The Annals of Statistics
2007, Vol. 35, No. 6, 2385–2391
DOI: 10.1214/009053607000000479
Main article DOI: 10.1214/009053606000001523

DISCUSSION: THE DANTZIG SELECTOR: STATISTICAL ESTIMATION WHEN *p* IS MUCH LARGER THAN *n*

BY MICHAEL P. FRIEDLANDER AND MICHAEL A. SAUNDERS

University of British Columbia and Stanford University

1. Computational considerations. When Lasso [11] was proposed, it was a computational challenge to solve the associated quadratic program

$$\text{Lasso}(t) \qquad \min_{\beta} \tfrac{1}{2}\|y - X\beta\|_2^2 \quad \text{s.t.} \quad \|\beta\|_1 \le t$$

given just a single parameter t. Two active-set methods were described in [11], with some concern about efficiency if p were large, where X is $n \times p$. Later when basis pursuit de-noising (BPDN) was introduced [2], the intention was to deal with p very large and to allow X to be a sparse matrix or a fast operator. A primal–dual interior method was used to solve the associated quadratic program, but it remained a challenge to deal with a single parameter.

The authors' new Dantzig Selector (DS) also assumes a specific parameter. It is helpful to state the BPDN and DS models together:

$$\text{BPDN}(\lambda) \qquad \min_{\beta,r} \lambda\|\beta\|_1 + \tfrac{1}{2}\|r\|_2^2 \quad \text{s.t.} \quad r = y - X\beta,$$

$$\text{DS}(\lambda) \qquad \min_{\beta,r} \|\beta\|_1 \quad \text{s.t.} \quad \|X^T r\|_\infty \le \lambda, \qquad r = y - X\beta.$$

For reference purposes we also state the corresponding dual problems:

$$\text{BPdual}(\lambda) \qquad \min_{r} -y^T r + \tfrac{1}{2}\|r\|_2^2 \quad \text{s.t.} \quad \|X^T r\|_\infty \le \lambda,$$

$$\text{DSdual}(\lambda) \qquad \min_{r,z} -y^T r + \lambda\|z\|_1 \quad \text{s.t.} \quad \|X^T r\|_\infty \le \lambda, \qquad r = Xz.$$

We congratulate the authors on justifying their Dantzig Selector on detailed statistical grounds while also investigating a primal–dual interior method suitable for a sparse or fast-operator X and making codes available through ℓ_1-magic [1]. The attraction of a pure linear programming (LP) formulation is understandable. Our aim here is to help explore the prospects for both interior and simplex implementations of DS, and to compare with BPDN.

The vectors $r = y - X\beta$ and $s = -X^T r$ are used often below.

We now know that the Homotopy [5, 8, 7] and LARS [6] algorithms can solve BPDN(λ) for all $\lambda \ge 0$, and their active-set continuation approaches are remarkably efficient if the computed β remains sufficiently sparse. Nevertheless, most

Received January 2007.