

Filières MP/MPI - ENS de Paris-Saclay, Lyon, Rennes et Paris - Session 2024
Page de garde du rapport de TIPE

NOM : DUFFAUD	Prénoms : Martin
Classe :	
Lycée :	Numéro de candidat : 10736
Ville :	

Concours auxquels vous êtes admissible, dans la banque MP inter-ENS (les indiquer par une croix) :

ENS Cachan	MP - Option MP	<input checked="" type="checkbox"/>	MP - Option MPI	<input type="checkbox"/>
	Informatique	<input type="checkbox"/>		
ENS Lyon	MP - Option MP	<input checked="" type="checkbox"/>	MP - Option MPI	<input type="checkbox"/>
	Informatique - Option M	<input type="checkbox"/>	Informatique - Option P	<input type="checkbox"/>
ENS Rennes	MP - Option MP	<input checked="" type="checkbox"/>	MP - Option MPI	<input type="checkbox"/>
	Informatique	<input type="checkbox"/>		
ENS Paris	MP - Option MP	<input type="checkbox"/>	MP - Option MPI	<input type="checkbox"/>
	Informatique	<input type="checkbox"/>		

Matière dominante du TIPE (la sélectionner d'une croix inscrite dans la case correspondante) :

Informatique	<input type="checkbox"/>	Mathématiques	<input checked="" type="checkbox"/>	Physique	<input type="checkbox"/>
--------------	--------------------------	---------------	-------------------------------------	----------	--------------------------

Titre du TIPE :

Débruitage et décomposition en ondelettes, Transformée de Box-Muller

Nombre de pages (à indiquer dans les cases ci-dessous) :

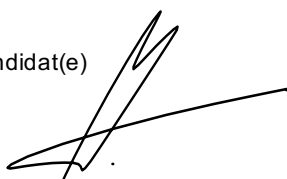
Texte	5	Illustration	1	Bibliographie	1
-------	---	--------------	---	---------------	---

Attention, les illustrations doivent figurer dans le corps du texte et non en fin du document !

Résumé ou descriptif succinct du TIPE (6 lignes, maximum) :

Ce travail tente d'apporter une approche théorique au problème du contrôle du bruit en théorie du signal, par l'intermédiaire d'outils empruntés à la théorie de la mesure et des probabilités. Après une rapide introduction sur les ondelettes, on présentera l'approche classique de Donoho et Johnstone, pour enfin s'interroger sur la distribution du bruit observé grâce à un test de normalité. La notion de vecteur gaussien et de loi normale sera au cœur de l'étude.

À	Signature du professeur responsable de la classe préparatoire dans la discipline	Cachet de l'établissement
Le		
Signature du (de la) candidat(e)		



Débruitage et décomposition en ondelettes, Transformée de Box-Muller

TIPE 2024

Table des matières

1. INTRODUCTION.....	0
2. LES ONDELETTES DE HAAR.....	1
2.1. GENERALITES	1
2.2. AVANTAGES DE LA TRANSFORMEE EN ONDELETTES DE HAAR	1
3. ALGORITHME DE SEUILLAGE DE DONOHO ET JOHNSTONE	2
3.1. ÉNONCE DU PROBLEME	2
3.2. CALCUL PRATIQUE DE L'ECART-TYPE : LA DEMARCHE DE STEPHANE MALLAT	2
4. TEST DE NORMALITE DU BRUIT DETECTE DANS UN SIGNAL	3
4.1. TRANSFORMEE DE BOX-MULLER	3
4.2. APPLICATION.....	5
5. CONCLUSION.....	5
BIBLIOGRAPHIE	5
ANNEXE 1 APPROCHE NUMERIQUE, TEST DE NORMALITE.....	6
CALCUL DE LA MOYENNE.....	6
REPARTITION.....	6

1. Introduction

Le débruitage est une méthode qui permet d'éliminer l'erreur – appelée *bruit* – qui apparaît lorsqu'un signal subit une dégradation suite à une conversion (par exemple, lors de la transformation d'un signal sonore en un signal électrique par un microphone) ou une transmission de mauvaise qualité (comme des variations brusques de courants dans les circuits imprimés). Tout d'abord, nous présenterons les ondelettes de Haar et certaines de leurs propriétés, afin de justifier leur utilisation dans une version simplifiée de la démarche de Donoho et Johnstone : à partir de la décomposition en ondelettes d'un signal, il est possible de réaliser un débruitage de celui-ci. Il s'agira, à travers une application pratique sur un signal bruité, de réaliser un test de normalité du bruit, hypothèse majeure de cette théorie. On aura recours à la transformée de Box-Muller, qui permet de générer numériquement des échantillons gaussiens grâce à des variables aléatoires uniformes, et qu'il s'agira de démontrer.

2. Les ondelettes de Haar

2.1. Généralités

Comme toutes les familles d'ondelettes, les *ondelettes de Haar* sont définies à partir d'une ondelette *primitive*, appelée ondelette *mère* de Haar définie comme suit :

$$\psi = \mathbf{1}_{[0, \frac{1}{2}[} - \mathbf{1}_{[\frac{1}{2}, 1[}$$

Puis on définit les ondelettes filles de Haar par translation et homothétie :

$$\forall j \in \mathbb{N}, \forall k \in \llbracket 0, 2^j - 1 \rrbracket, \psi_{j,k} : x \mapsto \psi(2^j x - k)$$

On définit par abus les ondelettes fils de Haar à partir de l'*ondelette père* $\phi = \mathbf{1}_{[0,1[}$, toujours par homothétie et fenêtrage, avec les mêmes conventions. On notera E_n le \mathbb{R} -espace vectoriel des fonctions en escalier s pour lesquelles il existe une subdivision régulière sur $[0,1]$ de taille 2^n , subordonnée à s , où n est un entier naturel. Avec ces conventions, on vérifie que :

$$\forall s \in E_n, \quad s = \sum_{k=0}^{2^n-1} s\left(\frac{k}{2^n}\right) \phi_{n,k}$$

Ainsi, les ondelettes fils $(\phi_{n,k})_{k \in \llbracket 0, 2^n-1 \rrbracket}$ forment une famille génératrice de E_n , et même une base, leurs supports étant deux à deux disjoints. Dans la pratique, on aura toujours $n \geq 2$.

2.2. Avantages de la transformée en ondelettes de Haar

L'objectif de cette partie est de présenter la démarche qui permet d'écrire E_n comme somme directe de sous espaces engendrés par des ondelettes filles. On vérifie que, si $n > 0$, $(\phi_{n-1,k})_{k \in \llbracket 0, 2^{n-1}-1 \rrbracket}$ et $(\psi_{n-1,k})_{k \in \llbracket 0, 2^{n-1}-1 \rrbracket}$ sont deux familles libres de E_n car les $\phi_{n-1,k}$ (respectivement les $\psi_{n-1,k}$) sont à supports deux à deux disjoints, lorsque k varie. De plus, on vérifie que la réunion de ces deux familles forme une famille orthogonale de vecteurs de E_n , pour le produit scalaire intégral sur $[0,1]$. Cette réunion constitue donc une famille libre maximale de E_n , on a ainsi :

$$E_n = E_{n-1} \oplus F_{n-1}$$

où F_{n-1} est l'espace engendré par la famille $(\psi_{n-1,k})_{k \in \llbracket 0, 2^{n-1}-1 \rrbracket}$. Par itérations successives, on a finalement :

$$E_n = E_{n-1} \oplus F_{n-1} = (E_{n-2} \oplus F_{n-2}) \oplus F_{n-1} = \dots = E_0 \oplus \left(\bigoplus_{p=0}^{n-1} F_p\right)$$

ce qui constitue l'objectif recherché, à savoir pouvoir décomposer tout signal de E_n en fonction des ondelettes filles de Haar, à plusieurs échelles. On dira que l'échelle $n - 1$ est l'échelle la plus fine, l'échelle 0 est la plus grossière, la valeur moyenne de s est sa coordonnée dans E_0 :

$$\forall s \in E_n, \quad s = \sum_{p=0}^{n-1} \left(\sum_{k=0}^{2^p-1} \frac{(s|\psi_{p,k})}{\|\psi_{p,k}\|^2} \psi_{p,k} \right) + (s|\phi)\phi$$

Cette première étape est systématique quelle que soit la famille d'ondelettes choisie. Cependant, les ondelettes de Haar présentent l'avantage, par leur simplicité, de pouvoir réduire considérablement le temps de calcul des coefficients d'ondelettes. En effet, si $p \in \llbracket 1, n \rrbracket$, la donnée de la projection orthogonale de s sur E_p , que l'on notera s_p , permet de calculer les coefficients d'ondelettes à l'échelle $p - 1$:

$$\forall k \in \llbracket 0, 2^{p-1} - 1 \rrbracket, \quad (s|\psi_{p-1,k}) = (s_p|\psi_{p-1,k}) = \frac{s_p\left(\frac{k}{2^{p-1}}\right) - s_p\left(\frac{k}{2^{p-1}} + \frac{1}{2^p}\right)}{2^p}$$

Et on détermine également très aisément s_{p-1} , qui n'est autre que la projection orthogonale de s_p sur E_{p-1} :

$$s_{p-1} = \sum_{k=0}^{2^{p-1}-1} \frac{(s_p|\phi_{p-1,k})}{\|\phi_{p-1,k}\|^2} \phi_{p-1,k} = \sum_{k=0}^{2^{p-1}-1} \left(\frac{s_p\left(\frac{k}{2^{p-1}}\right) + s_p\left(\frac{k}{2^{p-1}} + \frac{1}{2^p}\right)}{2^p} \right) \phi_{p-1,k}$$

3. Algorithme de seuillage de Donoho et Johnstone

3.1. Énoncé du problème

Supposons donné un signal s dans E_n bruité, *id est* de la forme $s = \lambda + W$, où W est le bruit, et λ le signal original, que l'on cherche à déterminer. Donoho et Johnstone montrent que, sous l'hypothèse où W est un bruit blanc gaussien d'écart-type σ , alors une bonne approximation de λ est donnée par application de la fonction de seuillage θ définie sur les coefficients d'ondelettes de s par :

$$\left(\left((s|\psi_{j,k}) \right)_{k=0}^{2^j-1} \right)_{j=0}^n \xrightarrow{\theta} \left(\left(\theta_{j,k}(s|\psi_{j,k}) \right)_{k=0}^{2^j-1} \right)_{j=0}^n$$

où les $\theta_{j,k}$ sont définis par la relation :

$$\forall j \in \llbracket 0, n \rrbracket, \forall k \in \llbracket 0, 2^j - 1 \rrbracket, \theta_{j,k} = \mathbf{1}_{|(s|\psi_{j,k})| > T}$$

et T , le seuil de Donoho et Johnstone, vaut $\sigma\sqrt{2\log(n)}$. Le choix de T fait par Donoho et Johnstone s'appuie sur le fait que, pour cette méthode de seuillage, T doit nécessairement vérifier :

$$T > \max_{j,k} |(W|\psi_{j,k})|$$

afin de traiter le cas où $\lambda = 0$ et $s = W$. Ainsi, pour le choix d'un seuil optimal, le choix de la valeur $\sigma\sqrt{2\log(n)}$ vient du résultat suivant que l'on admettra, montré par Donoho et Johnstone, et qui donne *un intervalle de confiance* pour le choix d'une valeur seuil cohérente :

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\sigma\sqrt{2\log(n)} - \frac{\sigma \log(\log(n))}{\log(n)} < \max_{j,k} |(W|\psi_{j,k})| < \sigma\sqrt{2\log(n)} \right) = 1 \quad (1)$$

Ainsi, l'algorithme de seuillage de Donoho et Johnstone, par une sélection de chaque coefficient d'ondelettes de s , donne une approximation des coefficients de λ , que l'on peut alors reconstruire par transformée inverse en ondelette.

3.2. Calcul pratique de l'écart-type : la démarche de Stéphane Mallat

A priori, le bruit W , que l'on suppose gaussien, est inconnu. On ne connaît pas son écart-type σ . Néanmoins, Stéphane Mallat explique que, pour un signal assez régulier (dans le sens où elle ne possède pas de sauts importants), alors la valeur des coefficients d'ondelettes de λ sur F_{n-1} (l'échelle la plus fine, qui mesure justement les aspérités du signal) sont négligeables devant ceux de W . Ainsi, on a l'approximation :

$$\forall k \in \llbracket 0, 2^{n-1} - 1 \rrbracket, (s|\psi_{n-1,k}) = (W|\psi_{n-1,k}) + (\lambda|\psi_{n-1,k}) \cong (W|\psi_{n-1,k})$$

En vertu du fait que W est un vecteur gaussien, et que les $\frac{\psi_{n-1,k}}{\sqrt{2^{n-1}}}$ forment une base orthonormale de vecteurs, les $\left(W \left| \frac{\psi_{n-1,k}}{\sqrt{2^{n-1}}} \right. \right)$ sont, des variables aléatoires gaussiennes, indépendantes (car les $\psi_{n-1,k}$ sont à supports disjoints) identiquement distribués, et toujours d'écart-type σ , que l'on cherche à déterminer. Or, si $(W_i)_{i=1}^d$ est une famille de $d \in \mathbb{N}$ variables aléatoires gaussiennes de moyennes nulles et indépendantes, leur écart type commun σ peut-être approximé selon la relation :

$$\sigma \cong \frac{\text{Med}_{i \in \llbracket 1, d \rrbracket} (|W_i|)}{0.6745} \quad (2)$$

On déduit de ces deux approximations une estimation de σ , que l'on notera $\tilde{\sigma}$:

$$\tilde{\sigma} := \frac{\text{Med}_{k \in \llbracket 0, 2^{n-1} - 1 \rrbracket} \left(\left| \left(s \left| \frac{\psi_{n-1,k}}{\sqrt{2^{n-1}}} \right. \right) \right| \right)}{0.6745} \quad (3)$$

Dès lors, la mise en place de l'algorithme de Donoho et Johnstone, est possible, dans la mesure où une valeur approchée du seuil T peut-être obtenue grâce à la démarche de Stéphane Mallat, que l'on admettra également.

4. Test de normalité du bruit détecté dans un signal

Étant donné un signal WAV, que l'on suppose bruité par un bruit blanc gaussien W d'écart-type σ , on détermine $\tilde{\sigma}$ grâce à la relation (3) et on en déduit une valeur seuil $T = \tilde{\sigma}\sqrt{2\log(n)}$. Ainsi, lors de la mise en place de l'algorithme de Donoho et Johnstone, on va conserver uniquement les coefficients d'ondelettes éliminés, c'est-à-dire ceux qui caractérisent le bruit \tilde{W} détecté par l'algorithme, afin de réaliser un test de normalité de \tilde{W} . Ainsi, on pourra conclure concernant la justesse de l'hypothèse de normalité de $W \cong \tilde{W}$.

4.1. Transformée de Box-Muller

Le test de normalité du bruit \tilde{W} consiste en deux étapes : le calcul de sa moyenne et l'étude de l'histogramme de répartition de ses valeurs. Cependant, pour pouvoir dire que « \tilde{W} est de moyenne quasiment nulle » ou encore que l'histogramme de répartition des valeurs de \tilde{W} « ressemble » à une gaussienne d'écart-type $\tilde{\sigma}$, on va directement comparer sa moyenne et son histogramme de répartition avec ceux d'un bruit de référence $W_r = (W_{r,k})_{k=0}^{2^n-1}$ d'écart-type $\tilde{\sigma}$ généré à partir de deux échantillons de loi uniforme, indépendants et identiquement distribués. Une telle construction est donnée par George Edward Pelham Box et Mervin Edgar Muller, en 1958.

L'objectif de cette partie est la démonstration du théorème ci-dessous. Pour ce faire, on va faire un détour par la théorie de la mesure, en s'appuyant notamment sur les théorèmes de Beppo-Levi et de changement de variables dans \mathbb{R}^2 .

Théorème (Box-Muller) : Soient U et V deux variables aléatoires indépendantes identiquement distribuées de loi uniforme sur $[0,1]$ et σ un réel strictement positif. Alors $W := \sigma\sqrt{-2\ln(U)}\cos(2\pi V)$ suit une loi normale centrée d'écart-type σ .

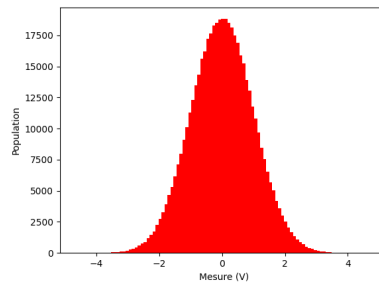


Figure 1 : Histogramme de répartition d'un échantillon gaussien (loi centrée réduite) obtenu grâce à une transformée de Box-Muller

Comme les résultats préliminaires vont être appliqués dans \mathbb{R} (respectivement \mathbb{R}^2) muni de la tribu des boréliens $\mathcal{B}(\mathbb{R})$ (resp. $\mathcal{B}(\mathbb{R}^2)$), on se contentera de manipuler la mesure de Lebesgue λ (resp. $\lambda_2 := \lambda \otimes \lambda$), que l'on ne reconstruira pas, et on rappelle que les boréliens de \mathbb{R}^d ($d \geq 1$) sont engendrés par les pavés $\prod_{i=1}^d]a_i, b_i[$, où les $a_i \leq b_i$ sont réels. Plus précisément, il s'agit de l'ensemble des parties de \mathbb{R}^d qui s'écrivent comme réunion et intersection (éventuellement vide) de tels ensembles ou de leur complémentaires.

Définition : On appelle fonction étagée (positive) toute application de \mathbb{R}^2 dans \mathbb{R} de la forme $\sum_{i=1}^n \alpha_i \mathbf{1}_{B_i}$ où n est un entier, les α_i des réels positifs et B_i des éléments de $\mathcal{B}(\mathbb{R}^2)$. On note \mathcal{E} l'ensemble des fonctions étagées. Si $h = \sum_{i=1}^n \alpha_i \mathbf{1}_{B_i}$ est dans \mathcal{E} , on définit son intégrale par $\int_{\mathbb{R}^2} h d\lambda_2 := \sum_{i=1}^n \alpha_i \lambda_2(B_i)$.

Définition : On dit d'une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ qu'elle est mesurable si pour tout B dans $\mathcal{B}(\mathbb{R})$, $f^{-1}(B)$ est dans $\mathcal{B}(\mathbb{R}^2)$. Par exemple, les fonctions continues et les fonctions étagées sont mesurables.

Proposition : Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction mesurable positive. Il existe une suite $(f_n)_{n \in \mathbb{N}^*}$ de fonctions étagées positives qui converge simplement vers f .

Preuve. Il s'agit de poser, pour $n \in \mathbb{N}^*$ et $i \in \llbracket 0, n2^n - 1 \rrbracket$,

$$A_n := \{x \in \mathbb{R}^2 \mid f(x) \geq n\} \text{ et } B_{n,i} := \{x \in \mathbb{R}^2 \mid i2^{-n} \leq f(x) < (i+1)2^{-n}\}$$

Puis, on vérifie que la suite de fonctions $(f_n)_{n \in \mathbb{N}}$ définie pour $n \geq 1$ comme suit convient.

$$f_n := \sum_{i=0}^{n2^n-1} \frac{i}{2^n} \mathbf{1}_{B_{n,i}} + n \mathbf{1}_{A_n}$$

■

Cette construction permet d'étendre naturellement la notion d'intégrale aux fonctions mesurables positives :

Définition : Soit f une fonction mesurable positive. On définit l'intégrale de f par $\int_{\mathbb{R}^2} f d\lambda := \sup_{h \in \mathcal{E} \mid h \leq f} \int_{\mathbb{R}^2} h d\lambda$.

En particulier, on peut généraliser cette définition aux fonctions mesurables (en regardant leur partie positive et négative, qui sont mesurables positives) et on vérifie facilement les propriétés usuelles de l'intégrale (linéarité, croissance, etc.).

Théorème (Beppo-Lévi) : Soit f une fonction mesurable positive et $(f_n)_{n \in \mathbb{N}}$ une suite croissante de fonctions mesurables qui converge simplement vers f . Alors, pour tout borélien A , on a :

$$\lim_{n \rightarrow +\infty} \int_A f_n(x) dx = \int_A f(x) dx$$

Preuve. On montre le résultat par double inégalité.

Par croissance de l'intégrale, $\int_A f_n(x) dx \leq \int_A f(x) dx$ pour $n \in \mathbb{N}$. Ceci fournit également l'existence d'une limite par croissance de $(f_n)_{n \in \mathbb{N}}$.

Montrons l'inégalité réciproque en revenant à la définition de $\int_A f(x) dx$. Soit h une fonction étagée plus petite que f et a un réel compris dans $[0,1[$. Posons, pour $n \in \mathbb{N}$:

$$C_n := \{x \in A \mid ah(x) \leq f_n(x)\}$$

Comme $a < 1$, $ah < f$, on a $\bigcup_{n \geq 0} C_n = A$ par convergence simple de la suite des f_n vers f .

Soit $n \in \mathbb{N}$. $C_n = (f_n - ah)^{-1}(\mathbb{R}_+)$ est mesurable comme image réciproque d'un borélien par une fonction mesurable. Ainsi, on peut intégrer sur C_n et on trouve par croissance de l'intégrale :

$$a \int_A h(x) \mathbf{1}_{C_n} dx \leq \int_A f_n(x) \mathbf{1}_{C_n} dx \leq \int_A f_n(x) dx \leq \lim_{n \rightarrow +\infty} \int_A f_n(x) dx$$

Or, h est étagée, donc $h \mathbf{1}_{C_n}$ aussi. En mettant h sous sa forme canonique $\sum_{i=1}^n \alpha_i \mathbf{1}_{B_i}$ on conclut alors par croissance de la suite $(C_n)_{n \in \mathbb{N}}$ (les f_n forment une suite croissante), du fait qu'alors $\lim_{n \rightarrow +\infty} \lambda(B_i \cap C_n) = \lambda(B_i)$ (par continuité croissante des mesures, vue en classe préparatoire pour les mesures de probabilités), puis par linéarité de la limite, s'agissant d'une somme finie.

$$a \int_A h(x) dx \leq \lim_{n \rightarrow +\infty} \int_A f_n(x) dx$$

En faisant tendre a vers 1 et par passage au sup, on obtient le résultat voulu. ■

Enfin, voici un dernier résultat, que l'on admettra dans ce rapport, qui étend la notion de changement de variables vue en classe préparatoires pour les fonctions réelles d'une variable réelle :

Théorème (Changement de variable) : Soient U et V deux boréliens de \mathbb{R}^2 et φ un C^1 -difféomorphisme de U dans V . Alors, pour $f : U \rightarrow \mathbb{R}$ mesurable, on a, en notant $J_{\varphi^{-1}}(v)$ la Jacobienne de φ^{-1} en $v \in V$:

$$\int_U f(u) du = \int_V f(\varphi^{-1}(v)) \left| \det \left(J_{\varphi^{-1}}(v) \right) \right| dv$$

Le théorème de changement de variable se montre en utilisant les résultats démontrés précédemment : étant donnée une fonction mesurable f , quitte à écrire f comme différence de sa partie positive et négative, qui sont toutes deux mesurables positives, et en utilisant la linéarité de l'intégrale, on peut supposer que f est positive. Il suffit alors de démontrer le théorème de changement de variable pour une fonction étagée, et donc par linéarité pour une fonction indicatrice d'un borélien de \mathbb{R}^2 . Cette démarche est assez habituelle en théorie de la mesure, et est par exemple utilisée pour démontrer le théorème de transfert. Passons à la preuve du Théorème de Box-Muller :

Preuve du Théorème de Box-Muller : Posons $X := \sigma \sqrt{-2 \ln(U)} \cos(2\pi V)$ et $Y := \sigma \sqrt{-2 \ln(U)} \sin(2\pi V)$. Comme U et V sont des variables aléatoires, X et Y sont également des variables aléatoires comme composée de (U, V) avec une fonction continue donc mesurable. En particulier, si on note $R := \sigma \sqrt{-2 \ln(U)}$, alors comme U suit une loi uniforme sur $[0,1]$, on a, pour tout segment $]a, b[$ inclus dans \mathbb{R}_+ (R est positive p.s. c'est-à-dire $\mathbb{P}(R > 0) = 1$) :

$$\mathbb{P}(R \in]a, b]) = \mathbb{P} \left(U \in \left[e^{-\frac{b^2}{2\sigma^2}}, e^{-\frac{a^2}{2\sigma^2}} \right] \right) = e^{-\frac{a^2}{2\sigma^2}} - e^{-\frac{b^2}{2\sigma^2}} = \int_a^b \frac{t}{\sigma^2} e^{-\frac{t^2}{2\sigma^2}} dt$$

Donc admet $t \mapsto \frac{t}{\sigma^2} e^{-\frac{t^2}{2\sigma^2}} \mathbf{1}_{\mathbb{R}_+}(t)$ comme densité par rapport à la mesure de Lebesgue. De même, si on note $\Phi := 2\pi V$, on montre que Φ admet la fonction $t \mapsto \frac{1}{2\pi} \mathbf{1}_{[0,2\pi[}$ comme densité par rapport à la mesure de Lebesgue. Comme U et V

sont indépendantes, le lemme des coalitions assure que R et Φ sont indépendantes. En termes de densités, cela se traduit par le fait que (R, Φ) possède une densité par rapport à la mesure de Lebesgue qui est par simple produit la fonction :

$$f : (r, \theta) \mapsto \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \mathbf{1}_{[0, 2\pi[\times \mathbb{R}_+}(r, \theta)$$

Procédons au changement de variable, en posant $\varphi : (r, \theta) \mapsto (r\cos(\theta), r\sin(\theta))$ définie de $[0, 2\pi[\times \mathbb{R}_+$ dans \mathbb{R}^2 .

φ est un \mathcal{C}^1 -difféomorphisme et $|\det(J_{\varphi^{-1}}(x, y))| = \frac{1}{\sqrt{x^2 + y^2}}$ pour (x, y) réels. Donc (X, Y) admet une densité par rapport à la mesure de Lebesgue, donnée par

$$g : (x, y) \mapsto f(\varphi^{-1}(x, y)) |\det(J_{\varphi^{-1}}(x, y))| = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}}$$

En intégrant par rapport à y , on retrouve la densité de X , qui est alors caractéristique d'une loi normale. ■

4.2. Application

Ainsi, si $(U_k)_{k=0}^{2^n-1}$ et $(V_k)_{k=0}^{2^n-1}$ sont des séquences de variables aléatoires indépendantes, identiquement distribuées, suivant une loi uniforme sur $[0, 1]$. La séquence $W_r = (W_{r,k})_{k=0}^{2^n-1}$ de variables aléatoires définie par :

$$\forall k \in \llbracket 0, 2^n - 1 \rrbracket \quad W_{r,k} := \tilde{\sigma} \sqrt{-\ln(U_k)} \cos(2\pi V_k) \quad (4)$$

suit une loi normale centrée d'écart-type $\tilde{\sigma}$.

Dès lors, on pourra dire que le test de normalité de \tilde{W} est un succès lorsque sa moyenne est du même ordre de grandeur que celle du bruit de référence (qui est de moyenne presque nulle) et que les histogrammes de répartition des deux bruits se transposent (celui du bruit de référence étant une gaussienne d'écart-type $\tilde{\sigma}$). L'annexe 1 donne une description des résultats obtenus, dont un bref aperçu est donné en conclusion.

5. Conclusion

Par ce travail, il s'agissait de confronter une approche théorique du traitement du signal, par l'intermédiaire de la théorie des ondelettes, et s'appuyant sur des résultats aux hypothèses fortes, avec une application pratique, pour laquelle on obtient des résultats aux enseignements notables. D'une part, l'hypothèse du bruit blanc gaussien n'est pas illégitime, lorsque l'on compare la moyenne et la répartition d'un bruit de référence, gaussien et de même écart-type, obtenu par application de la transformée de Box-Muller, avec celui que l'algorithme de seuillage de Donoho et Johnstone, pour un signal donné, a permis d'identifier. Toutefois, même si l'intervalle de confiance donné par Donoho et Johnstone par le choix du seuil T permet d'imaginer, pour un signal de taille encore plus grande, des résultats encore plus convaincants et précis, ces résultats sont à relativiser, dans la mesure où la détection du bruit est un problème difficile. En effet, dans ce cadre, il a été utilisé pour un signal audio, vérifiant ainsi l'hypothèse de régularité de Mallat, et qui permet d'estimer l'écart-type du bruit, supposé gaussien. De fait, pour des signaux moins réguliers, le bruit serait alors plus difficile à détecter au sein des disparités déjà importantes du signal de départ.

Bibliographie

- [1] S. MALLAT : *A Wavelet Tour of Signal Processing, The Sparse Way : 3rd Edition* (2008)
- [2] C. CHARLES : *Introduction aux applications des ondelettes*. Unité de Statistique, Informatique et Mathématiques Appliquées (SIMa). Université de Liège (2011)
- [3] J.F. LE GALL : *Intégration, Probabilités et Processus Aléatoires*. Département de Mathématiques et Applications. École Normale Supérieure de Paris (2006).
- [4] J. GOODMAN : *Simple sampling of Gaussians*. Courant Institute of Mathematical Sciences, New York University (2005).
- [5] D. DONOHO, I. JOHNSTONE : *Ideal spatial adaptation by wavelet shrinkage. Biometrika*. Department of Statistics, Stanford University (1994).

Annexe 1 Approche numérique, Test de normalité

Calcul de la moyenne

Par un simple calcul de l'intégrale de W_r et \tilde{W} , on en déduit leur moyenne sur leur intervalle de définition commun. On trouve qu'elles sont du même ordre (10^{-2}) :

	\tilde{W}	W_r
Moyenne	-0.0046	0.0039 ≈ 0

Figure 2 : Comparaison de la moyenne du bruit détecté et du bruit de référence

Répartition

En regardant l'histogramme de la variable \tilde{W} , on s'aperçoit rapidement de la présence d'une anomalie sur la distribution du bruit, avec un nombre important de valeurs nulles. Celle-ci s'explique, comme le précise Stéphane Mallat, par le fait que la fonction de seuillage de Donoho et Johnstone favorisent l'élimination des coefficients d'ondelettes presque nuls, assimilés incessamment au bruit.

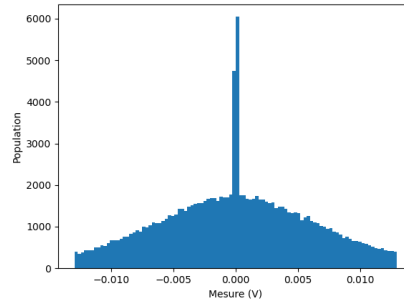


Figure 3 : Histogramme de la variable \tilde{W}

Ainsi, une première étape consiste à éliminer aléatoirement une partie des valeurs presque nulles de la variable \tilde{W} . Puis, on reconstruit avec la formule de Rice le bruit de référence W_r , mais cette fois, avec un nombre de valeurs moins important, afin qu'il soit égale au nombre de valeurs conservées pour la variable \tilde{W} . On met alors en évidence le résultat souhaité :

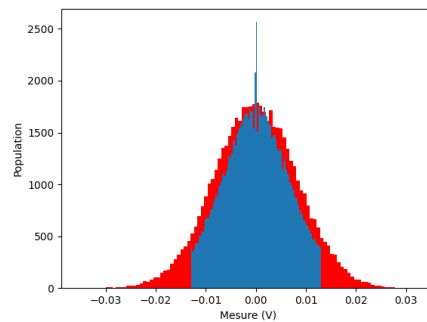


Figure 4 : Comparaison de l'histogramme de \tilde{W} (en bleu) avec celui de W_r (en rouge) après prise en compte des valeurs pathogènes

On remarquera que le bruit prend bien sûr des valeurs plus petites de le seuil T , expliquant l'arrêt net de la gaussienne en dehors de l'intervalle $[-T, T]$, valant ici environ 0.013 (la console affichant la valeur 0.01298870919743046). La valeur de $\frac{\sigma \log(\log(n))}{\log(n)}$, qui caractérise l'intervalle de confiance autour de l'encadrement $[-T, T]$ dans le résultat énoncé (1), est de 0.0030, sachant qu'ici $n = 17$.