Statistical learning for structured large dimensional data

Advisors: ***Florent Chatelain***†, Pierre Comon†, Romain Couillet† and Nicolas Le Bihan†.

†: GIPSA-Lab, Grenoble, France.

Email : `firstname.lastname@gipsa-lab.fr`

Motivation. In many statistical learning applications to engineering, data (or signals) have specific features: their being complex- or quaternionic-valued vectors, their assuming a tensor (rather than vector or matrix) form, or their resulting from projective amplitude observations of a phased complex vector. Specific applications of these scenarios are found in wireless communications, seismology, astrophysics, etc.

The recent advances in large dimensional statistics, and particularly in random matrix theory, have largely generalized the field of asymptotic statistics by assuming, not only that the data are numerous ($n \to \infty$) but also that their size is large ($p \to \infty$). The double asymptotics $n, p \to \infty$ disrupts with past paradigms and brings forward a new set of tools and algorithms to deal with large dimensional data.

Yet, conventional studies in random matrix theory and large dimensional statistics at large are mostly focused on *linear real (or complex circularly symmetric) operators and data models.* This may be largely limiting from an engineering perspective when data are intrinsically modelled in a more intricate (non linear, multi-modal, complex non circular) manner.

PhD description. The objective of the PhD is to extend the recent theoretical findings of large dimensional statistics, primarily in the context of statistical detection and estimation, for a wider penetration of these ideas into practical engineering data sciences. Several directions will be explored:

1. the asymptotic (large n, p) analysis of signal detection procedures (i.e., test statistics such as likelihood or generalized likelihood ratio tests) for non-circularly symmetric complex and simplectic (quaternionic) data and for tensor data models (in the large $n_1, \ldots, n_k$ dimensions of order-k tensors);
2. the further generalization of these procedures to statistical estimation, starting with principle component analysis in large dimensions;
3. the asymptotic analysis of spectral phase-retrieval methods, whereby the signals observed are the amplitudes $a_i = |w_i^* x|^2$ of projected phased-vectors $x \in \mathbb{C}^p$ onto (random) directions w_i.

Applications of these methods and proposed new algorithms to estimation, detection, properness testing, filtering and decomposition/denoising will be performed both on synthetic and real applied datasets.

Organization. The position is held within the GIPSA-lab at the University Grenoble-Alpes, as part of the MIAI chair on “Large Dimensional Statistics for AI” (LargeDATA). The LargeDATA chair develops expertise in large dimensional statistics for AI, notably focusing on random matrix theory, statistical physics and graphs.

The PhD will be advised by **Florent Chatelain**, assistant professor at Grenoble INP, expert in statistics and signal processing, and co-advised by Romain Couillet (professor, head of MIAI chair), Pierre Comon (research director at CNRS) and Nicolas Le Bihan (research director at CNRS).

Expected Outcomes. Results will be published in international journals (IEEE Transactions on Signal Processing, Journal of Machine Learning Research, and equivalent) and presented in international conferences (NeurIPS, ICML, IEEE ICASSP, MLSP and SSP).

Profile. The candidate should hold a MSc in statistical signal processing/applied mathematics/data science. Elementary notions of large dimensional statistics (possibly random matrix theory) is a plus.