**NOUREDDINE EL KAROUI**, Department of Statistics, UC Berkeley, 367 Evans Hall, Berkeley, CA 94720-3860, USA
*On the spectral properties of large-dimensional kernel random matrices*

Prompted by the recent explosion of the size of datasets statisticians are working with, there is currently renewed interest in the statistics literature for questions concerning the spectral properties of large-dimensional random matrices.

Most of the efforts so far has focused on understanding sample covariance matrices in the "large $n$, large $p$" setting, where the number of samples, $n$, and the number of variables in the problem, $p$, are of the same magnitude—say a few 10s or 100s. A basic message is that in this high-dimensional setting, the sample covariance matrix is a very poor estimator of the population covariance, especially from a spectral point of view. This naturally raises questions about the behavior of spectral methods of multivariate analysis such as the widely used principal component analysis (PCA).

The statistical learning literature has a number of "kernel analogs" to classical multivariate techniques, in which the sample covariance matrix is replaced by a kernel matrix, often times of the form $f(X_i'X_j)$ or $f(\|X_i - X_j\|)$. It is therefore natural to ask what can be said about the spectral properties of these kernel random matrices, and in particular, what is the impact of the choice of the function $f$ and related questions.

In this talk I will discuss these questions when:

(a) the data is assumed to be genuinely high-dimensional, and

(b) when it is a noisy version of data sampled from a low-dimensional structure and the noise is high-dimensional.

In both cases, we will see that standard heuristics can lead us astray and that careful analysis yields perhaps surprising results.