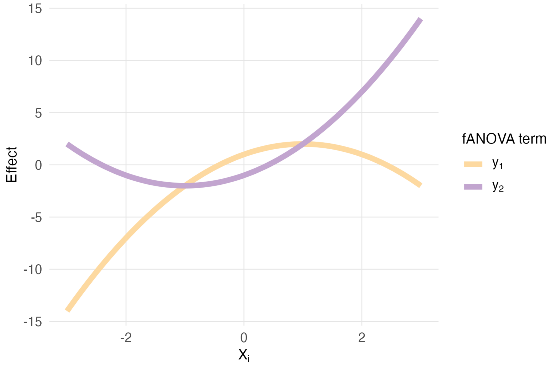
Functional ANOVA Decomposition

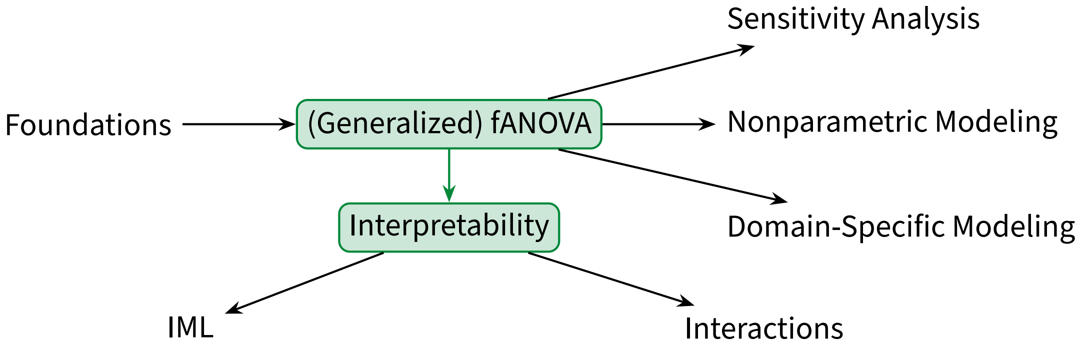
Juliet Fleischer

August 3, 2025



Outline

Overview of the fANOVA Research Field



Outline

Decompose model into interpretable components

General Form

$$y(\boldsymbol{X}) = \sum_{u \subseteq \{1,\ldots,N\}} y_u(\boldsymbol{X}_u)$$

Decompose model into interpretable components

General Form

$$y(\boldsymbol{X}) = \sum_{u \subseteq \{1,\ldots,N\}} y_u(\boldsymbol{X}_u)$$

- y : Model
- y_u : Component functions for subvector $\boldsymbol{X}_u$
- $\boldsymbol{X} := X_1, \ldots, X_N$ independent input variables

Decompose model into interpretable components

General Form

$$y(\boldsymbol{X}) = \sum_{u \subseteq \{1,\dots,N\}} y_u(\boldsymbol{X}_u)$$
$$= y_\emptyset + \left(y_{\{1\}}(\boldsymbol{X}_1) + \cdots + y_{\{N\}}(\boldsymbol{X}_N)\right)$$

- y : Model
- y_u : Component functions for subvector $\boldsymbol{X}_u$
- $\boldsymbol{X} := X_1, \dots, X_N$ independent input variables

Decompose model into interpretable components

General Form

$$y(\boldsymbol{X}) = \sum_{u \subseteq \{1,\ldots,N\}} y_u(\boldsymbol{X}_u)$$

$$= y_\emptyset + \left(y_{\{1\}}(\boldsymbol{X}_1) + \cdots + y_{\{N\}}(\boldsymbol{X}_N)\right)$$

$$+ \left(y_{\{1,2\}}(\boldsymbol{X}_1, \boldsymbol{X}_2) + y_{\{1,3\}}(\boldsymbol{X}_1, \boldsymbol{X}_3) + \ldots\right)$$

- y : Model
- y_u : Component functions for subvector $\boldsymbol{X}_u$
- $\boldsymbol{X} := X_1, \ldots, X_N$ independent input variables

Decompose model into interpretable components

General Form

$$
\begin{aligned}
y(\boldsymbol{X}) &= \sum_{u \subseteq \{1,\ldots,N\}} y_u(\boldsymbol{X}_u) \\
&= y_\emptyset + \left(y_{\{1\}}(\boldsymbol{X}_1) + \cdots + y_{\{N\}}(\boldsymbol{X}_N)\right) \\
&\quad + \left(y_{\{1,2\}}(\boldsymbol{X}_1, \boldsymbol{X}_2) + y_{\{1,3\}}(\boldsymbol{X}_1, \boldsymbol{X}_3) + \ldots\right) \\
&\quad + \left(y_{\{1,2,3\}}(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3) + \ldots\right) + \cdots + y_{\{1,\ldots,N\}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N)
\end{aligned}
$$

- y : Model
- y_u : Component functions for subvector $\boldsymbol{X}_u$
- $\boldsymbol{X} := X_1, \ldots, X_N$ independent input variables

Ensure Uniqueness and Interpretability of the Components

Strong Annihilating Conditions

$$\int_{\mathbb{R}} y_u(\boldsymbol{x}_u) f_{\{i\}}(x_i)\, d\nu(x_i) = 0 \quad \text{for } i \in u \neq \emptyset.$$

Ensure Uniqueness and Interpretability of the Components

Strong Annihilating Conditions

$$\int_{\mathbb{R}} y_u(\boldsymbol{x}_u) f_{\{i\}}(x_i)\, d\nu(x_i) = 0 \quad \text{for } i \in u \neq \emptyset.$$

- $f_{\{i\}}$: marginal density of variable X_i
- $d\nu(x_i)$: measure on $\mathbb{R}$

Ensure Uniqueness and Interpretability of the Components

Strong Annihilating Conditions

$$\int_{\mathbb{R}} y_u(\boldsymbol{x}_u) f_{\{i\}}(x_i)\, d\nu(x_i) = 0 \quad \text{for } i \in u \neq \emptyset.$$

- $f_{\{i\}}$: marginal density of variable X_i
- $d\nu(x_i)$: measure on $\mathbb{R}$

It follows:

$$\mathbb{E}[y_u(\boldsymbol{X}_u)] = 0$$

Ensure Uniqueness and Interpretability of the Components

Strong Annihilating Conditions

$$\int_{\mathbb{R}} y_u(\boldsymbol{x}_u) f_{\{i\}}(x_i)\, d\nu(x_i) = 0 \quad \text{for } i \in u \neq \emptyset.$$

- $f_{\{i\}}$: marginal density of variable X_i
- $d\nu(x_i)$: measure on $\mathbb{R}$

It follows:

$$\mathbb{E}[y_u(\boldsymbol{X}_u)] = 0$$

$$\mathbb{E}[y_u(\boldsymbol{X}_u) y_v(\boldsymbol{X}_v)] = 0 \quad (u \neq v)$$

Recursive Form of the Components

$$y_\emptyset = \int_{\mathbb{R}^N} y(\boldsymbol{x}) \prod_{i=1}^{N} f_{\{i\}}(x_i)\, d\nu(x_i) = \mathbb{E}[y(\boldsymbol{X})].$$

$$\begin{aligned} y_u(\boldsymbol{X}_u) &= \int_{\mathbb{R}^{N-|u|}} y(\boldsymbol{X}_u, \boldsymbol{x}_{-u}) \prod_{i=1, i\notin u}^{N} f_{\{i\}}(x_i)\, d\nu(x_i) - \sum_{v \subsetneq u} y_v(\boldsymbol{X}_v) \\ &= \mathbb{E}[y(\boldsymbol{X}_u, \boldsymbol{X}_{-u}) \mid \boldsymbol{X}_u = \boldsymbol{x}_u] - \sum_{v \subsetneq u} y_v(\boldsymbol{X}_v) \end{aligned}$$

Under independence:

- $f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}) = \prod_{i=1}^{N} f_{\{i\}}(x_i)\, d\nu(x_i)$
- $f_{\boldsymbol{X}}(\boldsymbol{x}) d\nu(\boldsymbol{x}_{-u}) = \prod_{i=1, i\notin u}^{N} f_{\{i\}}(x_i)\, d\nu(x_i)$

Recursive Form Example

- $N = 3, u = \emptyset$

Recursive Form Example

- $N = 3, u = \emptyset$

$$y_\emptyset = \int_{\mathbb{R}^3} y(\boldsymbol{x}) \prod_{i=1}^{3} f_{\{i\}}(x_i)\, d\nu(x_i) = \mathbb{E}[y(\boldsymbol{X})].$$

Recursive Form Example

- $N = 3, u = \emptyset$

$$y_{\emptyset} = \int_{\mathbb{R}^3} y(\boldsymbol{x}) \prod_{i=1}^{3} f_{\{i\}}(x_i)\, d\nu(x_i) = \mathbb{E}[y(\boldsymbol{X})].$$

- $u = \{1\} \rightarrow v = \emptyset$

Recursive Form Example

- $N = 3, u = \emptyset$

$$y_{\emptyset} = \int_{\mathbb{R}^3} y(\boldsymbol{x}) \prod_{i=1}^{3} f_{\{i\}}(x_i)\, d\nu(x_i) = \mathbb{E}[y(\boldsymbol{X})].$$

- $u = \{1\} \to v = \emptyset$

$$y_{\{1\}}(x_1) = \int_{\mathbb{R}^2} y(x_1, x_2, x_3) \prod_{i=2}^{3} f_{\{i\}}(x_i)\, d\nu(x_i) - y_{\emptyset} = \mathbb{E}[y(X_1, X_2, X_3) | X_1 = x_1] - y_{\emptyset}.$$

Recursive Form Example

- $N = 3, u = \emptyset$

$$y_{\emptyset} = \int_{\mathbb{R}^3} y(\boldsymbol{x}) \prod_{i=1}^{3} f_{\{i\}}(x_i)\, d\nu(x_i) = \mathbb{E}[y(\boldsymbol{X})].$$

- $u = \{1\} \to v = \emptyset$

$$y_{\{1\}}(x_1) = \int_{\mathbb{R}^2} y(x_1, x_2, x_3) \prod_{i=2}^{3} f_{\{i\}}(x_i)\, d\nu(x_i) - y_{\emptyset} = \mathbb{E}[y(X_1, X_2, X_3) | X_1 = x_1] - y_{\emptyset}.$$

- $u = \{1, 2\} \to v = \{1\}, \{2\}, \emptyset$

Recursive Form Example

- $N = 3, u = \emptyset$

$$y_\emptyset = \int_{\mathbb{R}^3} y(\boldsymbol{x}) \prod_{i=1}^{3} f_{\{i\}}(x_i)\, d\nu(x_i) = \mathbb{E}[y(\boldsymbol{X})].$$

- $u = \{1\} \to v = \emptyset$

$$y_{\{1\}}(x_1) = \int_{\mathbb{R}^2} y(x_1, x_2, x_3) \prod_{i=2}^{3} f_{\{i\}}(x_i)\, d\nu(x_i) - y_\emptyset = \mathbb{E}[y(X_1, X_2, X_3) | X_1 = x_1] - y_\emptyset.$$

- $u = \{1, 2\} \to v = \{1\}, \{2\}, \emptyset$

$$\begin{aligned} y_{\{1,2\}}(x_1, x_2) &= \int_{\mathbb{R}} y(x_1, x_2, x_3) f_{\{3\}}(x_3)\, d\nu(x_3) - y_{\{1\}}(x_1) - y_{\{2\}}(x_2) - y_\emptyset \\ &= \mathbb{E}[y(X_1, X_2, X_3) | X_1 = x_1, X_2 = x_2] - y_{\{1\}}(x_1) - y_{\{2\}}(x_2) - y_\emptyset. \end{aligned}$$

Reminder: Definition of Orthogonal Projection

$$\Pi_{\mathcal{G}} y = \arg\min_{g \in \mathcal{G}} \|y - g\|^2 = \arg\min_{g \in \mathcal{G}} \mathbb{E}[(y(\boldsymbol{X}) - g(\boldsymbol{X}))^2].$$

- $\mathcal{G}$: linear subspace of $\mathcal{L}^2$ we project onto
- g all functions in the subspace

fANOVA as Orthogonal Projection

$$
\begin{aligned}
y_\emptyset &= \mathbb{E}[y(\boldsymbol{X})] \\
&= \arg\min_{a \in \mathbb{R}} \mathbb{E}[(y(\boldsymbol{X}) - a)^2] \\
&= \arg\min_{g_0 \in \mathcal{G}_0} \|y - g_0\|^2 = \Pi_{\mathcal{G}_0} y,
\end{aligned}
$$

$$
\begin{aligned}
y_u(.) &= \mathbb{E}[y(\boldsymbol{X}) \mid X_u = .] - \sum_{v \subsetneq u} y_v(.) \\
&= \arg\min_{g_u \in \mathcal{G}_u} \mathbb{E}[(y(\boldsymbol{X}) - g_u(.))^2] - \sum_{v \subsetneq u} y_v(.) \\
&= (\Pi_{\mathcal{G}_u} y)(.) - \sum_{v \subsetneq u} y_v(.)
\end{aligned}
$$

Example: Independent MVN

$$y(x_1, x_2) = 2x_1 + x_2^2 + x_1x_2$$

Example: Independent MVN

$$y(x_1, x_2) = 2x_1 + x_2^2 + x_1x_2$$

$$(X_1, X_2)^\mathsf{T} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Example: Independent MVN

$$y(x_1, x_2) = 2x_1 + x_2^2 + x_1 x_2$$

$$(X_1, X_2)^\top \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

$$X_1 \mid X_2 = x_2 \sim \mathcal{N}(0, 1), \quad X_2 \mid X_1 = x_1 \sim \mathcal{N}(0, 1).$$

Example: Independent MVN

$$y(x_1, x_2) = 2x_1 + x_2^2 + x_1 x_2$$

$$(X_1, X_2)^\mathsf{T} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

$$X_1 \mid X_2 = x_2 \sim \mathcal{N}(0, 1), \quad X_2 \mid X_1 = x_1 \sim \mathcal{N}(0, 1).$$

Components:

Example: Independent MVN

$$y(x_1, x_2) = 2x_1 + x_2^2 + x_1x_2$$

$$(X_1, X_2)^\mathsf{T} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

$$X_1 \mid X_2 = x_2 \sim \mathcal{N}(0, 1), \quad X_2 \mid X_1 = x_1 \sim \mathcal{N}(0, 1).$$

Components:

$$y_\emptyset = 1, \qquad y_{\{1\}}(x_1) = 2x_1, \qquad y_{\{2\}}(x_2) = x_2^2 - 1, \qquad y_{\{1,2\}}(x_1, x_2) = x_1x_2.$$

Example: Only Linear Terms

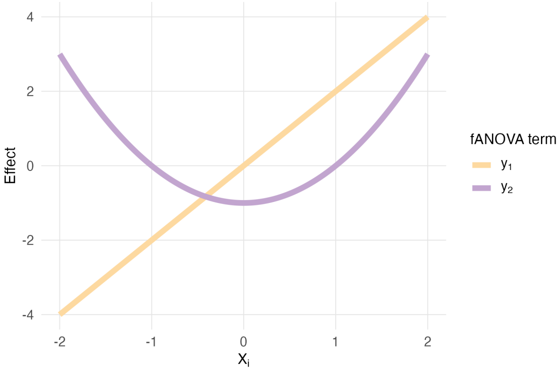


Figure: Main effects

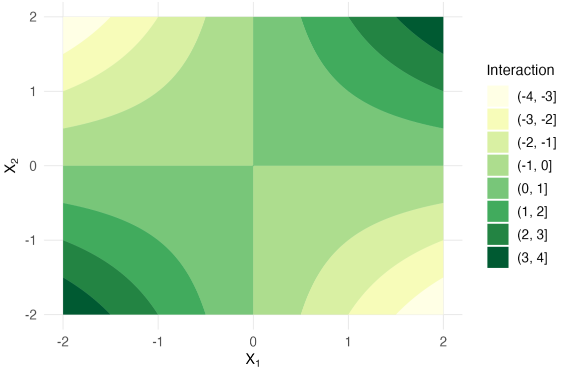


Figure: Interaction

Equality to Hoeffding Decomposition

Hoeffding Decomposition

$$y(\boldsymbol{X}) = \sum_{A \subseteq D} y_A(\boldsymbol{X}_A), \qquad D := \{1, \ldots, N\}, \tag{1}$$

where, for each $A \subseteq D$, the component function y_A is defined by:

$$y_A(\boldsymbol{X}_A) = \sum_{B \subseteq A} (-1)^{|A|-|B|} \, \mathbb{E}[y(\boldsymbol{X}) \mid \boldsymbol{X}_B] \,, \tag{2}$$

where y_u are orthogonal components.

- Classical fANOVA and Hoeffding decomposition yield same components under zero-centered inputs
- Both assume independence of input variables

Hoeffding Decomposition Example

$$y(x_1, x_2) = 2x_1 + x_2^2 + x_1 x_2$$

$$y_\emptyset = \mathbb{E}[y(X_1, X_2)] = 2\,\mathbb{E}[X_1] + \mathbb{E}[X_2^2] + \mathbb{E}[X_1 X_2] = 1,$$

$$\begin{aligned} y_{\{1\}}(x_1) &= \sum_{B \subseteq \{1\}} (-1)^{1-|B|}\,\mathbb{E}[y(\boldsymbol{X}) \mid X_B] = -\,\mathbb{E}[y] + \mathbb{E}[y|X_1 = x_1] \\ &= -1 + (2x_1 + \mathbb{E}[X_2^2] + x_1 \mathbb{E}[X_2]) = 2x_1, \end{aligned}$$

$$\begin{aligned} y_{\{2\}}(x_2) &= \sum_{B \subseteq \{2\}} (-1)^{1-|B|}\,\mathbb{E}[y(\boldsymbol{X}) \mid X_B] - \mathbb{E}[y] + \mathbb{E}[y|X_2 = x_2] \\ &= -1 + (2\mathbb{E}[X_1] + x_2^2 + x_2 \mathbb{E}[X_1]) = x_2^2 - 1. \end{aligned}$$

$$
\begin{aligned}
y_{\{1,2\}}(x_1, x_2) &= \sum_{B \subseteq \{1,2\}} (-1)^{2-|B|}\, \mathbb{E}[y(\boldsymbol{X}) \mid X_B] \\
&= (+1)\, \mathbb{E}[y] - \mathbb{E}[y|X_1 = x_1] - \mathbb{E}[y|X_2 = x_2] + y(x_1, x_2) \\
&= 1 - (2x_1 + 1) - (x_2^2) + (2x_1 + x_2^2 + x_1x_2) \\
&= x_1x_2.
\end{aligned}
$$

$$
y(x_1, x_2) = y_\emptyset + y_{\{1\}}(x_1) + y_{\{2\}}(x_2) + y_{\{1,2\}}(x_1, x_2) = 1 + 2x_1 + (x_2^2 - 1) + x_1x_2
$$

Outline

Example with Dependent Inputs

$$y(x_1, x_2) = 2x_1 + x_2^2 + x_1 x_2, \qquad \rho = 0.8$$

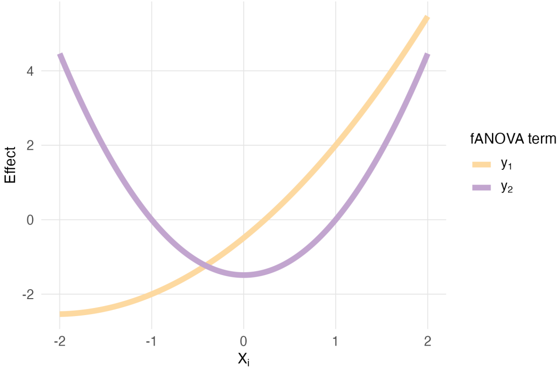


Figure: Main effects

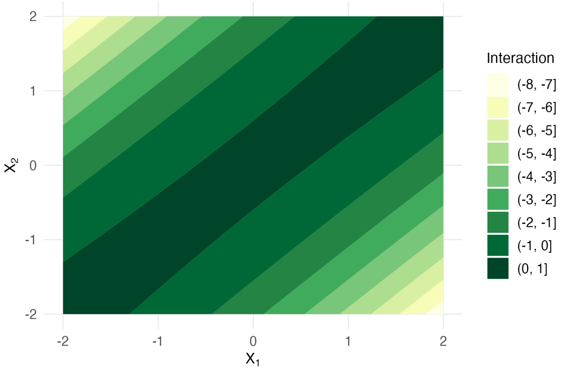


Figure: Interaction

Ensuring Interpretability under Dependent Inputs

Weak Annihilating Conditions

$$\int_{\mathbb{R}^{N-|u|}} y_{u,G}(\boldsymbol{x}_u) f_{\boldsymbol{X}_u}(\boldsymbol{x}_u)\, d\nu(x_i) = 0 \quad \text{for} \quad i \in u \neq \emptyset.$$

Ensuring Interpretability under Dependent Inputs

Weak Annihilating Conditions

$$\int_{\mathbb{R}^{N-|u|}} y_{u,G}(\boldsymbol{x}_u) f_{\boldsymbol{X}_u}(\boldsymbol{x}_u)\, d\nu(x_i) = 0 \quad \text{for} \quad i \in u \neq \emptyset.$$

It follows:

Ensuring Interpretability under Dependent Inputs

Weak Annihilating Conditions

$$\int_{\mathbb{R}^{N-|u|}} y_{u,G}(\boldsymbol{x}_u) f_{\boldsymbol{X}_u}(\boldsymbol{x}_u)\, d\nu(x_i) = 0 \quad \text{for} \quad i \in u \neq \emptyset.$$

It follows:

$$\mathbb{E}[y_{u,G}(\boldsymbol{X}_u)] := \int_{\mathbb{R}^N} y_{u,G}(\boldsymbol{x}_u) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}) = 0.$$

Ensuring Interpretability under Dependent Inputs

Weak Annihilating Conditions

$$\int_{\mathbb{R}^{N-|u|}} y_{u,G}(\boldsymbol{x}_u) f_{\boldsymbol{X}_u}(\boldsymbol{x}_u)\, d\nu(x_i) = 0 \quad \text{for} \quad i \in u \neq \emptyset.$$

It follows:

$$\mathbb{E}[y_{u,G}(\boldsymbol{X}_u)] := \int_{\mathbb{R}^N} y_{u,G}(\boldsymbol{x}_u) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}) = 0.$$

$$\mathbb{E}[y_{u,G}(\boldsymbol{X}_u) y_{v,G}(\boldsymbol{X}_v)] := \int_{\mathbb{R}^N} y_{u,G}(\boldsymbol{x}_u) y_{v,G}(\boldsymbol{x}_v) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}) = 0 \quad (v \subsetneq u)$$

Component Definition (Coupled System)

$$y_{\emptyset,G} = \int_{\mathbb{R}^N} y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \, d\nu(\boldsymbol{x})$$

Component Definition (Coupled System)

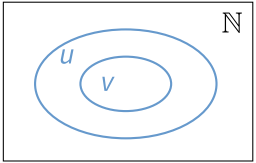
$$y_{\emptyset,G} = \int_{\mathbb{R}^N} y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) \, d\nu(\boldsymbol{x})$$

$$y_{u,G}(\boldsymbol{X}_u) = \int_{\mathbb{R}^{N-|u|}} y(\boldsymbol{X}_u, \boldsymbol{x}_{-u}) f_{-u}(\boldsymbol{x}_{-u}) \, d\nu(\boldsymbol{x}_{-u}) - \sum_{v \subsetneq u} y_{v,G}(\boldsymbol{X}_v)$$
$$- \sum_{\substack{\emptyset \neq v \subseteq \{1,\dots,N\} \\ v \cap u \neq \emptyset, \; v \not\subset u}} \int_{\mathbb{R}^{|v \cap -u|}} y_{v,G}(\boldsymbol{X}_{v \cap u}, \boldsymbol{x}_{v \cap -u}) f_{v \cap -u}(\boldsymbol{x}_{v \cap -u}) \, d\nu(\boldsymbol{x}_{v \cap -u})$$

Component Definition (Coupled System)

$$y_{\emptyset,G} = \int_{\mathbb{R}^N} y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x})$$

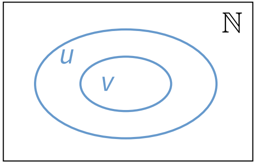
$$y_{u,G}(\boldsymbol{X}_u) = \int_{\mathbb{R}^{N-|u|}} y(\boldsymbol{X}_u, \boldsymbol{x}_{-u}) f_{-u}(\boldsymbol{x}_{-u})\, d\nu(\boldsymbol{x}_{-u}) - \sum_{v \subsetneq u} y_{v,G}(\boldsymbol{X}_v)$$
$$- \sum_{\substack{\emptyset \neq v \subseteq \{1,\dots,N\} \\ v \cap u \neq \emptyset,\ v \not\subset u}} \int_{\mathbb{R}^{|v \cap -u|}} y_{v,G}(\boldsymbol{X}_{v\cap u}, \boldsymbol{x}_{v\cap -u}) f_{v\cap -u}(\boldsymbol{x}_{v\cap -u})\, d\nu(\boldsymbol{x}_{v\cap -u})$$

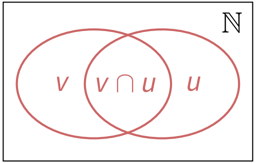


Component Definition (Coupled System)

$$y_{\emptyset,G} = \int_{\mathbb{R}^N} y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x})$$

$$y_{u,G}(\boldsymbol{X}_u) = \int_{\mathbb{R}^{N-|u|}} y(\boldsymbol{X}_u, \boldsymbol{x}_{-u}) f_{-u}(\boldsymbol{x}_{-u})\, d\nu(\boldsymbol{x}_{-u}) - \sum_{v \subsetneq u} y_{v,G}(\boldsymbol{X}_v)$$
$$- \sum_{\substack{\emptyset \neq v \subseteq \{1,\dots,N\} \\ v \cap u \neq \emptyset,\ v \not\subset u}} \int_{\mathbb{R}^{|v \cap -u|}} y_{v,G}(\boldsymbol{X}_{v\cap u}, \boldsymbol{x}_{v\cap -u}) f_{v\cap -u}(\boldsymbol{x}_{v\cap -u})\, d\nu(\boldsymbol{x}_{v\cap -u})$$





- $N = 3, u = \emptyset$

$$y_{\emptyset,G} = \mathbb{E}[y(\boldsymbol{X})]$$

- $u = \{1\} \to v \subsetneq u = \emptyset$ and $(\emptyset \neq v \subseteq \{1,\dots,N\},\ v \cap u \neq \emptyset,\ v \not\subset u) = \{1,2\}, \{1,2,3\}$

$$y_{\{1\},G}(\boldsymbol{X}_u) = \int_{\mathbb{R}^2} y(x_1, x_2, x_3) f_{\{2,3\}}(x_2, x_3)\, d\nu(x_2, x_3) - y_{\emptyset,G}$$

$$- \int_{\mathbb{R}} y_{\{1,2\},G}(x_1, x_2) f_{\{2\}}(x_2)\, d\nu(x_2) - \int_{\mathbb{R}} y_{\{1,3\},G}(x_1, x_3) f_{\{3\}}(x_3)\, d\nu(x_3)$$

$$- \int_{\mathbb{R}^2} y_{\{1,2,3\},G}(x_1, x_2, x_3) f_{\{2,3\}}(x_2, x_3)\, d\nu(x_2, x_3)$$

Construct components via Fourier-polynomial expansion

$$
\begin{aligned}
y(x_1, x_2) &= a_0 + a_1 x_1 + a_2 x_2 + a_{11} x_1^2 + a_{22} x_2^2 + a_{12} x_1 x_2 \\
&= c_0 + c_{1,1}\,\psi_{1,1}(x_1) + c_{2,1}\,\psi_{2,1}(x_2) \\
&\quad + c_{1,2}\,\psi_{1,2}(x_1) + c_{2,2}\,\psi_{2,2}(x_2) + c_{12,11}\,\psi_{12,11}(x_1, x_2) \\
&= \underbrace{c_0}_{y_0} + \underbrace{\left(c_{1,1}\,\psi_{1,1}(x_1) + c_{1,2}\,\psi_{1,2}(x_1)\right)}_{y_1(x_1)} \\
&\quad + \underbrace{\left(c_{2,1}\,\psi_{2,1}(x_2) + c_{2,2}\,\psi_{2,2}(x_2)\right)}_{y_2(x_2)} \\
&\quad + \underbrace{c_{12,11}\,\psi_{12,11}(x_1, x_2)}_{y_{12}(x_1, x_2)}.
\end{aligned}
$$

Basis Functions proposed by Rahman (2014)

In [6] Hermite polynomial basis functions are proposed

$$\psi_{\emptyset}(x_1, x_2) = 1,$$

$$\psi_{1,1}(x_1) = x_1,$$

$$\psi_{2,1}(x_2) = x_2,$$

$$\psi_{1,2}(x_1) = x_1^2 - 1,$$

$$\psi_{2,2}(x_2) = x_2^2 - 1,$$

$$\psi_{12,11}(x_1, x_2) = \frac{\rho(x_1^2 + x_2^2)}{1 + \rho^2} - x_1 x_2 + \frac{\rho(\rho^2 - 1)}{1 + \rho^2},$$

Substituting the basis functions:

$$
\begin{aligned}
y(x_1, x_2) = \underbrace{c_0}_{y_0} &+ \underbrace{\left(c_{1,1}\, x_1 + c_{1,2}\,(x_1^2 - 1)\right)}_{y_1(x_1)} \\
&+ \underbrace{\left(c_{2,1}\, x_2 + c_{2,2}\,(x_2^2 - 1)\right)}_{y_2(x_2)} \\
&+ \underbrace{c_{12,11}\,\left(\frac{\rho(x_1^2 + x_2^2)}{1+\rho^2} - x_1 x_2 + \frac{\rho(\rho^2 - 1)}{1+\rho^2}\right)}_{y_{12}(x_1, x_2)}.
\end{aligned}
$$

Find weights to recover original polynomial while fulfilling zero-mean and hierarchical orthogonality:

$$
y(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2 + a_{11} x_1^2 + a_{22} x_2^2 + a_{12} x_1 x_2
$$

Coefficient Matching

The corresponding weights can be found via coefficient matching. Start from the interaction term:

$$
\begin{aligned}
-c_{12,11} &= a_{12} & \Rightarrow \quad c_{12,11} &= -a_{12} \\
c_{1,2} + c_{12,11}\,\tfrac{\rho}{1+\rho^2} &= a_{11} & \Rightarrow \quad c_{1,2} &= a_{11} + \tfrac{\rho}{1+\rho^2} a_{12} \\
c_{2,2} + c_{12,11}\,\tfrac{\rho}{1+\rho^2} &= a_{22} & \Rightarrow \quad c_{2,2} &= a_{22} + \tfrac{\rho}{1+\rho^2} a_{12} \\
c_{1,1} &= a_1 & & \\
c_{2,1} &= a_2 & & \\
c_0 - c_{1,2} - c_{2,2} + c_{12,11}\,\frac{\rho(\rho^2-1)}{1+\rho^2} &= a_0 & \Rightarrow \quad c_0 &= a_0 + a_{11} + a_{22} + \rho\, a_{12}
\end{aligned}
$$

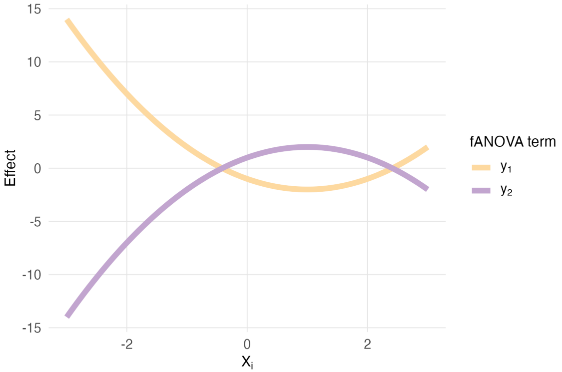
General Form of fANOVA Components

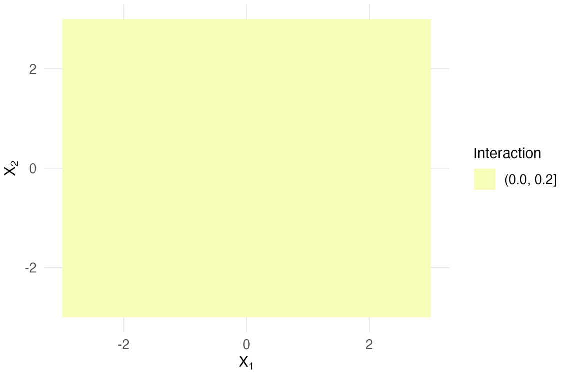
- Yields fANOVA components for MVN Inputs
- Works for polynomials of degree up to $d = 2$

$$
\begin{aligned}
y_{\emptyset,G} &= a_0 + a_{11} + a_{22} + \rho\, a_{12}, \\
y_{\{1\},G}(x_1) &= a_1\, x_1 + \left(a_{11} + \frac{\rho}{1+\rho^2} a_{12}\right) \left(x_1^2 - 1\right), \\
y_{\{2\},G}(x_2) &= a_2\, x_2 + \left(a_{22} + \frac{\rho}{1+\rho^2} a_{12}\right) \left(x_2^2 - 1\right), \\
y_{\{1,2\},G}(x_1, x_2) &= -a_{12} \left(\frac{\rho(x_1^2 + x_2^2)}{1+\rho^2} - x_1 x_2 + \frac{\rho(\rho^2 - 1)}{1+\rho^2} \right).
\end{aligned}
\tag{3}
$$

Example: Only Main

$$y(x_1, x_2) = -2x_1 - 2x_2 + x_1^2 + x_2^2 \qquad \rho = 0$$





Decomposition for different correlations

$$z(x_1, x_2) = 2x_1 + x_1^2 + 0.5x_1x_2$$

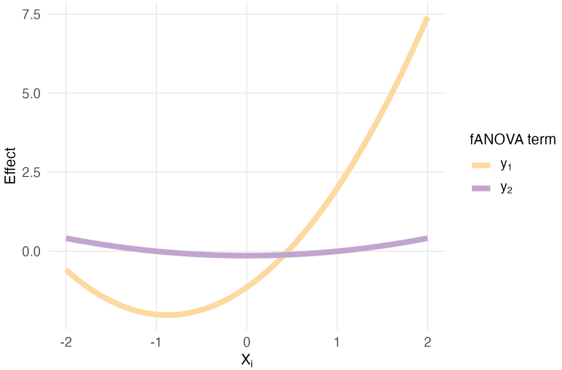


Figure: Main effect for $\rho = 0.3$.

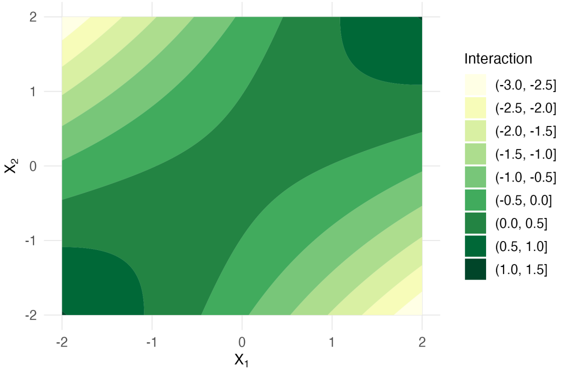


Figure: Interaction effect for $\rho = 0.3$.

Decomposition for different correlations

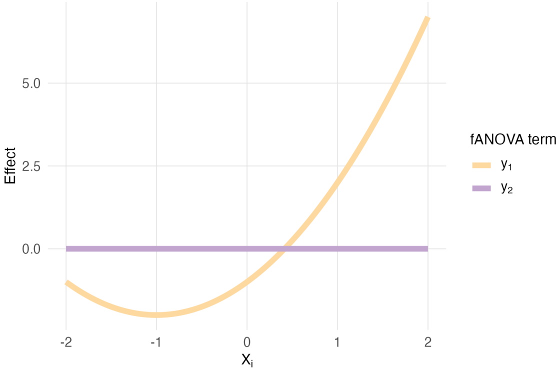


Figure: Main effect for $\rho = 0$.

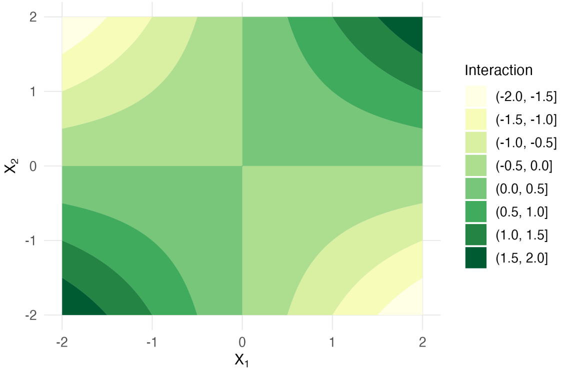


Figure: Interaction effect for $\rho = 0$.

$\Rightarrow$ nonzero main effect of X_2 only present under correlation.

Alternative Generalization of fANOVA

In [2] Hooker originally proposed different formulation of generalized fANOVA components:

$$\{y_{u,G}(\boldsymbol{x}_u) \mid u \subseteq d\} = \arg \min_{\{g_u \in \mathcal{L}^2(\mathbb{R}^{|u|})\}} \int_{\mathbb{R}^N} \left(\sum_{u \subseteq d} g_u(\boldsymbol{x}_u) - y(\boldsymbol{x}) \right)^2 f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}),$$

subject to hierarchical orthogonality conditions:

$$\forall v \subseteq u,\ \forall g_v : \int_{\mathbb{R}^N} y_u(\boldsymbol{x}_u) g_v(\boldsymbol{x}_v) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}) = 0.$$

Variance Decomposition

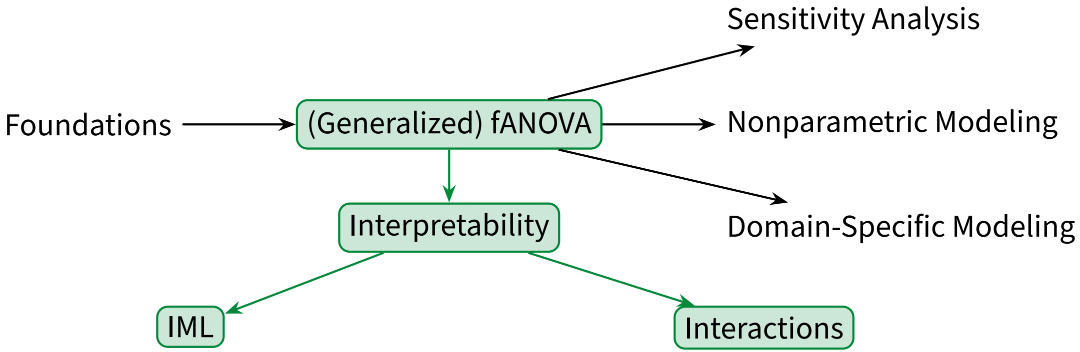
First proposed by [7], Sobol' indices build on variance decomposition:

$$\mu := \mathbb{E}[y(\boldsymbol{X})] = y_{\emptyset,G}$$

$$\begin{aligned}
\sigma^2 &:= \mathbb{E}\left[(y(\boldsymbol{X}) - \mu)^2\right] \\
&= \mathbb{E}\left[\left(y_{\emptyset,G} + \sum_u y_{u,G}(\boldsymbol{X}_u) - y_{\emptyset,G}\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_u y_{u,G}(\boldsymbol{X}_u)\right)^2\right] \\
&= \sum_u \mathbb{E}\left[y_{u,G}^2(\boldsymbol{X}_u)\right] + \sum_{u \not\subseteq v,\, v \not\subseteq u} \mathbb{E}\left[y_{u,G}(\boldsymbol{X}_u) y_{v,G}(\boldsymbol{X}_v)\right],
\end{aligned}$$

Outline

Summary and Future Research



- Purify Interactions [5, 4]
- Model-agnostic tool for effect quantification and visualization [1, 2, 3]
- But mainly theoretical application so far

Outline

Example: Only Linear Terms

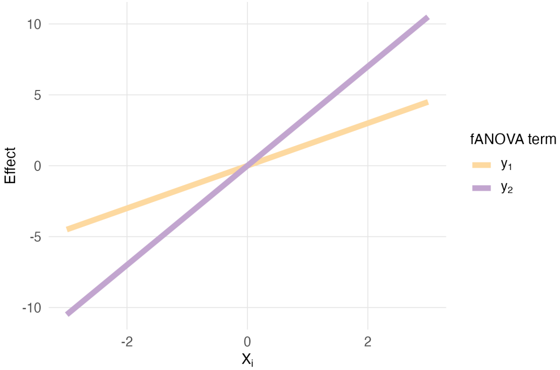


Figure: $q(x_1, x_2) = 1.5x_1 + 3.5x_2$

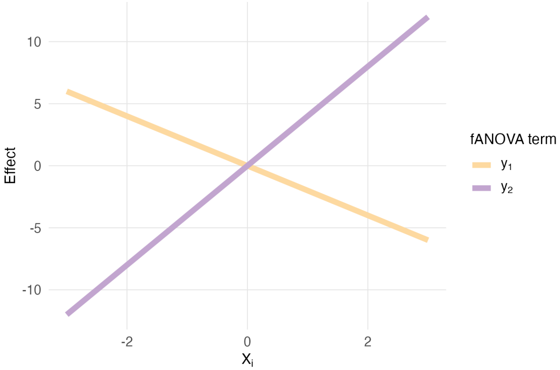
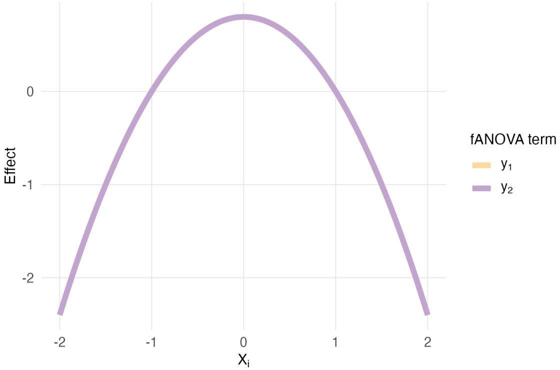
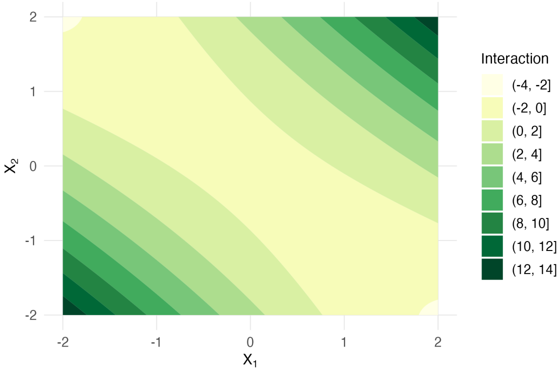


Figure: $q(x_1, x_2) = -2x_1 + 4x_2$

Example: Interaction

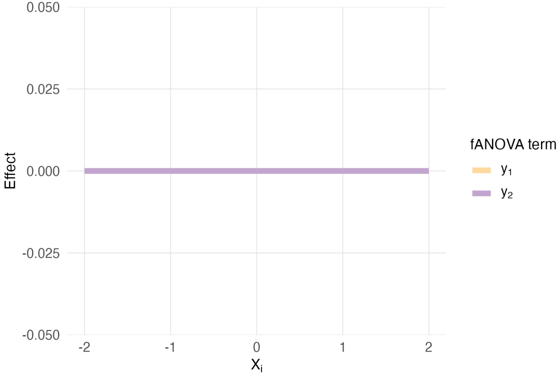
$$y(x_1, x_2) = x_1 x_2 \qquad \rho = -0.5$$

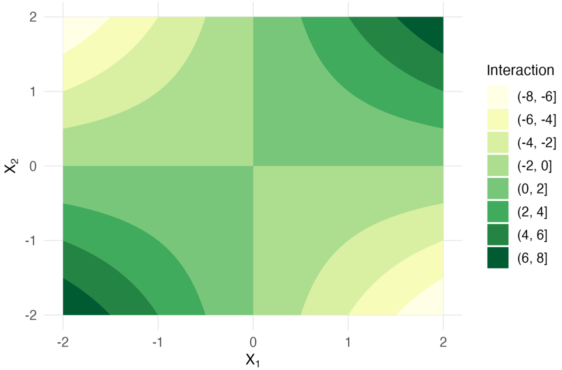




Example: Interaction

$$y(x_1, x_2) = x_1 x_2 \qquad \rho = 0$$





Sobol Indices

Formula for classical Sobol' indices?

Decomposition of linear functions

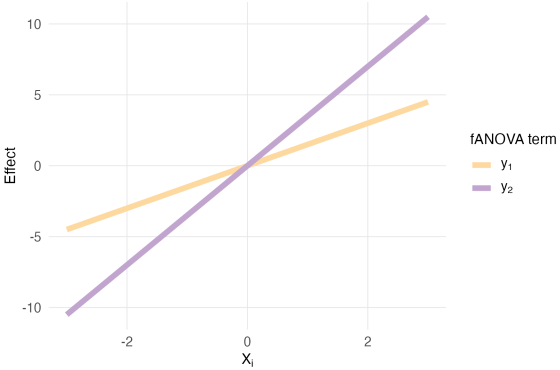


Figure: $q(x_1, x_2) = 1.5x_1 + 3.5x_2$

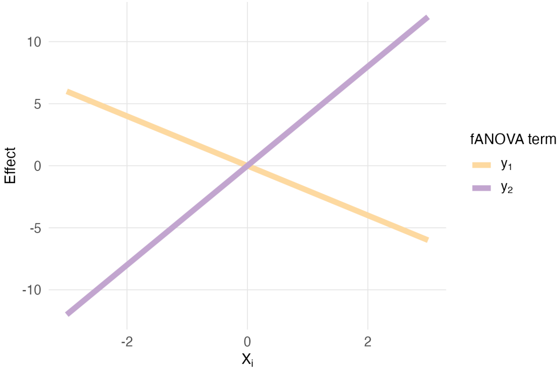


Figure: $q(x_1, x_2) = -2x_1 + 4x_2$

Classical fANOVA Proofs

- Zero mean property: factorized density, Fubinis Theorem, strong annihilating conditions
- Mutual orthogonality: factorized density, Fubinis Theorem, strong annihilating conditions

Generalized fANOVA Proofs

- Zero mean property: separating x into subvectors, marginal density, Fubinis Theorem, weak annihilating conditions
- Hierarchical orthogonality: set the scene, u is a proper subset of v $u \subsetneq v$, so there is an index in u which is not in v; divide x_u into subvectors, marginal density, Fubini and weak annihilating conditions
- Weak annihilating becomes strong under independence: assume the weak ones, product density, factor out
- Three integration cases: distinguish between different relationships u and v, depending on the relationship the integral w.r.t. to marginal density simplifies
- Generalized fANOVA components by Rahman: first build constant term; for nonconstant terms use integration cases
- Integration constraint Hooker: show that hierarchical orthogonality is fulfilled if the conditions hold, show that it is not fulfilled if they do not hold; but why exactly these conditions a bit unclear
- Take a look at Sobols proof again

Relevant External Links

- https://docs.google.com/spreadsheets/d/1K5ECL6hDPDnHwM_k342xa29H-vHWzdk27PTgDHUwfFE/edit?usp=sharing - Table with fANOVA-related literature

References I

Giles Hooker.
Discovering additive structure in black box functions.
In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 575–580, Seattle WA USA, August 2004. ACM.

Giles Hooker.
Generalized Functional ANOVA Diagnostics for High-Dimensional Functions of Dependent Variables.
Journal of Computational and Graphical Statistics, 16(3):709–732, September 2007.

Linwei Hu, Vijayan N. Nair, Agus Sudjianto, Aijun Zhang, Jie Chen, and Zebin Yang.
Interpretable Machine Learning Based on Functional ANOVA Framework: Algorithms and Comparisons.
Applied Stochastic Models in Business and Industry, 41(1):e2916, January 2025.

References II

Gunnar König, Eric Günther, and Ulrike von Luxburg.
Disentangling Interactions and Dependencies in Feature Attribution, October 2024.
arXiv:2410.23772.

Benjamin Lengerich, Sarah Tan, Chun-Hao Chang, Giles Hooker, and Rich Caruana.
Purifying Interaction Effects with the Functional ANOVA: An Efficient Algorithm for Recovering Identifiable Additive Models.
In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pages 2402–2412. PMLR, June 2020.
ISSN: 2640-3498.

Sharif Rahman.
A Generalized ANOVA Dimensional Decomposition for Dependent Probability Measures.
SIAM/ASA Journal on Uncertainty Quantification, 2(1):670–697, 2014.

References III

I. M. Sobol.
Sensitivity Estimates for Nonlinear Mathematical Models.
Mathematical Modelling and Computational Experiments, 1:407–414, 1993.