The sparse Polynomial Chaos expansion: a fully Bayesian approach with joint priors on the coefficients and global selection of terms

Paul Bürkner, Ilja Kröker, Sergey Oladyshkin, Wolfgang Nowak

Function Approximation

General setup:

$$y = f(x) + e$$

with input variables $x \in \mathbb{R}^D$ and response $y \in \mathbb{R}$

Given observed data $(\tilde{y}, \tilde{x})$ find a function $f_A(x)$ with

$$f_A(x) \approx f(x)$$

Polynomial Chaos Expansion (PCE)

Polynomial approximation:

$$f_A(x) = \sum_{m=0}^{M} c_m \Psi_m(x)$$

with polynomials $\Psi_m(x)$ and coefficients c_m

In PCE $\Psi_m(x)$ are orthonormal:

$$\int \Psi_m(x)\, \Psi_{m'}(x)\, p(x)\, dx = \delta_{m,m'}$$

with $\delta_{m,m'} = 1$ if $m = m'$ and $\delta_{m,m'} = 0$ otherwise

All we need is unidimensional PCE

Assume independence of input variables x_d

If the polynomials $\psi_{d,s}(x_d)$ for are orthonormal for $p(x_d)$, then the tensor product polynomials

$$\Psi_{\alpha}(x) = \prod_{d=1}^{D} \psi_{d,\alpha_k}(x_d)$$

are orthonormal for $p(x) = \prod_{d=1}^{D} p(x_d)$

Combinatorial Explosion

Fix the maximal joint polynomial order to P

Then we have

$$M = \binom{P+D}{P} = \frac{(P+D)!}{P!D!}$$

D-variate polynomials of order P or less:

$$\Psi_\alpha(x) = \prod_{d=1}^{D} \psi_{d,\alpha_k}(x_d) \qquad \text{with} \qquad \sum_{d=1}^{D} \alpha_k \leq P$$

Examples:

- For $D = 3$ and $P = 10$ we have $M = 286$
- For $D = 6$ and $P = 8$ we have $M = 3003$

Bayesian PCE

Assume normally distributed errors $e \sim \text{normal}(0, \sigma^2)$

Then the standard Bayesian PCE model is given by:

$$
\begin{aligned}
y &\sim \text{normal}(f_A(x), \sigma^2) \\
f_A(x) &= \sum_{m=0}^{M} c_m \Psi_m(x) \\
c &\sim p(c) \\
\sigma^2 &\sim p(\sigma^2),
\end{aligned}
$$

Flexible estimation with MCMC, for example with Stan and brms

References: Carpenter et al. (2017), Bürkner (2017)

Percentage of Variance Explained

The coefficient of determination R^2 (percantage of variance explained by the model) can be written as:

$$R^2 = \frac{\mathrm{var}(f_A(x))}{\mathrm{var}(f_A(x)) + \sigma^2}$$

where the variance of the PCE approximation is

$$\mathrm{var}(f_A(x)) = \sum_{m=1}^{M} c_m^2.$$

Accordingly, a prior on R^2 implies a joint prior on the c_m

The R2D2 Prior: A global-local shrinkage prior

The R2D2 prior is specified as follows:

$$
\begin{aligned}
R^2 &\sim \text{Beta}(a_1, a_2) \\
\tau^2 &= \frac{R^2}{1 - R^2} \\
c_m &\sim \text{normal}\left(0, \sigma^2 \tau^2 \phi_m\right) \\
\phi_m &\geq 0 \text{ and } \sum_{m=1}^{M} \phi_m = 1 \\
\phi &\sim \text{Dirichlet}(\theta) \\
c_0 &\sim p(c_0) \\
\sigma^2 &\sim p(\sigma^2)
\end{aligned}
$$

Reference: Zhang et al. (2020)

(Bayesian) Variable Selection

Choose a number M_{sel} of polynomials to be selected

Option 1: Choose the polynomials with the largest Sobol indices:

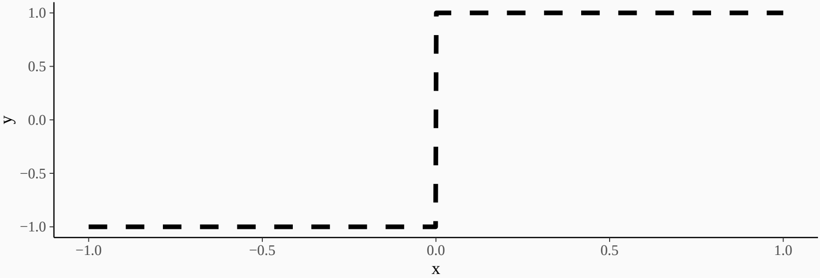
$$S_m = \frac{c_m^2}{\sum_{m'=1}^{M} c_{m'}^2},$$

Option 2 (Projpred): Choose the polynomials that imply the largest reduction in KL-divergence from the posterior predictive distribution of the full model:

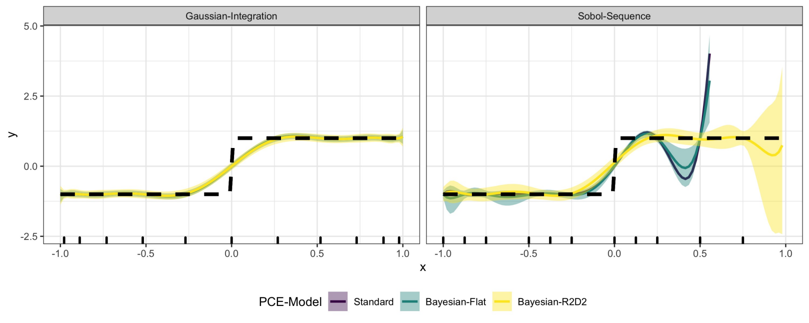
$$\text{KL}\left[p(y|\tilde{y}),\, q_{\text{sel}}(y)\right]$$

Reference: Piironen et al. (2020)

1D Case Study: Sign Function

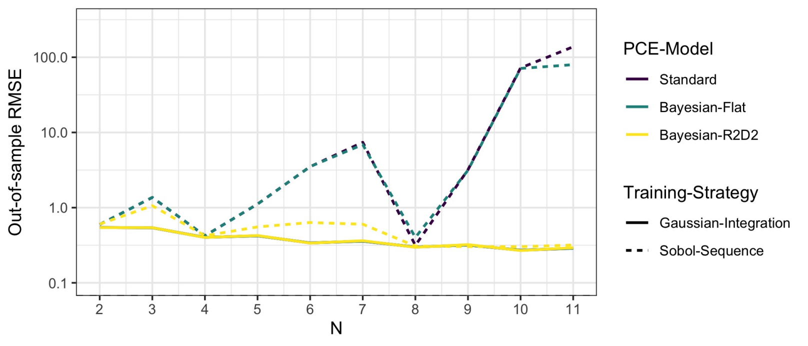


Sign Function: Conditional Predictions



Conditional predictions for different PCE models of the Signum function based on $P = 10$ polynomials and $N = 11$ training points.

Sign Function: Results Overview



Results for varying number of training points N with $P = N + 1$.

3D Case Study: Ishigami Function

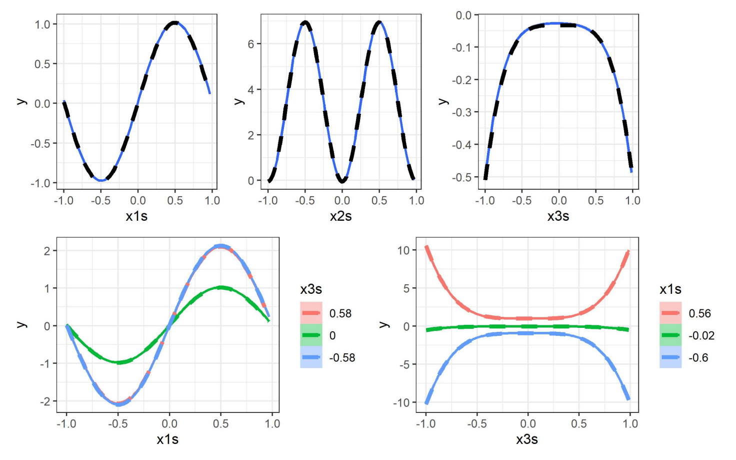
Ishigami function:

$$y = f(x) = \sin(x_1) + a\sin(x_2)^2 + bx_3^4\sin(x_1)$$

- Hyperparamerters set to $a = 7$, $b = 0.1$
- Input variables distributed as $x_1, x_2, x_3 \sim \text{uniform}(-1, 1)$
- Mean and variance known analytically

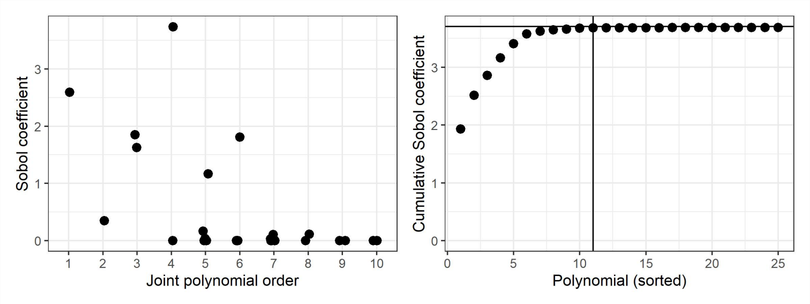
Referenece: Ishigami et al. (1990)

Ishigami Function: Conditional Predictions

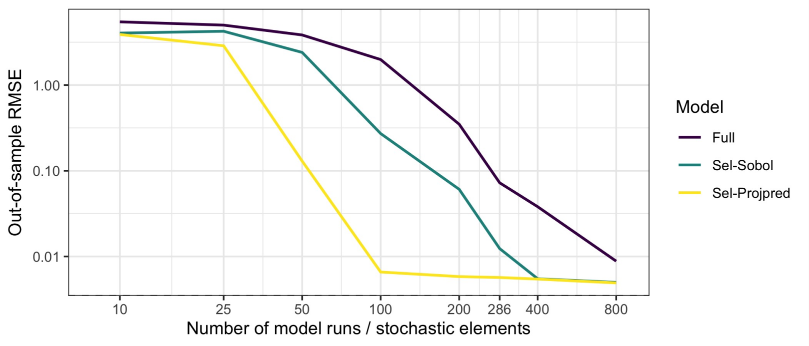


Conditional prediction for the sparse projpred model of the Ishigami function based on $N = 100$ training points and the $P_S = 25$ most important polynomials.

Ishigami Function: Sobol Indices



Posterior mean Sobol indices and total Sobol indices for the sparse projpred model on the Ishigami function based on based on $N = 100$ training points and the $P_S = 25$ most important non-constant polynomials.



Summarized results for the Ishigami function by the size of the training data and model type.

Conclusion

- PCE is an general-purpose function approximation approach
- ... but suffers heavily from the curse of dimensionality
- Sparse or regularized PCEs can help reduce this problem
- Our approach combines regularized PCE with exact sparsity
- ... and achieves highly precise results in several benchmarks

Contact details:

- Email: paul.buerkner@gmail.com
- Website: https://paul-buerkner.github.io
- Twitter: @paulbuerkner

References

Zhang, Y. D., Naughton, B. P., Bondell, H. D., & Reich, B. J. (2020). Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association*.

Piironen, J., Paasiniemi, M., & Vehtari, A. (2020). Projective inference in high-dimensional problems: Prediction and feature selection. *Electronic Journal of Statistics.*

Carpenter B., Gelman A., Hoffman M. D., Lee D., Goodrich B., Betancourt M., Brubaker M., Guo J., Li P., and Riddell A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*.

Bürkner P. C. (2017). brms: An R Package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software*.

Ishigami T. and Homma T. (1990). An importance quantification technique in uncertainty analysis for computer models. *First International Symposium on Uncertainty Modeling and Analysis*.