

Measuring the additivity of functions with variance-based global sensitivity analysis

John Darges

August 2022

1 Distance from additivity with ANOVA decomposition

We can find a means of measuring additivity by looking to the ANOVA decomposition. ANOVA allows us to separate f into different components corresponding to groups of inputs. The additive portion of f is then the sum of the univariate ANOVA terms. Owen defines an additive function in this manner, using the ANOVA decomposition [5].

Definition 1.1 (Owen [5]). The **additive portion** of f is

$$f_{\text{add}}(\mathbf{x}) = \mathbb{E}(f) + \sum_{i=1}^d f_{\{i\}}(\mathbf{x}),$$

where $f_{\{i\}}$ is the first-order ANOVA term $f_{\{i\}}(\mathbf{x}) := \mathbb{E}(f(\mathbf{x})|\mathbf{x}_{-i})$.

Note, then, that taking $f - f_{\text{add}}$ extracts only the interaction ANOVA terms of f . If $f = f_{\text{add}}$, then f is an additive function, since $f - f_{\text{add}} = 0$.

Using Definition 1.1 as a frame for viewing the additivity of f , the relative magnitude of $f = f_{\text{add}}$ could serve to measure 'distance' from being fully additive. In practice, however, we cannot get our hands on $f = f_{\text{add}}$, since finding the ANOVA decomposition is impractical. The following result, which shows that f_{add} is the best additive approximation to f , could point us towards a strategy to approximate $f = f_{\text{add}}$.

Lemma 1.1 (Owen [5]). Let $f \in L^2([-1, 1]^d)$. Then, for any additive function $g \in L^2([-1, 1]^d)$,

$$\|f - g\|_{L^2} \geq \|f - f_{\text{add}}\|_{L^2}. \quad (1)$$

By slightly modifying (1) in Lemma 1.1, we can define distance from additivity in terms of first-order Sobol' indices known from [? ?].

Corollary 1.2. Let $f \in L^2([-1, 1]^d)$. Then, for any additive function $g \in L^2([-1, 1]^d)$,

$$\frac{\|f - g\|_{L^2}^2}{\text{var}(f)} \geq 1 - \sum_{i=1}^d S_i(f) = \sum_{|\mathbf{u}| > 1} S_{\mathbf{u}}(f), \quad (2)$$

where $S_i(f)$ are the first-order Sobol' indices and $S_{\mathbf{u}}(f)$ are the higher-order Sobol' indices of f [? ?].

Proof. Here, we show that $\frac{\|f - f_{\text{add}}\|_{L^2}^2}{\text{var}(f)} = 1 - \sum_{i=1}^d S_i(f)$. Recalling Definition 1.1, we have

$f_{\text{add}} = \mathbb{E}(f) + \sum_{i=1}^d f_{\{i\}}$. Therefore, $f - f_{\text{add}} = \sum_{|\mathbf{u}|>1} f_{\mathbf{u}}$. Since the terms in the ANOVA decomposition are orthogonal,

$$\|f - f_{\text{add}}\|_{L^2}^2 = \int \left(\sum_{|\mathbf{u}|>1} f_{\mathbf{u}} \right)^2 d\mathbf{x} \quad (3)$$

$$= \sum_{|\mathbf{u}|>1} \int f_{\mathbf{u}}^2 d\mathbf{x} \quad (4)$$

$$= \sum_{|\mathbf{u}|>1} \sigma_{\mathbf{u}}^2 \quad (5)$$

When we divide by the variance of f , we arrive at

$$\frac{\|f - f_{\text{add}}\|_{L^2}^2}{\text{var}(f)} = \sum_{|\mathbf{u}|>1} S_{\mathbf{u}}(f) = 1 - \sum_{i=1}^d S_i(f)$$

□

The left-hand side of (2) provides a bound on the sum of the higher-order Sobol' indices, which measure the importance of interactions.

Lemma 1.1 suggests that we can find an upper bound $\|f - g\|_{L^2}$ on the L^2 -norm of the non-additive portion of f by choosing $g \approx f_{\text{add}}$. We define distance from additivity in the following manner.

Definition 1.2. Let $f \in L^2([-1, 1]^d)$. The **distance from additivity** of f , denoted Δ_{add} is

$$\Delta_{\text{add}} := \frac{\|f - f_{\text{add}}\|_{L^2}}{\|f\|_{L^2}}.$$

Therefore, choosing $g \approx f_{\text{add}}$ should mean

$$\frac{\|f - g\|_{L^2}}{\|f\|_{L^2}} \geq \Delta_{\text{add}} \quad (6)$$

provides a tight bound on the distance from additivity.

2 Surrogates for bound on distance from additivity

As a practical approach to approximating the bound in (6), we propose to compute a surrogate $\hat{f} \approx f$ for which the ANOVA terms can readily be computed [2, 4?]. Therefore, we should be able to get our hands on a close approximation for f_{add} using \hat{f}_{add} . By choosing $g = \hat{f}_{\text{add}}$, we can approximate the integrals in (6) by sampling.

Remark 2.1. For a given function f , define

$$\mathbf{P}_{\text{add}}(f) = f_{\text{add}}$$

We can either compute the additive portion of a surrogate for f

$$\|f - \mathbf{P}_{\text{add}}(\hat{f})\|,$$

where \hat{f} is a surrogate for f , or we can compute a surrogate for the additive portion of f

$$\|f - \widehat{\mathbf{P}_{\text{add}}(f)}\|,$$

where here $\widehat{\mathbf{P}_{\text{add}}(f)}$ is a surrogate for $\mathbf{P}_{\text{add}}(f)$. Note that we know both the following inequalities are valid:

$$\|f - \mathbf{P}_{\text{add}}(f)\| \leq \|f - \mathbf{P}_{\text{add}}(\hat{f})\| \quad \text{and} \quad \|f - \mathbf{P}_{\text{add}}(f)\| \leq \|f - \widehat{\mathbf{P}_{\text{add}}(f)}\|.$$

The key questions are (i) which bound is better? And (ii) which one is easier to compute?

2.1 Approach 1: Compute surrogate for model and take its additive portion

The first approach described in Remark 2.1 entails constructing a surrogate $\hat{f} \approx f$ for the model f .

We project \hat{f} onto its additive portion $\mathbf{P}_{\text{add}}(\hat{f})$. Setting $g = \mathbf{P}_{\text{add}}(\hat{f})$, the upper bound on (6) becomes

$$\frac{\|f - \mathbf{P}_{\text{add}}(\hat{f})\|_{L^2}}{\|f\|_{L^2}} \geq \Delta_{\text{add}}.$$

The surrogate \hat{f} can be constructed by a method of choice. The term $\|f - \mathbf{P}_{\text{add}}(\hat{f})\|_{L^2}$ must be approximated by numerical integration. We can approximate $\|f\|_{L^2}$ either by numerical integration or use the surrogate to approximate it.

2.2 Approach 2: Compute surrogate for additive portion of model

The second approach in Remark 2.1 entails constructing a surrogate $\widehat{\mathbf{P}_{\text{add}}(f)}$ for the model's additive portion $\mathbf{P}_{\text{add}}(f)$.

We can compute a PCE surrogate for $\mathbf{P}_{\text{add}}(f)$ by finding the coefficients on the additive terms in the PCE of f . Gaussian quadrature can accurately compute the coefficients, but

is costly in higher dimensions. Recall, the PCE coefficient corresponding to ϕ_i^k , univariate polynomial of degree k , is given by

$$c_{i,k}^2 = \int \phi_i^k f d\mathbf{x},$$

where we assume the basis polynomials are orthonormal. The quadrature formula for integrating a multivariate function h is

$$\int h d\mathbf{x} \approx \sum_{i=1}^d \sum_{n_i=1}^{N_i} w_{n_i} h(\mathbf{x}_{n_i}),$$

where N_i are the levels of accuracy in each direction and w_{n_i} and \mathbf{x}_{n_i} are the respective quadrature weights and nodes.

We can estimate the coefficient $c_{i,k}$, using level of accuracy N , by

$$\int \phi_i^k f d\mathbf{x} \approx w_0^{d-1} \sum_{n=1}^N w_n \phi_i^k(\mathbf{x}_n) f(\mathbf{x}_n),$$

where \mathbf{x}_n has the n th quadrature node in the i th entry and zero in all other entries.

3 Numerical experiments

Here, we test the method detailed above for the following test example:

$$f_\delta(\mathbf{x}) = \sum_{i=1}^5 x_i + \delta \prod_{j=1}^5 (1 + x_j), \quad \mathbf{x} \in [0, 1]^5, \quad (7)$$

where δ controls the influence of interactions.

Here, we take the first approach in Remark 2.1, where we construct a surrogate \hat{f} of f and project on the additive portion of surrogate $\mathbf{P}_{\text{add}}(\hat{f})$. In experiments, we estimate $\|f - \mathbf{P}_{\text{add}}(\hat{f})\|_{L^2}$ and $\|f\|_{L^2}$ using 10^2 samples.

We construct PCE [1, 6], standard ELM [3], and SW-ELM surrogates [?]. The surrogates use progressively larger basis sizes. We use basis sizes (or numbers of neurons for ELM) $N = \frac{(n+d)!}{n!d!}$ corresponding to the size of the degree n PCE basis for $n = 1, \dots, 5$. We always use $2N$ training points (two training points for every basis function) to perform regression. At each basis size, the three surrogates use the same training sets for regression.

We compare surrogate-estimated bounds to the true distance from additivity given in Definition 1.2. Since we have access to values for the first-order Sobol' indices, mean, and variance of f_δ , we use Corollary 1.2 to say $\Delta_{\text{add}} = \frac{\sqrt{\text{var}(f)}}{\|f\|_{L^2}} \sqrt{1 - \sum_{i=1}^d S_i(f)}$.

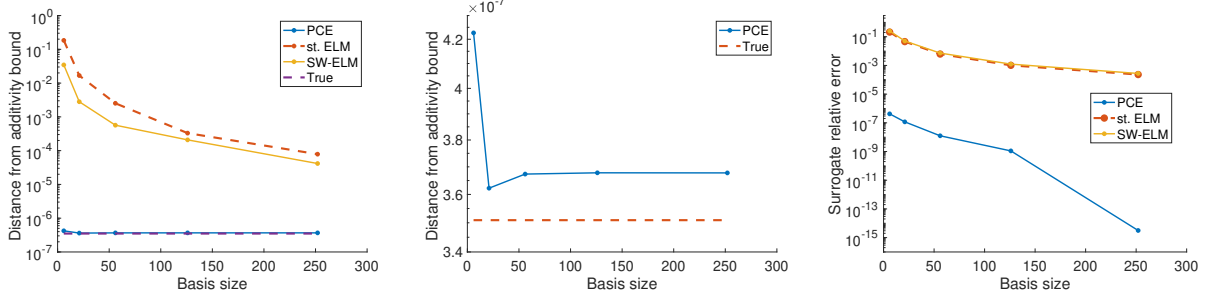


Figure 1: Low interaction case (10^{-6}) where $\Delta_{\text{add}} = 3.5091 \times 10^{-7}$. **Left:** Distance from additivity bounds estimated with PCE, standard ELM, and SW-ELM. **Middle:** Distance from additivity bound for PCE only compared to true distance from additivity. **Right:** Relative error of surrogates.

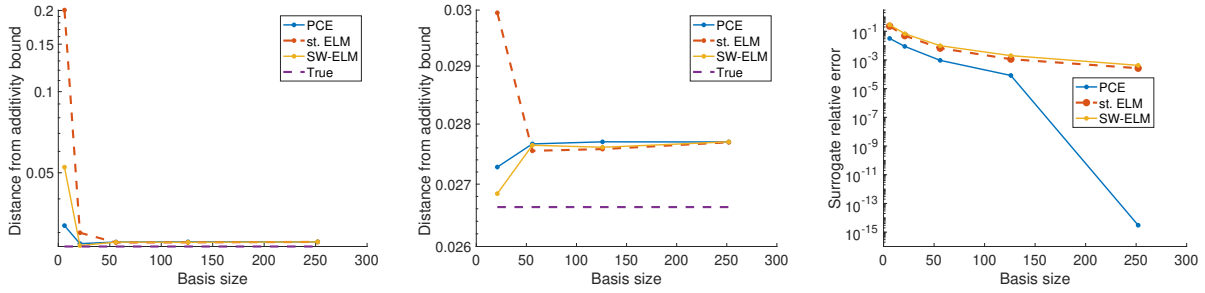


Figure 2: Medium interaction case ($\delta = 10^{-1}$) where $\Delta_{\text{add}} = 0.0266$. **Left:** Distance from additivity bounds estimated with PCE, standard ELM, and SW-ELM. **Middle:** Same as Left figure, but zoomed in. **Right:** Relative error of surrogates.

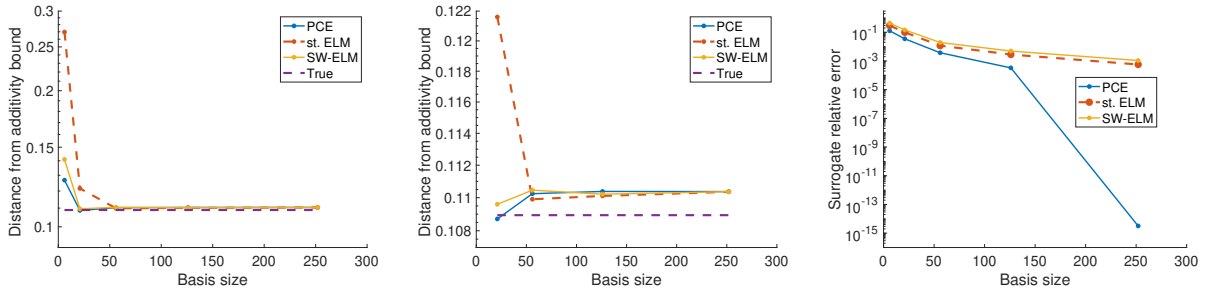


Figure 3: High interaction case ($\delta = 10^4$) where $\Delta_{\text{add}} = 0.1089$. **Left:** Distance from additivity bounds estimated with PCE, standard ELM, and SW-ELM. **Middle:** Same as Left figure, but zoomed in. **Right:** Relative error of surrogates.

References

- [1] T. Crestaux, O. L. Maitre, and J.-M. Martinez. Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94(7):1161 – 1172, 2009. Special Issue on Sensitivity Analysis.
- [2] J. Darges, A. Alexanderian, and P. Gremaud. Extreme learning machines for variance-based global sensitivity analysis. *preprint*, 2022.
- [3] L. L. Gratiet, S. Marelli, and B. Sudret. Metamodel-based sensitivity analysis: polynomial chaos expansions and gaussian processes. In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Handbook of Uncertainty Quantification*. Springer, 2017.
- [4] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1):489–501, 2006. Neural Networks.
- [5] R. Jin, W. Chen, and A. Sudjianto. Analytical metamodel-based global sensitivity analysis and uncertainty propagation for robust design. *SAE transactions*, pages 121–128, 2004.
- [6] A. B. Owen. *Monte Carlo theory, methods and examples*. Art B. Owen, 2013.
- [7] C. Prieur and S. Tarantola. Variance-based sensitivity analysis: Theory and estimation algorithms. In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Handbook of Uncertainty Quantification*, pages 1217–1239. Springer, 2017.
- [8] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.
- [9] A. Saltelli and I. Sobol’. Sensitivity analysis for nonlinear mathematical models: numerical experience. *Matematicheskoe Modelirovanie*, 7(11):16–28, 1995.
- [10] I. Sobol’. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiments*, 1(4):407–414, 1993.
- [11] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979, 2008. Bayesian Networks in Dependability.