Consortium Industrie Recherche pour l'Optimisation et la QUantification d'incertitude pour les données Onéreuses

*POST-DOCTORAL PROPOSAL*

# *Réduction de dimension semi-supervisée*

**INRIA Grenoble**

Miguel MUNOZ ZUNIGA*,
Thomas PERRILLAT-BOTTONET†,

Olivier ZAHM‡

Dans le cadre de l'etude et de l'utilisation de codes de calcul, un besoin récurrent concerne la réduction de dimension supervisée des entrées du code afin de réduire les besoins computationnels et améliorer la qualité des analyses. En effet, déterminer les entrées pertinentes à l'explication des sorties d'intérêts permet notamment de réduire la taille des plans d'expériences nécessaire aux analyses (réduction du besoin en simulations couteuses), formuler des analyses de sensibilité sur des entrées influentes, construire des méta-modèles plus efficacement et de meilleure qualité, ou encore de donner de meilleurs taux de convergence des algorithmes d'optimisation.

Dans la littérature il est récurrent d'utiliser plutôt des stratégies non-supervisées (PCA, KL) pour réduire la dimension des entrées avec potentiellement dans un sec- ond temps des analyses de sensibilités comme complément supervisé. En effet l'approche non-supervisé a l'avantage de ne pas nécessiter de simulation (approche *a priori* ). Les approches supervisées peuvent nécessiter un nombre de simulations relativement conséquent pour obtenir des résultats pertinents ce qui n'est pas toujours accessible en début d'analyse.

On peut citer par exemple les cas où les simulations arrivent de manière séquentielle : assimilation de données en temps réelle, enrichissement adaptatif d'un plan d'expériences.

Dans ce contexte, nous proposons une approche semi-supervisée dans laquelle un compromis entre réduction de dimension non-supervisée et supervisée est mise en place en fonction du nombre courant de simulations. Cette approche pourra par exemple se formuler comme un problème d'optimisation du type

$$\min_{B\in\mathbb{R}^{d\times m}} \left( \min_{\substack{f:\mathbb{R}^m\to\mathbb{R}\\ \text{any function}}} \mathbb{E}\left[(Y-f(B^TX))^2\right] + \lambda \min_{\substack{h:\mathbb{R}^m\to\mathbb{R}^d\\ \text{affine function}}} \mathbb{E}\left[\|X-h(B^TX)\|^2\right] \right) \quad (1)$$

avec $\lambda > 0$ un paramètre de régularisation qu'il conviendra de calibrer en fonction du

nombre d'évaluation du modèle $Y = u(X)$ disponible. Alors que l'optimisation sur $h$ peut être réalisé analytiquement (PCA), l'optimisation sur $f$ est plus délicate. Nous envisageons d'utiliser des approches à noyau [1,2] ou bien des approches qui utilisent les gradients du modèle $\nabla u(X)$ [3,4,5].

*miguel.munoz-zuniga@ifpen.fr
†thomas.perrillat-bottonet@cea.fr

‡olivier.zahm@inria.fr