The Annals of Statistics
2007, Vol. 35, No. 6, 2352–2357
DOI: 10.1214/009053607000000424
Main article DOI: 10.1214/009053606000001523

DISCUSSION: THE DANTZIG SELECTOR: STATISTICAL ESTIMATION WHEN p IS MUCH LARGER THAN n[1]

BY PETER J. BICKEL

University of California, Berkeley

1. A personal comparative review. Reading this very interesting paper prompted me to review the current literature on the Lasso and sparsity and realize that there are at least three different sets of conditions on the behavior of the predictors $\mathbf{X}_1, \dots, \mathbf{X}_p$ under which the coefficients of the sparsest representation of a general regression model are recovered with an l_2 error of order $\sqrt{\frac{s}{n}\log p}$, where s is the dimension of the sparsest model. These are, respectively, the conditions of this paper using the Dantzig selector and those of Bunea, Tsybakov and Wegkamp [2] and Meinshausen and Yu [9] using the Lasso. Strictly speaking, Bunea, Tsybakov and Wegkamp consider only prediction, not l_2 loss, but in a paper in preparation with Ritov and Tsybakov we show that the spirit of their conditions is applicable for l_2 loss as well. Since these authors emphasize different points and use different normalizations, I thought it would be useful to present them together. Write the model as

$$Y = X_{n\times p}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where for simplicity we take $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n)$, a case falling under all the authors' conditions, and

$$X = (\mathbf{X}_1, \dots, \mathbf{X}_p).$$

We begin by assuming that

$$X\boldsymbol{\beta} \in [\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_s}] \equiv V,$$

an unknown unique s-dimensional linear subspace of $[\mathbf{X}_1, \dots, \mathbf{X}_p]$ and no lower-dimensional subspace. For simplicity, we follow Knight and Fu [8] and Meinshausen and Yu [9] and assume $|\mathbf{X}_j|^2 = n$, $1 \le j \le p$, which puts the problem on the same scale as the familiar: X a matrix of n i.i.d. p vectors.

In its current form, a goal of all three authors is to state conditions on X under which, for the method of estimation they propose,

$$|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\circ| = O_p\left(\sqrt{\frac{s}{n}\log p}\right), \tag{1}$$

Received February 2007.

[1]Supported in part by NSF Grant DMS-06-05236.