The Functional LiNGAM

Tianle Yang yang@sigmath.es.osaka-u.ac.jp

Joe Suzuki prof.joe.suzuki@gmail.com

Graduate School of Engineering Science, Osaka University, Japan

Abstract

We consider a causal order such as the cause and effect among variables. In the Linear Non-Gaussian Acyclic Model (LiNGAM), we can only identify the order if at least one of the variables is non-Gaussian. This paper extends the notion of variables to functions (Functional Linear Non-Gaussian Acyclic Model, Func-LiNGAM). We first prove that we can identify the order among random functions if and only if one of them is a non-Gaussian process. In the actual procedure, we approximate the functions by random vectors. To improve the correctness and efficiency, we propose to optimize the coordinates of the vectors in such a way as functional principal component analysis. The experiments contain an order identification simulation among multiple functions for given samples. In particular, we apply the Func-LiNGAM to recognize the brain connectivity pattern with fMRI data. We can see the improvements in accuracy and execution speed compared to existing methods.

Keywords: Functional data, Darmois-Skitovich theorem, LiNGAM

1. Introduction

The application areas for causal discovery are immense, from its use in climate research (Ebert-Uphoff and Deng, 2012) to biomedical (Spirtes et al., 2000) and gene expression (Friedman et al., 2000). Suppose we have two variables, X, and Y, and wish to identify their cause and effect. Although there are many ways to detect the independence, there is no statistical way to determine the causal direction, i.e., either $X \to Y$ or $Y \to X$. Shimizu et al. (2006) proposed the LiNGAM to identify the causal directions when the relations are linear, i.e., $Y = aX + \epsilon$ and $X = a'Y + \epsilon'$, for some $a, a' \in \mathbb{R}$ and $X \perp\!\!\!\perp \epsilon$ and $Y \perp\!\!\!\perp \epsilon'$[1]. Moreover, they find the necessary and sufficient condition of identifiability: one of X, Y should be non-Gaussian. Note that for the Gaussian variables, zero correlation is equivalent to independence, which means that we cannot distinguish the two linear models if X, Y are Gaussian. Consequently, the identifiability of earlier work (Spirtes et al., 2000) in the setting of linear Gaussian fails even when satisfying the causal faithfulness (Pearl, 2000; Spirtes et al., 2000) without considering the latent confounders. For this reason, it is a significant advance that the LiNGAM can uniquely identify all the causal ordering from observational data, even without assuming faithfulness. Even if we have three variables X, Y, Z, it is not often possible to identify the direction of the DAG that connects them via any statistical procedure, such as $X \to Y \to Z$, $Z \to Y \to X$ and $Y \to X$, $Y \to Z$. They share the distribution $P(XY)P(YZ)/P(Y)$, which means that they are Markov equivalent. For the other direction of the proof, we use the Darmois-Skitovich theorem (D-S): if the

1. We write $X \perp\!\!\!\perp Y$ when X, Y are independent.

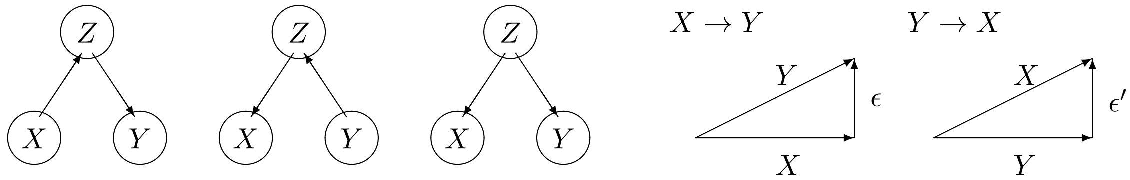


Figure 1: The three DAGs share the same distribution $P(XY)P(YZ)/P(Y)$. However, structure learning such as the PC algorithm cannot distinguish them (Left) but the LiNGAM can do based on the assumption that the noise being independent of the covariates (Right).

linear combinations $\sum_i \alpha_i X_i$ and $\sum_i \beta_i Y_i$ are independent, then X_i, Y_i are Gaussian for i such that $\alpha_i, \beta_i \neq 0$.

The original D-S theorem dealt with one-dimensional Gaussian random variables. Later, Ghurye and Olkin (1962) extended it into the case of random vectors. Myronyuk (2008) generalized the D-S into a Banach space. We call the Banach space valued random variables, conventionally as the random elements taking values in a Banach space, which we call random functions. This paper is interested in estimating a causal order in a more complicated setting.

We aim to identify the causal relationships among p random functions, from which we can sample the functional data. The functional data is fast gaining traction in a variety of fields, including genetics (Wei and Li, 2008), finance (Tsay and Pourahmadi, 2017), and neuroimaging (Luo et al., 2019). Investigating the directional links among the random functions is a significant challenge in multivariate functional data analysis. Our motivation is brain effective connection (Friston, 2001), which refers to the directional influence of one neural system on another. The Effective connectivity analysis estimates directional effects among brain regions through electrocorticographic imaging (ECoG) or task-based functional magnetic resonance imaging (fMRI). The imaging data is usually summarized in part by a time matrix for each individual. The columns represent time points, whereas the rows represent brain areas or locations. Given the continuous structure of the data and the short-time period between adjacent sample points, treating the data at each site as a function is an obvious choice. Modeling such multivariate processes and estimating the brain's effective connection between various areas is a significant challenge. The previous work, such as Qiao et al. (2019) considered the functional situation to calculate the inverse covariance matrix for the Gaussian graphical model. Lee and Li (2022) proposed the functional structure equation model with Gaussian noises, which can only determine the DAG up to its equivalence class under the linear setting. This paper presents the model to estimate the directional relationships among the functional data with non-Gaussian errors. We name it Functional LiNGAM (Func-LiNGAM). The Func-LiNGAM inherits the advantages of LiNGAM without the necessity of faithfulness.

Our contributions are as follows:

• We present a framework for identifying causal orders of random functions.

• We show the identifiability of causal relationships if and only if one of the random functions is non-Gaussian.

• Empirically, the results of Functional LiNGAM can identify the causal ordering among the non-Gaussian random functions. The Func-LiNGAM is not only able to identify the correct solution as DirectLiNGAM and it is much faster than the original approach DirectLiNGAM (Shimizu et al., 2011) because of using the FPCA (Ramsay and Silverman, 2005). The FPCA is a dimension reduction approach to approximate each infinite-dimensional function by a finite representation. We also consider an application to the fMRI dataset.

We organize this paper as follows. Section 2 introduces the background required for understanding the current paper, such as the LiNGAM, Hilbert spaces, and random functions. Section 3 proves the main theorem (extension of the LiNGAM) and describes the procedure. Section 4 gives experiments of the proposed procedure. Section 5 summarizes the paper and states future works.

2. Preliminaries

2.1 LiNGAM

We introduce the LiNGAM (Linear non-Gaussian acyclic model) that infers the causal order among variables.

Suppose that we have random variables X, Y that take values on $\mathbb{R}$, and that we wish to identify the causal order such as $X \to Y$ or $Y \to X$. More precisely, we assume that X, Y are linearly related and have zero averages, and that we have two models

$$Y = aX + \epsilon \ , \tag{1}$$

$$X = a'Y + \epsilon' \tag{2}$$

to choose from, where $a, a' \in \mathbb{R}$ and ϵ, ϵ' are zero-mean random variables. For simplicity, we assume

$$a \neq 0 \ , \ \text{or } a' \neq 0 \ , \tag{3}$$

to exclude the trivial cases, which means that X and Y are not independent. In particular, LiNGAM assumes that the noise such as ϵ and ϵ' should be independent of the covariate such as X and Y in (1) and (2), respectively. Thus, we identify the true model as (1) or (2) depending on[2] $X \perp\!\!\!\perp \epsilon$ or $Y \perp\!\!\!\perp \epsilon'$.

One might think that we cannot distinguish (1) and (2), i.e. X, Y may satisfy both of (1) and (2) for some a, a' and ϵ, ϵ' such that $X \perp\!\!\!\perp \epsilon$ and $Y \perp\!\!\!\perp \epsilon'$. LiNGAM claims that such an inconvenience occurs if and only if X, Y are Gaussian.

Suppose X, Y are Gaussian, and that (1) with $X \perp\!\!\!\perp \epsilon$ for some a and ϵ is true. Let $e_1 := X$, $e_2 := \epsilon$, $e_1' := Y$, $e_2' := \epsilon'$, and σ_1^2, σ_2^2 be the variances of e_1 and e_2. Then, from $\mathbb{E}[e_1 e_2] = 0$, we have

$$e_1' = ae_1 + e_2 \tag{4}$$

$$e_2' = e_1 - a'e_1' = e_1 - a'(ae_1 + e_2) = (1 - aa')e_1 - a'e_2 \ , \tag{5}$$

and $\mathbb{E}[e_1' e_2'] = a(1 - aa')\sigma_1^2 - a'\sigma_2^2$, which means that choosing

$$a' = \frac{a\sigma_1^2}{a^2\sigma_1^2 + \sigma_2^2} \tag{6}$$

2. We write $X \perp\!\!\!\perp Y$ when X, Y are independent.

makes $\mathbb{E}[e_1' e_2']$ zero as well. We say that variables Z, W are jointly Gaussian if they can be expressed by

$$\begin{bmatrix} Z \\ W \end{bmatrix} = A \begin{bmatrix} U \\ V \end{bmatrix}$$

for some $A \in \mathbb{R}^{2\times 2}$ and independent Gaussian U, V. It is known that zero-correlation is equivalent to independence for jointly Gaussian variables[3]. We can check that e_1', e_2' are jointly Gaussian. Therefore, We have $e_1' \perp\!\!\!\perp e_2'$ as well. Thus, we have (2) with $Y \perp\!\!\!\perp \epsilon'$ for the a' and ϵ'.

On the other hand, suppose that (1) with $X \perp\!\!\!\perp \epsilon$ and (2) with $Y \perp\!\!\!\perp \epsilon'$ occur simultaneously for some a, a' and ϵ, ϵ', which requires (6). Thus, the assumption implies that none of a, a' are zero due to (6). We notice the following statement:

Proposition 1 (Darmois (1953); Skitivic (1953)) Let $n \geq 2$ and $\xi_1, \ldots, \xi_n$ independent random variables that take values on $\mathbb{R}$. Suppose that $\sum_{i=1}^m \alpha_i \xi_i$ and $\sum_{i=1}^m \beta_i \xi_i$ are independent for some $\alpha_1, \ldots, \alpha_m$, $\beta_1, \ldots, \beta_m \in \mathbb{R}$. Then, ξ_j such that $\alpha_j \beta_j \neq 0$ is Gaussian for $j = 1, \ldots, m$.

Authors have extended the DS theorem into several cases, such as when the random variables are replaced by random functions in a Banach space (Myronyuk, 2008) and random vectors (Ghurye and Olkin, 1962).

Comparing (4)(5) and Proposition 1, we see that each term in

$$(e_1, e_2, a, 1, 1-aa', -a') = (X, \epsilon, a, 1, \frac{\sigma_2^2}{a^2\sigma_1^2 + \sigma_2^2}, -\frac{a\sigma_1^2}{a^2\sigma_1^2 + \sigma_2^2}) = (\xi_1, \xi_2, \alpha_1, \alpha_2, \beta_1, \beta_2) \quad (7)$$

corresponds, where we have applied (6). Combining with (3), we find that X, ϵ are Gaussian, which means that Y is Gaussian as well.

Proposition 2 (Shimizu et al. (2011)) Suppose we assume either $a \neq 0$ in (1) or $a' \neq 0$ in (2). We can identify the order between random variables via the LiNGAM if and only if one of them is non-Gaussian.

We may apply the above inference for the multivariate causal order identification. Suppose we have linearly related variables X, Y, Z and have zero mean. Then, there are six orders, such as $Y \to X \to Z$, and $Z \to Y \to X$. We first find the top variable among the three. Suppose that X and $\{Y - aX, Z - a'X\}$ are independent for some $a, a' \in \mathbb{R}$. Then, we regard X as the top. Suppose also that $Y - aX$ and $Z - a'X - a''(Y - aX)$ are independent for some $a'' \in \mathbb{R}$. Then, we regard Y and Z as the middle and bottom variables, respectively, which means $X \to Y \to Z$. In such a way, we can determine the order among X, Y, Z. Extending the notion, we can determine the order among any number of variables such as

$$X_i = \sum_{j=1}^{i-1} b_{i,j} X_j + e_i$$

with non-Gaussian noise e_i and $b_{i,j} \in \mathbb{R}$ for p variables $X_1, \ldots, X_p$.

3. Let Z and W be zero-mean Gaussian and binary taking ± 1 equiprobably. Then, Z and ZW are Gaussian but not jointly Gaussian. The correlation $\mathbb{E}[Z \cdot ZW] = \mathbb{E}[Z^2] \cdot \mathbb{E}[W] = 0$ but they are not independent.

2.2 Hilbert Spaces

In the following, we regard the set of functions that we deal with makes a Hilbert space[4].

We say that $T_{21} : H_1 \to H_2$ is a linear operator over $\mathbb{R}$ when $T(\alpha f + \beta g) = \alpha T f + \beta T g$ for $f, g \in H_1$ and $\alpha, \beta \in \mathbb{R}$, and bounded when there exists $C > 0$ such that $\|T_{21} f\|_2 \leq C \|f\|_1$ for $f \in H_1$, where $\|\cdot\|_1$ and $\|\cdot\|_2$ are the norms in H_1 and H_2, respectively. For each bounded operator $T_{21} : H_1 \to H_2$, there exists a unique bounded and linear operator $T_{21}^* : H_2 \to H_1$ such that $\langle T_{21} f_1, f_2 \rangle_2 = \langle f_1, T_{21}^* f_2 \rangle_1$ for $f_1 \in H_1$ and $f_2 \in H_2$. We call such an operator T_{21}^* the adjoint operator of T_{21}, and if $T_{21} = T_{21}^*$, we say that T_{21} is self-adjoint, and it is symmetric in particular if the dimension of H is finite.

2.3 Random functions

Formally speaking, we say that $X : \Omega \to \mathbb{R}$ is a random variable if it is measurable from a probability space $(\Omega, \mathcal{F}, \mu)$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

$$\{\omega \in \Omega | X(\omega) \in \mathcal{B}(\mathbb{R})\} \in \mathcal{F} \ ,$$

where $\mathcal{B}(\mathbb{R})$ is the whole Borel sets. Similarly, we say that $\chi : \Omega \to H$ is a random function in a Hilbert space H if it is measurable from $(\Omega, \mathcal{F}, \mu)$ to $(H, \mathcal{B}(H))$:

$$\{\omega \in \Omega | \chi(\omega) \in \mathcal{B}(H)\} \in \mathcal{F} \ ,$$

where $\mathcal{B}(\mathcal{H})$ is the whole Borel sets w.r.t. the norm of H. Letting E be a set, we assume that each element f of H is a function $f : E \ni x \mapsto f(x) \in \mathbb{R}$.

We define the mean function and covariance operator of a random function $\chi : \Omega \to H$. If the expectation of $\|\chi\|$ is bounded, we define the mean of χ by the Bochner integral[5] $\int_\Omega \chi d\mu$. If the means of χ_1, χ_2 in H are m, then we define the covariance operator $\mathscr{K} : H \to H$ of random functions χ_1, χ_2 when $H := H_1 = H_2$ by

$$\langle \mathscr{K} g_1, g_2 \rangle = \langle \int_\Omega \langle \chi_1 - m, g_1 \rangle (\chi_2 - m) \rangle d\mu, g_2 \rangle = \int_\Omega \langle \chi_1 - m, g_1 \rangle \langle \chi_2 - m, g_2 \rangle d\mu \ ,$$

for $g_1, g_2 \in H$. Let $\{e_i\}$ be an orthonormal basis of H. Then, we can obtain the covariance values $\langle \mathscr{K} e_i, e_j \rangle$ for each i, j. In general, if the random functions χ_1 and χ_2 are independent, then $\langle \mathscr{K} \ g_1, g_2 \rangle = 0$ for $g_1, g_2 \in H$.

Note that random function $\chi : \Omega \to H$ takes values $\chi(\omega, x) \in \mathbb{R}$ with $\omega \in \Omega$ and $x \in E$ if each element in H is $E \to \mathbb{R}$. If we fix $\omega \in \Omega$, then $\chi(\omega, \cdot)$ is a function $E \to \mathbb{R}$. Hereafter, we abbreviate the random function $\chi(\omega, \cdot)$ by $\chi(\cdot)$, which is similar to write a random variable $X(\omega)$ as X. We say that a random function χ with mean m is a Gaussian process if the random vector $[\chi(x_1), \ldots, \chi(x_n)]$ of length n is Gaussian with mean $[m(x_1), \ldots, m(x_n)]$ for any $n \geq 1$ and $x_1, \ldots, x_n \in E$.

If H is of (finite) dimension d, then the covariance operator becomes the covariance matrix expressed by a positive definite matrix $\Sigma \in \mathbb{R}^{d \times d}$. Then, we can define the eigen-values $\{\lambda_i\}$ and eigen-vectors $\{\phi_i\}$ of Σ. In particular, we can express each vector by

4. The Banach space is a linear space with a norm that gives completeness (any Cauchy sequence converges). The Hilbert space is a Banach space with an inner-product that induces the norm that gives completeness.
5. For the exact definition of the Bochner integral, see the reference HSING and EUBANK (2015).

$\sum_{i=1}^{d}\langle X,\phi_i\rangle\phi_i$, and see that the variance of $\langle X,\phi_i\rangle$ is λ_i. Similarly, we have the following statement: if H is a function space (of an infinite dimension),

Proposition 3 (HSING and EUBANK (2015)) *Suppose that $\{\lambda_i\}$ and $\{\phi_i\}$ are the eigen-values and eigen-functions obtained via $\mathscr{K}\phi_i = \lambda_i\phi_i$ for $i = 1, 2, \ldots$. Then, the random function χ is expressed by*

$$\chi = m + \sum_{i=1}^{\infty}\langle\chi,\phi_i\rangle_H\phi_i$$

with probability one, where the mean and variance of $\langle\chi,\phi_i\rangle_H$ are zero and λ_i.

Note that the notion of stochastic processes is close to that of random functions: $\{X(t)\}_{t\in E}$ is a stochastic process if $X : \Omega\times E \to \mathbb{R}$ is measurable from $(\Omega,\mathcal{F},\mu)$ to $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ for each $t \in E$. Some stochastic processes are random functions as well (HSING and EUBANK, 2015).

3. Functional LiNGAM

This section extends the LiNGAM of random variables to that of the random functions.

3.1 Identifiability

In this subsection, we show that the order identification is possible via the LiNGAM if and only if one of the random functions is not a Gaussian process.

To this end, We shall establish the problem of the LiNGAM in functional spaces (Hilbert spaces). Let H_1, H_2 be Hilbert spaces. Suppose that we have two models $f_1 \in H_1$ and $f_2 \in H_2$,

$$\begin{aligned} f_1 &= h_1, \qquad f_2 = T_{21}f_1 + h_2,\\ f_2 &= h_1', \qquad f_1 = T_{12}f_2 + h_2'. \end{aligned} \tag{8}$$

where h_1, h_2' and h_1', h_2 are random functions of H_1 and H_2, respectively, and $T_{12} : H_2 \to H_1$, $T_{21} : H_1 \to H_2$ are linear bounded operators bewteen H_1, H_2, and that we wish to identify the order by examining whether $h_2 \perp\!\!\!\perp f_1$ or $h_1 \perp\!\!\!\perp f_2$.

Before proceeding with our discussion, we confirm two statements for deriving the claim.

- the equivalence between non-correlation and independence for jointly Gaussian random functions in a Hilbert space (Proposition 4), and
- the DS theorem for random functions in Hilbert spaces (Proposition 5).

The following (Proposition 4) claims the equivalence for the Banach spaces that contain the Hilbert spaces. Let χ_1, χ_2 be random functions that randomly take functions $E \to \mathbb{R}$ in Banach spaces B_1, B_2. Note that the dimension of the linear space is infinity when E is an infinite set while the dimension is finite for vectors of finite dimension.

Proposition 4 (van Neerven (2020)) *Let χ, χ' be jointly Gaussian random functions in Banach spaces. Then, χ, χ' are independent if and only if they are uncorrelated.*

We say that a bounded linear oprator $T : H_1 \to H_2$ is continuous if the set $\{T(f)|f \in U\} \subseteq H_2$ is open for a subset $U \subseteq H_1$, so is the inverse image U, and that $T : H_1 \to H_2$ is invertible if it is one-to-one.

Proposition 5 (Darmois-Skitovich in Banach spaces (Myronyuk, 2008)) *Let $n \geq 2$ and $\xi_1, \ldots, \xi_n$ are random functions in a Banach space. Suppose that $\sum_{i=1}^n A_i\xi_i$ and $\sum_{i=1}^n B_i\xi_i$ are independent for some continuous linear bounded operators $A_1, \ldots, A_n$, and $B_1, \ldots, B_n$. Then, ξ_j such that A_j, B_j are invertible is a Gaussian process for $j = 1, \ldots, n$.*

Based on Propositions 4 and 5 above, we derive the main theorem as below.

Theorem 6 *Suppose that either T_{12} or T_{21} is invertible. We can identify the order between random functions via the LiNGAM if and only if one of them is not a Gaussian process.*

Proof. From (8), we have

$$\begin{aligned} h'_1 &= f_2 = T_{21}h_1 + h_2 \\ h'_2 &= f_1 - T_{12}f_2 = h_1 - T_{12}(T_{21}h_1 + h_2) = (I - T_{12}T_{21})h_1 - T_{12}h_2 \ . \end{aligned} \tag{9}$$

then h'_1, h'_2 are jointly Gaussian due to the linear combinations of independent Gaussian random functions van Neerven (2020). Then the zero-correlation means independence from Proposition 4. Since $h_1 \perp\!\!\!\perp h_2$ and $h_1 \in H_1$, $h_2 \in H_2$, the cross-covariance operator $\mathscr{K}_{12}$ between them is zero: $\langle \mathscr{K}_{12}g_1, g_2\rangle_{H_2} = \int_\Omega \langle h_1, g_1\rangle_{H_1}\langle h_2, g_2\rangle_{H_2} = 0$ for any $g_1 \in H_1$,$g_2 \in H_2$. Then, the cross-covariance operator $\mathscr{K}'_{12}$ between h'_1 and h'_2 is

$$\begin{aligned} \langle \mathscr{K}'_{12}g_1, g_2\rangle_{H_2} &= \int_\Omega \langle (I - T_{12}T_{21})h_1 - T_{12}h_2, g_1\rangle_{H_1}\langle T_{21}h_1 + h_2, g_2\rangle_{H_2} d\mu \\ &= \int_\Omega \langle (I - T_{12}T_{21})h_1, g_1\rangle_{H_1}\langle T_{21}h_1, g_2\rangle_{H_2} d\mu + \int_\Omega \langle -T_{12}h_2, g_1\rangle_{H_1}\langle h_2, g_2\rangle_{H_2} d\mu \\ &= \int_\Omega \langle h_1, (I - T_{12}T_{21})^* g_1\rangle_{H_1}\langle h_1, T^*_{21}g_2\rangle_{H_1} d\mu - \int_\Omega \langle h_2, T^*_{12}g_1\rangle_{H_2}\langle h_2, g_2\rangle_{H_2} d\mu \\ &= \langle \mathscr{K}_{11}(I - T_{12}T_{21})^* g_1, T^*_{21}g_2\rangle_{H_1} - \langle \mathscr{K}_{22}T^*_{12}g_1, g_2\rangle_{H_2} \\ &= \langle T_{21}\mathscr{K}_{11}(I - T^*_{21}T^*_{12})g_1, g_2\rangle_{H_2} - \langle \mathscr{K}_{22}T^*_{12}g_1, g_2\rangle_{H_2} \end{aligned} \tag{10}$$

for any $g_1 \in H_1$,$g_2 \in H_2$, where $\mathscr{K}_{11}, \mathscr{K}_{22}$ are the covariance operators of h_1, h_2, respectively. We assume that $\mathscr{K}_{11}, \mathscr{K}_{22}$ are not zero. If $\mathscr{K}'_{12} = 0$, then we require

$$\mathscr{K}_{11}T^*_{21} = T_{12}\{T_{21}\mathscr{K}_{11}T^*_{21} + \mathscr{K}_{22}\} \ . \tag{11}$$

In fact, we have

$$(10) = 0 \Leftrightarrow T_{21}\mathscr{K}_{11}(I - T^*_{21}T^*_{12}) = \mathscr{K}_{22}T^*_{12} \Leftrightarrow T_{21}\mathscr{K}_{11} = (T_{21}\mathscr{K}_{11}T^*_{21} + \mathscr{K}_{22})T^*_{12} \Leftrightarrow (11)$$

Then, we claim that there exists T_{12} that satisfies (11). It is sufficient to show $(T_{21}\mathscr{K}_{11}T^*_{21} + \mathscr{K}_{22}\}f = 0 \Longrightarrow \mathscr{K}_{11}T^*_{21}f = 0$ for $f \in H_2$. In order to show it, we note that there exists a unique nonnegative linear bounded operator $S : H_1 \to H_1$ such that $S^2 = \mathscr{K}_{11}$ (Theorem 3.4.3 in HSING and EUBANK (2015)) since $\mathscr{K}_{11}$ is nonnegative. Thus, we have

$$\begin{aligned} &(T_{21}\mathscr{K}_{11}T^*_{21} + \mathscr{K}_{22})f = 0 \Longrightarrow \langle (T_{21}\mathscr{K}_{11}T^*_{21} + \mathscr{K}_{22})f, f\rangle_{H_2} = 0 \\ \Longrightarrow \ &\langle T_{21}\mathscr{K}_{11}T^*_{21}f, f\rangle_{H_2} = 0 \Longrightarrow ST^*_{21}f = 0 \Longrightarrow \mathscr{K}_{11}T^*_{21}f = S \cdot ST^*_{21}f = 0 \ . \end{aligned}$$

For the converse, suppose that $h_1 \perp\!\!\!\perp h_2$ and $h'_1 \perp\!\!\!\perp h'_2$ in (8) with occur simultaneously for some T_{12},T_{21}, and that we wish to derive under (11) that h_1, h_2, h'_1, h'_2 are Gaussian. Because a Hilbert space is a Banach space, we apply Proposition 5. Without loss of generality, we assume that T_{12} is invertible. We show that the eigenvalue of $T_{12}T_{21}$ is less than 1, which means that $I - T_{12}T_{21}$ is invertible (Theorem 3.5.5 in HSING and EUBANK (2015)). To this end, if we multiply (11) by T_{21} from the left, we have

$$T_{21}\mathscr{K}_{11}T_{21}^* = T_{21}T_{12}\{T_{21}\mathscr{K}_{11}T_{21}^* + \mathscr{K}_{22}\} ,$$

which means that the eigenvalue of $T_{21}T_{12}$ is less than 1. Noting that $T_{21}T_{12}$ and $T_{12}T_{21}$ share the eigenvalues:

$$T_{21}T_{12}u = \lambda u \Longrightarrow T_{12}T_{21}T_{12}u = \lambda T_{12}u \Longrightarrow T_{12}T_{21}v = \lambda v$$

for $\lambda \neq 0$, $u \in H_2$, and $v := T_{12}u \in H_1$, we have proved that the eigenvalue of $T_{12}T_{21}$ is less than 1. Then, as we did in (7), we correspond

$$(h_1.h_2, T_{21}, I, I - T_{21}T_{12}, -T_{12}) = (\xi_1, \xi_2, A_1, A_2, B_1, B_2)) \ ,$$

where A_1, A_2, B_1, B_2 are invertible. □

3.2 The Procedure

We consider the first model in (8):

$$f_2 = T_{21}f_1 + h_2 \ , \tag{12}$$

and notice the following statement:

Proposition 7 (HSING and EUBANK (2015)) *Let the $T : H_1 \to H_2$ be a compact*[6] *bounded linear operator. Let $\{e_{1,j}\}$ and $\{e_{2,j}\}$ be the orthonormal eigenvector sequences of T^*T and TT^*, respectively, and $\{\lambda_j\}$ the eigenvalues of them. Then, we can express the T by*

$$Tf = \sum_{j=1}^{\infty} \lambda_j \langle f, e_{1,j}\rangle_{H_1} e_{2,j}$$

for $f \in H_1$.

Using the same notation as Proposition 7, suppose we express $f_1 = \sum_{j=1}^{\infty} f_{1,j}e_{1,j}$, $f_2 = \sum_{j=1}^{\infty} f_{2,j}e_{2,j}$, $h_2 = \sum_{j=1}^{\infty} h_{2,j}e_{2,j}$. Then, (12) becomes the following relation:

Theorem 8 *Suppose that $T_{21} : H_1 \to H_2$ is compact. If we choose the bases of H_1 and H_2 as $\{e_{1,j}\}$ and $\{e_{2,j}\}$, respectively, then we have*

$$f_{2,j} = \lambda_j f_{1,j} + h_{2,j} \tag{13}$$

for $j = 1, 2, \ldots$, where $\lambda_1 \geq \lambda_2 \geq \cdots$.

6. We say a bounded linear operator $T : H_1 \to H_2$ is compact if $\{Tf_n\}$ has a convergent subsequence in H_2 for any bounded infinite sequence $\{f_n\}$ in H_1.

We assumed that the T_{21} is compact, otherwise the eigenvalue sequence $\{\lambda_j\}$ would not converge. In reality, we approximate the random functions $f_1 \in H_1,\ f_2, h_2 \in H_2$ by random vectors of finite length M. We choose the $\{e_{1,j}\}_{j=1}^M$ and $\{e_{2,j}\}_{j=1}^M$ for the bases, so that we minimize the approximation loss.

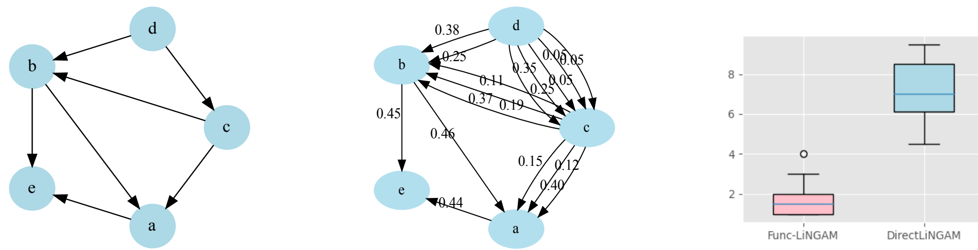
The other merit of using the FPCA (functional principal component analysis) approach is its efficiency. We assume the following procedure: first, we approximate the functions by the L coefficients of the basis functions (B-spline). Then, we transform it by the M coefficients of the basis functions defined above. As observed in the next section, $M \ll L$ and the time complexity $C(M)$ of the proposed procedure is much less than $C(L)$ for the B-spline. For example, Shimizu et al. (2011) evaluated the complexity of their method as $C(L) = O(n(Lp)^3q^2 + (Lp)^4q^3)$, where q ($\ll n$) is the maximal rank found by the low-rank decomposition used in the kernel-based independence measure, although the proposed procedure requires additional $O(nL^2 + L^3)$ complexity for computing the covariance matrix $O(nL^2)$ and eigenvalue decomposition $O(L^3)$. Extending the notion, we can determine the order among any number of random functions such as

$$f_i = \sum_{j=1}^{i-1} T_{i,j} f_j + h_i$$

with non-Gaussian h_i and bounded linear operators $T_{i,j}; H_j \to H_i$ for p random functions $f_1 \in H_1, \ldots, f_p \in H_p$.

4. Simulation

For the synthetic data, we follow the setting in Qiao et al. (2019) by generating $n \times p$ functional variables via $X_{ij}(t) = \phi(t)^T \delta_{ij}$, where $\phi(t)$ is a five-dimensional Fourier basis function, and $\delta_{ij} \in \mathbb{R}^5$ is a mean zero non-Gaussian random vector. The δ_{ij} can belong to any non-Gaussian distribution, we let them be the $\mathcal{N}(\mathbf{0}, I_5)^2$ here. Hence, $\delta_i = \left(\delta_{i1}^T, \ldots, \delta_{ip}^T\right)^T \in \mathbb{R}^{5p}$ followed from a multivariate non-Gaussian distribution. We generate $n = 1000$ observations of δ_i, and the observed values, g_{ijk}, were sampled using $g_{ijk} = X_{ij}(t_k) + e_{ijk}$, where $i = 1, \ldots, n$, $j = 1, \ldots, p$ and e_{ijk} belongs to non-Gaussian distribution with mean 0 and variance 0.25. Each function was observed at $T = 100$ equally spaced time points, $0 = t_1, \ldots, t_{100} = 1$. We denote the five random functions as $\{a, b, c, d, e\}$, and the true causal ordering of them is in Figure 2(a). To mimic real data, we fit each function using a L-dimensional B-spline basis (rather than the Fourier basis used to generate the data). Then, we compute the first M estimated principal component scores of X_{ij} from the L basis coefficients ($M \leq L$). We choose $L = 50$ for the B-spline and $M = 5$ (explained variance ratio greater than 99%). Actually, we can choose the optimal L and M by using cross-validation. Here, we choose the parameters which can contain the information as most as possible. The causal result of the principal component scores is in Figure 2(b), where $M = 5$. By comparing Figure 2(a) and Figure 2(b), we find the Func-LiNGAM successfully recovers the causal order of the five random functions. Moreover, we compare the running time of Func-LiNGAM and DirectLiNGAM for the functional data in Table 1. Due to the FPCA, the Func-LiNGAM is much faster than the DirectLiNGAM ($M \ll T$). Whereas the Direct-LiNGAM faces all the number T of time points. We also compare the Func-LiNGAM



(a) True order. (b) Simulated order. (c) SHD (T=50)

Figure 2: The left is the true causal order of five functions, the middle is the simulated order from Func-LiNGAM, and the right is the Structural Hamming Distances (SHD) of two models.

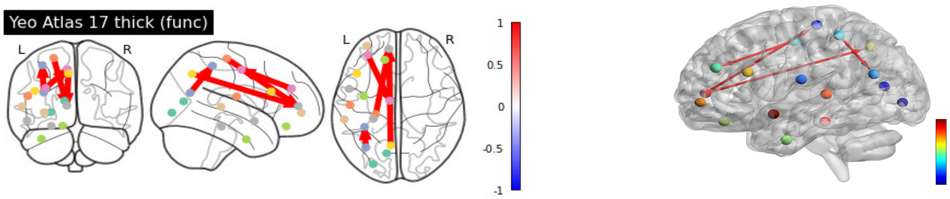
and DirectLiNGAM by computing the Structural Hamming Distance (SHD) (Tsamardinos et al., 2006) of the estimated DAG to the true one in 10 times. In simple terms, this is the number of edge insertions, deletions, or flips to transform one graph into another. Thus the smaller the SHD, the more precise it is. From Figure 2(c), we find the Func-LiNGAM performs better than the DirectLiNGAM.

	T=10	T=30	T=50	T=70
Func-LiNGAM	2.81	3.29	3.19	3.23
DirectLiNGAM	16.09	417.61	1940.87	5689.06

Table 1: CPU (i7-12700KF) running time (in seconds) for $p = 5$.

5. Real Data

This section illustrates the application of our method to a brain connectome analysis using fMRI. The data (Richardson et al., 2018) is downsampled to 4mm resolution for convenience with a repetition time (TR) of 2 secs. The dataset contains 155 subjects ($n = 155$), 17 parcels ($p = 17$), and 168 time points ($T = 168$), where 122 children and a reference group of 33 adults, watched a short, animated movie that included events evoking the mental states and physical sensations of the characters, while undergoing fMRI. To make sure the data has non-Gaussian elements, we use Henze-Zirkler Multivariate Normality Test (Henze and Zirkler, 1990) for the estimated principal component scores. The p-values of the test of 17 parcels (we denote them as $1, \ldots, 17$) are almost less than 0.05 except two parcels $(13, 17)$, which means most of them reject the null hypothesis and they are non-Gaussian. Then We estimate the adjacent matrix between the parcels with $M = 5$. The adjacent matrix has values at $5 \to 6$, $16 \to 3$, $16 \to 11$, $7 \to 8$. We also give the 2D graph of brain connectivity of the causal relationships by Python package Nilearn in Figure 3(a), and the 3D graph in 3(b) with the BrainNet Viewer (Xia et al., 2013).



(a) 2D. (b) 3D.

Figure 3: 2D and 3D Brain connectivity graphs.

6. Conclusion

We proposed a new framework (Func-LiNGAM) for identifying causal relationships of random functions. Theoretically, We proved the identifiability of non-Gaussian processes in Theorem 6 and proposed the approximation of random functions by random vectors in a FPCA manner. The simulation results show that the proposed procedure can identify the causal order among the non-Gaussian random functions precisely and quickly. Moreover, we successfully identified the brain connectivity architecture from fMRI data with Func-LiNGAM. In future work, we need to consider the optimum estimation given a finite number of functions. Furthermore, exploring more applications of Func-LiNGAM will make us excited.

References

G. Darmois. Analyse generale des liaisons stochastiques. *Rev. Inst. Intern. Stat*, 21:2–8, 1953.

I. Ebert-Uphoff and Y. Deng. Causal Discovery for Climate Research Using Graphical Models. *Journal of Climate*, 25(17):5648–5665, 2012.

N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using Bayesian Networks to analyze expression data. *J. Comput. Biol.*, 7:601–620, 2000.

K. J. Friston. Functional and effective connectivity: A review. *Brain Connectivity*, 1(1): 13–36, 2001.

S. Ghurye and I. Olkin. A characterization of the multivariate normal distribution. *Ann. Math. Statist.*, 33(2):533–541, June 1962.

N. Henze and B. Zirkler. A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods*, 19(10):3595–3617, 1990.

T. HSING and R. EUBANK. *Theoretical Foundations of Functional Data Analysis,with anIntroduction to Linear Operators.* Wiley, Chichester, 2015.

K. Y. Lee and L. Li. Functional structural equation model. *Journal of the Royal Statistical Society: Series B*, 84(2), 2022.

S. Luo, R. Song, M. Styner, J. Gilmore, and H. Zhu. FSEM: Functional structural equation models for twin functional data. *Journal of the American Statistical Association*, 114 (525):344–357, 2019.

M. V. Myronyuk. On the Skitovich-Darmois theorem and Heyde theorem in a Banach space. *Ukr Math J*, 60:1437–1447, 2008.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge, U.K.: Cambridge University Press, 2000.

X. Qiao, S. Guo, and G. M. James. Functional Graphical Models. *J. Amer. Statist. Assoc.*, 114(525):211–222, 2019.

J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, 2005.

H. Richardson, G. Lisandrelli, A. Riobueno-Naylor, and R. Saxe. Development of the social brain from age three to twelve years. *Nature Communications*, 9(1):1027, 2018.

S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen. A Linear Non-Gaussian Acyclic Model for Causal Discovery. *Journal of Machine Learning Research*, 7(72):2003–2030, 2006.

S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. O. Hoyer, and K. Bollen. DirectLiNGAM: A Direct Method for Learning a Linear Non-Gaussian Structural Equation Model. *Journal of Machine Learning Research*, 12(33):1225–1248, 2011.

V. P. Skitivic. On a property of the normal distribution. *Dokl. Akad. Nauk SSSR (N.S.)*, (89):217–219, 1953.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. The MIT Press, 2000.

I. Tsamardinos, L. Brown, and C. Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Mach Learn*, 65:31–78, 2006.

R. Tsay and M. Pourahmadi. Modelling structured correlation matrices. *Biometrika*, 104 (1):237–242, 2017.

J. van Neerven. Stochastic Evolution Equations, 3 2020. URL `https://ocw.tudelft.nl/courses/stochastic-evolution-equations/subjects/lecture-4-gaussian-random-variables/`.

Z. Wei and H. Li. A hidden spatial-temporal markov random field model for network-based analysis of time course gene expression data. *The Annals of Applied Statistics*, 2(1): 408–429, 2008.

M. Xia, J. Wang, and Y. He. BrainNet Viewer: A network visualization tool for human brain connectomics. *PLoS ONE*, 8(7):e68910, 2013.