Journal of Classification 14:331-333 (1997)

Peter Verboon, *A Robust Approach to Nonlinear Multivariate Analysis,* Leiden: DSWO, 1994, pp. 189.

Many problems in data analysis can be formulated in terms of minimizing a function of *residuals*. Residuals measure deviation from the ideal situation, in which situation they vanish. In most cases, the residuals are a function of unknown *parameters*. Observe that we have not stipulated that there is precisely one residual per observation, or that residuals are deviations between observed and expected values. This is true in some techniques, and not in others. For our purposes, residuals are merely a bunch of quantities $r_i(\theta)$, depending on the parameter θ, which we would like to be small. The job of the statistician is to choose θ in such a way that the residuals are as small as possible. In the *optimal scaling* approach to multivariate analysis, which is sometimes described as *nonlinear multivariate analysis*, the parameters include transformations of the variables.

If there was only a single residual, most people would agree on the definition of what is "small." It is given by the modulus of the number. But if we have many residuals, we need some way of combining them into a single number, which we then can proceed to call "small" or "large." Classical recipes of making these combinations are the sum of squared residuals (Gauss-Legendre), the sum of absolute deviations (Boskovitch, Laplace), and (in approximation theory) the largest absolute deviation (Chebyshev). In robust statistics, which emphasizes outliers and distributions with heavy tails, combination rules usually downweight large residuals before combining them additively. This leads to a class of combination rules of the form

$$L(\theta) = \sum_{i=1}^{n} \phi(r_i(\theta)),$$

where ϕ is some function reweighting the residuals. Estimators of θ minimizing a loss function of this form are called *M-estimators*, at least in the context of location and regression. Usually $\phi(r) \geq 0$, and $\phi(0) = 0$. In many cases ϕ is chosen to be *redescending*, which means that its derivative goes to zero if r goes to plus or minus infinity.

Given this terminology and notation, we can now discuss what is in the book by Verboon, his 1994 dissertation at the Department of Data Theory of

Reviewer's Address: Jan de Leeuw, UCLA Department of Statistics, 8118 Math Sciences, 405 Hilgard Ave, Los Angeles, CA 90095-1554, USA, email: deleeuw@stat.ucla.edu

the University of Leiden. It can be interpreted as yet another volume in the series of books on nonlinear multivariate analysis, or the *Gifi techniques.* Chapter 1 indicates that outliers are not uncommon in NLMVA, and they can dominate the outcome of the analysis. At least I think that is what it *intends* to convey. The chapter is a quick and rather superficial discussion of some of the key concepts used in the book, without indicating precisely what makes these concepts important. It follows that the book is mostly useful for readers who already know about robust statistics and downweighting of outliers.

Chapter 2 is the most original and the most important part of the book. It discusses the iterative majorization algorithms which are used in many of the more complicated *Gifi system.* Majorization is used to replace an optimization problem by a sequence of simpler problems, which are local approximations to the original problem. Solving the sequence of simpler problems produces a sequence converging to the solution of the original problem. It has been well known since the work of Holland and Welch in the mid-seventies that robust regression problems can be solved by constructing a sequence of iterative least squares problems. To prove global convergence of such procedures, one needs the concept of majorization.

In Chapter 2 majorizations are derived for the Huber, Tukey, and Hampel reweighting functions. Unfortunately, the derivations seem to be somewhat ad hoc, and is unclear what the general principle behind them is. Thus it is not trivial to generalize to other reweighting functions. A more general approach would be based on the observation that all reweighting functions have a bounded second derivative. Thus $\phi''(r) \leq M$ for all r, and consequently

$$\phi(r) \leq \phi(\tilde{r}) + \phi'(\tilde{r})(r - \tilde{r}) + \frac{1}{2}M(r - \tilde{r})^2$$

$$= \phi(\tilde{r}) - \frac{1}{2M}[\phi'(\tilde{r})]^2 + \frac{1}{2}M(r - \bar{r})^2,$$

where $\bar{r} = \tilde{r} - \frac{1}{M}\phi'(\tilde{r})$. This automatically gives a quadratic majorizer. It can often be improved if we have an inequality of the form $\phi''(r) \leq \lambda(r)$ for all r, where λ is some simple function. I think most of the majorizations proposed in the book fit into this general framework. Quadratic majorizations always lead to least squares problems, i.e. to iterative reweighted least squares. Thus the general approach outlined in the book makes it possible to replace additive reweightings of the residuals by sequences of least squares functions (as long as there are bounded second derivatives).

This is what Verboon does in the rest of the book for some examples important in multivariate data anlysis. Thus there is robust Procrustus matching, robust canonical discriminant analysis, robust multiple regression, and robust principal component analysis. These applied chapters are well-written, with nice examples, and with questions of stability and inference briefly discussed. The approach is not always successfull, and in some cases does not really improve the non-robust least squares methods, but the advantages of having a general framework in which to incorporate redescending reweighting of residuals is clear.

In summary, the book is useful for people who already know quite a bit about robust regression, nonlinear multivariate analysis according to *Gifi*, and majorization algorithms. It cannot be used as an introduction to these topics. For this (presumably rather small) class of researchers Verboon presents a well-written and nicely illustrated extension of the basic least squares approach. He does not emphasize, unfortunately, that the gain in generality comes with a non-ignorable price-tag. Many aspects of least squares, such as the connection with Euclidean geometry, with projections, and with linearity, are lost — with a corresponding loss in interpretational possibilities and analytical results. It is highly doubtful, given the relatively minor gains summarized in Chapter 7, that applied reseachers are willing to pay the price.

University of California Jan de Leeuw