One popular robust technique is the so-called M-estimators. Let ri be the residual of the ith datum, i.e. the difference between the ith observation and its fitted value. The standard least-squares method tries to minimize∑i r2i , which is unstable if there are outliers present in the data. Outlying data give an effect so strong in the minimization that the parameters thus estimated are distorted. The M-estimators try to reduce the effect of outliers by replacing the squared residuals r2i by another function of the residuals, yielding min ∑i ρ(r2) where ρ is a symmetric, positive-definite function with a unique minimum at zero, and is chosen to be less increasing than square.

Let p = [p1; : : : ; pm]T be the parameter vector to be estimated. The M-estimator of p based on the function ρ (ri) is the vector p which is the solution of the following *m* equations where the derivative Ψ(x) = d ρ (x)/dx is called the influence function:

∑i Ψ(ri)dri/dpj = 0 for j=1,2 ….m

The influence function Ψ (x) measures the influence of a datum on the value of the parameter estimate. For the least-squares with ρ (x) = x2/2, the influence function is Ψ (x) = x, that is, the influence of a datum on the estimate increases linearly with the size of its error, which confirms the non-robustness of the least-squares estimate. When an estimator is robust, it may be inferred that the influence of any single observation (datum) is insufficient to yield any significant offset.

Huber's function [13] is a parabola in the vicinity of zero, and increases linearly at a given level |x| > k. The 95% asymptotic efficiency on the standard normal distribution is obtained with the tuning constant k = 1:345. This estimator is so satisfactory that it has been recommended for almost all situations; very rarely it has been found to be inferior to some other ρ -function. However, from time to time, difficulties are encountered, which may be due to the lack of stability in the gradient values of the ρ-function because of its discontinuous second derivative:

d2 ρ (x)/dx2 = (1 if |x| <= k,0 if |x| > k.)

Cauchy's function does not guarantee a unique solution. With a descending first derivative, such a function has a tendency to yield erroneous solutions in a way which cannot be observed. The 95% asymptotic efficiency on the standard normal distribution is obtained with the tuning constant c = 2:3849.The other remaining functions have the same problem as the Cauchy function. As can be seen from the influence function, the influence of large errors only decreases linearly with their size. The Geman-McClure and Welsh functions try to further reduce the

effect of large errors.

All these functions do not eliminate completely the influence of large gross errors.

The four last functions do not guarantee unique solution, but reduce considerably, or even eliminate completely, the influence of large gross errors.

Parameter Estimation Techniques:

A Tutorial with Application to Conic Fitting

Zhengyou Zhang

INRIA, 2004 route des Lucioles, BP 93, F-06902 Sophia-Antipolis Cedex, France

Email: zzhang@sophia.inria.fr

Fax: (+33) 93 65 78 45

This paper appeared in Image and Vision Computing, Vol.15, No.1, pages 59-76, January

1997.

1