

DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **09 juillet 2025**

Nom de famille et prénom de l'auteur. e : **Monsieur Luca CASTELLI**

Titre de la thèse : Inférence statistiques pour la régression PLS et ses variantes

Résumé



Cette thèse est consacrée aux propriétés statistiques de l'estimateur de régression des moindres carrés partiels (PLS). La régression PLS est une technique de réduction de la dimension connue pour traiter les cas de haute dimension et de multicollinéarité. Cette méthode projette les covariables sur un sous-espace bien choisi, en considérant les corrélations successives avec la variable à expliquer, dans le but d'améliorer la qualité prédictive. Nous décrivons les différentes formes d'algorithmes PLS, en détaillant leurs propriétés mathématiques et en soulignant le lien avec la méthode du gradient conjugué. Par ailleurs, les propriétés algébriques du sous-espace calculé seront analysées et approfondies dans notre contexte. Certains liens avec la structure des valeurs propres de la matrice de covariance seront également établis. Nous commençons par fournir une borne supérieure non asymptotique pour l'estimateur PLS sur la perte quadratique dans le cas d'une seule composante. Nous étendons ces résultats dans un contexte parcimonieux pour la sparse PLS (sPLS) où les bornes obtenues sont similaires à celles obtenues pour l'algorithme Lasso. Nous généralisons ensuite la borne donnée au cas de l'estimateur PLS avec K composantes. La borne est obtenue grâce à une hypothèse sur la norme des composantes PLS de Krylov. Nous assouplissons cette hypothèse en introduisant une étape de régularisation de Ridge dans l'estimateur PLS afin d'obtenir des bornes similaires. Enfin, nous nous concentrerons sur l'estimation des projections aléatoires. Ce travail découle de l'intérêt porté à la régression PLS, où le sous-espace calculé est aléatoire et représente une estimation d'une version théorique. Nous fournissons une borne supérieure non asymptotique sur la norme de l'opérateur de la différence entre l'opérateur de projection et son estimation.

Mots-clés : régression, moindres carrés partiels, parcimonie, Grande dimension, régularisation Ridge