Bachelor's Thesis

fANOVA for Interpretable Machine Learning

Department of Statistics
Ludwig-Maximilians-Universität München

Juliet Fleischer

Munich, Month Dayth, Year

Submitted in partial fulfillment of the requirements for the degree of B. Sc.
Supervised by Prof. Dr. Thomas Nagler

Abstract

This article studies the functional ANOVA decomposition (fANOVA) in the context of model interpretability. We start with the historical background and providing context to how the method has evolved over time and is used in different domains. This is followed by the formal definition of the method, where we distinguish between the classical fANOVA, which assumes independent inputs, and the generalized fANOVA, which allows for dependency between variables. For both cases, we unite different notations and approaches to the formalization of fANOVA. When generalizing fANOVA to dependent inputs, we encounter the Hoeffding decomposition, which is closely related to fANOVA. Illustrated by examples we clearly distinguish between both methods and experiment with visualizing the decomposition of different functions. We conclude with some words on current estimation approaches and suggestions for future work.

Contents

1 Introduction

Interpretability and sensitivity analysis are increasingly important in machine learning (ML) and in the study of complex systems. One of the foundational mathematical tools supporting these goals is the functional ANOVA decomposition (fANOVA).

At its core, the fANOVA decomposition provides a method which allows decomposing integrable functions into a sum of orthogonal components. It is a foundational method, useful in interpretability of black box models (Hooker (2004), Molnar (2025)), uncertainty quantification of complex systems Rahman (2014), non-parametric statistical modelling (see for example Stone et al. (1997)), sensitivity analysis Sobol (1993)), and many more fields. Given this wide range of applications, fANOVA is an essential concept worth understanding in depth.

However, learning about fANOVA is not straightforward. A problem is the mix of formalizations and definitions around the method, partly due to its long history and the different streams of science that have used it. This already starts with the name of the method. It has been called decomposition into summands of different order (Sobol, 1993), ANOVA representation (Sobol, 1993), functional ANOVA decomposition (Hooker, 2004), ANOVA dimensional decomposition (Rahman, 2014) – in this thesis we will refer to it as the fANOVA decomposition. The diversity does not stop at naming. Different authors formalize the decomposition using different notation, slightly different sets of assumptions, and either interpret fANOVA from a probabilistic perspective using expectations or from a more deterministic mathematical viewpoint using integrals. While these approaches are mathematically equivalent and can be unified under the concept of orthogonal projections, this connection is often not obvious when first encountering the literature.

Given this state of affairs, there is a clear need for a comprehensive overview of fANOVA-related work and for a unification of the various notations and definitions that ultimately express the same concepts. Bringing clarity into the fANOVA landscape is more relevant than ever as the method has recently attracted renewed attention in interpretable machine learning (IML) literature (see for example Hu et al. (2025)). It is being used to build inherently interpretable models, yet the theoretical foundation is often mentioned only briefly or left implicit.

This thesis addresses that gap by providing an accessible and intuitive introduction to the fANOVA decomposition while remaining mathematically rigorous. It can be viewed as a handbook of the fANOVA decomposition that will help researchers and practitioners to understand the mathematical background of this method as well as its more applied aspects.

This work is organized as follows: It starts with historical context and related work. This

is followed by the central part in which we give the formal definition of the classical and generalized fANOVA decomposition. Next we illustrate characteristics of the method based on analytical examples, before briefly outlining current estimation schemes and concluding with a discussion and possible future research directions.

2 Background and Related Work

The related literature on fANOVA can be grouped into several thematic clusters. Each highlights a different angle on why fANOVA has proven useful and points to why a unified presentation is needed.

The underlying principle of the hierarchical, additive decomposition of a function dates back to Hoeffding (1948). In his seminal work on U-statistics, he introduced the Hoeffding decomposition. Though originally framed around estimators, this decomposition laid the groundwork for fANOVA by showing how a symmetric function can be written as a sum of orthogonal components of increasing dimensionality. Independently, Sobol (1993) proved that any square integrable function on the unit hypercube can be decomposed into a sum of orthogonal components. He originally called it "decomposition into summands of different dimension" and later renames it "ANOVA-representation" (Sobol, 2001) now referred to as the fANOVA decomposition. The foundational work on fANOVA shows, that it is rooted in rigorous mathematical theory, and provides a principled way to break down complex multivariate functions into interpretable, orthogonal parts.

A second strand of work explores how fANOVA underlies non-parametric modeling approaches. Takemura (1983) introduced tensor-analysis of ANOVA decompositions, laying the theoretical foundation. Stone (1994) applied fANOVA ideas to polynomial splines and generalized additive models. Gu (2013) extended this into smoothing-spline ANOVA frameworks for flexible regression estimation. fANOVA not only provides a theoretical decomposition—but also serves as a basis for widely-used non-parametric statistical models featuring additive structure and controlled interactions.

Perhaps the most well-known application of fANOVA is in variance-based sensitivity analysis. Sobol's original decomposition led directly to variance partitioning and interpretation via Sobol indices. Work from Owen (2013, 2014) modernized this framework, introducing efficient estimation strategies and generalized indices suited to quasi-Monte Carlo methods. Borgonovo et al. (2022) further advanced the field with mixture-based generalizations of fANOVA for uncertainty quantification.

Classical fANOVA requires independent input variables, which is a strong assumption in many real-world applications. Therefore, a stream of literature is concerned with the generalization of fANOVA to dependent variables. WWhile Hooker (2007) was the first to present a generalized fANOVA framework, many other researchers were inspired by his work to create modifications of this Rahman (2014), Chastaing et al. (2012), Il Idrissi et al. (2025). We see the generalization as central part of the basis of the fANOVA decomposition and therefore will also present it in this thesis.

A vibrant recent cluster of literature studies fANOVA for model interpretability. There

is work of Lengerich et al. (2020), König et al. (2024), Choi et al. (2025) that all enhance interpretability by using fANOVA to identify and disentangle variable interactions. Then there is work done in the explicit context of interpretable ML. Here fANOVA can be used as a model-agnostic tool Hooker (2004), Fumagalli et al. (2025) are relevant or as foundational principle to build inherently interpretable models Hu et al. (2025). TThis is probably the most recent field of fANOVA in which research is actively ongoing.
Finally, there are specific domains of statistics, such as geostatistics, that explicitly build models on fANOVA framework (see Muehlenstaedt et al. (2012) for fANOVA Kriging models) or use it to study functions arising in computational finance Liu and Owen (2006).

3 Formalization of fANOVA

The formal setup is based on Chastaing et al. (2012) and Rahman (2014).
Let $\mathbb{N}, \mathbb{N}_0, \mathbb{R},$ and $\mathbb{R}_0^+$ denote the sets of positive integer (natural), nonnegative integer, real, and nonnegative real numbers, respectively. Throughout this thesis, we represent the k-dimensional Euclidean space by $\mathbb{R}^k$ and the set of all $k \times k$ real-valued matrices by $\mathbb{R}^{k\times k}$.
Let $(\Omega, \mathcal{F}, \nu)$ be a measure space, where Ω is a sample space, $\mathcal{F}$ is a $\sigma-$algebra on Ω and $\nu : \mathcal{F} \to [0, 1]$ is a probability measure. $\mathcal{B}^N$ is the Borel σ-algebra on $\mathbb{R}^N, N \in \mathbb{N}$. $\boldsymbol{X} = (X_1, \ldots, X_N) : (\Omega, \mathcal{F}) \to (\mathbb{R}^N, \mathcal{B}^N)$ denotes a $\mathbb{R}^N$-valued random vector.
We assume that the probability distribution of $\boldsymbol{X}$ is continuous and completely defined by the joint probability density function $f_{\boldsymbol{X}} : \mathbb{R}^N \to \mathbb{R}_0^+$.
Let u denote a subset of $\{1, \ldots, N\}$ with the complementary set $-u := \{1, \ldots, N\}\backslash u$ and cardinality $0 \leq |u| \leq N$. We denote strict inclusion of a subset by $\subsetneq$ and $\subseteq$ allows for equality. $\boldsymbol{X_u} = (X_{i_1}, \ldots, X_{i_{|u|}}), u \neq \emptyset, 1 \leq i_1 < \cdots < i_{|u|} \leq N$ is a sub-vector of $\boldsymbol{X}$ and $\boldsymbol{X}_{-u} = \boldsymbol{X}_{\{1,\ldots,N\}\backslash u}$ is the complementary subvector.
The marginal density function of $\boldsymbol{X_u}$ is $f_u(\boldsymbol{x_u}) := \int_{\mathbb{R}^{N-|u|}} f_{\boldsymbol{X}}(\boldsymbol{x}) d\nu(\boldsymbol{x}_{-u})$ for a given set $\emptyset \neq u \subseteq \{1, \ldots N\}$.
Let $y(\boldsymbol{X}) := y(X_1, \ldots, X_N)$ be a real-valued, measurable transformation on $(\Omega, \mathcal{F})$, which represents a probabilistic model with random variables as inputs. The Hilbert space of square-integrable functions y with respect to the induced generic measure $f_{\boldsymbol{X}}(\boldsymbol{x})d\nu(\boldsymbol{x})$ supported on $\mathbb{R}^N$ is given by:

$$\mathcal{L}^2(\Omega, \mathcal{F}, \nu) = \left\{ y : \Omega \to \mathbb{R} \; s.t. \; \mathbb{E}[y^2(\boldsymbol{X})] < \infty \right\}.$$

The inner product defined as:

$$\langle y, g \rangle = \int_{\mathbb{R}^N} y(\boldsymbol{x}) g(\boldsymbol{x}) \, f_{\boldsymbol{X}} d\nu(\boldsymbol{x}) = \mathbb{E}[y(\boldsymbol{X}) g(\boldsymbol{X})],$$

and the norm, denoted as $\|.\|$, is defined by:

$$\|y\| = \sqrt{\langle y, y \rangle} = \sqrt{\int_{\mathbb{R}^N} y^2(\boldsymbol{x}) \, d\nu(\boldsymbol{x})} = \mathbb{E}[y(\boldsymbol{X})^2], \quad \forall y \in \mathcal{L}^2.$$

We start by defining the fANOVA decomposition in a general form, which is independent of distribution assumptions about the input variables or anything of the sort. The decomposition consists of 2^N terms and its specific form is determined by the assumptions about the input variables and integration measure.

Definition 3.1. *Let y denote a mathematical model with input given by $X_1, \ldots, X_N$. The functional ANOVA (fANOVA) decomposition of y takes the form:*

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

3.1 Classical fANOVA

For his original fANOVA decomposition, Sobol only considered function defined on the unit hypercube, but later work shows that it is no problem to work within the measure space $(\mathbb{R}^N, \mathcal{B}^N, \nu)$. In any case, we assume that the coordinates $X_1, \ldots, X_N$ are independent of each other. Under independence, we work with a product-type probability measure of $\boldsymbol{X}$ given by $f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}) = \prod_{i=1}^{N} f_{X_i}(x_i)\, d\nu(x_i)$, where $f_{X_i} : \mathbb{R} \to \mathbb{R}_0^+$ is the marginal probability density function of X_i defined on $(\Omega_i, \mathcal{F}_i, \nu_i)$ with a bounded or an unbounded support on $\mathbb{R}$.
Given this setup, we formulate a condition, proposed by Rahman (2014), which we would like to hold for the fANOVA terms to be well-defined and interpretable.

Condition 3.1 (Strong annihilating conditions, (Rahman, 2014))**.** *For the classical fANOVA decomposition we require, that all the nonconstant component functions y_u integrate to zero w.r.t. the marginal probability density of each random variable in u, i.e.*

$$\int_{\mathbb{R}} y_u(\boldsymbol{x}_u) f_{X_i}(x_i)\, d\nu(x_i) = 0 \quad \textit{for } i \in u \neq \emptyset. \tag{2}$$

Proposition 3.1. *Given the strong annihilating conditions, the nonconstant fANOVA components are centred around zero. This means for all $\emptyset \neq u \subseteq \{1, \ldots, N\}$ it holds that:*

$$\int_{\mathbb{R}^N} y_u(\boldsymbol{x}_u) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}) := \mathbb{E}[y_u(\boldsymbol{X}_u)] = 0. \tag{3}$$

Proof.

$$\begin{aligned}
\mathbb{E}[y_u(\boldsymbol{X}_u)] &:= \int_{\mathbb{R}^N} y_u(\boldsymbol{x_u}) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^{|u|}} y_u(\boldsymbol{x_u}) f_{\boldsymbol{X}_u}(\boldsymbol{x}_u)\, d\nu(\boldsymbol{x}_u) \\
&= \int_{\mathbb{R}^{|u|}} y_u(\boldsymbol{x_u}) \prod_{j\in u} f_{X_j}(x_j)\, d\nu(\boldsymbol{x}_u) \\
&= \int_{\mathbb{R}^{|u|-1}} \int_{\mathbb{R}} y_u(\boldsymbol{x_u}) f_{X_i}(x_i)\, d\nu(x_i) \prod_{j\in u, j\neq i} f_{X_j}(x_j)\, d\nu(x_{u\setminus\{i\}}) = 0
\end{aligned}$$

□

Proposition 3.2. *Given the strong annihilating conditions, the fANOVA terms are orthogonal to each other. If two sets of indices are not completely equivalent, i.e.* $\emptyset \neq u \subseteq \{1,\ldots,N\}, \emptyset \neq v \subseteq \{1,\ldots,N\}$, *and* $u \neq v$, *then it holds that:*

$$\int_{\mathbb{R}^N} y_u(\boldsymbol{x_u}) y_v(\boldsymbol{x_v}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\nu(\boldsymbol{x}) = \mathbb{E}[y_u(\boldsymbol{X}_u) y_v(\boldsymbol{X}_v)] = 0. \tag{4}$$

Proof. Since $u \neq v$, there exists at least one index contained in exactly one of the sets. Without loss of generality, we pick $i \in u \setminus v$. Then $y_v(\boldsymbol{x}_v)$ is independent of x_i, and by the strong annihilating conditions

$$\int_{\mathbb{R}} y_u(\boldsymbol{x}_u) f_{X_i}(x_i)\, d\nu(x_i) = 0 \quad \text{for all fixed } \boldsymbol{x}_{u\setminus\{i\}}.$$

Hence,

$$\begin{aligned}
\mathbb{E}[y_u(\boldsymbol{X}_u) y_v(\boldsymbol{X}_v)] &= \int_{\mathbb{R}^N} y_u(\boldsymbol{x_u}) y_v(\boldsymbol{x_v}) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^N} y_u(\boldsymbol{x}_u) y_v(\boldsymbol{x}_v) \prod_{j=1}^{N} f_{X_j}(x_j)\, d\nu(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^{N-1}} \left(\int_{\mathbb{R}} y_u(\boldsymbol{x}_u) f_{X_i}(x_i)\, d\nu(x_i) \right) y_v(\boldsymbol{x}_v) \prod_{j\neq i} f_{X_j}(x_j)\, d\nu(\boldsymbol{x}_{-i}) \\
&= 0.
\end{aligned}$$

□

As we have seen, the fANOVA terms are "fully orthogonal" to each other, meaning not only terms of different order are orthogonal to each other but also terms of the same order are. Zero-mean and orthogonality are desirable and important properties because

they ensure that a fANOVA term can be interpretate as isolated effect of the specific variable or isolated effect of an interaction. The term y_1, for example, captures the isolated main effects of X_1; there is no other effect mixed into it, which X_1 might have through interactions with other variables. The term y_{12} on the other hand captures the interaction effect of X_1 and X_2, while the solo effect of X_1 is already captured by y_1 and does not merge into y_{12}. From the lense of interpretability, this distinguishes the fANOVA decomposition from methods such as partial dependence (PD) or Shapley values.

3.1.1 Construction of the fANOVA Terms

The individual fANOVA terms for the variables with indices in u are constructed by integrating the original function $y(\boldsymbol{X})$ w.r.t all variables expect for the ones in u, and subtracting the lower order terms. Intuitively the integral is averaging the original function over all other variables expect the ones of interest, which makes sense as we are then left with a function of the variables of interest only. Subtracting lower order terms corresponds to accounting for effects that are already explained by other variables or interactions so that we obtain the isolated effects.
Since $u = \emptyset$ for the constant term, we integrate w.r.t. all variables:

$$y_\emptyset = \int_{\mathbb{R}^N} y(\boldsymbol{x}) \prod_{i=1}^{N} f_{X_i}(x_i)\, d\nu(x_i) = \mathbb{E}[y(\boldsymbol{X})]. \tag{5}$$

For all other effects $\emptyset \neq u \in \{1, \ldots, N\}$ we can calculate:

$$y_u(\boldsymbol{X}_u) = \int_{\mathbb{R}^{N-|u|}} y(\boldsymbol{X}_u, \boldsymbol{x}_{-u}) \prod_{i=1, i\notin u}^{N} f_{X_i}(x_i)\, d\nu(x_i) - \sum_{v \subsetneq u} y_v(\boldsymbol{X}_v). \tag{6}$$

Notice that this definition relies on a product-type measure rooted in the independence assumption. We will see what changes when we let got of this assumption in the second part of this section.
As suggested earlier, the fANOVA components offer a clear interpretation of the model, decomposing it into main effects, two-way interaction effects, and so on. This is why fANOVA decomposition has received increasing attention in the IML and XAI literature, holding the potential for a global model-agnostic explanation method of black box models.

3.1.2 Running Example: Multivariate Normal Inputs

Throughout this thesis, we will use the following simple function as a running example:

$$h(x_1, x_2) = a + X_1 + 2X_2 + X_1X_2,$$

which contains both main effects and an interaction term.
We assume the input vector

$$\boldsymbol{X} = (X_1, X_2)^T$$

follows a bivariate standard normal distribution

$$\boldsymbol{X} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right).$$

This general formulation includes both independent inputs ($\rho = 0$) and correlated inputs ($\rho \neq 0$).
From properties of the multivariate normal distribution, the marginal distributions are

$$X_1 \sim \mathcal{N}(0, 1), \qquad X_2 \sim \mathcal{N}(0, 1),$$

and the conditional distributions are given by

$$X_1 \mid X_2 = x_2 \sim \mathcal{N}(\rho x_2, 1 - \rho^2), \quad X_2 \mid X_1 = x_1 \sim \mathcal{N}(\rho x_1, 1 - \rho^2).$$

This example will allow us to compute and compare the classical and generalized fANOVA decompositions for different correlation structures.

3.1.3 Case 1: Independent Inputs

The classical fANOVA decomposition we covered so far assumes independence, i.e., $\rho = 0$. Here, X_1 and X_2 are independent and standard normal, so the conditional means vanish, and the classical fANOVA decomposition simplifies considerably. Computing the constant component via expectation gives:

$$\begin{aligned} y_\emptyset &= \mathbb{E}[h(X_1, X_2)] \\ &= \mathbb{E}[a + X_1 + 2X_2 + X_1X_2] \\ &= a + \mathbb{E}[X_1] + 2\mathbb{E}[X_2] + \mathbb{E}[X_1]\mathbb{E}[X_2] = a. \end{aligned}$$

Under zero-mean constraint and independence, the main effects and the interaction effect can be computed as follows:

$$\begin{aligned}
y_1(x_1) &= \mathbb{E}_{X_2}[h(x_1, X_2)] - y_\emptyset \\
&= \mathbb{E}_{X_2}[a + x_1 + 2X_2 + x_1 X_2] - a \\
&= x_1 + 2\mathbb{E}_{X_2}[X_2] + x_1 \mathbb{E}_{X_2}[X_2] = x_1, \\
y_2(x_2) &= \mathbb{E}_{X_1}[h(X_1, x_2)] - y_\emptyset \\
&= \mathbb{E}_{X_1}[a + X_1 + 2x_2 + X_1 x_2] - a \\
&= \mathbb{E}_{X_1}[X_1] + 2x_2 + x_2 \mathbb{E}_{X_1}[X_1] = 2x_2, \\
y_{12}(x_1, x_2) &= \mathbb{E}[h(x_1, x_2)] - y_\emptyset - y_1(x_1) - y_2(x_2) \\
&= a + x_1 + 2x_2 + x_1 x_2 - a - x_1 - 2x_2 = x_1 x_2.
\end{aligned}$$

It comes as no surprise that in this simple case the fANOVA decomposition does not provide any additional insights, as the isolated effects can be directly seen from the function. We show this simple example nevertheless to illustrate at which step which assumption is used. This will make clearer what breaks down when we generalize to dependent variables.

3.1.4 fANOVA as projection

In the following we revisit the fANOVA decomposition from the view of orthogonal projections. For this section the parallel between the (conditional) expected value and orthogonal projections formulated in Van der Vaart (1998) is crucial. Having this perspective on the fANOVA decomposition is useful helps in bridging different notations of the method (e.g. via expected value or via integral) and also supports in understanding the generalization of fANOVA later in this section. First we define generally what an orthogonal projection is, and then we will use the idea in the context of fANOVA.

Definition 3.2. *Let $\mathcal{G} \subset \mathcal{L}^2$ denote a linear subspace. The projection of y onto $\mathcal{G}$ is defined by the function $\Pi_{\mathcal{G}} y$ which minimizes the distance to y in $\mathcal{L}^2$:*

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

When we define the constant term $y_\emptyset$ our goal is to best approximate the original function y by a constant function. In other words, we want to minimize the squared difference between y and a constant function $g_0(x) = a$ over all possible constant functions. The solution is the orthogonal projection of y onto the linear subspace of all constant functions $\mathcal{G}_0 = \{g(x) = a; a \in \mathbb{R}\}$. In a probabilistic context, we want to minimize the expected

squared different between the random variables $y(\boldsymbol{X})$ and a, which turns out to be equivalent to the expected value of the random variable (Van der Vaart, 1998). So intuitively, in the absence of any additional information, the expected value is our best approximation of y. More formally we can write:

$$\begin{aligned}\Pi_{\mathcal{G}_0} y &= \arg\min_{g_0 \in \mathcal{G}_0} \|y - g_0\|^2 \\ &= \arg\min_{a_0 \in \mathbb{R}} \mathbb{E}[(y(\boldsymbol{X}) - a)^2] \\ &= \mathbb{E}[y(\boldsymbol{X})] = y_\emptyset\end{aligned}$$

The main effect $y_i(x_i)$ is the projection of y onto the subspace of all functions that only depend on x_i, i.e. $\mathcal{G}_i = \{g(x) = g_i(x_i)\}$. There is no need for additional constraints since subtracting lower order terms ensures that orthogonality and zero mean are fulfilled. The conditional expected value of $\mathbb{E}[y(\boldsymbol{X}) \mid X_i = x_i]$ is the solution to the minimization problem (Van der Vaart, 1998), and the conditional expected value is also a way to express the fANOVA terms (Muehlenstaedt et al., 2012):

$$\begin{aligned}(\Pi_{\mathcal{G}_i} y)(.) - y_\emptyset &= \arg\min_{g_i \in \mathcal{G}_i} \|y - g_i\|^2 - y_\emptyset \\ &= \arg\min_{g_i \in \mathcal{G}_i} \mathbb{E}[(y(\boldsymbol{X}) - g_i(X_i))^2] - y_\emptyset \\ &= \mathbb{E}[y(\boldsymbol{X}) \mid X_i = .] - y_\emptyset = y_i(.)\end{aligned}$$

The two-way interaction effect $y_{ij}(.,.)$ is the projection of y onto the subspace of all functions that depend on x_i and x_j. i.e. $\mathcal{G}_{i,j} = \{g(x) = g_{ij}(x_i, x_j)\}$. Again, we account for lower-order effects by subtracting the constant term and all main effects:

$$\begin{aligned}(\Pi_{\mathcal{G}_{ij}} y)(.;.) - (y_\emptyset + y_i(.) + y_j(.)) &= \arg\min_{g_{ij} \in \mathcal{G}_{ij}} \|y - g_{i,j}\|^2 - (y_\emptyset + y_i(.) + y_j(.)) \\ &= \arg\min_{g_{ij} \in \mathcal{G}_{ij}} \mathbb{E}[(y(\boldsymbol{X}) - g(.,.))^2] - (y_\emptyset + y_i(.) + y_j(.)) \\ &= \mathbb{E}[y(\boldsymbol{X}) | X_j = x_j, X_i = x_i] - (y_\emptyset + y_i(.) + y_j(.)) = y_{ij}(.;.)\end{aligned}$$

In general, we can write for a subset of indices $u \subseteq \{1, \ldots, N\}$ and the subspace $\mathcal{G}_u =$

$\{g(\boldsymbol{x}) = g_u(\boldsymbol{x}_u)\}$:

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

which means that we project y onto the subspace spanned by the own terms of the fANOVA component to be defined, while accounting for all lower-order terms.
On this note, we want to highlight that instead of subtracting the lower order terms from the projection, it is just as valid to first subtract lower order terms and project y on what is left. We can find both formulations in the literature. For example, Muehlenstaedt et al. (2012) subtracts from the projection and defines:

$$\begin{aligned}y_u(\boldsymbol{x}_u) &:= \mathbb{E}[y(\boldsymbol{X}) | \boldsymbol{X}_u = \boldsymbol{x}_u] - \sum_{v \subsetneq u} y_v(\boldsymbol{x}) \\ &= \int_{-u} y(\boldsymbol{x}) d\nu(\boldsymbol{x}_{-u}) - \sum_{v \subsetneq u} y_v(\boldsymbol{x}).\end{aligned}$$

Hooker (2004) takes the alternative view and defines the fANOVA components via the integral, which can be rewritten as the expected value:

$$\begin{aligned}y_u(\boldsymbol{x}_u) &:= \int_{-u} (y(\boldsymbol{x}) - \sum_{v \subsetneq u} y_v(\boldsymbol{x})) d\nu(\boldsymbol{x}_{-u}) \\ &= \mathbb{E}[y(\boldsymbol{X}) - \sum_{v \subsetneq u} y_v(\boldsymbol{x}) | \boldsymbol{X}_u = \boldsymbol{x}_u].\end{aligned}$$

The first equivalence in each formulation is simply the definition in each original paper, while the second equivalence holds under the assumption of independent inputs.

3.1.5 Second-moment statistics

No handbook on fANOVA is complete without at least mentioning *Sobol indices*. This requires us to observe the second moment statistics of the decomposition. We already established that:

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

We can also compute the variance of $y(\boldsymbol{X})$ via the fANOVA decomposition. We write the sum over u for the sum over $\emptyset \neq u \subseteq \{1, \ldots, N\}$ and the sum over $u \neq v$ for the sum over

$\emptyset \neq u \subseteq \{1, \ldots, N\}, \emptyset \neq v \subseteq \{1, \ldots, N\}, u \neq v$. The variance of y is then given by:

$$\begin{aligned}
\sigma^2 := \mathbb{E}[(y(\boldsymbol{X}) - \mu])^2] &= \mathbb{E}[(y_\emptyset + \sum_u y_u(\boldsymbol{X}_u) - y_\emptyset)^2] \\
&= \mathbb{E}[(\sum_u y_u(\boldsymbol{X}_u))^2] \\
&= \mathbb{E}[\sum_u y_u^2(\boldsymbol{X}_u)] + 2\mathbb{E}[\sum_{u \neq v} y_u(\boldsymbol{X}_u) y_v(\boldsymbol{X}_v)] \\
&= \sum_u \mathbb{E}[y_u^2(\boldsymbol{X}_u)]
\end{aligned}$$

We can verify that the variance decomposition holds for our example:

$$\begin{aligned}
Var(a + X_1 + 2X_2 + X_1X_2) &= Var(X_1) + 4Var(X_2) + Var(X_1X_2) + 2\mathcal{C}\wr\sqsubseteq(X_1, X_2) \\
&= 1 + 4 \cdot 1 + 1 \cdot 1 + 2 \cdot 0 = 6 \\
&= \mathbb{E}[X_1^2] + 4\mathbb{E}[X_2^2] + \mathbb{E}[X_1^2]\mathbb{E}[X_2^2] + 2\mathcal{C}\wr\sqsubseteq(X_1, X_2) \\
&= \mathbb{E}[X_1^2] + 4\mathbb{E}[X_2^2] + \mathbb{E}[X_1^2 X_2^2] \\
&= \mathbb{E}[y_1^2(X_1)] + \mathbb{E}[y_2^2(X_2)] + \mathbb{E}[y_{12}^2(X_1, X_2)]
\end{aligned}$$

Studying the variance of the decomposition was the main focus in early works on this method (see e.g. Sobol (1993)). From the variance decomposition Sobol (1993) construct the *Sobol indices*, which are well-known in sensitivity analysis. As it is only one application of the fANOVA decomposition, we will not go into depth here, but we should keep in mind that the presentation of fANOVA is closely linked to the Sobol indices in many works.

3.1.6 Motivating Example

For the classical fANOVA we make the assumption of independent inputs, which is often violated in practice. In the remainder of this section, we therefore investigate what happens, when we allow for dependency between variables. First, let us test with our running example:

$$h(x_1, x_2) = a + X_1 + 2X_2 + X_1X_2,$$

3.1.7 Case 2: Dependent Inputs

Now $\rho \neq 0$, while keeping everything else the same. When we follow the exact same logic as above we obtain the following terms:

$$\begin{aligned}
\tilde{y}_\emptyset &= \mathbb{E}[g(X_1, X_2)] = a + \mathbb{E}[X_1] + 2\mathbb{E}[X_2] + \mathbb{E}[X_1X_2] \\
&= a + \mathbb{E}[X_1X_2] = a + (Cov(X_1, X_2) + \mathbb{E}[X_1]\mathbb{E}[X_2]) \\
&= a + \rho \\
\tilde{y}_1(x_1) &= \mathbb{E}[g(X_1, X_2)|X_1 = x_1] - \tilde{y}_\emptyset \\
&= \mathbb{E}[a + X_1 + 2X_2 + X_1X_2|X_1 = x_1] - (a + \rho) \\
&= a + x_1 + 2\mathbb{E}[X_2|X_1 = x_1] + x_1\mathbb{E}[X_2|X_1 = x_1] - a - \rho \\
&= x_1 + \rho(2x_1 + x_1^2 - 1) \\
\tilde{y}_2(x_2) &= \mathbb{E}[g(X_1, X_2) \mid X_2 = x_2] - \tilde{y}_\emptyset \\
&= \mathbb{E}[a + X_1 + 2X_2 + X_1X_2 \mid X_2 = x_2] - (a + \rho) \\
&= a + 2x_2 + x_2\mathbb{E}[X_1 \mid X_2 = x_2] - a - \rho \\
&= 2x_2 + \rho(x_2 + x_2^2 - 1) \\
\tilde{y}_{12}(x_1, x_2) &= g(x_1, x_2) - \tilde{y}_\emptyset - \tilde{y}_1(x_1) - \tilde{y}_2(x_2) \\
&= a + x_1 + 2x_2 + x_1x_2 - (a + \rho) - (x_1 + \rho(2x_1 + x_1^2 - 1)) - (2x_2 + \rho(x_2 + x_2^2 - 1)) \\
&= x_1x_2 - 2\rho x_1 - \rho x_2 - \rho x_1^2 - \rho x_2^2 + \rho
\end{aligned}$$

The fANOVA components are characterized by two central properties zero mean and orthogonality which follow from Equation 2. When we check if the components $\tilde{y}_0, \tilde{y}_1, \tilde{y}_2, \tilde{y}_{12}$ satisfy these properties, we find out that all components are zero-centred, but not all are orthogonal to each other. We can, for example, immediately see that checking orthogonality between $\tilde{y}_1, \tilde{y}_{1,2}$ will yield the expectation over the constant term ρ exactly once, meaning even if all the other expectations cancel out, this constant will remain and the

entire expression will be unequal to zero:

$$\begin{aligned}\mathbb{E}(\tilde{y}_1(X_1)\tilde{y}_{1,2}(X_1,X_2)) &= \mathbb{E}[(X_1 + 2\rho X_1 + \rho X_1^2 - \rho) \\ &\quad \cdot (X_1X_2 - 2\rho X_1 - \rho X_2 - \rho X_1^2 - \rho X_2^2 + \rho)] \\ &= \mathbb{E}[X_1^2X_2] \ldots - \mathbb{E}[\rho^2] \neq 0.\end{aligned}$$

It turns out that naively computing the "fANOVA decomposition" under dependent inputs, results in components that lack orthogonality, which is a crucial property for interpretability. What we performed in this example is not the fANOVA decomposition for dependent variables but the Hoeffding decomposition (Hoeffding, 1948). This shows the need for a more involved approach for a generalization of this method.

3.2 Generalized fANOVA

Stone (1994) inspired the pioneering work of Hooker (2007) who offers a first solution to the problem of dependent inputs. Work by Chastaing et al. (2012) and Rahman (2014) build on his framework with modifications and extensions.
The generalized fANOVA decomposition still follows the overarching form in Equation 1. However, we no longer work with a product-type probability measure but now $f_{\boldsymbol{X}} : \mathbb{R} \to \mathbb{R}_0^+$ denotes an arbitrary probability density function and $f_{\boldsymbol{X}_u} : \mathbb{R}^u \to \mathbb{R}_0^+$ the marginal probability density function of the variables with indices in $u \subseteq d$.
Instead of enforcing the strong annihilating conditions for desirable properties of the components, Rahman (2014) proposed to formulate a milder version. The milder version fulfils the same function as the strong annihilating conditions in the classical case but works with the joint density of the variables of interest, instead of the individual marginal probability density functions.

Condition 3.2 (Weak annihilating conditions)**.** *For the generalized fANOVA decomposition we require, that all the non-constant fANOVA terms of variables in u integrate to zero w.r.t. the joint probability density function of variables in u:*

$$\int_{\mathbb{R}} y_{u,G}(\boldsymbol{x}_u) f_{\boldsymbol{X}_u}(\boldsymbol{x}_u) d\nu(x_i) = 0 \quad \textit{for} \quad i \in u \neq \emptyset \tag{7}$$

If components are defined under the weak annihilating conditions, we can ensure that they have zero mean and satisfy a milder from of orthogonality - hierarchical orthogonality, which means that components of different order are orthogonal to each other while components of the same order are not. Hierarchical orthogonality is the best we can do when independence cannot be assumed.

Proposition 3.3. *Given the weak annihilating conditions, the generalized fANOVA components $y_{u,G}$, with $\emptyset \neq u \subseteq \{1, \ldots, N\}$, are centred around zero:*

$$\mathbb{E}[y_{u,G}(\boldsymbol{X}_u)] := \int_{\mathbb{R}^N} y_{u,G}(\boldsymbol{x}_u) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}) = 0. \tag{8}$$

Proof. For any subset $\emptyset \neq u \subseteq \{1, \ldots, N\}$, let $i \in u$. We assume that the weak annihilating conditions hold. Then

$$\begin{aligned}
\mathbb{E}[y_{u,G}(\mathbf{X}_u)] &:= \int_{\mathbb{R}^N} y_{u,G}(\mathbf{x}_u) f_{\mathbf{X}}(\mathbf{x})\, d\nu(\mathbf{x}) \\
&= \int_{\mathbb{R}^{|u|}} y_{u,G}(\mathbf{x}_u) \left(\int_{\mathbb{R}^{N-|u|}} f_{\mathbf{X}}(\mathbf{x})\, d\nu(\mathbf{x}_{-u}) \right) d\nu(\mathbf{x}_u) \\
&= \int_{\mathbb{R}^{|u|}} y_{u,G}(\mathbf{x}_u) f_u(\mathbf{x}_u)\, d\nu(\mathbf{x}_u) \\
&= \int_{\mathbb{R}^{|u|-1}} \left(\int_{\mathbb{R}} y_{u,G}(\mathbf{x}_u) f_u(\mathbf{x}_u)\, d\nu(\mathbf{x}_i) \right) \prod_{j \in u,\, j \neq i} d\nu(\mathbf{x}_j) \\
&= 0,
\end{aligned}$$

where we make use of Fubini's theorem and the last line follows from using the weak annihilating condition □

Proposition 3.4. *Given the weak annihilating conditions, the fANOVA components are hierarchically orthogonal. This means that for two components $y_{u,G}$ and $y_{v,G}$ with $u \subsetneq v, \emptyset \neq u \subseteq \{1, \ldots, N\}, \emptyset \neq v \subseteq \{1, \ldots, N\}$ it holds that:*

$$\mathbb{E}[y_{u,G}(\boldsymbol{X}_u) y_{v,G}(\boldsymbol{X}_v)] := \int_{\mathbb{R}^N} y_{u,G}(\boldsymbol{x}_u) y_{v,G}(\boldsymbol{x}_v) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}) = 0. \tag{9}$$

Proof. For any two subsets $\emptyset \neq u \subseteq \{1, \ldots, N\}$ and $\emptyset \neq v \subseteq \{1, \ldots, N\}$, where $v \subsetneq u$,

the subset $u = v \cup (u \setminus v)$. Let $i \in (u \setminus v) \subseteq u$. Then

$$\begin{aligned}
\mathbb{E}[y_{u,G}(\mathbf{X}_u)\, y_{v,G}(\mathbf{X}_v)] &:= \int_{\mathbb{R}^N} y_{u,G}(\mathbf{x}_u) y_{v,G}(\mathbf{x}_v) f_{\mathbf{X}}(\mathbf{x})\, d\nu(\mathbf{x}) \\
&= \int_{\mathbb{R}^{|u|}} y_{u,G}(\mathbf{x}_u) y_{v,G}(\mathbf{x}_v) \left(\int_{\mathbb{R}^{N-|u|}} f_{\mathbf{X}}(\mathbf{x})\, d\nu(\mathbf{x}_{-u}) \right) d\nu(\mathbf{x}_u) \\
&= \int_{\mathbb{R}^{|u|}} y_{u,G}(\mathbf{x}_u) y_{v,G}(\mathbf{x}_v) f_u(\mathbf{x}_u)\, d\nu(\mathbf{x}_u) \\
&= \int_{\mathbb{R}^{|v|}} y_{v,G}(\mathbf{x}_v) \int_{\mathbb{R}^{|u\setminus v|}} y_{u,G}(\mathbf{x}_u) f_u(\mathbf{x}_u)\, d\nu(\mathbf{x}_{u\setminus v})\, d\nu(\mathbf{x}_v) \\
&= \int_{\mathbb{R}^{|v|}} y_{v,G}(\mathbf{x}_v) \int_{\mathbb{R}^{|u\setminus v|-1}} \left(\int_{\mathbb{R}} y_{u,G}(\mathbf{x}_u) f_u(\mathbf{x}_u)\, d\nu(\mathbf{x}_i) \right) \prod_{\substack{j \in (u\setminus v) \\ j \neq i}} d\nu(\mathbf{x}_j)\, d\nu(\mathbf{x}_v) \\
&= 0.
\end{aligned}$$

Repeatedly using Fubini's theorem and the weak annihilating conditions the equality to zero follows. □

A key contribution from Hooker (2007) and Rahman (2014) is that they construct a generalization of the fANOVA decomposition method as a whole, not only parts, such as the Sobol indices. This means it is important that Rahman's generalized statements are coherent with the classical fANOVA decomposition.

Proposition 3.5. *The weak annihilating conditions become the strong annihilating conditions under independence assumption.*

Proof. Assume that the random variables $\{X_j\}_{j\in u}$ are independent. Then we can factorize the marginal density $f_u(\boldsymbol{x}_u)$ as

$$f_u(\boldsymbol{x}_u) = \prod_{j\in u} f_{X_j}(x_j).$$

Now consider the weak annihilating condition (4.2) for some $i \in u \neq \emptyset$:

$$\int_{\mathbb{R}} y_{u,G}(\boldsymbol{x}_u) f_u(\boldsymbol{x}_u)\, d\nu(\boldsymbol{x}_i) = 0.$$

Since we assume independence, we can substitute the joint marginal density with the product of the marginal densities:

$$\int_{\mathbb{R}} y_{u,G}(\boldsymbol{x}_u) \left(\prod_{j\in u} f_{X_j}(x_j) \right) d\nu(\boldsymbol{x}_i) = 0.$$

For fixed x_j with $j \neq i$, the terms $f_{X_j}(x_j)$ are constant with respect to x_i, and can therefore be pulled out of the integral:

$$\left(\prod_{j \in u,\, j \neq i} f_{X_j}(x_j) \right) \int_{\mathbb{R}} y_{u,G}(\boldsymbol{x}_u) f_{X_i}(x_i)\, d\nu(x_i) = 0.$$

As product of probability density functions the prefactor is strictly positive for all x_j with $j \neq i$. Therefore, the integral must be zero for the equality to hold:

$$\int_{\mathbb{R}} y_{u,G}(\boldsymbol{x}_u) f_{X_i}(x_i)\, d\nu(x_i) = 0,$$

which are the strong annihilating conditions Equation 2. □

3.2.1 Construction of the Generalized fANOVA Terms

Recall the construction of the classical fANOVA components Equation 6. The equation tells us that the non-constant classical fANOVA components are defined via the integral of the original function w.r.t. to the product-type probability density function, minus effects by other terms. So ideally for a well-aligned generalization, we would want that the general fANOVA terms can be understood in a similar way, as the integral of y w.r.t. *a smartly chosen* probability density function, minus effects explained by other terms. This is exactly what Rahman (2014) accomplishes. To understand this, we first need to distinguish three cases of integration that will occure in the construction of the generalized components.

Proposition 3.6. *Consider the generalized fANOVA components $y_{v,G}$, $\emptyset \neq v \subseteq \{1, \ldots, N\}$, of a square-integrable function $y : \mathbb{R}^N \to \mathbb{R}$. When integrated w.r.t. the probability measure $f_{-u}(\boldsymbol{x}_{-u})\, d\boldsymbol{x}_{-u}$, $u \subseteq \{1, \ldots, N\}$, one can distinguish three cases:*

$$\int_{\mathbb{R}^{N-|u|}} y_{v,G}(\mathbf{x}_v) f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u} = \begin{cases} \displaystyle\int_{\mathbb{R}^{|v \cap -u|}} y_{v,G}(\mathbf{x}_v) f_{v \cap -u}(\mathbf{x}_{v \cap -u})\, d\mathbf{x}_{v \cap -u}, & \text{if } v \cap u \neq \emptyset \text{ and } v \not\subset u, \\ y_{v,G}(\mathbf{x}_v), & \text{if } v \cap u \neq \emptyset \text{ and } v \subseteq u, \\ 0, & \text{if } v \cap u = \emptyset. \end{cases}$$

Proof. Let $u \subseteq \{1, \ldots, N\}$ and $\emptyset \neq v \subseteq \{1, \ldots, N\}$. We distinguish between three types of relationship between v and u.
Before analysing the first case, note that for any such u and v, it is possible to write

$$(v \cap -u) \subseteq -u \quad \text{and} \quad -u = (-u \setminus (v \cap -u)) \cup (v \cap -u),$$

which will be used in the integral decomposition below.

Case 1: $v \cap u \neq \emptyset$ **and** $v \not\subset u$**.** We use the decomposition of $-u$ stated above to decompose the integration over $\mathbf{x}_{-u}$ as:

$$\int_{\mathbb{R}^{N-|u|}} y_{v,G}(\mathbf{x}_v) f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u} = \int_{\mathbb{R}^{|v\cap -u|}} y_{v,G}(\mathbf{x}_v) \left(\int_{\mathbb{R}^{N-|u|-|v\cap -u|}} f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u\setminus(v\cap -u)} \right) d\mathbf{x}_{v\cap -u}.$$

The inner integral gives the marginal density $f_{v\cap -u}(\mathbf{x}_{v\cap -u})$, so we obtain:

$$= \int_{\mathbb{R}^{|v\cap -u|}} y_{v,G}(\mathbf{x}_v) f_{v\cap -u}(\mathbf{x}_{v\cap -u})\, d\mathbf{x}_{v\cap -u}.$$

Case 2: $v \cap u \neq \emptyset$ **and** $v \subseteq u$**.** Since the sets v and $-u$ are then completely disjoint, $y_{v,G}(\mathbf{x}_v)$ is independent of $\mathbf{x}_{-u}$ and can be pulled out of the integral:

$$\int_{\mathbb{R}^{N-|u|}} y_{v,G}(\mathbf{x}_v) f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u} = y_{v,G}(\mathbf{x}_v) \int_{\mathbb{R}^{N-|u|}} f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u} = y_{v,G}(\mathbf{x}_v),$$

which works because f_{-u} is a probability density function.

Case 3: $v \cap u = \emptyset$**.** In this case, we have $v \subseteq -u$, so $v \cap -u = v$. Then we can write:

$$\begin{aligned}
\int_{\mathbb{R}^{N-|u|}} y_{v,G}(\mathbf{x}_v) f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u} &= \int_{\mathbb{R}^{|v|}} y_{v,G}(\mathbf{x}_v) \left(\int_{\mathbb{R}^{N-|u|-|v|}} f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u\setminus v} \right) d\mathbf{x}_v \\
&= \int_{\mathbb{R}^{|v|}} y_{v,G}(\mathbf{x}_v) f_v(\mathbf{x}_v)\, d\mathbf{x}_v \\
&= \int_{\mathbb{R}^{|v|-1}} \left(\int_{\mathbb{R}} y_{v,G}(\mathbf{x}_v) f_v(\mathbf{x}_v)\, dx_i \right) \prod_{\substack{j\in v \\ j\neq i}} dx_j \\
&= 0,
\end{aligned}$$

while we again split the interval in such a way that we recognize the marginal density f_v and make use of the zero mean property from the strong annihilating conditions.

□

As we will see in the following, we will encounter all of these three integration cases in the definition of the generalized fANOVA components via Rahman (2014) principle. In 3.6 we also already see that the smartly chosen probability density function is $f_{-u}(\boldsymbol{x}_{-u})$.

Theorem 3.1. *The generalized fANOVA component functions $y_{u,G}(\mathbf{x}_u)$ can be recursively*

defined via the following set of equations:

$$y_{\emptyset,G} = \int_{\mathbb{R}^N} y(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}, \tag{4.5a}$$

$$\begin{aligned} y_{u,G}(\mathbf{X}_u) = &\int_{\mathbb{R}^{N-|u|}} y(\mathbf{X}_u, \mathbf{x}_{-u}) f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u} - \sum_{v \subset u} y_{v,G}(\mathbf{X}_v) \\ &- \sum_{\substack{\emptyset \neq v \subseteq \{1,\ldots,N\} \\ v \cap u \neq \emptyset,\ v \not\subset u}} \int_{\mathbb{R}^{|v \cap -u|}} y_{v,G}(\mathbf{X}_{v \cap u}, \mathbf{x}_{v \cap -u}) f_{v \cap -u}(\mathbf{x}_{v \cap -u})\, d\mathbf{x}_{v \cap -u}. \end{aligned} \tag{4.5b}$$

Proof. We begin by integrating both sides of the generalized fANOVA decomposition

$$y(\mathbf{x}) = \sum_{v \subseteq \{1,\ldots,N\}} y_{v,G}(\mathbf{x}_v)$$

w.r.t. $f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u}$, replacing $\boldsymbol{X}$ by $\boldsymbol{x}$, and changing the dummy index from u to v. This yields:

$$\int_{\mathbb{R}^{N-|u|}} y(\mathbf{x}) f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u} = \sum_{v \subseteq \{1,\ldots,N\}} \int_{\mathbb{R}^{N-|u|}} y_{v,G}(\mathbf{x}_v) f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u}.$$

Case $u = \emptyset$: computing the constant term. We set $u = \emptyset$, so $-u = \{1, \ldots, N\}$ and $f_{-u}(\boldsymbol{x}_{-u}) = f_{\boldsymbol{X}}(\boldsymbol{x})$. The above integral can then be written as:

$$\int_{\mathbb{R}^N} y(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = \sum_{v \subseteq \{1,\ldots,N\}} \int_{\mathbb{R}^N} y_{v,G}(\mathbf{x}_v) f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

$$= \int_{\mathbb{R}^N} y_{\emptyset,G}\, f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} + \sum_{\emptyset \neq v \subseteq \{1,\ldots,N\}} \int_{\mathbb{R}^N} y_{v,G}(\mathbf{x}_v) f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

$$= y_{\emptyset,G} + \sum_{\emptyset \neq v \subseteq \{1,\ldots,N\}} \mathbb{E}[y_{v,G}(\mathbf{X}_v)] = y_{\emptyset,G},$$

where the last sum vanishes under the weak annihilating condition.

Case $\emptyset \neq u \subseteq \{1, \ldots, N\}$: computing nonconstant terms. We return to the integrated decomposition

$$\int_{\mathbb{R}^{N-|u|}} y(\mathbf{x}) f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u} = \sum_{v \subseteq \{1,\ldots,N\}} \int_{\mathbb{R}^{N-|u|}} y_{v,G}(\mathbf{x}_v) f_{-u}(\mathbf{x}_{-u})\, d\mathbf{x}_{-u},$$

and apply Lemma 4.3 to evaluate each term in the sum according to the relationship between v and u:

(A) $v \cap u \neq \emptyset$ and $v \not\subset u$:
This is Case 1 of Lemma 4.3. The integral becomes:

$$\sum_{\substack{\emptyset \neq v \subseteq \{1,\dots,N\} \\ v \cap u \neq \emptyset,\ v \not\subset u}} \int_{\mathbb{R}^{|v \cap -u|}} y_{v,G}(\mathbf{x}_v) f_{v \cap -u}(\mathbf{x}_{v \cap -u})\, d\mathbf{x}_{v \cap -u}.$$

(B) $v \subsetneq u$:
This is Case 2 of Lemma 4.3. The integrals reduce to the component functions themselves:

$$\sum_{v \subsetneq u} y_{v,G}(\mathbf{x}_v).$$

(C) $v = u$:
Also part of Case 2 of Lemma 4.3. The integral becomes:

$$y_{u,G}(\mathbf{x}_u).$$

(D) $v \cap u = \emptyset$:
Case 3 of Lemma 4.3. These terms vanish:

$$\sum_{\substack{v \subseteq \{1,\dots,N\} \\ v \cap u = \emptyset}} 0 = 0.$$

Putting everything together:

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

Rearranging gives the almost final expression for $y_{u,G}(\mathbf{x}_u)$:

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

As a last step, we only have to write $v = (v \cap u) \cup (v \cap -u)$ to obtain the expression of Theorem 5.1.

□

3.2.2 Alternative Definition of Generalized fANOVA Components

Hooker (2007) approaches his generalization of the fANOVA decomposition from the angle of orthogonal projections. Instead of the more recursive definition of the components functions as in Rahman (2014), he defines the fANOVA components as a joint set which simultaneously minimizes the squared difference to the original function y under certain constraints. The constraints he sets for the optimization problem should ensure that the generalized components satisfy the desired properties of zero mean and hierarchical orthogonality.
The generalized fANOVA terms $\{y_u(x_u)|u \subseteq d\}$ jointly satisfiy:

$$\{y_{u,G}(\boldsymbol{x}_u) \,|\, u \subseteq d\} = \arg \min_{\{g_u \in L^2(\mathbb{R}^u)\}_{u \subseteq d}} \int \left(\sum_{u \subseteq d} g_u(\boldsymbol{x}_u) - y(\boldsymbol{x}) \right)^2 f_{\boldsymbol{X}}(\boldsymbol{x}) \, d\nu(\boldsymbol{x}) \tag{10}$$

under the hierarchical orthogonality conditions:

$$\forall v \subseteq u, \ \forall g_v : \int y_u(x_u) g_v(x_v) w(x) \, dx = 0. \tag{4.2}$$

In Hookers definition we recognize a projection. We are simultaneously finding the set of components functions g_u that minimize the weighted squared difference to the original function y (under zero mean and hierarchical orthogonality constraint), which is exactly the definition of a projection of y onto a specific subspace $\mathcal{G}$.
However, the constraint in Equation 4.2 is infeasible to enforce in practice. Therefore, Hooker formulated the following proposition, which ensures hierarchical orthogonality of the fANOVA components and thus forms the building block of his approach. It can be compared to the weak annihilating conditions in Rahman (2014).

Proposition 3.7. *The hierarchical orthogonality of the fANOVA components is ensured if and only if the following integral condition holds:*

$$\forall u \subseteq N, \ \forall i \in u : \int y_u(x_u) w(x) \, dx_i \, dx_{-u} = 0. \tag{4.3}$$

Proof. The proof is organized in two parts. First, Hooker needs to show that, if the integral conditions hold, the hierarchical orthogonality is true, and second, that if the hierarchical orthogonality does not hold, the integral conditions do not hold either. For the first part, assume that (4.3) holds. Let $i \in u \setminus v$, then $y_v(x_v)$ is independent of x_i and

x_{-u}, so we can write:

$$\int y_v(x_v)\, y_u(x_u)\, w(x)\, dx_i\, dx_{-u} = y_v(x_v) \int y_u(x_u)\, w(x)\, dx_i\, dx_{-u} = 0. \tag{11}$$

For the second part, assume that there exists a subset u and an index i for which (4.3) does not hold, i.e.

$$\int y_u(x_u)\, w(x)\, dx_i\, dx_{-u} \neq 0. \tag{12}$$

Further, assume that (4.3) does hold for a subset $v \neq u$ and an index $j \in v$. Hooker then constructs a fANOVA term y_v with lower order than y_u, which is not orthogonal to y_u. He sets $v = u \setminus \{i\}$, so y_v is one order lower than y_u and defined as:

$$y_v(x_v) := \int f_u(x_u)\, w(x)\, dx_i\, dx_{-u}. \tag{13}$$

y_v is a valid fANOVA component, which is unequal to zero by assumption of (4.3) being false, while it itself satisfies (4.3):

$$\forall j \in v, \quad \int y_v(x_v)\, w(x)\, dx_j\, dx_{-v} = 0 \tag{14}$$

Lastly, Hooker verifies that f_v is not orthogonal to f_u:

$$\begin{aligned} \langle y_u, y_v \rangle_w &= \int y_u(x_u)\, y_v(x_v)\, f_X(x)\, dx \\ &= \int y_u(x_u) \left(\int y_u(x_u)\, f_X(x)\, dx_i\, dx_{-u} \right) w(x)\, dx \\ &= \int \left(\int y_u(x_u)\, f_X(x)\, dx_i\, dx_{-u} \right)^2 dx_{u \setminus \{i\}} \\ &\neq 0. \end{aligned} \tag{15}$$

□

A crucial difference to the classical case is that both versions of the generalized components are defined in dependence of each other (**??**, Equation 10). This makes it in general difficult to compute the generalized fANOVA components analytically, even for simple functions.

3.2.3 Second-moment statistics

Given that the fANOVA decomposition changes under dependent inputs, it is interesting to examine what happens to the variance decomposition in this case. The mean of y re-

mains unchanged and is still given by the constant term $y_{\emptyset,G}$. The variance decomposition looks different, as cross-terms do not cancel out anymore:

$$\begin{aligned}
\sigma^2 &:= \mathbb{E}\left[(y(\mathbf{X}) - \mu)^2\right] \\
&= \mathbb{E}\left[\left(y_{\emptyset,G} + \sum_{\emptyset \neq u \subseteq \{1,\dots,N\}} y_{u,G}(\mathbf{X}_u) - \mu\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{\emptyset \neq u \subseteq \{1,\dots,N\}} y_{u,G}(\mathbf{X}_u)\right)^2\right] \\
&= \sum_{\emptyset \neq u} \mathbb{E}\left[y_{u,G}^2(\mathbf{X}_u)\right] + \sum_{\substack{\emptyset \neq u,\, v \subseteq \{1,\dots,N\} \\ u \neq v,\, u \not\subset v,\, v \not\subset u}} \mathbb{E}\left[y_{u,G}(\mathbf{X}_u) y_{v,G}(\mathbf{X}_v)\right]. \qquad (6.2)
\end{aligned}$$

The first term is the sum of the variances of the components, while the second term is the sum of the covariances between components that are not hierarchically orthogonal. The indices under the second component capture precisely the cross-terms that do not vanish under hierarchical orthogonality. For the classical fANOVA decomposition, the second term is zero for any relationship between u and v, and we are left with only the sum of the individual variances.

3.2.4 Correction of the Example: Dependent Multivariate Normal Inputs

For the end of this section, it remains to answer how the true generalized fANOVA decomposition looks like for our running example. While the interdependence of the generalized components makes it difficult to arrive at an analytical solution, Rahman (2014) provides a way to obtain the closed-from solution for any polynomial of maximum two degree under normally distributed input variables.

The idea in Rahman (2014) is based on Fourier-Polynomial expansion, which allows to write each generalized fANOVA component as a weighted sum of basis functions. The problem shifts from finding the fANOVA component functions to finding the basis functions which allow us to express the fANOVA terms. Rahman chooses Hermite polynomials as basis functions, which are by construction zero mean and hierarchical orthogonal. Satisfying these properties, the Hermite polynomial basis functions ensure fANOVA components that are also zero mean and hierarchical orthogonal. The challenge that remains is to find the weights for the basis functions, which can be done via coefficient matching; at least for a polynomial of degree two.

A polynomial of degree two has the general form

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

We know that any such polynomial may be expressed as a sum of weighted basis functions Nagler (2024) of the form:

$$y(x_1, x_2) = c_0 + c_{1,1}\,\psi_{1,1}(x_1) + c_{2,1}\,\psi_{2,1}(x_2) + c_{1,2}\,\psi_{1,2}(x_1) + c_{2,2}\,\psi_{2,2}(x_2) + c\,\psi(x_1, x_2)$$

where the $\psi_{i,j}$ are the basis functions with corresponding weights $c_0, \ldots, c_{12,11} \in \mathbb{R}$.
The idea is to carefully construct a set of basis functions which fulfill zero mean and hierarchical orthogonality. Then the expansion in these basis functions is already the fANOVA decomposition of a quadratic polynomial, i.e.

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

Coming from the probability density of a multivariate normal distribution, Rahman (2014) chooses multivariate Hermite polynomials. We use a slightly simplified (unscaled) version of the proposed basis functions to find an explicit solution for our running example. The basis functions we work with are:

$$\begin{aligned}
\psi_{\emptyset}(x_1, x_2) &= 1, \\
\psi_{1,1}(x_1) &= x_1, \\
\psi_{2,1}(x_2) &= x_2, \\
\psi_{1,2}(x_1) &= x_1^2 - 1, \\
\psi_{2,2}(x_2) &= x_2^2 - 1, \\
\psi(x_1, x_2) &= \frac{\rho(x_1^2 + x_2^2)}{1 + \rho^2} - x_1 x_2 + \frac{\rho(\rho^2 - 1)}{1 + \rho^2},
\end{aligned}$$

where ρ is the correlation coefficient between x_1 and x_2. So this formula will work for dependent as well independent inputs.

What remains it to find the coefficients $c_0, c_{1,1}, \ldots, c_{12,11}$ such the weighted sum of the basis really recovers the original polynomial. To find the correct weights, we plug in the basis functions and rearrange terms to recognize the groups more easily:

$$\begin{aligned} y(x_1, x_2) &= c_0 + c_{1,1}x_1 + c_{2,1}x_2 + c_{1,2}(x_1^2 - 1) + c_{2,2}(x_2^2 - 1) \\ &\quad + c_{12,11}\left(\frac{\rho(x_1^2 + x_2^2)}{1+\rho^2} - x_1x_2 + \frac{\rho(\rho^2-1)}{1+\rho^2}\right) \\ &= \left(c_0 - c_{1,2} - c_{2,2} + c_{12,11}\,\tfrac{\rho(\rho^2-1)}{1+\rho^2}\right) + c_{1,1}\,x_1 + c_{2,1}\,x_2 \\ &\quad + \left(c_{1,2} + c_{12,11}\,\tfrac{\rho}{1+\rho^2}\right)x_1^2 + \left(c_{2,2} + c_{12,11}\,\tfrac{\rho}{1+\rho^2}\right)x_2^2 - c_{12,11}\,x_1x_2 \end{aligned}$$

Now we can use monomial matching to find the coefficients. It is best to start with the interaction term and work backwards from there to the constant term, plugging in the current solutions along the way:

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

Hence, the generalized fANOVA decomposition of a two-degree polynomial is given by:

$$
\begin{aligned}
y(x_1, x_2) &= c_0 + c_{1,1}\,\psi_{1,1}(x_1) + c_{2,1}\,\psi_{2,1}(x_2) + c_{1,2}\,\psi_{1,2}(x_1) + c_{2,2}\,\psi_{2,2}(x_2) + c\,\psi(x_1, x_2) \\
&= \underbrace{\left(a_0 + a_{11} + a_{22} + \rho\, a_{12}\right)}_{c_0} + \underbrace{a_1}_{c_{1,1}} x_1 + \underbrace{a_2}_{c_{2,1}} x_2 \\
&\quad + \underbrace{\left(a_{11} + \frac{\rho}{1+\rho^2} a_{12}\right)}_{c_{1,2}} (x_1^2 - 1) + \underbrace{\left(a_{22} + \frac{\rho}{1+\rho^2} a_{12}\right)}_{c_{2,2}} (x_2^2 - 1) \\
&\quad + \underbrace{(-a_{12})}_{c_{12,11}} \left(\frac{\rho(x_1^2 + x_2^2)}{1+\rho^2} - x_1 x_2 + \frac{\rho(\rho^2 - 1)}{1+\rho^2} \right) \\
&= (a_0 + a_{11} + a_{22} + \rho a_{12}) + a_1 x_1 + a_2 x_2 \\
&\quad + \left(a_{11} + \frac{\rho}{1+\rho^2} a_{12}\right) (x_1^2 - 1) + \left(a_{22} + \frac{\rho}{1+\rho^2} a_{12}\right) (x_2^2 - 1) \\
&\quad - a_{12} \left(\frac{\rho(x_1^2 + x_2^2)}{1+\rho^2} - x_1 x_2 + \frac{\rho(\rho^2 - 1)}{1+\rho^2} \right),
\end{aligned}
$$

with individual components

$$
\begin{aligned}
y_0 &= a_0 + a_{11} + a_{22} + \rho\, a_{12}, \\
y_1(x_1) &= a_1\, x_1 + \left(a_{11} + \frac{\rho}{1+\rho^2} a_{12}\right) \left(x_1^2 - 1\right), \\
y_2(x_2) &= a_2\, x_2 + \left(a_{22} + \frac{\rho}{1+\rho^2} a_{12}\right) \left(x_2^2 - 1\right), \\
y_{12}(x_1, x_2) &= -a_{12} \left(\frac{\rho(x_1^2 + x_2^2)}{1+\rho^2} - x_1 x_2 + \frac{\rho(\rho^2 - 1)}{1+\rho^2} \right).
\end{aligned}
\tag{16}
$$

This set of component functions is true under the assumption of Gaussian inputs. The basis representation is still correct for other distribution assumptions in the sense that it recovers the original function g; however, the component function would not be hierarchically orthogonal.

With this we are able to give the fANOVA terms for our running example in a generalized from, which allows for dependent inputs variables, assumed to be Gaussian. For $g(x_1, x_2) = x_1 + 2x_2 + x_1 x_2$ we have $a_0 = 0, a_1 = 1, a_2 = 2, a_{11} = 0, a_{22} = 0, a_{12} = 1$ and

therefore obtain:

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

4 Visualization and Estimation

In the final section of this thesis, we will explore the fANOVA decomposition visually. This will provide better understanding for how fANOVA components behave in different scenarios; also if fANOVA should enhance interpretability, it is important to have visualizations of it. We will first revisit our running example and then explore some other functions.

4.1 Comparison of Decompositions

Recall our running example:

$$h(x_1, x_2) = x_1 + 2x_2 + x_1x_2,$$

with polynomial coefficients: $a_0 = 0$, $a_1 = 1$, $a_2 = 2$, $a_{11} = 0$, $a_{22} = 0$, $a_{12} = 1$. Under independent inputs ($\rho = 0$), the fANOVA components are given by:

$$\begin{aligned} y_\emptyset &= 0, \\ y_1(x_1) &= x_1 \\ y_2(x_2) &= 2x_2 \\ y_{12}(x_1, x_2) &= x_1x_2, \end{aligned}$$

visualized in Figure 1. As expected, we observe simple linear functions and a regular symmetric contour plot.

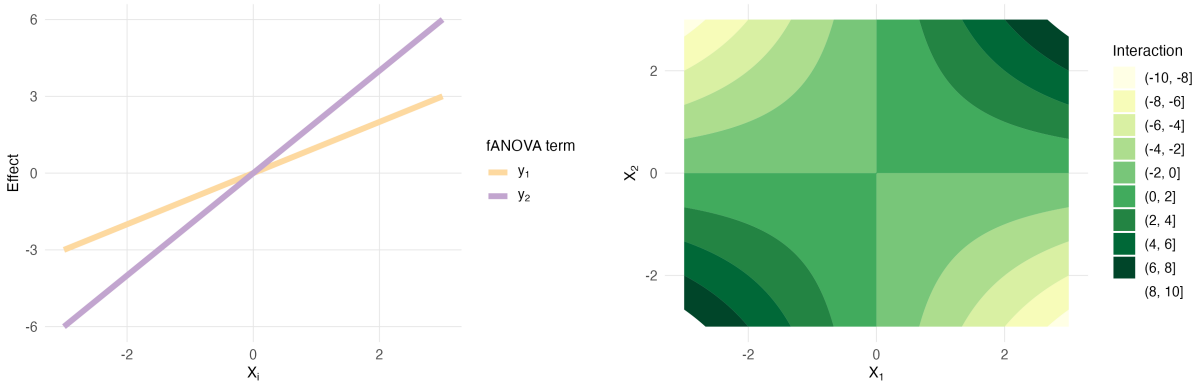


Figure 1: Main effects (left) and interaction effect (right) of the fANOVA decomposition for $h(x_1, x_2) = x_1 + 2x_2 + x_1x_2$ with independent inputs.

Now we assume $\rho = 0.5$. In attempt to compute the fANOVA terms under dependent inputs, we calculated the following effects, which are in reality the components of the

Hoeffding decomposition:

$$
\begin{aligned}
\tilde{y}_0 &= a + 0.5 \\
\tilde{y}_1(x_1) &= 2x_1 + 0.5x_1^2 - 0.5 \\
\tilde{y}_2(x_2) &= 2.5x_2 + 0.5x_2^2 - 0.5 \\
\tilde{y}_{12}(x_1, x_2) &= x_1x_2 - x_1 - 0.5x_2 - 0.5x_1^2 - 0.5x_2^2 + 0.5,
\end{aligned}
$$

which are visualized in Figure 2. The main effects are parabolic, and the interaction term seems to be non-symmetric. The true fANOVA components under $\rho = 0.5$ are given by:

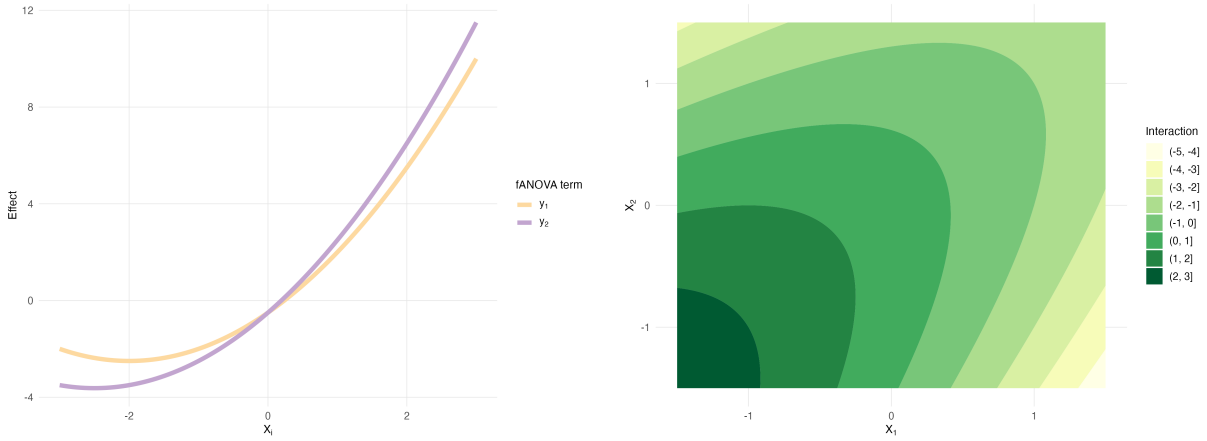


Figure 2: Main effects (left) and interaction effect (right) of the Hoeffding decomposition for $g(x_1, x_2) = x_1 + 2x_2 + x_1x_2$ with dependent inputs, $\rho = 0.5$.

$$
\begin{aligned}
y_0 &= 0.5 \\
y_1(x_1) &= x_1 + 0.4\,(x_1^2 - 1) = x_1 + 0.4x_1^2 - 0.4 \\
y_2(x_2) &= 2x_2 + 0.4\,(x_2^2 - 1) = 2x_2 + 0.4x_2^2 - 0.4 \\
y_{12}(x_1, x_2) &= -\Big(0.4(x_1^2 + x_2^2) - x_1x_2 - 0.3\Big) \\
&= -0.4x_1^2 - 0.4x_2^2 + x_1x_2 + 0.3.
\end{aligned}
$$

These are visualized in Figure 3. Interestingly, the parabolic form of the main effects is similar between Hoeffding and fANOVA, but the interaction effects diverge notably.
Our running example included linear effects of both input variables and the interaction term. For the remainder of this section, we will explore other representative scenarios, we can build within the scaffold of a bivariate two degree polynomial.

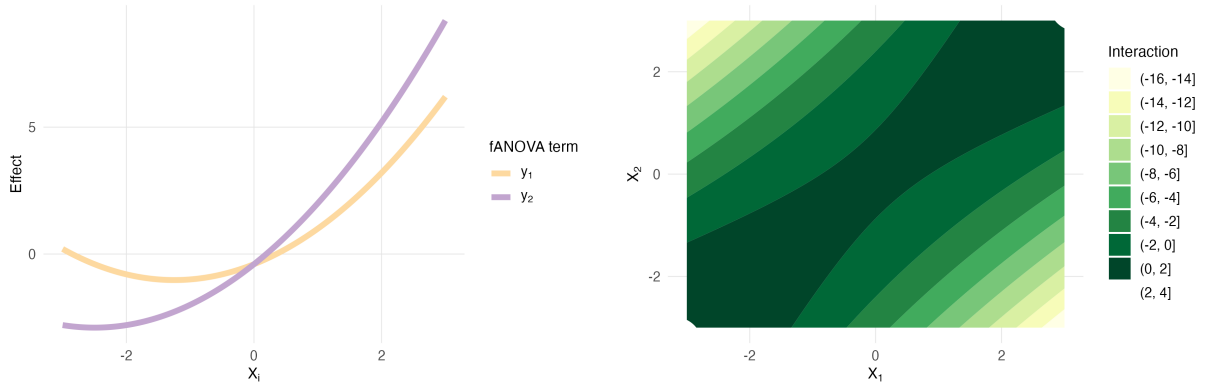


Figure 3: Main effects (left) and interaction effect (right) of the generalized fANOVA decomposition for $g(x_1, x_2) = x_1 + 2x_2 + x_1x_2$ with dependent inputs, $\rho = 0.5$.

4.2 Comparison of Functions

4.2.1 Scenario: Linear

First, we consider two-degree polynomials of the form:

$$g_1(x_1, x_2) = a_1x_1 + a_2x_2.$$

We can immediately read of the fANOVA components or use the general set of fANOVA components for a two degree polynomial in Equation 16 which simplify for g_1 to:

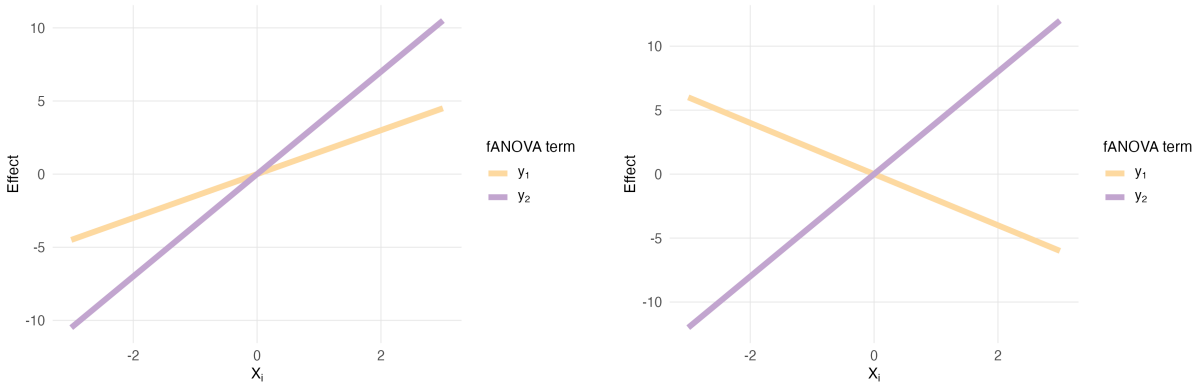
$$y_1(x_1) = a_1x_1$$
$$y_2(x_2) = a_2x_2.$$

The function g_1 can solely be described by linear main effects (Figure 4). Since no interaction effect is present varying ρ has no impact on the main effects.

4.2.2 Scenario: Mixed, only main

Slightly more complex is a two-degree polynomials, which allow for effects of linear and quadratic nature:

$$g_4(x_1, x_2) = a_1x_1 + a_2x_2 + a_{11}x_1^2 + a_{22}x_2^2.$$



(a) $a_1 = 1.5$, $a_2 = 3.5$ (b) $a_1 = -2$, $a_2 = 4$

Figure 4: Main fANOVA components for linear terms with different coefficients. The components are given by: $y_1(x_1) = a_1 x_1$, $y_2(x_2) = a_2 x_2$.

The fANOVA components for g_4 are given by:

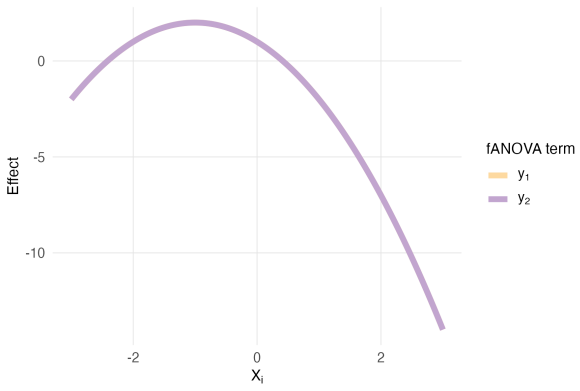
$$\begin{aligned} y_\emptyset &= a_{11} + a_{22}, \\ y_1(x_1) &= a_1 x_1 + a_{11}(x_1^2 - 1), \\ y_2(x_2) &= a_2 x_2 + a_{22}(x_2^2 - 1). \end{aligned}$$

The main effects are parabolas now. In Figure 5, we vary the coefficients a_1, a_2, a_{11}, and a_{22}, while the interaction term is still absent. The coefficients of the quadratic terms determine whether the parabola is facing downwards or upwards; when a_{11} and a_{22} are both negative or both positive the parabola is open downwards or upwards respectively, and when they have opposite signs the parabolas are open in different directions. Alongside the quadratic coefficients, the linear ones a_1 and a_2 influence how stretched or compressed the parabola is.

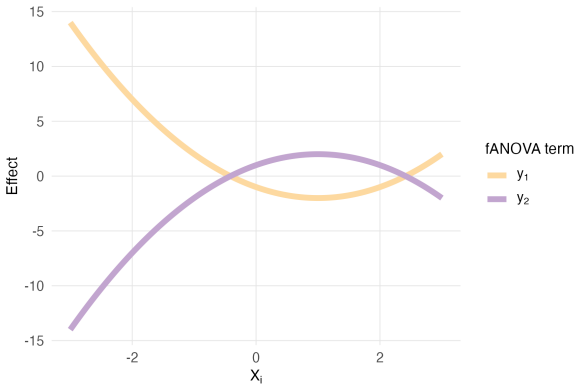
4.2.3 Scenario: Interaction only

Next, we consider a model, which solely consists of an interaction term:

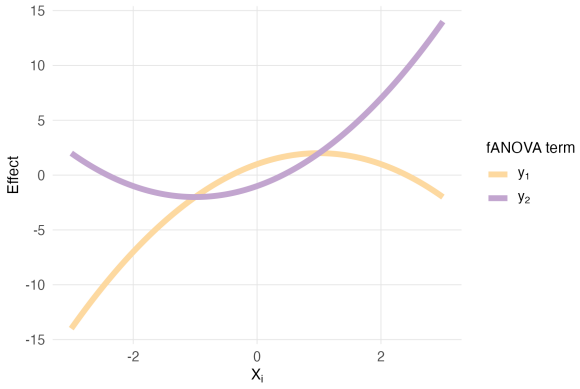
$$g_3(x_1, x_2) = a_{12} x_1 x_2.$$



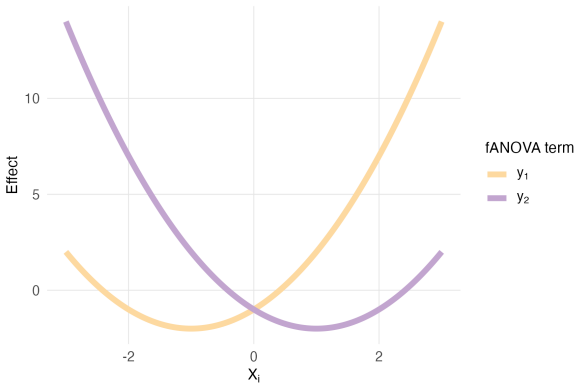
(a) $a_1 = -2$, $a_2 = -2$, $a_{11} = -1$, $a_{22} = -1$



(b) $a_1 = -2$, $a_2 = 2$, $a_{11} = 1$, $a_{22} = -1$



(c) $a_1 = 2$, $a_2 = 2$, $a_{11} = -1$, $a_{22} = 1$



(d) $a_1 = 2$, $a_2 = -2$, $a_{11} = 1$, $a_{22} = 1$

Figure 5: Main effects for different coefficient combinations of mixed main effects. The commponents are given by: $y_1(x_1) = a_1 x_1 + a_{11}(x_1^2 - 1)$, $y_2(x_2) = a_2 x_2 + a_{22}(x_2^2 - 1)$.

The fANOVA components for g_3 are given by:

$$\begin{aligned} y_\emptyset &= a_{12}\rho, \\ y_1(x_1) &= a_{12}\frac{\rho}{1+\rho}(x_1^2 - 1), \\ y_2(x_2) &= a_{12}\frac{\rho}{1+\rho}(x_2^2 - 1), \\ y_{12}(x_1, x_2) &= -a_{12}\left(\frac{\rho(x_1^2 + x_2^2)}{1+\rho^2} - x_1x_2 + \frac{\rho(\rho^2-1)}{1+\rho^2}\right). \end{aligned}$$

The main effects y_1 and y_2, as well as the interaction term y_{12}, are influenced by ρ and a_{12}. In our example we keep $a_{12} = 2$ fixed and show the interaction effect as a contour plot for varying ρ with the corresponding main effects next to it Figure 6. The main effects have the same form for every case of ρ and a_{12} and thus overlap. This example is simple yet interesting because it shows that in the case where the true function consists solely of an interaction term, fANOVA still attributes something to the isolated effect of each variable. Only when the variables are uncorrelated, all the effect is attributed to the interaction term. This functionality hints to why Lengerich et al. (2020) build an algorithm around fANOVA to purify interaction effects[1].

4.2.4 Scenario: All

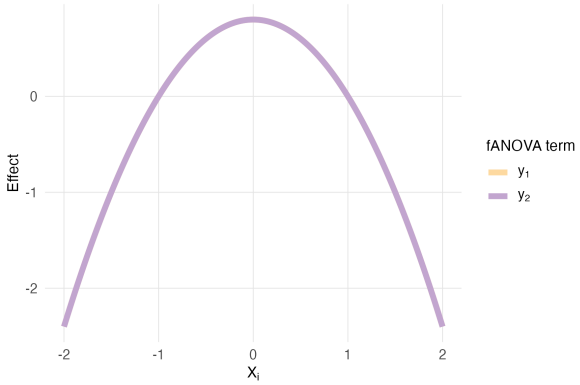
Finally, a full example, including all main and interaction effects:

$$g_5(x_1, x_2) = a_1x_1 + a_2x_2 + a_{11}x_1^2 + a_{22}x_2^2 + a_{12}x_1x_2.$$

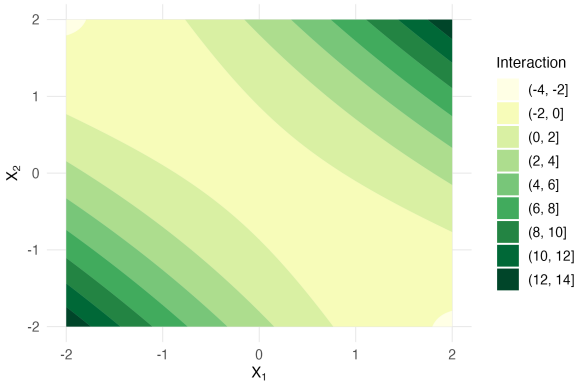
Now the fANOVA components are given by Equation 16, where $a_0 = 0$. We can vary the coefficients as well as ρ.

When the true function has no interaction term, as in our first two scenarios, varying ρ is uninteresting because there is no way it could influence the form of the main effects. In this full scenario, however, there is an interaction term present, and therefore it is most interesting to compare pairs of coefficient sets under $\rho = 0$ versus $\rho \neq 0$. With this we want to essentially ask how effects are distorted by performing the classical fANOVA decomposition when a true interaction effect is present and variables exhibit dependency (or is this nonsense because we essentially did this with our running example?). In Figure 7 we make this comparison for a weak linear correlation between variables and in Figure 8 we show the same for a strong linear correlation between variables. Similar to our running

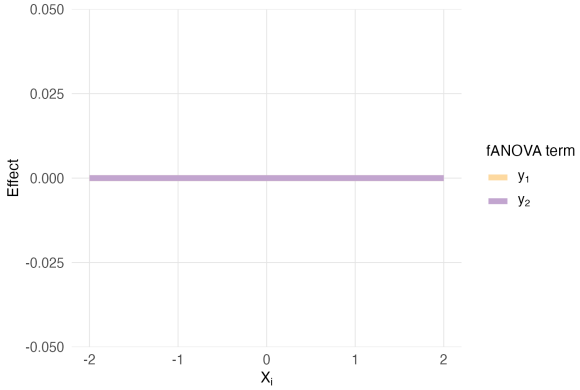
[1]Because they see a pure interaction as an effect which cannot be attributed to lower order terms; this means when identifying interactions we want to attribute all we can to lower order terms and what is left is the true interaction effect.



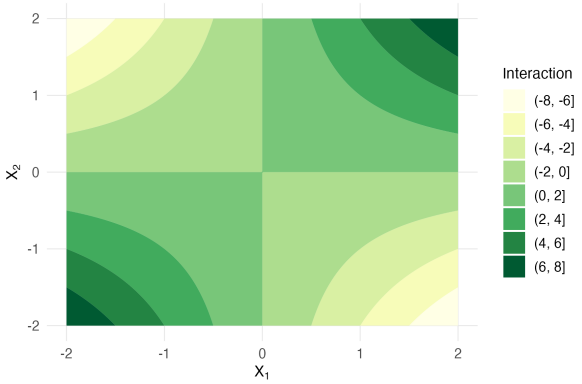
(a) Main effects for $\rho = -0.5$



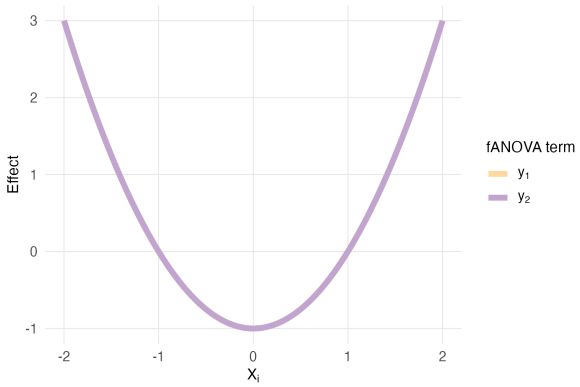
(b) Interaction contour for $\rho = -0.5$



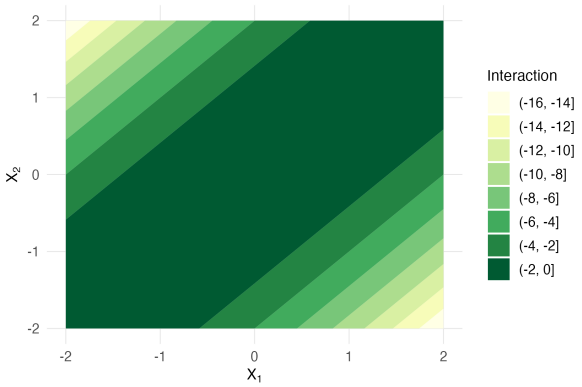
(c) Main effects for $\rho = 0$



(d) Interaction contour for $\rho = 0$



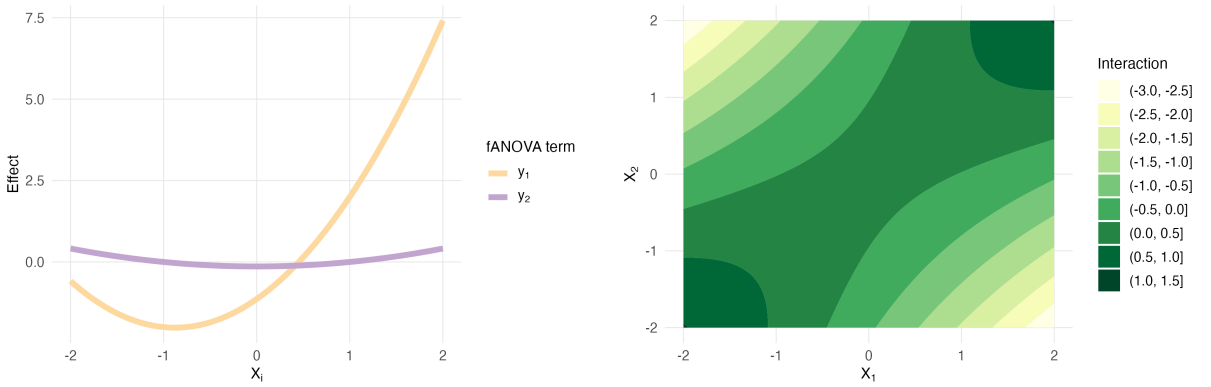
(e) Main effects for $\rho = 1$



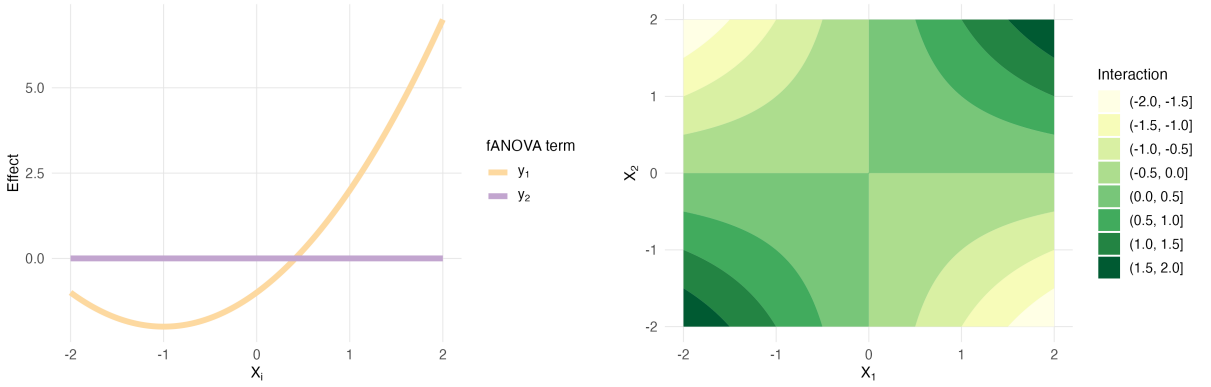
(f) Interaction contour for $\rho = 1$

Figure 6: Main effects (left column) and interaction contour plots (right column) for different values of ρ. The effects are given by $y_1(x_1) = a_{12}\frac{\rho}{1+\rho}(x_1^2-1)$, $y_2(x_2) = a_{12}\frac{\rho}{1+\rho}(x_2^2-1)$, $y_{12}(x_1, x_2) = -a_{12}\left(\frac{\rho(x_1^2+x_2^2)}{1+\rho^2} - x_1x_2 + \frac{\rho(\rho^2-1)}{1+\rho^2}\right)$.

example at the beginning of this section, we see that main effects are distorted slightly, while interaction effects look substantially different under dependent inputs.



(a) Main and interaction effects for (1) $a_1 = 2$, $a_2 = 0$, $a_{11} = 1$, $a_{22} = 0$, $a_{12} = 0.5$, $\rho = 0.3$.



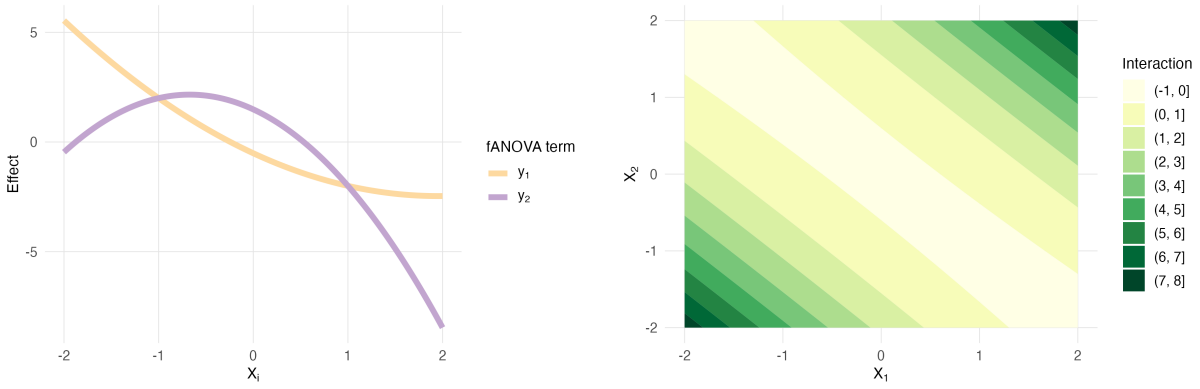
(b) Main and interaction effects for (2) $a_1 = 0$, $a_2 = 2$, $a_{11} = 0$, $a_{22} = 1$, $a_{12} = -0.5$, $\rho = 0$.

Figure 7: Main effects (left) and interaction contours (right) for four different coefficient sets.

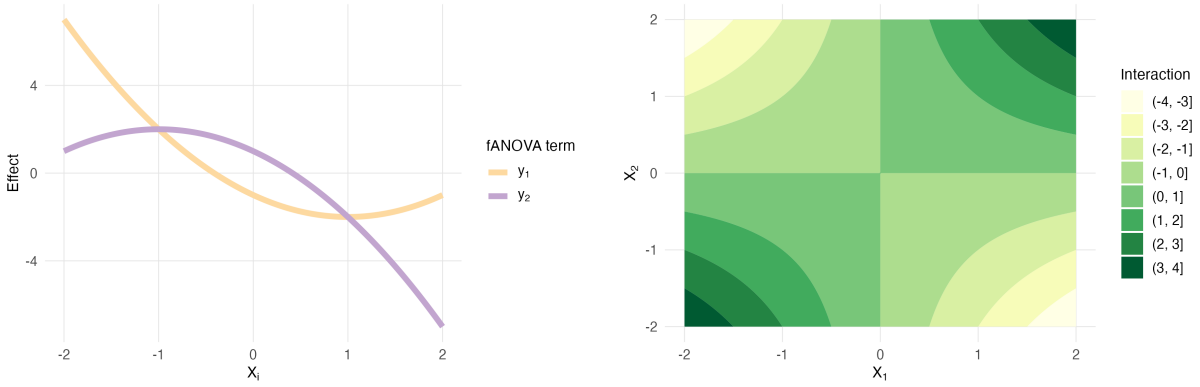
4.2.5 Estimation of fANOVA components

This was all pretty theoretical and the examples we used foster understanding, but they are toy examples and in reality the true function is unknown and more complex. So to become a more widely used, established interpretability method, an estimation scheme is inevitable. We already encountered one estimation scheme proposed by Rahman (2014) when computing the generalized fANOVA components for our running example; we refer to section 3 the conceptual idea behind it.

In Hooker (2004) an estimation framework based on partial dependence is proposed, which makes use of the formulation of fANOVA via projections. To obtain the component estimate for y_u, Hooker proposed to estimate the projections of y onto the subspace of variables spanned by u empirically. One does so by first estimating the conditional



(a) Main and interaction effects for (3) $a_1 = -2$, $a_2 = -2$, $a_{11} = 1$, $a_{22} = -1$, $a_{12} = 1$, $\rho = -0.8$.



(b) Main and interaction effects for (4) $a_1 = -2$, $a_2 = 2$, $a_{11} = -1$, $a_{22} = 1$, $a_{12} = -1$, $\rho = 0$.

Figure 8: Main effects (left) and interaction contours (right) for four different coefficient sets.

expected value of the variables in u. This is a simple Monte Carlo estimation, which results in the partial dependence function (PD Function) for the variables in u (Hooker, 2004). The PD Function can then be used to estimate the empirical projection of interest. He states that his method works well for functions that have nearly additive true structure and purely additive functions are exactly recoverable with this approach. However, the approach suffers from extrapolation issues or artefacts when the true function involves interactions and inputs are dependent.
Therefore, in Hooker (2007) a new estimation scheme is proposed for his version of the generalized fANOVA decomposition (see section 3). Hooker rewrites his proposed system of equations as restricted weighted least squares problem and solves it via Lagrange multiplier for the exact solution of the simultaneously defined generalized components. The function is evaluated at a grid of points to reduce computational costs. Because of the parallel to weighted least squares, it is also possible to compute a weighted standard ANOVA with existing software; however, like so it is difficult to incorporate the system constraints and one might obtain components that are not hierarchical orthogonal.
None of these estimation approaches has a standard software implementation or published code. Some existing unfinished implementations are numerically instable or yield illogical results. This underpins the need for a more robust estimation scheme with stable software implementation.

5 Conclusion

We started by working through the historical context of the fANOVA decomposition. We explored the origins of the fANOVA method rooted in mathematical work by Hoeffding (1948) and Sobol (1993). We saw how the method was picked up by following researchers in different contexts.

Clear contribution of this work: brought clarity and unity to the various different formulations of fANOVA. We see trend in recent ML literature (cite all these ML papers with the fancy models), pick up the methods but the theoretical background and clean formalism often left aside. This work serves as a reference to practitioners who seek a unified and clean formalization of the fANOVA method. Filled the void of visualizations and intuitions around the method due to the lack of software implementations.

Outlook, work that could follow from this thesis: Examine the different approaches to estimate fANOVA components (how do they scale? what is their accuracy? etc.) Write software implementation for fANOVA decomposition; current landscape is sparse but the method has great potential for IML; with current practicability it is however clear that fANOVA will not be accepted, it is not convenient to use the method fANOVA powerful theory, sound mathematical foundation, but without standardized software implementation application to IML difficult. Parallels to Shapley values, unified under a game theoretic approach; Fumagalli et al. (2025) recently established this parallel, would be very interesting to investigate further.

A Appendix

Proof of classical fANOVA decomposition

Here we show the proof of Theorem 1 in Sobol (1993).

Theorem A.1. *Any function y, which is integrable over the unit hypercube $[0,1]^k$, has a unique fANOVA expansion of the form:*

$$y(\boldsymbol{X}) = \sum_{u \subseteq \{1,\ldots,N\}} y_u(\boldsymbol{X}_u), \tag{17}$$

subject to the zero mean constraint Equation 3.

Sobol proofs existence and uniqueness of the fANOVA decomposition by showing how the summands of the desired decomposition look and showing that they satisfy the desired zero mean property.

Proof. Assume that $\boldsymbol{X}$ is an N-dimensional vector of independent random variables and that the zero mean property holds for the - still abstract - fANOVA terms. We define the integral w.r.t. all variables except for the ones with indices in v:

$$g_v(\boldsymbol{x}_v) = \int_{[0,1]^{N-|v|}} y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}_{-v}) \tag{18}$$

The very first term in the decomposition is the integral of y with respect to all variables:

$$y_\emptyset = \int_{\mathbb{R}^N} y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}) \tag{19}$$

This integral exists because $y \in L^2(\mathbb{R}^N, f_{\boldsymbol{X}} d\nu)$, and the product measure is finite on the domain.
Next, Sobol derives the one dimensional fANOVA terms. For this, take the integral of

Equation 1 w.r.t. all variables except for the one with index i, so $v_1 = \{i\}$:

$$\int_{\mathbb{R}^{N-1}} y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\nu(\boldsymbol{x}_{-v_1}) = \int_{\mathbb{R}^{N-1}} \sum_{u \subseteq \{1,\dots,N\}} y_u(\boldsymbol{x}_u) f_{\boldsymbol{X}}(\boldsymbol{x}) d\nu(\boldsymbol{x}_{-v_1}) \tag{20}$$

$$= \sum_{u \subseteq \{1,\dots,N\}} \int_{\mathbb{R}^{N-1}} y_u(\boldsymbol{x}_u) f_{\boldsymbol{X}}(\boldsymbol{x}) d\nu(\boldsymbol{x}_{-v_1}) \tag{21}$$

$$= \sum_{u \subseteq \{1,\dots,N\}} \int_{\mathbb{R}^{N-1}} y_u(\boldsymbol{x}_u) \left(\prod_{j=1}^{N} f_{X_j}(x_j) \right) d\nu(\boldsymbol{x}_{-v_1}) \tag{22}$$

For every summand $y_u(\boldsymbol{x}_u)$ with $u \not\ni i$, the integrand does not depend on x_i, and thus vanished due to the zero mean constraint. Similarly, for any term $y_u(\boldsymbol{x}_u)$ with $i \in u$ and $|u| > 1$, the integration will include at least one other variable in u, again causing the integral to vanish. In the end, only the constant term $y_\emptyset$ and the one-dimensional term $y_{\{i\}}(x_i)$ remain, which depend only on x_i and are not integrated. Therefore, we can derive the simplified expression:

$$\int_{\mathbb{R}^{N-1}} y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\nu(\boldsymbol{x}_{-v_1}) = y_\emptyset + y_{\{i\}}(x_i). \tag{23}$$

This equation allows us to define the one-dimensional term $y_{\{i\}}$ explicitly as:

$$y_{\{i\}}(x_i) = \int_{\mathbb{R}^{N-1}} y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\nu(\boldsymbol{x}_{-v_1}) - y_\emptyset. \tag{24}$$

Now consider $v_2 = \{1, 2\}$. We integrate the ANOVA decomposition over all variables except x_1 and x_2:

$$\int_{\mathbb{R}^{N-2}} y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\nu(\boldsymbol{x}_{-\{1,2\}}) = \int_{\mathbb{R}^{N-2}} \sum_{u \subseteq \{1,\dots,N\}} y_u(\boldsymbol{x}_u) f_{\boldsymbol{X}}(\boldsymbol{x}) d\nu(\boldsymbol{x}_{-\{1,2\}}) \tag{25}$$

$$= \sum_{u \subseteq \{1,\dots,N\}} \int_{\mathbb{R}^{N-2}} y_u(\boldsymbol{x}_u) \left(\prod_{j=1}^{N} f_{X_j}(x_j) \right) d\nu(\boldsymbol{x}_{-\{1,2\}}) \tag{26}$$

$$= y_\emptyset + y_{\{1\}}(x_1) + y_{\{2\}}(x_2) + y_{\{1,2\}}(x_1, x_2) \tag{27}$$

Hence, the two-dimensional component is given by:

$$y_{\{1,2\}}(x_1, x_2) = \int_{\mathbb{R}^{N-2}} y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\nu(\boldsymbol{x}_{-\{1,2\}}) - y_\emptyset - y_{\{1\}}(x_1) - y_{\{2\}}(x_2) \tag{28}$$

We can continue this process for all combinations of indices $v \subseteq \{1, \ldots, N\}$ to derive the corresponding fANOVA terms $y_v(\boldsymbol{x}_v)$. Now let $v \subseteq \{1, \ldots, N\}$. The general expression for the component $y_v(\boldsymbol{x}_v)$ is given by:

$$y_v(\boldsymbol{x}_v) = \int_{\mathbb{R}^{N-|v|}} y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}) d\nu(\boldsymbol{x}_{-v}) - \sum_{u \subsetneq v} y_u(\boldsymbol{x}_u) \tag{29}$$

The last term is the decomposition integrated with respect to no variables, i.e., the function itself:

$$y_{\{1,\ldots,N\}}(\boldsymbol{x}) = y(\boldsymbol{x}) - \sum_{u \subsetneq \{1,\ldots,N\}} y_u(\boldsymbol{x}_u) \tag{30}$$

Finally, we verify that the constructed component functions satisfy the zero mean constraint. Let $v \subseteq \{1, \ldots, N\}$, and let $i \in v$. Then:

$$\int y_v(\boldsymbol{x}_v) f_{X_i}(x_i)\, d\nu(x_i) = \int \left(\int y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}_{-v}) - \sum_{u \subsetneq v} y_u(\boldsymbol{x}_u) \right) f_{X_i}(x_i)\, d\nu(x_i) \tag{31}$$

$$= \int \left(\int y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}_{-v}) \right) f_{X_i}(x_i)\, d\nu(x_i) - \sum_{u \subsetneq v} \int y_u(\boldsymbol{x}_u) f_{X_i}(x_i)\, d\nu(x_i) \tag{32}$$

$$= \int y(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\nu(\boldsymbol{x}_{-v}) d\nu(x_i) - \sum_{u \subsetneq v} \int y_u(\boldsymbol{x}_u) f_{\boldsymbol{X}_u}(\boldsymbol{x}_u)\, d\nu(\boldsymbol{x}_u) \tag{33}$$

The first term integrates out all of $\boldsymbol{x}_v$, leaving $y_\emptyset$. Each term in the sum vanishes by the zero-mean property of lower-order components:

$$\int y_v(\boldsymbol{x}_v) f_{X_i}(x_i)\, d\nu(x_i) = y_\emptyset - y_\emptyset = 0 \tag{34}$$

Thus, every component $y_v(\boldsymbol{x}_v)$ satisfies:

$$\int y_v(\boldsymbol{x}_v) f_{X_i}(x_i)\, d\nu(x_i) = 0, \quad \text{for all } i \in v. \tag{35}$$

This completes the proof of the existence and uniqueness of the fANOVA decomposition,

as well as the verification of the zero mean property for all components. □

Square Integrability of $f_1(x_1)$

For now we want to show that the single fANOVA term $f_1(x_1)$ is square integrable, given that the original function $f(x) \in \mathcal{L}^2$. We need to show that:

$$\int |f_1(x_1)|^2 \, dx_1 < \infty$$

The single fANOVA term is defined as:

$$f_1(x_1) = \int f(x) \, dx_{-1} - f_0$$

We take the squared norm, and integrate w.r.t. x_1 to use the Cauchy-Schwarz inequality:

$$\int |f_1(x_1)|^2 \, dx_1 = \int \left| \int f(x) \, dx_{-1} - f_0 \right|^2 dx_1$$

$$= \int |(\int f(x) \, dx_{-1})^2 - 2 \int f(x) \, dx_{-1} f_0 + f_0^2 | dx_1$$

Break this into three terms:

$$(1): \quad \int \left| \int f(x) \, dx_{-1} \right|^2 dx_1 \leq \int \left(\int 1^2 \, dx_{-1} \right) \left(\int |f(x)|^2 \, dx_{-1} \right) dx_1 = \int |f(x)|^2 \, dx < \infty$$

$$(2): \quad 2 \int \left(\int f(x) \, dx_{-1} \right) f_0 \, dx_1 = 2 f_0 \int \left(\int f(x) \, dx_{-1} \right) dx_1 = 2 f_0^2 < \infty$$

$$(3): \quad \int f_0^2 \, dx_1 = f_0^2 < \infty$$

Since each term (1)–(3) is finite, and $\int |f_1(x_1)|^2 dx_1$ is a linear combination of them: $\int |f_1(x_1)|^2 dx_1 < \infty$

B Electronic appendix

Data, code and figures are provided in electronic form.

References

Borgonovo, E., Li, G., Barr, J., Plischke, E. and Rabitz, H. (2022). Global Sensitivity Analysis with Mixtures: A Generalized Functional ANOVA Approach, *Risk Analysis* **42**(2): 304–333.
URL: *https://onlinelibrary.wiley.com/doi/10.1111/risa.13763*

Chastaing, G., Gamboa, F. and Prieur, C. (2012). Generalized Hoeffding-Sobol decomposition for dependent variables - application to sensitivity analysis, *Electronic Journal of Statistics* **6**(none).

Choi, Y., Park, S., Park, C., Kim, D. and Kim, Y. (2025). Meta-anova: screening interactions for interpretable machine learning, *Journal of the Korean Statistical Society* .
URL: *https://link.springer.com/10.1007/s42952-024-00302-2*

Fumagalli, F., Muschalik, M., Hüllermeier, E., Hammer, B. and Herbinger, J. (2025). Unifying Feature-Based Explanations with Functional ANOVA and Cooperative Game Theory. arXiv:2412.17152 [cs].
URL: *http://arxiv.org/abs/2412.17152*

Gu, C. (2013). *Smoothing Spline ANOVA Models*, Vol. 297 of *Springer Series in Statistics*, Springer New York, New York, NY.
URL: *https://link.springer.com/10.1007/978-1-4614-5369-7*

Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution, *The Annals of Mathematical Statistics* **19**(3): 293–325. Publisher: Institute of Mathematical Statistics.
URL: *https://www.jstor.org/stable/2235637*

Hooker, G. (2004). Discovering additive structure in black box functions, *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, Seattle WA USA, pp. 575–580.
URL: *https://dl.acm.org/doi/10.1145/1014052.1014122*

Hooker, G. (2007). Generalized Functional ANOVA Diagnostics for High-Dimensional Functions of Dependent Variables, *Journal of Computational and Graphical Statistics* **16**(3): 709–732.
URL: *http://www.tandfonline.com/doi/abs/10.1198/106186007X237892*

Hu, L., Nair, V. N., Sudjianto, A., Zhang, A., Chen, J. and Yang, Z. (2025). Interpretable Machine Learning Based on Functional ANOVA Framework: Algorithms and Comparisons, *Applied Stochastic Models in Business and Industry* **41**(1): e2916.
URL: *https://onlinelibrary.wiley.com/doi/10.1002/asmb.2916*

Il Idrissi, M., Bousquet, N., Gamboa, F., Iooss, B. and Loubes, J.-M. (2025). Hoeffding decomposition of functions of random dependent variables, *Journal of Multivariate Analysis* **208**: 105444.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0047259X25000399*

König, G., Günther, E. and Luxburg, U. v. (2024). Disentangling Interactions and Dependencies in Feature Attribution. arXiv:2410.23772 [cs].
URL: *http://arxiv.org/abs/2410.23772*

Lengerich, B., Tan, S., Chang, C.-H., Hooker, G. and Caruana, R. (2020). Purifying Interaction Effects with the Functional ANOVA: An Efficient Algorithm for Recovering Identifiable Additive Models, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, PMLR, pp. 2402–2412. ISSN: 2640-3498.
URL: *https://proceedings.mlr.press/v108/lengerich20a.html*

Liu, R. and Owen, A. B. (2006). Estimating Mean Dimensionality of Analysis of Variance Decompositions, *Journal of the American Statistical Association* **101**(474): 712–721.
URL: *https://www.tandfonline.com/doi/full/10.1198/016214505000001410*

Molnar, C. (2025). *Interpretable Machine Learning*, 3 edn.
URL: *https://christophm.github.io/interpretable-ml-book*

Muehlenstaedt, T., Roustant, O., Carraro, L. and Kuhnt, S. (2012). Data-driven Kriging models based on FANOVA-decomposition, *Statistics and Computing* **22**(3): 723–738.
URL: *http://link.springer.com/10.1007/s11222-011-9259-7*

Nagler, T. (2024). Methoden der Linearen Algebra in der Statistik – Vorlesungsskript, `https://tnagler.github.io/linalg-2024.pdf`. Version Sommersemester 2024.

Owen, A. B. (2013). Variance components and generalized sobol' indices, *SIAM/ASA Journal on Uncertainty Quantification* **1**(1): 19–41. tex.eprint: https://doi.org/10.1137/120876782.
URL: *https://doi.org/10.1137/120876782*

Owen, A. B. (2014). Sobol' Indices and Shapley Value, *SIAM/ASA Journal on Uncertainty Quantification* **2**(1): 245–251.

Rahman, S. (2014). A generalized anova dimensional decomposition for dependent probability measures, *SIAM/ASA Journal on Uncertainty Quantification* **2**(1): 670–697.
URL: *https://doi.org/10.1137/120904378*

Sobol, I. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation* **55**(1-3): 271–280.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0378475400002706*

Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models, *Mathematical Modelling and Computational Experiments* **1**: 407–414.

Stone, C. J. (1994). The Use of Polynomial Splines and Their Tensor Products in Multivariate Function Estimation, *The Annals of Statistics* **22**(1): 118–171. Publisher: Institute of Mathematical Statistics.
URL: *https://www.jstor.org/stable/2242446*

Stone, C. J., Hansen, M. H., Kooperberg, C. and Truong, Y. K. (1997). Polynomial Splines and their Tensor Products in Extended Linear Modeling, *The Annals of Statistics* **25**(4): 1371–1425. Publisher: Institute of Mathematical Statistics.
URL: *https://www.jstor.org/stable/2959054*

Takemura, A. (1983). Tensor Analysis of ANOVA Decomposition, *Journal of the American Statistical Association* **78**(384): 894–900. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].
URL: *https://www.jstor.org/stable/2288201*

Van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Declaration of authorship

I hereby declare that the report submitted is my own unaided work. All direct or indirect sources used are acknowledged as references. I am aware that the Thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the report as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future Theses submitted. Further rights of reproduction and usage, however, are not granted here. This paper was not previously presented to another examination board and has not been published.

Location, date

Name