Journal of Computational Mathematics, Vol.13, No.2, 1995, 130-143.

HUBER'S M-ESTIMATOR ON UNDERDETERMINED PROBLEMS*1)

Wang Jia-song Tang Sheng-rong
(*Department of Mathematics, Nanjing University, Nanjing, China*)

Abstract

After surveying the theoretical aspects of Huber's M-estimator on underdetermined problems, two finite algorithms are presented. Both proceed in a constructive manner by moving from one partition to an adjacent one. One of the algorithm, which uses the tuning constant as a continuation parameter, also has the facility to simultaneously estimate the tuning constant and scaling factor. Stable and efficient implementation of the algorithms is presented together with numerical results. The L_1-norm problem is mentioned as a special case.

1. Introduction

Huber's M-estimator on overdetermined problems has been surveyed by many authors such as Clark[3,4], Madsen and Nielsen[12,13] using various schemes. But the underdetermined problems have not attracted much attention, although they are often met in engineering problems. Here we want to use the most popular approaches on the basis of iterative schemes to solve this kind of problems. In the algorithm a scaling factor can be estimated either at the beginning of the computation or by the algorithm at each iteration.

We are concerned with the following problem:

Problem 1.

$$\min \quad f(X) = \|X\|_2 = \sum_{i=1}^{n} x_i^2 \ , \tag{1}$$

$$\text{s.t.} \qquad AX = b \ , \tag{2}$$

$$A \in R^{m\times n}, m < n \ .$$

The estimator X is called the least square or L_2-estimator and was shown by Gauss[6] in 1821 to be the most probable value under the assumption that the model has independent identical normal distribution. However, as illustrated by Tuckey[15] in

* Received October 22, 1993.
1) The Project Supported by National Natural Science Foundation of China.

1960 , the L_2 estimator is very sensitive to quite small deviation from that assumption, and, in particular, a few gross errors can have a marked effect.

In an effect to find a more robust estimator, Huber[10] suggested replacing the square terms in (1) with a less rapidly increasing function:

$$\rho(x_i) = \begin{cases} \frac{1}{2}x_i^2 & \text{for } |x_i| \le \gamma \\ \gamma|x_i| - \frac{1}{2}\gamma^2 & \text{for } |x_i| > \gamma \end{cases} \tag{3}$$

where γ is a parameter to be estimated from the data. The resulting estimator was shown by Huber [10] to be a maximum likelihood estimator for a perturbed normal distribution and has become known as Huber's M-estimator.

Many iterative methods can be used to obtain the M-estimator. Among those are Huber's method[10], Newton method[12], Beaton and Tuckey's method[15], Clark's method[4]. We find the last one most attractive because of its efficiency and finiteness.

Now, let us consider Problem 1 with the replacement of (3) for the estimator. Then Problem 1 becomes

Problem 2.

$$\min \quad F(X) = \sum_{i=1}^{n} \rho(x_i) = \frac{1}{2}\sum_{\sigma} x_i^2 + \sum_{\sigma_+}(\gamma x_i - \frac{1}{2}\gamma^2) + \sum_{\sigma_-}(-\gamma x_i - \frac{1}{2}\gamma^2) \ , \tag{4}$$

$$s.t. \qquad AX = b \ , \tag{5}$$

$$A \in R^{m\times n}, m \le n \ .$$

where $\sigma = \{i| \quad |x_i| \le \gamma\}, \sigma_- = \{i|x_i < -\gamma\}, \sigma_+ = \{i|x_i > \gamma\}$.

To solve this problem, Lagrange multiplier is used to transform the constrained problem into an unconstrained one:

$$F(X,\Lambda) = \sum_{\sigma}\frac{1}{2}x_i^2 + \sum_{\bar{\sigma}}(|x_i|\gamma - \frac{1}{2}\gamma^2) + \Lambda^T(AX - b)$$

where Λ is a Lagrange multiplier and $\bar{\sigma} = \sigma_+ \cup \sigma_-$.

Since $F(X,\Lambda)$ is convex, the necessary and sufficient condition of minimum is

$$\bigtriangledown F(X,\Lambda) = 0$$

while

$$\bigtriangledown F(X,\Lambda) = \begin{pmatrix} D_\sigma & A^T \\ A & 0 \end{pmatrix}\begin{pmatrix} X^* \\ \Lambda^* \end{pmatrix} + \begin{pmatrix} \gamma e_\sigma \\ -b \end{pmatrix}$$

where D_σ is a diagonal matrix,

$$(D)_{ii} = \begin{cases} 1, \text{if } i \in \sigma \\ 0, \text{if } i \in \bar{\sigma} \end{cases} \qquad (e_\sigma)_i \begin{cases} = 0 \ , & \text{if } i \in \sigma \\ = \theta_i \ , & \text{if } i \in \bar{\sigma}, x_i^* \ne 0 \\ \in [-1,1] \ , & \text{if } i \in \bar{\sigma}, x_i^* = 0 \end{cases}$$

and $\theta_i = sign(x_i)$, $\quad X^*, \Lambda^*$ are the optimum.