_1}{f_{\max}} = 33.28$ j^{-1} : éclairés par les considérations du § 4.4.1 sur le blocage de l'algorithme d'échantillonnage dans une configuration de ce type et par l'estimation obtenue avec

les autres méthodes, nous pensons que cette configuration traduit l'existence d'une unique raie dont la fréquence est encadrée par f_{k_1-1} et f_{k_1+1}, ce que nous supposerons par la suite. Nous reviendrons bien entendu sur ce point problématique lors de la conclusion de ce chapitre.

Les quatre raies de plus grande amplitude détectées par l'approche pénalisée coïncident approximativement avec les quatre raies principales obtenues par Böhm *et al.*, et l'écart maximal sur l'estimation des quatre fréquences entre ces deux approches est de l'ordre de 0.04 j^{-1}. L'estimation des deux raies de plus faible amplitudes diffèrent cependant entre ces deux approches. De manière similaire, les quatre raies principales détectées par l'approche par modélisation B-$\mathbf{G_0}$ correspondent à l'approximation la plus proche des fréquences détectées par les deux méthodes précédentes sur la grille de reconstruction employée, qui est de « seulement » $P + 1 = 1001$ fréquences positives dans le cas du modèle B-$\mathbf{G_0}$.

Les résultats obtenus par la modélisation B-$\mathbf{G_{df}}$, où l'estimation des décalages est effectuée de la manière décrite au § 5.3.3, concordent encore de manière approximative pour les quatre premières raies, et la différence maximale sur l'estimation de leurs fréquences est inférieure à 0.02 j^{-1} par rapport à l'estimation par pénalisation et de l'ordre de 0.06 j^{-1} par rapport aux estimations de Böhm *et al.*. Notons que ces quatre raies communes à toutes les méthodes de détection se voient associer, pour les deux modélisations Bernoulli-Gaussiennes, une probabilité estimée très proche de 1, alors que les autres raies détectées par ces deux approches B-G le sont avec une probabilité plus faible. **Cette analyse nous incite donc à ne conserver avec une confiance très élevée que les quatre raies $\nu_1 \dots \nu_4$.**

Notons enfin que la fréquence ν_5, non détectée par la modélisation B-$\mathbf{G_0}$, est présente pour l'approche pénalisée, la modélisation B-$\mathbf{G_{df}}$ (avec une probabilité de l'ordre de 98%) et aussi chez Böhm *et al.*, où l'estimation de sa fréquence varie d'au maximum 0.22 j^{-1} entre les différentes méthodes.

5.3.5 Comparaison avec les résultats sur les données corrigées

Nous présentons pour finir les résultats d'estimation obtenus à partir des données auxquelles l'essentiel des basses fréquences a été retiré [Böhm *et al.*, 2004], représentées sur la figure 5.1 à la première page de ce chapitre. Rappelons une fois encore que cette correction n'est pas effectuée de manière exacte et que, si elle permet de supprimer la majeure partie des perturbations à basse fréquence (et donc d'appliquer des algorithmes de type CLEAN), elle induit aussi nécessairement des erreurs supplémentaires.

Pour les résultats présentés ci-après, le nombre d'inconnues utilisé pour chaque méthode est identique à celui utilisé pour le traitement des données brutes et les différentes options réglant le fonctionnement des algorithmes sont conservées (critères d'arrêt pour l'approche par pénalisation, nombre d'itérations et choix des lois *a priori* sur les hyperparamètres pour les modélisations B-$\mathbf{G_0}$ et B-$\mathbf{G_{df}}$).

Estimation par pénalisation

La figure 5.25 présente l'estimation obtenue par la pénalisation ℓ^1 pour une grille de $P = 5000$ fréquences positives et un paramètre de régularisation λ entre 10 et 25. Dans chaque cas, l'estimateur présente très peu de composantes à basses fréquences et des raies localisées dans l'intervalle $[27\ \mathrm{j}^{-1}, 37\ \mathrm{j}^{-1}]$. Une démarche similaire à celle effectuée au § 5.3.1 conduit

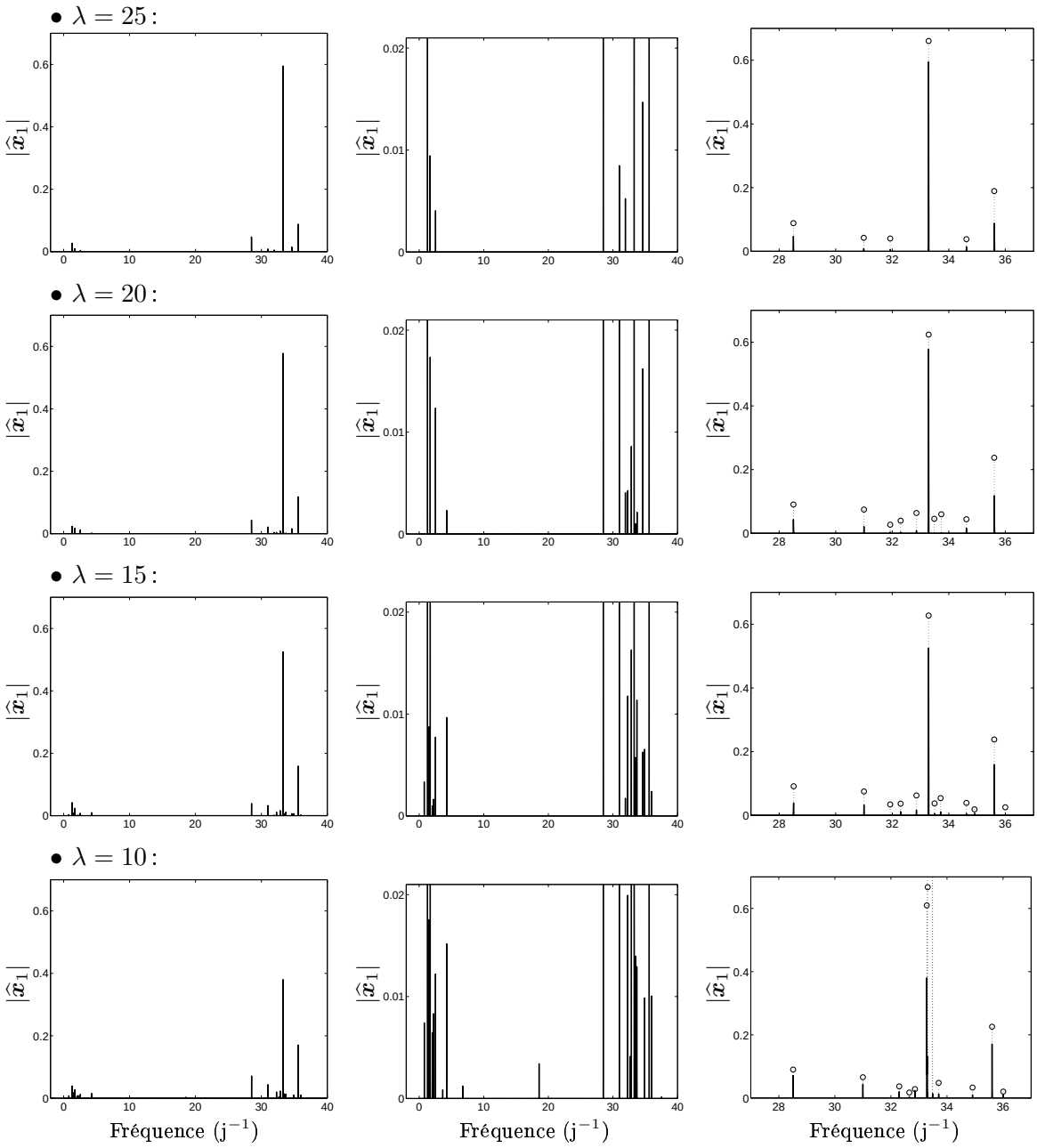


FIG. 5.25 – *Données corrigées : résultats d'estimation par minimisation d'un critère pénalisé pour* $\lambda = 25$ *à* $\lambda = 10$. *À gauche, estimé en module sur l'ensemble de la grille fréquentielle de* $P + 1 = 5001$ *points. Au centre, zoom en amplitude pour les faibles valeurs. À droite, zoom en fréquence sur l'intervalle* $[27\ j^{-1}, 37\ j^{-1}]$. *Les traits pleins représentent le module du spectre estimé et les* $\circ$ *correspondent à la réestimation a posteriori par moindres carrés régularisés des amplitudes localisées par les pics de* $|\widehat{\boldsymbol{x}}_1|$.

à choisir $\lambda = 15$, valeur en-dessous de laquelle le spectre estimé présente des pics « parasites » en dehors de la plage fréquentielle où les raies sont recherchées.

Un agrandissement de l'estimation obtenue dans l'intervalle $[27\ \mathrm{j}^{-1}, 37\mathrm{j}^{-1}\ ]$ est présenté sur la figure 5.26, où la hauteur des raies localisées par $\widehat{\boldsymbol{x}}_1$ correspond à l'amplitude de la sinusoïde associée à la détection de chaque raie, réestimée *a posteriori* par moindres carrés régularisés. Cette solution fait apparaître douze raies, ce qui est supérieur au nombre de raies détectées sur les données brutes. Le comparatif établi à la fin de ce paragraphe montrera cependant que seules les six raies de plus forte amplitude sont également détectées par les autres méthodes.

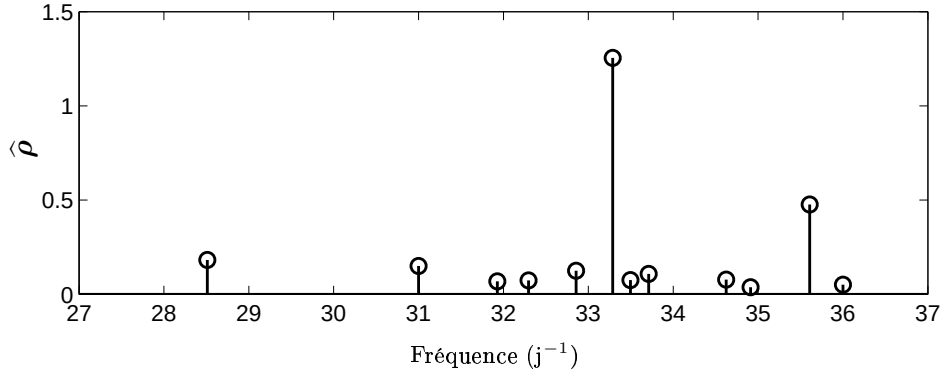


FIG. 5.26 – *Données corrigées et pénalisation ℓ^1 : réestimation par moindres carrés régularisés des amplitudes associées aux fréquences détectées, représentée dans l'intervalle* $[27\ j^{-1}, 37\ j^{-1}]$.

Estimation par modélisation B-G

Pour la modélisation Bernoulli-Gaussienne, l'estimation de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ est représentée en figure 5.27 pour une grille de $P+1 = 1001$ fréquences positives. Cette estimation présente une allure similaire à celle obtenue en figure 5.18 sur les données brutes, comportant à la fois des valeurs égales à 1 (raies détectées sur l'ensemble des échantillons) et des valeurs adjacentes *complémentaires*. Nous notons donc en particulier que la « suppression » des basses fréquences ne réduit pas la variabilité des paramètres de fréquence le long de la chaîne d'échantillons $\boldsymbol{q}^{(t)}$.

Comme au § 5.3.2, une étape postérieure de regroupement de composantes adjacentes est donc nécessaire pour effectuer un seuillage adéquat de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ en termes de probabilité. Afin d'éviter des redondances avec l'analyse effectuée sur les données brutes au § 5.3.2 – la démarche est ici identique – les détails de ce regroupement ne sont pas exposés ici. La figure 5.28 présente l'estimation des amplitudes (ainsi que les variances associées) correspondant aux raies pour lesquelles nous pouvons associer une probabilité supérieure à 60%. Huit raies sont ainsi détectées, dont deux sont localisées en des positions adjacentes autour de 33.28 j^{-1}. Cependant, la variance associée aux amplitudes de ces deux raies est plus élevée que celle des autres raies spectrales, laissant entrevoir un comportement « suspect » de l'estimation pour cette fréquence : ce phénomène a déjà été rencontré sur les simulations du § 5.2.2, traduisant l'existence d'une unique raie aux alentours de 33.28 j^{-1}.

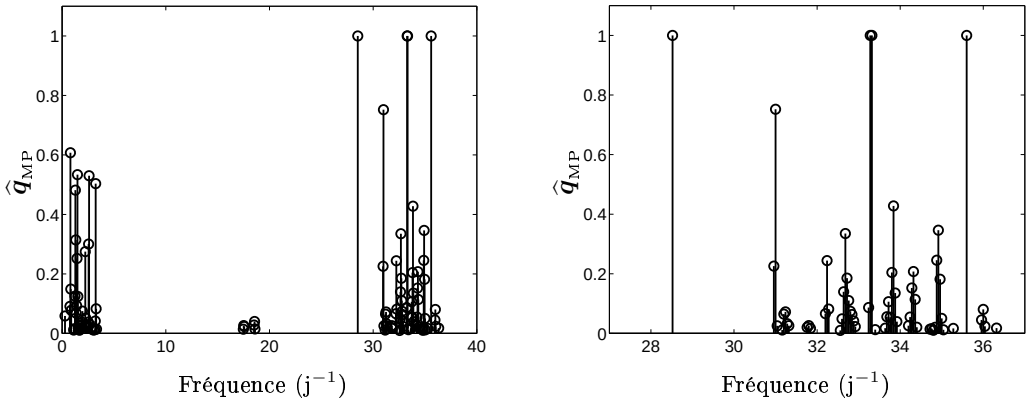


FIG. 5.27 – *Données corrigées : moyenne estimée $\widehat{\boldsymbol{q}}_{MP}$ pour la modélisation B-$G_{\mathbf{0}}$. À gauche, sur l'ensemble de la grille fréquentielle de $P+1 = 1001$ points. À droite, entre 27 et 37 j^{-1}. Afin d'améliorer la lisibilité, seules les valeurs supérieures à 10^{-2} ont été représentées.*

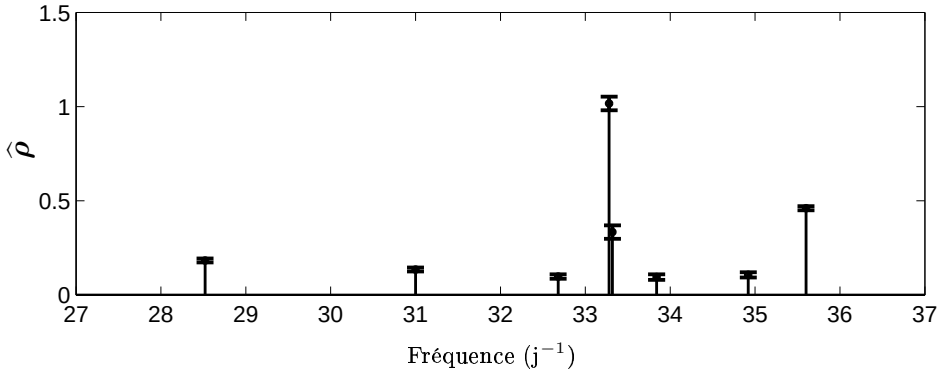


FIG. 5.28 – *Données corrigées et modèle B-$G_{\mathbf{0}}$: estimation au sens du maximum a posteriori des amplitudes associées aux fréquences sélectionnées et barres d'erreur à plus ou moins l'écart-type, représentées dans l'intervalle* $[27\ j^{-1}, 37\ j^{-1}]$.

Estimation par modélisation B-G avec décalages en fréquence

L'estimation de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ obtenue pour une modélisation Bernoulli-Gaussienne prenant en compte des décalages en fréquence est représentée sur la figure 5.29 et présente une alllure similaire à celle obtenue sur les données brutes en figure 5.22. De nouveau, un regroupement *a posteriori* des

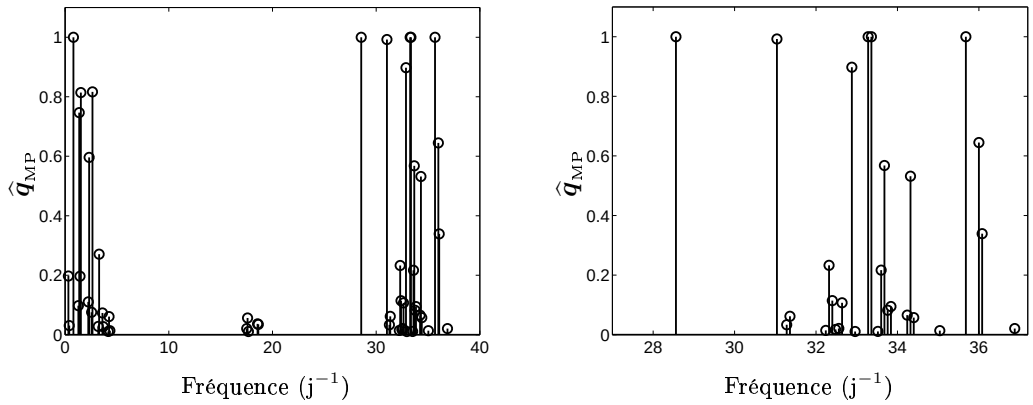


FIG. 5.29 – *Données corrigées : moyenne estimée $\widehat{\boldsymbol{q}}_{MP}$ pour la modélisation B-G$_{\boldsymbol{df}}$. À gauche, sur l'ensemble de la grille fréquentielle de $P = 501$ points. À droite, entre 27 et 37 j^{-1}. Afin d'améliorer la lisibilité, seules les valeurs supérieures à 10^{-2} ont été représentées.*

valeurs adjacentes significatives de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ est nécessaire. L'estimation finale des fréquences et des amplitudes est représentée sur la figure 5.30, où seules sont conservées les raies auxquelles sont associées des probabilités supérieures à 60 %. Les variances en amplitude sont également représentées, mais pas celles sur les paramètres fréquentiels : comme nous l'avons expliqué au § 5.3.3, une trop grande variabilité de ces paramètres (dont l'amplitude dépasse le pas de la grille) rend délicate l'estimation de cette variance à partir des échantillons $\boldsymbol{df}^{(t)}$.

Huit raies sont ainsi détectées dont deux correspondent, comme dans le cas de la modélisation B-G$_{\mathbf{0}}$ appliquée aux mêmes données, à des localisations adjacentes sur la grille fréquentielle (c'est-à-dire à deux valeurs consécutives de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ égales à 1). Ici encore, des variances plus élevées pour les amplitudes de ces deux raies laissent supposer l'existence d'une unique raie de fréquence approximative 33.28 j^{-1}.

Bilan

Le tableau 5.3 compare les résultats d'estimation des raies dans l'intervalle fréquentiel [27 j^{-1}, 37 j^{-1}] obtenus sur les données corrigées ainsi que les résultats obtenus par Böhm *et al.* [2004] à partir de ces mêmes données, utilisant les programmes du logiciel PERIOD04 basé sur le principe de l'algorithme CLEAN.

En comparant les résultats, pour chaque méthode, à ceux obtenus à partir des données brutes, consignés dans le tableau 5.2, nous pouvons conclure sur les points suivants :

- pour l'approche pénalisée, les cinq raies détectées avec les plus grandes amplitudes correspondent à celles obtenues à partir des données brutes : les localisations sur la grille sont identiques pour ν_1, ν_2 et ν_3, diffèrent de 0.008 j^{-1} pour ν_4 et de 0.016 j^{-1} pour ν_5. Les raies d'amplitude plus faible, par contre, ne sont « détectées » que dans le cas des données

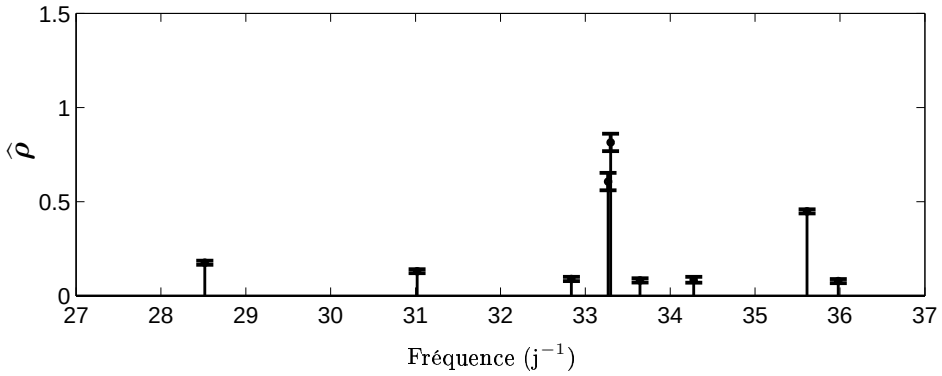


FIG. 5.30 – *Données corrigées et modèle B-G$_{df}$: estimation au sens du maximum a posteriori des amplitudes associées aux fréquences sélectionnées et barres d'erreur en amplitude à plus ou moins l'écart-type, représentées dans l'intervalle* $[27\ j^{-1}, 37\ j^{-1}]$.

	Pénalisation ℓ^1 $\frac{f_{\max}}{P} = 0.008$ j^{-1}		Modélisation B-G$_0$ $\frac{f_{\max}}{P} = 0.04$ j^{-1}			Modélisation B-G$_{df}$			Böhm *et al.* [2004] précision 0.04 j^{-1}		
	Fréq. (j^{-1})	Amp. (km.s^{-1})	Fréq. (j^{-1})	Amp. (km.s^{-1})	Proba. (%)	Fréq. (j^{-1})	Amp. (km.s^{-1})	Proba. (%)	Fréq. (j^{-1})	Amp. (km.s^{-1})	Conf. ($\geq$ %)
ν_1	33.288	1.255	33.28	1.016	100	33.298	0.815	100	33.289	1.320	99.9
			33.32	0.334	100	33.268	0.607	100			
ν_2	35.608	0.476	35.60	0.461	100	35.612	0.449	100	35.606	0.474	99.9
ν_3	28.512	0.182	28.52	0.183	100	28.517	0.176	100	28.503	0.195	99.9
ν_4	31	0.150	31	0.135	100	31.018	0.131	100	30.954	0.139	99
ν_5	32.856	0.125	32.68	0.010	90	32.836	0.090	91	32.616	0.105	99
ν_6	33.712	0.108	33.84	0.095	91	33.645	0.082	96	33.862	0.099	99
	34.624	0.078	34.92	0.107	86	34.278	0.086	65			
	33.496	0.075				35.980	0.079	98			
	32.296	0.0735									
	31.928	0.069									
	36.000	0.05									
	34.912	0.038									

TAB. 5.3 – *Bilan comparatif de l'estimation obtenue par les différentes méthodes sur les données réelles corrigées.*

corrigées. Par ailleurs, celles-ci ne sont pas détectées par l'ensemble des autres méthodes, ce qui peut donc correspondre à une sous-régularisation de la solution.

- pour la modélisation B-G$_{\mathbf{0}}$, six des sept raies conservées sont de fréquences similaires pour les deux jeux de données : les fréquences ν_1 à ν_4 sont localisées aux mêmes positions de la grille, les deux estimations de ν_5 diffèrent de 0.12 j^{-1} et celles de ν_6 diffèrent de 0.08 j^{-1}. Seule une raie, localisée en 34.24 j^{-1} à partir des données brutes, laisse place ici à une raie en 32.68 j^{-1}.

- pour la modélisation B-G$_{\boldsymbol{df}}$, les valeurs des fréquences ν_1 à ν_4 sont estimées dans les deux cas à des valeurs très proches : hormis le problème de dédoublement pour ν_1, l'écart maximal en fréquence est de 0.009 j^{-1}. Parmi les quatre autres raies détectées (avec une probabilité estimée plus faible), trois étaient déjà présentes à des fréquences semblables lors du traitement des données brutes, et une raie supplémentaire à 33.64 j^{-1} apparaît ici.

Ces résultats permettent, de manière similaire à l'analyse des données brutes, de distinguer **quatre raies « presque sûres » de fréquences ν_1 à ν_4** (une probabilité estimée à 1 est associée à leur détection par les deux modélisations B-G). Celles-ci sont détectées par l'ensemble des méthodes proposées, ainsi que par l'application d'un algorithme de type CLEAN proposée par Böhm *et al.* [2004]. Ensuite, deux raies (notées ν_5 et ν_6) sont également présentes pour chaque méthode (pour les modélisations B-G, la probabilité estimée est inférieure à celle des quatre premières raies) mais dont la localisation varie de manière non négligeable (± 0.2 j^{-1}) selon la méthode utilisée. En raison de la grande complexité des données d'une part, et du faible niveau des amplitudes correspondantes d'autre part, il semble cependant plus raisonnable de ne pas tenir compte de ces détections secondaires. *A fortiori* les raies correspondant aux dernières lignes du tableau 5.3, qui ne sont détectées que de manière propre à chacune des méthodes d'estimation, sont plus probablement imputables à des erreurs de modélisation.

Nous remarquons que le problème du dédoublement de la raie en ν_1 apparaît encore pour les modélisations B-G. À la différence du cas des données brutes (pour lesquelles la configuration locale de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ est de type 1-0-1), ce dédoublement se manifeste ici par des valeurs adjacentes non nulles de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$. Si dans le cas de la modélisation B-G$_{\mathbf{0}}$ ce dédoublement peut s'expliquer par l'inexactitude de la discrétisation fréquentielle, nous avons vu que ce phénomène était encore présent pour la modélisation B-G$_{\boldsymbol{df}}$, aussi bien sur le traitement des données brutes que sur celui des données corrigées. À notre sens, ce problème est imputable aux différents prétraitements subis en amont par les données : rappelons que les données brutes ne sont pas directement des observations, mais résultent de différentes calibrations et corrections, aussi bien sur l'axe des temps que sur celui des amplitudes. Il n'est alors pas surprenant que ces modifications produisent par exemple un « glissement » de la fréquence de la raie principale au cours des observations, pouvant être responsable de ce dédoublement apparent.

La complexité de ces données ne réside donc finalement pas, pour les méthodes proposées, dans la présence des perturbations à basse fréquence : ces résultats confirment les résultats des simulations du § 5.2, à savoir une bonne robustesse devant ce genre de perturbations. Les difficultés proviennent plutôt du jeu de données en lui-même qui, en raison des différentes phases de prétraitement, est probablement assez éloigné du modèle « multi-sinusoïdal + bruit blanc » sous-jacent aux modélisations utilisées. Par conséquent, si elle permet l'application de méthodes de type CLEAN, il ne semble pas que la « correction » des basses fréquences améliore l'estimation obtenue par nos différentes méthodes. Elle induit même probablement des perturbations supplémentaires dans des données déjà passablement délicates à traiter.

5.4 Conclusion

Nous avons comparé dans ce chapitre le comportement des méthodes d'estimation proposées pour le traitement de données complexes simulées, puis en application à des données réelles. Sur l'ensemble des algorithmes, nous avons dû adapter les tests d'arrêt afin de les rendre plus « permissifs » devant la complexité des données. Si le critère de convergence des algorithmes d'optimisation pour l'approche par pénalisation peut être associé à une tolérance numérique sur l'évolution négligeable des paramètres ou du gradient, le contrôle du nombre d'échantillons MCMC nécessaires pour l'estimation de la moyenne *a posteriori* s'est avéré plus problématique.

Dans un premier temps, l'étude sur un signal artificiel construit sur la forme des données réelles a permis de mettre en évidence la robustesse des estimateurs proposés qui, malgré la présence de composantes à basse fréquence dominant le contenu spectral des données, détectent correctement les quatre raies. Notons que l'utilisation d'algorithmes dérivés de la méthode CLEAN tels que ceux du programme PERIOD04 utilisé par Böhm *et al.* [2004] est compromise pour l'analyse de telles données, de par la prédominance des basses fréquences dans leur spectre de Fourier. Cette robustesse vis-à-vis des perturbations de basses fréquences représente un atout important des méthodes proposées, puisque ce type de perturbations est fréquemment rencontré dans les séries temporelles astrophysiques.

Sur ces simulations, l'approche pénalisée s'est révélée la moins coûteuse en temps de calcul et a permis d'obtenir une haute résolution fréquentielle en considérant une grille de discrétisation très fine. Le choix du paramètre de régularisation est cependant apparu comme un point délicat, et « la courbe en L » ne permet qu'une détermination grossière d'une valeur acceptable. Une technique *ad hoc* consiste à choisir la valeur la plus faible ne débouchant pas sur une solution sous-régularisée, qui s'est montrée sur ces données assez facilement identifiable par l'apparition de pics disséminés tout au long de la grille fréquentielle. Par opposition, les approches basées sur une modélisation Bernoulli-Gaussienne sont entièrement non supervisées sur le plan de l'algorithme d'échantillonnage, mais ont parfois nécessité une étape postérieure d'interprétation des résultats qui n'est pas automatique. Celle-ci a pu néanmoins être effectuée de manière satisfaisante en préservant l'interprétation en termes de probabilités propre à l'estimation de la moyenne *a posteriori* de la séquence de Bernoulli. Si la précision de la modélisation B-$G_{\mathbf{0}}$ est limitée par le pas de la grille, qui ne peut être réduit dans les mêmes proportions que pour l'approche pénalisée, l'extension au modèle B-$G_{\boldsymbol{df}}$ a permis d'obtenir sur ces simulations les meilleures performances en termes d'estimation des fréquences, bien qu'à un coût considérablement plus élevé que l'approche pénalisée. La mise en œuvre des algorithmes MCMC s'est cependant révélée d'une grande lourdeur et la difficulté à atteindre la stationnarité des chaînes de Markov générées a révélé la difficulté qui réside dans le choix d'un critère au sens duquel l'exploration de la distribution *a posteriori* échantillonnée est jugée suffisante.

Le passage au traitement des données réelles a mis en évidence plusieurs sortes de complications : rappelons que la série temporelle étudiée, de par la présence de composantes de basses fréquences mais aussi en raison des différents prétraitements subis en amont, est vraisemblablement encore plus éloignée d'un modèle multi-sinusoïdal que ne l'est le signal simulé utilisé au début de ce chapitre. Si l'approche par pénalisation a présenté des caractéristiques similaires à celles obtenues en simulation ainsi que des résultats compatibles, pour les quatre raies principales sur les six détectées, avec les résultats obtenus par Böhm *et al.* [2004] sur les données corrigées, l'estimation des paramètres des modèles Bernoulli-Gaussiens classique et avec décalages en fréquence s'est avérée plus délicate.

D'une part, les deux estimations B-G ont montré (du moins sur la réalisation de l'algorithme stochastique présentée) une configuration critique de dédoublement d'une raie alors que les algorithmes d'échantillonnage développés ont pour objectif d'éviter de tels problèmes. D'autre part, une difficulté majeure est causée par une trop grande variabilité de la localisation des fréquences détectées le long de la chaîne générée. L'alternance de cette localisation en des fréquences adjacentes de la grille fréquentielle pour B-$G_{\mathbf{0}}$ (en des intervalles adjacents dans le cas de B-$G_{\boldsymbol{df}}$) complique alors considérablement l'estimation par rapport au cas de données simulées. Une interprétation adéquate *a posteriori* nous a cependant permis de pallier en partie ces difficultés et de procéder à une étape de détection pour laquelle l'association d'une probabilité a encore un sens, et les deux modélisations B-G ont également permis de détecter quatre raies principales avec une probabilité estimée très proche de 1, et dont les paramètres coïncident approximativement avec ceux obtenus par l'approche pénalisée ainsi que par Böhm *et al.* [2004] sur les données corrigées.

Que ce soit sur les données brutes ou corrigées, nous obtenons ainsi la détection de quatre raies de manière « presque sûre », alors que les *niveaux de confiance* obtenus par PERIOD04 classent seulement trois raies avec une confiance très élevée, la quatrième étant détectée au même niveau de confiance que deux autres raies, la détection de celles-ci étant plus amibuë à notre sens.

Vu la difficulté représentée par le traitement des données brutes – sur lesquelles, rappelons-le, aucune méthode d'analyse spectrale classique n'a fourni de résultat – les résultats obtenus sur cet exemple sont donc finalement très encourageants.

Chapitre 6

Conclusion et perspectives

Ce travail de recherche a été consacré à l'étude de méthodes d'estimation de raies spectrales dans un contexte astrophysique, où les données recueillies sont généralement à échantillonnage irrégulier. Nous avons abordé l'analyse spectrale à haute résolution sous la forme d'un problème inverse sous-déterminé, où le spectre à reconstruire est discrétisé sur une grille de fréquences arbitrairement fine. La recherche de spectres de raies s'effectue alors par la régularisation de la solution, en recherchant une solution de forme *parcimonieuse* à ce problème. Deux formes de régularisation ont été étudiées : la minimisation d'un critère pénalisé et la modélisation probabiliste Bernoulli-Gaussienne, pour laquelle une extension prenant en compte des décalages des localisations fréquentielles a également été envisagée. Alors que la contrainte d'un échantillonnage irrégulier réduit l'efficacité de certains algorithmes d'optimisation pour l'approche pénalisée, la procédure d'estimation proposée pour la modélisation Bernoulli-Gaussienne est indifférente à la forme de l'échantillonnage sur le plan calculatoire ; l'espacement irrégulier des données rend toutefois le problème d'analyse spectrale plus complexe.

Pour la première approche, la pénalisation par la norme ℓ^1 opérant sur les modules des amplitudes complexes traduit de manière très satisfaisante l'information de parcimonie : les spectres obtenus dans ce cas présentent des valeurs exactement nulles en la plupart des points de la grille. Alors que la pénalisation ℓ^1 est habituellement appliquée à des variables réelles, nous avons montré l'avantage de son utilisation sur les modules des amplitudes complexes qui représente une meilleure information *a priori* pour l'estimation de spectres de raies. Notons que la pénalisation introduit un biais sur l'estimation des amplitudes et nécessite une ré-estimation postérieure de ces dernières une fois les fréquences détectées ; cette phase est cependant immédiate et ne pose généralement aucun problème.
Du point de vue algorithmique, l'utilisation d'une pénalisation strictement convexe de type $\ell^2\ell^1$ permet de traiter un problème d'optimisation différentiable et de prendre en compte, par des algorithmes de type gradient, les propriétés structurelles de l'opérateur dans le cas de l'échantillonnage irrégulier. Cependant, l'optimisation du critère pénalisé par la norme ℓ^1 s'est également avérée plus avantageuse sur le plan calculatoire. La mise en œuvre d'un algorithme de type Gauss-Seidel, réalisant successivement des minimisations sur chaque variable scalaire, revêt pour une telle pénalisation une forme très simple, qui n'avait pourtant jamais été exploitée à notre connaissance, et qui nous a permis d'envisager une analyse à très haute résolution pour un coût de calcul raisonnable. Pour cette même pénalisation, nous avons montré que l'algorithme IRLS pouvait bénéficier d'une mise en œuvre efficace pour une dimension très élevée de l'espace des paramètres. Bien que la convergence de cet algorithme ne soit pas garantie dans ce cadre,

nous avons établi une condition nécessaire et suffisante complémentaire permettant, aux erreurs numériques près, de tester *a posteriori* l'optimalité de la solution.
Le réglage automatique du paramètre de régularisation lorsque les caractéristiques du bruit sont inconnues reste cependant un point délicat. Le choix de la valeur produisant visuellement l'estimation la plus satisfaisante, s'il s'est avéré aisé lors des simulations, s'est montré d'interprétation plus ambiguë dans le cas des données astrophysiques plus complexes étudié au dernier chapitre.

La modélisation Bernoulli-Gaussienne des amplitudes spectrales, conjuguée à une estimation bayésienne au sens de la moyenne *a posteriori*, fournit quant à elle une procédure totalement non supervisée. Bien que l'estimation de la séquence de Bernoulli moyenne, localisant les fréquences des raies, ne soit pas rigoureusement parcimonieuse, elle peut être traduite en termes de probabilités *a posteriori* de détection des raies spectrales. Cette interprétation, ainsi que l'estimation de la variance sur chaque paramètre, représente un apport informationnel non disponible pour l'approche pénalisée, et s'est notamment montrée utile pour sélectionner les raies significatives lors de l'application à des données réelles.
L'estimation des fréquences et des amplitudes à partir d'un tel modèle s'est révélée satisfaisante sur des données simulées. Si le pas de discrétisation de la grille ne peut être réduit dans les mêmes proportions que pour l'approche pénalisée en raison d'un coût calculatoire plus élevé, l'extension à un modèle prenant en compte des décalages en fréquence a permis d'améliorer sensiblement la précision fréquentielle de l'estimation. Cette méthode fournit ainsi les meilleures performances parmi les approches proposées en termes de qualité de l'estimation, donnant de surcroît accès à des estimations de la variance sur chacune des fréquences et des amplitudes estimées.
Sur le plan algorithmique, nous avons proposé dans le cas du modèle Bernoulli-Gaussien classique deux variantes de l'échantillonneur de Gibbs permettant d'en améliorer sensiblement les performances. La version G_3 notamment, réalisant des étapes d'échantillonnage par blocs de trois composantes adjacentes, est une alternative élégante évitant le piégeage de l'échantillonneur autour de minima locaux. Ici encore, l'extension à un modèle Bernoulli-Gaussien avec décalages en fréquence nous a permis, par la mise en œuvre d'un algorithme d'échantillonnage hybride efficace, de réduire la dimension de la grille fréquentielle (et donc le coût calculatoire de l'échantillonnage) tout en améliorant la précision de l'estimation.

En pratique, nous avons cependant mis en évidence deux sources de complications venant entacher la clarté de ce cadre théorique. D'une part, les algorithmes MCMC proposés restent bien plus coûteux en temps de calcul que les algorithmes d'optimisation de l'approche pénalisée pour un problème de même taille. Nous avons également montré l'ambiguïté du choix d'un critère d'arrêt pour la procédure d'échantillonnage, notamment lorsque l'algorithme reste piégé au voisinage d'un mode local de la distribution échantillonnée. D'autre part, l'estimation à partir de l'échantillonnage de la distribution *a posteriori* n'est en général pas totalement « automatique ». Ainsi dans le cas du modèle Bernoulli-Gaussien classique, l'utilisation d'une grille fréquentielle fixe provoque l'alternance des échantillons en des positions adjacentes ne localisant qu'approximativement les fréquences théoriques et nécessite une étape d'interprétation postérieure, qui est toutefois facilement réalisable en général. L'introduction des décalages en fréquence, qui vise en partie à pallier ces problèmes, s'est montrée satisfaisante à ce propos sur l'ensemble des simulations effectuées. Cependant, l'application à des données réelles particulièrement délicates a également débouché sur l'alternance des localisations fréquentielles en des intervalles adjacents.
La contrainte de restauration d'une composante fréquentielle pour chaque point de la grille (ou chaque intervalle dans le cas du modèle à décalages), qui apporte une certaine robustesse de l'estimateur – lors des simulations du dernier chapitre, nous avons ainsi pu estimer de manière très satisfaisante des raies noyées dans un continuum spectral à basse fréquence – a donc également

son revers : un niveau de bruit trop fort ou un écart trop important des données au modèle multi-sinusoïdal provoque une grande variabilité des paramètres de fréquence, d'amplitude supérieure au pas de la grille, rendant alors plus délicate la phase postérieure d'estimation de l'ensemble des paramètres. Dans ce cas, l'estimation obtenue par la pénalisation ℓ^1 est préférable, qui présente des caractéristiques plus robustes : les algorithmes d'optimisation, dont la vitesse de convergence est certes ralentie par la difficulté de modéliser le spectre continu à basse fréquence avec un *a priori* de parcimonie, ne présentent pas ce caractère pathologique et l'estimateur obtenu, *pour un choix adéquat du paramètre de régularisation*, a toujours une structure rigoureusement parcimonieuse.

En terme d'outils d'analyse spectrale exploitables sur des données astrophysiques, les différentes procédures que nous avons proposées présentent donc un réel apport méthodologique. Pour des données perturbées par des basses fréquences comme celles que nous avons traitées au chapitre 5, les deux approches envisagées ont permis de réaliser une estimation assez satisfaisante, ne nécessitant pas le recours à une préalable correction qui n'est pas toujours réalisable en pratique. Sur de telle données brutes, l'application de procédures de type CLEAN n'ont pas fourni de résultat satisfaisant. Dans le cas de données plus proches d'un modèle multi-sinusoïdal, l'estimation par un algorithme CLEAN *correctement supervisé* permet en général d'extraire des raies spectrales de manière satisfaisante. Cependant, les méthodes proposées présentent un avantage certain par rapport à de tels algorithmes. L'approche pénalisée fournit une méthode assez rapide qui n'est supervisée que par le réglage du paramètre de régularisation, cette étape étant assez intuitive lorsque les données contiennent suffisamment d'information. Si elle est plus lourde à mettre en œuvre, l'approche par modélisation Bernoulli-Gaussienne, notamment dans sa version avec décalages en fréquence, permet d'atteindre la précision du continu (qui est celle des algorithmes CLEAN) et fournit les informations supplémentaires de niveau de confiance et de variance sur l'ensemble des paramètres spectraux estimés. En particulier, la sélection du nombre de composantes significatives (qui correspond au test d'arrêt de CLEAN dont nous avons vu l'ambiguïté de l'interprétation) peut être ici effectuée en termes probabilistes plus rigoureux.
En plus des améliorations sur le plan technique que nous décrivons plus loin, ces outils demandent encore quelques étapes de perfectionnement « esthétique » afin de pouvoir être directement utilisés par nos collègues astrophysiciens. Leur codage dans un langage plus performant devrait notamment permettre de réduire le temps de calcul qui reste une forte pénalisation des méthodes développées. Une interface graphique serait également bienvenue afin d'assister l'utilisateur dans différentes phases de l'estimation comme le réglage du paramètre de régularisation pour l'approche pénalisée ou la sélection des fréquences significatives pour les modélisations Bernoulli-Gaussienne.

Sur le plan méthodologique, certains points demanderaient une étude plus approfondie, parmi lesquels :

- le réglage du paramètre de régularisation pour l'approche pénalisée, qui jusqu'à présent nécessite l'interprétation de l'utilisateur. L'estimation de ce paramètre, par exemple en exploitant une approche bayésienne de type maximum de vraisemblance généralisée ou par maximum de vraisemblance au moyen d'algorithmes EM, pourrait être envisagée, bien qu'au prix d'un accroissement conséquent du coût calculatoire ;
- le couplage des algorithmes IRLS et de type Gauss-Seidel pour la pénalisation ℓ^1. Ces deux algorithmes ont présenté des caractéristiques complémentaires en termes de coûts calculatoires. Une stratégie consistant à fusionner les phases avantageuses de chaque procédure doit être étudiée, mais paraît cependant délicate de par la nature différente des deux

schémas algorithmiques.
- la recherche d'un meilleur critère de convergence des algorithmes d'échantillonnage stochastique proposés dans le cadre de la modélisation Bernoulli-Gaussienne. Même si un tel problème n'est pas encore résolu de façon théorique dans un cadre général [Robert, 1996], il doit être possible de trouver un compromis satisfaisant pour les applications qui nous intéressent ;
- enfin dans le cas défavorable, identifié sur les données réelles, où les fréquences détectées varient au cours de la chaîne sur plusieurs positions de la grille fréquentielle, une estimation prenant en compte l'ensemble de cette évolution est également à envisager.

Les méthodes que nous avons appliquées ici au cas de l'analyse spectrale reposent sur une formulation très générale du problème initial, qui peut être adaptée pour de nombreux problèmes différents. Plus particulièrement dans le cas de l'approche pénalisée, les bonnes performances en termes algorithmiques et informationnels obtenues par la pénalisation ℓ^1 encouragent la poursuite de cette approche dans plusieurs directions :

- Les algorithmes d'optimisation proposés dans le cas de la pénalisation ℓ^1, valables pour des variables réelles ou complexes, sont capables de traiter des problèmes à un nombre très élevé de paramètres. Il faudrait à présent envisager l'utilisation de tels algorithmes pour des problèmes plus généraux de type *Basis Pursuit* utilisant des dictionnaires plus complets que les seules familles d'exponentielles complexes utilisées ici. Les nombreux travaux récents sur ce thème considèrent des fonctions à valeurs réelles, pour lesquelles l'optimisation peut être réalisée par des algorithmes de programmation quadratique. Il serait donc intéressant de poursuivre nos travaux dans cette direction et d'étudier aussi bien leur apport en termes de modélisation (pénalisation sur les modules dans le cas de fonctions à valeurs complexes) qu'en termes d'algorithmes : pour des variables complexes, la programmation quadratique n'est pas satisfaisante et, dans le cas de variables réelles, une comparaison des coûts calculatoires doit être effectuée entre les algorithmes que nous avons proposés et les algorithmes de programmation quadratique habituellement utilisés.
- Dans un contexte astrophysique et toujours pour des données à échantillonnage irrégulier, la recherche de décompositions prenant en compte des non stationnarités du signal (glissement d'une fréquence au fil du temps, excitation temporaire de modes d'oscillation) est d'un grand intérêt. L'extension de l'approche pénalisée au cas de l'analyse temps-fréquence a été étudiée lors du stage de DEA de H. Rabeson [2004] en utilisant une pénalisation hyperbolique sur chaque spectre ainsi qu'une pénalisation traduisant la variation relativement douce des raies en fonction du temps. Une telle régularisation s'est notamment montrée efficace pour des données présentant de larges trous dans l'échantillonnage. L'optimisation, effectuée par des algorithmes de gradients conjugués, s'est cependant révélée très coûteuse. L'utilisation d'une pénalisation ℓ^1 et la construction d'algorithmes adaptés est une alternative pouvant permettre la réduction du coût calculatoire.
- Un travail plus théorique serait également souhaitable sur l'étude des propriétés de parcimonie de la pénalisation ℓ^1 pour des variables complexes. En effet, si nous avons constaté l'obtention de solutions rigoureusement parcimonieuses dans ce cas – qui semble naturelle, au vu de la description des algorithmes d'optimisation associés – les travaux récents sur les représentations parcimonieuses à partir de la norme ℓ^1 considèrent exclusivement des variables réelles. Rappelons également que la convergence de l'algorithme IRLS pour une telle pénalisation n'a toujours pas été établie. Si J.-J. Fuchs semble l'avoir démontrée [Fuchs, 2005, communication privée], cette preuve n'est valable que dans le cas de variables réelles. Il faudrait donc étudier sa généralisation au cas de variables complexes.

Nous signalons pour finir deux perspectives envisageables à partir d'une modélisation de type probabiliste semblable à celle développée dans la deuxième approche :

- Nous avons évoqué au cours du premier chapitre la décomposition en série de Fourier (tronquée) d'un signal périodique pour l'étude des systèmes multiples. Des algorithmes MCMC ont récemment été proposés pour des applications musicales, et il serait intéressant d'étudier leur possible adaptation aux séries temporelles très particulières caractérisant l'étude des systèmes multiples.
- La modélisation Bernoulli-Gaussienne pourrait quant à elle être étendue en traitement d'images – les coordonnées de chaque pixel jouant alors le rôle de la grille – afin d'être appliquée à des problèmes de déconvolution impulsionnelle, fréquemment rencontrés en astrophysique où les étoiles sont assimilables à des objets ponctuels. Si la mise en œuvre résultante risque d'être extrêmement lourde, l'enjeu est en conséquence : l'utilisation d'un modèle prenant en compte des décalages par rapport aux pixels de l'image permet en théorie d'atteindre une précision illimitée sur la localisation des objets lumineux, fournissant également la variance associée à cette estimation, tant en position qu'en amplitude.

Annexe A

Calculs pour l'approche pénalisée

A.1 Critères et gradients

Pour la pénalisation hyperbolique, nous récapitulons ici l'expression des critères et de leurs gradients, permettant de faire apparaître les structures matricielles exploitables dans le cas de l'échantillonnage irrégulier. Nous ne détaillons que la partie quadratique, sur laquelle porte la différence entre les différents critères. Les notations utilisées sont celles du chapitre 3.

La dérivation d'une fonction $f : \mathbb{C} \to \mathbb{R}$ s'écrit $f'(u) = 2\dfrac{\partial g(u,u^*)}{\partial u^*}$, où $f(u) = g(u,u^*)$ et en considérant indépendamment u et u^* [Brandwood, 1983]. Les expressions des gradients obtenues par la suite s'obtiennent à partir des égalités suivantes, où $\boldsymbol{l}$, $\boldsymbol{c}$ et $\mathbf{A}$ désignent respectivement un vecteur ligne, un vecteur colonne et une matrice carrée de la taille de $\boldsymbol{x}$ et $\boldsymbol{l}^*$ désigne le conjugué, $\boldsymbol{l}^{\mathrm{t}}$ le transposé et $\boldsymbol{l}^\dagger$ le transposé conjugué de $\boldsymbol{l}$:

$$\begin{array}{llll}
\dfrac{\partial}{\partial \boldsymbol{x}^*} & \boldsymbol{b}\boldsymbol{x} = \mathbf{0} & \dfrac{\partial}{\partial \boldsymbol{x}^*} & \boldsymbol{b}\boldsymbol{x}^* = \boldsymbol{b}^{\mathrm{t}} \\
\dfrac{\partial}{\partial \boldsymbol{x}^*} & \boldsymbol{x}^\dagger\mathbf{A}\boldsymbol{x} = \mathbf{A}\boldsymbol{x} & \dfrac{\partial}{\partial \boldsymbol{x}^*} & \boldsymbol{x}^\dagger\mathbf{A}\boldsymbol{x}^* = (\mathbf{A}+\mathbf{A}^{\mathrm{t}})\boldsymbol{x}^* \\
\dfrac{\partial}{\partial \boldsymbol{x}^*} & \boldsymbol{x}^\dagger\boldsymbol{c} = \boldsymbol{c} & \dfrac{\partial}{\partial \boldsymbol{x}^*} & \boldsymbol{x}'\boldsymbol{c} = \mathbf{0} \\
\dfrac{\partial}{\partial \boldsymbol{x}^*} & \boldsymbol{x}'\mathbf{A}\boldsymbol{x} = 0 & \dfrac{\partial}{\partial \boldsymbol{x}^*} & \boldsymbol{x}'\mathbf{A}\boldsymbol{x}^* = \mathbf{A}'
\end{array}$$

Le critère pénalisé par la branche d'hyperbole s'écrit :

$$\begin{aligned}
J_{\mathrm{hyp}}(\boldsymbol{x}) &= \frac{1}{2}\left\|\boldsymbol{y}-\mathbf{W}\boldsymbol{x}\right\|^2 + \lambda R_{\mathrm{hyp}}(\boldsymbol{x}), \text{ avec } R_{\mathrm{hyp}}(\boldsymbol{x}) = \sum_k \sqrt{s^2+|x_k|^2} \qquad \text{(A.1)} \\
\text{et } \overrightarrow{\nabla J_{\mathrm{hyp}}}(\boldsymbol{x}) &= \mathbf{W}^\dagger(\boldsymbol{y}-\mathbf{W}\boldsymbol{x}) + \lambda\overrightarrow{\nabla R}_{\mathrm{hyp}}(\boldsymbol{x}), \text{ avec } \overrightarrow{\nabla R}_{\mathrm{hyp}}(\boldsymbol{x})|_k = \frac{x_k}{\sqrt{s^2+|x_k|^2}}.
\end{aligned}$$

On obtient de la même manière pour les critères prenant en compte la symétrie hermitienne :

$$\begin{aligned}
J_+(\boldsymbol{x}_+) &= \frac{1}{2}\left\|\boldsymbol{y}-(\mathbf{W}_+\boldsymbol{x}_+ + \mathbf{W}_+^*\boldsymbol{x}_+^*)\right\|^2 + \lambda R(\boldsymbol{x}_+) \\
\overrightarrow{\nabla J_+}(\boldsymbol{x}_+) &= -2\mathbf{W}_+^\dagger\left(\boldsymbol{y}-\mathbf{W}_+\boldsymbol{x}_+ - \mathbf{W}_+^*\boldsymbol{x}_+^*\right) + \lambda\overrightarrow{\nabla R}_{\mathrm{hyp}}(\boldsymbol{x}_+).
\end{aligned}$$

A.2 Test d'optimalité pour un minimiseur du critère pénalisé par la norme ℓ^1

Cette partie est consacrée à la démonstration de la condition nécessaire et suffisante d'optimalité proposée pour tout minimiseur du critère pénalisé par la norme ℓ^1 :

$$J_1(\boldsymbol{x}) = \frac{1}{2}||\boldsymbol{y} - \mathbf{W}\boldsymbol{x}||^2 + \lambda \sum_k |x_k|.$$

Il s'agit de démontrer l'équivalence suivante :

$$\boldsymbol{x} \text{ est un minimum de } J_1 \text{ si et seulement si } \begin{cases} (i) & \forall k \text{ tq. } |x_k| = 0 : \quad |r_k| \leq \lambda, \\ (ii) & \forall k \text{ tq. } |x_k| \neq 0 : \quad r_k + \lambda \dfrac{x_k}{|x_k|} = 0, \end{cases}$$

où $\boldsymbol{r} = \mathbf{W}^\dagger(\mathbf{W}\boldsymbol{x} - \boldsymbol{y})$.

D'après Alliney et Ruzinsky [1994] on a pour une fonction[1] convexe J :

$$\boldsymbol{x} \text{ est un minimum de } J \text{ si et seulement si } \forall \boldsymbol{p} \in \mathbb{C}^N, J'(\boldsymbol{x}\,;\boldsymbol{p}) \geq 0$$

où $J'(\boldsymbol{x}\,;\boldsymbol{p})$ est le gradient directionnel : $J'(\boldsymbol{x}\,;\boldsymbol{p}) = \lim_{\epsilon \to 0} \frac{1}{\epsilon}(J(\boldsymbol{x} + \epsilon\boldsymbol{p}) - J(\boldsymbol{x}))$. En écrivant $J_1(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\dagger\mathbf{W}^\dagger\mathbf{W}\boldsymbol{x} - \Re\{\boldsymbol{y}^\dagger\mathbf{W}\boldsymbol{x}\} + \frac{1}{2}\boldsymbol{y}^\dagger\boldsymbol{y} + \lambda \|\boldsymbol{x}\|_1$, on a :

$$\begin{aligned} \Delta J_1(\epsilon) &= \frac{1}{\epsilon}\left(J_1(\boldsymbol{x} + \epsilon\boldsymbol{p}) - J_1(\boldsymbol{x})\right) \\ &= \frac{1}{\epsilon}\left(\frac{\epsilon^2}{2}\boldsymbol{p}^\dagger\mathbf{W}^\dagger\mathbf{W}\boldsymbol{p} + \epsilon\,\Re\{\boldsymbol{p}^\dagger\boldsymbol{r}\} + \lambda(\|\boldsymbol{x} + \epsilon\boldsymbol{p}\|_1 - \|\boldsymbol{x}\|_1)\right) \end{aligned}$$

où $\boldsymbol{r} = \mathbf{W}^\dagger(\mathbf{W}\boldsymbol{x} - \boldsymbol{y})$. Le gradient directionnel s'écrit donc :

$$J_1'(\boldsymbol{x}\,;\boldsymbol{p}) = \lim_{\epsilon \to 0} \Delta J_1(\epsilon) = \Re\{\boldsymbol{p}^\dagger\boldsymbol{r}\} + \lambda \lim_{\epsilon \to 0} \frac{1}{\epsilon}(\|\boldsymbol{x} + \epsilon\boldsymbol{p}\|_1 - \|\boldsymbol{x}\|_1).$$

Pour calculer cette dernière limite, on va séparer les indices pour lesquels $\boldsymbol{x}_k$ est nul ou non. Soit $\mathcal{I}$ l'ensemble des indices pour lesquels $\boldsymbol{x}_k$ est nul et $\mathcal{N} - \mathcal{I}$ l'ensemble des indices pour lesquels $\boldsymbol{x}_k$ est non nul. On a alors :

$$\|\boldsymbol{x} + \epsilon\boldsymbol{p}\|_1 - \|\boldsymbol{x}\|_1 = \epsilon \sum_{k \in \mathcal{I}} |\boldsymbol{p}_k| + \sum_{k \in \mathcal{N} - \mathcal{I}} |\boldsymbol{x}_k + \epsilon\boldsymbol{p}_k| - |\boldsymbol{x}_k|.$$

Il suffit donc d'étudier la limite : $\lim_{\epsilon \to 0} \frac{1}{\epsilon}(|x + \epsilon p| - |x|)$ pour $p \in \mathbb{C}$ et $x \in \mathbb{C}^*$; il vient par un développement limité d'ordre 1 lorsque $\epsilon \to 0$: $|x + \epsilon p| = |x| + \epsilon\,\Re\{p^* \frac{x}{|x|}\} + \mathcal{O}(\epsilon^2)$ et donc $\lim_{\epsilon \to 0} \frac{1}{\epsilon}(|x + \epsilon p| - |x|) = \Re\{p^* \frac{x}{|x|}\}$. D'où :

$$\begin{aligned} J_1'(\boldsymbol{x}\,;\boldsymbol{p}) &= \Re\{\boldsymbol{p}^\dagger\boldsymbol{r}\} + \lambda\left(\sum_{k \in \mathcal{I}} |\boldsymbol{p}_k| + \sum_{k \in \mathcal{N} - \mathcal{I}} \Re\{\boldsymbol{p}_k^* \frac{\boldsymbol{x}_k}{|\boldsymbol{x}_k|}\}\right) \\ &= \sum_{k \in \mathcal{I}} (\Re\{\boldsymbol{p}_k^*\boldsymbol{r}_k\} + \lambda|\boldsymbol{p}_k|) + \sum_{k \in \mathcal{N} - \mathcal{I}} \Re\{\boldsymbol{p}_k^*(\boldsymbol{r}_k + \lambda \frac{\boldsymbol{x}_k}{|\boldsymbol{x}_k|})\} \end{aligned}$$

1. Ce résultat est donné par les auteurs pour une fonction de $\mathbb{R}^N$ dans $\mathbb{R}$, on peut l'utiliser pour une fonction de $\mathbb{C}^N$ dans $\mathbb{R}$, en la considérant comme une fonction de $\mathbb{R}^{2N}$ dans $\mathbb{R}$.

La condition $\forall \boldsymbol{p} \in \mathbb{C}^N, J'(\boldsymbol{x}\,;\boldsymbol{p}) \geq 0$ se met alors sous la forme :

$$\begin{aligned} \forall k \in \mathcal{N}-\mathcal{I}, \quad \forall \boldsymbol{p}_k \in \mathbb{C}, \quad \Re\{\boldsymbol{p}_k^*(\boldsymbol{r}_k + \lambda \frac{\boldsymbol{x}_k}{|\boldsymbol{x}_k|})\} \geq 0 \\ \forall k \in \mathcal{I}, \quad \forall \boldsymbol{p}_k \in \mathbb{C}, \quad \Re\{\boldsymbol{p}_k^* \boldsymbol{r}_k\} + \lambda |\boldsymbol{p}_k| \geq 0. \end{aligned}$$

La première condition se ramène de façon évidente à $\forall k \in \mathcal{N}-\mathcal{I}, \boldsymbol{r}_k + \lambda \frac{\boldsymbol{x}_k}{|\boldsymbol{x}_k|} = 0$. En posant $\Re\{\boldsymbol{p}_k^* \boldsymbol{r}_k\} = |p_k r_k| \cos(\phi_k)$, la seconde peut s'écrire $\forall k \in \mathcal{I}, \forall \phi_k \in [0, 2\pi], |r_k| \cos(\phi_k) + \lambda \geq 0$, qui est vérifié si et seulement si $|r_k| < \lambda$. □

A.3 Minimisations scalaires pour les algorithmes de type Gauss-Seidel

La minimisation du critère pénalisé (A.1) en la variable complexe x_k s'écrit sous la forme :

$$\mathring{x_k} = \arg\min_{x_k} J(\boldsymbol{x}) \Leftrightarrow \mathring{x_k} = \arg\min_{x_k} \frac{1}{2} \|\boldsymbol{e}_k - \boldsymbol{w}_k x_k\|^2 + \lambda \varphi(|x_k|)$$

et est équivalente à celle de $\frac{1}{2}\boldsymbol{w}_k^\dagger \boldsymbol{w}_k |x_k|^2 - \Re(\boldsymbol{e}_k^\dagger \boldsymbol{w}_k x_k) + \lambda\varphi(|x_k|)$. En posant $a = \boldsymbol{w}_k^\dagger \boldsymbol{w}_k$, $b = \boldsymbol{w}_k^\dagger \boldsymbol{e}_k$ et $x_k = \rho e^{j\theta}$, le problème est celui de la minimisation de la fonction de deux variables $(\rho,\theta) \in \mathbb{R}^+ \times [0,2\pi]$:

$$f(\rho,\theta) = \frac{1}{2} a\rho^2 - \rho(b_r \cos\theta + b_i \sin\theta) + \lambda\varphi(\rho)$$

où b_r et b_i représentent les parties réelle et imaginaire du nombre complexe b.

La démarche suivante est employée. Nous établissons dans un premier temps que la minimisation du critère J en la variable x_k est équivalente à la résolution des équations (A.2) et (A.3). La vérification simultanée de ces deux équations permet alors d'établir que la minimisation de la fonction f en (ρ,θ) débouche sur $\mathring{\rho} = \arg b \ \forall \rho \in \mathbb{R}^+$, *pour toute fonction de pénalisation* φ *dérivable*. La minimisation en ρ est ensuite détaillée pour les pénalisations hyperbolique et ℓ^1 aux paragraphes A.3.1 et A.3.2, respectivement.

Nous cherchons donc à minimiser la fonction de deux variables $f(\rho,\theta)$. Lorsque $b = 0$, la solution $\rho = 0$ et donc $x_k = 0$ est immédiate. Nous supposons par la suite $b \neq 0$. On a alors :

$$\begin{aligned} \frac{\partial f}{\partial \theta}(\rho,\theta) = 0 \quad &\Leftrightarrow \quad \rho(-b_r \sin\theta + b_i \cos\theta) = 0 \\ &\Leftrightarrow \quad \rho = 0 \ \text{ ou } \ \tan\theta = b_i/b_r \\ &\Leftrightarrow \quad \rho = 0 \ \text{ ou } \ \theta = \theta_1 = \operatorname{atan} b_i/b_r \ \text{ ou } \ \theta = \theta_2 = \operatorname{atan} b_i/b_r + \pi \end{aligned} \qquad \text{(A.2)}$$

où atan est la fonction arc-tangente, pour laquelle nous incluons les cas limites $\theta_1 = \operatorname{atan}(-\infty) = -\pi/2$ et $\theta_2 = \operatorname{atan}(+\infty) = \pi/2$ correspondant aux solutions dans le cas particulier $b_r = 0$. Par ailleurs :

$$\frac{\partial f}{\partial \rho}(\rho,\theta) = 0 \Leftrightarrow a\rho - (b_r \cos\theta + b_i \sin\theta) + \lambda\varphi'(\rho) = 0.$$

Détaillons alors les deux cas suivants :

- Pour $\theta = \theta_1 = \operatorname{atan} \frac{b_i}{b_r}$ et $b_r \neq 0$, on a $\cos\theta = \frac{1}{\sqrt{1+b_i^2/b_r^2}} = \frac{|b_r|}{|b|}$ et $\sin\theta = \frac{b_i/b_r}{\sqrt{1+b_i^2/b_r^2}} = \frac{b_i}{b_r}\frac{|b_r|}{|b|}$ et donc :

$$b_r \cos\theta_1 + b_i \sin\theta_1 = \frac{b_r|b_r|}{|b|} + \frac{b_i^2|b_r|}{b_r|b|} = \operatorname{sgn}(b_r)\left(\frac{b_r^2}{|b|} + \frac{b_i^2}{|b|}\right) = \operatorname{sgn}(b_r)|b|.$$

– De manière similaire, pour $\theta = \theta_2 = \operatorname{atan}\frac{b_i}{b_r} + \pi$ et $b_r \neq 0$, on a $\cos\theta = -\frac{1}{\sqrt{1+b_i^2/b_r^2}} = -\frac{|b_r|}{|b|}$ et $\sin\theta = -\frac{b_i/b_r}{\sqrt{1+b_i^2/b_r^2}} = -\frac{b_i}{b_r}\frac{|b_r|}{|b|}$ et donc :

$$b_r\cos\theta_2 + b_i\sin\theta_2 = -\frac{b_r|b_r|}{|b|} - \frac{b_i^2|b_r|}{b_r|b|} = -\operatorname{sgn}(b_r)\left(\frac{b_r^2}{|b|} + \frac{b_i^2}{|b|}\right) = -\operatorname{sgn}(b_r)|b|.$$

– Dans le cas particulier $b_r = 0$, on a $b = b_i$, $\theta_1 = -\pi/2$ et $\theta_2 = \pi/2$. D'où :
$b_r\cos\theta_1 + b_i\sin\theta_1 = -b_i = -\operatorname{sgn}(b_i)|b|$ et $b_r\cos\theta_2 + b_i\sin\theta_2 = b_i = \operatorname{sgn}(b_i)|b|$.

Pour $\theta_k = \theta_1, \theta_2$, on peut donc écrire $b_r\cos\theta_k + b_i\sin\theta_k = \pm|b|$ et il vient :

$$\frac{\partial f}{\partial\rho}(\rho,\theta_k) = 0 \Leftrightarrow a\rho + \epsilon|b| + \lambda\varphi'(\rho) = 0 \text{ avec } \epsilon = \pm 1.$$

Pour les différentes pénalisations envisagées, la fonction φ est croissante, *i.e.*, $\phi'(\rho) \geq 0$ pour tout $\rho \geq 0$. Par conséquent, dans le cas où $\epsilon = 1$, ce terme ne s'annule jamais (nous avons supposé $b \neq 0$) et $f(\rho,\theta_k)$ est clairement minimal en $\rho = 0$. Nous recherchons donc par la suite les solutions de :

$$\text{pour } \rho \neq 0, \frac{\partial f}{\partial\rho}(\rho,\theta_k) = 0 \Leftrightarrow a\rho - |b| + \lambda\varphi'(\rho) = 0 \tag{A.3}$$

où $\theta_k = \theta_1$ ou θ_2 selon que la partie réelle de b est positive ou négative, respectivement : les cas $(\theta = \theta_1, b_r < 0)$ ou $(\theta = \theta_2, b_r > 0)$ débouchent sur un minimum égal à 0 (ils correspondent au cas où $\epsilon = 1$). Dans le cas particulier $b_r = 0$, on aura $\theta = \theta_1 = -\pi/2$ lorsque $b_i < 0$ et $\theta = \theta_1 = -\pi/2$ lorsque $b_i > 0$. Rappelons que nous supposons ici $b \neq 0$, donc on ne peut pas avoir simultanément $b_r = b_i = 0$.

Nous pouvons donc prouver dès à présent que **la minimisation de f en (θ,ρ) débouche sur $\mathring{\theta} = \arg b$, et ceci $\forall\rho \in \mathbb{R}^+$.** En effet, nous pouvons écrire :

– lorsque $b_r > 0$, on a $\mathring{\theta} = \theta_1 = \operatorname{atan} b_i/b_r$. On a donc $\tan\mathring{\theta} = \tan\arg b$ et les deux angles en question appartiennent à l'intervalle $[-\pi/2, \pi/2]\,(2\pi)$ (θ_1 est une arc-tangente et $\Re(b) > 0$). Par conséquent, $\mathring{\theta} = \arg b$.

– de même, lorsque $b_r < 0$, on a $\mathring{\theta} = \theta_2 = \operatorname{atan} b_i/b_r + \pi$. On a donc $\tan\mathring{\theta} = \tan\arg b$ et les deux angles en question appartiennent à l'intervalle $[\pi/2, 3\pi/2]\,(2\pi)$ (θ_2 est une arc-tangente et $\Re(b) < 0$). Par conséquent, $\mathring{\theta} = \arg b$.

– lorsque $b_r = 0$, on a $\mathring{\theta} = -\pi/2$ si $b_i < 0$: dans ce cas, on a aussi $\arg b = \arg jb_i = -\pi/2$. De même, on a $\mathring{\theta} = \pi/2$ si $b_i > 0$: dans ce cas, on a aussi $\arg b = \arg jb_i = \pi/2$. □

Notons que cette propriété est valable pour toute fonction de pénalisation de la forme $\phi(x_k) = \varphi(|x_k|)$ avec φ croissante et dérivable. Il reste donc maintenant à étudier la minimisation en ρ de $f(\rho,\mathring{\theta})$.

A.3.1 Cas de la pénalisation hyperbolique

Pour la pénalisation hyperbolique, on a $\varphi(\rho) = \sqrt{s^2+\rho^2}$ et l'équation (A.3) s'écrit :

$$\begin{aligned} a\rho - |b| + \lambda\frac{\rho}{\sqrt{s^2+\rho^2}} = 0 \quad &\Leftrightarrow \quad \sqrt{s^2+\rho^2}\,(|b|-a\rho) = \lambda\rho \\ &\Leftrightarrow \quad \rho \le |b|/a \text{ et } (s^2+\rho^2)\,(|b|-a\rho)^2 = \lambda^2\rho^2 \\ &\Leftrightarrow \quad \rho \le |b|/a \text{ et } P(\rho) = 0 \text{ où :} \\ &\qquad P(\rho) = a^2\rho^4 - 2a|b|\rho^3 + \left(|b|^2+s^2-\lambda^2\right)\rho^2 - 2s^2a|b|\rho + s^2|b|^2. \end{aligned}$$

La recherche du minimum de la fonction $f(\rho,\theta)$ peut donc être menée par le schéma suivant :

- si $b = 0$, le minimum est atteint pour $\mathring{\rho} = 0$ et θ quelconque. Dans le cas contraire, prendre $\mathring{\theta} = \arg b$ et procéder aux itérations suivantes :
- rechercher les racines du polynôme P : pour un polynôme de degré quatre, celles-ci peuvent être obtenues de manière explicite [Abramowitz et Stegun, 1964, pages 17–18] ;
- ne conserver que les racines $\mathring{\rho}_j$ vérifiant $0 \le \mathring{\rho}_j \le |b|/a$, puis calculer les valeurs de $f(\mathring{\rho}_j,\mathring{\theta})$. Sélectionner alors le minimum, et le comparer finalement à la valeur de $f(0,\theta) = \lambda s$, car ($\rho = 0, \theta$ quelconque) est également un possible minimum.

A.3.2 Cas de la pénalisation ℓ^1

Pour la pénalisation par la norme ℓ^1, on a $\varphi(\rho) = \rho$ et l'équation (A.3) prend la forme directe :

$$a\rho - |b| + \lambda = 0.$$

Lorsque $|b| \le \lambda$, cette équation n'a pas de solution pour $\rho > 0$; par conséquent, le minimum de f est atteint en $\rho = 0$. Lorsque $|b| > \lambda$, par contre, celle-ci a une unique solution et l'unique solution non nulle de $\left(\frac{\partial f}{\partial \rho}(\rho,\theta) = 0, \frac{\partial f}{\partial \theta}(\rho,\theta) = 0\right)$ est $\mathring{\rho} = \frac{|b|-\lambda}{a}$ et $\mathring{\theta} = \arg b$. Finalement, il convient de noter que, dans le cas $|b| > \lambda$, le couple correspondant $(\mathring{\rho},\mathring{\theta})$ donne lieu à une valeur de f plus faible que $\rho = 0$. La justification est immédiate en notant que $\mathring{\rho}$ est l'unique minimum de la parabole $f(\rho,\mathring{\theta})$: $f(\mathring{\rho},\mathring{\theta}) < f(0,\mathring{\theta})$. Par conséquent, nous avons donc prouvé que :

$$\begin{cases} \text{si } |b| \le \lambda,\ (\mathring{\rho},\mathring{\theta}) \text{ est un minimum de } f \quad \Leftrightarrow \quad \mathring{\rho} = 0\ ; \\ \text{si } |b| > \lambda,\ (\mathring{\rho},\mathring{\theta}) \text{ est un minimum de } f \quad \Leftrightarrow \quad \mathring{\theta} = \arg b \text{ et } \mathring{\rho} = \frac{|b|-\lambda}{a} \end{cases}$$

où la variable $b = \boldsymbol{w}_k^\dagger \boldsymbol{e}_k$ et $a = \boldsymbol{w}_k^\dagger \boldsymbol{w}_k = N$.

Annexe B

Calculs pour la modélisation Bernoulli-Gaussienne

B.1 Loi du module en fonction des lois des parties réelle et imaginaire

Nous notons ici l'expression de la distribution de probabilité et deux deux premiers moments (non centrés) de la variable aléatoire correspondant au module du nombre complexe dont les parties réelle et imaginaires sont supposées gaussiennes, de même variance, mais de moyenne différente. Soient $x \sim \mathcal{N}(\mu_1,\sigma^2)$ et $y \sim \mathcal{N}(\mu_2,\sigma^2)$. D'après Goodman [1985], la variable aléatoire $a = \sqrt{x^2+y^2}$ suit la loi :

$$p_A(a) = \frac{a}{\sigma^2} \exp\left(-\frac{a^2+\mu^2}{2\sigma^2}\right) I_0\left(\frac{a\mu}{\sigma^2}\right) \mathbb{1}_{\mathbb{R}^+}(a)$$

où $\mu = \sqrt{\mu_1^2+\mu_2^2}$ et I_0 est la fonction de Bessel modifiée de première espèce d'ordre 0 [Abramowitz et Stegun, 1964, p. 376]. Remarquons que dans le cas de deux variables x et y centrées ($\mu_1 = \mu_2 = 0$) on a $\mu = 0$, $I_0(0) = 0$ et l'expression ci-dessus correspond à la loi de Rayleigh.

L'expression des moments non centrés est alors donnée par [Goodman, 1985] :

$$\begin{aligned} \mathbb{E}[a] &= \sqrt{\frac{\pi}{2}}\sigma \exp -\frac{\mu^2}{4\sigma^2}\left[\left(1+\frac{\mu^2}{2\sigma^2}\right) I_0\left(\frac{\mu^2}{4\sigma^2}\right) + \frac{\mu^2}{2\sigma^2} I_1\left(\frac{\mu^2}{4\sigma^2}\right)\right] \\ \text{et } \mathbb{E}[a^2] &= 2\sigma^2+\mu^2 \end{aligned}$$

où I_1 est la fonction de Bessel modifiée de première espèce d'ordre 1 [Abramowitz et Stegun, 1964, p.376].

B.2 Calculs relatifs à l'échantillonnage de Gibbs

Nous établissons ici les expressions des lois conditionnelles intervenant dans l'algorithme de Gibbs du tableau (4.2). Rappelons tout d'abord l'expression de la loi jointe *a posteriori* $p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y})$:

$$\begin{aligned} p(\boldsymbol{q},\boldsymbol{s},\boldsymbol{\theta}|\boldsymbol{y}) \propto\ & \lambda^{N_e}(1-\lambda)^{P+1-N_e}\ \frac{1}{\sigma_\epsilon^N}\exp-\frac{1}{2\sigma_\epsilon^2}\left\|\boldsymbol{y}-\mathbf{H}\boldsymbol{s}\right\|^2 \\ & \times\ p(\boldsymbol{s}_0|q_0)\prod_{k\geq 1,q_k=1}\left(\frac{1}{\sigma^2}\exp-\frac{1}{2\sigma^2}\left\|\boldsymbol{s}_k\right\|^2\right)\prod_{k\geq 1,q_k=0}\delta_2(\boldsymbol{s}_k) \\ & \times\ p(\lambda)\ p(\sigma_\epsilon^2)\ p(\sigma^2). \end{aligned} \tag{B.1}$$

où la spécificité du cas $k=0$, que nous avons éludée dans la description de ce manuscrit, conduit à une loi *a priori* légèrement différente puisque $\boldsymbol{s}_0=[a_0,0]$:

$$p(\boldsymbol{s}_0|q_0=1)=g_1(a_0,\sigma^2)\delta(b_0)\text{ et }p(\boldsymbol{s}_0|q_0=0)=\delta_2(\boldsymbol{s}_0).$$

Cette différence aura son importance pour l'établissement de la loi conditionnelle de $\sigma^2|reste$, puisque le facteur de normalisation de la gaussienne g_1 (pour $k=0,q_0=1$) est en $1/\sigma$ alors que celui de la gaussienne g_2 (pour $k\geq 1,q_k=1$) est en $1/\sigma^2$.

B.2.1 Lois conditionnelles des hyperparamètres

Loi de λ

Pour une loi *a priori* $p(\lambda)$ uniforme, l'expression (B.1) permet d'écrire :

$$p(\lambda|reste)\propto\lambda^{N_e}\,(1-\lambda)^{P+1-N_e}$$

où N_e est le nombre de composantes non nulles de la séquence $\boldsymbol{q}$. Nous reconnaissons l'expression d'une loi beta: $\lambda|reste\sim\mathcal{B}e\,(N_e,P+1-N_e)$.

Loi de σ_ϵ^2

Nous considérons trois types de lois *a priori* sur σ_ϵ^2, à savoir :

- une loi uniforme impropre $p(\sigma_\epsilon^2)\propto\mathbb{1}_{[0,+\infty[}(\sigma_\epsilon^2)$,
- une loi *a priori* de Jeffreys, également impropre : $p(\sigma_\epsilon^2)\propto 1/\sigma_\epsilon^2$,
- une loi inverse-gamma $\mathcal{IG}(\alpha,\beta)$: $p(\sigma_\epsilon^2)\propto\frac{1}{\sigma_\epsilon^{2(\alpha+1)}}e^{-\beta/\sigma_\epsilon^2}$.

Les deux premières peuvent également s'exprimer analytiquement sous la forme de lois (impropres) de type inverse-gamma, de paramètres $\alpha=-1,\beta=0$ pour la loi uniforme et $\alpha=0,\beta=0$ pour la loi de Jeffreys. De manière générale, nous tirons alors de (B.1) :

$$p(\sigma_\epsilon^2|reste)\propto\frac{1}{\sigma_\epsilon^{2(N/2+\alpha+1)}}e^{-(\beta+\|\boldsymbol{y}-\mathbf{H}\boldsymbol{s}\|^2/2)/\sigma_\epsilon^2},$$

qui est l'expression d'une loi $\mathcal{IG}(N/2+\alpha,\left\|\boldsymbol{y}-\mathbf{H}\boldsymbol{s}\right\|^2/2+\beta)$.

Loi de σ^2

Nous considérons deux types de lois *a priori* sur σ^2, à savoir :

- une loi *a priori* de Jeffreys : $p(\sigma^2) \propto 1/\sigma^2$,
- une loi inverse-gamma $\mathcal{IG}(\nu,\eta)$: $p(\sigma^2) \propto \frac{1}{\sigma^{2(\mu+1)}} e^{-\eta/\sigma^2}$.

La première peut également s'exprimer analytiquement sous la forme d'une loi de type inverse-gamma, de paramètres $\mu = 0, \eta = 0$. Il vient alors de (B.1) les expressions suivantes :

$$\begin{aligned} p(\sigma^2|reste) &\propto \frac{1}{\sigma^{2(N_e+\mu+1)}} e^{-(\eta+\|\boldsymbol{s}\|^2/2)/\sigma^2} \text{si } q_0 = 0, \\ p(\sigma^2|reste) &\propto \frac{1}{\sigma^{2(N_e+\mu+1/2)}} e^{-(\eta+\|\boldsymbol{s}\|^2/2)/\sigma^2} \text{si } q_0 = 1, \end{aligned}$$

qui sont les expressions d'une loi $\mathcal{IG}(N_e + \mu, \|\boldsymbol{s}\|^2/2 + \eta)$ lorsque $q_0 = 0$ et d'une loi $\mathcal{IG}(N_e - 1/2 + \mu, \|\boldsymbol{s}\|^2/2 + \eta)$ lorsque $q_0 = 1$.

B.2.2 Lois conditionnelles des variables Bernoulli-Gaussiennes

Nous notons $\mathbf{H}_k$ la matrice $N \times 2$ formée par les deux colonnes d'indices $2k$ et $2k+1$ de $\mathbf{H}$, c'est-à-dire la matrice formée par les N lignes $[\cos(2\pi f_k t_n), \sin(2\pi f_k t_n)]_{n=1\ldots N}$, et $\boldsymbol{e}_k = \boldsymbol{y} - \sum_{j\neq k} \mathbf{H}_j \boldsymbol{s}_j$. Nous remplaçons pour mener à bien le calcul suivant la distribution de Dirac $\delta_2(\boldsymbol{s}_k)$ par une loi gaussienne bidimensionnelle de covariance $\sigma_0^2\mathbf{I}_2$ où l'on fera tendre σ_0 vers 0, et notons par analogie σ_1^2 la variance jusqu'ici notée σ^2. La loi conditionnelle de $(q_k,\boldsymbol{s}_k|reste)$ s'écrit alors :

$$p(q_k,\boldsymbol{s}_k|reste) \propto \exp -\frac{\|\boldsymbol{e}_k - \mathbf{H}_k\boldsymbol{s}_k\|^2}{2\sigma_\epsilon^2} \left(\lambda g_2(\boldsymbol{s}_k,\sigma_1^2\mathbf{I}_2) + (1-\lambda) g_2(\boldsymbol{s}_k,\sigma_0^2\mathbf{I}_2)\right)$$

qui s'écrit encore comme une somme de deux distributions gaussiennes :

$$\begin{aligned} p(q_k,\boldsymbol{s}_k|reste) \propto\ & \frac{\lambda}{|\sigma_1^2\mathbf{I}_2|}\, e^{\boldsymbol{v}_{1,k}^{\mathrm{t}}\boldsymbol{\Gamma}_{1,k}^{-1}\boldsymbol{v}_{1,k}/2}\, e^{-(\boldsymbol{s}_k - \boldsymbol{v}_{1,k})^{\mathrm{t}}\boldsymbol{\Gamma}_{1,k}^{-1}(\boldsymbol{s}_k - \boldsymbol{v}_{1,k})/2} \\ & + \frac{1-\lambda}{|\sigma_0^2\mathbf{I}_2|}\, e^{\boldsymbol{v}_{0,k}^{\mathrm{t}}\boldsymbol{\Gamma}_{0,k}^{-1}\boldsymbol{v}_{0,k}/2}\, e^{-(\boldsymbol{s}_k - \boldsymbol{v}_{0,k})^{\mathrm{t}}\boldsymbol{\Gamma}_{0,k}^{-1}(\boldsymbol{s}_k - \boldsymbol{v}_{0,k})/2} \end{aligned}$$

soit sous forme condensée :

$$p(q_k,\boldsymbol{s}_k|reste) = \lambda_{1,k}\, g_2(\boldsymbol{s}_k - \boldsymbol{v}_{1,k},\boldsymbol{\Gamma}_{1,k}) + \lambda_{0,k}\, g_2(\boldsymbol{s}_k - \boldsymbol{v}_{0,k},\boldsymbol{\Gamma}_{0,k})$$

où $\boldsymbol{\Gamma}_{i,k}^{-1} = \frac{1}{\sigma_\epsilon^2}\mathbf{H}_k^{\mathrm{t}}\mathbf{H}_k + \frac{1}{\sigma_i^2}\mathbf{I}_2$, $\boldsymbol{v}_{i,k} = \frac{1}{\sigma_\epsilon^2}\boldsymbol{\Gamma}_{i,k}\mathbf{H}_k^{\mathrm{t}}\boldsymbol{e}_k$ et les coefficients de normalisation $\lambda_{i,k}$ sont tels que $\lambda_{1,k} + \lambda_{0,k} = 1$, soit :

$$\lambda_{i,k} = \frac{\widetilde{\lambda}_{i,k}}{\widetilde{\lambda}_{1,k} + \widetilde{\lambda}_{0,k}} \text{ avec } \widetilde{\lambda}_{1,k} = \frac{\lambda}{\sigma_1^2}|\boldsymbol{\Gamma}_{1,k}|^{1/2} e^{\boldsymbol{v}_{1,k}^{\mathrm{t}}\boldsymbol{\Gamma}_{1,k}^{-1}\boldsymbol{v}_{1,k}/2} \text{ et } \widetilde{\lambda}_{0,k} = \frac{1-\lambda}{\sigma_0^2}|\boldsymbol{\Gamma}_{0,k}|^{1/2} e^{\boldsymbol{v}_{0,k}^{\mathrm{t}}\boldsymbol{\Gamma}_{0,k}^{-1}\boldsymbol{v}_{0,k}/2}.$$

Le passage à la limite $\sigma_0 \to 0$ donne alors :

$$\begin{aligned} \boldsymbol{\Gamma}_{0,k} &\to \mathbf{0}_{2\times 2}, \text{ la matrice } 2\times 2 \text{ identiquement nulle}, \\ \boldsymbol{v}_{0,k} &\to \mathbf{0}_2, \text{ le vecteur } 2\times 1 \text{ nul}, \\ \widetilde{\lambda}_{0,k} &\to 1-\lambda. \end{aligned}$$

Par conséquent, $(q_k,\boldsymbol{s}_k|reste)$ est une variable Bernoulli-Gaussienne telle que :

- $\Pr(q_k = 1|reste) = \lambda_{1,k}$
- $p(\boldsymbol{s}_k|q_k = 1,reste) = g_2(\boldsymbol{s}_k - \boldsymbol{v}_{1,k},\boldsymbol{\Gamma}_{1,k})$ et $p(\boldsymbol{s}_k|q_k = 0,reste) = \delta_2(\boldsymbol{s}_k)$.

B.2.3 Lois conditionnelles des variables Bernoulli-Gaussiennes par blocs de trois composantes

Nous désignons dans ce paragraphe par $(\boldsymbol{q}_{k,3}, \boldsymbol{s}_{k,3})$ les blocs de trois variables B-G correspondant aux indices k, $k+1$ et $k+2$. (Nous les avons notés $q_{k\to k+2}$,$\boldsymbol{s}_{k\to k+2}$ dans le corps de ce manuscrit, mais une notation plus compacte permet ici de réduire la longueur des expressions suivantes.)

Nous notons $\mathbf{H}_{k,3}$ la matrice $N \times 6$ formée par la concaténation des trois matrices $\mathbf{H}_k$, $\mathbf{H}_{k+1}$ et $\mathbf{H}_{k+2}$, c'est-à-dire des colonnes d'indices $2k$ à $2k+5$ de la matrice $\mathbf{H}$. Nous notons également $\boldsymbol{e}_{k,3} = \boldsymbol{y} - \sum_{j\neq k,k+1,k+2} \mathbf{H}_j \boldsymbol{s}_j$. Comme au paragraphe précédent, nous remplaçons la distribution de Dirac $\delta_2(\boldsymbol{s}_k)$ par une loi gaussienne bidimensionnelle de covariance $\sigma_0^2\mathbf{I}_2$ où l'on fera tendre σ_0 vers 0, et notons par analogie σ_1^2 la variance σ^2. La loi conditionnelle de $(\boldsymbol{q}_{k,3},\boldsymbol{s}_{k,3}|reste)$ s'écrit :

$$p(\boldsymbol{q}_{k,3},\boldsymbol{s}_{k,3}|reste) \propto \exp\left(-\frac{\|\boldsymbol{e}_{k,3} - \mathbf{H}_{k,3}\,\boldsymbol{s}_{k,3}\|^2}{2\sigma_\epsilon^2}\right) \underbrace{\prod_{j=k}^{k+2} \left(\lambda g_2(\boldsymbol{s}_j,\sigma_1^2\mathbf{I}_2) + (1-\lambda) g_2(\boldsymbol{s}_j,\sigma_0^2\mathbf{I}_2)\right)}_{p(\boldsymbol{q}_{k,3},\boldsymbol{s}_{k,3})}$$

où le terme *a priori* $p(\boldsymbol{q}_{k,3},\boldsymbol{s}_{k,3})$ se décompose en la somme de huit termes correspondant chacun à une configuration de $\boldsymbol{q}_{k,3} \in \{0;1\}^3$. De manière analogue à la démarche réalisée au paragraphe précédent, la loi *a posteriori* s'écrit comme un mélange de huit gaussiennes :

$$p(\boldsymbol{q}_{k,3},\boldsymbol{s}_{k,3}|reste) = \sum_{(u,v,w)\in\{0;1\}^3} \lambda_{u,v,w}\; g_6(\boldsymbol{s}_{k,3} - \boldsymbol{v}_{u,v,w}, \boldsymbol{\Gamma}_{u,v,w}) \tag{B.2}$$

où $\lambda_{u,v,w}$ est la probabilité *a posteriori*, connaissant les données et le reste des paramètres B-G, que les variables de Bernoulli $\boldsymbol{q}_{k,3}$ soient dans la configuration (u,v,w) et $g_6(\boldsymbol{s}_{k,3} - \boldsymbol{v}_{u,v,w}, \boldsymbol{\Gamma}_{u,v,w})$ est la distribution *a posteriori* gaussienne correspondante des amplitudes. Par souci de clarté, nous omettons ici la dépendance en k des quantités $\lambda_{u,v,w}$, $\boldsymbol{v}_{u,v,w}$ et $\boldsymbol{\Gamma}_{u,v,w}$. Par un raisonnement similaire à celui effectué dans le cas d'une seule variable B-G au paragraphe précédent, on montre que :

– les matrices de covariance *a posteriori* sont les inverses de matrices de la forme :

$$\boldsymbol{\Gamma}_{u,v,w}^{-1} = \frac{1}{\sigma_\epsilon^2}\mathbf{H}_{k,3}^{\mathrm{t}}\mathbf{H}_{k,3} + \mathbf{I}_{u,v,w}(\sigma_1,\sigma_0)$$

où $\mathbf{I}_{u,v,w}(\sigma_1,\sigma_0)$ est une matrice 6×6 diagonale formée par trois blocs de type $\frac{1}{\sigma_1^2}\mathbf{I}_2$ lorsque l'indicateur correspondant (u, v ou w) vaut 1 et $\frac{1}{\sigma_0^2}\mathbf{I}_2$ lorsque celui-ci vaut 0 ;

– les moyennes *a posteriori* s'écrivent :

$$\boldsymbol{v}_{u,v,w} = \frac{1}{\sigma_\epsilon^2}\boldsymbol{\Gamma}_{u,v,w}\mathbf{H}_{k,3}^{\mathrm{t}}\boldsymbol{e}_{k,3}\ ;$$

– les probabilités *a posteriori* prennent la forme :

$$\lambda_{u,v,w} = \frac{\widetilde{\lambda}_{u,v,w}}{S_{\widetilde{\lambda}}} \text{ avec } \widetilde{\lambda}_{u,v,w} = \frac{\lambda^{N_{u,v,w}}(1-\lambda)^{(3-N_{u,v,w})}}{\sigma_1^{2N_{u,v,w}}\sigma_0^{2(3-N_{u,v,w})}}|\mathbf{\Gamma}_{u,v,w}|^{1/2}e^{\boldsymbol{v}_{u,v,w}^{\mathrm{t}}\mathbf{\Gamma}_{u,v,w}^{-1}\boldsymbol{v}_{u,v,w}/2}$$

où $N_{u,v,w}$ est le nombre de 1 dans la séquence $\boldsymbol{q}_{k,3}$ ($N_{u,v,w} = u+v+w$) et $S_{\widetilde{\lambda}}$ est la somme des probabilités $\widetilde{\lambda}_{u,v,w}$.

Ainsi pour $\boldsymbol{q}_{k,3} = [1,1,1]$, on a :

$$([\boldsymbol{s}_k,\boldsymbol{s}_{k+1},\boldsymbol{s}_{k+2}]|\boldsymbol{q}_{k,3} = [1,1,1],\mathit{reste}) \sim \mathcal{N}_6(\boldsymbol{v}_{111},\mathbf{\Gamma}_{111}) \text{ avec}$$

$$\left\{ \begin{array}{rcl} \mathbf{\Gamma}_{111}^{-1} & = & \frac{1}{\sigma_\epsilon^2}\mathbf{H}_{k,3}^{\mathrm{t}}\mathbf{H}_{k,3} + \frac{1}{\sigma^2}\mathbf{I}_6, \\ \boldsymbol{v}_{111} & = & \frac{1}{\sigma_\epsilon^2}\mathbf{\Gamma}_{111}\mathbf{H}_{k,3}^{\mathrm{t}}\boldsymbol{e}_{k,3}, \end{array} \right.$$

et la probabilité *a posteriori* associée à cette configuration s'écrit :

$$\lambda_{111} = \frac{\widetilde{\lambda}_{111}}{S_{\widetilde{\lambda}}}, \text{ avec } \widetilde{\lambda}_{111} = \frac{\lambda^3}{\sigma_1^6}|\mathbf{\Gamma}_{111}|^{1/2}e^{\boldsymbol{v}_{111}^{\mathrm{t}}\mathbf{\Gamma}_{111}^{-1}\boldsymbol{v}_{111}/2}.$$

$$\widetilde{\lambda}_{111} = \frac{\lambda^3}{\sigma_1^6}|\mathbf{\Gamma}_{111}|^{1/2}e^{\boldsymbol{v}_{111}^{\mathrm{t}}\mathbf{\Gamma}_{111}^{-1}\boldsymbol{v}_{111}/2}.$$

Dans les sept autres configurations, le passage à la limite $\sigma_0 \to 0$ débouche sur des matrices de covariance comportant des blocs de zéros et des moyennes correspondantes nulles, c'est-à-dire sur des triplets de variables de type B-G. Plus précisément, nous détaillons les calculs dans trois cas, selon que $N_{u,v,w} = 2$, 1 ou 0. Nous utiliserons pour cela le **lemme d'inversion des matrices partitionnées** qui s'écrit, en supposant les matrices A_{ij} carrées et inversibles :

$$\begin{pmatrix} A_{12} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \nabla^{-1} & -\nabla^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}\nabla^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}\nabla^{-1}A_{12}A_{22}^{-1} \end{pmatrix}$$

$$\text{avec } \nabla = A_{11} - A_{12}A_{21}A_{22}^{-1}A_{21}. \tag{B.3}$$

Cas $\mathbf{N_{u,v,w} = 2}$: par exemple, (u,v,w) = (0,1,1)

Dans ce cas, la matrice $\mathbf{\Gamma}_{011}^{-1} = \frac{1}{\sigma_\epsilon^2}\mathbf{H}_{k,3}^{\mathrm{t}}\mathbf{H}_{k,3} + \begin{pmatrix} \frac{1}{\sigma_0^2}\mathbf{I}_2 & \mathbf{0}_{2\times4} \\ \mathbf{0}_{4\times2} & \frac{1}{\sigma_1^2}\mathbf{I}_4 \end{pmatrix}$ s'écrit sous la forme $\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, où A_{11} est 2×2, A_{12} est 2×4, A_{21} est 4×2 et A_{22} est 4×4. En particulier :

$$A_{22} = \frac{1}{\sigma_\epsilon^2}\mathbf{H}_{k+1,2}^{\mathrm{t}}\mathbf{H}_{k+1,2} + \frac{1}{\sigma_1^2}\mathbf{I}_4$$

où $\mathbf{H}_{k+1,2}$ est la matrice formée par la concaténation des deux matrices $\mathbf{H}_{k+1}$ et $\mathbf{H}_{k+2}$. Avec les notations du lemme (B.3), on a alors : $\nabla = \frac{1}{\sigma_0^2}\mathbf{I}_2 + \mathbf{K}_{2\times2}$, où $\mathbf{K}_{2\times2}$ ne dépend pas de σ_0. Par conséquent, lorsque σ_0 tend vers 0, on a $\nabla^{-1} \to \mathbf{0}_{2\times2}$ et, en vertu de ce lemme d'inversion :

$$\mathbf{\Gamma}_{011} \longrightarrow \begin{pmatrix} \mathbf{0}_{2\times2} & \mathbf{0}_{2\times4} \\ \mathbf{0}_{4\times2} & A_{22}^{-1} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{2\times2} & \mathbf{0}_{2\times4} \\ \mathbf{0}_{4\times2} & \left(\frac{1}{\sigma_\epsilon^2}\mathbf{H}_{k+1,2}^{\mathrm{t}}\mathbf{H}_{k+1,2} + \frac{1}{\sigma_1^2}\mathbf{I}_4\right)^{-1} \end{pmatrix}.$$

On déduit alors facilement que :

$$\boldsymbol{v}_{011} \longrightarrow \frac{1}{\sigma_\epsilon^2}\begin{pmatrix} \mathbf{0}_{2\times 1} \\ \left(\frac{1}{\sigma_\epsilon^2}\mathbf{H}_{k+1,2}^{\mathrm{t}}\mathbf{H}_{k+1,2} + \frac{1}{\sigma_1^2}\mathbf{I}_4\right)^{-1}\mathbf{H}_{k+1,2}^{\mathrm{t}}\boldsymbol{e}_{k,3} \end{pmatrix}.$$

et pour cette configuration, $\boldsymbol{s}_k = \mathbf{0}$ et $[\boldsymbol{s}_{k+1},\boldsymbol{s}_{k+2}]$ est une variable gaussienne (de dimension 4) dont la covariance et la moyenne sont données par les blocs correspondants des deux expressions précédentes :

$$\boxed{\begin{array}{c}\left([\boldsymbol{s}_{k+1},\boldsymbol{s}_{k+2}]|\boldsymbol{q}_{k,3} = [0,1,1],\mathit{reste}\right) \sim \mathcal{N}_4(\boldsymbol{v}_{k+1,2},\boldsymbol{\Gamma}_{k+1,2}) \text{ avec} \\ \left\{\begin{array}{rcl} \boldsymbol{\Gamma}_{k+1,2} & = & \left(\frac{1}{\sigma_\epsilon^2}\mathbf{H}_{k+1,2}^{\mathrm{t}}\mathbf{H}_{k+1,2} + \frac{1}{\sigma_1^2}\mathbf{I}_4\right)^{-1} \\ \boldsymbol{v}_{k+1,2} & = & \frac{1}{\sigma_\epsilon^2}\boldsymbol{\Gamma}_{k+1,2}\mathbf{H}_{k+1,2}^{\mathrm{t}}\boldsymbol{e}_{k,3} \end{array}\right. \end{array}}$$

et la probabilité *a posteriori* associée à une telle configuration s'écrit :

$$\lambda_{011} = \frac{\widetilde{\lambda}_{011}}{S_{\widetilde{\lambda}}}, \text{ avec } \widetilde{\lambda}_{011} = \lambda^2(1-\lambda)\frac{|\boldsymbol{\Gamma}_{011}|^{1/2}}{\sigma_1^4\sigma_0^2}e^{\boldsymbol{v}_{011}^\dagger\boldsymbol{\Gamma}_{011}^{-1}\boldsymbol{v}_{011}/2}.$$

On a alors, d'une part, $\frac{|\boldsymbol{\Gamma}_{011}|^{1/2}}{\sigma_0^2} = \left|\sigma_0^{2/3}\boldsymbol{\Gamma}_{011}^{-1}\right|^{-1/2}$, avec :

$$\sigma_0^{2/3}\boldsymbol{\Gamma}_{011}^{-1} = \frac{\sigma_0^{2/3}}{\sigma_\epsilon^2}\mathbf{H}_{k,3}^\dagger\mathbf{H}_{k,3} + \begin{pmatrix} \frac{\sigma_0^{2/3}}{\sigma_0^2}\mathbf{I}_2 & \mathbf{0}_{2\times 4} \\ \mathbf{0}_{4\times 2} & \frac{\sigma_0^{2/3}}{\sigma_1^2}\mathbf{I}_4 \end{pmatrix}$$

et donc lorsque $\sigma_0 \to 0$,

$$\left|\sigma_0^{2/3}\boldsymbol{\Gamma}_{011}^{-1}\right|^{-1/2} \sim \left|\begin{array}{cc} \frac{\sigma_0^{2/3}}{\sigma_0^2}\mathbf{I}_2 & \mathbf{0}_{2\times 4} \\ \mathbf{0}_{4\times 2} & \frac{\sigma_0^{2/3}}{\sigma_1^2}\mathbf{I}_4 \end{array}\right|^{-1/2} = \left(\left(\sigma_0^{-4/3}\right)^2\left(\frac{\sigma_0^{2/3}}{\sigma_1^2}\right)^4\right)^{-1/2} = \sigma_1^4.$$

Par ailleurs, il est facile de voir que :

$$\boldsymbol{v}_{011}^\dagger\boldsymbol{\Gamma}_{011}^{-1}\boldsymbol{v}_{011} \to \frac{1}{\sigma_\epsilon^4}\boldsymbol{e}_{k,3}^\dagger\mathbf{H}_{k+1,2}\boldsymbol{\Gamma}_{k+1,2}\mathbf{H}_{k+1,2}^\dagger\boldsymbol{e}_{k,3}$$

d'où finalement $\boxed{\widetilde{\lambda}_{011} \longrightarrow \lambda^2(1-\lambda)\exp\left(\frac{1}{2\sigma_\epsilon^4}\boldsymbol{e}_{k,3}^\dagger\mathbf{H}_{k+1,2}\boldsymbol{\Gamma}_{k+1,2}\mathbf{H}_{k+1,2}^\dagger\boldsymbol{e}_{k,3}\right).}$

Les configurations $(u,v,w) = (1,0,1)$ et $(u,v,w) = (1,1,0)$ donnent lieu à des expressions similaires en ne considérant que les colonnes de $\mathbf{H}$ correspondant aux « 1 » de la séquence de Bernoulli $\boldsymbol{q}_{k,3}$.

Cas $\mathbf{N_{u,v,w}} = \mathbf{1}$: par exemple, (u,v,w) = (0,0,1)

Par une démarche analogue à celle effectuée pour $(u,v,w) = (0,1,1)$, on montre que :

$$\begin{aligned} \boldsymbol{\Gamma}_{001} &\longrightarrow \begin{pmatrix} \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 2} \\ \mathbf{0}_{2\times 4} & \left(\frac{1}{\sigma_\epsilon^2}\mathbf{H}_{k+2}^{\mathrm{t}}\mathbf{H}_{k+2} + \frac{1}{\sigma_1^2}\mathbf{I}_2\right)^{-1} \end{pmatrix}. \\ \boldsymbol{v}_{001} &\longrightarrow \frac{1}{\sigma_\epsilon^2}\begin{pmatrix} \mathbf{0}_{4\times 1} \\ \left(\frac{1}{\sigma_\epsilon^2}\mathbf{H}_{k+2}^{\mathrm{t}}\mathbf{H}_{k+2} + \frac{1}{\sigma_1^2}\mathbf{I}_2\right)^{-1}\mathbf{H}_{k,2}^{\mathrm{t}}\boldsymbol{e}_{k,3} \end{pmatrix}. \end{aligned}$$

Pour cette configuration $\boldsymbol{s}_k = \boldsymbol{s}_{k+1} = \mathbf{0}$ et $\boldsymbol{s}_{k+2}$ est une variable gaussienne (de dimension 2) dont la covariance et la moyenne sont données par les blocs correspondants des deux expressions précédentes, à savoir respectivement :

$$\boxed{\begin{array}{c}(\boldsymbol{s}_{k+2}]|\boldsymbol{q}_{k,3} = [0,0,1], reste) \sim \mathcal{N}_2(\boldsymbol{v}_{k+2}, \boldsymbol{\Gamma}_{k+2}) \text{ avec} \\ \left\{ \begin{array}{rcl} \boldsymbol{\Gamma}_{k+2} & = & \left(\frac{1}{\sigma_\epsilon^2}\mathbf{H}_{k+2}^{\mathrm{t}}\mathbf{H}_{k+2} + \frac{1}{\sigma_1^2}\mathbf{I}_2\right)^{-1} \\ \boldsymbol{v}_{k+2} & = & \frac{1}{\sigma_\epsilon^2}\boldsymbol{\Gamma}_{k+2}\mathbf{H}_{k+2}^{\mathrm{t}}\boldsymbol{e}_{k,3} \end{array} \right. \end{array}}$$

et la probabilité *a posteriori* associée à une telle configuration s'écrit :

$$\lambda_{001} = \frac{\widetilde{\lambda}_{001}}{S_{\widetilde{\lambda}}}, \text{ avec } \widetilde{\lambda}_{001} = \lambda(1-\lambda)^2 \frac{|\boldsymbol{\Gamma}_{001}|^{1/2}}{\sigma_1^4\sigma_0^2} e^{\boldsymbol{v}_{001}^\dagger \boldsymbol{\Gamma}_{001}^{-1} \boldsymbol{v}_{001}/2}.$$

et on montre que $\boxed{\widetilde{\lambda}_{001} \longrightarrow \lambda(1-\lambda)^2 \exp\left(\frac{1}{2\sigma_\epsilon^4}\boldsymbol{e}_{k,3}^\dagger \mathbf{H}_{k+2}\boldsymbol{\Gamma}_{k+2}\mathbf{H}_{k+2}^\dagger \boldsymbol{e}_{k,3}\right).}$

Les configurations $(u,v,w) = (0,1,0)$ et $(u,v,w) = (1,0,0)$ donnent lieu à des expressions similaires en ne considérant que la colonne de $\mathbf{H}$ correspondant au « 1 » de la séquence de Bernoulli $\boldsymbol{q}_{k,3}$, c'est-à-dire respectivement $\mathbf{H}_{k+1}$ et $\mathbf{H}_k$.

Cas $\mathbf{N_{u,v,w} = 0}$: (u,v,w) = (0,0,0)

Enfin, le cas $u = v = w = 0$ correspond bien évidemment à $\Gamma_{000} = \mathbf{0}_{6\times 6}$ et $\boldsymbol{v}_{000} = \mathbf{0}_{3\times 1}$, et la probabilité associée vaut $\lambda_{000} = \frac{\widetilde{\lambda}_{000}}{S_{\widetilde{\lambda}}}$, avec $\widetilde{\lambda}_{000} \longrightarrow (1-\lambda)^3$.

Bibliographie

M. Abramowitz et I.A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, 1964.

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, décembre 1974.

S. Alliney et S. A. Ruzinsky. An algorithm for the minimization of mixed ℓ_1 and ℓ_2 norms with application to Bayesian estimation. *IEEE Transactions on Signal Processing*, 42(3):618–627, mars 1994.

C. Andrieu et A. Doucet. Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC. *IEEE Transactions on Signal Processing*, 47(10):2667–2676, octobre 1999.

J.M. Bernardo et A.F.M. Smith. *Bayesian Theory*. Wiley series in probability and statistics, New York, 1994.

D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 1995.

T. Böhm, C. Catala, L. Balona, et B. Carter. Spectroscopic monitoring of the Herbig Ae star HD 104237. I. Multiperiodic stellar oscillations. *Astronomy and Astrophysics*, 427:907–922, décembre 2004.

C. Bouman et K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Transactions on Image Processing*, 2(3):296–310, juillet 1993.

S. Bourguignon. Analyse spectrale à haute résolution de signaux à données manquantes : application à l'Astrophysique. Rapport de stage de DEA, Laboratoire d'Astrophysique de l'Observatoire Midi-Pyrénées, 2002.

S. Bourguignon, H. Carfantan, et L. Jahan. Regularized spectral analysis of unevenly spaced data. Dans *Proceedings of ICASSP*, volume 4, pages 421–424, Philadelphia, USA, mars 2005.

D.H. Brandwood. A complex gradient operator and its application in adaptive array theory. *IEE Proceedings*, 130(1):11–16, février 1983.

M. Breger, J. Stich, R. Garrido, B. Martin, S. Y. Jiang, Z. P. Li, D. P. Hube, W. Ostermann, M. Paparo, et M. Scheck. Nonradial pulsation of the Delta-Scuti star Bu-Cancri in the Praesepe cluster. *Astronomy and Astrophysics*, 271:482–486, avril 1993.

S. Brette et J. Idier. Optimized single site update algorithms for image deblurring. Dans *Proceedings of ICIP*, volume 3, pages 65–68, Lausanne, Switzerland, septembre 1996.

G. L. Bretthorst. *Bayesian Spectrum Analysis and Parameter Estimation*. Lecture Notes in Statistics, 48. Springer-Verlag, New York, 1988.

G. L. Bretthorst. Nonuniform Sampling: Bandwidth and Aliasing. Dans *Maximum Entropy and Bayesian Methods in Science and Engineering*, pages 1–28. Joshua Rychert, Gary Erickson and C. Ray Smith, American Institute of Physics, USA, 2001.

K.P. Bube et R.T. Langan. Hybrid ℓ^1/ℓ^2 minimization with applications to tomography. *Geophysics*, 62(4):1183–1195, 1997.

R.H. Byrd et D.A. Payne. Convergence of the iterative reweighted least squares algorithm for robust regression. Rapport Technique 313, The John Hopkins University, Baltimore, MD, USA, juin 1979.

S.D. Cabrera et T.W. Parks. Extrapolation and spectral estimation with iterative weighted norm modification. *IEEE Transactions on Signal Processing*, 39(4):842–851, avril 1991.

O. Cappé, A. Doucet, M. Lavielle, et É. Moulines. Simulation-based methods for blind maximum-likelihood filter identification. *Signal Processing*, 73(1–2):3–25, janvier 1999.

J.-M. Carquillat, N. Ginestet, J.-L. Prieur, et S. Udry. Contribution to the search for binaries among Am stars - III. HD 7119: a double-lined spectroscopic binary and a triple system. *Monthly Notices of the Royal Astronomical Society*, 336:1043–1048, novembre 2002.

F. Champagnat. *Déconvolution impulsionnelle et extensions pour la caractérisation des milieux inhomogènes en échographie*. PhD thesis, Université de Paris-Sud, décembre 1993.

F. Champagnat, Y. Goussard, et J. Idier. Unsupervised deconvolution of sparse spike trains using stochastic approximation. *IEEE Transactions on Signal Processing*, 44(12):2988–2998, décembre 1996.

F. Champagnat et J. Idier. An alternative to standard maximum likelihood for Gaussian mixtures. Dans *Proceedings of ICASSP*, pages 2020–2023, Detroit, USA, mai 1995.

S.S. Chen et D.L. Donoho. Application of basis pursuit in spectrum estimation. Dans *Proceedings of ICASSP*, volume 3, pages 1865–1868, Seattle, USA, mai 1998.

S.S. Chen, D.L. Donoho, et M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.

Q. Cheng, R. Chen, et T.-H. Li. Simultaneous wavelet estimation and deconvolution of reflection seismic signals. *IEEE Transactions on Geoscience and Remote Sensing*, 34(2):377–384, mars 1996.

C.-Y. Chi et J.M Mendel. Improved maximum-likelihood detection and estimation of Bernoulli-Gaussian processes. *IEEE Transactions on Information Theory*, 30(2):429–435, mars 1984.

P. Ciuciu et J. Idier. A half-quadratic block-coordinate descent method for spectral estimation. *Signal Processing*, 82:941 – 959, juillet 2002.

P. Ciuciu, J. Idier, et J.-F. Giovannelli. Regularized estimation of mixed spectra using a circular Gibbs-Markov model. *IEEE Transactions on Signal Processing*, 49(10):2201–2213, octobre 2001.

M. Davies et L. Daudet. Sparsifying subband decompositions. Dans *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, octobre 2003.

M. Davy et J. Idier. Fast MCMC computations for the estimation of sparse processes from noisy observations. Dans *Proceedings of ICASSP*, volume II, pages 1041–1044, Montréal, QC, Canada, mai 2004.

G. Demoment. Image reconstruction and restoration: overview of common estimation structures and problems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12):2024–2036, décembre 1989.

P.M. Djurić. A model selection rule for sinusoids in white gaussian noise. *IEEE Transactions on Signal Processing*, 44:1744–1751, juillet 1996.

D.L. Donoho et B.F. Logan. Signal recovery and the large sieve. *SIAM Journal of Applied Mathematics*, 52:577–591, 1992.

A. DOUCET et P. DUVAUT. Bayesian estimation of state-space models applied to deconvolution of Bernoulli-Gaussian processes. *Signal Processing*, 57(2):147–161, mars 1997.

F. DUBLANCHET. *Contribution de la méthodologie bayésienne à l'analyse spectrale de raies pures et à la goniométrie haute résolution.* PhD thesis, Université de Paris-Sud, octobre 1996.

P. DUVAUT et F. DUBLANCHET. Une méthode d'analyse de sinusoïdes complexes par déconvolution du périodogramme. *Traitement du Signal*, 12(3):239–253, 1995.

M. M. DWORETSKY. A period-finding method for sparse randomly spaced observations or 'How long is a piece of string?'. *Monthly Notices of the Royal Astronomical Society*, 203:917–924, juin 1983.

L. EYER et P. BARTHOLDI. Variable stars: which Nyquist frequency? *Astronomy and Astrophysics Supplement Series*, 135:1–3, février 1999.

L. FATHI, G. JOURDAIN, et M. ARNDT. A Bernoulli-Gaussian model for multipath channel estimation in downlink WCDMA. Dans *IEEE 13th workshop on Statistical Signal Processing*, Papier #204, Bordeaux, France, juillet 2005.

S. FERRAZ-MELLO. Estimation of periods from unequally spaced observations. *The Astronomical Journal*, 86:619–624, avril 1981.

M. FIGUEIREDO. Adaptive sparseness for supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1150–1159, septembre 2003.

G. FOSTER. The Cleanest Fourier spectrum. *The Astronomical Journal*, 109(4):1889–1902, avril 1995.

G. FOSTER. Time series analysis by projection. I. Statistical properties of Fourier analysis. *The Astronomical Journal*, 111:541–554, janvier 1996.

J.-J. FUCHS. Une approche à l'estimation et à l'identification simultanées. Dans *Actes du seizième colloque GRETSI*, pages 1273–1276, Grenoble, France, 1997.

J.-J. FUCHS. Multipath time-delay detection and estimation. *IEEE Transactions on Signal Processing*, 47:237–243, janvier 1999.

J.-J. FUCHS. On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50:1341–1344, juin 2004.

J.-J. FUCHS. Communication personnelle. 2005.

E. GASSIAT, F. MONFRONT, et Y. GOUSSARD. On simultaneous signal estimation and parameter identification using a generalized likelihood approach. *IEEE Transactions on Information Theory*, 38(1):157–162, 1992.

J.-F. GIOVANNELLI et J. IDIER. Bayesian interpretation of periodograms. *IEEE Transactions on Signal Processing*, 49(7):1988–1996, juillet 2001.

S. GODSILL et M. DAVY. Bayesian harmonic models for musical pitch estimation and analysis. Dans *Proceedings of ICASSP'02*, Orlando, USA, mai 2002.

J. W. GOODMAN. *Statistical Optics.* Pure and Applied Optics. Wiley, New York, 1985.

Y. GOUSSARD. *Déconvolution de processus aléatoires non-gaussiens par maximisation de vraisemblances.* PhD thesis, Université de Paris-Sud, septembre 1989.

Y. GOUSSARD, G. DEMOMENT, et J. IDIER. A new algorithm for iterative deconvolution of sparse spike trains. Dans *Proceedings of ICASSP*, pages 1547–1550, Albuquerque, NM, USA, avril 1990.

J. GOUTSIAS et J.M. MENDEL. Maximum likelihood deconvolution: an optimization theory perspective. *Geophysics*, 51(6):1206–1220, juin 1986.

K. N. GRANKIN. Evolution of the light curve of V410 Tau during 1986-1997. *Astronomy Letters*, 25:526–535, août 1999.

R. GRAY. Toeplitz and circulant matrices: A review. Rapport Technique 6504-1, Information Systems Laboratory, Stanford University, Stanford, CA, USA, 1977.

J. A. HÖGBOM. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement Series*, 15:417–426, juin 1974.

P.C. HANSEN. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34:561–580, 1992.

J. IDIER, éditeur. *Approche bayésienne pour les problèmes inverses.* Traité IC2, Série traitement du signal et de l'image, Hermès, Paris, novembre 2001.

J. IDIER. Convex half-quadratic criteria and interacting auxiliary variables for image restoration. *IEEE Transactions on Image Processing*, 10(7):1001–1009, juillet 2001.

J. IDIER, H. PIET-LAHANIER, G LE BESNERAIS, et F. CHAMPAGNAT. Cours de traitement numérique du signal. ONERA, Paris, 2004.

L. JAHAN. Analyse spectrale haute-résolution de données pauvres : application à l'Astérosismologie. Rapport de stage de DEA, Laboratoire d'Astrophysique de l'Observatoire Midi-Pyrénées, 2001.

K.F. KAARESEN. Deconvolution of sparse spike trains by iterated window maximization. *IEEE Transactions on Signal Processing*, 45:1173–1183, mai 1997.

S.M. KAY. *Modern Spectral Estimation.* Prentice-Hall, Englewood Cliffs, 1988.

J.-P. KNEIB, P. HUDELOT, R. S. ELLIS, T. TREU, G. P. SMITH, P. MARSHALL, O. CZOSKE, I. SMAIL, et P. NATARAJAN. A wide-field Hubble Space Telescope study of the cluster Cl 0024+1654 at z=0.4. II. The cluster mass distribution. *The Astrophysical Journal*, 598:804–817, décembre 2003.

J. KORMYLO et J.M. MENDEL. Maximum-likelihood detection and estimation of Bernoulli-Gaussian processes. *IEEE Transactions on Information Theory*, 28:482–488, 1982.

M. KUNT. *Traitement numérique des signaux.* Presses Polytechniques et Universitaires Romandes, 1984.

R. KUSCHNIG, W. W. WEISS, R. GRUBER, P. Y. BELY, et H. JENKNER. Microvariability survey with the Hubble Space Telescope Fine Guidance Sensors. Exploring the instrumental properties. *Astronomy and Astrophysics*, 328:544–550, décembre 1997.

J. LAFLER et T. D. KINMAN. An RR Lyrae star survey with the Lick 20-INCH Astrograph II. The calculation of RR Lyrae periods by electronic computer. *Astrophysical Journal Supplement*, 11:216–222, juin 1965.

M. LAVIELLE. Bayesian deconvolution of Bernoulli-Gaussian processes. *Signal Processing*, 33(1):67–79, juillet 1993.

M. LAVIELLE. A stochastic procedure for parametric and non-parametric estimation in the case of incomplete data. *Signal Processing*, 42(1):3–17, février 1995.

P. LENZ et M BREGER. *Period04 package.* Communications in Asteroseismology, 146, 2005.

P. LÉNA, éditeur. *Les Sciences du Ciel.* Flammarion, 1996.

N. R. LOMB. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39:447–462, février 1976.

S. MALLAT et Z. ZHANG. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

R.J. MARTIN. Autoregression and irregular sampling: spectral estimation. *Signal Processing*, 77:139–157, 1999.

R. J. MARTIN. *Irregularly sampled signals: theories and techniques for analysis.* PhD thesis, University College, London, janvier 1998.

R. E. NATHER, D. E. WINGET, J. C. CLEMENS, C. J. HANSEN, et B. P. HINE. The Whole Earth Telescope: a new astronomical instrument. *The Astrophysical Journal*, 361:309–317, septembre 1990.

W.H. PRESS, S.A. TEUKOLSKY, W.T. VETTERLING, et B.P. FLANNERY. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 1992.

J.-L. PRIEUR. Habilitation à diriger des recherches, en cours de rédaction. 2005.

H. RABESON. Analyse temps-fréquence de signaux irrégulièrement échantillonés : application à l'Astrophysique. Rapport de stage de DEA, Laboratoire d'Astrophysique de l'Observatoire Midi-Pyrénées, 2004.

S. RICHARDSON et P.J. GREEN. On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, B*, 59:731–792, 1997.

J. RISSANEN. Modeling by shortest data description. *Automatica*, 14:468–478, 1978.

A. RIVOIRA, Y. MOUDDEN, et G. FLEURY. Real time continuous AR parameter estimation from randomly sampled observations. Dans *Proceedings of ICASSP*, volume 2, pages 1725–1728, Orlando, USA, mai 2002.

C. ROBERT. *Méthodes de Monte-Carlo par Chaînes de Markov*. Economica, Paris, 1996.

D. H. ROBERTS, J. LEHAR, et J. W. DREHER. Time series analysis with CLEAN. I. Derivation of a spectrum. *The Astronomical Journal*, 93(4):968–989, avril 1987.

M.D. SACCHI, T.J. ULRYCH, et C.J. WALKER. Interpolation and extrapolation using a high-resolution discrete Fourier transform. *IEEE Transactions on Signal Processing*, 46(1):31–38, janvier 1998.

K. SAUER et C. BOUMAN. A local update strategy for iterative reconstruction from projections. *IEEE Transactions on Signal Processing*, 41(2):534–548, février 1993.

J. D. SCARGLE. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853, décembre 1982.

U. J. SCHWARTZ. Mathematical-statistical description of the iterative beam removing technique (method CLEAN). *Astronomy and Astrophysics*, 65:345–356, avril 1978.

P. STOICA, R. L. MOSES, B. FRIEDLANDER, et T. SÖDERSTRÖM. Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(3):378–392, mars 1989.

J. SUN et J. ZHANG. Global convergence of conjugate gradient methods without line search. *Annals of Operations Research*, 103:161–173, 2001.

A.N. TIKHONOV et V.Y. ARSENIN. *Solutions of ill-posed problems*. V.H. Winston & Sons, Washington, DC, 1977.

J.-Y. TOURNERET. Estimation-détection. Notes de cours du DEA Signal, Image et Acoustique, INPT, Toulouse, 2002.

R. VIO, T. STROHMER, et W. WALSTEKER. On the reconstruction of irregularly sampled time series. *Publications of the Astronomical Society of the Pacific*, 112:74–90, janvier 2000.

R. YARLAGADDA, J.B. BEDNAR, et T.L. WATT. Fast algorithms for ℓ^p deconvolution. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33(1):174–182, février 1985.

Table des matières

Remerciements **v**

Introduction **1**

1 L'estimation de spectres de raies dans un contexte astrophysique **5**

1.1 Différents champs d'application 5

1.1.1 L'étude d'étoiles variables 6

1.1.2 Les systèmes multiples 6

1.1.3 Des problèmes d'analyse spectrale 7

1.2 Spécificités des données 8

1.2.1 Une couverture temporelle incomplète 8

1.2.2 Nature des perturbations 11

1.3 Conséquences pour l'analyse spectrale 12

1.3.1 Définition d'une fréquence limite de Nyquist 13

1.3.2 Une fenêtre spectrale complexe 16

1.3.3 Conséquences sur le périodogramme 20

2 Les outils classiques et leurs limites **25**

2.1 Les périodogrammes 28

2.1.1 Le périodogramme de Lomb-Scargle 28

2.1.2 La méconnue DCDFT et le périodogramme à date compensée 31

2.1.3 Interprétation statistique 33

2.1.4 Prise en compte de perturbations non *i.i.d.* 36

2.2 Déconvolution itérative : méthodes CLEAN 36

2.2.1 Principe général 37

2.2.2 Nécessité d'une stabilisation 38

2.2.3 Extensions 40

2.2.4 Conclusion 40
2.3 Problèmes à une période : techniques spécifiques 44
2.4 Méthodologie adoptée : un problème inverse mal posé 45

3 Approche par pénalisation **49**
3.1 Analyse de l'opérateur $\mathbf{W}$ 52
3.1.1 Échantillonnage régulier 52
3.1.2 Échantillonnage régulier à données manquantes 53
3.1.3 Échantillonnage irrégulier 53
3.1.4 Conséquences sur l'inversion généralisée 54
3.2 Différents types de pénalisation 54
3.2.1 Régularisation quadratique 54
3.2.2 Régularisations non quadratiques 56
3.2.3 Bilan 61
3.3 Sélection des hyperparamètres 63
3.3.1 Choix de s pour la pénalisation hyperbolique 63
3.3.2 Choix du paramètre de régularisation 65
3.4 Algorithmes d'optimisation pour une pénalisation hyperbolique 69
3.4.1 Gradients Conjugués 69
3.4.2 L'algorithme IRLS 70
3.4.3 Algorithme de type Gauss-Seidel 71
3.5 Algorithmes d'optimisation pour une pénalisation ℓ^1 73
3.5.1 Programmation quadratique 74
3.5.2 Algorithme IRLS 74
3.5.3 Un test d'optimalité 77
3.5.4 Algorithme de type Gauss-Seidel 78
3.6 Prise en compte de la symétrie hermitienne 79
3.6.1 Des estimateurs « souvent » identiques 79
3.6.2 Un modèle à deux fois moins d'inconnues 80
3.6.3 Intérêt pour les algorithmes de gradients conjugués 81
3.7 Performances algorithmiques 82
3.7.1 Pénalisation hyperbolique 82
3.7.2 Pénalisation ℓ^1 85
3.8 Conclusions 89

4 Approche par modélisation Bernoulli-Gaussienne 93

4.1 Description du modèle . . . 95

4.1.1 Une structure hiérarchique 95

4.1.2 ... adaptée à l'estimation bayésienne . . . 97

4.1.3 Loi jointe ou loi marginale? . . . 98

4.1.4 Quid des hyperparamètres? . . . 99

4.2 Estimation au sens du maximum *a posteriori* (MAP) . . . 100

4.2.1 Un algorithme performant : SMLR . . . 101

4.2.2 Estimation délicate des hyperparamètres . . . 102

4.2.3 Optimisations alternatives . . . 104

4.3 Estimation au sens de la moyenne *a posteriori* (MP) . . . 104

4.3.1 La puissance de l'échantillonnage stochastique . . . 105

4.3.2 Avantages techniques de la loi jointe . . . 106

4.3.3 L'intérêt de $\widehat{\boldsymbol{q}}_{\mathrm{MP}}$ . . . 108

4.3.4 Échantillonnage des hyperparamètres . . . 110

4.4 Mise en œuvre de l'estimation MP . . . 114

4.4.1 L'échantillonnage de Gibbs classique . . . 115

4.4.2 Accélérations de l'algorithme . . . 118

4.5 Résultats d'estimation sur le modèle B-G . . . 123

4.5.1 Comportement algorithmique . . . 124

4.5.2 Influence des hyperparamètres . . . 126

4.5.3 Bilan de l'estimation . . . 128

4.6 Prise en compte de décalages fréquentiels . . . 130

4.6.1 Description du modèle . . . 131

4.6.2 Principe de l'estimation . . . 132

4.6.3 Mise en œuvre : algorithme hybride . . . 135

4.7 Résultats d'estimation sur le modèle BG avec prise en compte de décalages . . . 138

4.8 Conclusion . . . 140

5 Application au traitement de données réelles 143

5.1 Présentation des données . . . 145

5.2 Comparaison des méthodes sur des données simulées . . . 146

5.2.1 Estimation par pénalisation . . . 147

5.2.2 Estimation par modélisation B-G . . . 152

5.2.3 Estimation par modélisation B-G avec décalages en fréquence 154
5.2.4 Bilan 156
5.3 Application aux données réelles 157
5.3.1 Estimation par pénalisation 157
5.3.2 Estimation par modélisation B-G 160
5.3.3 Estimation par modélisation B-G avec décalages en fréquence 164
5.3.4 Bilan 167
5.3.5 Comparaison avec les résultats sur les données corrigées 168
5.4 Conclusion 175

6 Conclusion et perspectives **177**

Annexes

A Calculs pour l'approche pénalisée **183**
A.1 Critères et gradients 183
A.2 Test d'optimalité pour un minimiseur du critère pénalisé par la norme ℓ^1 184
A.3 Minimisations scalaires pour les algorithmes de type Gauss-Seidel 185
A.3.1 Cas de la pénalisation hyperbolique 187
A.3.2 Cas de la pénalisation ℓ^1 187

B Calculs pour la modélisation Bernoulli-Gaussienne **189**
B.1 Loi du module en fonction des lois des parties réelle et imaginaire 189
B.2 Calculs relatifs à l'échantillonnage de Gibbs 190
B.2.1 Lois conditionnelles des hyperparamètres 190
B.2.2 Lois conditionnelles des variables Bernoulli-Gaussiennes 191
B.2.3 Lois conditionnelles des variables Bernoulli-Gaussiennes par blocs de trois composantes 192

Bibliographie **196**

AUTHOR : Sébastien BOURGUIGNON

TITLE : HIGH RESOLUTION SPECTRAL ANALYSIS OF IRREGULARLY SAMPLED DATA : APPLICATION TO ASTROPHYSICS

THESIS SUPERVISOR : Hervé CARFANTAN

SUMMARY :

The study of many astrophysical phenomena is based on the search for periodicities from time series, as light or radial velocity curves. Because of observation constraints, astrophysical data generally suffer missing data and irregular sampling. Thus, Fourier-based spectral analysis may not be satisfactory, and widespread heuristic CLEAN deconvolution methods may lack accuracy. This thesis addresses spectral analysis as an inverse problem, where the spectrum is discretized on an arbitrarily thin frequency grid. Regularization is then addressed by taking into account the prior sparseness of the solution, as we focus on line spectra estimation.

A first approach considers the minimization of a penalized least-squares criterion, where the penalization function is designed to retrieve sparse solutions. In particular, penalization by the ℓ^1-norm is studied in application to complex variables, that shows a satisfactory behavior in terms of prior modeling. Several powerful optimization algorithms are developed that allow a very high spectral resolution.

Second, a probabilistic regularization is studied by modeling the spectral amplitudes as the realization of a Bernoulli-Gaussian process. Bayesian posterior mean estimation is then addressed using Monte-Carlo Markov Chain methods, which enable a fully unsupervised procedure. The probabilistic interpretation of the estimator combined with variance information for each estimated parameter then provides confidence levels, which is crucial information for astronomy. Significant algorithmic improvements are proposed to accelerate the classic Gibbs sampling algorithm. Then, continuous-valued frequency shifts are introduced that substantially improve the frequency precision at a reasonable computational cost.

Simulations illustrate the estimation quality for each method and the performances of the proposed algorithms. An application to astrophysical experimental data is finally presented that brings out the advantage of this methodology compared to classic spectral analysis methods.

KEYWORDS : Spectral analysis, irregular sampling, astrophysics, regularization, ℓ^1 penalization, optimization, Bernoulli-Gaussian processes, MCMC methods.

AUTEUR : Sébastien BOURGUIGNON

TITRE : ANALYSE SPECTRALE À HAUTE RÉSOLUTION DE SIGNAUX IRRÉGULIÈREMENT ÉCHANTILLONNÉS : APPLICATION A L'ASTROPHYSIQUE

DIRECTEUR DE THÈSE : Hervé CARFANTAN

DATE ET LIEU DE SOUTENANCE : mercredi 14 décembre 2005 à 10h30, salle Pyrénées de l'Observatoire Midi-Pyrénées, Toulouse

---

RÉSUMÉ : L'étude de nombreux phénomènes astronomiques repose sur la recherche de périodicités dans des séries temporelles (courbes de lumière ou de vitesse radiale). En raison des contraintes observationnelles, la couverture temporelle des données résultantes est souvent incomplète, présentant des trous périodiques ainsi qu'un échantillonnage irrégulier. L'analyse du contenu fréquentiel de telles séries basée sur le spectre de Fourier s'avère alors inefficace et les méthodes heuristiques de déconvolution de type CLEAN, couramment utilisées en astronomie, ne donnent pas entière satisfaction. Cette thèse s'inscrit dans le formalisme fréquemment rencontré depuis les années 1990 abordant l'analyse spectrale sous la forme d'un problème inverse, le spectre étant discrétisé sur une grille fréquentielle arbitrairement fine. Sa régularisation est alors envisagée en traduisant la nature a priori parcimonieuse de l'objet à reconstruire: nous nous intéressons ici à la recherche de raies spectrales.

Une première approche envisagée a trait au domaine de l'optimisation et consiste à minimiser un critère de type moindres carrés, pénalisé par une fonction favorisant les solutions parcimonieuses. La pénalisation par la norme ℓ^1 est en particulier étudiée en extension à des variables complexes et s'avère satisfaisante en termes de modélisation. Nous proposons des solutions algorithmiques particulièrement performantes permettant d'envisager une analyse à très haute résolution fréquentielle.

Nous étudions ensuite la modélisation probabiliste des amplitudes spectrales sous la forme d'un processus Bernoulli-Gaussien, dont les paramètres sont estimés au sens de la moyenne a posteriori à partir de techniques d'échantillonnage stochastique, permettant d'envisager une estimation totalement non supervisée. L'interprétation probabiliste du résultat ainsi que l'obtention conjointe des variances associées, sont alors d'un intérêt astrophysique majeur, s'interprétant en termes de niveaux de confiance sur les composantes spectrales détectées. Nous proposons dans un premier temps des améliorations de l'algorithme échantillonneur de Gibbs permettant d'accélérer l'exploration de la loi échantillonnée. Ensuite, nous introduisons des variables de décalage fréquentiel à valeur continue, permettant d'augmenter la précision de l'estimation sans trop pénaliser le coût calculatoire associé.

Pour chaque méthode proposée, nous illustrons sur des simulations la qualité de l'estimation ainsi que les performances des algorithmes développés. Leur application à un jeu de données issu d'observations astrophysiques est enfin présentée, mettant en évidence l'apport d'une telle méthodologie par rapport aux méthodes d'analyse spectrale habituellement utilisées.

MOTS-CLEFS : Analyse spectrale, échantillonnage irrégulier, astrophysique, régularisation, pénalisation ℓ^1, optimisation, processus Bernoulli-Gaussien, méthodes MCMC.

DISCIPLINE ADMINISTRATIVE : Traitement du Signal

INTITULÉ ET ADRESSE DU LABORATOIRE :

Laboratoire d'Astrophysique de Toulouse et de Tarbes – UMR 5572

Observatoire Midi–Pyrénées

14, avenue Edouard Belin

31400 Toulouse, FRANCE