

Étude statistique de la distribution des rayons cosmiques d'ultra haute énergie

Antoine DUVAL

13 septembre 2010

Rapport du stage effectué au laboratoire Astroparticules et cosmologie
(Université Paris VII) sous la direction de Jacques Delabrouille et Gilles Faÿ en
juillet 2010

Remerciements

Je tiens à remercier toutes les personnes de l'APC qui m'ont accueilli pendant mon stage et plus particulièrement Gilles et Jacques qui m'ont encadré et qui ont relu ce rapport avec attention

Table des matières

1	Les rayons cosmiques	3
1.1	Qu'est-ce qu'un rayon cosmique?	3
1.2	Détection des rayons cosmiques	3
1.3	Sources probables des rayons cosmiques	4
1.4	Analyse statistique	5
2	Présentation du problème statistique	6
2.1	Modélisation	6
2.2	Notations	6
2.3	Estimation non-paramétrique	7
2.3.1	Généralités sur les estimateurs	7
2.3.2	Exemples d'estimateurs	7
3	Estimateur à noyau	8
3.1	Etude de l'estimateur à noyau	8
3.1.1	Premier exemple sur \mathbb{R}	8
3.1.2	Le cas de la sphère	8
3.2	Validation croisée	9
4	Estimateur à ondelettes	15
4.1	Motivation et rappels sur la décomposition de Fourier	15
4.2	Décomposition en ondelettes et en needlets	15
4.3	Seuillage et estimateur à needlets	17
4.4	Comparaison noyaux-needlets	19
5	Simulations	19
5.1	Particularités liées à l'algorithmique	19
5.2	L'estimateur à noyau	19
5.3	L'estimateur à needlets	22
A	Exemple d'estimations	24

1 Les rayons cosmiques

1.1 Qu'est-ce qu'un rayon cosmique ?

Les rayons cosmiques sont des particules chargées, qui proviennent de l'espace. Il s'agit principalement de protons mais on trouve également quelques noyaux plus lourds (voir [4]). Ils ont une énergie qui peut aller jusqu'à 10^{22} eV. Il existe probablement des rayons d'énergie supérieure mais ceux-ci sont en nombre très faible et interagissent fortement avec le fond diffus cosmologique ce qui fait qu'ils ne sont pas observés (la distribution en énergie des rayons cosmiques est donnée en figure 2). L'objet de ce stage est d'étudier les méthodes qui permettent d'analyser statistiquement la distribution spatiale des sources de ces particules afin de tenter de comprendre la nature de ces sources qui reste encore largement inconnue.

Nous nous intéresserons ici aux particules ayant une énergie supérieure à 10^{19} eV. Pourquoi cette limite? Tout d'abord il faut savoir que si les processus d'accélération de ces particules jusqu'à 10^{15} eV semblent compris (il s'agit du mécanisme de Fermi voir section 1.3) en revanche la présence de rayons à des énergies supérieures reste inexplicée. Un des problèmes astrophysique d'aujourd'hui est donc de comprendre l'origine de ces particules de hautes énergies. Pour cela on veut étudier les zones de production de ces rayons. Or il se trouve que ces rayons sont déviés par les champs magnétiques extragalactiques. Ces champs étant inconnus cela pose un problème pour déduire la position de la source qui n'est pas forcément dans la direction d'arrivée du rayon. C'est pourquoi on étudie les rayons de plus hautes énergies qui sont les moins déviés par ces champs magnétiques. En effet les champs magnétiques intergalactiques sont uniformes sur des cellules d'environ 1 Mpc (1 pc = 3,26 années-lumière) et ont une valeur d'environ 10^{-9} gauss (voir [9]). La force magnétique exercée sur les particules cosmiques vaut alors $q(\vec{v} \wedge \vec{B})$ ce qui donne un rayon de courbure $R = \frac{mv}{qB} = \frac{E}{qcB}$ en utilisant la formule relativiste $E = mc^2$. Les particules sont alors d'autant moins déviées que ce rayon de courbure est grand devant la taille des cellules de champs magnétiques. En pratique avec une énergie de 10^{19} eV on a un rayon de courbure d'environ 10 Mpc et une déviation de quelques degrés entre la direction d'arrivée du rayon et celle de la source qui l'a produit (voir figure 1).

Une des difficultés de cette étude est le faible nombre de données, en effet au delà de 10^{19} eV on ne reçoit plus qu'une particule par kilomètre carré et par an. De plus on n'observe pas tout le ciel. C'est pourquoi il faut développer des outils d'analyse localisée comme la décomposition en ondelettes sur la sphère pour analyser statistiquement les données récoltées.

1.2 Détection des rayons cosmiques

Le plus grand observatoire dédié à l'observation des rayons cosmiques est actuellement l'observatoire Pierre Auger en Argentine. Il ne détecte pas directement les rayons cosmiques mais les particules secondaires produites lors de l'entrée de la particule ultra-énergétique dans la haute atmosphère. Ces particules secondaires peuvent ensuite être captées dans des détecteurs individuels. Ces détecteurs sont distants d'un kilomètre et demi et couvrent au total une surface de 3 000 km². L'intérêt de ce dispositif est que chaque cascade induite par une particule est détectée par plusieurs capteurs. On peut ensuite, en mesurant les temps où la cascade atteint chaque détecteur, déterminer la direction d'arrivée du rayon (actuellement l'ordre de l'erreur sur la direction est inférieure à un degré). En comparant les énergies des particules détectées aux modèles de cascades simulées on peut également estimer l'énergie de la particule initiale (on se reportera à [10] pour une présentation succincte de l'observatoire et des techniques de détection des rayons cosmiques).

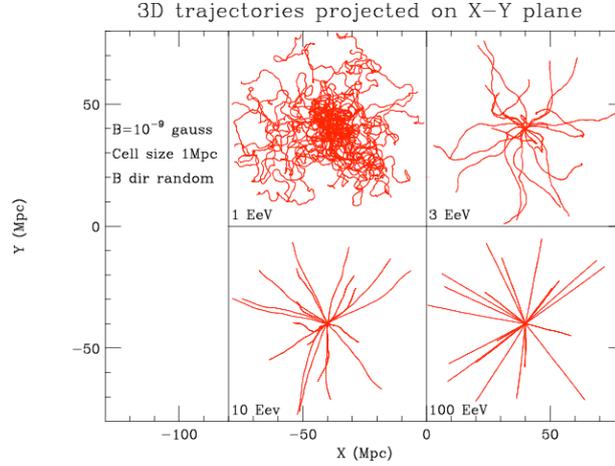


FIGURE 1 – Trajectoire des rayons cosmiques en présence de champs magnétiques. Les trajectoires sont tracées jusqu’à 40 Mpc de la source. D’après Cronin ([4])

1.3 Sources probables des rayons cosmiques

Plusieurs hypothèses pourraient expliquer l’accélération de ces particules aux hautes énergies. La plus ancienne est celle formulée par Fermi : il pensait que ces énergies étaient acquises après plusieurs chocs successifs avec des nuages comportant un important champ magnétique. Par exemple si une particule rencontre un nuage magnétisé, elle va être déviée jusqu’à faire demi-tour. L’énergie de la particule sera conservée lors de la collision ; mais cette conservation a lieu dans le référentiel du nuage. Ainsi si le nuage va à contresens de la particule, cette dernière sera accélérée par la collision, réciproquement elle sera ralentie si le nuage va dans le même sens. On peut alors imaginer que deux nuages magnétisés se rencontrent ; cette situation peut se produire aux abords des supernovae qui émettent de tels nuages qui peuvent ensuite se rattraper ; dans ces conditions une particule placée entre les deux nuages va subir des chocs successifs qui l’accéléreront à chaque fois. Une variante de ce phénomène peut apparaître si de nombreux nuages circulent dans un milieu avec des directions réparties aléatoirement. Dans ce cas une particule donnée a statistiquement plus de chances de rencontrer un nuage qui va dans le sens opposé à sa trajectoire qu’un nuage qui va dans le même sens. Elle sera donc progressivement accélérée jusqu’à atteindre des vitesses très supérieures à celles des nuages, à ce moment le déséquilibre statistique ne sera plus significatif et le processus deviendra moins efficace. Cette hypothèse a été étudiée par les physiciens et aujourd’hui il semble que les conditions observées dans différentes structures astrophysiques peuvent accélérer des particules jusqu’à environ 10^{15} eV via ce mécanisme (voir [3]).

L’origine de particules de plus hautes énergies nécessitent selon la plupart des hypothèses l’existence d’intenses champs magnétiques sur une grande étendue spatiale car les particules doivent rester confinées dans la zone magnétique pour être accélérées. Pour une zone de taille L avec un champ B l’énergie maximale pouvant être atteinte par une particule de charge q est environ $E_{max} = qcBL$. Il faut donc chercher des zones où le produit BL est le plus élevé

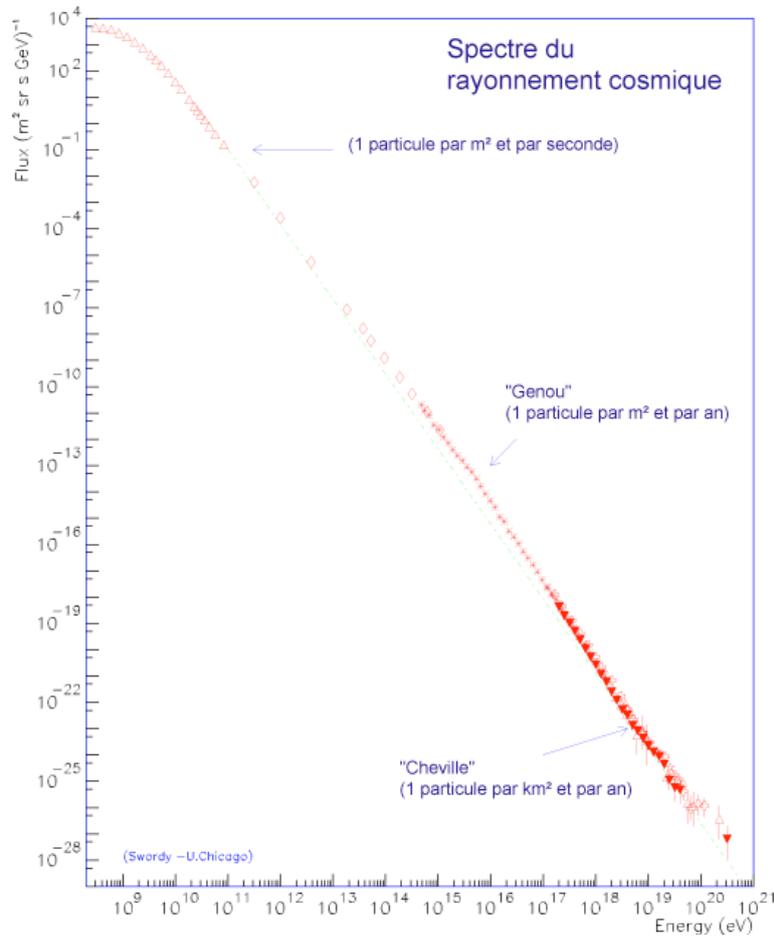


FIGURE 2 – Flux de particules en fonction de leur énergie. On observe un décrochage à partir de 10^{15} eV ce qui suggère un mécanisme de formation différent pour les rayons au-dessous et au-dessus de cette limite. Source *Wikipédia* (Simon Swordy —Université de Chicago)

possible. Un type d'astres possibles sont les noyaux actifs de galaxies (AGN). Ce sont des galaxies possédant en leur centre un trou noir supermassif (d'un million à un milliard de masses solaires) qui attire de la matière et où règne un intense champ magnétique. Ces astres engendrent des jets de matière qui en rencontrant le milieu intergalactique pourraient accélérer des particules grâce à un mécanisme de Fermi (voir [9] pour des explications plus détaillées). D'autres astres possédant des caractéristiques adéquates sont les étoiles à neutrons (champ magnétique de 10^{10} Tesla pour un rayon de 10 km) qui pourraient être à l'origine des particules de hautes énergies.

1.4 Analyse statistique

Un problème qui se pose est donc de connaître la position des sources des rayons cosmiques pour savoir quels peuvent être les mécanismes à l'origine de leur production. Ensuite par exemple si l'on suppose que ce sont les AGNs qui produisent ces rayons et si l'on connaît la position des AGNs on peut regarder les corrélations entre les catalogues d'AGNs et les sources de rayons

cosmiques et confirmer ou infirmer l'hypothèse (cette question a été traitée en détail dans [11] et [10] à partir des données présentées sur la figure 3). On voudrait donc analyser la répartition spatiale des sources de rayons cosmiques à l'aide des observations. Cette question nécessite des outils mathématiques complexes à cause de la nature des données à analyser.

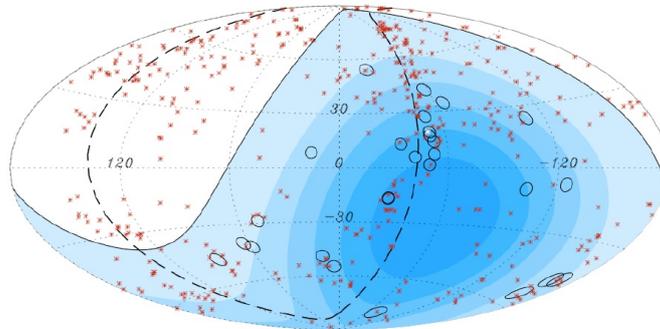


FIGURE 3 – Etude des corrélations entre AGN et rayons cosmiques. Les niveaux de bleus représentent le temps d’observation par l’observatoire Pierre Auger : plus la région est bleue, plus elle est observée, la région blanche n’est jamais observée. Les points rouges représente les AGNs et les cercles noir les rayons cosmiques. D’après Santos ([11])

2 Présentation du problème statistique

2.1 Modélisation

La première idée consiste à considérer l’observation de ces rayons cosmiques comme des réalisations d’événements aléatoires indépendants et de même loi sur la sphère \mathbb{S}^2 , cette loi étant à densité par rapport à la mesure de Lebesgue; on cherche alors à estimer cette densité f . Il existe plusieurs méthodes de résolution de ce problème d’estimation selon les hypothèses que l’on formule sur f . Le cas le plus simple est l’estimation paramétrique, il s’agit du cas où on postule que f est dans un espace vectoriel de dimension finie (par exemple on suppose que f est un polynôme de degré inférieur ou égal à n); l’estimation de f se réduit alors à l’estimation d’un nombre fini de paramètres. Cela nécessite d’avoir déjà une connaissance précise de la fonction or, pour le problème des rayons cosmiques, les connaissances physiques sont trop imprécises pour cela. On doit donc utiliser des méthodes dites non-paramétriques dans lesquelles f est supposée appartenir à un certain espace de fonctions de dimension infinie (on doit quand même supposer une régularité minimale sur f pour obtenir des résultats sur l’efficacité de ces méthodes).

Une partie du travail effectué lors de ce stage a consisté à simuler des données aléatoires suivant une loi à densité connue, puis à établir un algorithme pour reconstituer cette densité à l’aide des seules données et enfin à en tester l’efficacité.

2.2 Notations

Avant d’aller plus loin définissons quelques notations utilisées dans ce rapport :

- \mathbb{S}^2 désigne la sphère unité de \mathbb{R}^3 ;
- χ_A désigne la fonction indicatrice de l’ensemble A ;
- i.o. est une abréviation signifiant infiniment souvent (pour infinity often);

- i.i.d. signifie indépendant et identiquement distribué ;
- p.s. signifie presque sûrement ;
- lorsque les intégrales ne précisent pas d'ensemble d'intégration elles portent sur tout \mathbb{R} ou toute la sphère \mathbb{S}^2 selon le contexte (parfois la même formule s'applique aux deux cas).
- μ désigne la mesure de Lebesgue sur \mathbb{S}^2 ou sur \mathbb{R} ;

2.3 Estimation non-paramétrique

On observe un échantillon X_1, X_2, \dots, X_n de points sur la sphère (ou parfois sur la droite réelle), ces points sont répartis aléatoirement de manière indépendante en suivant une loi à densité f par rapport à la mesure de Lebesgue. On cherche à estimer f par des méthodes non-paramétriques. On commencera par remarquer que l'espace de probabilité de ces événements a une mesure qui dépend de f , en conséquence de quoi on devrait en toute rigueur noter \mathbb{P}_f et \mathbb{E}_f les probabilités et espérances associées, mais on omettra l'indice f pour alléger les notations.

2.3.1 Généralités sur les estimateurs

Un estimateur d'une fonction f sur \mathbb{S}^2 (généralement noté \hat{f}) est en fait une fonction des observations à valeur dans l'ensemble des fonctions sur \mathbb{S}^2 . Autrement dit à chaque série de n observations on associe une fonction de \mathbb{S}^2 . On demande généralement à un estimateur de converger en un certain sens vers la fonction qu'il estime lorsque n devient grand.

Il y a plusieurs manières d'évaluer la performance d'un estimateur. Une des plus courantes est de considérer l'Erreur Quadratique Moyenne (EQM) qui est la quantité : $\mathbb{E}((\hat{f} - f)^2)$

On peut décomposer cette EQM en une somme de deux termes :

$$\begin{aligned} \int \mathbb{E}((\hat{f} - f)^2) d\mu &= \int \mathbb{E}((\hat{f} - \mathbb{E}(\hat{f}))^2) d\mu + \int \mathbb{E}((\mathbb{E}(\hat{f}) - f)^2) d\mu + 2 \int \mathbb{E}((\hat{f} - \mathbb{E}(\hat{f}))(\mathbb{E}(\hat{f}) - f)) d\mu \\ &= \int \mathbb{E}((\hat{f} - \mathbb{E}(\hat{f}))^2) d\mu + \int (\mathbb{E}(\hat{f}) - f)^2 d\mu \\ &= \text{Variance} + \text{Biais}^2 + 0 \end{aligned} \tag{1}$$

car $(\mathbb{E}(\hat{f}) - f)$ est constant et $(\hat{f} - \mathbb{E}(\hat{f}))$ est par définition de moyenne nulle .

Le plus souvent, comme on l'explicitera pour l'un des estimateurs étudiés, il y a un compromis à trouver entre la variance et le biais de l'estimateur

2.3.2 Exemples d'estimateurs

Deux méthodes d'estimation seront étudiées dans ce rapport : la première est l'estimation à noyau qui pour chaque particule observée ajoute une densité estimée aux points proches de l'observation. La deuxième est celle des estimateurs à ondelettes, il s'agit d'un cas particulier de la classe plus générale des estimateurs de projection. Par exemple si on suppose que f est de carré intégrable, on peut la décomposer en harmoniques sphériques : $f = \sum a_{l,m} Y_{l,m}$. Une idée est alors de projeter f sur les harmoniques de basses fréquences et d'estimer les coefficients correspondants :

$$\hat{f} = \sum_{l < l_{\max}} \hat{a}_{l,m} Y_{l,m}$$

en considérant que l'on ne peut bien estimer les hautes fréquences de f qu'avec un nombre suffisant d'observations. Le problème, dans ce cas précis, est que les harmoniques sphériques sont

à support dans toute la sphère alors que l'on n'observe pas tout le ciel (en effet avec un seul observatoire une région importante reste inobservée), ce qui pose des difficultés pour estimer le produit scalaire de f avec les harmoniques sphériques. Une solution est alors de trouver une autre décomposition des fonctions de carré intégrable, c'est ce qui sera étudiée dans ce rapport où on introduira la décomposition en ondelettes. L'application d'un seuillage non-linéaire des coefficients estimés est connu pour donner des résultats théoriques optimaux

Ces deux méthodes d'estimation utilisées (ondelettes et noyaux) donnent des fonctions continues, plus précisément le caractère plus ou moins lisse des fonctions estimées va dépendre d'un paramètre (largeur de bande pour le noyau et fréquence de coupure pour les ondelettes). La performance de l'estimateur dépendra alors du choix de ce paramètre, mais l'absence d'hypothèse sur la fonction inconnue ne nous permet pas de choisir une valeur a priori, il faudra donc développer des méthodes dites adaptatives pour choisir la valeur du paramètre en fonction des observations afin d'optimiser l'estimation.

3 Estimateur à noyau

3.1 Etude de l'estimateur à noyau

3.1.1 Premier exemple sur \mathbb{R}

Plaçons nous sur \mathbb{R} . Soit X_1, X_2, \dots, X_n un échantillon i.i.d de n points tirés au hasard suivant une densité inconnue f , alors Parzen et Rosenblatt proposent d'estimer f à l'aide d'un noyau. Ils définissent :

$$\hat{f}_{\lambda,n}(x) = \frac{1}{n} \sum_{i=1}^n \lambda K(\lambda(x - X_i))$$

où K est une fonction d'intégrale 1 appelée *noyau* et $1/\lambda$ *largeur de bande*.

Souvent K est fixé une fois pour toute tandis que λ varie avec n . L'enjeu est alors de trouver la dépendance qui optimise la vitesse de convergence de l'EQM vers 0. Intuitivement si λ augmente le noyau sera plus fin et la variance de l'estimateur augmentera tandis que le biais diminuera et réciproquement. Il y aura donc un compromis à trouver (voir figure 4). Ce compromis dépendra de n , en effet si n augmente à λ fixé alors le biais reste constant car les observations sont indépendantes tandis que la variance diminue, ainsi l'équilibre biais-variance se trouve modifié.

C'est donc le choix de λ en fonction de n qui est le problème crucial pour l'estimation de densité. Plusieurs méthodes ont été développées pour cela; nous en aborderons une à la section suivante.

3.1.2 Le cas de la sphère

Dans le cas de la sphère on peut définir de même un estimateur à noyau à partir d'un noyau réel L . (Cela a été étudié entre autre par Klemelä dans [6]) Avec les notations précédentes l'estimateur s'écrit :

$$\hat{f}_{n,\lambda}(x) = \frac{c(\lambda)}{n} \sum_{i=1}^n K(\lambda \arccos(x \cdot X_i))$$

Remarque 1. $\arccos(x \cdot y)$ est la distance riemannienne, il s'agit de la longueur de l'arc de grand cercle reliant deux points de la sphère.

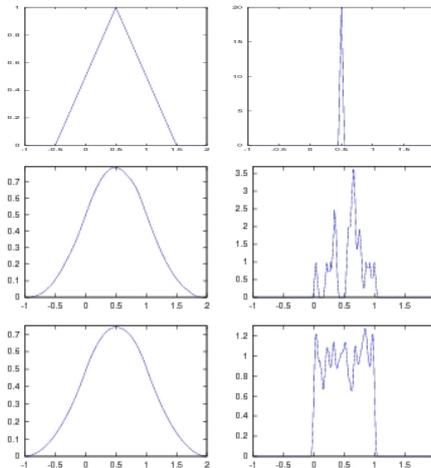


FIGURE 4 – Illustration du compromis biais-variance : On a tiré 20 points au hasard uniformément entre 0 et 1 puis on applique l’estimateur à noyau (le noyau est représenté en haut, une estimation au milieu, une moyenne de 20 estimations en bas). À gauche on a choisi $\lambda = 1$, on voit un fort biais mais une faible variance; à droite $\lambda = 20$ on voit que l’estimateur dépend de la position précise des points donc il y a une forte variance, en revanche en moyennant l’estimateur se rapproche de la vraie densité, le biais est donc faible

Remarque 2. $c(\lambda)$ est une constante de normalisation :

$$c(\lambda) = \left(\int_{\mathbb{S}^2} K(\lambda \arccos(x \cdot y)) d\mu(y) \right)^{-1}$$

où μ est la mesure de Lebesgue sur \mathbb{S}^2

Dans le cas réel on avait simplement $c(\lambda) = \left(\int K(\lambda(x - y)) dy \right)^{-1} = \lambda$

À partir de là, avec des hypothèses de régularité sur la fonction f , il est possible de montrer que le biais tend vers 0 lorsque λ devient infiniment grand, qu’à λ fixé la variance tend vers 0 lorsque n tend vers l’infini et d’évaluer les vitesses de convergences respectives. On peut alors en déduire une dépendance optimale de λ en fonction de n (c’est ce qui est réalisé par Klemelä dans [6]). Mais cette dépendance optimale ainsi que l’estimation de l’erreur de l’estimateur dépendent de la classe de régularité de f , or cette fonction f est inconnue. C’est pour cela qu’on utilise une autre méthode qui détermine une largeur de bande à partir des données.

3.2 Validation croisée

Cette section s’inspire fortement de [2] où est étudiée la validation croisée.

La méthode de la validation croisée permet de choisir une largeur de bande à partir des observations, ce qui est utile pour coder un estimateur à noyau par exemple. Le principe est de regarder si chaque observation est bien prédite par les $n - 1$ autres. On rappelle que l’estimateur à noyau est donné par :

$$\hat{f}_{\lambda,n}(x) = \frac{1}{n} \sum_{i=1}^n \lambda K(\lambda(x - X_i))$$

On note alors $\hat{f}_{\lambda,n}^i$ l'estimateur obtenu après suppression de la i -ème donnée.

$$\hat{f}_{\lambda,n}^i(x) = \frac{1}{n-1} \sum_{j \neq i} \lambda K(\lambda(x - X_j))$$

Cet estimateur ne dépend pas de X_i et on interprète $\hat{f}_{\lambda,n}^i(X_i)$ comme la vraisemblance de l'observation X_i connaissant les $n-1$ autres données pour une valeur fixée du paramètre λ .

Ainsi on va chercher à maximiser la quantité :

$$L_\lambda = \prod_{i=1}^n \hat{f}_{\lambda,n}^i(X_i)$$

et on appellera λ^* le paramètre qui maximise L_λ .

Le but sera alors de montrer que sous de bonnes hypothèses on a

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} |\hat{f}_{\lambda^*,n}(x) - f(x)| dx = 0 \quad \text{p.s.} \quad (2)$$

Arguments heuristiques

On rappelle que l'on veut minimiser la quantité

$$\begin{aligned} R(\lambda) &= \mathbb{E} \int (\hat{f}_{\lambda,n}(x) - f(x))^2 dx \\ &= \int f^2 dx + \mathbb{E} \int \hat{f}_{\lambda,n}(x)^2 dx - 2\mathbb{E} \int \hat{f}_{\lambda,n}(x) f(x) dx \end{aligned} \quad (3)$$

Pour cela on va estimer la quantité $R(\lambda)$ à l'aide des données. Le premier terme est constant donc ne pose pas de problème, le deuxième terme est facilement estimable par $\int \hat{f}_{\lambda,n}(x)^2 dx$. Pour le troisième terme la première idée est de dire que l'échantillon X_1, X_2, \dots, X_n est distribué selon la densité f et donc d'évaluer $\frac{1}{n} \sum_{i=1}^n \hat{f}_{\lambda,n}(X_i)$. Mais $\hat{f}_{\lambda,n}$ dépend de l'observation X_i , ce qui induit donc un biais potentiel. C'est pourquoi on évalue $\frac{1}{n-1} \sum_{i=1}^n \hat{f}_{\lambda,n}^i(X_i)$.

En résumé on veut minimiser la quantité

$$\hat{R}(\lambda) = \int \hat{f}_{\lambda,n}(x)^2 dx - \frac{2}{n-1} \sum_{i=1}^n \hat{f}_{\lambda,n}^i(X_i)$$

dont on va montrer qu'elle est un estimateur sans biais de $R(\lambda)$

Nous voulons montrer que la validation croisée donne un estimateur convergent, c'est-à-dire que (2) est vérifiée. Le schéma de la preuve est le suivant :

- Montrer que le paramètre de bande choisi λ^* diverge vers $+\infty$;
- Etablir une borne supérieure pour la vitesse de divergence ;
- Montrer que n'importe quelle séquence γ_n de paramètres de bande qui diverge en respectant cette borne supérieure donne un estimateur convergent.

Donnons d'abord quelques arguments en faveur de cette démarche.

Si f est une densité de probabilité telle que $|\int f(x) \log(f(x)) dx| < \infty$ et si g est une autre densité de probabilité alors l'inégalité de Jensen assure

$$\int f(x) \log(g(x)) dx \leq \int f(x) \log(f(x)) dx$$

La quantité $\int f(x) \log\left(\frac{f(x)}{g(x)}\right) dx$ s'appelle la distance de Kullback-Leibler et mesure un "éloignement" entre les deux densités (on notera qu'elle n'est pas symétrique).

On remarque alors qu'en vertu de la loi des grands nombres on a :

$$\frac{1}{n} \log L_\lambda = \frac{1}{n} \sum_{i=1}^n \log \hat{f}_{\lambda, n-1}^i(X_i) \simeq \frac{1}{n} \sum_{i=1}^n \log \hat{f}_{\lambda, n}(X_i) \simeq \int f(x) \log \hat{f}_{\lambda, n}(x) dx$$

ainsi si γ_n est une suite qui vérifie

$$\int f(x) \log\left(\frac{f(x)}{f_{\gamma_n, n}(x)}\right) dx \rightarrow 0$$

alors comme $L_{\gamma_n} \leq L_{\lambda_n^*}$ par définition on peut espérer au moins approximativement

$$\int f(x) \log(f_{\gamma_n, n}(x)) dx \leq \int f(x) \log(f_{\lambda_n^*, n}(x)) dx \leq \int f(x) \log(f(x)) dx$$

et donc

$$\int f(x) \log\left(\frac{f(x)}{f_{\gamma_n, n}(x)}\right) dx > \int f(x) \log\left(\frac{f(x)}{f_{\lambda_n^*, n}(x)}\right) dx > 0$$

Théorème 1. On prend $\pi \in]0, 1[$ une constante et on pose $\Lambda_n = \{\lambda ; L_\lambda > \pi \sup_{\gamma > 0} L_\gamma\}$.

Alors si f est à support compact et si K est à support compact, strictement croissant sur $] -\infty, 0]$, strictement décroissant sur $[0, \infty[$ et que pour un certain $\delta > 0$ on a $\min(K(-\delta), K(\delta)) > 0$

Alors

$$\sup_{\lambda \in \Lambda_n} \int |f_{\lambda, n}(x) - f(x)| dx \rightarrow 0 \quad \text{p.s.}$$

Démonstration. On énonce d'abord quatre lemmes, le premier sera admis et on reprendra pour les trois derniers les preuves de [2].

Lemme 1. Pour $n > 1$ Λ_n est presque sûrement non vide et

$$\text{i) } \mathbb{P}(\inf_{\lambda \in \Lambda_n} \lambda \rightarrow 0) = 1$$

$$\text{ii) } \text{il existe } k \text{ tel que } \mathbb{P}(\sup_{\lambda \in \Lambda_n} \lambda > \frac{kn}{\ln n} \text{ i.o.}) = 0$$

Lemme 2. Pour $\tilde{K} = \chi_{[0,1]}$, posons

$$\tilde{f}_{\lambda, n}(x) = \frac{1}{n} \sum_{i=1}^n \lambda \tilde{K}(\lambda(x - X_i))$$

et

$$\tilde{f}_\lambda(x) = \int f(y) \lambda \tilde{K}(\lambda(x - y)) dy$$

Si λ_n est une suite divergeant vers $+\infty$, constante sur $]n_{k-1}, n_k]$ avec $n_k = [e^k]$ et telle que $\lambda_n = o\left(\frac{n}{\ln \ln n}\right)$, alors

$$\int |\tilde{f}_{\lambda_n, n}(x) - \tilde{f}_{\lambda_n}(x)| dx \rightarrow 0 \quad \text{p.s.}$$

Lemme 3.

$$\sup_{\lambda \in \Lambda_n} \int |\tilde{f}_{\lambda_n, n}(x) - \tilde{f}_\lambda(x)| dx \rightarrow 0 \quad \text{p.s.}$$

Lemme 4. Si $\tilde{K} = \sum_{i=1}^m \alpha_i \chi_{[a_i, b_i]}$ alors de même

$$\sup_{\lambda \in \Lambda_n} \int |\tilde{f}_{\lambda_n, n}(x) - \tilde{f}_{\lambda_n}(x)| dx \rightarrow 0 \quad \text{p.s.}$$

Preuve du théorème Définissons

$$f_\lambda(x) = \int f(y) \lambda K(\lambda(x-y)) dy$$

On peut voir que $f_\lambda \rightarrow f$ dans L^1 quand $\lambda \rightarrow \infty$. Remarquons alors

$$\sup_{\lambda \in \Lambda_n} \int |f_{\lambda, n}(x) - f(x)| dx \leq \sup_{\lambda \in \Lambda_n} \int |f_{\lambda, n}(x) - f_\lambda(x)| dx + \sup_{\lambda \in \Lambda_n} \int |f_\lambda(x) - f(x)| dx$$

par le lemme 1 le dernier terme tend vers 0.

Pour le deuxième terme : Soit $\epsilon > 0$, par densité on peut choisir un noyau \tilde{K} de la forme du lemme 4 tel que $\|K - \tilde{K}\|_{L^1} < \epsilon$, alors

$$\begin{aligned} \sup_{\lambda \in \Lambda_n} \int |f_{\lambda, n}(x) - f_\lambda(x)| dx &\leq \sup_{\lambda \in \Lambda_n} \int |f_{\lambda, n}(x) - \tilde{f}_{\lambda, n}(x)| dx \\ &+ \sup_{\lambda \in \Lambda_n} \int |f_\lambda(x) - \tilde{f}_\lambda(x)| dx \\ &+ \sup_{\lambda \in \Lambda_n} \int |\tilde{f}_{\lambda, n}(x) - \tilde{f}_\lambda(x)| dx \end{aligned} \quad (4)$$

Le dernier terme tend vers 0 par le lemme 4.

Pour le premier on écrit

$$\begin{aligned} \int |f_{\lambda, n}(x) - \tilde{f}_{\lambda, n}(x)| dx &= \int \left| \frac{1}{n} \sum_{i=1}^n (\lambda K(\lambda(x - X_i)) - \lambda \tilde{K}(\lambda(x - X_i))) \right| dx \\ &\leq \frac{1}{n} \sum_{i=1}^n \int |(\lambda K(\lambda(x - X_i)) - \lambda \tilde{K}(\lambda(x - X_i)))| dx = \|K - \tilde{K}\|_{L^1} \leq \epsilon \end{aligned} \quad (5)$$

Le deuxième terme se traite de la même façon en remplaçant $\frac{1}{n} \sum$ par \int

Preuve du lemme 2 Montrons d'abord que pour presque tout x

$$\tilde{f}_{\lambda_n, n}(x) - \tilde{f}_{\lambda_n}(x) \rightarrow 0 \text{ p.s.} \quad (6)$$

En posant $A_n = \{\tilde{f}_{\lambda_n}(x) - \tilde{f}_{\lambda_n, n}(x) \geq 0\}$

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \int |\tilde{f}_{\lambda_n, n}(x) - \tilde{f}_{\lambda_n}(x)| dx \\ &= 2 - 2 \liminf_{n \rightarrow \infty} \int [\tilde{f}_{\lambda_n}(x) - \chi_{A_n}(x)(\tilde{f}_{\lambda_n}(x) - \tilde{f}_{\lambda_n, n}(x))] dx \\ &\leq 2 - 2 \int f(x) dx = 0 \text{ p.s. par Fatou} \end{aligned} \quad (7)$$

Ce qui montre (6)

Fixons un x qui vérifie (1). Posons $Z_i(\lambda) = \lambda \tilde{K}(\lambda(x - X_i)) - \tilde{f}_\lambda(x)$ et $S_n(\lambda) = \sum_{i=1}^n Z_i(\lambda)$
Alors (1) se réécrit $(1/n)S_n(\lambda_n) \rightarrow 0$ p.s.

Fixons $\epsilon > 0$ On veut montrer que

$$\mathbb{P}\left(\frac{1}{n}S_n(\lambda_n) > \epsilon \text{ i.o.}\right) = 0$$

pour $k > 3$, $n_{k-1} > n_k/9$ Ainsi

$$\begin{aligned} \mathbb{P}\left(\frac{1}{n}S_n(\lambda_n) > \epsilon \text{ i.o.}\right) &= \mathbb{P}\left(\max_{n_{k-1} < n < n_k} (S_n(\lambda_n) - n\epsilon) > 0 \text{ i.o. } (k)\right) \\ &\leq \mathbb{P}\left(\max_{n_{k-1} < n < n_k} (S_n(\lambda_n) - n_{k-1}\epsilon) > 0 \text{ i.o. } (k)\right) \\ &\leq \mathbb{P}\left(\max_{n_{k-1} < n < n_k} (S_n(\lambda_n) - \frac{n_k}{9}\epsilon) > 0 \text{ i.o. } (k)\right) \\ &= \mathbb{P}\left(\max_{n_{k-1} < n < n_k} S_n(\lambda_n) > \frac{n_k}{9}\epsilon > 0 \text{ i.o. } (k)\right) \end{aligned} \quad (8)$$

Donc il suffit de montrer que pour tout $\epsilon > 0$

$$\mathbb{P}(\max_{n_{k-1} < n < n_k} S_n(\lambda_n) > n_k \epsilon > 0 \text{ i.o. } (k))$$

Observons que $\mathbb{E}(Z_i(\lambda_n)) = 0$ et

$$\mathbb{E}(Z_i(\lambda_n))^2 = \int f(y) \lambda_n^2 \tilde{K}(\lambda_n(x - y))^2 dy - (\tilde{f}_{\lambda_n}(x))^2 \leq \lambda_n \tilde{f}_{\lambda_n}(x) = O(\lambda_n)$$

Posons $\phi_n(t) = \mathbb{E}(e^{t(Z_1(\lambda_n) - \epsilon)})$ Alors comme λ_n est constante sur $[n_{k-1}, n_k]$ et donc $S_n(\lambda_n)$ est une martingale sur $[n_{k-1}, n_k]$ on peut appliquer l'inégalité des martingales

$$\mathbb{P}(\max_{n_{k-1} < n < n_k} S_n(\lambda_n) > n_k \epsilon > 0) \leq e^{-tn_k \epsilon} \mathbb{E}(e^{t(S_{n_k}(\lambda_{n_k}) - \epsilon)} = (\phi_{n_k}(t))^{n_k}$$

En particulier

$$\mathbb{P}(\max_{n_{k-1} < n < n_k} S_n(\lambda_n) > n_k \epsilon > 0) \leq \left(1 - \frac{\delta}{\lambda_{n_k}}\right)^{n_k}$$

pour un certain δ

Par conséquent en utilisant le lemme de Borel-Cantelli il suffit de prouver que la série des $a_k = \left(1 - \frac{\delta}{\lambda_{n_k}}\right)^{n_k}$ est sommable. Or $a_k < \exp(-\delta n_k / \lambda_{n_k})$ qui est sommable puisque $\lambda_{n_k} / n_k = o(1/\log k)$

Preuve du lemme 3 Le lemme 1 affirme que l'on peut trouver une suite (λ_n) qui satisfait les conditions du lemme 2 et qui vérifie

$$\lim_{n \rightarrow \infty} \sup_{\lambda \in \Lambda_n} \frac{\lambda}{\lambda_n} = 0 \quad (9)$$

Définissons maintenant pour $\lambda \in \Lambda_n$, $N_n(\lambda) = \inf\{n ; n\lambda \geq \lambda_n\}$ comme $\lambda_n \rightarrow \infty$ et par 9 on a :

$$\sup_{\lambda \in \Lambda_n} \left| \frac{N_n(\lambda)\lambda}{\lambda_n} - 1 \right| \rightarrow 0 .$$

Une conséquence est que

$$\sup_{\lambda \in \Lambda_n} \int |N_n(\lambda)\lambda\chi_{[0,1]}(N_n(\lambda)\lambda x) - \lambda_n\chi_{[0,1]}(\lambda_n x)| dx \rightarrow 0$$

et donc

$$\sup_{\lambda \in \Lambda_n} \int |\tilde{f}_{N_n(\lambda)\lambda,n}(x) - \tilde{f}_{\lambda_n,n}(x)| dx \rightarrow 0 \quad (10)$$

$$\sup_{\lambda \in \Lambda_n} \int |\tilde{f}_{N_n(\lambda)\lambda}(x) - \tilde{f}_{\lambda_n}(x)| dx \rightarrow 0 .$$

Notons que

$$\lambda\chi_{[0,1]}(\lambda x) = \frac{1}{N} \sum_{j=1}^N N\lambda\chi_{[0,1]}(N\lambda(x - \frac{j-1}{N\lambda}))$$

ce qui permet d'écrire

$$\tilde{f}_{\lambda,n}(x) = \frac{1}{N} \sum_{j=1}^N \tilde{f}_{N\lambda,n}(x - \frac{j-1}{N\lambda})$$

$$\tilde{f}_{\lambda}(x) = \frac{1}{N} \sum_{j=1}^N \tilde{f}_{N\lambda}(x - \frac{j-1}{N\lambda}) .$$

Maintenant on peut appliquer 10

$$\begin{aligned} \sup_{\lambda \in \Lambda_n} \int |\tilde{f}_{\lambda_n,n}(x) - \tilde{f}_{\lambda_n}(x)| dx &\leq \sup_{\lambda \in \Lambda_n} \frac{1}{N_n(\lambda)} \sum_{j=1}^{N_n(\lambda)} \int |\tilde{f}_{\lambda N_n(\lambda),n}(x - \frac{j-1}{N\lambda}) - \tilde{f}_{\lambda N_n(\lambda)}(x - \frac{j-1}{N\lambda})| dx \\ &= \sup_{\lambda \in \Lambda_n} \int |\tilde{f}_{\lambda N(\lambda),n}(x) - \tilde{f}_{\lambda N(\lambda)}(x)| dx \\ &\leq \sup_{\lambda \in \Lambda_n} \int |\tilde{f}_{\lambda_n,n}(x) - \tilde{f}_{\lambda_n}(x)| dx \\ &\quad + \sup_{\lambda \in \Lambda_n} \int |\tilde{f}_{\lambda N(\lambda),n}(x) - \tilde{f}_{\lambda_n,n}(x)| dx \\ &\quad + \sup_{\lambda \in \Lambda_n} \int |\tilde{f}_{\lambda N(\lambda),n}(x) - \tilde{f}_{\lambda N(\lambda)}(x)| dx \rightarrow 0 \end{aligned} \quad (11)$$

Preuve du lemme 4 Il faut passer de $\tilde{K}(x) = \chi_{[0,1]}(x)$ à $\tilde{K}(x) = \sum_{i=1}^m \alpha_i \chi_{[a_i, b_i]}$

On écrit alors

$$\hat{f}_\lambda(x) = \sum_{i=1}^m \alpha_i (b_i - a_i) \tilde{f}_{\lambda/(b_i - a_i)}(x - \frac{a_i}{\lambda})$$

La suite ne pose aucune difficulté. □

4 Estimateur à ondelettes

4.1 Motivation et rappels sur la décomposition de Fourier

Un des problèmes de l'estimation à noyau est son manque de souplesse. En effet le paramètre de bande étant fixé globalement on a du mal à estimer des densités qui présenteraient des parties lisses et des parties rapidement variables. Pour remédier à ce problème une idée est d'analyser la décomposition en fréquence de la densité. Sur \mathbb{R} on connaît bien la décomposition de Fourier et si l'on se restreint à un compact on a même une base hilbertienne dénombrable composée des fonctions vecteurs propres du laplacien. Pour estimer une densité, il suffit alors d'estimer le produit scalaire de cette densité avec les éléments de cette base. Or cela est très simple par l'égalité

$$\int f(x)\psi(x)dx = \mathbb{E}(\psi(X_1))$$

Ce que l'on estime naturellement par $\frac{1}{n} \sum_{i=1}^n \psi(X_i)$ puisque l'échantillon X_1, X_2, \dots, X_n est i.i.d.

Sur la sphère (qui est compacte) l'analogie de cette décomposition est donnée par les harmoniques sphériques qui sont également des vecteurs propres du laplacien.

Le laplacien (ou opérateur de Laplace-Beltrami) sur la sphère étant défini par : $\Delta f = \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} (\sin \theta \frac{\partial f}{\partial \theta}) + \frac{1}{\sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2}$ Les harmoniques sphériques usuelles sont notés $Y_{l,m}$ où l et m sont des entiers tels que $-l \leq m \leq l$ et $l \geq 0$ On les a choisies de telle sorte que la dépendance en θ et ϕ se factorise :

$$Y_{l,m}(\theta, \phi) = k P_{l,m}(\cos \theta) e^{im\phi}$$

où les $P_{l,m}$ sont des polynômes de degré l construits à partir des polynômes de Legendre et k est une constante de normalisation.

On peut donc comme dans le cas réel estimer le produit scalaire de la densité avec les harmoniques sphériques pour estimer cette densité. Le problème est que ces harmoniques sphériques sont à support dans toute la sphère, or on se rappelle que dans notre problème physique on n'observe pas certaines régions du ciel ce qui pose problème pour estimer le produit scalaire. L'idée est donc de trouver une autre décomposition avec des fonctions localisées en espace mais également en fréquence.

4.2 Décomposition en ondelettes et en needlets

On se place d'abord sur la droite réelle. Un exposé complet sur l'analyse à l'aide des ondelettes a été publié par Stéphane Mallat (voir [7]).

Une ondelette est une fonction ψ intégrable et de moyenne nulle (La démarche de Mallat présentée ici est de donner ces deux conditions minimales pour définir une décomposition en ondelettes, on peut imposer d'autres conditions selon les propriétés que l'on souhaite obtenir). A l'aide de cette ondelette-mère on définit une famille d'ondelettes à double indice $u \in \mathbb{R}, s \in \mathbb{R}_+^*$

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right)$$

et on a alors $\hat{\psi}_{u,s}(\omega) = e^{-iu\omega} \sqrt{s} \hat{\psi}(s\omega)$ si $\hat{\psi}$ représente la décomposition de Fourier : $\hat{\psi}(\omega) = \int \psi(t)e^{-it\omega} dt$

On définit alors pour toute fonction f la transformée en ondelettes à l'échelle s et la position u par

$$Wf(u, s) = \int_{\mathbb{R}} f(t) \frac{1}{\sqrt{s}} \psi^*\left(\frac{t-u}{s}\right) dt$$

Remarque. *L'égalité de Parseval permet alors d'écrire*

$$Wf(u, s) = \frac{1}{2\pi} \int \hat{f}(\omega) \hat{\psi}_{u,s}^*(\omega) d\omega$$

Exemple En prenant pour ψ la fonction de Haar ($\psi = \chi_{[0, \frac{1}{2}]} - \chi_{[\frac{1}{2}, 1]}$ voir fig 5), la famille $(\psi_{2^j, 2^j n} ; n, j \in \mathbb{N})$ forme une base hilbertienne de $L^2(\mathbb{R})$. Néanmoins dans de nombreux cas on n'obtient pas tout-à-fait une base mais une famille génératrice avec un peu de redondance

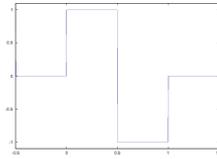


FIGURE 5 – fonction de Haar

Sur la sphère cette décomposition est un peu plus complexe, en effet on ne peut pas dilater aussi simplement une fonction de base comme on le fait dans le cas réel, de plus on n'a pas de notions de grille isotrope de points régulièrement espacés. On utilise des needlets dont la construction est due à Narcowich, Petrushev et Ward (voir [8]). Nous allons en présenter les grandes lignes mais sans démontrer les propriétés. Il faut d'abord construire un filtre en fréquence. Pour cela on utilise une fonction réelle C^∞ b à support sur $[\frac{1}{2}, 2]$ présentant une symétrie particulière en 1, plus exactement vérifiant $\forall \xi \in [\frac{1}{2}, 1] \quad b^2(\xi) + b^2(2\xi) = 1$.

Si \mathcal{H}_l représente l'espace vectoriel engendré par les $(Y_{l,m} ; m \in [-l, l])$ alors on peut écrire l'opérateur de projection sur cet espace de la façon suivante :

$$P_{\mathcal{H}_l} f(x) = \int_{\mathbb{S}^2} L_l(x \cdot y) f(y) dy$$

où les L_l sont les polynômes de Legendre qui vérifient la condition d'orthogonalité suivante

$$\int_{-1}^1 L_l(t) L_k(t) (1-t^2) dt = \frac{2l+1}{8\pi^2} \delta_{l,k} \quad .$$

Ils vérifient également la propriété $\int_{\mathbb{S}^2} L_l(x \cdot y) L_k(y \cdot z) dy = \delta_{l,k} L_l(x \cdot z)$

On définit ensuite l'opérateur $\Lambda_j = \sum_{l \geq 0} b^2(\frac{l}{2^j}) L_l$ et le noyau associé $\Lambda_j(x, y) = \sum_{l \geq 0} b^2(\frac{l}{2^j}) L_l(x \cdot y)$
On remarque que $\Lambda_j(x, x) = 1$

Proposition Pour tout $f \in L^2$

$$f = \lim_{J \rightarrow \infty} L_0(f) + \sum_{j=0}^J \Lambda_j(f)$$

De plus si $M_j(x, y) = \sum_{l \geq 0} b(\frac{l}{2^j}) L_l(x \cdot y)$

$$\Lambda_j(x, y) = \int M_j(x, z) M_j(z, y) dz$$

Si l'on veut pouvoir décomposer les fonctions selon une famille dénombrable, on a alors besoin d'un ensemble discret de points de la sphère sur lesquels pourront s'appuyer nos needlets (dans le cas de l'ondelette réelle de Haar par exemple cet ensemble était $\{n2^j ; n \in \mathbb{N}\}$, voir la section précédente). On utilise alors le résultat suivant énoncé dans [1]. Si \mathcal{P}_l désigne l'ensemble engendré par $(Y_{k,m} ; m \in [-k, k], k \leq l)$ alors il existe un ensemble fini $\mathcal{E}_l \subset \mathbb{S}^2$ de points et des réels positifs $\{\lambda_\eta, \eta \in \mathcal{E}_l\}$ tel que

$$\forall f \in \mathcal{P}_l \quad \int_{\mathbb{S}^2} f(x) dx = \sum_{\eta \in \mathcal{E}_l} \lambda_\eta f(\eta) \quad (12)$$

Cet ensemble n'est pas unique mais son cardinal doit être suffisamment élevé pour que la formule (appelée formule de quadrature) soit valable.

Alors la fonction $z \mapsto M_j(x, z)$ appartenant à $\mathcal{P}_{2^{j+1}}$ donc on a

$$\Lambda_j(x, y) = \sum_{\eta \in \mathcal{E}_{2^{j+2}}} \lambda_\eta M_j(x, \eta) M_j(\eta, y)$$

Ce qui implique

$$\Lambda_j f(x) = \int \Lambda_j(x, y) f(y) dy = \sum_{\eta \in \mathcal{E}_{2^{j+2}}} \sqrt{\lambda_\eta} M_j(x, \eta) \int \sqrt{\lambda_\eta} M_j(y, \eta) f(y) dy$$

On note alors $\psi_{j,\eta} = \sqrt{\lambda_\eta} M_j(x, \eta)$ pour $\eta \in \mathcal{E}_{2^{j+2}}$ On a alors

$$f = L_0(f) + \sum_j \sum_{\eta \in \mathcal{E}_{2^{j+2}}} (f \cdot \psi_{j,\eta})_{L^2} \psi_{j,\eta}$$

On a donc bien obtenu la décomposition souhaitée

Le principal résultat obtenu par Narcowich, Petrushev et Ward dans [8] concerne la localisation de ces needlets $\psi_{j,\eta}$: Pour tout k il existe une constante c_k telle que

$$|\psi_{j,\eta}(\xi)| \leq \frac{c_k 2^j}{(1 + 2^j d(\eta, \xi))^k} \quad (13)$$

où $d(\eta, \xi)$ est la distance riemannienne.

4.3 Seuillage et estimateur à needlets

Passons maintenant à la pratique, nous avons des observations X_1, X_2, \dots, X_n réparties selon une densité f . Le problème est toujours d'estimer f . Cette fois-ci nous la décomposons avec nos needlets

$$f = \frac{1}{4\pi} + \sum_{j \geq 0} \sum_{\eta} \beta_{j,\eta} \psi_{j,\eta} \cdot$$

Ici encore on commence par estimer les $\beta_{j,\eta}$

$$\hat{\beta}_{j,\eta} = \frac{1}{n} \sum_{i=1}^n \psi_{j,\eta}(X_i) .$$

Mais ensuite, au lieu de remplacer directement $\beta_{j,\eta}$ par $\hat{\beta}_{j,\eta}$ pour recomposer f , on applique un seuillage sur les coefficients de needlets :

$$\hat{f} = \frac{1}{4\pi} + \sum_{j=0}^{J(n)} \sum_{\eta} \hat{\beta}_{j,\eta} \chi_{\{|\hat{\beta}_{j,\eta}| \geq \kappa c_n\}} \psi_{j,\eta}$$

En fait intuitivement on juge que les coefficients needlets qui seraient trop faibles par rapport à leur écart-type sont dus à des fluctuations statistiques et donc on les ramène à 0 (cela est justifié théoriquement dans [5])

Il y a alors trois paramètres à définir

- la fréquence maximale $J(n)$ à laquelle on évalue les coefficients needlets, c'est une fonction croissante de n , en effet si on a trop peu de points il est clair que les hautes fréquences seront inaccessibles. Son choix exact résulte d'un compromis biais-variance car un $J(n)$ élevé augmente la variance et diminue le biais et réciproquement. On a en effet toujours un biais au moins égal à $\sum_{j>J(n)} \sum_{\eta} \beta_{j,\eta} \psi_{j,\eta}$
- le facteur d'échelle du seuillage c_n . Vu l'interprétation intuitive du seuillage on s'attend à ce qu'il varie comme l'écart-type des coefficients, (donc en $\sqrt{\frac{1}{n}}$ avec une densité uniforme par exemple). En fait des travaux théoriques suggèrent plutôt une vitesse en $\sqrt{\frac{\log n}{n}}$ (voir [1]).
- la constante de seuillage κ ; elle sera déterminée empiriquement à l'aide des simulations. Intuitivement on s'attend encore une fois à devoir trouver un compromis biais-variance; un seuillage plus fort augmentant le biais et diminuant la variance.

Exemple Prenons l'exemple de la densité uniforme pour comprendre les effets de la localisation des needlets. Nous pouvons par exemple calculer la covariance des différents coefficients

$$\begin{aligned} \mathbb{E}(\hat{\beta}_{j,\eta} \hat{\beta}_{j,\xi}) &= \int_{\mathbb{S}^2} \psi_{j,\eta}(x) \psi_{j,\xi}(x) dx \\ &= \sqrt{\lambda_\eta \lambda_\xi} \int M_j(x, \eta) M_j(x, \xi) dx \\ &= \sqrt{\lambda_\eta \lambda_\xi} \Lambda_j(\eta, \xi) \\ &= \sqrt{\lambda_\eta \lambda_\xi} \sum_{l \geq 0} b^2 \left(\frac{2}{2^j}\right) L_l(\eta \cdot \xi) \end{aligned} \tag{14}$$

On peut calculer cette quantité à l'aide de (13) et voir qu'elle décroît rapidement avec la distance entre η et ξ

Remarque. Bien que ces coefficients soit faiblement corrélés, ils ne sont pas du tout indépendants. En fait ils satisfont toujours la relation

$$\sum_{\eta \in \mathcal{E}_{2^j+2}} \sqrt{\lambda_\eta} \hat{\beta}_{j,\eta} = 0$$

En effet comme pour $l < 2^{j+1}$ $y \mapsto L_l(y \cdot x)$ est un polynôme de degré $\leq 2^{2j+2}$ on a

$$\sum_{\eta \in \mathcal{E}_{2j+2}} \lambda_\eta L_l(\eta \cdot x) = \int L_l(x \cdot y) dy = 0$$

et par conséquent

$$\sum_{\eta \in \mathcal{E}_{2j+2}} \sqrt{\lambda_\eta} \hat{\beta}_{j,\eta} = \frac{1}{n} \sum_{i=1}^n \sum_{l>0} b\left(\frac{l}{2^j}\right) \sum_{\eta \in \mathcal{E}_{2j+2}} \lambda_\eta L_l(\eta \cdot x) = 0$$

4.4 Comparaison noyaux-needlets

Il existe un résultat théorique sur la comparaison des deux estimateurs développé par Donoho, Johnstone, Kerkyacharian et Picard dans [5], ils affirment que l'estimateur à needlets avec seuillage est meilleur que celui à noyau selon un critère dit minimax, c'est-à-dire que l'on compare chaque estimateur dans le pire des cas. Cependant ces résultats ne sont pas uniformes ; pour une densité particulière on peut donc avoir un estimateur à noyau plus performant. Nous avons donc voulu comparer ces estimateurs pour quelques densités particulières.

Intuitivement on a vu que l'estimateur à noyau se comportait bien vis-à-vis des fonctions régulières mais avait plus de mal avec les discontinuités. En fait l'avantage des ondelettes est qu'elles permettent une résolution multi-échelles en analysant les données à différentes fréquences. Ce n'est pas le cas du noyau : en effet pour estimer une fonction constante par morceaux il faudrait une largeur de bande étroite pour estimer les discontinuités et une largeur de bande élevée pour estimer les régions où la densité est constante ; on doit donc se contenter d'un compromis qui n'est pas idéal. Un estimateur à noyau pourrait être intéressant si on avait des raisons physiques de penser que la densité soit continue mais ce n'est pas le cas pour le problème des rayons cosmiques.

5 Simulations

Cette dernière section présente les résultats obtenus lors des simulations ainsi que les algorithmes utilisés.

5.1 Particularités liées à l'algorithmique

L'implémentation numérique des estimateurs impose de définir les fonctions sur la sphère à l'aide de leurs valeurs en un nombre fini de points. C'est pourquoi on utilise un système de pixelisation —*Healpix*¹— qui découpe la sphère en pixels. Le nombre de pixels n_{pix} est modifiable à l'aide d'un paramètre n_{side} qui est une puissance de 2 ; on a alors $n_{pix} = 12(n_{side}^2)$ pixels. On peut alors définir des cartes qui sont des fonctions sur la sphère et dont la valeur est constante sur chaque pixel. On remarquera en fait que l'ensemble des centres de ces pixels forme un ensemble de quadrature évoqué dans la formule (12). La formule de quadrature est alors valable jusqu'à la fréquence $l_{max} = 2n_{side}$.

5.2 L'estimateur à noyau

Le codage de l'estimateur à noyau nous permet de vérifier un certain nombre de résultats énoncés dans ce rapport. Tout d'abord on peut regarder si la validation croisée donne bien la

1. <http://healpix.jpl.nasa.gov/>

largeur de bande optimale, ensuite vérifier que la décomposition de l'EQM en biais et variance est bien conforme à l'intuition, à savoir une variance croissante et un biais décroissant avec λ . Enfin on veut évaluer le risque L^2 et regarder ses variations en fonction du nombre de points n .

La première chose à faire est de simuler l'observation de points répartis aléatoirement selon une densité f . Nous avons étudié dans toutes les simulations trois densités différentes : la densité uniforme, une densité constante par morceaux (appelée dans la suite densité discontinue) et une densité lentement variable (appelée dans la suite densité lisse) obtenue en additionnant des harmoniques sphériques de basses fréquences. Pour simuler les données j'ai choisi d'exploiter le caractère discret des cartes et de répartir les points au hasard dans les pixels suivant une fonction de répartition calculée à l'aide de la densité. Le calcul de l'estimateur en fonction des données est ensuite très simple. J'ai choisi le noyau : $K(t) = (1 - t)\chi_{[0,1]}(t)$. Le calcul du λ optimal par validation croisée ne pose pas non plus de soucis une fois que l'on a codé l'estimateur.

La suite de la procédure consiste à évaluer numériquement les performances de l'estimateur. On utilise pour cela l'algorithme de Monte-Carlo. Il s'agit tout simplement de simuler les données et de faire tourner l'algorithme un grand nombre n_M de fois, de calculer l'erreur à chaque itération, puis d'estimer l'espérance par la moyenne empirique des données (le nombre n_M est appelé nombre de Monte-Carlo). Cette méthode permet également de calculer séparément le biais et la variance. Les résultats sont globalement conformes aux attentes (voir les figures 6, 7 et 8). On voit également que le λ choisi par validation croisée est proche de celui qui minimise l'erreur. Une remarque doit cependant être faite : pour le cas de la densité uniforme le biais n'est pas décroissant avec λ il tend même vers 0 lorsque λ tend vers 0 (figure 8). En effet lorsque λ tend vers 0 l'estimateur va lisser la densité de plus en plus. Quelle que soit la densité de départ l'estimateur va tendre vers la densité uniforme, donc dans le cas particulier où la densité est la densité uniforme, le biais devient nul.

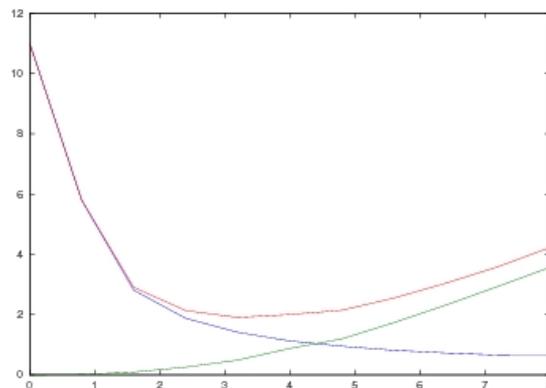


FIGURE 6 – erreur de l'estimateur à noyau en fonction de λ pour la densité discontinue et décomposition en biais et variance. La validation croisée donne $\lambda_{opt} = 2,8$

La dernière étape consiste à regarder la diminution du risque en fonction du nombre de points. Cela ne pose a priori pas de difficultés, cependant le problème du temps de calcul a commencé à se poser. Nous avons donc légèrement changé le principe du code. Au lieu de calculer directement le noyau nous avons effectué un calcul par Fourier. En effet l'estimateur peut s'écrire comme le produit de convolution de la carte des points par le noyau. Il suffit donc

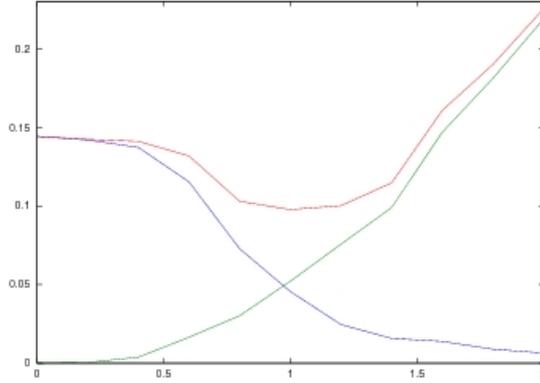


FIGURE 7 – erreur de l’estimateur à noyau en fonction de λ pour la densité lisse et décomposition en biais et variance. La validation croisée donne $\lambda_{opt} = 1$

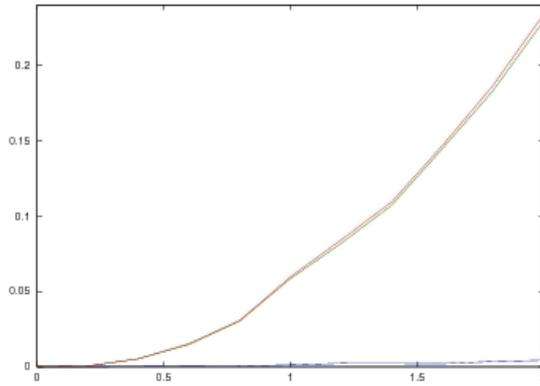


FIGURE 8 – erreur de l’estimateur à noyau en fonction de λ pour la densité uniforme et décomposition en biais et variance. La validation croisée donne $\lambda_{opt} = 0$

d’effectuer une multiplication dans la décomposition en harmoniques sphériques. Concrètement, le noyau étant à symétrie sphérique, seules les harmoniques sphériques avec $m = 0$ entrent dans la décomposition, le noyau peut ainsi s’écrire $L = \sum_{l \geq 0} a_l Y_{l,0}$, la carte des points se décompose elle aussi en harmoniques sphériques : $C = \sum_{\{l \geq 0, -l \leq m \leq l\}} b_{l,m} Y_{l,m}$. Alors la convolution du noyau et de cette carte (qui est l’estimée de la densité) s’exprime simplement en terme d’harmoniques sphériques par un résultat énoncé dans [1]

$$\tilde{f} = \sum_{\{l \geq 0, -l \leq m \leq l\}} a_l b_{l,m} Y_{l,m}$$

Cette procédure de convolution est avantageuse car les algorithmes de décomposition et recombinaison en harmoniques sphériques sont déjà codés au sein de Healpix de manière efficace avec un temps de calcul proportionnel à $n_{pix}^{3/2}$, alors que la procédure classique demande un temps en n_{pix}^2 .

On a pu alors voir que l’erreur suivait approximativement une loi de puissance en n (voir figure 9).

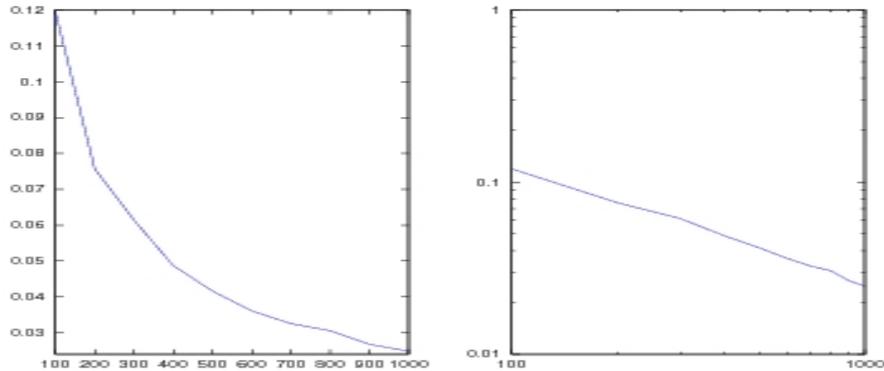


FIGURE 9 – EQM de l’estimateur à noyau pour la densité lisse en fonction du nombre de points avec λ choisi par validation croisée. La courbe de gauche est représenté en échelle classique et celle de droite en échelle logarithmique. On voit que cela suit une loi de puissance de coefficient $\alpha = 0,68$

5.3 L’estimateur à needlets

Pour coder l’estimateur à needlets il nous faut d’abord un filtre en fréquence b . Nous avons décidé de nous passer de l’hypothèse C^∞ dans un souci de simplicité et de gain de temps. En pratique cela donne des needlets moins localisées dans l’asymptotique des hautes fréquences. Nous avons donc pris des arches de cosinus. Ensuite pour calculer les estimés des coefficients needlets de la densité nous sommes encore passés par Fourier. Il faut bien comprendre ici qu’il ne s’agit que d’un intermédiaire de calcul qui permet de calculer facilement le produit scalaire de la densité avec les needlets. Mais nous cherchons bien les coefficients en needlets et non ceux en harmoniques sphériques.

L’autre difficulté concerne le choix du seuil dans l’estimateur. La littérature nous suggère une constante de seuillage qui dépend du nombre de points à analyser de manière fixée puis d’ajuster ensuite une constante multiplicative. Or ceci ne marche que si les needlets sont correctement normalisées. Nous n’avons pas vraiment réussi avec cette méthode à obtenir des résultats intéressants. Nous avons alors décidé de modifier quelque peu la procédure de seuillage. S’inspirant de l’interprétation intuitive qui dit que le seuillage doit être proportionnel à l’écart-type, nous avons pensé à un seuillage adaptatif qui calculerait par un algorithme de Monte-Carlo la variance des coefficients avant d’appliquer un seuillage adéquat. Cependant cette procédure est un peu longue à exécuter et n’est pas réalisable dans un problème concret où on ne connaît pas la densité. On s’est alors inspiré de l’équation (14) de la section 4.3 qui permet de calculer la variance des coefficients dans le cas de la densité uniforme. Celle-ci est alors égale à la norme L^2 de la needlet. Le programme calcule donc les normes des needlets pour pouvoir seuiller. Il reste toujours une constante de seuillage multiplicative κ qui est choisie empiriquement.

La dernière étape est alors de comparer les deux estimateurs. Dans la plupart des résultats la comparaison était moins nette que ce que l’on espérait et les needlets ne semblaient pas beaucoup plus performantes que le noyau. Il faut dire que les needlets présentent un défaut : le filtrage des hautes fréquences fait parfois apparaître des zones où la densité est estimée négative. Cela pose un problème pour des densités qui sont nulles sur une portion importante de la sphère. La dernière correction a donc été de ramener les densités estimées négatives à 0 puis de renormaliser

le tout. D'autres approches sont possibles pour éviter cet écueil comme par exemple estimer le logarithme de la densité puis en prendre l'exponentielle mais nous n'avons pas eu le temps ni de les étudier ni de les tester algorithmiquement.

paramètre	Noyaux								
	$\lambda = 1$			$\lambda = 2$			$\lambda = 4$		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
$n = 40$	0,12	10,1	0,17	0,54	7,69	0,58	2,33	6,15	2,34
$n = 100$	0,058	9,99	0,099	0,24	7,34	0,24	0,95	4,64	0,94
$n = 400$	0,014	9,96	0,063	0,060	7,21	0,055	0,24	4,04	0,24
biais	0	9,93	0,042	0	7,13	0,004	0	3,74	0,0003

paramètre	Needlets								
	$J = 3$			$J = 4$			$J = 5$		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
$n = 40$	0,24	8,27	0,29	1,05	6,18	1,10	5,94	11,9	6,15
$n = 100$	0,071	7,79	0,14	0,28	4,73	0,37	1,50	6,58	1,60
$n = 400$	0,008	7,49	0,054	0,030	4,28	0,090	0,15	2,89	0,24
biais	10^{-26}	7,41	10^{-9}	10^{-26}	3,96	10^{-9}	10^{-26}	1,73	10^{-9}

FIGURE 10 – Résultats des simulations : Risque quadratique des deux estimateurs en fonction du nombre de points et du paramètre ajustable. Les erreurs ont été estimées avec 50 réalisations de Monte-Carlo pour trois densités différentes : uniforme (a), discontinue(b) et lisse(c). En gras sont représentées les meilleures performances pour chaque couple de densité et de nombre de points donné. La dernière ligne représente le biais incompressible qui est obtenu en mesurant la distance quadratique entre la densité et celle lissée par le noyau ou celle projetée sur les ondelettes de basses fréquences. On voit que ce terme est dominant dans l'erreur de l'estimation de la densité discontinue

A Exemple d'estimations

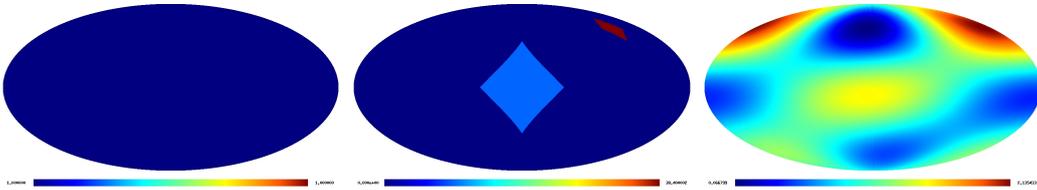


FIGURE 11 – Les trois densités : uniforme, discontinue et lisse

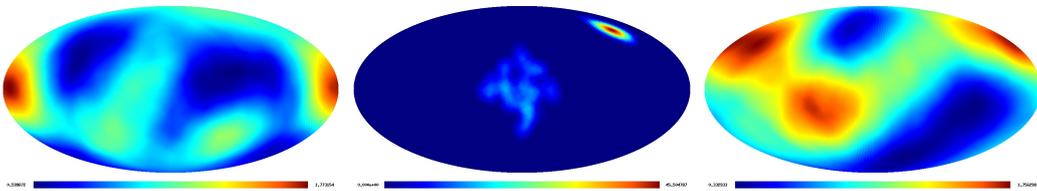


FIGURE 12 – Les estimées à noyaux des trois densités ; 100 points, $\lambda = 1$ pour les densités lisse et uniforme, $\lambda = 5$ pour la densité discontinue

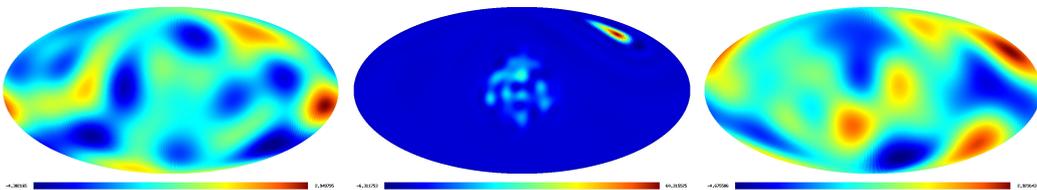


FIGURE 13 – Les estimées à needlets des trois densités ; 100 points, $J = 3$ pour les densités lisse et uniforme, $J = 5$ pour la densité discontinue

Références

- [1] P. Baldi, G. Kerkyacharian, D. Marinucci, and D. Picard. Adaptive density estimation for directional data using needlets. *The Annals of Statistics*, 37(6A), July 2009.
- [2] Y-S. Chow, S. German, and L-D. Wu. Consistent cross-validated density estimation. *The Annals of Statistics*, 11(1), 1983.
- [3] J. W. Cronin. Cosmic rays : the most energetic particles in the universe. *Reviews of Modern Physics Supplement*, 71 :165–+, March 1999.
- [4] James W. Cronin. The highest-energy cosmic rays. Center for Cosmological Physics Enrico Fermi Institute, University of Chicago, 2004.
- [5] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Density estimation by wavelet thresholding. *The Annals of Statistics*, 24(2), 1996.
- [6] J. Klemelä. Estimation of densities and derivatives of densities with directional data. *Journal of Multivariate Analysis*, 73, march 1997.
- [7] S. Mallat. *A Wavelet Tour of signal processing*. Academic Press, 2 edition, 1998.
- [8] F.J. Narcowich, P. Petrushev, and J. D. Ward. Localized tight frames on spheres. *SIAM Journal Mathematics Analysis*, 38(2), 2006.
- [9] A. V. Olinto. Ultra high energy cosmic rays : the theoretical challenge. *Physics Report*, 333 :329–348, August 2000.
- [10] Pierre AUGER Collaboration. Correlation of the highest-energy cosmic rays with the positions of nearby active galactic nuclei. *Astroparticle Physics*, 29 :188–204, August 2008.
- [11] E. M. Santos and f. t. P. Auger Colaboration. Anisotropy Studies with the Pierre Auger Observatory. *ArXiv e-prints*, January 2009.