
Generalized Functional ANOVA: A Complete Theoretical Framework

Baptiste Ferrere · Nicolas Bousquet · Fabrice Gamboa

Jean-Michel Loubes · Joseph Muré

ETICS 2026



Authors



B. Ferrere*

EDF R&D, SINCLAIR
Univ. Toulouse



N. Bousquet†

EDF R&D, SINCLAIR
Univ. Sorbonne



F. Gamboa†

Univ. Toulouse
ANITI
Univ. Medellín



JM. Loubes†

Univ. Toulouse
ANITI
Inria REGALIA



J. Muré†

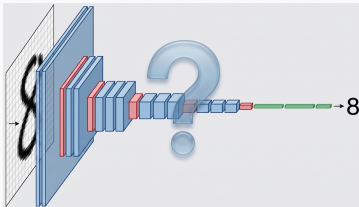
EDF R&D, SINCLAIR

* Corresponding author † PhD Advisors

✉ baptiste.ferrere@edf.fr · [bapfr.github.io](https://github.com/bapfr) · [BapFr](https://www.bapfr.fr) · [Google Scholar](https://scholar.google.com/citations?user=bapfr)

eXplainable Artificial Intelligence (XAI)

Millions of non-linear operations



LLM hallucinations

User: How many *r*'s in *strawberry*?

Model: There are **two** *r*'s in strawberry.

!

Feature importance

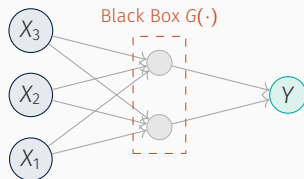


Misclassification



Consider a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ and a random vector $\mathbf{X} := (X_1, \dots, X_d) \in \mathbb{R}^d$.

- 📦 In standard ML, f is a **black-box** model and we observe n realizations of $\mathbf{X}$.
- ⚙️ f can be highly complex, and even when its structure is accessible, it is rarely interpretable.
- 🔗 $\mathbf{X}$ typically follows an **unknown distribution** with highly correlated components.



The many additive explanations

Additive explanation. Decompose the prediction $f(\mathbf{x})$ for a given input $\mathbf{x} := (x_1, \dots, x_d)$ in term of **feature contributions**:

$$f(\mathbf{x}) = \text{Contrib.}(x_1) + \text{Contrib.}(x_2) + \dots + \text{Contrib.}(x_d)$$

⇒ we analyse the impact of each feature on the output.

Some well known methods:

- **Sampling-based:** LIME (Ribeiro'16), SHAP (Lundberg & Lee'17)
- **Gradient-based:** DeepLIFT (Shrikumar'17), LRP (Bach'15), IG (Sundararajan'17)

The many additive decompositions

Additive decomposition. Assume that the function f has an **additive representation** and learn the following representation.

Some well known methods:

- **Generalized Additive Models:** GAM (Tibshirani), SpAM (Raskutti), NAM (Agarwal)

$$f(\mathbf{x}) = f_0 + f_1(x_1) + \cdots + f_d(x_d)$$

- **Nonparametric interaction models:** (Stone; Huang), EBM (Nori), NIM with DNN (Bhattacharaya)

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^d f_i(x_i) + \sum_{i < j} f_{i,j}(x_i, x_j)$$

The Independent Case

For mutually independent $X_1, \dots, X_d$, any $f \in L^2$ admits a **unique** decomposition

$$f(\mathbf{X}) = \sum_{S \subseteq D} f_S(\mathbf{X}_S), \quad D = \{1, \dots, d\}.$$

Orthogonality

$$S \neq T \implies \mathbb{E}[f_S(\mathbf{X}_S) f_T(\mathbf{X}_T)] = 0$$

Components pairwise orthogonal in L^2 .

Closed-form (Möbius)

$$f_S(\mathbf{x}_S) = \sum_{T \subseteq S} (-1)^{|S|-|T|} \mathbb{E}[f(\mathbf{X}) \mid \mathbf{X}_T = \mathbf{x}_T]$$

Fourier–ANOVA representation on the continuous cube

Representation theorem

Let $(\phi_k)_{k \in \mathbb{N}}$ be an orthonormal basis of $L^2([0, 1], \lambda)$ such that $\phi_0 \equiv 1$, and let $\mathbb{N}_+ = \{1, 2, \dots\}$. Every $f \in L^2([0, 1]^d, \lambda^{\otimes d})$ admits the unique ANOVA decomposition

$$f(\mathbf{x}) = \sum_{S \subseteq [d]} f_S(\mathbf{x}_S),$$

where, for every nonempty $S \subseteq [d]$,

$$f_S(\mathbf{x}_S) = \sum_{\alpha \in \mathbb{N}_+^{|S|}} \hat{f}_S(\alpha) \prod_{j \in S} \phi_{\alpha_j}(x_j),$$

with

$$\hat{f}_S(\alpha) = \int_{[0, 1]^d} f(\mathbf{x}) \prod_{j \in S} \phi_{\alpha_j}(x_j) d\mathbf{x}.$$

Local Explainability: SHAP

 **Players:** local input features $\{\mathbf{x}_1^*, \dots, \mathbf{x}_d^*\}$.

 **Value function:** conditional expectation of the prediction

$$v_{\mathbf{x}^*}(S) = \mathbb{E}[f(\mathbf{X}) \mid \mathbf{X}_S = \mathbf{x}_S^*] \stackrel{\text{indep}}{=} \sum_{T \subseteq S} f_T(\mathbf{x}_T^*)$$

From Functional ANOVA to SHAP

$$\text{shap}_i(\mathbf{x}) := \sum_{\substack{S \subseteq D \\ i \in S}} \frac{f_S(\mathbf{x}_S)}{|S|} \quad \implies \quad f(\mathbf{x}^*) = \underbrace{f_\emptyset}_{\text{shap}_0} + \sum_{i=1}^d \text{shap}_i(\mathbf{x}^*)$$

Global Explainability: Shapley Effects

 **Players:** the random variables $\{X_1, \dots, X_d\}$.

 **Value function:** proportion of explained variance (i.e. standard Sobol' indices)

$$v(S) \stackrel{\text{indep}}{=} \frac{1}{\text{Var}(Y)} \sum_{T \subseteq S} \text{Var}(f_T(X_T))$$

Global Analysis via Variance Shares

$$Sh_i = \frac{1}{\text{Var}(Y)} \sum_{\substack{S \subseteq D \\ i \in S}} \frac{\text{Var}(f_S(X_S))}{|S|}, \quad \sum_{i=1}^d Sh_i = 1.$$

Each Sh_i is the share of total variance attributable to X_i .

Loss of Independence $\Rightarrow$ Loss of Properties

When dependencies arise, the desirable properties break

✗ **Orthogonality lost:** components no longer uncorrelated.

$$S \neq T \implies \mathbb{E}[f_S(\mathbf{X}_S) f_T(\mathbf{X}_T)] = ?$$

✗ **Variance decomposition broken:** cross-covariance terms arise.

$$\text{Var}(Y) \neq \sum_{S \subseteq D} \text{Var}(f_S(\mathbf{X}_S))$$

✗ **ANOVA $\leftrightarrow$ SHAP link:** the equivalence $\text{shap}_i = \sum \frac{f_S}{|S|}$ no longer holds.

Two cases for Functional ANOVA

Independent case

Hoeffding'48 – Stone'94

Möbius transform:

$$f_S(\mathbf{X}_S) = \sum_{T \subseteq S} (-1)^{|S|-|T|} \mathbb{E}[f(\mathbf{X}) \mid \mathbf{X}_T]$$

- Closed form ✓
- Pairwise orthogonality ✓
- Variance decomposition ✓

Dependent case

Hooker'07, Chastaing'12, Il Idrissi'25

$$f_S(\mathbf{X}_S) = ?$$

no explicit construction

- Existence & uniqueness only
- Hierarchical orthogonality required
- Costly sampling-based approximations

The Variational Approach

Constrained Optimization (Hooker'07)

$$\min_{\{f_S\}_S} \mathbb{E} \left[\left(f(\mathbf{X}) - \sum_{S \subseteq D} f_S(\mathbf{X}_S) \right)^2 \right] \quad \text{s.t.} \quad \forall T \subsetneq S, \forall g_T(\mathbf{X}_T), \mathbb{E}[f_S(\mathbf{X}_S) g_T(\mathbf{X}_T)] = 0$$

Each $f_S(\mathbf{X}_S)$ must provide new information over any function $g_T(\mathbf{X}_T)$ of a lower set $T \subsetneq S$.

Generalization & Challenges

- ✓ If $\mathbf{X}_1, \dots, \mathbf{X}_d$ are independent, the solution recovers standard Functional ANOVA.
- ⚠ No closed-form solution, no Möbius inversion; exponential dimensionality.
- ⚠ Uniqueness only under specific conditions (Stone'94, Hooker'07, Chastaing'12, Il Idrissi'23).

Research Question

Given a d -dimensional random vector $\mathbf{X}$ with distribution $\mathbb{P}$, can we find a **decomposition basis** for $L^2(\mathbb{P})$ satisfying **hierarchical orthogonality**?

Algebraic Reformulation

Explicitly characterize the functional subspaces $\mathcal{H}_S$ such that

$$L^2(\mathbb{P}) = \bigoplus_{S \subseteq [d]} \mathcal{H}_S, \quad \mathcal{H}_S \subseteq L^2_S(\mathbb{P}) \quad \forall T \subsetneq S \subseteq D, \quad \mathcal{H}_S \perp \mathcal{H}_T$$

The Key Mechanism: Inverse Likelihood Weighting

Inverse Likelihood Weighting

$$\text{Hierarchical Orthogonality} = \frac{\text{Orthogonal Basis}}{\text{Likelihood}}$$

Idea

If $\mathbb{P} \ll \mu$, take $\mathcal{B}$ an orthogonal basis for $L^2(\mu)$, then when computing the scalar products, the density will simplify and lead to hierarchical orthogonality.

Bounded case

Let $\mathbf{X} \in [0, 1]^d$ a continuous random vector of joint density p satisfying:

$$0 < p_{\min} \leq p \leq p_{\max} < \infty.$$

Denote by e_j the j^{th} element of a basis of $L^2([0, 1], \lambda)$ such that $e_0 \equiv 1$ and the tensorized product of this basis is **an orthonormal basis** of $L^2([0, 1]^d, \lambda^{\otimes d})$ (e.g. Trigonometric basis or Legendre polynomials).

ANOVA-adapted Riesz basis

$$\phi_S^{\alpha}(\mathbf{X}) := \frac{\prod_{j \in S} e_{\alpha_j}(X_j)}{p_S(\mathbf{X}_S)}, \quad \alpha \in \mathbb{N}_+^{|S|}, S \subseteq D$$

Main Result

Main Theorem

$$\forall f \in L^2([0, 1]^d, P), \exists ! \boldsymbol{\theta} := ((\theta_S^\alpha)_{\alpha \in \mathbb{N}_+^{|S|}})_{S \subseteq [d]} : f(\mathbf{X}) = \sum_{S \subseteq [d]} \sum_{\alpha \in \mathbb{N}_+^{|S|}} \theta_S^\alpha \phi_S^\alpha(\mathbf{X}_S)$$

Hilbertian view point

We characterized all the underlying decomposing spaces $\mathcal{H}_S$ and showed that

$$\mathcal{H}_S = \overline{\text{span}} \left\{ \phi_S^\alpha \mid \alpha \in \mathbb{N}_+^{|S|} \right\}.$$

Nonparametric least-squares problem

If we obtain a basis with the *desirable* properties of Functional ANOVA, then computing the decomposition reduces to a **least-squares problem**.

Empirical Formulation

Given n realizations $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}$, with design matrix Φ_n and $\mathbf{y}_n := (f(\mathbf{X}^{(1)}), \dots, f(\mathbf{X}^{(n)}))^T$, we solve

$$\underbrace{\frac{1}{2n} \arg \min_{\boldsymbol{\theta}} \|\Phi_n \boldsymbol{\theta} - \mathbf{y}_n\|_2^2}_{\text{LS}} + \underbrace{\lambda_n \Omega(\boldsymbol{\theta})}_{\text{sparsity pen}} + \underbrace{\rho_n \zeta(\boldsymbol{\theta})}_{\text{smoothness pen}}.$$

Proof of concept: Hierarchical orthogonality

Let $T \subsetneq S \subseteq D$, let $\alpha \in \mathbb{N}_+^{|T|}$ and $\beta \in \mathbb{N}_+^{|S|}$.

$$\begin{aligned}\mathbb{E} \left[\left(\phi_T^{(\alpha)} \phi_S^{(\beta)} \right) (\mathbf{X}) \right] &= \mathbb{E} \left[\left(\phi_T^{(\alpha)} \phi_S^{(\beta)} \right) (\mathbf{X}_S) \right] \\&= \int \phi_T^{(\alpha)}(\mathbf{x}) \phi_S^{(\beta)}(\mathbf{x}) p_S(\mathbf{x}_S) d\mathbf{x}_S \\&= \int \frac{\prod_{i \in T} e_{\alpha_i}(\mathbf{x}_i)}{p_T(\mathbf{x}_T)} \cdot \frac{\prod_{j \in S} e_{\beta_j}(\mathbf{x}_j)}{p_S(\mathbf{x}_S)} \cdot p_S(\mathbf{x}_S) d\mathbf{x}_S \\&= \int \left(\frac{\prod_{i \in T} (e_{\alpha_i} \cdot e_{\beta_i})(\mathbf{x}_i)}{p_T(\mathbf{x}_T)} \right) \cdot \prod_{j \in S \setminus T} e_{\beta_j}(\mathbf{x}_j) d\mathbf{x}_S \\&= \int u_T(\mathbf{x}_T) d\mathbf{x}_T \cdot \prod_{j \in S \setminus T} \underbrace{\int e_{\beta_j} d\lambda}_{=0}\end{aligned}$$

The Key Equivalence

When the standard assumptions hold (non-vanishing pmf):

Generalized Functional ANOVA = Representation Theorem

In practice, it is an empirical least squares.

Technical Challenges

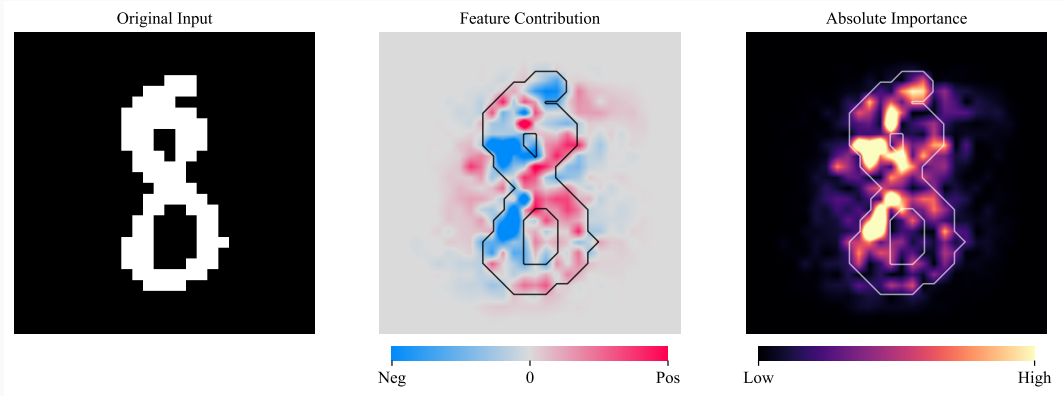
⚠ Zero probabilities (ill-posed ANOVA)

⚠ Curse of dimensionality

⚠ Numerical stability

⚠ Estimation of the pmf / density

Visualization on the MNIST database



Analytical continuous setting

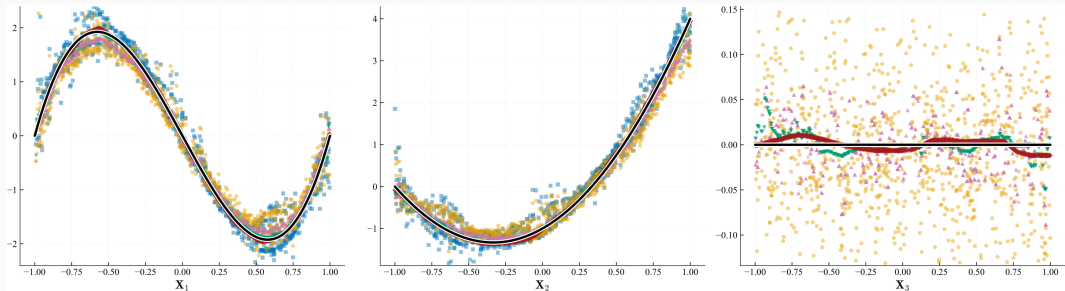


Figure 1: Estimated main effects in the analytical setting for X_1 , X_2 , and the irrelevant variable X_3 . All methods ([KernelSHAP](#), [TreeSHAP](#), [TreeHFD](#), [NAM](#), [EBM](#)) closely track the **theoretical ANOVA component** on X_1 and X_2 , and correctly assign a near-zero contribution to X_3 .

Conclusion

Under standard regularity assumption, we derive an **explicit decomposition basis** characterizing the formulation of the Generalized Functional ANOVA.

Limitations & Future Work

- Degenerate cases ($p = 0$).
- The curse of dimensionality.
- Numerical stability.
- Minimax analysis.

Thank you!

Papers

- [Fourier Analysis on the Boolean Hypercube via Hoeffding Functional Decomposition](#) (Preprint)
- [Exact Functional ANOVA Decomposition for Categorical Inputs Models](#) (ICML 2026, Oral)
- [Generalized Functional ANOVA: A Complete Theoretical Framework](#) (Preprint)



 Website

 baptiste.ferrere@edf.fr