Mollifiers to enhance gradient-based dimension reduction

R. Verdière[†,1], C. Prieur[§,1] O. Zahm[§,1],

[†] PhD student (presenting author). [§] PhD supervisor
PhD expected duration: Sept. 2022 – Aug. 2025

[1] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK
{romain.verdiere,olivier.zahm}@inria.fr ; clementine.prieur@univ-grenoble-alpes.fr

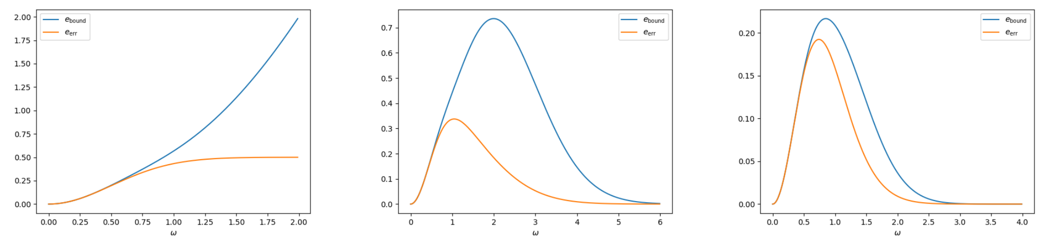
Abstract

Modern computational models for scientific and engineering applications typically involve a large number of input parameters and are expensive-to-evaluate both in time and resources. Replacing the model with an accurate and fast-to-evaluate surrogate (or approximation) offers a viable workaround in many applications. Approximating such high-dimensional functions with classical approximation tools such as polynomials, wavelets or neural networks is, however, a difficult task. This is even aggravated in the small sample regime where one only has access to a little number of model evaluations. One way to address this challenge is to reduce the input dimension beforehand. This consists in approximating the *model* $x \mapsto u(x)$ as the composition of two functions: a *feature map* $x \mapsto z = g(x)$ which extracts the relevant features of the input variables, and a *profile function* $z \mapsto f(z)$ which regresses the model output on the features. The feature map can be built by minimizing an upper bound of the reconstruction error $\min_f \mathbb{E}[(u(\boldsymbol{X}) - f \circ g(\boldsymbol{X}))^2]$ obtained with Poincaré-type inequalities. When the feature map is linear this strategy reduces to Active Subspace [4, 2]. The case of nonlinear feature maps has been studied in [1] for polynomial feature maps and in [3, 5] for diffeomorphism-based feature maps. The bound derived from Poincaré inequality is proportional to the L^2-norm of model gradients, therefore, this strategy works well for slowly varing functions for which the bound is tight. For oscillatory model with large gradient norms, however, the bound reveals too loose to build a meaningfull feature map and the method fails.

In this talk we demonstrate that working with a mollified version of the model $(u \star \rho_\sigma)$ is a good strategy to circumvent this issue as it allows to obtain sharper Poincaré error bounds and to reduce the dimension efficiently using gradient-based techniques. Here ρ_σ is the gaussian kernel with 0 mean and $\sigma^2 I_d$ covariance, $\star$ is the convolution operator and we call σ the mollifying parameter. We demonstrate that the reconstruction error when using a mollified version of the model can be bounded by the sum of two terms: one that vanishes when the mollifying parameter goes to zero and one that is proportional to the Poincaré error bound of the mollified model. This bound shows the trade-off between mollification and dimension reduction: for strongly mollified models the first term is large and the second one quite small and the other way around when the model is less mollified. Based on this result, we propose an iterative algorithm for dimension reduction. More precisely, we introduce a sequence $\sigma_1 > \sigma_2 > \ldots > \sigma_p \geq 0$ of decreasing mollifying parameters. Then at the first iteration we approximate a strongly mollified version of the model $u_1^* = u \star \rho_{\sigma_1}$ with a feature map g_1 and a profile function f_1. At the next interation the algorithm approximates a slightly less mollified version of the residual model $u_2^* = (u - f_1 \circ g_1) \star \rho_{\sigma_2}$ with a feature map g_2 and a profile function f_2. This proccess iterates p times and at the end the original model u is approximated by $\sum_{1 \leq i \leq p} f_i \circ g_i$.

Let us illustrate on some example the impact of the mollification step on the accuracy of Poincaré error bound. We consider the analytical toy model $u(x) = \sum_{i=1}^{d} a_i \sin(\omega_i x_i)$, where a_i, ω_i, x_i are

respectivelly the ith components of vectors $a \in \mathbb{R}^d, \omega \in \mathbb{R}^d, x \in \mathbb{R}^d$. We aim at approximating u by $f \circ g$ with g a projector onto $\{e_1, ..., e_d\}$ the canonical basis of $\mathbb{R}^d$. Here g is a linear feature map and $g = U^\top \sum_{i \in \tau} e_i e_i^\top$ where $\tau \subset \{1, \ldots, d\}$, $\#\tau = m$ and $U = [e_i]_{i\in\tau} \in \mathbb{R}^{d\times m}$. In this framework, and for $\boldsymbol{X} \sim \mathcal{N}(0, I_d)$, we compare the minimal reconstruction error for $u_\sigma = u * \rho_\sigma$ with the one obtained by minimizing Poincaré error bound. We perform the comparison for $a_i = 1$, $i = 1, ..., d$ and for different values of σ. In this situation the reconstruction error is equal to $\frac{1}{2}\sum_{i\in-\tau} e^{-\omega_i^2\sigma^2}(1 - e^{-2\omega_i^2})$ and the Poincaré error bound is equal to $\frac{1}{2}\sum_{i\in-\tau} \omega_i^2 e^{-\omega_i^2\sigma^2}(1 + e^{-2\omega_i^2})$, where $-\tau$ is the complementary set of τ in $\{1, \ldots, d\}$. We can compare the 2 functions $e_{\text{err}}(\omega) = \frac{1}{2}e^{-\omega^2\sigma^2}(1 - e^{-2\omega^2})$ and $e_{\text{bound}}(\omega) = \frac{1}{2}\omega^2 e^{-\omega^2\sigma^2}(1 + e^{-2\omega^2})$ to understand how the error and the bound behave for different values of the mollifying parameter σ. Figure 1 clearly shows that the reconstruction error and the error bound become closer together as the value of σ grows.



(a) $\sigma = 0$ (b) $\sigma = 0.5$ (c) $\sigma = 1$

Figure 1: Plots of e_{err} and e_{bound} according to ω for different values of σ

Short biography

I am a former student of ENSTA Paris where I studied mathematics and computer science. At the end of my engineering studies, to deepen my knowledge in general mathematics, I prepared for and passed mathematical aggregation. Finally, after a short experience as a teacher, I moved to Grenoble to work as PhD student with Clémentine Prieur and Olivier Zahm on nonlinear dimension reduction for function approximation. This thesis is funded by the French Research Agency (ANR).

References

[1] Daniele Bigoni, Youssef Marzouk, Clémentine Prieur, and Olivier Zahm. Nonlinear dimension reduction for surrogate modeling using gradient information. *Information and Inference: A Journal of the IMA*, 11(4):1597–1639, 2022.

[2] Paul G. Constantine, Eric Dow, and Qiqi Wang. Active subspace methods in theory and practice: Applications to kriging surfaces. *SIAM Journal on Scientific Computing*, 36(4):A1500–A1524, January 2014.

[3] Romain Verdière, Clémentine Prieur, and Olivier Zahm. Diffeomorphism-based feature learning using Poincaré inequalities on augmented input space. working paper or preprint, December 2023.

[4] Olivier Zahm, Paul Constantine, Clémentine Prieur, and Youssef Marzouk. Gradient-based dimension reduction of multivariate vector-valued functions. 2019.

[5] Guannan Zhang, Jiaxin Zhang, and Jacob Hinkle. Learning nonlinear level sets for dimensionality reduction in function approximation. *Advances in Neural Information Processing Systems*, 32, 2019.