

# Stanford University

## Department of Statistics

### DEPARTMENTAL SEMINAR

4:30pm, Tuesday, January 28, 2020  
Sloan Mathematics Center Room 380C

Refreshments served at 4pm in Sequoia Lounge.

**Speaker:** Didong Li, *Duke University*

**Title:** **Efficient Manifold and Density Estimation with Spherelets**

**Abstract:**

Data lying in a high-dimensional ambient space are commonly thought to have a much lower intrinsic dimension. In particular, the data may be concentrated near a lower-dimensional subspace or manifold. There is an immense literature focused on approximating the unknown subspace and the unknown density, and exploiting such approximations in clustering, data compression, and building of predictive models. Most of the literature relies on approximating subspaces and densities using a locally linear, and potentially multiscale, dictionary with Gaussian kernels. In this talk, we propose a simple and general alternative, which instead uses pieces of spheres, or spherelets, to locally approximate the unknown subspace. I will also introduce a curved kernel called the the Fisher–Gaussian (FG) kernel which outperforms multivariate Gaussians in many cases. Theory is developed showing that spherelets can produce lower covering numbers and mean square errors for many manifolds, as well as the posterior consistency of the Dirichlet process mixture of FG kernels. Results relative to state-of-the-art competitors show gains in ability to accurately approximate the subspace and the density with fewer components and parameters. Time permitting, I will also present some applications of spherelets, including classification, geodesic distance estimation and clustering.