The Annals of Statistics
2008, Vol. 36, No. 4, 1534–1541
DOI: 10.1214/07-AOS0316A
Main article DOI: 10.1214/009053607000000802

DISCUSSION: ONE-STEP SPARSE ESTIMATES IN NONCONCAVE PENALIZED LIKELIHOOD MODELS

BY PETER BÜHLMANN AND LUKAS MEIER

ETH Zürich

Hui Zou and Runze Li ought to be congratulated for their nice and interesting work which presents a variety of ideas and insights in statistical methodology, computing and asymptotics.

We agree with them that one- or even multi-step (or -stage) procedures are currently among the best for analyzing complex data-sets. The focus of our discussion is mainly on high-dimensional problems where $p \gg n$: we will illustrate, empirically and by describing some theory, that many of the ideas from the current paper are very useful for the $p \gg n$ setting as well.

1. Nonconvex objective function and multi-step convex optimization. The paper demonstrates a nice, and in a sense surprising, connection between difficult nonconvex optimization and computationally efficient Lasso-type methodology which involves one- (or multi-) step convex optimization. The SCAD-penalty function [5] has been often criticized from a computational point of view as it corresponds to a nonconvex objective function which is difficult to minimize; mainly in situations with many covariates, optimizing SCAD-penalized likelihood becomes an awkward task.

The usual way to optimize a SCAD-penalized likelihood is to use a local quadratic approximation. Zou and Li show here what happens if one uses a local *linear* approximation instead. In 2001, when Fan and Li [5] proposed the SCAD-penalty, it was probably easier to work with a quadratic approximation. Nowadays, and because of the contribution of the current paper, a local linear approximation seems as easy to use, thanks to the homotopy method [12] and the LARS algorithm [4]. While the latter is suited for linear models, more sophisticated algorithms have been proposed for generalized linear models; cf. [6, 8, 13].

In addition, and importantly, the local linear approximation yields sparse model fits where quite a few or even many of the coefficients in a linear or generalized linear model are zero, that is, the method does variable selection. From this point of view, the local linear approximation is often to be preferred. In fact, it closely corresponds to the adaptive Lasso [17] which is, in our view, very useful for variable selection with Lasso-type technology. The rigorous convergence results in Section 2.3 of the paper, with a nice ascent property as for the EM-algorithm, are

Received November 2007; revised November 2007.