Towards a sparse, scalable, and stably positive definite (inverse) covariance estimator

Joong-Ho (Johann) Won
Department of Statistics, Seoul National University

Abstract: High-dimensional covariance estimation and graphical model selection is a contemporary topic in statistics and machine learning, and has widespread applications. The problem is notoriously difficult in high dimensions as the traditional estimate is not even positive definite, let alone sufficiently stable. An important line of research is to shrink the spectrum to yield stable well-conditioned estimators. A separate line of research has considered sparse estimation using nonsmooth regularization methods and provides interpretable models with fewer parameters. Though an estimator which is both stable and sparse is often desirable in numerous downstream applications, obtaining such estimators is inherently challenging in modern high-dimensional regimes due to the very different nature of the two approaches. In this talk we propose a unifying and scalable framework which addresses this problem. Our general methodology takes an arbitrary covariance loss functions (such as the ones which have been proposed in the literature) and yields estimates that are both spectrally regularized and sparse. The framework leads to an enriched class of estimators which are computationally tractable and enjoy good asymptotic properties. In addition, when the covariance loss function is orthogonally invariant, we further demonstrate that a solution path algorithm can be derived, involving a series of ordinary differential equations. The path algorithm is attractive because it provides the entire family of estimates for all possible values of the regularization parameter, at the same computational cost of a single estimate with a fixed parameter. An important finding is

that an iterative path algorithm can be devised even when the loss function is not orthogonally invariant, utilizing modern operator splitting techniques. We illustrate the efficacy of our approach on both real and simulated data.

This is a collaboration with Sang-Yun Oh (UC Santa Barbara) and Bala Rajaratnam (UC Davis).