
Statistique en grande dimension: en quête des méthodes optimales

Alexandre Tsybakov*¹

¹Centre de Recherche en Économie et Statistique (CREST) – INSEE, École Nationale de la Statistique et de l'Administration Économique – France

Résumé

Le but de cet exposé est de présenter quelques principes fondateurs qui émergent dans l'analyse des problèmes statistiques en grande dimension. Ces principes sont communs à de nombreux problèmes apparus récemment, tels que la régression linéaire sparse, l'estimation de grandes matrices sparses, la complétion de matrices, l'apprentissage de dictionnaires ainsi que les modèles de réseaux, par exemple, les modèles stochastiques à blocs (stochastic block models) ou à participation mixte (mixed membership models). L'accent sera mis sur l'optimalité des méthodes au sens minimax non-asymptotique ainsi que la prise en compte d'erreur d'approximation par l'intermédiaire d'inégalités d'oracles exactes. Ces résultats sont obtenus grâce à des outils universels tel que l'agrégation. Le décalage entre les vitesses de convergence des méthodes réalisables en temps polynomial et les vitesses optimales au sens minimax donne lieu à une nouvelle classe de problèmes, pour lesquels on ne dispose aujourd'hui que de solutions partielles.

*Intervenant