The Annals of Statistics
1999, Vol. 27, No. 4, 1164–1177

BREAKDOWN POINTS AND VARIATION EXPONENTS OF ROBUST *M*-ESTIMATORS IN LINEAR MODELS[1]

BY IVAN MIZERA AND CHRISTINE H. MÜLLER

Comenius University and Georg-August-University

The breakdown point behavior of M-estimators in linear models with fixed designs, arising from planned experiments or qualitative factors, is characterized. Particularly, this behavior at fixed designs is quite different from that at designs which can be corrupted by outliers, the situation prevailing in the literature. For fixed designs, the breakdown points of robust M-estimators (those with bounded derivative of the score function), depend on the design and the variation exponent (index) of the score function. This general result implies that the highest breakdown point within all regression equivariant estimators can be attained also by certain M-estimators: those with slowly varying score function, like the Cauchy or slash maximum likelihood estimator. The M-estimators with variation exponent greater than 0, like the L_1 or Huber estimator, exhibit a considerably worse breakdown point behavior.

1. Introduction. We consider a *general linear model*,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_N)^T \in \mathbb{R}^N$ is a *vector of observations*, $\boldsymbol{\beta} \in \mathbb{R}^p$ is an unknown *parameter vector*, $\boldsymbol{\epsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)^T \in \mathbb{R}^N$ a *vector of errors*, and $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)^T \in \mathbb{R}^{N\times p}$ is the known matrix of *design points*, which together constitute a *design*. Let $\hat{\boldsymbol{\beta}}$ be an estimator of $\boldsymbol{\beta}$. If $\mathbf{X}$ is given by an experimenter, it can rightly be assumed that the design points do not contain any gross errors, outliers and similar phenomena; in other words, they are without errors and, particularly, they are nonstochastic. The same can be said for factors of linear models which are of qualitative nature (in ANOVA models, for instance). In such a context, it is natural to define the *breakdown point* $\varepsilon^*(\hat{\boldsymbol{\beta}}, \mathbf{y}, \mathbf{X})$ of $\hat{\boldsymbol{\beta}}$ as [He, Jurečková, Koenker and Portnoy (1990), Ellis and Morgenthaler (1992) or Müller (1995, 1997)]

$$\varepsilon^*(\hat{\boldsymbol{\beta}}, \mathbf{y}, \mathbf{X}) = \frac{1}{N}\min\Big\{M\colon \sup_{\tilde{\mathbf{y}}\in B(\mathbf{y}, M)} \|\hat{\boldsymbol{\beta}}(\tilde{\mathbf{y}}, \mathbf{X})\| = \infty\Big\},$$

where $B(\mathbf{y}, M) = \{\tilde{\mathbf{y}}\colon \operatorname{card}\{n\colon \tilde{y}_n \neq y_n\} \leq M\}$. This approach to the breakdown point (with *fixed design*) differs from that prevailing in the literature (with *moving design*). The latter assumes that not only the y_i's, but also the design points $\mathbf{x}_i$ are vulnerable to errors; as a consequence, the definition of the

Received October 1996; revised April 1999.

[1]Supported by Slovak VEGA Grants 1/1489/94, 1/4196/97 and by Grants 436 SLK 17/3/95, 17/8/96 and 17/9/96 of the Deutsche Forschungsgemeinschaft.

AMS 1991 *subject classifications*. Primary 62F35, 62F10; secondary 62J05, 62J10, 62K99.

Key words and phrases. Breakdown point, L_1 estimator, linear model, M-estimator, planned experiments, regular variation.

breakdown point allows also for perturbations of the $\mathbf{x}_i$'s [see, e.g., Rousseeuw and Leroy (1987)].

In the present paper, we support the view that for planned experiments and models with qualitative factors, the approach with fixed design is more appropriate. In this vein, an upper bound for the breakdown point was given in Müller (1995, 1997): for any regression equivariant estimator [an estimator such that $\hat{\boldsymbol{\beta}}(\mathbf{y}+\mathbf{X}\boldsymbol{\theta},\mathbf{X})=\hat{\boldsymbol{\beta}}(\mathbf{y},\mathbf{X})+\boldsymbol{\theta}$ for all $\mathbf{y}$, $\mathbf{X}$ and $\boldsymbol{\theta}$],

$$\varepsilon^*(\hat{\boldsymbol{\beta}},\mathbf{y},\mathbf{X}) \le \frac{1}{N}\left\lfloor \frac{N-\mathscr{N}(\mathbf{X})+1}{2}\right\rfloor; \tag{1.1}$$

here $\mathscr{N}(\mathbf{X})=\max_{\boldsymbol{\beta}\neq\mathbf{0}} \operatorname{card}\{n\colon \mathbf{x}_n^T\boldsymbol{\beta}=0\}$ is the maximal number of regressors $\mathbf{x}_n$ in a subspace of $\mathbb{R}^p$ ($\lfloor u\rfloor$ denotes the largest integer $\le u$). Note that (1.1) is the same upper bound as in the moving design setting (which allows for more general perturbations of the data points); see Rousseeuw and Leroy (1987) for $\mathscr{N}(\mathbf{X})=p-1$ and Mili and Coakley (1993) for the general case.

It is of interest whether in fixed design setting the upper bound (1.1) is attainable and which estimators can achieve it. Müller (1995) showed that the bound (1.1) can be attained by some trimmed L_p estimators. In the present paper, we address the question whether the bound can be attained also by certain M-estimators.

An *M-estimator* is defined via minimization,

$$\hat{\boldsymbol{\beta}}(\mathbf{y},\mathbf{X}) = \arg\min_{\boldsymbol{\beta}} D(\boldsymbol{\beta},\mathbf{y},\mathbf{X}),$$

where

$$D(\boldsymbol{\beta},\mathbf{y},\mathbf{X}) = \sum_{n=1}^{N} \varphi(y_n-\mathbf{x}_n^T\boldsymbol{\beta})$$

is an *objective function* and φ is a given score function from $\mathbb{R}$ to $\mathbb{R}$; in the sequel, we suppose that φ is absolutely continuous, a primitive function of ψ. All M-estimators are regression equivariant.

For the special case of the *location model*, that is, the simplest regression model with $\mathbf{x}_1=\mathbf{x}_2=\cdots=\mathbf{x}_N=1$ and $\boldsymbol{\beta}\in\mathbb{R}^1$, it is known that M-estimators with bounded ψ attain the maximum possible breakdown point of approximately 50%, whenever ψ is nondecreasing, corresponding to a convex φ [Huber (1981), page 54], or φ is unbounded for redescending ψ [Huber (1984)]. In regression models with moving design, all M-estimators with nondecreasing ψ have the same low breakdown point $1/N$ as the least squares estimator. This can be shown along the lines of Maronna, Bustos and Yohai (1979), whose finding created an overall impression that M-estimators possess "bad" breakdown point behavior. However, their result was based on the convexity of φ and used a moving design definition of the breakdown point.

In the fixed design setting, the performance of M-estimators radically changes. As shown below, M-estimators can exhibit high breakdown points, including the highest possible ones. In general linear models, the breakdown point of an M-estimator depends on the asymptotic behavior of the function

φ, described in terms of the exponent r of regular variation of φ. Together with the design matrix $\mathbf{X}$, r determines the breakdown behavior of the M-estimator.

The paper is organized as follows. In Section 2 we briefly introduce regularity conditions and state the main result: a dependence of the breakdown point of an M-estimator on its variation exponent r and the design matrix $\mathbf{X}$. In Section 3 we discuss the regularity conditions, in Section 4 the effect of unknown scale and some computational aspects. Section 5 deals with applications; the explicit breakdown points are computed for special linear models and it is shown how choosing an appropriate design, for a given φ, can improve the breakdown point of the M-estimator. Finally, Section 6 contains the proofs.

2. Main result. A measurable function $f\colon \mathbb{R}^+ \to \mathbb{R}^+$ is called *regularly varying*, if there is a function h such that for all $u > 0$,

$$\lim_{t\to\infty} \frac{f(tu)}{f(t)} = h(u)$$

[see Bingham, Goldie and Teugels (1987), Chapter 1, or Resnick (1987), Chapter 0]. In such a case, $h(u) = u^r$ for some $r \in \mathbb{R}$, which is called the *exponent* or *index of variation*. If $r = 0$, then f is called *slowly varying*.

We introduce the following assumptions about φ:

(A) Shape: φ is symmetric ($\varphi(t) = \varphi(-t)$), nondecreasing on $[0, +\infty]$ and nonnegative.
(B) Unboundedness: φ is unbounded.
(C) Subadditivity: there exists $L > 0$ such that $\varphi(t + s) \le \varphi(t) + \varphi(s) + L$ for all $t, s \ge 0$.
(D) Regular variation: φ is regularly varying with an exponent $r \ge 0$.

Given $r \ge 0$, we define, using the convention $0^0 = 0$,

$$\mathscr{M}(\mathbf{X}, r) = \min\left\{\text{card } E\colon \sum_{n\in E} |\mathbf{x}_n^T\boldsymbol{\beta}|^r \ge \sum_{n\notin E} |\mathbf{x}_n^T\boldsymbol{\beta}|^r \text{ for some } \boldsymbol{\beta} \ne \mathbf{0}\right\},$$

where E runs over the subsets of $\{1, 2, \ldots, N\}$. For $r = 1$, $\mathscr{M}(\mathbf{X}, 1)$ coincides with the minimal possible cardinality of a set $E \subset \{1, 2, \ldots, N\}$ such that

$$\max_{\boldsymbol{\beta}\ne\mathbf{0}} \frac{\sum_{n\in E} |\mathbf{x}_n^T\boldsymbol{\beta}|}{\sum_{n=1}^N |\mathbf{x}_n^T\boldsymbol{\beta}|} \ge \frac{1}{2}.$$

In this form, $\mathscr{M}(\mathbf{X}, 1)$ was introduced by He, Jurečková, Koenker and Portnoy (1990) in the context of breakdown points of L_1 type estimators. Ellis and Morgenthaler (1992) employed it to derive a lower bound for the exact fit degree of the L_1 estimator. They also pointed out its diagnostic value in the assessment of leverage points in L_1 regression.

PROPOSITION 1. *If* $q \geq r \geq 0$ *then*

$$\mathscr{M}(\mathbf{X}, q) \leq \mathscr{M}(\mathbf{X}, r) \leq \mathscr{M}(\mathbf{X}, 0) = \left\lfloor \frac{N - \mathscr{N}(\mathbf{X}) + 1}{2} \right\rfloor. \tag{2.1}$$

The following general theorem shows how the breakdown point of an M-estimator depends only on r and $\mathbf{X}$, via $\mathscr{M}(\mathbf{X}, r)$.

THEOREM 1. *If* $\hat{\boldsymbol{\beta}}$ *is an* M*-estimator with* φ *satisfying* (A), (B), (C) *and* (D) *with variation exponent* $r \in [0, 1]$, *then*

$$\frac{\mathscr{M}(\mathbf{X}, r)}{N} \leq \varepsilon^*(\hat{\boldsymbol{\beta}}, \mathbf{y}, \mathbf{X}) \leq \frac{\mathscr{M}(\mathbf{X}, r) + 1}{N} \tag{2.2}$$

for all $\mathbf{y} \in \mathbb{R}^N$.

The proof of Theorem 1 shows that the upper bound is still true when φ satisfies only (A), (B) and (D) with $r > 1$. Then, in general, the upper bound is no longer sharp, as the case $r = 2$ shows: the least squares estimator has the breakdown point $1/N$, which is in general not equal to $(\mathscr{M}(\mathbf{X}, 2) + 1)/N$. Hence the connection between the breakdown point and variation exponent holds only for robust M-estimators (those with bounded ψ).

Theorem 1 shows that for a large sample size N, the breakdown point is approximately equal to $\mathscr{M}(\mathbf{X}, r)/N$. If $r = 0$ in (D), the exact equality is the immediate consequence of the upper bound (1.1) and Proposition 1. In particular, M-estimators with slowly varying φ, like the Cauchy or slash maximum likelihood estimator, have the highest breakdown point possible for regression equivariant estimators. Proposition 1 also implies that this is not true, in general, for M-estimators with $r = 1$ (see also Section 5).

For certain special cases, for the L_1-estimator, for instance, a sharpened version of Theorem 1 can be proved, leading to an improvement of Theorem 5.3 of He, Jurečková, Koenker and Portnoy (1990).

THEOREM 2. *For the* L_1 *estimator* $\hat{\boldsymbol{\beta}}_1$,

$$\varepsilon^*(\hat{\boldsymbol{\beta}}_1, \mathbf{y}, \mathbf{X}) = \frac{\mathscr{M}(\mathbf{X}, 1)}{N}$$

holds for all $\mathbf{y} \in \mathbb{R}^N$ *and* $\mathbf{X}$.

3. Discussion of regularity conditions. Concerning shape, it is clear that nonnegativity of φ can be always achieved by adding a constant. The other assumptions in (A) are satisfied by all φ used in practice.

A standard property of regularly varying functions implies that if (D) holds with $r > 0$, then (B) is satisfied. This implication does not hold for $r = 0$; the location case shows that for $r = 0$, the assumption (B) in Theorem 1 is essential.

Since φ and $\varphi + K$ yield the same M-estimates, (C) really means a kind of subadditivity: a constant can be added to φ such that for the resulting φ,

$\varphi(t+s) \le \varphi(t) + \varphi(s)$ holds, not only for $t, s \ge 0$, but for all t, s, as shown by Lemma 1 in Section 6. Subadditivity holds for $\varphi(u) = |u|$; for other functions φ this is not that obvious, nevertheless, (C) can be proved relatively easily from simpler assumptions.

A careful inspection of the proof of Lemma 4.2 in Huber (1984) reveals that if:

(i) (A) is fulfilled and $\varphi(0) = 0$;

(ii) there is an $u_0 \le 0$ such that ψ is nondecreasing for $0 < u < u_0$ and nonincreasing for $u_0 < u < \infty$, with $\psi(u) \to 0$ for $u \to \infty$; then $\eta(t) = \sup_s |\varphi(t+s) - \varphi(s)| - \varphi(t)$ is bounded. This immediately implies (C) [see also Mizera (1996)]. Assumptions (i) and (ii) are satisfied, for instance, by the Cauchy log-likelihood $\varphi(u) = \log(1+u^2)$. Another example is provided by the slash likelihood [for the definition, see Morgenthaler and Tukey (1991) or Hoaglin, Mosteller and Tukey (1983)].

For convex φ, (C) can be easily verified and subadditivity implies that φ is bounded. Thus, our regularity conditions hold for the Huber as well as the logistic maximum likelihood estimator [for the definitions, see Hampel, Ronchetti, Rousseeuw and Stahel (1986)].

The easiest way to verify (D) is a direct one, using the l'Hospital rule. Also, all functions with a nonzero limit at ∞ are slowly varying; as a consequence, any nondecreasing bounded ψ is slowly varying and this implies, in view of Proposition 1.5.8 of Bingham, Goldie and Teugels (1987) [see also Resnick (1987), Theorem 0.6a], that the corresponding φ is regularly varying with $r = 1$. There are also other criteria for regular variation: for instance, the von Mises property asserts that φ is regularly varying with the exponent r whenever

$$\lim_{u\to\infty} \frac{u\psi(u)}{\varphi(u)} = r. \tag{D1}$$

If $u\psi(u)$ is bounded, then (B) immediately implies the slow variation of φ.

If φ is the log-likelihood corresponding to a density f, that is, if $\varphi(u) = -\log f(u)$, then (A) is satisfied whenever f is symmetric and unimodal. Condition (D) in this case reflects the tail behavior of f: it is implied by a stronger condition:

(D2) There exist $r \ge 0$, $c > 0$ and $m > 0$ such that

$$\lim_{t\to\infty} \frac{\varphi(t)}{m\log(t) + c\,t^r} = 1.$$

Condition (D2) is itself implied by a further stronger condition:

(D3) There exist $L^* > 0$, $r \ge 0$, $m \ge 0$ and $c \ge 0$, $m + c > 0$, such that $|\varphi(t) - m\,\log(t) - c\,t^r| \le L^*$ for all $t > L^*$.

A special class of functions satisfying (D3) appeared in He, Jurečková, Koenker and Portnoy (1990); they called those φ satisfying (D3) with $m = 0$

and $r = 1$ "score functions of L_1 type." The log-likelihood functions φ satisfying (D2) with $r > 0$ correspond to light-tailed distributions (distributions with exponential tails, for instance, logistic and exponential distribution); the slowly varying φ with $r = 0$ correspond to heavy-tailed distributions (distributions with algebraic tails, like the Student t-distributions with k degrees of freedom, including the Cauchy distribution with $k = 1$). For more background on tail behavior, see Jurečková (1981) or He, Jurečková, Koenker and Portnoy (1990).

The following proposition lists some interrelations among the regularity conditions. For the proof, the interested reader is referred to Mizera and Müller (1996).

PROPOSITION 2. *Suppose that* (A) *holds*. *Then*:

(a) (D1) *implies* (D);
(b) (D2) *implies* (D) *and* (B);
(c) (D3) *implies* (D2), *and also* (C) *if* $r \in [0, 1]$;
(d) (B), (C) *and* (D) *imply* $r \in [0, 1]$.

Summarizing the discussion above, we have (among others) the following examples of estimators satisfying (A), (B), (C) and (D): the L_1, Huber and logistic maximum likelihood estimator ($r = 1$), the Cauchy and slash maximum likelihood estimator ($r = 0$).

4. Estimators with unknown scale and computational aspects. In Section 1, we defined a regression M-estimator without involving a scale parameter; for simplicity, the scale was set equal to one. This does not make any difference for the L_1 or the least squares estimators; however, it is well known that the scale should not be ignored in the practical use of other M-estimators. Moreover, scale can have also impact on computational issues; this point was illustrated by Mizera (1994) and is also discussed below.

There are two commonly used approaches to scale adjustment. The Studentizing approach replaces $D(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ by

$$D_S(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X}) = \sum_{n=1}^{N} \varphi\left(\frac{y_n - \mathbf{x}_n^T \boldsymbol{\beta}}{S(\mathbf{y}, \mathbf{X})}\right),$$

where $S(\mathbf{y}, \mathbf{X})$ is a suitable regression and scale equivariant estimator. To maintain the high breakdown point, S should have high breakdown itself. Suitable estimators are easily found in the location/scale model (the most popular choice is the MAD, median deviation from the median, but there are also better options). In regression, finding a suitable S may be more difficult, though we admit that recently a considerable work has been done in this direction.

The second approach to scale adjustment, the simultaneous approach, arises naturally in the likelihood setting. We simply compute the simultaneous maximum likelihood regression/scale estimator: $D(\boldsymbol{\beta}, \mathbf{y}, \mathbf{X})$ is now

replaced by

$$D(\boldsymbol{\beta}, \sigma, \mathbf{y}, \mathbf{X}) = N \log \sigma + \sum_{n=1}^{N} \varphi\left(\frac{y_n - \mathbf{x}_n^T \boldsymbol{\beta}}{\sigma} \right), \tag{4.1}$$

and $\boldsymbol{\beta}$ and σ simultaneously minimizing this new objective function are sought. Differentiation leads to first-order equations,

$$\begin{aligned} \sum_{n=1}^{N} \mathbf{x}_n \psi\left(\frac{y_n - \mathbf{x}_n^T \boldsymbol{\beta}}{\sigma} \right) &= \mathbf{0}, \\ \sum_{n=1}^{N} \chi\left(\frac{y_n - \mathbf{x}_n^T \boldsymbol{\beta}}{\sigma} \right) &= 0, \end{aligned} \tag{4.2}$$

with $\psi(u) = (\partial/\partial u)\varphi(u)$ and $\chi(u) = u\psi(u) - 1$. Note that our χ is usually monotone, but ψ not.

The advantage of the simultaneous approach is the possible implication of unique roots of (4.2). Any known iterative method for minimization ends with some solution of (4.2). In the presence of multiple roots, the notorious problem is that this solution may correspond just to a local minimum, or even to a saddle-point or a local maximum. Surprisingly however, for the Cauchy maximum likelihood location/scale estimator this problem vanishes; the result of Copas (1975) shows that (4.2) have a unique root in this case, unless the data are equidistributed between just two distinct points, a rather exceptional configuration. Gabrielsen (1982) complemented this result by showing that this unique root corresponds to the global minimum; he also noted that the property extends to t distributions with degrees of freedom greater or equal to 1. We do not know whether this result extends to certain other redescending estimators, but would find that plausible.

Unfortunately, the result of Copas does not extend to the regression case, as was also noted by Gabrielsen (1982). Even in the simplest regression setting (simple regression with or without intercept), there are examples exhibiting more than one local minimum. Though we feel that in "typical cases" this does not occur, it is not clear whether such a statement can be given a suitable formalization.

On the positive side, we know that our Theorem 1 extends, with some minor modifications, to the Cauchy maximum likelihood regression/scale estimator (the proof, contrary to the proof of Theorem 1, is lengthy and tedious, and will appear elsewhere). That is, for $\varphi(u) = \log(1 + u^2)$, the breakdown remains approximately the same in regression/scale setting. We conjecture that the same is true for other φ satisfying Theorem 1. We underline, however, that it is crucial to define the M-estimator as the *global* minimizer of (4.1). The breakdown behavior of the other roots of (4.2) may be different. While the global minimizer remains stable under almost 50% contamination, there are examples where a solution of (4.2) (corresponding to a local minimum), moves to infinity as a consequence of moving just one data point in the y direction.

Therefore, any computational method should seek global minimizers, not just stationary points. While this might be perceived as a drawback, we recall that common high-breakdown estimators (like the LMS or LTS) exhibit much worse behavior in this respect. In this context, Ruppert (1992) pointed out that S-estimators behave better, thanks to the smoothness of their objective functions. It is known that first-order conditions for S-estimators also lead to (4.2), with $\chi(u) = \varphi(u) - c$ now (c usually equal to 1/2); however, it is also stressed that it is the minimization definition of S-estimators which ensures their desired breakdown properties (if χ is bounded). Hence, there is a connection between regression/scale M-estimators and S-estimators. In particular, the first-order conditions coincide for the L_p estimators, when $\varphi(u) = |u|^p$.

The similarity to S-estimators indicates that the algorithm of Ruppert (1992) is easily adaptable also to our setting. In any case, its performance should be at least as good as for S-estimators. Moreover, we may expect improvements benefiting from the likelihood-maximization nature of the problem; for instance, employing the EM algorithm in the local step, we can eliminate local maxima of (4.1). Lange, Little and Taylor (1989) used the EM algorithm for computing regression estimators with t errors (including the Cauchy), apparently without a concern for multiple extremes (they report not finding any such problem in their computations).

We conclude this section by turning attention to a different problem: the computation of $\mathscr{M}(\mathbf{X}, r)$, especially when $r = 1$, which may be of interest in evaluating designs for the L_1 regression. Some aspects of this are addressed in the next section; more can be found in the Appendix of Mizera and Müller (1996).

5. Determination of $\mathscr{M}(\mathbf{X}, r)$ for special models. For designs with minimum support, like the classical A- and D-optimal designs for polynomial regression and the designs for the one-way layout, the breakdown points of all M-estimators with φ satisfying (A), (B), (C) and (D) with $r \in [0, 1]$ are equal and attain the maximum value if the numbers of repetitions are equal.

PROPOSITION 3. *Suppose that* $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subseteq \{\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_p\} \subset \mathbb{R}^p$. *Let* $N_i = \sum_{n=1}^{N} 1_{\{\boldsymbol{\chi}_i\}}(x_n)$ *for* $i = 1, \ldots, p$. *If* $\boldsymbol{\chi}_1, \ldots, \boldsymbol{\chi}_p$ *are linearly independent, then*

$$\mathscr{M}(\mathbf{X}, r) = \left\lfloor \frac{N - \mathscr{N}(\mathbf{X}) + 1}{2} \right\rfloor = \left\lfloor \frac{1 + \min_{i=1,\ldots,p} N_i}{2} \right\rfloor \leq \left\lfloor \frac{N + p}{2p} \right\rfloor$$

for every $r \in [0, 1]$.

The next proposition provides an upper bound for $\mathscr{M}(\mathbf{X}, 1)$, in particular for polynomial and multiple linear regression. It confirms the upper bound 1/4 given by Ellis and Morgenthaler (1992), who showed that in multiple regression (of any dimension p) there are designs attaining this upper bound. The proofs of both propositions can be found in Mizera and Müller (1996).

PROPOSITION 4. *Let, for* $n = 1, \dots, N$, $\mathbf{x}_n = (x_n^1, \dots, x_n^p)^T$ *with* $x_1^i \le x_2^i \le \cdots \le x_N^i$ *for some* $i \in \{1, \dots, p\}$. *If there exists* $\mathbf{A} \in \mathbb{R}^{2\times p}$ *such that* $\mathbf{A}\mathbf{x}_n = (1, x_n^i)^T$ *for* $n = 1, \dots, N$, *then*

$$\mathscr{M}(\mathbf{X}, 1) \le \min\Bigg\{M\colon \sum_{n=1}^{N-M}(x_n^i - x_1^i) \le \sum_{n=N-M+1}^{N}(x_n^i - x_1^i) \quad \text{or} \quad \sum_{n=M+1}^{N}(x_N^i - x_n^i) \le \sum_{n=1}^{M}(x_N^i - x_n^i)\Bigg\} \le \left\lfloor \frac{N+3}{4} \right\rfloor.$$

For linear regression, that is, $\mathbf{x}_n = (1, t_n)^T \in \mathbb{R}^2$, the first inequality becomes an equality [see also Müller (1997), page 67]. The upper bound of $N/4$ is attained by designs with equally spaced design points t_n, as well as by the classical D-optimal design which puts a half of the observation at each endpoint of the design region. There are also other designs attaining this upper bound, for instance, the design with L observations at 1 and 3 and $2L$ observations at 2.

If we have linear regression through the origin, that is, $\mathbf{x}_n = t_n \in \mathbb{R}$ with $|t_1| \le |t_2| \le \cdots \le |t_N|$, then

$$\mathscr{M}(\mathbf{X}, 1) = \min\Bigg\{M\colon \sum_{n=1}^{N-M}|t_n| \le \sum_{n=N-M+1}^{N}|t_n|\Bigg\} \le \left\lfloor \frac{N+1}{2} \right\rfloor,$$

with equality if and only if $|t_1| = |t_2| = \cdots = |t_N|$ [see also He, Jurečková, Koenker and Portnoy (1990)].

6. Proofs.

PROOF OF PROPOSITION 1. We first prove the inequality part of (2.1). Take an arbitrary $\boldsymbol{\beta} \neq 0$ and $E \subset \{1, 2, \dots, N\}$ with $\operatorname{card} E = \mathscr{M}(\mathbf{X}, r)$ and

$$\sum_{n\in E}|\mathbf{x}_n^T\boldsymbol{\beta}|^r - \sum_{n\notin E}|\mathbf{x}_n^T\boldsymbol{\beta}|^r \ge 0. \tag{6.1}$$

If there are $i \in E$ and $j \notin E$ such that $|\mathbf{x}_i^T\boldsymbol{\beta}| < |\mathbf{x}_j^T\boldsymbol{\beta}|$, then we can exchange i and j, that is, we can consider $(E \cup \{j\}) \setminus \{i\}$ instead of E. Thus, without loss of generality we can assume that

$$d = \min\{|\mathbf{x}_n^T\boldsymbol{\beta}|\colon n \in E\} \ge \max\{|\mathbf{x}_n^T\boldsymbol{\beta}|\colon n \notin E\}.$$

Then

$$\begin{aligned} 0 &\le d^{q-r}\Bigg(\sum_{n\in E}|\mathbf{x}_n^T\boldsymbol{\beta}|^r - \sum_{n\notin E}|\mathbf{x}_n^T\boldsymbol{\beta}|^r\Bigg) \\ &\le \sum_{n\in E}|\mathbf{x}_n^T\boldsymbol{\beta}|^r|\mathbf{x}_n^T\boldsymbol{\beta}|^{q-r} - \sum_{n\notin E}|\mathbf{x}_n^T\boldsymbol{\beta}|^r|\mathbf{x}_n^T\boldsymbol{\beta}|^{q-r}, \end{aligned}$$

hence $\mathscr{M}(\mathbf{X}, q) \le \mathscr{M}(\mathbf{X}, r)$.

Now we show the equality in (2.1). By definition of $\mathscr{N}(\mathbf{X})$, there exists $\boldsymbol{\beta} \neq \mathbf{0}$ and $E_0 \subset \{1, 2, \ldots, N\}$ with card $E_0 = \mathscr{N}(\mathbf{X})$ such that $|\mathbf{x}_n^T\boldsymbol{\beta}|^0 = 0$ for all $n \in E_0$ and $|\mathbf{x}_n^T\boldsymbol{\beta}|^0 = 1$ for all $n \notin E_0$. Then for any subset $E \subset \{1, 2, \ldots, N\} \setminus E_0$ with card $E = \lfloor (N - \mathscr{N}(\mathbf{X}) + 1)/2 \rfloor$ we have

$$\sum_{n\in E} |\mathbf{x}_n^T\boldsymbol{\beta}|^0 = \left\lfloor \tfrac{1}{2}(N - \mathscr{N}(\mathbf{X}) + 1) \right\rfloor$$
$$\geq N - \left\lfloor \tfrac{1}{2}(N - \mathscr{N}(\mathbf{X}) + 1) \right\rfloor - \mathscr{N}(\mathbf{X}) = \sum_{n\notin E} |\mathbf{x}_n^T\boldsymbol{\beta}|^0;$$

hence $\mathscr{M}(\mathbf{X}, 0) \leq \lfloor (N - \mathscr{N}(\mathbf{X}) + 1)/2 \rfloor$. To see the reverse inequality, take any $\boldsymbol{\beta} \neq 0$ and $E \subset \{1, 2, \ldots, N\}$ with card $E = \mathscr{M}(\mathbf{X}, 0)$ such that (6.1) holds with $r = 0$. Since at most $\mathscr{N}(\mathbf{X})$ of the $\mathbf{x}_n^T\boldsymbol{\beta}$'s satisfy $|\mathbf{x}_n^T\boldsymbol{\beta}|^0 = 0$, we have

$$\mathscr{M}(\mathbf{X}, 0) \geq \sum_{n\in E} |\mathbf{x}_n^T\boldsymbol{\beta}|^0 \geq \sum_{n\notin E} |\mathbf{x}_n^T\boldsymbol{\beta}|^0 \geq N - \mathscr{M}(\mathbf{X}, 0) - \mathscr{N}(\mathbf{X}),$$

and this implies that $2\mathscr{M}(\mathbf{X}, 0) \geq N - \mathscr{N}(\mathbf{X})$; that is, $\mathscr{M}(\mathbf{X}, 0) \geq \lfloor (N - \mathscr{N}(\mathbf{X}) + 1)/2 \rfloor$.

To prove Theorem 1, we require three lemmas.

LEMMA 1. *Conditions* (A) *and* (C) *imply* $\varphi(t + s) \leq \varphi(t) + \varphi(s) + L$ *for all* $t, s \in \mathbb{R}$.

PROOF. We prove the equivalent assertion: for any $s, t \in \mathbb{R}$, $\varphi(s - t) \geq \varphi(s) - \varphi(t) - L$. If $0 \leq t < s$, then $\varphi(s) = \varphi(s - t + t) \leq \varphi(s - t) + \varphi(t) + L$. If $s < t \leq 0$, then, due to the symmetry of φ, we have $\varphi(s) = \varphi(-s) = \varphi(-(s - t) - t) \leq \varphi(s - t) + \varphi(t) + L$. Finally, if $|s| \leq |t|$, then $\varphi(s - t) - \varphi(s) \geq 0 - \varphi(t) - L$, since φ is nonnegative and nondecreasing.

LEMMA 2. *Suppose that* φ *satisfies conditions* (A) *and* (D) *and let* $t_k \to \infty$ *and* $u_k \to u$ *be arbitrary. Then*

$$\lim_{k\to\infty} \frac{\varphi(u_k t_k)}{\varphi(t_k)} = u^r,$$

for $k \to \infty$, *whenever* (a) $r > 0$, $u_k \geq 0$ *and* $u \geq 0$, *or* (b) $r = 0$ *and* $u > 0$.

The lemma follows from the standard property of regularly varying functions; see, for instance, Bingham, Goldie and Teugels [(1987), Theorem 1.5.7] or Resnick [(1987), Proposition 0.5].

LEMMA 3. *If* $\mathscr{N}(\mathbf{X}) < N$, *then for all* $r \geq 0$ *there is a* $\boldsymbol{\beta} \neq \mathbf{0}$ *and a subset* $E \subseteq \{1, 2, \ldots, N\}$ *with* card $E = \mathscr{M}(\mathbf{X}, r) + 1$ *such that*

$$\sum_{n\in E} |\mathbf{x}_n^T\boldsymbol{\beta}|^r > \sum_{n\notin E} |\mathbf{x}_n^T\boldsymbol{\beta}|^r.$$

PROOF. Let $\boldsymbol{\beta} \neq \mathbf{0}$ and $E \subset \{1, 2, \ldots, N\}$ with $\operatorname{card} E = \mathscr{M}(\mathbf{X}, r)$ so that

$$\sum_{n \in E} |\mathbf{x}_n^T \boldsymbol{\beta}|^r \geq \sum_{n \notin E} |\mathbf{x}_n^T \boldsymbol{\beta}|^r \tag{6.2}$$

is satisfied. If the inequality in (6.2) is sharp then it remains sharp when we add an element to E. If we have equality in (6.2) and there exists $n \notin E$ with $|\mathbf{x}_n^T \boldsymbol{\beta}| \neq 0$ then we can add this n to E so that for $E \cup \{n\}$ the inequality in (6.2) is sharp. If $|\mathbf{x}_n^T \boldsymbol{\beta}| = 0$ for all $n \notin E$ then equality in (6.2) implies $|\mathbf{x}_n^T \boldsymbol{\beta}| = 0$ for all $n = 1, 2, \ldots, N$ so that $\mathscr{N}(\mathbf{X}) = N$. This contradicts the assumptions.

PROOF OF THEOREM 1. *Lower bound*. Due to the regression equivariance of M-estimators, we may assume that $\hat{\boldsymbol{\beta}}(\mathbf{y}, \mathbf{X}) = \mathbf{0}$, without loss of generality. Let $M = N\varepsilon^*(\hat{\boldsymbol{\beta}}, \mathbf{y}, \mathbf{X})$. Then for every $k \in \mathbb{N}$ there is a $\mathbf{y}^k \in B(\mathbf{y}, M)$, such that $\|\hat{\boldsymbol{\beta}}(\mathbf{y}^k, \mathbf{X})\| > k$. Again without restricting generality we may suppose that $y_n^k = y_n$ for $n = 1, 2, \ldots, N - M$. For brevity, $\hat{\boldsymbol{\beta}}(\mathbf{y}^k, \mathbf{X})$ is denoted by $\boldsymbol{\beta}^k$. We have that $\|\boldsymbol{\beta}^k\| \to \infty$ and, in view of the compactness of the unit sphere in $\mathbb{R}^p$, we may also suppose that $\boldsymbol{\beta}^k / \|\boldsymbol{\beta}^k\| \to \boldsymbol{\beta}^0$, passing to a subsequence otherwise. Since

$$\begin{aligned} 0 &\geq D(\boldsymbol{\beta}^k, \mathbf{y}^k, \mathbf{X}) - D(\mathbf{0}, \mathbf{y}^k, \mathbf{X}) \\ &= \sum_{n=1}^{N-M} \varphi(y_n - \mathbf{x}_n^T \boldsymbol{\beta}^k) + \sum_{n=N-M+1}^{N} \varphi(y_n^k - \mathbf{x}_n^T \boldsymbol{\beta}^k) \\ &\quad - \sum_{n=1}^{N-M} \varphi(y_n) - \sum_{n=N-M+1}^{N} \varphi(y_n^k), \end{aligned}$$

we have by Lemma 1 and the symmetry of φ,

$$0 \geq \sum_{n=1}^{N-M} \varphi(\mathbf{x}_n^T \boldsymbol{\beta}^k) - 2 \sum_{n=1}^{N-M} \varphi(y_n) - \sum_{n=N-M+1}^{N} \varphi(\mathbf{x}_n^T \boldsymbol{\beta}^k) - NL. \tag{6.3}$$

Dividing (6.3) by $\varphi(\|\boldsymbol{\beta}^k\|)$ and letting $k \to \infty$, we obtain

$$\begin{aligned} 0 \geq \sum_{n=1}^{N-M} \liminf_{k \to \infty} \frac{\varphi\big(\|\boldsymbol{\beta}^k\| \, \big|\mathbf{x}_n^T \frac{\boldsymbol{\beta}^k}{\|\boldsymbol{\beta}^k\|}\big|\big)}{\varphi(\|\boldsymbol{\beta}^k\|)} \\ - \sum_{n=N-M+1}^{N} \limsup_{k \to \infty} \frac{\varphi\big(\|\boldsymbol{\beta}^k\| \, \big|\mathbf{x}_n^T \frac{\boldsymbol{\beta}^k}{\|\boldsymbol{\beta}^k\|}\big|\big)}{\varphi(\|\boldsymbol{\beta}^k\|)}. \end{aligned} \tag{6.4}$$

If $r \in (0, 1]$ then, in view of Lemma 2(a), inequality (6.4) acquires the form

$$0 \geq \sum_{n=1}^{N-M} |\mathbf{x}_n^T \boldsymbol{\beta}^0|^r - \sum_{n=N-M+1}^{N} |\mathbf{x}_n^T \boldsymbol{\beta}^0|^r,$$

which implies that $M \geq \mathscr{M}(\mathbf{X}, r)$. Thus, $\varepsilon^*(\hat{\boldsymbol{\beta}}, \mathbf{y}, \mathbf{X}) \geq \mathscr{M}(\mathbf{X}, r)/N$ in this case.

Suppose now that $r = 0$. If $t_k \to \infty$ and $u_k \to 0$, then

$$0 \le \liminf_{k\to\infty} \frac{\varphi(u_k t_k)}{\varphi(t_k)} \le \limsup_{k\to\infty} \frac{\varphi(u_k t_k)}{\varphi(t_k)} \le \lim_{k\to\infty} \frac{\varphi(t_k)}{\varphi(t_k)} = 1, \tag{6.5}$$

by (A). Since $\mathbf{x}_n^T \boldsymbol{\beta}^0 = 0$ for at most $\mathscr{N}(\mathbf{X})$ of the $\mathbf{x}_n$'s, (6.4), (6.5) and Lemma 2(b) entail $0 \ge N - M - \mathscr{N}(\mathbf{X}) - M$, hence $M \ge (N - \mathscr{N}(\mathbf{X}))/2$, and consequently, in view of Proposition 1,

$$\varepsilon^*(\hat{\boldsymbol{\beta}}, \mathbf{y}, \mathbf{X}) \ge \frac{1}{N}\left\lfloor \frac{N - \mathscr{N}(\mathbf{X}) + 1}{2} \right\rfloor = \frac{\mathscr{M}(\mathbf{X}, 0)}{N}.$$

Upper bound. We consider the case $r = 0$ first. Here we see at once that the upper bound (1.1) and Proposition 1 imply that

$$\varepsilon^*(\hat{\boldsymbol{\beta}}, \mathbf{y}, \mathbf{X}) \le \frac{1}{N}\left\lfloor \frac{N - \mathscr{N}(\mathbf{X}) + 1}{2} \right\rfloor = \frac{\mathscr{M}(\mathbf{X}, 0)}{N}.$$

Now, let $r \in (0, 1]$. Suppose that $M = \mathscr{M}(\mathbf{X}, r) + 1$. If $\mathscr{N}(\mathbf{X}) = N$, the upper bound (1.1) shows that there is nothing to prove; hence suppose that $\mathscr{N}(\mathbf{X}) < N$. By Lemma 3, there is a set $E \subseteq \{1, 2, \ldots, N\}$, with $\operatorname{card} E = M$, and $\boldsymbol{\beta}^1 \ne \mathbf{0}$ such that

$$\sum_{n\in E} |\mathbf{x}_n^T \boldsymbol{\beta}^1|^r > \sum_{n\notin E} |\mathbf{x}_n^T \boldsymbol{\beta}^1|^r. \tag{6.6}$$

Without restriction of generality we may now assume that $E = \{N - M + 1, N - M + 2, \ldots, N\}$. Let $\boldsymbol{\beta}^k = k\boldsymbol{\beta}^1$. Define $\mathbf{y}^k$ to satisfy $y_n^k = y_n$ for $n = 1, 2, \ldots, N - M$ and $y_n^k = \mathbf{x}_n^T \boldsymbol{\beta}^k$ for $n \in E$.

Suppose that there is a compact $\mathscr{K}$ such that $\tilde{\boldsymbol{\beta}}^k = \hat{\boldsymbol{\beta}}(\mathbf{y}^k, \mathbf{X}) \in \mathscr{K}$ for all k. Then,

$$\begin{aligned} 0 < d_k &= D(\boldsymbol{\beta}^k, \mathbf{y}^k, \mathbf{X}) - D(\tilde{\boldsymbol{\beta}}^k, \mathbf{y}^k, \mathbf{X}) \\ &= \sum_{n=1}^{N-M} \varphi(y_n - \mathbf{x}_n^T \boldsymbol{\beta}^k) - \sum_{n=1}^{N-M} \varphi(y_n - \mathbf{x}_n^T \tilde{\boldsymbol{\beta}}^k) \\ &\quad + \sum_{n=N-M+1}^{N} \varphi(0) - \sum_{n=N-M+1}^{N} \varphi(\mathbf{x}_n^T \boldsymbol{\beta}^k - \mathbf{x}_n^T \tilde{\boldsymbol{\beta}}^k). \end{aligned}$$

Since $\tilde{\boldsymbol{\beta}}^k$, and consequently $\mathbf{x}_n^T \tilde{\boldsymbol{\beta}}^k$ are in a compact set, dividing by $\varphi(k)$ and letting $k \to \infty$ yields, in view of (A) and Lemma 2(a), that

$$\begin{aligned} 0 &\le \lim_{k\to\infty} \frac{d_k}{\varphi(k)} \\ &= \sum_{n=1}^{N-M} \lim_{k\to\infty} \frac{\varphi(k(\mathbf{x}_n^T \boldsymbol{\beta}^1 - (1/k) y_n))}{\varphi(k)} - \sum_{n=N-M+1}^{N} \lim_{k\to\infty} \frac{\varphi(k(\mathbf{x}_n^T \boldsymbol{\beta}^1 - (1/k)\mathbf{x}_n^T \tilde{\boldsymbol{\beta}}^k))}{\varphi(k)} \\ &= \sum_{n=1}^{N-M} |\mathbf{x}_n^T \boldsymbol{\beta}^1|^r - \sum_{n=N-M+1}^{N} |\mathbf{x}_n^T \boldsymbol{\beta}^1|^r. \end{aligned}$$

This is in contradiction with (6.6); hence, $\varepsilon^*(\hat{\boldsymbol{\beta}}, \mathbf{y}, \mathbf{X}) \le M/N$.

PROOF OF THEOREM 2. We have only to show that, in view of (2.2),

$$\varepsilon^*(\hat{\boldsymbol{\beta}}_1, \mathbf{y}, \mathbf{X}) \leq \frac{\mathscr{M}(\mathbf{X}, 1)}{N}. \tag{6.7}$$

A way to do that is to establish (6.7) first for $\mathbf{y} = \mathbf{0}$ and then finish the proof of Theorem 2 by verifying that the L_1 estimator is of uniform variation. We say that an estimator $\hat{\boldsymbol{\beta}}$ is of *uniform variation*, if, given $\mathbf{X}$, for every $\delta > 0$ there is a $K > 0$ such that whenever $\|\mathbf{y}^1 - \mathbf{y}^2\| \leq \delta$, then for any $\boldsymbol{\beta}^1 \in \hat{\boldsymbol{\beta}}(\mathbf{y}^1, \mathbf{X})$ there is $\boldsymbol{\beta}^2 \in \hat{\boldsymbol{\beta}}(\mathbf{y}^2, \mathbf{X})$ such that $\|\boldsymbol{\beta}^1 - \boldsymbol{\beta}^2\| \leq K$. Uniform variation implies that the breakdown point $\varepsilon^*(\hat{\boldsymbol{\beta}}_1, \mathbf{y}, \mathbf{X})$ does not exceed the breakdown point at $\mathbf{y} = \mathbf{0}$. For the details of the proof, see Mizera and Müller (1996). □

Acknowledgments. We are very grateful to Jana Jurečková for valuable suggestions and discussions, and to her, Xuming He and Stephen Portnoy for exhaustive replies to our queries. The authors express their thanks to Viktor Kurotschka and Andrej Pázman and to the Deutsche Forschungsgemeinschaft and Slovak Grant Agency for Science for supporting their mutual visits of the Free University of Berlin and Comenius University.

REFERENCES

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation.* Cambridge Univ. Press.

COPAS, J. B. (1975). On the unimodality of the likelihood for the Cauchy distribution. *Biometrika* **62** 701–704.

ELLIS, S. P. and MORGENTHALER, S. (1992). Leverage and breakdown in L_1 regression. *J. Amer. Statist. Assoc.* **87** 143–148.

GABRIELSEN, G. (1982). On the unimodality of the likelihood for the Cauchy distribution: some comments. *Biometrika* **69** 677–678.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics—The Approach Based on Influence Functions.* Wiley, New York.

HE, X., JUREČKOVÁ, J., KOENKER, R. and PORTNOY, S. (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica* **58** 1195–1214.

HOAGLIN, D. C., MOSTELLER, F. and TUKEY, J. W. (eds.) (1983). *Understanding Robust and Exploratory Data Analysis.* Wiley, New York.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

HUBER, P. J. (1984). Finite sample breakdown of M- and P-estimators. *Ann. Statist.* **12** 119–126.

JUREČKOVÁ, J. (1981). Tail-behavior of location estimators. *Ann. Statist.* **9** 578–585.

LANGE, K. L., LITTLE R. J. A. and TAYLOR, J. M. G. (1989). Robust statistical modeling using the t distribution. *J. Amer. Statist. Assoc.* **84** 881–896.

MARONNA, R. A., BUSTOS, O. H. and YOHAI, V. J. (1979). Bias- and efficiency-robustness of general M-estimators for regression with random carriers. In *Smoothing Techniques for Curve Estimation. Lecture Notes in Math.* **757** 91–116. Springer, Berlin.

MILI, L. and COAKLEY, C. W. (1993). Robust estimation in structured linear regression. *Ann. Statist.* **24** 2593–2607.

MIZERA, I. (1994). On consistent M-estimators: tuning constants, unimodality and breakdown. *Kybernetika* **30** 289–300.

MIZERA, I. (1996). Weak continuity of redescending M-estimators of location with an unbounded objective function. In *PROBASTAT '94, Proceedings of the Second International Conference on Mathematical Statistics* (A. Pázman and V. Witkovský, eds.) 343–347.

MIZERA, I. and MÜLLER, CH. H. (1996). Breakdown points and variation exponents of robust *M*-estimators in linear models. Preprint A-22-96, Freie Univ. Berlin, Fachbereich Mathematik und Informatik, Ser. A (Mathematik). Available as No. 22 at ftp://ftp.math.fu-berlin.de/pub/math/publ/pre/1996/index.html.

MORGENTHALER, S. and TUKEY, J. W. (1991). *Configural Polysampling.* Wiley, New York.

MÜLLER, CH. H. (1995). Breakdown points for designed experiments. *J. Statist. Plann. Inference* **45** 413–427.

MÜLLER, CH. H. (1997). *Robust Planning and Analysis of Experiments. Lecture Notes in Statist.* **124**. Springer, New York.

RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

ROUSSEEUW, P. J. and LEROY, A. M. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.

RUPPERT, D. (1992). Computing *S* estimators for regression and multivariate location/dispersion. *J. Comput. Graph. Statist.* **1** 253–270.

DEPARTMENT OF PROBABILITY
AND STATISTICS
COMENIUS UNIVERSITY
MLYNSKÁ DOLINA
BRATISLAVA SK-84215
SLOVAKIA
E-MAIL: mizera@fmph.uniba.sk

INSTITUTE OF MATHEMATICAL
STOCHASTICS
GEORG-AUGUST-UNIVERSITY GÖTTINGEN
LOTZESTRASSE 13
GÖTTINGEN D-37083
GERMANY
E-MAIL: chmuelle@math.uni-goettingen.de