MascotNum2017 conference - A specific kriging kernel for dimensionality reduction: Isotropic by group kernel

M. RIBAUD
ICJ, Centrale Lyon

Supervisor(s): C. Vial (ICJ, université Claude Bernard Lyon 1), C. Helbert (ICJ, Centrale Lyon)

Ph.D. expected duration: 2015-2018

Adress: Centrale Lyon, 36 Avenue Guy de Collongue, 69134 Ecully

Email: melina.ribaud@doctorant.ec-lyon.fr

Abstract:

In this work, we focus on kriging based metamodel and more precisely we propose a new covariance kernel adapted to high dimension problem. In general, people use an *anisotropic* kernel that is a tensor product of as many 1D kernels as the number of inputs, each kernel being parameterized by a spatial correlation length, called range parameter. In high dimension with a very restricted number of data points the estimation of range parameters becomes quite difficult. That's why in this context, an *isotropic kernel* could be a good alternative. This kernel is a function of the euclidean distance in the entire input space and depends on only one range parameter. However *isotropic* kernel are too restrictive. The idea of this paper is to construct a kernel between this two extremal choices. A review of the literature shows that one way to improve performance of kriging is to adapt the covariance structure to each specific case see e.g. Durrande et al. ([2]) and Padonou et al. ([5]). In general the covariance function only depends on the range parameters and on the sill parameter (variance) that have to be estimated. Data points range parameters estimation becomes very tricky in high dimension. Some answers could be found in the literature. Binois ([1]) propose different sparse and specific covariance kernels based on expert information or on the work of Muehlenstaedt et al. ([4]) and Durrande et al. ([3]). This construction is not possible in high dimension. Moreover, we can also find in literature the work of Welch et al ([[7]]) that proposes a forward strategy that aims at identifying the active factors. However, the proposed kernel is too separator and can be improved. In this paper, we propose a data-driven construction of a covariance kernel adapted to high dimension. This kernel is a tensor product of a few *isotropic* kernels built on well-chosen subgroups of variables. It has the following expression:

$$r_{\boldsymbol{\theta}}(\mathbf{x}-\mathbf{x}') = \prod_{\ell=1}^{q} \rho_{\theta_\ell}\left(\|\mathbf{x}-\mathbf{x}'\|_{\mathcal{I}_\ell}\right) \tag{1}$$

where, $q < p$ is the number of groups and $\theta \in \mathbb{R}^q$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$. Let $\mathcal{I}_\ell$ be the set of indexes in the ℓth group with cardinal $|\mathcal{I}_\ell| = p_\ell$, $p_\ell = 1, \ldots, p$. Then we have $p_1 + \ldots + p_q = p$. We note the norm of the subvector $(x_j)_{j\in\mathcal{I}_\ell}$ by $\|\mathbf{x}\|^2_{\mathcal{I}_\ell} = \left(\sum_{j\in\mathcal{I}_\ell} x_j^2\right)$ $\forall \mathbf{x} \in [0;1]^p$. The difficulty is to find the number and the composition of groups. The idea is to start the algorithm with a full *isotropic* kernel. It depends on only one parameter to gain in robustness in high dimensional problems. In the procedures that we introduce, the final kernel is an *isotropic by group*. Then, at each stage we compare different models and we choose the best under a BIC criterion. We test 4 different algorithms and "algW" that is described in ([7]).

Algorithms	alg1	alg2	alg3	alg4	algW
Strategy	Quasi exhaustive	Separate	Separate + Cluster at the end	Separate + Cluster at each step	Separate
Direction	Forward	Both	Both	Both	Forward

The table summarizes the procedures characteristics. We apply the algorithms to a simulated case. We study the algorithms behavior with learning set size and compare their prediction quality and their time of estimation. It results that algorithm 4 is the most efficient. The estimation time does not grow to fast with the number of experiments and the quality of prediction is always the best. The prediction quality increases with the number of experiments. Finally, we consider the Sobol function composed of 6 active variables and 9 non-active ones. This function is popular in sensitivity analysis see Statelli et al. ([6]). We compare the predictive power of the three models, the kriging model with the *isotropic* kernel is poor. kriging with an *isotropic by group* kernel improves prediction power compared to kriging with an *anisotropic* kernel.

References

[1] Mickael Binois. *Uncertainty quantification on pareto fronts and high-dimensional strategies in bayesian optimization, with applications in multi-objective automotive design.* PhD thesis, Saint-Etienne, EMSE, 2015.

[2] Nicolas Durrande. *Étude de classes de noyaux adaptées à la simplification et à l'interprétation des modèles d'approximation. Une approche fonctionnelle et probabiliste.* PhD thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne, 2001.

[3] Nicolas Durrande, David Ginsbourger, and Olivier Roustant. Additive kernels for gaussian process modeling. *arXiv preprint arXiv:1103.4023*, 2011.

[4] Thomas Muehlenstaedt, Olivier Roustant, Laurent Carraro, and Sonja Kuhnt. Data-driven kriging models based on fanova-decomposition. *Statistics and Computing*, 22(3):723–738, 2012.

[5] Espéran Padonou and O Roustant. Polar gaussian processes and experimental designs in circular domains. 2016.

[6] Andrea Saltelli, Karen Chan, E Marian Scott, et al. *Sensitivity analysis*, volume 1. Wiley New York, 2000.

[7] William J Welch, Robert J Buck, Jerome Sacks, Henry P Wynn, Toby J Mitchell, and Max D Morris. Screening, predicting, and computer experiments. *Technometrics*, 34(1):15–25, 1992.

Short biography – I have studied applied mathematics during my engineering school and my research master. The main subject of my thesis is metamodeling and heuristic optimization. The funding of my thesis is given by the French National Research Agency (ANR PEPITO fund). We work on computer experiments and optimization for the Industry of transportation. The project lead is Valeo and we work with other companies and academics partner.