Le modèle d'innovation en traitement du signal

Julien Fageot

A partir d'un stage réalisé au sein du Biomedical Imaging Group de l'EPFL,
coencadré par M. Unser et P. Tafti.

1 Introduction

Le traitement du signal se donne pour but d'analyser et interpréter des messages, simplifiés et généralement codés, appelés signaux. Ce domaine de recherche a connu à partir des années cinquante un rapide développement autour de la théorie de Shannon via l'analyse de Fourier. Ainsi, Shannon démontra qu'un signal continu dont la transformée de Fourier est à support compact peut être échantillonné (c'est-à-dire discrétisé) puis reconstruit parfaitement sans perte d'information. Cette théorie fut à la base de la conversion analogique-digitale avec comme application essentielle les - aujourd'hui antiques - CD audios. De la même période datent les travaux de Wiener offrant une approche probabiliste permettant de restaurer optimalement un signal gaussien, corrompu par un bruit gaussien : on parle de filtrage de Wiener. Ce cadre théorique est aujourd'hui bien compris, tant dans l'usage pratique que l'on peut en faire que dans ses limitations.

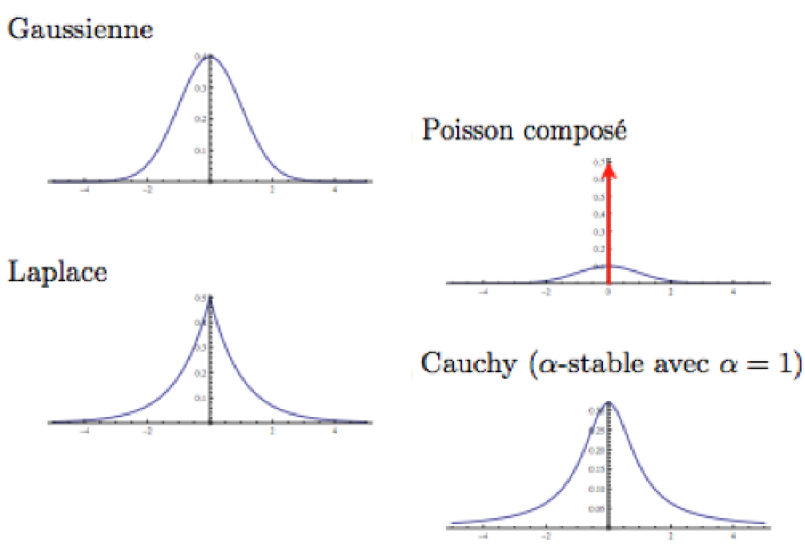
Notamment, il serait optimal dans un univers “stationnaire et gaussien”, c’est-à-dire dans lequel les signaux que nous étudions seraient stationnaires (le signal $s_{t+\tau}$ suit les mêmes lois statistiques que le signal s_t) et de statistiques gaussiennes.

Dans les années huitante se développèrent notamment par les travaux de Stéphane Mallat de nouvelles approches pour le traitement des signaux numériques : la théorie des ondelettes. Si les ondelettes en tant que concept mathématique sont bien antérieures, Mallat fit date en reliant la transformation en ondelette d’un signal avec son analyse multiéchelle. Nous présenterons et utiliserons ces idées par la suite (cf. partie (3)). La force des ondelettes et de certaines nouvelles approches, qui les rendent plus à même de traiter les signaux numériques, est la *parcimonie* de tels signaux. Est parcimonieux un signal s_n dont l’énergie est contenue dans peu de coefficients, c’est-à-dire dont la plupart des coefficients sont nuls ou quasi-nuls et ou l’information importante est donc portée par une faible fraction du signal. Remarquons que cette parcimonie, donc cette non-gaussiannité des signaux réels est une excellente nouvelle pour nos sociétés numériques : pour ne citer qu’un exemple, un signal gaussien est en effet par essence moins compressible qu’un signal parcimonieux, et ainsi le stockage et la transmission d’information sont considérablement facilités.

Cette notion de parcimonie a également un sens probabiliste. Pour une variable aléatoire X de densité de probabilité p, on parlera ici de parcimonie par comparaison avec la densité gaussienne p_g si

$$\frac{p}{p_g} \underset{\infty}{\longrightarrow} \infty$$

c’est-à-dire si la densité p décroit moins vite que la densité gaussienne (qui évolue en $\mathrm{O}(e^{-\frac{x^2}{2}})$). C’est notamment le case de la densité de Laplace ($p(x) = \frac{1}{2}e^{-|x|}$) ou des lois de puissance ($p(x) \sim \frac{c}{|x|^a}$ avec $a, c > 0$). De manière générale, ce sera le case pour toutes les variables aléatoires non-gaussiennes et *infiniment divisibles,* sur lesquelles reposera l’aspect probabiliste du modèle d’innovation.



Notre but sera ici de présenter un cadre probabiliste adapté à l’étude de signaux parcimonieux à domaine continu. Plus précisément, nous chercherons à définir une classe de processus stochastiques englobant les processus gaussiens et les processus parcimonieux. Cette classe peut moralement se résumer par l’équation suivante, présentant le modèle d’innovation :

$$Ls = w$$

où l’on a noté :

- s le processus stochastique que l’on veut étudier en dimension d.

– w un *bruit blanc*, c'est-à-dire un processus stochastique stationnaire totalement indépendant, représentant l'innovation du processus s,
– L un opérateur dit de *blanchiment*, déterministe.

Les fondements théoriques du modèle d'innovation $Ls = w$ posent deux questions. Tout d'abord, l'opérateur L doit être précisé. En pratique, on considérera essentiellement des opérateurs de dérivation (éventuellement fractionnaire). Typiquement, parce qu'un bruit blanc w est stationnaire, les propriétés statistiques du processus s sont reliées aux invariances géométriques de L. Par exemple, le processus s sera stationnaire pour un opérateur L invariant par translation. L'étude des opérateurs dans le cadre du modèle d'innovation fut faite notamment par Tafti et Unser et ne sera décrite ici que rapidement.

La seconde question, qui sera étudiée dans un premier temps, est celle du bruit blanc w : il n'est pas possible en effet dans un cadre classique de définir un bruit blanc à domaine continu avec une généralité satisfaisante. Afin de définir une classe de bruit blanc suffisamment riche, nous présenterons la théorie des processus stochastiques généralisés développée par Gel'fand et Vilenkin. Puis, nous étendrons cette théorie pour la rendre utilisable afin de définir les processus dits d'innovation s tels que $Ls = w$.

Dans un second temps, nous tenterons d'appliquer le modèle d'innovation dans le cadre de l'analyse en ondelette des signaux. Ce modèle induit des prédictions théoriques pour le comportement statistique des coefficients d'ondelette d'un signal, prédictions que nous présenterons. Enfin, ces prédictions théoriques sont testées sur des simulations et des données réelles.

2 Processus d'innovation

2.1 Processus stochastique généralisé et bruit blanc à domaine continu

Il n'est pas possible comme on l'a dit de définir un bruit blanc à domaine continu très général dans le cadre classique. Ainsi, Gel'fand dans sa théorie des fonctions généralisées utilise les distributions de Laurent Schwartz (qu'on appelle donc fonctions généralisées pour ne pas confondre avec la notion de distribution en théorie des probabilités) pour généraliser la notion de processus stochastique dans le même sens qu'on généralise les fonctions classiques via les distributions.

Ainsi, un processus stochastique généralisé est une fonctionnelle (stochastique) s définit sur un ensemble $\mathcal{T}$ de fonctions dites tests (typiquement, $\mathcal{T} = \mathcal{D}$, l'ensemble des fonctions infiniment dérivables à support compact ou $\mathcal{S}$, l'espace de Schwartz des fonctions infiniment dérivables à décroissance rapide) linéaire et continue (pour la topologie naturelle sur $\mathcal{T}$). On note $\varphi \mapsto < s, \varphi >$ cette fonctionnelle. Par analogie avec le théorème de Lévy, qui assure que la fonction caractéristique caractérise la loi d'une variable aléatoire, nous allons définir les processus stochastiques généralisés à partir de la notion de fonctionnelle caractéristique. Le résultat fondamental sera pour nous le :

Théorème : (Minlos-Bochner)
Soit Z une fonctionnelle définite sur $\mathcal{T} = \mathcal{D}$ ou $\mathcal{S}$, continue, nulle en 0 et définie-positive, c'est-à-dire telle que :

$$\sum_{n,m=1}^{N} a_n a_m^* Z(\varphi_m - \varphi_n) \geq 0$$

pour tout $N \in \mathbb{N}^*$, $\varphi_1, ..., \varphi_N \in \mathcal{T}$, $a_1, ..., a_N \in \mathbb{C}$.

Alors, il existe une unique mesure de probabilité μ sur $\mathcal{T}'$, dual topologique de $\mathcal{T}$, telle que :

$$Z(\varphi) = \mathbb{E}\left[e^{j\langle s,\varphi\rangle}\right] = \int_{\mathcal{T}} e^{j\langle s,\varphi\rangle} d\mu(s)$$

Inversement, toute mesure de probabilité μ sur $\mathcal{T}'$ définit par la formule précédente une fonctionnelle continue, définie-positive et nulle en 0. On note alors $Z(\varphi) = Z_s(\varphi)$.

Un processus généralisé est alors donné par une mesure μ sur l'ensemble des distributions (tempérées ou non). En particulier, si $X = (\langle s, \varphi_1 \rangle, ..., \langle s, \varphi_n \rangle)$ (variable aléatoire dans $\mathbb{R}^n$), alors :

$$\Phi_X(\omega) = \mathbb{E}[e^{j\omega.X}] = Z_w(\omega_1 \varphi_1 + ... + \omega_n \varphi_n)$$

La propriété d'indépendance d'un processus ne peut plus être décrite point à point (puisqu'une fonction généralisée est définie à un ensemble de mesure nulle près), et on définit alors l'indépendance d'un processus par l'indépendance des variables aléatoires $X_1 = \langle s, \varphi_1 \rangle$ et $X_2 = \langle s, \varphi_2 \rangle$ dès que φ_1 et φ_2 sont à supports disjoints. Un bruit blanc a en outre des propriétés de stationnarité.

Etant donnée une fonction f continue de $\mathbb{R}$ dans $\mathbb{C}$, on définit sur $\mathcal{T}$:

$$Z^f(\varphi) = \exp\left(\int_{\mathbb{R}^d} f(\varphi(\mathbf{r}))d\mathbf{r}\right)$$

Une telle fonctionnelle a bien formellement les propriétés d'indépendance et de stationnarité requises. Le théorème suivant donne des conditions sur f pour qu'elle constitue bien une fonctionnelle caractéristique sur $\mathcal{D}$ (théorie de Gelfand) ou $\mathcal{S}$ (extension de la théorie).

Théorème et définition :

– *Sur $\mathcal{D}$, la fonctionnelle Z^f est une fonctionnelle caractéristique bien définie si et seulement si f s'écrit :*

$$f(\omega) = ja\omega - \frac{b\omega^2}{2} + \int_{\mathbb{R}\backslash(0)} \left(e^{jt\omega} - 1 - j\omega \mathbb{I}_{|t|<1}\right) v(t)dt$$

avec $a \in \mathbb{R}$, $b \in \mathbb{R}^+$ et v une mesure de Lévy, c'est-à-dire telle que $\int_{\mathbb{R}\backslash\{0\}} (1 \wedge t^2) v(t)dt < \infty$. f est appelée une fonction de Lévy.

– *Sur $\mathcal{S}$, il est nécessaire que f soit une fonction de Lévy, et il est alors suffisant que v vérifie $\int_{\mathbb{R}\backslash\{0\}} (|t|^\epsilon \wedge t^2) v(t)dt < \infty$ pour un certain $\epsilon > 0$.*

Dans ce cas, on note w le processus stochastique associé et on a ainsi défini un bruit blanc à domaine continu.

Comme $\mathcal{D} \subset \mathcal{S}$, il est plus difficile de définir un bruit blanc sur $\mathcal{S}$ que sur $\mathcal{D}$. La condition sur f est la même que dans la représentation de Lévy-Khintchine des processus de Lévy (processus à trajectoire càdlàg et à accroissement indépendants et stationnaire). D'ailleurs, un bruit blanc peut-être vu en un sens comme la dérivée au sens des distributions d'un processus de Lévy.

On trouve également un lien avec la notion de variable aléatoire infiniment divisible, c'est-à-dire pouvant se décomposer sous la forme d'une somme de n variables iid pour tout n. En effet, un bruit blanc voit toutes ses marginales $\langle w, \varphi \rangle$ être des variables aléatoires infiniment divisibles. La condition sur f est donc un résultat central en théorie des probabilités puisqu'il décrit des classes fondamentales de processus ou de variables aléatoires.

Si $f(\omega) = -\frac{\sigma^2\omega^2}{2}$, on a alors un bruit blanc *gaussien* et pour tout φ, $\langle w, \varphi \rangle \sim \mathcal{N}\left(0, \|\varphi\|_2^2\right)$. Si le bruit blanc n'est pas gaussien, on dira qu'il est *parcimonieux*. En effet, les variables infiniment divisibles et non-gaussiennes ont la propriété d'être plus parcimonieuses que les gaussiennes, et la connexion entre le bruit blanc et les variables infiniment divisibles justifie alors cette appellation.

2.2 Le processus d'innovation et ses propriétés

Nous sommes capables de définir un bruit blanc sur $\mathcal{T}$. On veut maintenant donner du sens au processus $s = Uw = L^{-1}w$. Tout d'abord, puisqu'il s'agit de processus généralisé, cette équation s'interprète formellement, sur $\mathcal{T}$:

$$\langle s, \varphi \rangle = \langle Uw, \varphi \rangle = \langle w, U^* \varphi \rangle$$

Autrement dit, c'est plutôt non sur les opérateurs U ou L, mais sur $T = U^*$ que nous nous concentrerons. En fait, la question cruciale pour la bonne définition du processus ne requiert pas tant qu'on donne une précise définition de l'opérateur T, mais simplement qu'on connaisse son ensemble d'arrivée. En effet, pour définir un processus s, il n'est pas requis de définir a priori un bruit blanc w si on se concentre sur la fonctionnelle caractéristique. Les équations précédentes se comprennent alors formellement par :

$$Z_s(\varphi) = Z_w(T\varphi) = \exp\left(\int_{\mathbb{R}^d} f(\{T\varphi\}(\mathbf{r}))d\mathbf{r}\right)$$

Définition des processus d'innovation :
Les questions centrales sont donc, si on définit pour un opérateur linéaire et continu de $\mathcal{T}$ dans $\mathcal{E} \subset \mathcal{D}'$ et une fonction f continue de $\mathbb{R}$ dans $\mathbb{C}$, de savoir si $Z^{f,T}$, définie par la formule ci-dessus a les propriétés suivantes :

- $Z^{f,T}$ est bien définie pour $\varphi \in \mathcal{T}$ et continue sur $\mathcal{T}$.
- $Z^{f,T}$ est définie-positive sur $\mathcal{T}$.
- $Z^{f,T}(0) = 1$.

Si tel est le cas, le théorème de Minlos-Bochner assure l'existence d'un tel processus.

Nous donnons ici les principaux résultats sur ces questions :

- $Z^{f,T}(0) = 1$ est toujours vérifié puisqu'on considère toujours des fonctions f nulle en 0.
- Si $\mathcal{T} = \mathcal{E} = \mathcal{D}$, alors la théorie de Gel'fand s'applique $Z^{f,T}$ est bien la fonctionnelle caractéristique d'un processus s.
- La continuité assure la définie-positivité, tant que $\mathcal{E}$ est inclu dans les fonctions mesurables, et à la condition que f soit une fonction de Lévy.
- Le problème de la continuité (et de bonne définition qui se résoud de la même façon) peut-être formulé différemment : on se demande plutôt si $\varphi \mapsto \int_{\mathbb{R}^d} f(\varphi(\mathbf{r}))d\mathbf{r}$ est continue sur $\mathcal{E}$, pour la norme naturelle. Ce fait dépendra des contraintes qu'on impose sur f : plus l'ensemble $\mathcal{E}$ est grand, plus les contraintes sur f sont importantes. On a vu par exemple avec Gel'fand que $\mathcal{E} \subset \mathcal{D}$ (le plus petit ensemble possible) imposait que f soit une fonction de Lévy. On a également vu que si $\mathcal{E} \subset \mathcal{S}$, il est suffisant d'ajouter la condition supplémentaire $\int_{|t|\geq 1} |t|^\epsilon v(t)dt < +\infty$ pour ϵ arbitrairement petit. On va maintenant donner quelques résultats pour des espaces $\mathcal{E}$ plus généraux (typiquement des intersections d'espace L^p) et donc plus conformes aux exigences des opérateurs T "intéressants". Avant cela, on donne quelques éléments sur les opérateurs T considérés.

Quelques mots sur les opérateurs et processus d'innovation
Soit T un opérateur de $\mathcal{T} = \mathcal{D}$ ou $\mathcal{S}$ dans $\mathcal{E}$ un espace vectoriel topologique. On dit que T est invariant par translation si pour tout $\mathbf{r}_0 \in \mathbb{R}^d$:

$$T\{\varphi(. - \mathbf{r}_0)\} = \{T\varphi\}(. - \mathbf{r}_0)$$

T est invariant par rotation si pour $\Omega \in O(\mathbb{R}^d)$,

$$T\{\varphi(\Omega.)\} = \{T\varphi\}(\Omega.)$$

On dit enfin qu'un opérateur est homogène d'ordre $-\gamma \in \mathbb{R}$ si

$$T\{\varphi(./\sigma)\} = |\sigma|^\gamma \{T\varphi\}(./\sigma)$$

L'homogénéité est un affaiblissement de l'hypothèse d'invariance par changement d'échelle, trop contraignante. On considèrera dans la suite des ordres $-\gamma \leq 0$.

De telles invariances sont particulièrement pertinentes pour le traitement de l'image puisqu'on peut considérer que les statistiques d'une image numérique sont invariantes par translation (stationnarité), rotation et changement d'échelle (renormalisé). On requiert essentiellement qu'un opérateur utilisé pour définir un processus d'innovation soit homogène d'un certain ordre. Cette hypothèse justifiée est indispensable pour espérer faire une analyse multiéchelle d'un signal par le modèle d'innovation.

La signification de l'égalité $s = L^{-1}w$ est, on le rappelle, à comprendre comme $\langle s, \varphi \rangle = \langle w, (L^{-1})^* \varphi \rangle$ pour $\varphi \in \mathcal{T}$. Typiquement, des opérateurs qu'on peut définir comme l'inverse d'opérateurs différentiels (dérivée fractionnaire, laplacien fractionnaire, ...) vérifie cette assertion c'est-à-dire qu'on choisit par exemple $L^* = (-\Delta)^{\gamma/2}$. Nous retiendrons simplement l'hypothèse d'homogénéité et d'invariance par translation (au moins dans une forme faible).

Résultats sur la continuité de Z^f :

Sans rentrer dans les détails techniques, on peut simplement mentionner que les opérateurs T pertinents envoient $\mathcal{D}$ dans des espaces L^p. Il est donc important de comprendre quand - sous certaines hypothèses sur f - on sait que Z^f est continue, sur des espaces L^p (ou des intersections). Si on reprend la forme générale d'une fonction de Lévy, on a notamment, formellement :

$$Z(\varphi) = \exp\left(F(\varphi)\right) = \exp\left(ja \int \varphi - \frac{b}{2} \int \varphi^2 + G(\varphi)\right)$$

où $G(\varphi) = \int_{\mathbb{R}^d} \int_{\mathbb{R}\setminus\{0\}} \left(e^{j\omega\varphi(\mathbf{r})t} - 1 - j\omega\varphi(\mathbf{r})t\mathbb{I}_{|t|\leq 1}\right) v(t) dt d\mathbf{r}$ est la partie non gaussienne de F.

Dans le cas général, on obtient les bornes sur G suivantes :

$$|G(\varphi) - G(\psi)| \leq m_2^0 \left(||\varphi - \psi||_2^2 + ||\varphi - \psi||_2 ||\psi||_2\right) + \max(1, 2^{1-p}) m_p^\infty ||\varphi - \psi||_{1\wedge p}^{1\wedge p}$$

où $m_q^0 = \int_{0<|t|\leq 1} |t|^q v(t) dt$ et $m_p^\infty = \int_{|t|\geq 1} |t|^p v(t) dt$ sont les moments en 0 ou en $+\infty$ de $v(t)dt$. Ainsi, on voit comment des convergences dans les espaces L^p et des hypothèses sur v (et donc sur f), typiquement par la finitude de certains moments. Une question cruciale est de comprendre le plus finement possible le lien entre les hypothèses sur v et les domaines qui sont alors "contrôlés" et dans lesquels peuvent arriver les opérateurs T, la précédente borne offrant un premier résultat exploitable.

Propriétés des processus d'innovation :

Une fois le processus défini, on cherche à savoir quelles propriétés de la fonction f se répercutent sur le processus, notamment sur les variables aléatoires $\langle s, \varphi \rangle$. Nous donnons ici certaines de ces propriétés :

- *$\langle s, \varphi \rangle$ est sous certaines conditions (portant de nouveau sur les moments de la fonction φ) infiniment divisible pour la fonction de Lévy $f_\varphi(\omega) = \int_{\mathbb{R}^d} f(\omega\varphi(\mathbf{r})) d\mathbf{r}$. On peut de plus exprimer le triplet de Lévy de $\langle s, \varphi \rangle$ en fonction de celui de f.*

- *Conditions de moments : Sous de nouvelles conditions de moments sur φ, les moments de $v_\varphi = \int_{Supp(\varphi)} \frac{1}{|\varphi(\mathbf{r})|} v\left(\frac{\cdot}{\varphi(\mathbf{r})}\right) d\mathbf{r}$ sont finis ssi les moments de v le sont. La conséquence importante est que les moments de la variable aléatoire $\langle s, \varphi \rangle$ ssi ceux de v le sont.*

– *La conservation de certaines propriétés essentielles (gaussianité, variance finie ou infinie, Poisson, α-stable...) est comprise avec des conditions sur φ.*

Nous avons ici présenté un survol de certaines interrogations théoriques qui furent abordées lors du stage. Il était en effet important de définir tout d'abord l'objet théorique qu'est un processus d'innovation avec un maximum de généralité (de ce point de vue, de nouveaux résultats sont attendus), et de comprendre alors ses principales propriétés.

3 Analyse en ondelettes et modèle d'innovation

Le but est maintenant de comprendre le comportement d'un processus d'innovation soumis à une analyse multiéchelle. On définit tout d'abord une ondelette comme une fonction $\psi \in L^1 \cap L^2$ de moyenne nulle : $\int_{\mathbb{R}^d} \psi(\mathbf{r}) d\mathbf{r} = 0$, telle que la famille $\psi_{j,\mathbf{n}}(\mathbf{r}) = \frac{1}{2^{jd/2}} \psi\left(\frac{r_1 - 2^j n_1}{2^j}, ..., \frac{r_d - 2^j n_d}{2^j}\right)$ forment une base orthonormées de l'ensemble $L^2(\mathbb{R}^d)$. Ainsi, une image (ie un fonction $s \in L^2$) est comprise par son comportement aux différentes échelles j, chacune porteuse d'une certaine information. Certains détails de l'image seront visibles aux échelles grossières et d'autres aux échelles fines. Les coefficients d'ondelette de l'image s à l'échelle j sont donc les $\langle s, \psi_{j,\mathbf{n}} \rangle$ pour $\mathbf{n} \in \mathbb{N}^d$.

Pour le modèle d'innovation, nous considérerons des versions continues des opérations en ondelette afin de mieux comprendre le comportement. Nous définissons donc :

$$\psi_{a,\mathbf{r}_0}(\mathbf{r}) = \frac{1}{a^{d/2}} \psi\left(\frac{\mathbf{r} - a\mathbf{r}_0}{a}\right)$$

Ainsi, on définit les coefficients d'ondelette $w_{a,\mathbf{r}_0}$ par :

$$w_{a,\mathbf{r}_0} = \langle s, \psi_{a,\mathbf{r}_0} \rangle$$

Nous oublierons dans la suite le paramètre $\mathbf{r}_0$ puisque nous nous concentrons sur un comportement en fonction de l'échelle. On définit alors le processus multi-échelle y par

$$y_\tau = a^{\frac{d}{2} - \gamma} w_{\tau^{1/d}}$$

La fonction caractéristique du processus multiéchelle est alors :

$$\Phi_\tau(\omega) = \Phi_{y_\tau}(\omega) = \exp(\tau f_{T\psi}(\omega))$$

Autrement dit, le processus multiéchelle a les même statistiques d'ordre 1 qu'un processus de Lévy de fonction de Lévy $f_{T\psi}$. Cependant y_τ n'est pas un processus de Lévy en général, et n'est d'ailleurs pas nécessairement markovien.

Nous donnons ici quelques résultats théoriques de l'étude de ce processus.

Proposition : Equations d'évolution dans les cas gaussiens et α-stables

– *Supposons f fonction de Lévy gaussienne de variance σ^2. Alors la fonction caractéristique de $\mathbf{y}_\tau = (y_\tau^1, ..., y_\tau^n)$ avec $\left(y_\tau^i\right)_\tau$ processus multiéchelle de l'ondelette ψ^i est*

$$\exp\left(-\frac{\tau\sigma^2\omega^T C w}{2}\right)$$

où $C \in M_n(\mathbb{R})$ est définie par $C(i,j) = \langle T\psi^i, T\psi^j \rangle_{L^2}$. Autrement dit, si on note p_τ la densité gaussienne de $\mathbf{y}_\tau$, et que l'on décompose la matrice symmetric $C = P^T D P$ (P orthogonale et D diagonale), on a :

$$\frac{\partial}{\partial\tau} p_\tau(P\mathbf{x}) - \frac{\sigma^2}{2} \sum_{i=1}^{n} \lambda_i \frac{\partial^2}{\partial x_i^2} p_\tau(P\mathbf{x}) = 0$$

On reconnait une équation de diffusion.

- *Supposons f fonction de Lévy α-stable pour $0 < \alpha < 2$. Alors, la fonction caractéristique de la variable aléatoire y_τ est*
$$\Phi_\tau(\omega) = \exp\left(-\tau \|\varphi\|_\alpha^\alpha |\omega|^\alpha\right)$$
Pour la densité de probabilité, cela impose que
$$\frac{\partial}{\partial \tau} p_\tau(x) + \|\varphi\|_\alpha^\alpha \left(-\Delta_x\right)^{\alpha/2} p_\tau(x) = 0$$
On reconnait cette fois une équation de diffusion fractionnaire.

Ces résultats sont importants car ils montrent que l'évolution a travers les échelles et liée au comportement "spatiale" des processus stables.

Théorème : Comportement aux échelles larges : cas symétrique

Supposons f fonction de Lévy symétrique.

- *Supposons que f corresponde à un bruit de variance finie (se traduit par $\int_{\mathbb{R}\setminus\{0\}} t^2 v(t)dt < +\infty$). Alors, on a :*
$$\frac{w_{a,\mathbf{r}_0}}{a^\gamma} \xrightarrow[a\to+\infty]{\mathcal{L}} \mathcal{N}(0, \|\varphi\|_2^2(\sigma^2 + m_2))$$
Notons que le même résultat peut être obtenu en multivarié.

- *Supposons que f, associé à la fonction de répartition F, soit de variance infinie avec l'hypothèse $x^\alpha F(x) \xrightarrow[x\to\infty]{} l > 0$. Alors, on a pour un certain $c > 0$:*
$$\frac{w_{a,\mathbf{r}_0}}{a^{\gamma + d\left(\frac{1}{\alpha} + \frac{1}{2}\right)}} \xrightarrow[a\to+\infty]{\mathcal{L}} S(\alpha, \|\varphi\|_\alpha c)$$
C'est-à-dire que les coefficients d'ondelette sont asymptotiquement gaussiens (avec une renormalisation dépendant de γ donc de T).

Conclusion :

Le développement de ce cadre théorique est essentiellement motivé, on l'a dit, par le fait de produire un modèle probabiliste pour les signaux parcimonieux. En ce qui concerne les images numériques, quelques prévisions théoriques, qu'il s'agirait d'étudier plus avant, peuvent être mentionnées :

- Les coefficients d'ondelette d'une image numérique sont un signal à variance finie, et de moyenne nulle. On l'a vu, dans le cadre du modèle d'innovation, cela impose une convergence vers une loi gaussienne pour les échelles grossières, fait expérimentalement vérifié (au moins qualitativement).

- Aux échelles supérieures, le comportement devient alors de plus en plus gaussien. Ce fait n'est pas exploité aujourd'hui dans les algorithmes de traitement de l'image. Le modèle rejette l'hypothèse aujourd'hui communément utilisée dans les modèles que la parcimonie témoignent d'un comportement statistique suivant une loi de Laplace ($p(x) = \frac{1}{2}e^{-|x|}$). Notamment : on ne peut suivre une loi de Laplace au plus qu'à une échelle (le signal n'étant plus Laplacien aux échelles supérieures).

- La question du choix de l'ondelette est pour l'ingénieur crucial : celle-ci doit s'adapter au signal considéré, suivant l'opération qu'on veut réaliser (débruitage, transmission, problème inverse...). La théorie du modèle d'innovation suggère que le choix de l'ondelette n'a - en ce qui concerne les comportements statistiques des coefficients - qu'une importante limitée . Ainsi, sous certaines hypothèses de compatibilité avec le signal (appartenance à des intersections d'espace L^p pour la fonction $\varphi = T\psi$ essentiellement), une ondelette ne diffère d'une autre que par la valeur des moments de φ, qui influent sur les paramètres "secondaires" du modèle (variance de la gaussienne par exemple).

- Au delà des questions théoriques, un modèle pour les signaux parcimonieux permet de développer de nouveaux algorithmes en traitement d'image, en interprétant le modèle suivant une approche bayesienne, débouchant sur des méthodes variationnelles exploitables concrètement.

D'un point de vue théorique, certaines questions mériteraient d'être plus approfondies :

- Sur la définition des processus d'innovation, on peut envisager de développer un cadre théorique plus général et plus unis permettant de mieux comprendre le lien entre les conditions sur les fonctions de Lévy et les propriétés des opérateurs (prendre en compte notamment pour les résultats de continuité l'ensemble des propriétés de T et non seulement l'ensemble d'arrivée $\mathcal{E}$). Notamment, la théorie se restreint aujourd'hui à des ondelettes infiniment dérivable et à support compact, deux hypothèses dont on aimerait s'éloigner.

- L'étude du processus multi-échelle : dans le cas général, peu de résultats sont aujourd'hui connus. Peut-on développer comme dans les cas stables des équations d'évolution ? Si oui, des techniques d'EDP peuvent-elles s'appliquer pour comprendre les propriétés qualitatives des densités de probabilité du phénomène. Peut-on étudier le processus d'innovation par ses marginales multivariées comme il a été fait dans le cas gaussien ?
- Il est également envisageable de combiner différents modèles d'innovation (par exemple en différentes parties d'une même image) et ainsi créer un modèle plus riche du comportement statistique des images.