

# Causalité de Granger pour des signaux à valeurs fonctionnelles

Pierre-Olivier AMBLARD<sup>1,2\*</sup> et Olivier J. J. MICHEL<sup>1</sup>

<sup>1</sup>GIPSA-lab, DIS(UMR CNRS 5216), ENSE3-BP 46 38402 Saint Martin d'Hères Cedex

<sup>2</sup>Dept. Math&Stat, The University of Melbourne, Parkville, VIC3010, Australie.

bidou.amblard@gipsa-lab.inpg.fr, olivier.michel@gipsa-lab.grenoble-inp.fr

**Résumé** – Tester l'influence d'un signal sur un autre peut être réalisé à l'aide de la causalité de Granger. Ce principe stipule qu'un premier signal est cause d'un deuxième si le passé du premier améliore la prédiction du deuxième. Ce principe est largement utilisé dans des situations stationnaires. Des extensions *ad-hoc* pour le non stationnaire existent, mais une véritable théorie fait défaut. Motivé par les signaux cyclostationnaires, nous proposons ici d'utiliser une modélisation de signaux prenant des valeurs fonctionnelles dans un espace de Hilbert, et nous étendons la causalité de Granger dans ce cadre. Diverses définitions sont proposées utilisant l'indépendance conditionnelle ou la prédiction linéaire. Enfin, nous donnons quelques pistes pour utiliser pratiquement le concept étendu.

**Abstract** – Granger causality is an answer to knowing whether a signal influences another signal or not. Its principle relies on prediction theory: a first signal causes a second one if the first helps in the prediction of the second. Granger causality was developed for economy applications but is now widely used across many fields, however in stationary situations. Possible extension have already been proposed to tackle nonstationarity, but these are rather *ad-hoc*. A general theory for Granger causality in nonstationary contexts is still lacking. Motivated by cyclostationary signals, we propose here to model signals as discrete time signals taking values in infinite dimensional Hilbert function spaces. We then extend Granger causality in this particular context, providing strong and weak definitions, and giving some hints to practically apply the theory.

## 1 Introduction

Une problématique très actuelle est celle de l'analyse de dépendance entre signaux aléatoires et plus particulièrement celle de la dépendance causale. Cette question est fondamentale dans des disciplines aussi diverses que la climatologie ou les neurosciences. Elle va par ailleurs de pair avec la notion de modèles graphiques pour les signaux. Une formalisation possible est proposée par la causalité de Granger [1, 2, 3], qui repose sur la notion de prédiction d'un signal par un autre. La théorie développée en économétrie a été répandue très largement dans de nombreux domaines. En général, elle repose sur des hypothèses de stationnarité des signaux. Toutefois, des questions importantes en neurosciences par exemple ou en climatologie, nécessitent d'étudier des dépendances causales entre signaux non stationnaires généralement, et cyclostationnaires en particulier. Rappelons qu'un signal aléatoire est cyclostationnaire si ses statistiques sont périodiques. Ainsi, dans les disciplines touchant au vivant ou au climat, les horloges naturelles (rotation terrestre, rythmes biologiques, *etc.*) impliquent l'émergence de nombreux processus stochastiques cyclostationnaires. Des efforts doivent être menés pour développer l'analyse des dépendances causales sous ces hypothèses de non ou de cyclostationnarité.

Dans beaucoup d'applications de la théorie du signal, les mesures à traiter sont données sous forme d'une suite de fonctions définies sur un intervalle de  $\mathbb{R}$ . Les exemples sont multiples. En physique, les spectres de la lumière solaire sont délivrés de manière journalière par des spectromètres à bord de satellites. Dans ce cas, le signal peut être vu comme une suite de fonctions de l'intervalle de fréquence considéré et à valeurs dans  $\mathbb{R}$ . Un modèle mathématique pertinent est de considérer

la mesure comme une suite  $X_n$ ,  $n \in \mathbb{Z}$ , de variables à valeurs dans  $L^2([\lambda_m, \lambda_M])$ , où  $\lambda_m$  et  $\lambda_M$  sont les longueurs d'onde respectivement minimale et maximale de l'intervalle d'indexation.  $n$  correspond au jour de la mesure.  $X_n$  est alors un signal à temps discret prenant des valeurs dans l'espace de fonctions  $L^2([\lambda_m, \lambda_M])$ . Dans d'autres domaines, les signaux sont cyclostationnaires (vibrations de machines tournantes, consommation d'électricité, température en un lieu donné, *etc.*) Un signal cyclostationnaire de période connue peut être transformé en un signal stationnaire en le regardant comme une suite de variables aléatoires à valeurs dans un espace fonctionnel (théorème de Gladyshev, [4]).

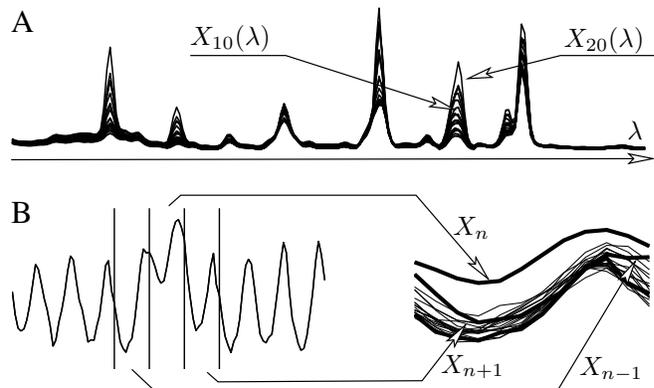


FIGURE 1 – A : Spectre solaire : chaque mesure journalière  $X_n(\lambda)$  est une fonction d'un intervalle de longueurs d'onde dans  $\mathbb{R}^+$  ; B : Consommation mensuelle d'électricité sur 9 années, période  $T = 12$  mois et sa transformation en un signal à valeurs fonctionnelles.

\*P.O.A. est soutenu par une bourse Marie Curie International Outgoing Fellowship de l'Union Européenne.

Dans ce papier, nous étendons la notion de causalité de Granger aux signaux à temps discret à valeurs fonctionnelles. Ceci permet d'envisager l'application de la causalité de Granger aux signaux à temps continu, aux signaux cyclostationnaires, mais également à des situations multicapteurs dans lesquelles chaque capteur mesure périodiquement des fonctions sur un intervalle de temps donné. Notre contribution est ici essentiellement conceptuelle ; nous ne faisons que définir l'extension et proposons quelques pistes pour son utilisation.

Nous donnons dans un premier paragraphe quelques éléments sur les variables aléatoires à valeurs dans des espaces fonctionnels de Hilbert et sur les signaux aléatoires à valeurs dans un tel espace. Dans un deuxième temps, nous présentons les définitions formelles de la causalité de Granger. Des définitions fortes reposant sur l'indépendance conditionnelle sont déclinées en des versions plus faibles reposant sur la prédiction. La notion de couplage instantané est également prise en compte. Dans un dernier paragraphe, nous donnerons des pistes pour l'application pratique de ces définitions.

## 2 Signaux à valeurs fonctionnelles

Nous utilisons dans ce travail la théorie des variables aléatoires à valeurs dans des espaces de Hilbert (séparables) de dimension infinie [5]. La théorie est donc un peu plus délicate et moins connue que dans le cas usuel de la dimension finie. On considère un espace de probabilité  $(\Omega, \mathcal{F}, P)$ . Soit  $H$  un espace de Hilbert séparable, de produit scalaire  $\langle \cdot | \cdot \rangle_H$ . L'indice  $H$  sera en général omis s'il n'y a pas de risque de confusion. On supposera ici que  $H$  est un espace de fonctions numériques définies sur un intervalle  $I$  de  $\mathbb{R}$  et à valeurs dans les réels ou les complexes. Des exemples sont l'espace  $L^2(I)$  ou les espaces de Sobolev  $W^{2,k}(I)$ .  $H$  est doté de la tribu borélienne.

**Variables.** Pour que  $X : \Omega \rightarrow H$  soit une variable aléatoire, il est nécessaire et suffisant que  $\langle x | X \rangle$  soit une variable aléatoire numérique  $\forall x \in H$ . On considère maintenant l'espace  $\Delta^2(H)$  des variables aléatoires du second ordre caractérisées par  $E[\|X\|^2] < +\infty$ , où  $\|\cdot\|$  est la norme dans  $H$  issue de  $\langle \cdot | \cdot \rangle$ .  $\Delta^2(H)$  est lui-même un espace de Hilbert pour le produit scalaire  $\langle X | Y \rangle_\Delta := E[\langle X | Y \rangle_H]$ . Si  $X$  est du second ordre, alors il existe un unique élément  $m$  de  $H$  tel que  $E[\langle u | X \rangle_H] = \langle u | m \rangle_H, \forall u \in H$ .  $m$  est appelé vecteur moyen et l'on écrit  $m = E[X]$ . On supposera par simplicité dans la suite des variables à moyenne nulle. On considère maintenant deux espaces de Hilbert séparables  $H_x$  et  $H_y$ . Soient  $X$  et  $Y$  deux variables aléatoires du second ordre sur  $(\Omega, \mathcal{F}, P)$  à valeurs dans respectivement  $H_x$  et  $H_y$ . On peut alors définir la covariance entre  $X$  et  $Y$  de diverses manières équivalentes. La plus simple consiste à étudier les corrélations entre  $\langle x | X \rangle$  et  $\langle y | Y \rangle$ . L'opérateur covariance entre  $X$  et  $Y$  est alors l'unique opérateur linéaire  $\Gamma_{XY}$  de  $H_y$  dans  $H_x$  défini par  $\langle \Gamma_{XY} y | x \rangle := E[\langle x | X \rangle \langle y | Y \rangle]$ . Une expression simple de  $\Gamma_{XY} y$  est donc  $E[\langle y | Y \rangle X]$ . Cette définition se particularise évidemment au cas de l'auto-covariance. Les opérateurs de covariance et autocovariance sont bornés pour la norme usuelle  $\|T\| = \sup_x \|Tx\|/\|x\|$ . Les opérateurs de covariance sont des opérateurs de Hilbert-Schmidt, i.e.,  $+\infty > (\sum_i \|Te_i\|^2)^{1/2} =: \|T\|_{HS}$  est la norme de Hilbert-Schmidt de l'opérateur  $T$ , où  $e_i$  est une base orthonormée quelconque de  $H$ . Nous utiliserons dans le dernier paragraphe l'exemple des opérateurs linéaires intégraux de  $L^2([0, 1])$  dans lui-même. Un tel opérateur est défini par un noyau  $k(s, t)$  de carré intégrable (qui constitue sa norme de Hilbert-Schmidt) et par la relation  $y(t) = (Tx)(t) = \int_0^1 k(s, t)x(s)ds$ .

La théorie de la prédiction en moyenne quadratique repose comme dans le cas de la dimension finie sur la notion d'orthogonalité et de sous-espace vectoriel clos (fermé). Un sous-espace de  $\mathcal{G}$  de  $\Delta^2(H)$  est

un sous-espace clos si  $\mathcal{G}$  est de Hilbert et si toute transformée linéaire bornée d'élément de  $\mathcal{G}$  reste dans  $\mathcal{G}$ . Cette notion est importante pour l'estimation puisque pour que  $X \in \Delta^2(H)$  soit orthogonal au sous-espace clos  $\mathcal{G}$  (i.e.  $E[\langle X | Z \rangle] = 0, \forall Z \in \mathcal{G}$ ), il est nécessaire et suffisant que  $X$  soit décorréolé de toute variable aléatoire de  $\mathcal{G}$ , i.e.  $\Gamma_{XZ}$  est l'opérateur nul  $\forall Z \in \mathcal{G}$ .

**Signaux.** Les signaux que nous allons maintenant considérer sont des signaux à temps discrets prenant des valeurs dans l'espace fonctionnel de Hilbert  $H$ . Un signal  $X_n$  est défini comme une suite de variables aléatoires  $X_n : \Omega \rightarrow H$  pour  $n \in \mathbb{Z}$ . On retrouve alors les mêmes concepts que pour les signaux aléatoires à valeurs réelles. Par exemple, si le vecteur  $m_n = E[X_n]$  est constant et si l'opérateur  $\Gamma_{XX}(n, k) : H \rightarrow H$  défini par  $\Gamma_{XX}(n, k)x = E[\langle x | X_n \rangle X_{n+k}]$  ne dépend que de  $k$ , nous dirons que le signal est stationnaire au second ordre. Cet opérateur est appelé opérateur de covariance du signal. Il s'étend évidemment à la notion d'opérateur de covariance croisée entre deux signaux  $X_n$  et  $Y_n$ . La notion de bruit blanc s'étend également. Au sens fort, il s'agit d'une suite de variables indépendantes et identiquement distribuées. Dans un sens faible, un signal stationnaire  $\varepsilon_n$  est blanc si deux échantillons distincts sont décorréolés. L'opérateur de covariance  $\Gamma_{\varepsilon\varepsilon}(n, k)$  est donc l'opérateur nul pour  $k \neq 0$ , et est un opérateur linéaire borné pour  $k = 0$ . Un exemple de signal à temps discret à valeurs fonctionnelles est le modèle autorégressif Hilbertien. Nous l'utiliserons dans le dernier paragraphe.

Pour démystifier ces notions un peu abstraites, il faut considérer qu'en pratique, la dimension infinie n'est pas manipulable, et que nous travaillons inévitablement sur des données échantillonnées. Dans ce contexte, les signaux à valeurs fonctionnelles deviennent des signaux à temps discrets multivariés, les opérateurs de covariances sont identifiables aux matrices de covariances, ... Les difficultés liées aux grandes dimensions ne restent pas moins présentes.

## 3 Causalité de Granger pour les signaux à valeurs fonctionnelles

Considérons trois signaux aléatoires  $X_t, Y_t$  et  $Z_t, t \in \mathbb{Z}$ , à valeurs dans trois espaces de Hilbert séparables *a priori* distincts  $H_x, H_y, H_z$ . Nous étudions ici la causalité de Granger de  $Y$  vers  $X$  en considérant également l'information supplémentaire apportée par  $Z$ . On suppose ici que ces trois signaux représentent l'ensemble des informations disponibles. Soit  $A^t$  la suite des variables aléatoires  $A_j$  depuis l'origine des dates jusqu'à  $j = t$  (on devrait en toute rigueur travailler avec les tribus engendrées).

La définition forte de la causalité de Granger repose sur l'indépendance conditionnelle.

**Définition 1. Causalité de Granger :**  $Y_t$  ne cause pas  $X_t$  relativement à l'information disponible  $X^t, Y^t, Z^t$  ssi  $X_t$  et  $Y^{t-1}$  sont indépendants conditionnellement à  $X^{t-1}$  et  $Z^{t-1}$ .

Cette définition formalise la causalité de Granger qui repose sur la prédiction. Le processus  $Y$  n'est pas une cause au sens de Granger du processus  $X$  si la prise en compte de son passé n'améliore pas la prédiction de  $X$ . Formellement, cette définition est évidemment la même que dans le cas usuel de signaux à valeurs numériques. Toutefois, sa nature est un peu différente, puisque dans le cas présent, la prédiction faite est celle d'une fonction complète, et non d'un seul échantillon. Par exemple, si le signal à valeurs fonctionnelles  $X_t$  représente la succession d'une mesure journalière, la prédiction considérée est celle de la mesure sur la période entière. Cette remarque renforce l'importance de la notion de couplage instantané définie par

**Définition 2. Couplage instantané :**  $Y_t$  et  $X_t$  ne sont pas instan-

tanément couplés relativement à l'information disponible  $X^t, Y^t, Z^t$  ssi  $X_t$  et  $Y_t$  sont indépendants conditionnellement à  $X^{t-1}, Y^{t-1}, Z^t$ .

Le couplage instantané mesure la dépendance (conditionnelle) entre le présent des deux signaux. Dans le cas de signaux à valeurs fonctionnelles construits à partir d'intervalles de signaux scalaires (voir l'exemple de la figure 1B), toute dépendance entre les signaux scalaires  $x(u)$  et  $y(u)$  de portée moindre que la longueur de l'intervalle de découpage  $T$  induit un couplage instantané entre les versions à valeurs fonctionnelles  $X_t(\cdot)$  et  $Y_t(\cdot)$ .

La définition forte est difficile à utiliser en pratique, même dans le cas usuel, puisqu'elle implique l'utilisation de mesures d'indépendance conditionnelle, comme par exemple celles issues de la théorie de l'information [6]. On décline donc la définition en des versions plus faibles, mais conduisant à des implantations *a priori* plus aisées de tests.

Ces déclinaisons implantent (explicitement ou implicitement) la prédiction. Le coût choisi pour évaluer la qualité de prédiction est le coût quadratique. On cherche alors  $\hat{X}_{t+1}(W^t)$  qui approche au mieux  $X_{t+1}$  au sens de la norme dans  $\Delta^2(H_x)$  (espace des variables aléatoires à valeurs dans  $H_x$  de norme  $E[\|X\|^2]$  finie), c'est-à-dire qui minimise  $R_{X_{t+1}|W^t} = E[\|X_{t+1} - \hat{X}_{t+1}(W^t)\|^2]$ .  $W^t$  représente soit  $(X^t, Z^t)$  soit  $(X^t, Y^t, Z^t)$ . La solution du problème de minimisation est donnée par l'espérance conditionnelle  $\hat{X}_{t+1}(W^t) = E[X_{t+1}|W^t]$ . Tester la causalité de Granger ou le couplage instantané revient alors à comparer les erreurs quadratique minimale  $\bar{R}$  obtenues. Les définitions deviennent alors :

**Définition 3. :**  $Y_t$  ne cause pas  $X_t$  relativement à l'information disponible  $X^t, Y^t, Z^t$  ssi  $\bar{R}_{X_t|X^{t-1}Y^{t-1}Z^{t-1}} = \bar{R}_{X_t|X^{t-1}Z^{t-1}}$ .

**Définition 4. :**  $Y_t$  et  $X_t$  ne sont pas couplés instantanément si relativement à l'information disponible  $X^t, Y^t, Z^t$  ssi  $\bar{R}_{X_t|X^{t-1}Y^tZ^t} = \bar{R}_{X_t|X^{t-1}Z^t}$ .

Ces erreurs quadratiques moyennes obtenues ne dépendent pas du temps lorsque l'hypothèse de stationnarité des signaux est vérifiées. Toutefois nous laissons la mention explicite du temps dans la notion pour garder trace des horizons temporels considérés. Nous verrons dans le paragraphe suivant comment utiliser cette définition pratiquement.

Une version plus pragmatique de ces déclinaisons repose sur la prédiction linéaire, dont la théorie générale pour des variables à valeurs dans des espaces de Hilbert est présentée dans [5]. Considérons  $\mathcal{G}_W^t$  le sous-espace vectoriel clos de  $H_x$  engendré par  $W^t$  [5, 7]. Par exemple, si  $\ell_{x,i} \in \mathcal{L}(H_x, H_x)$  et  $\ell_{z,i} \in \mathcal{L}(H_z, H_x)$  sont des opérateurs linéaires, alors  $\sum_{i \geq 0} (\ell_{x,i}(X_{t-i}) + \ell_{z,i}(Z_{t-i}))$  est un élément de  $\mathcal{G}_{x,z}^t$  (sous réserve de convergence des sommes). Le problème de prédiction linéaire consiste à trouver l'élément  $\hat{X}_{t+1}$  de  $\mathcal{G}_W$  qui minimise  $R_W = E[\|X_{t+1} - \hat{X}_{t+1}\|^2]$ . La solution est donnée par la projection orthogonale de  $X_{t+1}$  sur  $\mathcal{G}_W^t$ . Si  $\bar{R}_{X_t|X^{t-1}Y^{t-1}Z^{t-1}} = \bar{R}_{X_t|X^{t-1}Z^{t-1}}$  sont les erreurs minimales obtenues, tester la causalité de Granger de  $Y$  vers  $X$  relativement aux informations disponibles revient alors à les comparer. Autrement dit les définitions restent les mêmes dans le cas de la prédiction linéaires que les définitions 3 et 4. Si  $\bar{R}_{X_t|X^{t-1}Y^{t-1}Z^{t-1}} = \bar{R}_{X_t|X^{t-1}Z^{t-1}}$ , alors  $Y$  ne cause pas  $X$  relativement aux informations disponibles. Si par contre  $\bar{R}_{X_t|X^{t-1}Y^{t-1}Z^{t-1}} < \bar{R}_{X_t|X^{t-1}Z^{t-1}}$ , on dit que  $Y$  est une cause de  $X$ , relativement aux informations disponibles.

## 4 Tester la causalité de Granger

Dans le cas d'un espace hilbertien de dimension finie, des approches fondées sur des modélisations autorégressives usuelles sont

utilisées, comme par exemple dans [8, 9]. Dans le cas de la dimension infinie, les procédures d'identification sont beaucoup plus délicates à mettre en œuvre [7, 10]. Nous discutons ici deux pistes pour implanter des tests : la première repose sur la prédiction non paramétrique fonctionnelle, la deuxième sur une modélisation paramétrique du signal à temps discret à valeurs fonctionnelles.

**Prédiction non paramétrique.** Une procédure consiste à mettre en pratique un estimateur non paramétrique fonctionnel pour estimer l'espérance conditionnelle, solution optimale de la prédiction en moyenne quadratique lorsque la linéarité n'est pas imposée [7, 10, 11]. Pour tester la causalité de Granger, le problème est d'obtenir les erreurs quadratiques minimales de prédiction. Nous pouvons les obtenir de la manière suivante. Soit  $K : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  un noyau de somme unité et  $K_h(x) := K(x/h)$ . Supposons avoir à disposition  $N$  observations i.i.d.  $(X_{t+1,i}, W_i^t), i = 1, \dots, N$  où  $W$  est soit  $(X, Z)$  soit  $(X, Y, Z)$ . L'espérance conditionnelle recherchée  $\hat{X}_{t+1}(W^t) = E[X_{t+1}|W^t]$  est approchée par la quantité

$$\hat{X}_{t+1|t}(W^t) = \frac{\sum_i X_{t+1,i} K_h(\|W^t - W_i^t\|)}{\sum_i K_h(\|W^t - W_i^t\|)}$$

où la norme  $\|\cdot\|$  est une norme dans les espaces considérés.  $h$  règle la largeur du noyau. Cette dernière forme est la généralisation aux données à valeurs fonctionnelles de l'usuel régressogramme [11].

En pratique pour tester si  $Y$  cause  $X$  relativement à l'information disponible, il faut scinder les données en deux parties. La première permet par validation croisée d'optimiser  $h$ . Cette procédure est effectuée deux fois pour obtenir  $\hat{X}_{t+1|t}((X, Z)^t)$  et  $\hat{X}_{t+1|t}((X, Y, Z)^t)$ . Les estimateurs obtenus sont ensuite appliqués à la deuxième partie des données afin d'évaluer les erreurs minimales et conclure quant à la causalité de Granger.

**AR hilbertiens.** Une autre approche consiste à proposer un modèle paramétrique pour expliquer les données et à transférer la causalité éventuelle dans les paramètres. Imaginons devoir tester la causalité de Granger entre deux signaux  $X_n$  et  $Y_n$  en prenant en compte également la mesure d'un troisième  $Z_n$ , prenant des valeurs dans des espaces de Hilbert  $H_x, H_y, H_z$ . On modélise alors le processus trivarié comme un modèle AR. On se contente ici d'un modèle AR d'ordre 1, puisque des modèles d'ordre plus élevé peuvent se voir comme des modèles d'ordre 1 dans des espaces plus grands [7].

On suppose donc qu'il existe un opérateur linéaire  $\rho$  de l'espace produit  $H = H_x \times H_y \times H_z$  dans lui-même, tel que  $W_n = \rho(W_{n-1}) + \varepsilon_n$ , où  $\varepsilon_n$  est un bruit blanc à valeurs dans  $H$  et où  $W_n = (X_n, Y_n, Z_n)^\top$ . S'il existe un entier  $j_0 \geq 1$  tel que  $\|\rho^{j_0}\| \leq 1$ , alors le modèle est bien défini et représente un signal stationnaire [7]. Notons le fait étrange que la norme de  $\rho$  peut être plus grande que 1 !  $\rho$  peut être considéré comme une matrice d'opérateurs linéaires  $\rho_{\alpha\beta}$  de  $H_\beta$  dans  $H_\alpha$ ,  $(\alpha, \beta) \in \{x, y, z\}^2$ . Si nous testons la causalité de  $Y$  vers  $X$ , on écrit  $X_n = \rho_{xx}(X_{n-1}) + \rho_{xy}(Y_{n-1}) + \rho_{xz}(Z_{n-1}) + \varepsilon_{x,n}$ . Dans ce modèle, on constate que  $Y$  ne cause pas  $X$  relativement aux informations disponibles si et seulement si  $\rho_{xy}$  est identiquement nul. De plus,  $X$  et  $Y$  sont instantanément non couplés si et seulement si conditionnellement à  $X^{n-1}, Y^{n-1}$  et  $Z^n$ ,  $X_n$  et  $Y_n$  sont indépendants.

La stratégie consiste alors à identifier les opérateurs et tester leur nullité. Evidemment le contexte de la dimension infinie rend l'identification très difficile. Dans le cas d'un modèle AR, on montre que  $\rho$  satisfait  $\rho\Gamma(0) = \Gamma(1)$  où  $\Gamma(n)$  est l'opérateur de covariance entre  $W_k$  et  $W_{k+n}$ . En général l'image de  $\Gamma(0)$  n'est pas  $H$  et donc l'inverse n'est pas défini partout. De plus, cet opérateur est estimé à partir de  $n$  données, et l'inverse ne peut être défini que sur un espace

$n$ -dimensionnel. Des régularisations sont alors mises en places pour résoudre l'inversion et estimer  $\rho$  [7].

Toutefois, si l'on travaille sur des données échantillonnées temporellement, on peut se contenter de la réalisation des concepts précédents en dimension finie. Des approches usuelles peuvent alors s'appliquer, avec la difficulté des grandes dimensions. A titre d'illustration, on considère un modèle bidimensionnel ( $Z$  est absent) dans lequel les espaces de Hilbert sont  $L^2([0, 1])$  et les opérateurs  $\rho_{\alpha\beta}$  sont des opérateurs intégraux Hilbert-Schmidt. Pour l'exemple, on considère des noyaux  $k_{\alpha,\beta}(s, t) = a_{\alpha\beta} \exp(-c_{\alpha,\beta}|s - t|)$ , les constantes  $c_{\alpha,\beta}$  étant positives. On étudiera principalement le cas  $a_{xx} = a_{yy} = 1$ ,  $a_{xy} = 0$ ,  $a_{yx} \neq 0$  de sorte que  $Y$  cause  $X$  mais  $X$  ne cause pas  $Y$ . On a donc

$$X_n(t) = \int_0^1 k_{x,x}(s, t) X_{n-1}(s) ds + \int_0^1 k_{x,y}(s, t) Y_{n-1}(s) ds + \varepsilon_{n,x}(t)$$

$$Y_n(t) = \int_0^1 k_{y,y}(s, t) Y_{n-1}(s) ds + \varepsilon_{n,y}(t)$$

On travaille évidemment sur des données échantillonnées, et donc en dimension finie. Pour cette illustration, on discrétise  $[0; 1]$  en  $d$  échantillons équirépartis. On identifie brutalement l'opérateur par  $\hat{\rho} = \hat{\Gamma}_N(1)\hat{\Gamma}_N(0)^{-1}$ . Les indices de causalité envisagés sont les rapports des normes des opérateurs estimés,

$$i_{y \rightarrow x} = \frac{\|(\rho_{xx}\rho_{xy})\|}{\|\rho_{xx}\|} \quad \text{et} \quad i_{x \rightarrow y} = \frac{\|(\rho_{yx}\rho_{yy})\|}{\|\rho_{yy}\|}$$

où  $(\rho_{xx}\rho_{xy})$  est l'opérateur allant de  $H_x \times H_y \rightarrow H_x$  défini par  $(\rho_{xx}\rho_{xy})(U, V) = \rho_{xx}U + \rho_{xy}V$ .  $i_{y \rightarrow x}$  mesure la norme de l'opérateur estimé lorsque  $Y$  est pris en compte rapportée à la norme de l'opérateur lorsque  $Y$  n'est pas pris en compte. Ce rapport est plus grand que 1, et vaut 1 si et seulement si l'opérateur estimé  $\rho_{xy}$  est nul, c'est-à-dire si  $Y$  ne cause pas  $X$ . La figure (2 haut) montre l'évolution de ces rapports pour des signaux de longueurs  $N$  dont  $\log_2 N$  varie de 8 à 18, pour  $a_{xy} = 1$ , et pour  $d = 100$  et  $d = 500$ . Ces courbes illustrent le besoin gigantesque en nombre d'échantillons pour obtenir des estimations correctes en grandes dimensions  $d$ . Dans les figures (2 bas) le coefficient de couplage  $a_{xy}$  varie de 0 à 0.5. Nous traçons la moyenne et une bande de plus ou moins deux écarts-types (intervalles de confiance à 95% en supposant les mesures gaussiennes), estimées sur 500 réalisations, pour  $d = 100$  et  $N = 2^{12}$  et  $N = 2^{14}$  échantillons. Cette figure illustre la sensibilité des indices et montre encore une fois la nécessité d'une statistique de très grande taille pour pouvoir tester la causalité dans des couplages faibles avec une confiance raisonnable. Et ce sur un exemple synthétique simple !

## 5 Pour conclure

L'objet de ce papier est d'étendre le concept de causalité de Granger aux cas des processus à valeurs fonctionnelles. Cette extension permet de traiter des cas cyclostationnaires ou non stationnaires pour lesquels la définition de la causalité de Granger n'est pas claire. Nous n'avons proposé que des définitions formelles et deux idées pour tester, sans aborder les problèmes pratiques fondamentaux, à savoir les problèmes de statistiques. Dans la suite, nous envisageons de développer la formalisation en se reposant sur l'inversibilité de la décomposition de Wold qui existe également dans le cas fonctionnel. Ceci doit permettre de placer le problème dans un cadre autorégressif dont nous avons traité le cas le plus simple.

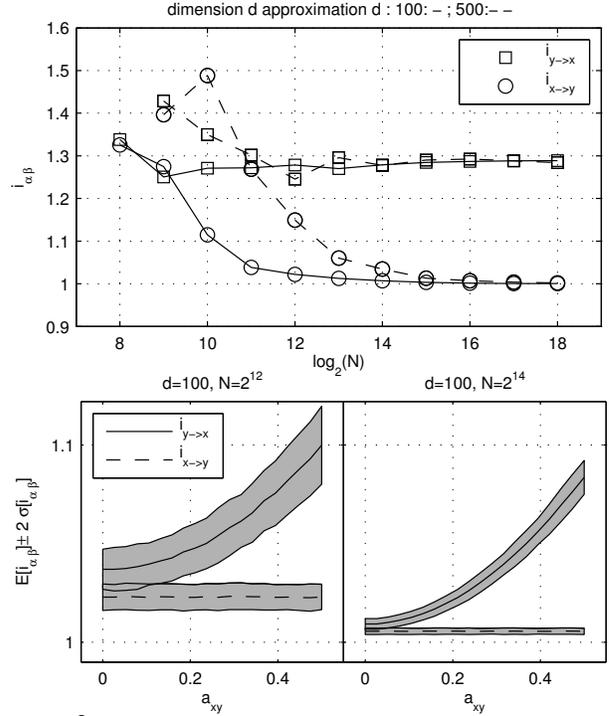


FIGURE 2 – Mesures  $i_{x \rightarrow y}$  et  $i_{y \rightarrow x}$  pour l'exemple traité. Haut : deux dimensions d'approximation, en fonction du nombre d'observations. Bas : en fonction du coefficient de couplage, pour  $2^{12}$  et  $2^{14}$  échantillons. Les bandes de confiance empiriques à 95% sous hypothèse gaussienne sont représentées en grisé.

## Références

- [1] N. Wiener. *Modern mathematics for the engineer*, chapter The theory of prediction, pages 165–190. MacGrawHill, 1956.
- [2] C. W. J. Granger. Economic processes involving feedback. *Information and control*, 6 :28–48, 1963.
- [3] C. W. J. Granger. Testing for causality : a personal viewpoint. *Journal of Economic Dynamics and Control*, 2 :329–352, 1980.
- [4] H. L. Hurd and A. Miamée. *Periodically correlated random sequences : Spectral theory and practice*. Wiley&Sons : Hoboken, NJ, USA, 2007.
- [5] R.M. Fortet. *Vecteurs, fonctions et distributions aléatoires dans les espaces de Hilbert (Random Vectors, functions and distributions in Hilbert spaces)*. (in French), Hermès, 1995.
- [6] P. O. Amblard and O. J. J. Michel. The relation between granger causality and directed information theory : A review. *Entropy*, 15(1) :113–143, 2013.
- [7] D. Bosq. *Linear processes in function spaces*. Springer : Berlin, Germany, 2000.
- [8] J. Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77 :304–313, 1982.
- [9] J. Geweke. Measures of conditional linear dependence and feedback between times series. *Journal of the American Statistical Association*, 79(388) :907–915, Dec. 1984.
- [10] D. Bosq and D. Blanke. *Inference and prediction in large dimensions*. John Wiley & Sons : Chichester, UK, 2007.
- [11] F. Ferraty and P. Vieu. *Nonparametric functional data analysis*. Springer, 2006.