## EMMANUEL CANDES
Department of Applied and Computational Mathematics
California Institute of Technology

## The Dantzig Selector: statistical estimation when $p$ is larger than $n$

### MONDAY, May 14, 2007 at 4:00 PM
### 133 Eckhart Hall, 5734 S. University Avenue
*Refreshments following the seminar in Eckhart 110.*

## ABSTRACT

In many important statistical applications, the number of variables or parameters is much larger than the number of observations. In radiology and biomedical imaging for instance, one is typically able to collect far fewer measurements about an image of interest than the unknown number of pixels. Examples in functional MRI and tomography immediately come to mind. Other examples of high-dimensional data in genomics, signal processing and many other fields abound. In the context of multiple linear regression for instance, a fundamental question is whether it is possible to estimate a vector of parameters of size $p$ from a vector of observations of size $n$ when $n \ll p$. This seems a priori hopeless.

This talk introduces a new estimator, dubbed the "Dantzig selector" in honor of the late George Dantzig as it invokes linear programming, and which enjoys remarkable statistical properties. Suppose that the data or design matrix obeys a uniform uncertainty principle and that the true parameter vector is sufficiently sparse or compressible which roughly guarantees that the model is identifiable. Then the estimator achieves an accuracy which nearly equals that one would achieve with an oracle that would supply perfect information about which coordinates of the unknown parameter vector are nonzero and which were above the noise level. Our results connect with the important model selection problem. In effect, the Dantzig Selector automatically selects the subset of covariates with nearly the best predictive power, by solving a convenient linear program.

The results are parts of a larger body of work perhaps best known as "Compressive Sampling" or "Compressed Sensing". If time allows, I will discuss connections with other fields such as signal processing and coding theory.