High-Dimensional Statistical Analysis: Non-Sparse Modeling, Geometric Representations and New PCAs

Makoto Aoshima

University of Tsukuba, Japan

Abstract

High-dimensional data is often much more non-sparse than you imagine. Non-sparsity about intrinsic/necessary information is useful to improve the accuracy of inference. However, non-sparsity about noise/unnecessary information is troublesome for inference because the intrinsic information is covered with huge noise and the accuracy of inference of the necessary information is heavily damaged by the huge noise. In this talk, I emphasize how non-sparse modeling is essential in high-dimensional statistical analysis. My messages are as follows: (1) Analyze the pattern of high-dimensional data. Key tools are the dual space and the geometric representation of high-dimensional data. (2) As for handling huge intrinsic information, powerful tools are new PCAs by Yata and Aoshima: the noise-reduction methodology (J. Multivariate Anal., 2012; Electron. J. Stat., 2016) and the cross-data-matrix methodology (J. Multivariate Anal., 2010). (3) As for handling huge noise, additional new techniques are required to ensure high accuracy of inference. In this talk, I will introduce the new PCAs and non-sparse modeling of inference on multiclass mean vectors, correlation tests and classification problems.

[The talk is based on joint work with Kazuyoshi Yata (University of Tsukuba).]