Recent computational advances in multivariate statistics with applications to target recognition and genomics

ABSTRACT: Multivariate statistical analysis is concerned with observations on several variables which are thought to possess some degree of interdependence. Driven by problems in genetics, social sciences, image processing, and wireless communication, one central computational problem in multivariate statistical testing is to efficiently compute the distributions of eigenvalues of random covariance matrices.

While these distributions are available explicitly as (extremely slowly converging) series of Schur and Jack functions, the computational efforts had been hampered by the fact that a straightforward evaluation of even a single Schur or Jack function takes exponential time.

The rich mathematical structure of these functions stemming from combinatorics and representation theory had apparently been missed for over 40 years.

Recently we developed new algorithms that exploit these very properties of the Schur and Jack functions to compute them extremely efficiently, thus solving a major open problem in computational multivariate statistics. Tasks that used to take weeks now take fractions of a second.

I will present the key algebraic ideas behind our algorithms (as well as certain remaining challenges), including connections with the representations of the general linear group and Fast Fourier Transforms. I will describe the new applications that have been enabled, particularly in Genomics and 3D Target Classification.