# Generalized Hoeffding Decomposition

## and the (surprising) linear nature of non-linearities


[1]**Université du Québec à Montréal** (UQÀM)
[2]**Institut Intelligence et Données** (IID) - Universite Laval
[3]**EDF R&D** - PRISME Department
[4]**Institut de Mathématiques de Toulouse**
[5]**SINCLAIR AI Lab**


***CRM-ISM Montreal Probability seminar***
*Burnside Hall - Montréal, QC, Canada.*
*September 12, 2024*


Marouane Il Idrissi[12(345)], Nicolas Bousquet[34], Fabrice Gamboa[4], Bertrand Iooss[345], Jean-Michel Loubes[4]

# Hi!

## Marouane IL IDRISSI

ilidrissi.m@gmail.com - marouaneilidrissi.com

☞ **Before:** Ph.D. Candidate (2021-2024)

**EDF R&D - Institut de Mathématiques de Toulouse**

Nicolas Bousquet, Fabrice Gamboa, Bertrand Iooss, Jean-Michel Loubes

*Interpretability methods for certifying machine learning models applied to critical systems*

# Hi!

## Marouane IL IDRISSI

ilidrissi.m@gmail.com - marouaneilidrissi.com

☞ **Before:** Ph.D. Candidate (2021-2024)

**EDF R&D - Institut de Mathématiques de Toulouse**

Nicolas Bousquet, Fabrice Gamboa, Bertrand Iooss, Jean-Michel Loubes

*Interpretability methods for certifying machine learning models applied to critical systems*

☞ **Now:** Postdoctoral Researcher (2024-2026+)

**UQÀM - IID** (ULaval)

Arthur Charpentier (UQÀM), Marie-Pier Côté (ULaval)

*Interpretability, fairness and causal inference of black-box models*

**(Some) Topics of interest:**

*XAI* ● *Uncertainty quantification* ● *Sensitivity analysis* ● *Industrial risks* ● *Probabilistic modelling* ● *Statistics* ● *Statistical Learning* ● *Applied cooperative game theory* ● *Functional analysis*

# Introduction

☞ Being able to **understand how black-box models behave under uncertainty** is crucial for **integrating AI models to critical systems**

Industrial processes, energy production, medicine, cars, planes, the stock market...

# Introduction

☞ Being able to **understand how black-box models behave under uncertainty** is crucial for **integrating AI models to critical systems**

Industrial processes, energy production, medicine, cars, planes, the stock market...

☞ The decision-making process must be **justifiable** and **justified**

**Empirical statements** (e.g. SOTA) are notoriously **not enough** to convince **safety/regulation authorities**

# Introduction

☞ Being able to **understand how black-box models behave under uncertainty** is crucial for **integrating AI models to critical systems**

Industrial processes, energy production, medicine, cars, planes, the stock market...

☞ The decision-making process must be **justifiable** and **justified**

**Empirical statements** (e.g. SOTA) are notoriously **not enough** to convince **safety/regulation authorities**

**We must rely on theoretical arguments**

# Introduction

☞ Being able to **understand how black-box models behave under uncertainty** is crucial for **integrating AI models to critical systems**

Industrial processes, energy production, medicine, cars, planes, the stock market...

☞ The decision-making process must be **justifiable** and **justified**

**Empirical statements** (e.g. SOTA) are notoriously **not enough** to convince **safety/regulation authorities**

**We must rely on theoretical arguments**

☞ **But**, some of the theoretical developments are based on **unrealistic assumptions**

Usually on the inputs' probabilistic structure...

# Introduction

☞ Being able to **understand how black-box models behave under uncertainty** is crucial for **integrating AI models to critical systems**

Industrial processes, energy production, medicine, cars, planes, the stock market...

☞ The decision-making process must be **justifiable** and **justified**

**Empirical statements** (e.g. SOTA) are notoriously **not enough** to convince **safety/regulation authorities**

**We must rely on theoretical arguments**

☞ **But**, some of the theoretical developments are based on **unrealistic assumptions**

Usually on the inputs' probabilistic structure...

**This is the case for Hoeffding's functional decomposition**

# Classical Hoeffding decomposition

Let $X$ be a random vector of **mutually independent inputs**, and let $G(X)$ be a square integrable random variable

# Classical Hoeffding decomposition

Let $X$ be a random vector of **mutually independent inputs**, and let $G(X)$ be a square integrable random variable

☞ **Hoeffding (1948)**: We can **uniquely** write

$$G(X) = \sum_{A \subseteq D} G_A(X_A),$$

where the summands are all **pairwise orthogonal**

# Classical Hoeffding decomposition

Let $X$ be a random vector of **mutually independent inputs**, and let $G(X)$ be a square integrable random variable

☞ **Hoeffding (1948)**: We can **uniquely** write

$$G(X) = \sum_{A \subseteq D} G_A(X_A),$$

where the summands are all **pairwise orthogonal**

☞ The terms $G_A(X_A)$ can be interpreted as **interactions induced by the model** $G$
This is great for **post-hoc interpretability purposes**!

# Classical Hoeffding decomposition

Let $X$ be a random vector of **mutually independent inputs**, and let $G(X)$ be a square integrable random variable

☞ **Hoeffding (1948)**: We can **uniquely** write

$$G(X) = \sum_{A \subseteq D} G_A(X_A),$$

where the summands are all **pairwise orthogonal**

☞ The terms $G_A(X_A)$ can be interpreted as **interactions induced by the model** $G$

This is great for **post-hoc interpretability purposes**!

**However**, the mutual independence assumption is often not respected in practice...

# Classical Hoeffding decomposition

Let $X$ be a random vector of **mutually independent inputs**, and let $G(X)$ be a square integrable random variable

☞ **Hoeffding (1948)**: We can **uniquely** write

$$G(X) = \sum_{A \subseteq D} G_A(X_A),$$

where the summands are all **pairwise orthogonal**

☞ The terms $G_A(X_A)$ can be interpreted as **interactions induced by the model** $G$

This is great for **post-hoc interpretability purposes**!

**However**, the mutual independence assumption is often not respected in practice...

Tackling such a generalization has already been attempted:

e.g., Rabitz and Aliş (1999), Peccati (2004), Hooker (2007), Kuo et al. (2009), Hart and Gremaud (2018), and Chastaing, Gamboa, and Prieur (2012)

# Classical Hoeffding decomposition

Let $X$ be a random vector of **mutually independent inputs**, and let $G(X)$ be a square integrable random variable

☞ **Hoeffding (1948)**: We can **uniquely** write

$$G(X) = \sum_{A \subseteq D} G_A(X_A),$$

where the summands are all **pairwise orthogonal**

☞ The terms $G_A(X_A)$ can be interpreted as **interactions induced by the model** $G$

This is great for **post-hoc interpretability purposes**!

**However**, the mutual independence assumption is often not respected in practice...

Tackling such a generalization has already been attempted:

e.g., Rabitz and Aliş (1999), Peccati (2004), Hooker (2007), Kuo et al. (2009), Hart and Gremaud (2018), and Chastaing, Gamboa, and Prieur (2012)

**But no definitive answer has been given**

# Classical Hoeffding decomposition

In this talk:

**We will explore how this result can be generalized to non-mutually independent inputs**

# Classical Hoeffding decomposition

In this talk:
**We will explore how this result can be generalized to non-mutually independent inputs**

What we mean by "non-mutually independent inputs":

- **Non-perfect functional dependence**
- **Non-degenerate stochastic dependence**

# Classical Hoeffding decomposition

In this talk:
**We will explore how this result can be generalized to non-mutually independent inputs**

What we mean by "non-mutually independent inputs":

- **Non-perfect functional dependence**
- **Non-degenerate stochastic dependence**

☞ Somewhat *unusual* way to **deal with dependence**

But it involves a lot of interesting mathematics!

At least in my opinion :)

# Random inputs, black-box model

☞ Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a **probability space**
☞ Let $(E_1, \mathcal{E}_1), \ldots, (E_d, \mathcal{E}_d)$ be **standard Borel measurable spaces**

# Random inputs, black-box model

☞ Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a **probability space**
☞ Let $(E_1, \mathcal{E}_1), \ldots, (E_d, \mathcal{E}_d)$ be **standard Borel measurable spaces**

☞ Let $D = \{1, \ldots, d\}$ and let $\mathcal{P}_D$ denote the **power-set** of $D$

# Random inputs, black-box model

☞ Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a **probability space**
☞ Let $(E_1, \mathcal{E}_1), \ldots, (E_d, \mathcal{E}_d)$ be **standard Borel measurable spaces**

☞ Let $D = \{1, \ldots, d\}$ and let $\mathcal{P}_D$ denote the **power-set** of $D$

☞ The **random inputs** are defined as a **measurable mapping**:

$$X : \Omega \to E,$$

where $E = \bigtimes_{i=1}^{d} E_i$ is the **cartesian product of the $d$ Polish spaces**

# Random inputs, black-box model

☞ Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a **probability space**

☞ Let $(E_1, \mathcal{E}_1), \dots, (E_d, \mathcal{E}_d)$ be **standard Borel measurable spaces**

☞ Let $D = \{1, \dots, d\}$ and let $\mathcal{P}_D$ denote the **power-set** of $D$

☞ The **random inputs** are defined as a **measurable mapping**:

$$X : \Omega \to E,$$

where $E = \bigtimes_{i=1}^{d} E_i$ is the **cartesian product of the $d$ Polish spaces**

☞ For any $A \in \mathcal{P}_D$, the **subset of inputs** $X_A = (X_i)_{i \in A}$ is the measurable **mapping** valued in $E_A := \bigtimes_{i \in A} E_i$

# Random inputs, black-box model

☞ Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a **probability space**
☞ Let $(E_1, \mathcal{E}_1), \dots, (E_d, \mathcal{E}_d)$ be **standard Borel measurable spaces**

☞ Let $D = \{1, \dots, d\}$ and let $\mathcal{P}_D$ denote the **power-set** of $D$

☞ The **random inputs** are defined as a **measurable mapping**:

$$X : \Omega \to E,$$

where $E = \bigtimes_{i=1}^{d} E_i$ is the **cartesian product of the $d$ Polish spaces**

☞ For any $A \in \mathcal{P}_D$, the **subset of inputs** $X_A = (X_i)_{i \in A}$ is the measurable **mapping** valued in $E_A := \bigtimes_{i \in A} E_i$

☞ The **black-box model** is a mapping $G : E \to \mathbb{R}$
☞ The **random output** is the ($\mathbb{R}$-valued) random variable $G(X)$

# Generated $\sigma$-algebras

☞ Denote by $\sigma_X$ the **$\sigma$-algebra generated by** $X$:

$$\sigma_X = \left\{ X^{-1}[B] \,:\, \forall B \in \bigotimes_{i \in D} \mathcal{E}_i \right\} \subset \mathcal{F}$$

# Generated $\sigma$-algebras

☞ Denote by $\sigma_X$ the $\sigma$**-algebra generated by** $X$:

$$\sigma_X = \left\{ X^{-1}[B] \ : \ \forall B \in \bigotimes_{i \in D} \mathcal{E}_i \right\} \subset \mathcal{F}$$

☞ For every $A \in \mathcal{P}_D$, denote by $\sigma_A$ the $\sigma$**-algebra generated by** $X_A$:

$$\sigma_A = \left\{ X_A^{-1}[B] \ : \ \forall B \in \bigotimes_{i \in A} \mathcal{E}_i \right\} \subset \mathcal{F}$$

## Generated $\sigma$-algebras

☞ Denote by $\sigma_X$ the $\sigma$**-algebra generated by** $X$:

$$\sigma_X = \left\{ X^{-1}\left[B\right] \; : \; \forall B \in \bigotimes_{i \in D} \mathcal{E}_i \right\} \subset \mathcal{F}$$

☞ For every $A \in \mathcal{P}_D$, denote by $\sigma_A$ the $\sigma$**-algebra generated by** $X_A$:

$$\sigma_A = \left\{ X_A^{-1}\left[B\right] \; : \; \forall B \in \bigotimes_{i \in A} \mathcal{E}_i \right\} \subset \mathcal{F}$$

☞ Denote by $\sigma_\emptyset$ the $\mathbb{P}$**-trivial** $\sigma$**-algebra**:

$$\sigma_\emptyset = \sigma\left[\left\{B \in \mathcal{F} \; : \; \mathbb{P}\left(B\right) = 0\right\}\right] \subset \mathcal{F}$$

**Lemma** *(Doob-Dynkin)*.

Let $Y$ be an $\mathbb{R}$-value random variable, and let $X$ be random inputs.
If $Y$ is $\sigma_X$-measurable, then there exists a function $f : E \to \mathbb{R}$ such that

$$Y = f(X) \text{ a.s.}$$

# Measurability and Lebesgue spaces

**Lemma** *(Doob-Dynkin)*.

Let $Y$ be an $\mathbb{R}$-value random variable, and let $X$ be random inputs.
If $Y$ is $\sigma_X$-measurable, then there exists a function $f : E \to \mathbb{R}$ such that

$$Y = f(X) \text{ a.s.}$$

**Lemma** *(Kallenberg (2021, Lemma 4.9))*.

Let $Y$ be an $\mathbb{R}$-value random variable.
If $Y$ is $\sigma_\emptyset$-measurable, then it is constant a.s.

# Measurability and Lebesgue spaces

**Lemma** *(Doob-Dynkin)*.

Let $Y$ be an $\mathbb{R}$-value random variable, and let $X$ be random inputs.
If $Y$ is $\sigma_X$-measurable, then there exists a function $f : E \to \mathbb{R}$ such that

$$Y = f(X) \text{ a.s.}$$

**Lemma** *(Kallenberg (2021, Lemma 4.9))*.

Let $Y$ be an $\mathbb{R}$-value random variable.
If $Y$ is $\sigma_\emptyset$-measurable, then it is constant a.s.

**Definition** *(Lebesgue spaces $\mathbb{L}^2$)*.

For a sub-$\sigma$-algebra $\mathcal{G} \subset \mathcal{F}$, denote by $\mathbb{L}^2(\mathcal{G})$ the **Lebesgue space of square-integrable, $\mathbb{R}$-valued, $\mathcal{G}$-measurable** random variables.

It is a **Hilbert space** with the inner product, $\forall Z_1, Z_2 \in \mathbb{L}^2(\mathcal{G})$:

$$\langle Z_1, Z_2 \rangle = \mathbb{E}[Z_1 Z_2] = \int_\Omega Z_1(\omega) Z_2(\omega) d\mathbb{P}(\omega)$$

# Non-perfect functional dependence

- ☞ $\mathbb{L}^2(\sigma_X)$ **contains random variables that are functions of** $X$.
- ☞ For every $A \subset D$, $\mathbb{L}^2(\sigma_A)$ **contains random variables that are functions of** $X_A$.
- ☞ $\mathbb{L}^2(\sigma_\emptyset)$ **contains constants**.

# Non-perfect functional dependence

☞ $\mathbb{L}^2(\sigma_X)$ **contains random variables that are functions of** $X$.

☞ For every $A \subset D$, $\mathbb{L}^2(\sigma_A)$ **contains random variables that are functions of** $X_A$.

☞ $\mathbb{L}^2(\sigma_\emptyset)$ **contains constants**.

**Theorem** *Sidák (1957, Theorem 2)*. Let $\mathcal{G}_1, \mathcal{G}_2 \subseteq \mathcal{F}$, then

- If $\mathcal{G}_1 \subset \mathcal{G}_2$, then $\mathbb{L}^2(\mathcal{G}_1) \subset \mathbb{L}^2(\mathcal{G}_2) \subseteq \mathbb{L}^2(\mathcal{F})$ ;
- $\mathbb{L}^2(\mathcal{G}_1) \cap \mathbb{L}^2(\mathcal{G}_2) = \mathbb{L}^2(\mathcal{G}_1 \cap \mathcal{G}_2)$.

# Non-perfect functional dependence

- ☞ $\mathbb{L}^2(\sigma_X)$ **contains random variables that are functions of $X$**.
- ☞ For every $A \subset D$, $\mathbb{L}^2(\sigma_A)$ **contains random variables that are functions of $X_A$**.
- ☞ $\mathbb{L}^2(\sigma_\emptyset)$ **contains constants**.

**Theorem** *Sidák (1957, Theorem 2)*. Let $\mathcal{G}_1, \mathcal{G}_2 \subseteq \mathcal{F}$, then

- If $\mathcal{G}_1 \subset \mathcal{G}_2$, then $\mathbb{L}^2(\mathcal{G}_1) \subset \mathbb{L}^2(\mathcal{G}_2) \subseteq \mathbb{L}^2(\mathcal{F})$ ;
- $\mathbb{L}^2(\mathcal{G}_1) \cap \mathbb{L}^2(\mathcal{G}_2) = \mathbb{L}^2(\mathcal{G}_1 \cap \mathcal{G}_2)$.

**Assumption 1** *(Non-perfect functional dependence)*.

- $\sigma_\emptyset \subset \sigma_i$, $i = 1, \dots, d$ (inputs are not constant).
- For $B \subset A$, $\sigma_B \subset \sigma_A$ (inputs add information).
- For every $A, B \in \mathcal{P}_D$, $A \neq B$,

$$\sigma_A \cap \sigma_B = \sigma_{A \cap B}.$$

# Output space

Consequences of Assumption 1:

☞ $\mathbb{L}^2(\sigma_\emptyset) \subset \mathbb{L}^2(\sigma_A)$, for every $A \in \mathcal{P}_D$

**There are non-constant random variables in the Lebesgue spaces**

☞ For $B \subset A$, $\mathbb{L}^2(\sigma_B) \subset \mathbb{L}^2(\sigma_A)$

**There are functions of $X_A$ that are not functions of $X_B$**

☞ For any $A, B \in \mathcal{P}_D$,

$$\mathbb{L}^2(\sigma_A) \cap \mathbb{L}^2(\sigma_B) = \mathbb{L}^2(\sigma_{A\cap B})$$

**The functions of $X_A$ and $X_B$ are in fact functions of $X_{A\cap B}$**

# Output space

Consequences of Assumption 1:

☞ $\mathbb{L}^2(\sigma_\emptyset) \subset \mathbb{L}^2(\sigma_A)$, for every $A \in \mathcal{P}_D$

**There are non-constant random variables in the Lebesgue spaces**

☞ For $B \subset A$, $\mathbb{L}^2(\sigma_B) \subset \mathbb{L}^2(\sigma_A)$

**There are functions of $X_A$ that are not functions of $X_B$**

☞ For any $A, B \in \mathcal{P}_D$,

$$\mathbb{L}^2(\sigma_A) \cap \mathbb{L}^2(\sigma_B) = \mathbb{L}^2(\sigma_{A \cap B})$$

**The functions of $X_A$ and $X_B$ are in fact functions of $X_{A \cap B}$**

**Proposition** . Suppose that Assumption 1 hold. Then, for any $A, B \in \mathcal{P}_D$ such that $A \cap B \notin \{A, B\}$, **there is no mapping $T$ such that**

$$X_B = T(X_A) \text{ a.e.}$$

☞ Hence the name **"non-perfect functional dependence"**

# Dependence and maximal angles between Lebesgue spaces

**Theorem** *(Malliavin 1995, Chapter 3)*. Let $X$ and $Y$ be two random elements. Then:

$$X \perp\!\!\!\perp Y \iff \forall f(X) \in \mathbb{L}^2_0(\sigma_X), \forall g(Y) \in \mathbb{L}^2_0(\sigma_Y), \quad \mathbb{E}[f(X)g(Y)] = 0,$$

$\mathbb{L}^2_0$ are Lebesgue subspaces of **centered random variables**.

**Theorem** *(Malliavin 1995, Chapter 3)*. Let $X$ and $Y$ be two random elements. Then:

$$X \perp\!\!\!\perp Y \iff \forall f(X) \in \mathbb{L}_0^2(\sigma_X), \forall g(Y) \in \mathbb{L}_0^2(\sigma_Y), \quad \mathbb{E}[f(X)g(Y)] = 0,$$

$\mathbb{L}_0^2$ are Lebesgue subspaces of **centered random variables**.

- $X$ and $Y$ are **independent** $\implies$ **All the functions of $X$ and the functions of $Y$ are uncorrelated.** ✓

# Dependence and maximal angles between Lebesgue spaces

**Theorem** *(Malliavin 1995, Chapter 3)*. Let $X$ and $Y$ be two random elements. Then:

$$X \perp\!\!\!\perp Y \iff \forall f(X) \in \mathbb{L}_0^2\left(\sigma_X\right), \forall g(Y) \in \mathbb{L}_0^2\left(\sigma_Y\right), \quad \mathbb{E}\left[f(X)g(Y)\right] = 0,$$

$\mathbb{L}_0^2$ are Lebesgue subspaces of **centered random variables**.

- $X$ and $Y$ are **independent** $\implies$ **All the functions of** $X$ **and the functions of** $Y$ **are uncorrelated.** ✓
- **All the functions of** $X$ **and the functions of** $Y$ **are uncorrelated** $\implies$ $X$ and $Y$ are **independent**. !

# Dependence and maximal angles between Lebesgue spaces

**Theorem** *(Malliavin 1995, Chapter 3)*. Let $X$ and $Y$ be two random elements. Then:

$$X \perp\!\!\!\perp Y \iff \forall f(X) \in \mathbb{L}^2_0(\sigma_X), \forall g(Y) \in \mathbb{L}^2_0(\sigma_Y), \quad \mathbb{E}[f(X)g(Y)] = 0,$$

$\mathbb{L}^2_0$ are Lebesgue subspaces of **centered random variables**.

- $X$ and $Y$ are **independent** $\implies$ **All the functions of** $X$ **and the functions of** $Y$ **are uncorrelated.** ✓
- **All the functions of** $X$ **and the functions of** $Y$ **are uncorrelated** $\implies$ $X$ and $Y$ are **independent**. **!**

## Our intuition:

**Control the dependence structure of** $X$ by **controlling the magnitude of the inner product between the functions of** $X_A$ for every $A \in \mathcal{P}_D$.

# Minimal angle, maximal correlation

Dixmier's angle: the **maximal value** the inner product can take **between the elements of two closed subspaces** of a Hilbert space

**Definition** *(Dixmier's angle (Dixmier 1949))*. Let $M, N$ be **closed** subspaces of a Hilbert space $H$. The cosine of Dixmier's angle between $M$ and $N$ is defined as

$$c_0(M, N) := \sup \{ |\langle x, y\rangle| : x \in M, \|x\| \leq 1, \quad y \in N, \|y\| \leq 1 \}.$$

# Minimal angle, maximal correlation

Dixmier's angle: the **maximal value** the inner product can take **between the elements of two closed subspaces** of a Hilbert space

**Definition** *(Dixmier's angle (Dixmier 1949))*. Let $M$, $N$ be **closed** subspaces of a Hilbert space $H$. The cosine of Dixmier's angle between $M$ and $N$ is defined as

$$c_0\left(M, N\right) := \sup\left\{ |\langle x, y\rangle| : x \in M, \|x\| \leq 1, \quad y \in N, \|y\| \leq 1 \right\}.$$

☞ Analogous to the **maximal correlation** in probability theory (Koyak 1987), as a dependence measure between **random elements**.

**Definition** *(Maximal correlation (Gebelein 1941))*. Let $X$, $Y$ be two **random elements**. The **maximal correlation** between $X$ and $Y$ is

$$\rho_0(X, Y) := c_0\left(\mathbb{L}_0^2\left(\sigma_X\right), \mathbb{L}_0^2\left(\sigma_Y\right)\right)$$

# Minimal angle, maximal correlation

Dixmier's angle: the **maximal value** the inner product can take **between the elements of two closed subspaces** of a Hilbert space

**Definition** *(Dixmier's angle (Dixmier 1949))*. Let $M$, $N$ be **closed** subspaces of a Hilbert space $H$. The cosine of Dixmier's angle between $M$ and $N$ is defined as

$$c_0\left(M, N\right) := \sup\left\{|\langle x, y\rangle| : x \in M, \|x\| \leq 1, \quad y \in N, \|y\| \leq 1\right\}.$$

☞ Analogous to the **maximal correlation** in probability theory (Koyak 1987), as a dependence measure between **random elements**.

**Definition** *(Maximal correlation (Gebelein 1941))*. Let $X$, $Y$ be two **random elements**. The **maximal correlation** between $X$ and $Y$ is

$$\rho_0(X, Y) := c_0\left(\mathbb{L}_0^2\left(\sigma_X\right), \mathbb{L}_0^2\left(\sigma_Y\right)\right)$$

**Remark** . The independence relation from the previous slide can be written as:

$$X \perp\!\!\!\perp Y \iff c_0\left(\mathbb{L}_0^2\left(\sigma_X\right), \mathbb{L}_0^2\left(\sigma_Y\right)\right) = 0.$$

# Maximal partial correlation

Friedrichs' angle: Restriction to the **elements orthogonal to the intersection of the subspaces**

**Definition** *(Friedrich's angle (Friedrichs 1937))*. The cosine of Friedrichs' angle is defined as

$$c\,(M,N) := \sup\left\{ |\langle x,y\rangle| : \begin{cases} x \in M \cap (M\cap N)^{\perp}, \|x\| \leq 1 \\ y \in N \cap (M\cap N)^{\perp}, \|y\| \leq 1 \end{cases} \right\},$$

where the orthogonal complement is taken w.r.t. to $H$.

# Maximal partial correlation

Friedrichs' angle: Restriction to the **elements orthogonal to the intersection of the subspaces**

**Definition** *(Friedrich's angle (Friedrichs 1937))*. The cosine of Friedrichs' angle is defined as

$$c\left(M,N\right) := \sup\left\{ |\langle x,y\rangle| : \begin{cases} x \in M \cap \left(M \cap N\right)^{\perp}, \|x\| \leq 1 \\ y \in N \cap \left(M \cap N\right)^{\perp}, \|y\| \leq 1 \end{cases} \right\},$$

where the orthogonal complement is taken w.r.t. to $H$.

☞ Analogous to the **maximal partial dependence** between random elements (Bryc 1984, 1996; Dauxois, Nkiet, and Romain 2004).

**Definition** *(Maximal partial correlation)*. The **maximal partial correlation** between $X$ and $Y$ is

$$\rho^*(X,Y) := c\left(\mathbb{L}^2\left(\sigma_X\right), \mathbb{L}^2\left(\sigma_Y\right)\right)$$

## Maximal partial correlation

Friedrichs' angle: Restriction to the **elements orthogonal to the intersection of the subspaces**

**Definition** *(Friedrich's angle (Friedrichs 1937))*. The cosine of Friedrichs' angle is defined as

$$c(M,N) := \sup\left\{ |\langle x,y\rangle| : \begin{cases} x \in M \cap (M\cap N)^{\perp}, \|x\| \le 1 \\ y \in N \cap (M\cap N)^{\perp}, \|y\| \le 1 \end{cases} \right\},$$

where the orthogonal complement is taken w.r.t. to $H$.

☞ Analogous to the **maximal partial dependence** between random elements (Bryc 1984, 1996; Dauxois, Nkiet, and Romain 2004).

**Definition** *(Maximal partial correlation)*. The **maximal partial correlation** between $X$ and $Y$ is

$$\rho^*(X,Y) := c\left(\mathbb{L}^2(\sigma_X), \mathbb{L}^2(\sigma_Y)\right)$$

**Remark** . It is closely related **to the commutativity of conditional expectations**.

$$c\left(\mathbb{L}^2(\sigma_X), \mathbb{L}^2(\sigma_Y)\right) = 0 \iff \mathbb{E}\left[\mathbb{E}\left[. \mid X\right] \mid Y\right] = \mathbb{E}\left[\mathbb{E}\left[. \mid Y\right] \mid X\right]$$

## Closure and complements

☞ They assess the **closedness of the sum of subspaces** (for an infinite-dimensional $H$):

- $c\,(M,N) < 1 \iff M + N$ is closed in $H$ ;
- $c_0\,(M,N) < 1 \iff M \cap N = \{0\}$ and $M + N$ is closed in $H$.

# Closure and complements

☞ They assess the **closedness of <u>the sum</u> of subspaces** (for an infinite-dimensional $H$):

- $c(M, N) < 1 \iff M + N$ is closed in $H$ ;
- $c_0(M, N) < 1 \iff M \cap N = \{0\}$ and $M + N$ is closed in $H$.

☞ In Hilbert spaces, **a closed subspace is always complemented**:
If $M$ is closed, then there exists **another subspace** $K$ such that

$$H = M + K \text{ and } M \cap K = \{0\}.$$

## Closure and complements

☞ They assess the **closedness of <u>the sum</u> of subspaces** (for an infinite-dimensional $H$):

- $c(M, N) < 1 \iff M + N$ is closed in $H$ ;
- $c_0(M, N) < 1 \iff M \cap N = \{0\}$ and $M + N$ is closed in $H$.

☞ In Hilbert spaces, **a closed subspace is always complemented**:
If $M$ is closed, then there exists **another subspace** $K$ such that

$$H = M + K \text{ and } M \cap K = \{0\}.$$

☞ ***K* is what's missing from *M* to span the ambient space *H*.**
**One** popular complement of closed subspaces are their **orthogonal complement (**$M^{\perp}$**)**.

# Feshchenko matrix

- ☞ For every $A \in \mathcal{P}_D$, $\mathbb{L}^2(\sigma_A)$ contains **the functions of** $X_A$
- ☞ Dixmier's and Friedrichs' angles to **pairwise** control the inner products in these spaces

## Feshchenko matrix

- ☞ For every $A \in \mathcal{P}_D$, $\mathbb{L}^2(\sigma_A)$ contains **the functions of** $X_A$
- ☞ Dixmier's and Friedrichs' angles to **pairwise** control the inner products in these spaces

Intuition: A **generalized precision matrix** to control the *global magnitude of all the angles*

# Feshchenko matrix

☞ For every $A \in \mathcal{P}_D$, $\mathbb{L}^2(\sigma_A)$ contains **the functions of** $X_A$
☞ Dixmier's and Friedrichs' angles to **pairwise** control the inner products in these spaces

Intuition: A **generalized precision matrix** to control the *global magnitude of all the angles*

**Definition** *(Maximal coalitional precision matrix)*.

Let $\Delta$ be the $(2^d \times 2^d)$, symmetric **set-indexed** matrix, defined element-wise, $\forall A, B \in \mathcal{P}_D$ as

$$\Delta_{AB} = \begin{cases} 1 & \text{if } A = B; \\ -c\left(\mathbb{L}^2(\sigma_A), \mathbb{L}^2(\sigma_B)\right) & \text{otherwise.} \end{cases}$$

We use **Friedrichs' angles** (partial correlation), hence the **precision** part

☞ These matrices closely resemble the ones used by **Feshchenko (2020)** to study the **closedness of an arbitrary sum of closed subspaces** of a Hilbert space

# Feshchenko matrix

☞ For every $A \in \mathcal{P}_D$, $\mathbb{L}^2(\sigma_A)$ contains **the functions of** $X_A$
☞ Dixmier's and Friedrichs' angles to **pairwise** control the inner products in these spaces

Intuition: A **generalized precision matrix** to control the *global magnitude of all the angles*

**Definition** *(Maximal coalitional precision matrix)*.

Let $\Delta$ be the $(2^d \times 2^d)$, symmetric **set-indexed** matrix, defined element-wise, $\forall A, B \in \mathcal{P}_D$ as

$$\Delta_{AB} = \begin{cases} 1 & \text{if } A = B; \\ -c\left(\mathbb{L}^2(\sigma_A), \mathbb{L}^2(\sigma_B)\right) & \text{otherwise.} \end{cases}$$

We use **Friedrichs' angles** (partial correlation), hence the **precision** part

☞ These matrices closely resemble the ones used by **Feshchenko (2020)** to study the **closedness of an arbitrary sum of closed subspaces** of a Hilbert space

## We're calling them "Feshchenko matrices".

**But why are Feshchenko matrices interesting?**

## But why are Feshchenko matrices interesting?

**Proposition** . Suppose that Assumption 1 hold. Then,

$$\Delta = I_{2d} \quad \iff \quad X \text{ is mutually independent.}$$

☞ $X$ is valued in a product of Polish spaces, with **an arbitrary law**

## But why are Feshchenko matrices interesting?

**Proposition** . Suppose that Assumption 1 hold. Then,

$$\Delta = I_{2d} \quad \iff X \text{ is mutually independent.}$$

☞ $X$ is valued in a product of Polish spaces, with **an arbitrary law**

**Assumption 2** *(Non-degenerate stochastic dependence)*.

The Feshchenko matrix $\Delta$ of the inputs $X$ is definite-positive.

It's a **restriction on the inner product of** $\mathbb{L}^2(\sigma_X) \implies$ **A restriction on the law of** $X$

☞ Hence the *stochastic dependence* (in opposition to *functional dependence*).

# Direct-sum decomposition

**Definition** *Direct-sum decomposition (Axler 2015)*.

Let $H$ be a vector space and let $H_1, \dots, H_n$ be proper subspaces of $H$.

$H$ is said to admit a **direct-sum decomposition** if any $h \in H$ can be written **uniquely** as

$$h = \sum_{i=1}^{n} h_i \text{ where } h_i \in H_i \text{ for } i = 1, \dots, n.$$

In this case, we write:

$$H = \bigoplus_{i=1}^{n} H_i.$$

# Direct-sum decomposition

**Definition** *Direct-sum decomposition (Axler 2015)*.

Let $H$ be a vector space and let $H_1, \dots, H_n$ be proper subspaces of $H$.

$H$ is said to admit a **direct-sum decomposition** if any $h \in H$ can be written **uniquely** as

$$h = \sum_{i=1}^{n} h_i \text{ where } h_i \in H_i \text{ for } i = 1, \dots, n.$$

In this case, we write:

$$H = \bigoplus_{i=1}^{n} H_i.$$

☞ (Infinite-dimensional) **Hilbert spaces** are **linear vector spaces**

# Direct-sum decomposition

**Definition** *Direct-sum decomposition (Axler 2015)*.

Let $H$ be a vector space and let $H_1, \ldots, H_n$ be proper subspaces of $H$.

$H$ is said to admit a **direct-sum decomposition** if any $h \in H$ can be written **uniquely** as

$$h = \sum_{i=1}^{n} h_i \text{ where } h_i \in H_i \text{ for } i = 1, \ldots, n.$$

In this case, we write:

$$H = \bigoplus_{i=1}^{n} H_i.$$

☞ (Infinite-dimensional) **Hilbert spaces** are **linear vector spaces**

**Goal: Find a direct-sum decomposition of** $\mathbb{L}^2(\sigma_X)$ **w.r.t. the subspaces** $\{\mathbb{L}^2(\sigma_A)\}_{A \in \mathcal{P}_D}$

Any real-valued $G(X)$ could be **uniquely** decomposed as a sum of **functions** of $X_A$, $A \in \mathcal{P}_D$

# Direct-sum decomposition

**Definition** *Direct-sum decomposition (Axler 2015)*.

Let $H$ be a vector space and let $H_1, \dots, H_n$ be proper subspaces of $H$.

$H$ is said to admit a **direct-sum decomposition** if any $h \in H$ can be written **uniquely** as

$$h = \sum_{i=1}^{n} h_i \text{ where } h_i \in H_i \text{ for } i = 1, \dots, n.$$

In this case, we write:

$$H = \bigoplus_{i=1}^{n} H_i.$$

☞ (Infinite-dimensional) **Hilbert spaces** are **linear vector spaces**

**Goal: Find a direct-sum decomposition of $\mathbb{L}^2(\sigma_X)$ w.r.t. the subspaces $\{\mathbb{L}^2(\sigma_A)\}_{A \in \mathcal{P}_D}$**
Any real-valued $G(X)$ could be **uniquely** decomposed as a sum of **functions** of $X_A$, $A \in \mathcal{P}_D$

☞ **Consider Hoeffding's decomposition as a direct-sum decomposition of $\mathbb{L}^2(\sigma_X)$**

# Generalized Hoeffding decomposition

**Theorem .**

Under Assumptions 1 and 2, for every $A \in \mathcal{P}_D$, one has that

$$\mathbb{L}^2(\sigma_A) = \bigoplus_{B \in \mathcal{P}_A} V_B.$$

where $V_\emptyset = \mathbb{L}^2(\sigma_\emptyset)$, and

$$V_B = \left[ \sum_{C \in \mathcal{P}_B, C \neq B} V_C \right]^{\perp_B},$$

where $\perp_B$ denotes the orthogonal complement in $\mathbb{L}^2(\sigma_B)$.

# Generalized Hoeffding decomposition

**Theorem .**

Under Assumptions 1 and 2, for every $A \in \mathcal{P}_D$, one has that

$$\mathbb{L}^2(\sigma_A) = \bigoplus_{B \in \mathcal{P}_A} V_B.$$

where $V_\emptyset = \mathbb{L}^2(\sigma_\emptyset)$, and

$$V_B = \left[ \sum_{C \in \mathcal{P}_B, C \neq B} V_C \right]^{\perp_B},$$

where $\perp_B$ denotes the orthogonal complement in $\mathbb{L}^2(\sigma_B)$.

**Intuition of the proof:**

**Inductive functional centering**

# Intuition of the proof: One input

**One input:**

1. Let $i \in D$, and **fix** $\mathbb{L}^2(\sigma_i)$ **as the ambient space**
2. We have that $V_\emptyset := \mathbb{L}^2(\sigma_\emptyset)$ **is a closed subspace of** $\mathbb{L}^2(\sigma_i)$ (it is **complemented**)
3. Denote $V_i = [V_\emptyset]^{\perp_i}$, **the orthogonal complement of** $V_\emptyset$ **in** $\mathbb{L}^2(\sigma_i)$
4. One has that $\mathbb{L}^2(\sigma_i) = V_\emptyset \oplus V_i$

We just showed that any $f(X_i) \in \mathbb{L}^2(\sigma_i)$ can be written as

$$f(X_i) = \underbrace{\mathbb{E}[f(X_i)]}_{\in V_\emptyset} + \underbrace{\mathbb{E}[f(X_i) - \mathbb{E}[f(X_i)]]}_{\in V_i = \mathbb{L}^2_0(\sigma_i)}$$

# Intuition of the proof: One input

**One input:**

1. Let $i \in D$, and **fix** $\mathbb{L}^2(\sigma_i)$ **as the ambient space**
2. We have that $V_\emptyset := \mathbb{L}^2(\sigma_\emptyset)$ **is a closed subspace of** $\mathbb{L}^2(\sigma_i)$ (it is **complemented**)
3. Denote $V_i = [V_\emptyset]^{\perp_i}$, **the orthogonal complement of** $V_\emptyset$ **in** $\mathbb{L}^2(\sigma_i)$
4. One has that $\mathbb{L}^2(\sigma_i) = V_\emptyset \oplus V_i$

We just showed that any $f(X_i) \in \mathbb{L}^2(\sigma_i)$ can be written as

$$f(X_i) = \underbrace{\mathbb{E}[f(X_i)]}_{\in V_\emptyset} + \underbrace{\mathbb{E}[f(X_i) - \mathbb{E}[f(X_i)]]}_{\in V_i = \mathbb{L}^2_0(\sigma_i)}$$

**And note that** $\mathbb{L}^2(\sigma_i) = V_\emptyset \oplus V_i$ **hold for any** $i \in D$ **(induction)**

# Intuition of the proof: Two inputs

**Two inputs:**

1. Let $i, j \in D$, and **fix** $\mathbb{L}^2(\sigma_{ij})$ **as the ambient space**
2. **Assumptions 1 and 2 imply that** $\mathbb{L}^2(\sigma_i) + \mathbb{L}^2(\sigma_j)$ **is closed in** $\mathbb{L}^2(\sigma_{ij})$ (it is **complemented**)
3. Notice **(previous step)** that $\mathbb{L}^2(\sigma_i) + \mathbb{L}^2(\sigma_j) = V_\emptyset + V_i + V_j$
4. Denote $V_{ij} = [V_\emptyset + V_i + V_j]^{\perp_{ij}}$, **the orthogonal complement in** $\mathbb{L}^2(\sigma_{ij})$
5. We thus have that $\mathbb{L}^2(\sigma_{ij}) = V_\emptyset + V_i + V_j + V_{ij}$

# Intuition of the proof: Two inputs

**Two inputs:**

1. Let $i, j \in D$, and **fix** $\mathbb{L}^2(\sigma_{ij})$ **as the ambient space**
2. **Assumptions 1 and 2 imply that** $\mathbb{L}^2(\sigma_i) + \mathbb{L}^2(\sigma_j)$ **is closed in** $\mathbb{L}^2(\sigma_{ij})$ (it is **complemented**)
3. Notice **(previous step)** that $\mathbb{L}^2(\sigma_i) + \mathbb{L}^2(\sigma_j) = V_\emptyset + V_i + V_j$
4. Denote $V_{ij} = [V_\emptyset + V_i + V_j]^{\perp_{ij}}$, **the orthogonal complement in** $\mathbb{L}^2(\sigma_{ij})$
5. We thus have that $\mathbb{L}^2(\sigma_{ij}) = V_\emptyset + V_i + V_j + V_{ij}$

**And note that the decomposition hold for any pair** $i, j \in D$

We **"centered" a bivariate function from its "univariate and constant parts"**

# Intuition of the proof: Two inputs

Two inputs:

1. Let $i, j \in D$, and **fix** $\mathbb{L}^2(\sigma_{ij})$ **as the ambient space**
2. **Assumptions 1 and 2 imply that** $\mathbb{L}^2(\sigma_i) + \mathbb{L}^2(\sigma_j)$ **is closed in** $\mathbb{L}^2(\sigma_{ij})$ (it is **complemented**)
3. Notice **(previous step)** that $\mathbb{L}^2(\sigma_i) + \mathbb{L}^2(\sigma_j) = V_\emptyset + V_i + V_j$
4. Denote $V_{ij} = [V_\emptyset + V_i + V_j]^{\perp_{ij}}$, **the orthogonal complement in** $\mathbb{L}^2(\sigma_{ij})$
5. We thus have that $\mathbb{L}^2(\sigma_{ij}) = V_\emptyset + V_i + V_j + V_{ij}$

**And note that the decomposition hold for any pair** $i, j \in D$

We **"centered" a bivariate function from its "univariate and constant parts"**

**We continue the induction up to $d$ inputs.**

# Orthocanonical decomposition

**Corollary** *(Orthocanonical decomposition)*.

Suppose that Assumptions 1 and 2 hold.

Then, any random variable $G(X) \in \mathbb{L}^2\left(\sigma_X\right)$ can be **uniquely decomposed** as

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A),$$

where each $G_A(X_A) \in V_A$.

# Orthocanonical decomposition

**Corollary** *(Orthocanonical decomposition)*.

Suppose that Assumptions 1 and 2 hold.

Then, any random variable $G(X) \in \mathbb{L}^2(\sigma_X)$ can be **uniquely decomposed** as

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A),$$

where each $G_A(X_A) \in V_A$.

The term **"orthocanonical"** comes from **the choice** of the **orthogonal complement** in the "centering process"

# Orthocanonical decomposition

**Corollary** *(Orthocanonical decomposition)*.

Suppose that Assumptions 1 and 2 hold.

Then, any random variable $G(X) \in \mathbb{L}^2(\sigma_X)$ can be **uniquely decomposed** as

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A),$$

where each $G_A(X_A) \in V_A$.

The term **"orthocanonical"** comes from **the choice** of the **orthogonal complement** in the "centering process"

The subspaces $V_A$ contain **random variables that are functions of exactly** $X_A$
If one of their element can be expressed with fewer inputs, it is necessarily equal to 0

# Orthocanonical decomposition

**Corollary** *(Orthocanonical decomposition)*.

Suppose that Assumptions 1 and 2 hold.

Then, any random variable $G(X) \in \mathbb{L}^2(\sigma_X)$ can be **uniquely decomposed** as

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A),$$

where each $G_A(X_A) \in V_A$.

The term **"orthocanonical"** comes from **the choice** of the **orthogonal complement** in the "centering process"

The subspaces $V_A$ contain **random variables that are functions of exactly** $X_A$
If one of their element can be expressed with fewer inputs, it is necessarily equal to 0

**Is it possible to characterize the *representants* $G_A(X_A)$?**

## Projectors

- ☞ Let $H$ be a Hilbert space, and let $P : H \to H$ be an operator
- ☞ $\mathrm{Ran}\,(P)$ is the range of $P$, and $\mathrm{Ker}\,(P)$ is its nullspace

# Projectors

- ☞ Let $H$ be a Hilbert space, and let $P : H \to H$ be an operator
- ☞ $\mathrm{Ran}\,(P)$ is the range of $P$, and $\mathrm{Ker}\,(P)$ is its nullspace

If $P$ is an idempotent ($P \circ P = P$), linear, and bounded operator, it is then called the **oblique projector onto $\mathbf{Ran}\,(P)$ parallel to $\mathbf{Ker}\,(P)$** and

$$H = \mathrm{Ran}\,(P) \oplus \mathrm{Ker}\,(P)$$

# Projectors

- ☞ Let $H$ be a Hilbert space, and let $P : H \to H$ be an operator
- ☞ $\mathrm{Ran}\,(P)$ is the range of $P$, and $\mathrm{Ker}\,(P)$ is its nullspace

If $P$ is an idempotent ($P \circ P = P$), linear, and bounded operator, it is then called the **oblique projector onto $\mathbf{Ran}\,(P)$ parallel to $\mathbf{Ker}\,(P)$** and

$$H = \mathrm{Ran}\,(P) \oplus \mathrm{Ker}\,(P)$$

Conversely, suppose that

$$H = M \oplus N$$

then there exists a projector $P$ such that $\mathrm{Ran}\,(P) = M$ and $\mathrm{Ker}\,(P) = N$ (Galántai 2004)

- ☞ $P$ is called the **canonical oblique projector** w.r.t. to this direct-sum decomposition of $H$

# Projectors

☞ Let $H$ be a Hilbert space, and let $P : H \to H$ be an operator
☞ $\mathrm{Ran}\,(P)$ is the range of $P$, and $\mathrm{Ker}\,(P)$ is its nullspace

If $P$ is an idempotent ($P \circ P = P$), linear, and bounded operator, it is then called the **oblique projector onto $\mathrm{Ran}\,(P)$ parallel to $\mathrm{Ker}\,(P)$** and

$$H = \mathrm{Ran}\,(P) \oplus \mathrm{Ker}\,(P)$$

Conversely, suppose that

$$H = M \oplus N$$

then there exists a projector $P$ such that $\mathrm{Ran}\,(P) = M$ and $\mathrm{Ker}\,(P) = N$ (Galántai 2004)
☞ $P$ is called the **canonical oblique projector** w.r.t. to this direct-sum decomposition of $H$

If in addition $N = M^{\perp}$ ($P$ is self-adjoint), then $P$ is called the **orthogonal projector** onto $M$
☞ The orthogonal projector onto $\mathbb{L}^2\,(\sigma_A)$ is the conditional expectation $\mathbb{E}\,[. \mid \sigma_A]$

# Orthocanonical projectors

From the direct-sum decomposition of $\mathbb{L}^2(\sigma_X)$:

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A).$$

# Orthocanonical projectors

From the direct-sum decomposition of $\mathbb{L}^2(\sigma_X)$:

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A).$$

**Oblique projection onto $V_A$**

The operator

$$Q_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \quad G(X) \mapsto G_A(X_A).$$

$Q_A$ is the (canonical) **oblique projection** with

$$\mathrm{Ran}(Q_A) = V_A \text{ , and } \mathrm{Ker}(Q_A) = \bigoplus_{B \in \mathcal{P}_D : B \neq A} V_B$$

# Orthocanonical projectors

From the direct-sum decomposition of $\mathbb{L}^2(\sigma_X)$:

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A).$$

## Oblique projection onto $V_A$

The operator

$$Q_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \quad G(X) \mapsto G_A(X_A).$$

$Q_A$ is the (canonical) **oblique projection** with

$$\mathrm{Ran}(Q_A) = V_A \text{ , and } \mathrm{Ker}(Q_A) = \bigoplus_{B \in \mathcal{P}_D : B \neq A} V_B$$

## Orthogonal projections onto $V_A$

The projector

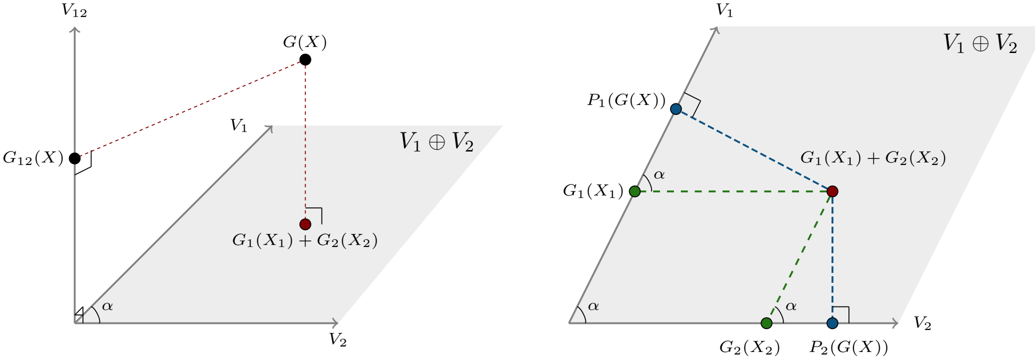
$$P_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \text{ with } \mathrm{Ran}(P_A) = V_A \text{ and } \mathrm{Ker}(P_A) = [V_A]^\perp$$

is the **orthogonal projection** onto $V_A$.

# Illustration* $\mathbb{L}_0^2(\sigma_{12})$

Hence, for any $G(X) \in \mathbb{L}^2(\sigma_X)$, one has that, $\forall A \in \mathcal{P}_D$

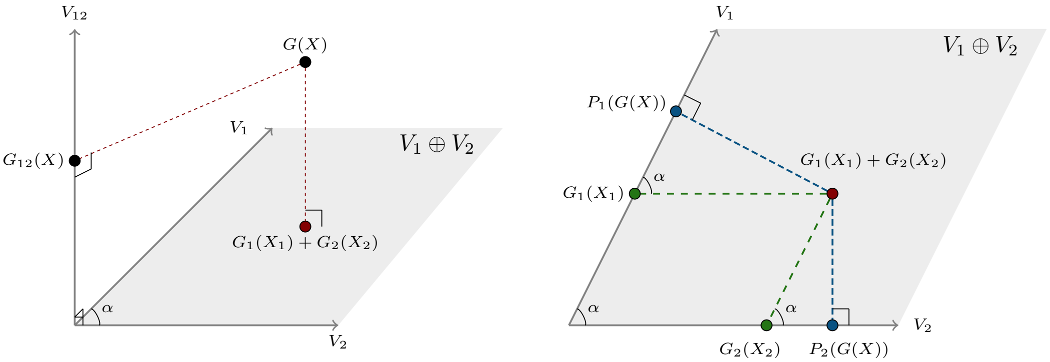
$$G_A(X_A) = Q_A(G(X)).$$



# Illustration* $\mathbb{L}^2_0(\sigma_{12})$

Hence, for any $G(X) \in \mathbb{L}^2(\sigma_X)$, one has that, $\forall A \in \mathcal{P}_D$

$$G_A(X_A) = Q_A(G(X)).$$




**The oblique projection $Q_A$ usually differ from the orthogonal projections $P_A$**

# Oblique and orthogonal projections

In fact,

**Proposition .**

Under Assumptions 1 and 2,

$$P_A(G(X)) = Q_A(G(X)) \text{ a.s. } , \forall A \in \mathcal{P}_D \quad \Longleftrightarrow \quad X \text{ is mutually independent.}$$

# Oblique and orthogonal projections

In fact,

**Proposition** .

Under Assumptions 1 and 2,

$$P_A(G(X)) = Q_A(G(X)) \text{ a.s. } , \forall A \in \mathcal{P}_D \quad \iff \quad X \text{ is mutually independent.}$$

This comes from the fact that **the subspaces $V_A$ are <u>all</u> pairwise orthogonal if and only if the inputs are mutually independent**

# Oblique and orthogonal projections

In fact,

**Proposition .**

Under Assumptions 1 and 2,

$$P_A(G(X)) = Q_A(G(X)) \text{ a.s. } , \forall A \in \mathcal{P}_D \quad \iff \quad X \text{ is mutually independent.}$$

This comes from the fact that **the subspaces $V_A$ are <u>all</u> pairwise orthogonal if and only if the inputs are mutually independent**

But, under Assumptions 1 and 2, they **<u>may not be all</u> orthogonal**

# Oblique and orthogonal projections

In fact,

**Proposition .**

Under Assumptions 1 and 2,

$$P_A(G(X)) = Q_A(G(X)) \text{ a.s. }, \forall A \in \mathcal{P}_D \quad \iff \quad X \text{ is mutually independent.}$$

This comes from the fact that **the subspaces $V_A$ are all pairwise orthogonal if and only if the inputs are mutually independent**

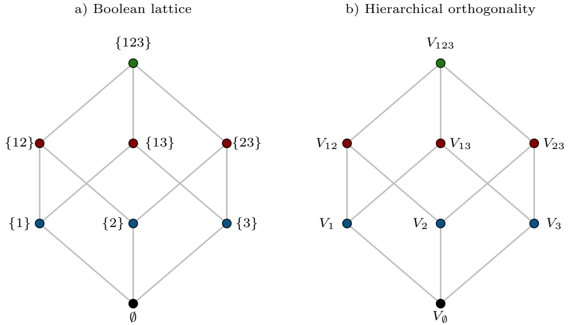
But, under Assumptions 1 and 2, they **may not be all orthogonal**

**There is a more visual way to illustrate that**

# Boolean lattice and hierarchical orthogonality

Our decomposition is **over the power-set $\mathcal{P}_D$**, **and this is not trivial**

☞ Endowed with the **binary relation** $\subseteq$, $(\mathcal{P}_D, \subseteq)$ forms **a Boolean lattice**




The subspaces $\{V_A\}_{A \in \mathcal{P}_D}$ are **hierarchically orthogonal** **by design**

☞ They form the same algebraic structure **w.r.t. to** $\perp$

## More projectors

Recall that:

- $Q_A$ is the **canonical oblique projection** onto $V_A$
- $P_A$ is the **orthogonal projection** onto $V_A$

## More projectors

Recall that:

- $Q_A$ is the **canonical oblique projection** onto $V_A$
- $P_A$ is the **orthogonal projection** onto $V_A$

**But we're more familiar with projections onto $\mathbb{L}^2(\sigma_A)$...**
☞ Conditional expectation operators, for example

# More projectors

Recall that:

- $Q_A$ is the **canonical oblique projection** onto $V_A$
- $P_A$ is the **orthogonal projection** onto $V_A$

**But we're more familiar with projections onto $\mathbb{L}^2(\sigma_A)$...**
☞ Conditional expectation operators, for example

**(Canonical) oblique projection onto $\mathbb{L}^2(\sigma_A)$:**

$$\mathbb{M}_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \quad G(X) \mapsto \sum_{B \in \mathcal{P}_A} G_B(X_B)$$

# More projectors

Recall that:

- $Q_A$ is the **canonical oblique projection** onto $V_A$
- $P_A$ is the **orthogonal projection** onto $V_A$

**But we're more familiar with projections onto $\mathbb{L}^2(\sigma_A)$...**
☞ Conditional expectation operators, for example

**(Canonical) oblique projection onto $\mathbb{L}^2(\sigma_A)$:**

$$\mathbb{M}_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \quad G(X) \mapsto \sum_{B \in \mathcal{P}_A} G_B(X_B)$$

**Orthogonal projection onto $\mathbb{L}^2(\sigma_A)$:**

$$\mathbb{E}_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \text{ with } \mathrm{Ran}(\mathbb{E}_A) = \mathbb{L}^2(\sigma_A) \text{ and } \mathrm{Ker}(P_A) = \mathbb{L}^2(\sigma_A)^{\perp},$$

a.k.a **the conditional expectation w.r.t. to $X_A$** (i.e., $\mathbb{E}[. \mid X_A]$).

# More projectors

Recall that:

- $Q_A$ is the **canonical oblique projection** onto $V_A$
- $P_A$ is the **orthogonal projection** onto $V_A$

**But we're more familiar with projections onto $\mathbb{L}^2(\sigma_A)$...**
☞ Conditional expectation operators, for example

**(Canonical) oblique projection onto $\mathbb{L}^2(\sigma_A)$:**

$$\mathbb{M}_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \quad G(X) \mapsto \sum_{B \in \mathcal{P}_A} G_B(X_B)$$

**Orthogonal projection onto $\mathbb{L}^2(\sigma_A)$:**

$$\mathbb{E}_A : \mathbb{L}^2(\sigma_X) \to \mathbb{L}^2(\sigma_X), \quad \text{with } \operatorname{Ran}(\mathbb{E}_A) = \mathbb{L}^2(\sigma_A) \text{ and } \operatorname{Ker}(P_A) = \mathbb{L}^2(\sigma_A)^{\perp},$$

a.k.a **the conditional expectation w.r.t. to $X_A$** (i.e., $\mathbb{E}[. \mid X_A]$).

**Can we characterize $Q_A$ w.r.t. $\mathbb{M}_A$?**

# Generalized Möbius inversion

Because $(\mathcal{P}_D, \subseteq)$ forms a **Boolean lattice**, yes!

**Corollary** *(Möbius inversion on power-sets (Rota 1964)).*

For any two set functions:

$$f : \mathcal{P}_D \to \mathbb{A}, \quad g : \mathcal{P}_D \to \mathbb{A},$$

valued in an abelian group $\mathbb{A}$, the following equivalence holds:

$$f(A) = \sum_{B \in \mathcal{P}_A} g(B), \quad \forall A \in \mathcal{P}_D \quad \iff \quad g(A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} f(B), \quad \forall A \in \mathcal{P}_D.$$

☞ Analogous to the *inclusion-exclusion principle*

In our case, we have, **by definition of the oblique projection onto** $\mathbb{L}^2(\sigma_A)$, that

$$\mathbb{M}_A(G(X)) = \sum_{B \in \mathcal{P}_A} G_B(X_B), \quad \forall A \in \mathcal{P}_D,$$

which is equivalent to

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B(G(X)), \quad \forall A \in \mathcal{P}_D$$

# Generalized Hoeffding decomposition

Hoeffding (1948) found that **for mutually independent inputs**:

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{E}_B \left[ G(X) \right], \quad \forall A \in \mathcal{P}_D$$

# Generalized Hoeffding decomposition

Hoeffding (1948) found that **for mutually independent inputs**:

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{E}_B \left[ G(X) \right], \quad \forall A \in \mathcal{P}_D$$

**Under Assumptions 1 and 2**, we have that:

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B \left[ G(X) \right], \quad \forall A \in \mathcal{P}_D$$

# Generalized Hoeffding decomposition

Hoeffding (1948) found that **for mutually independent inputs**:

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{E}_B \left[ G(X) \right], \quad \forall A \in \mathcal{P}_D$$

**Under Assumptions 1 and 2**, we have that:

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B \left[ G(X) \right], \quad \forall A \in \mathcal{P}_D$$

In addition:

**Proposition .**

Under Assumptions 1 and 2,

$$\mathbb{M}_A \left[ G(X) \right] = \mathbb{E}_A \left[ G(X) \right] \text{ a.s. } , \forall A \in \mathcal{P}_D \quad \Longleftrightarrow \quad X \text{ is mutually independent.}$$

# Generalized Hoeffding decomposition

Hoeffding (1948) found that **for mutually independent inputs**:

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{E}_B \left[ G(X) \right], \quad \forall A \in \mathcal{P}_D$$

**Under Assumptions 1 and 2**, we have that:

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B \left[ G(X) \right], \quad \forall A \in \mathcal{P}_D$$

In addition:

**Proposition .**

Under Assumptions 1 and 2,

$$\mathbb{M}_A \left[ G(X) \right] = \mathbb{E}_A \left[ G(X) \right] \text{ a.s. } , \forall A \in \mathcal{P}_D \quad \iff \quad X \text{ is mutually independent.}$$

**Our approach generalizes Hoeffding's original decomposition!**

# Illustrative example: Two Bernoulli inputs

Let $X = (X_1, X_2)$, where

$$X_1 \sim \mathcal{B}\left(q_1\right), \quad \text{and } X_2 \sim \mathcal{B}\left(q_2\right)$$

The joint law of $X$ can be characterized using **three parameters**:

$$p_{00} = 1 - q_1 - q_2 + \rho, \quad p_{01} = q_2 - \rho, \quad p_{10} = q_1 - \rho, \quad p_{11} = \rho$$

where $p_{ij} = \mathbb{P}\left(\{X_1 = i\} \cap \{X_2 = j\}\right)$

The functions of $X$ $G : \{0,1\}^2 \to \mathbb{R}$ can be expressed as a vector

$$G = \left(G_{00}, G_{01}, G_{10}, G_{11}\right)^\top, \text{ where } G_{ij} = G(i,j)$$

# Illustrative example: Two Bernoulli inputs

Let $X = (X_1, X_2)$, where

$$X_1 \sim \mathcal{B}\left(q_1\right), \quad \text{and } X_2 \sim \mathcal{B}\left(q_2\right)$$

The joint law of $X$ can be characterized using **three parameters**:

$$p_{00} = 1 - q_1 - q_2 + \rho, \quad p_{01} = q_2 - \rho, \quad p_{10} = q_1 - \rho, \quad p_{11} = \rho$$

where $p_{ij} = \mathbb{P}\left(\{X_1 = i\} \cap \{X_2 = j\}\right)$

The functions of $X$ $G : \{0,1\}^2 \to \mathbb{R}$ can be expressed as a vector

$$G = \left(G_{00}, G_{01}, G_{10}, G_{11}\right)^\top, \text{ where } G_{ij} = G(i,j)$$

**In this case, we can compute everything analytically**

It requires solving 13 equations with 13 unknowns*

*https://github.com/milidris/GeneralizedAnova

# Feshchenko matrix and the Fréchet bounds

For the **Feshchenko matrix** $\Delta$ to be definite positive, one has that:

$$\max\left\{0, q_1 q_2 - \sqrt{q_1 q_2 (1-q_1)(1-q_2)}\right\} < \rho < \min\left\{1, q_1 q_2 - \sqrt{q_1 q_2 (1-q_1)(1-q_2)}\right\}$$

# Feshchenko matrix and the Fréchet bounds

For the **Feshchenko matrix** $\Delta$ to be definite positive, one has that:

$$\max\left\{0, q_1 q_2 - \sqrt{q_1 q_2(1-q_1)(1-q_2)}\right\} < \rho < \min\left\{1, q_1 q_2 - \sqrt{q_1 q_2(1-q_1)(1-q_2)}\right\}$$

However, the **classical Fréchet bounds for $\rho$ for bivariate Bernoulli random variables** (Joe 1997, p.210) are equal to

$$\max\{0, q_1 + q_2 - 1\} \leq \rho \leq \min\{q1, q2\},$$

and are **more restrictive than the previous bounds**

# Feshchenko matrix and the Fréchet bounds

For the **Feshchenko matrix** $\Delta$ to be definite positive, one has that:

$$\max\left\{0, q_1q_2 - \sqrt{q_1q_2(1-q_1)(1-q_2)}\right\} < \rho < \min\left\{1, q_1q_2 - \sqrt{q_1q_2(1-q_1)(1-q_2)}\right\}$$

However, the **classical Fréchet bounds for $\rho$ for bivariate Bernoulli random variables** (Joe 1997, p.210) are equal to

$$\max\{0, q_1 + q_2 - 1\} \leq \rho \leq \min\{q1, q2\},$$

and are **more restrictive than the previous bounds**

**$\rho$ strictly contained in the Fréchet bounds $\implies$ Assumptions 1 and 2 hold**

**The generalized decomposition holds for any (non-comonotone) copula between two Bernoulli random variables**

# Main take-aways

☞ Under **mild assumptions** on the random inputs $X$, for any $G(X) \in \mathbb{L}^2(\sigma_X)$,

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A)$$

(Virtually) any model can be **decomposed** as a **sum of (pure) interactions of increasing dimensionality**

# Main take-aways

☞ Under **mild assumptions** on the random inputs $X$, for any $G(X) \in \mathbb{L}^2(\sigma_X)$,

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A)$$

(Virtually) any model can be **decomposed** as a **sum of (pure) interactions of increasing dimensionality**

☞ To generalize Hoeffding's decomposition, we relied on **Feshchenko matrices**
Which **control the maximal angles** between **Lebesgue spaces generated by the subsets of inputs**

# Main take-aways

☞ Under **mild assumptions** on the random inputs $X$, for any $G(X) \in \mathbb{L}^2(\sigma_X)$,

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A)$$

(Virtually) any model can be **decomposed** as a **sum of (pure) interactions of increasing dimensionality**

☞ To generalize Hoeffding's decomposition, we relied on **Feshchenko matrices**
Which **control the maximal angles** between **Lebesgue spaces generated by the subsets of inputs**

☞ $G_A(X_A)$ are characterized using **orthocanonical projections** onto $\mathbb{L}^2(\sigma_A)$:

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B(G(X))$$

Thanks to the **algebraic structure of the power-set**

# Main take-aways

☞ Under **mild assumptions** on the random inputs $X$, for any $G(X) \in \mathbb{L}^2(\sigma_X)$,

$$G(X) = \sum_{A \in \mathcal{P}_D} G_A(X_A)$$

(Virtually) any model can be **decomposed** as a **sum of (pure) interactions of increasing dimensionality**

☞ To generalize Hoeffding's decomposition, we relied on **Feshchenko matrices**
Which **control the maximal angles** between **Lebesgue spaces generated by the subsets of inputs**

☞ $G_A(X_A)$ are characterized using **orthocanonical projections** onto $\mathbb{L}^2(\sigma_A)$:

$$G_A(X_A) = \sum_{B \in \mathcal{P}_A} (-1)^{|A|-|B|} \mathbb{M}_B(G(X))$$

Thanks to the **algebraic structure of the power-set**

**Non-linear problem in $d$ dimensions becomes linear in $2^d$ dimensions**

# Some perspectives

☞ **Main challenge: Estimating the orthocanonical projections from observations**

They sure look a lot like conditional expectations...

# Some perspectives

☞ **Main challenge: Estimating the orthocanonical projections from observations**

They sure look a lot like conditional expectations...

☞ **Application to influence quantification**

Global and local sensitivity analysis, model interpretability, fairness assessment...

# Some perspectives

☞ **Main challenge: Estimating the orthocanonical projections from observations**

They sure look a lot like conditional expectations...

☞ **Application to influence quantification**

Global and local sensitivity analysis, model interpretability, fairness assessment...

☞ **Beyond orthocanonicality**

The properties of a different choice of complement in the "centering process"

e.g., Köhler, Rügamer, and Schmid (2024) with "stacked orthogonality" conditions

## Some perspectives

☞ **Main challenge: Estimating the orthocanonical projections from observations**

They sure look a lot like conditional expectations...

☞ **Application to influence quantification**

Global and local sensitivity analysis, model interpretability, fairness assessment...

☞ **Beyond orthocanonicality**

The properties of a different choice of complement in the "centering process"

e.g., Köhler, Rügamer, and Schmid (2024) with "stacked orthogonality" conditions

☞ **Beyond the power-set**

Other algebraic structures to model different data generating processes

Links with causality and probabilistic graphical models

# References i


Axler, S. 2015. *Linear Algebra Done Right* [in en]. Undergraduate Texts in Mathematics. Cham: Springer International Publishing. ISBN: 978-3-319-11079-0 978-3-319-11080-6. https://doi.org/10.1007/978-3-319-11080-6. https://link.springer.com/10.1007/978-3-319-11080-6.

Bryc, W. 1984. "Conditional expectation with respect to dependent sigma-fields." In *Proceedings of VII conference on Probability Theory,* 409–411. https://homepages.uc.edu/~brycwz/preprint/Brasov-1982.pdf.

———. 1996. "Conditional Moment Representations for Dependent Random Variables." Publisher: Institute of Mathematical Statistics and Bernoulli Society, *Electronic Journal of Probability* 1 (none): 1–14. ISSN: 1083-6489, 1083-6489. https://doi.org/10.1214/EJP.v1-7. https://projecteuclid.org/journals/electronic-journal-of-probability/volume-1/issue-none/Conditional-Moment-Representations-for-Dependent-Random-Variables/10.1214/EJP.v1-7.full.

Chastaing, G., F. Gamboa, and C. Prieur. 2012. "Generalized Hoeffding-Sobol decomposition for dependent variables - application to sensitivity analysis." Publisher: Institute of Mathematical Statistics and Bernoulli Society, *Electronic Journal of Statistics* 6, no. none (January): 2420–2448. ISSN: 1935-7524, 1935-7524. https://doi.org/10.1214/12-EJS749. https://projecteuclid.org/journals/electronic-journal-of-statistics/volume-6/issue-none/Generalized-Hoeffding-Sobol-decomposition-for-dependent-variables---application/10.1214/12-EJS749.full.

Da Veiga, S., F. Gamboa, B. Iooss, and C. Prieur. 2021. *Basics and Trends in Sensitivity Analysis. Theory and Practice in R.* SIAM. Computational Science / Engineering.

Dauxois, J, G. M Nkiet, and Y Romain. 2004. "Canonical analysis relative to a closed subspace." *Linear Algebra and its Applications,* Tenth Special Issue (Part 1) on Linear Algebra and Statistics, 388:119–145. ISSN: 0024-3795. https://doi.org/10.1016/j.laa.2004.02.036. https://www.sciencedirect.com/science/article/pii/S0024379504001107.

Dixmier, J. 1949. "Étude sur les variétés et les opérateurs de Julia, avec quelques applications" [in fre]. *Bulletin de la Société Mathématique de France* 77:11–101. http://eudml.org/doc/86830.

Feshchenko, I. 2020. *When is the sum of closed subspaces of a Hilbert space closed?* https://doi.org/10.48550/arXiv.2012.08688. arXiv: 2012.08688 [math.FA].

Friedrichs, K. 1937. "On Certain Inequalities and Characteristic Value Problems for Analytic Functions and For Functions of Two Variables." Publisher: American Mathematical Society, *Transactions of the American Mathematical Society* 41 (3): 321–364. ISSN: 0002-9947. https://doi.org/10.2307/1989786. https://www.jstor.org/stable/1989786.

Galántai, A. 2004. *Projectors and Projection Methods.* Boston, MA: Springer US. ISBN: 978-1-4613-4825-2. https://doi.org/10.1007/978-1-4419-9180-5. http://link.springer.com/10.1007/978-1-4419-9180-5.

Gebelein, H. 1941. "Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung" [in de]. *ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik* 21 (6): 364–379. ISSN: 00442267, 15214001. https://doi.org/10.1002/zamm.19410210604. https://onlinelibrary.wiley.com/doi/10.1002/zamm.19410210604.

Hart, J., and P. A. Gremaud. 2018. "An approximation theoretic perspective of Sobol' indices with dependent variables" [in English]. Publisher: Begel House Inc. *International Journal for Uncertainty Quantification* 8 (6). ISSN: 2152-5080, 2152-5099. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2018026498. https://www.dl.begellhouse.com/journals/52034eb04b657aea,23dc16a4645b89c9,61d464a51b6bf191.html.

Hoeffding, W. 1948. "A Class of Statistics with Asymptotically Normal Distribution." *The Annals of Mathematical Statistics* 19 (3): 293–325. ISSN: 0003-4851, 2168-8990. https://doi.org/10.1214/aoms/1177730196. https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-19/issue-3/A-Class-of-Statistics-with-Asymptotically-Normal-Distribution/10.1214/aoms/1177730196.full.

Hooker, G. 2007. "Generalized Functional ANOVA Diagnostics for High-Dimensional Functions of Dependent Variables" [in en]. *Journal of Computational and Graphical Statistics* 16 (3): 709–732. http://www.jstor.org/stable/27594267.

Joe, H. 1997. *Multivariate Models and Multivariate Dependence Concepts.* New York: Chapman / Hall/CRC. ISBN: 978-0-367-80389-6. https://doi.org/10.1201/9780367803896.

Kallenberg, O. 2021. *Foundations of modern probability.* Probability theory and stochastic modelling. Cham, Switzerland: Springer. ISBN: 978-3-030-61871-1. https://doi.org/10.1007/978-3-030-61871-1.

Koyak, R. A. 1987. "On Measuring Internal Dependence in a Set of Random Variables." Publisher: Institute of Mathematical Statistics, *The Annals of Statistics* 15 (3): 1215–1228. ISSN: 0090-5364, 2168-8966. https://doi.org/10.1214/aos/1176350501. https://projecteuclid.org/journals/annals-of-statistics/volume-15/issue-3/On-Measuring-Internal-Dependence-in-a-Set-of-Random-Variables/10.1214/aos/1176350501.full.

# References iv


Kuo, F. Y., I. H. Sloan, G. W. Wasilkowski, and H. Woźniakowski. 2009. "On decompositions of multivariate functions" [in en]. *Mathematics of Computation* 79, no. 270 (November): 953–966. ISSN: 0025-5718. https://doi.org/10.1090/S0025-5718-09-02319-9. http://www.ams.org/journal-getitem?pii=S0025-5718-09-02319-9.

Köhler, D., D. Rügamer, and M. Schmid. 2024. *Achieving interpretable machine learning by functional decomposition of black-box models into explainable predictor effects.* ArXiv:2407.18650 [cs, stat], July. Accessed August 26, 2024. http://arxiv.org/abs/2407.18650.

Lebrun, R., and A. Dutfoy. 2009a. "A generalization of the Nataf transformation to distributions with elliptical copula." *Probabilistic Engineering Mechanics* 24 (2): 172–178. ISSN: 0266-8920. https://doi.org/10.1016/j.probengmech.2008.05.001. https://www.sciencedirect.com/science/article/pii/S0266892008000507.

———. 2009b. "Do Rosenblatt and Nataf isoprobabilistic transformations really differ?" *Probabilistic Engineering Mechanics* 24 (4): 577–584. ISSN: 0266-8920. https://doi.org/10.1016/j.probengmech.2009.04.006. https://www.sciencedirect.com/science/article/pii/S0266892009000307.

Malliavin, P. 1995. *Integration and Probability.* Vol. 157. Graduate Texts in Mathematics. New York, NY: Springer. ISBN: 978-1-4612-8694-3. https://doi.org/10.1007/978-1-4612-4202-4. http://link.springer.com/10.1007/978-1-4612-4202-4.

Mara, Thierry A., Stefano Tarantola, and Paola Annoni. 2015. "Non-parametric methods for global sensitivity analysis of model output with dependent inputs" [in en]. *Environmental Modelling & Software* 72 (October): 173–183. ISSN: 13648152. https://doi.org/10.1016/j.envsoft.2015.07.010. https://linkinghub.elsevier.com/retrieve/pii/S1364815215300153.

Peccati, Giovanni. 2004. "Hoeffding-ANOVA decompositions for symmetric statistics of exchangeable observations." Publisher: Institute of Mathematical Statistics, *The Annals of Probability* 32 (3): 1796–1829. ISSN: 0091-1798, 2168-894X. https://doi.org/10.1214/009117904000000405. https://projecteuclid.org/journals/annals-of-probability/volume-32/issue-3/Hoeffding-ANOVA-decompositions-for-symmetric-statistics-of-exchangeable-observations/10.1214/009117904000000405.full.

Rabitz, H., and O. Aliş. 1999. "General foundations of high-dimensional model representations" [in en]. *Journal of Mathematical Chemistry* 25 (2): 197–233. ISSN: 1572-8897. https://doi.org/10.1023/A:1019188517934. https://doi.org/10.1023/A:1019188517934.

Rota, G. C. 1964. "On the foundations of combinatorial theory I. Theory of Möbius Functions." *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 2 (4): 340–368. ISSN: 1432-2064. https://doi.org/10.1007/BF00531932.

Sidák, Z. 1957. "On Relations Between Strict-Sense and Wide-Sense Conditional Expectations." *Theory of Probability & Its Applications* 2 (2): 267–272. ISSN: 0040-585X. https://doi.org/10.1137/1102020. https://epubs.siam.org/doi/abs/10.1137/1102020.

Thank you for your attention!

Any questions?

**Let's talk about variance decomposition.**

We propose **two complementary approaches** for decomposing $\mathbb{V}(G(X))$ based on this generalized decomposition.

**Organic variance decomposition:** separate **pure interaction effects** to **dependence effects**. The dependence structure of $X$ is **unwanted**, and one wishes to study its effects.

**Orthocanonical variance decomposition:** the dependence structure of $X$ is **inherent in the uncertainty modeling** of the studied phenomenon. It amounts to quantify **structural** and **correlative** effects.

# Organic variance decomposition: Pure interaction

The notion of pure interaction is intrinsically linked with the notion of mutual independence.

Let $\widetilde{X} = (\widetilde{X}_1, \dots, \widetilde{X}_d)^\top$ be the random vector such that

$$\widetilde{X}_i \stackrel{d}{=} X_i, \qquad \text{and } \widetilde{X} \text{ is mutually independent.}$$

**Definition** *Pure interaction*. For every $A \in \mathcal{P}_D$, define the **pure interaction of $X_A$ on $G(X)$ as**

$$S_A = \frac{\mathbb{V}\left(P_A(G(\widetilde{X}))\right)}{\mathbb{V}\left(G(\widetilde{X})\right)} \times \mathbb{V}\left(G(X)\right).$$

These indices are the **Sobol' indices** computed on the mutually independent version of $X$.

This approach **strongly resembles the "independent Sobol' indices"** proposed by Mara, Tarantola, and Annoni (2015).

(see, also, Lebrun and Dutfoy (2009a, 2009b))

# Organic variance decomposition: Dependence effects

Recall the following result:

**Proposition** . Under Assumptions 1 and 2,

$$P_A(G(X)) = Q_A(G(X)) \text{ a.s. } , \forall A \in \mathcal{P}_D \quad \iff \quad X \text{ is mutually independent.}$$

Which motivates the definition of dependence effects.

**Definition** *Dependence effects*. For every $A \in \mathcal{P}_D$, define the **dependence effects of $X_A$ on $G(X)$ as**

$$S_A^D = \mathbb{E}\left[(Q_A(G(X)) - P_A(G(X)))^2\right].$$

**Proposition** . Under Assumptions 1 and 2,

$$S_A^D = 0, \forall A \in \mathcal{P}_D, \quad \iff \quad X \text{ is mutually independent.}$$

**What do they sum up to ?...**

Probably some interesting global multivariate dependence effect measure!

# Canonical variance decomposition

The structural effects represent the variance of each of the $G_A(X_A)$. It amounts to perform a **covariance decomposition** (Hart and Gremaud 2018; Da Veiga et al. 2021).

**Definition** *Structural effects*. For every $A \in \mathcal{P}_D$, define the **structural effects of $X_A$ on $G(X)$ as**

$$S_A^U = \mathbb{V}\left(G_A(X_A)\right).$$

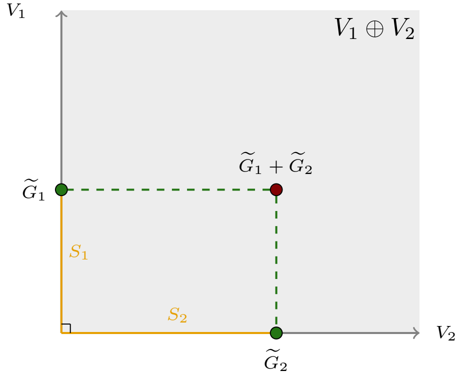
The **correlative effects** represent the part of variance that is due to the correlation between the $G_A(X_A)$

**Definition** *Correlative effects*. For every $A \in \mathcal{P}_D$, define the **correlative effects of $X_A$ on $G(X)$ as**

$$S_A^C = \mathrm{Cov}\left(G_A(X_A), \sum_{B \in \mathcal{P}_D : B \neq A} G_B(X_B)\right)$$

# Variance decomposition: Intuition

Pure interaction effects




Structural and dependence effects