On the Breakdown Properties of Two M-Functionals of Scatter

Lutz Duembgen* and David Tyler
University of Heidelberg and Rutgers University

draft, September 1997

Abstract. Let P be a probability distribution on $\mathbf{R}^q$. A detailed study of the breakdown properties of two M-functionals of scatter, $P \mapsto \Sigma(P)$ and $P \mapsto \Sigma(P \ominus P)$, is given. Here $\Sigma(\cdot)$ denotes Tyler's (1987) M-functional of scatter, taking values in the set of symmetric, positive definite $q \times q$ matrices. It assumes zero as a given center of the underlying distributions. The second functional avoids this assumption by operating on the symmetrized distribution $P \ominus P := \mathcal{L}(\mathbf{x} - \mathbf{y})$ with independent random vectors $\mathbf{x}, \mathbf{y} \sim P$. Let P be smooth in the sense of assigning probability zero to hyperplanes. Then:

(1) The (contamination) breakdown point of $P \mapsto \Sigma(P)$ equals $1/q$.

(2) The breakdown point of $P \mapsto \Sigma(P \ominus P)$ equals $1 - \sqrt{1 - 1/q} \in \,]1/(2q), 1/q[$.

(3) If we restrict attention to "tight contamination", then the breakdown point of $P \mapsto \Sigma(P \ominus P)$ equals $\sqrt{1/q} > 1/q$.

In all three cases the sources of breakdown are investigated. It turns out that breakdown is only caused by rather special contaminating distributions that are concentrated near low-dimensional subspaces.

Keywords and phrases: breakdown, coplanar contamination, tight contamination, M-functional, scatter matrix, symmetrization

*Research supported in part by European Union Human Capital and Mobility Program ERB CHRX-CT 940693.

1 Introduction

Let P, Q be nondegenerate probability distributions on $\mathbf{R}^q$. The covariance matrix of P,

$$\int (x-\mu(P))(x-\mu(P))^\top P(dx) \quad \text{with } \mu(P) := \int x\, P(dx),$$

is known to be very sensitive to small perturbations in P. Various robust surrogates for the covariance functional have been proposed. In the present paper we investigate the breakdown properties of two particular M-functionals of scatter.

The first one is Tyler's (1987) M-functional $\Sigma(P)$, which is defined as follows: Let $\mathbf{M}$ be the set of symmetric matrices in $\mathbf{R}^{q\times q}$, and let $\mathbf{M}^+$ be the set of all positive definite $M \in \mathbf{M}$. For $M \in \mathbf{M}^+$ let

$$G(P,M) \ := \ \frac{q}{P(\mathbf{R}^q \setminus \{0\})} \int_{\mathbf{R}^q\setminus\{0\}} \frac{M^{-1/2}xx^\top M^{-1/2}}{x^\top M^{-1}x}\, P(dx),$$

which is a nonnegative definite matrix in $\mathbf{M}$ with trace q. If there is a unique matrix $M \in \mathbf{M}^+$ with

$$G(P,M) \ = \ I \quad \text{and} \quad \operatorname{trace}(M) \ = \ q,$$

then we define $\Sigma(P) := M$. Otherwise we define arbitrarily $\Sigma(P) := 0$. In what follows we utilize the following two properties of Σ; see Kent and Tyler (1988) and Duembgen (1996).

Proposition 1.1 *Let $\mathcal{V}$ be the set of linear subspaces V of $\mathbf{R}^q$ with $1 \le \dim(V) < q$. Suppose that $P\{0\} = 0$. Then $\Sigma(P) \in \mathbf{M}^+$ if, and only if,*

(1.1) $$P(V) \ < \ \frac{\dim(V)}{q} \quad \text{for all } V \in \mathcal{V}.$$

If $G(P,M) = I$ for some matrix $M \in \mathbf{M}^+$ but $P(V) \ge \dim(V)/q$ for some space $V \in \mathcal{V}$, then there is a second space $W \in \mathcal{V}$ such that $V \cap W = \{0\}$ and $P(V \cup W) = 1$.

Proposition 1.2 *Suppose that $P\{0\} = 0$ and $\Sigma(P) \in \mathbf{M}^+$. Then*

(1.2) $$\Sigma(Q) \ \to \ \Sigma(P) \quad \text{as } Q \to P.$$

Here and throughout the sequel the space of probability measures on $\mathbf{R}^q$ is equipped with the topology of weak convergence.

The definition of $\Sigma(P)$ assumes zero as a given and known "center" of P. The second M-functional investigated here is $\Sigma(P \ominus P)$, where generally

$$P \ominus Q \ := \ \mathcal{L}(\mathbf{x} - \mathbf{y}) \quad \text{with independent random vectors } \mathbf{x} \sim P, \mathbf{y} \sim Q.$$

This modified functional, proposed in Duembgen (1996), avoids assumptions on or estimation of location parameters.

A quantity describing the robustness of the functional $P \mapsto \Sigma(P)$ is its contamination breakdown point (cf. Huber 1981). This is defined to be the supremum $\epsilon_*(P)$ of all $\epsilon \in \,]0,1[$ such that

$$\Sigma(Q) \in \mathbf{M}^+ \text{ for all } Q \in \mathcal{U}(P,\epsilon) \quad \text{and} \quad \sup_{Q \in \mathcal{U}(P,\epsilon)} \frac{\lambda_1(\Sigma(Q))}{\lambda_q(\Sigma(Q))} \ < \ \infty.$$

Here $\mathcal{U}(P,\epsilon)$ denotes the contamination neighborhood

$$\mathcal{U}(P,\epsilon) \ := \ \Big\{(1-\epsilon)P + \epsilon H : H \text{ some distribution on } \mathbf{R}^q\Big\}$$

of P, and $\lambda_1(M) \geq \lambda_2(M) \geq \cdots \geq \lambda_q(M)$ denote the ordered eigenvalues of $M \in \mathbf{M}$. If P is "smooth" in the sense that

$$P(H) \ = \ 0 \quad \text{for any hyperplane } H \subset \mathbf{R}^q, \tag{1.3}$$

then it turns out that

$$\epsilon_*(P) \ = \ \frac{1}{q}.$$

This result is known to several people, though it never appeared in a journal. Our purpose is not only to give a general expression for $\epsilon_*(P)$ and a precise proof but also to investigate the case $\epsilon = \epsilon_*(P)$ in more detail. Namely, in this case it turns out that for any sequence of distributions $Q_k = (1-\epsilon)P + \epsilon H_k \in \mathcal{U}(P,\epsilon)$ with $\Sigma(Q_k) \in \mathbf{M}^+$, the condition numbers $(\lambda_1/\lambda_q)(\Sigma(Q_k))$ tend to infinity if, and only if, the distributions H_k are concentrated near suitable linear subspaces of $\mathbf{R}^q$. In accordance with Tyler (1986) we call this "coplanar contamination".

Analogous considerations are made for $P \mapsto \Sigma(P \ominus P)$. The breakdown point $\epsilon_*^{\rm s}(P)$ of this functional is defined as $\epsilon_*(P)$ with $\Sigma(Q \ominus Q)$ in place of $\Sigma(Q)$. In case of smooth distributions P,

$$\epsilon_*^{\rm s}(P) \;=\; 1 - \sqrt{1 - \frac{1}{q}} \;\in\; \Big]\frac{1}{2q}, \frac{1}{q}\Big[.$$

In case of $\epsilon = \epsilon_*^{\rm s}(P)$, the conditions on a sequence of distributions $Q_k = (1-\epsilon)P + \epsilon H_k \in \mathcal{U}(P, \epsilon)$ in order to achieve unbounded condition numbers $(\lambda_1/\lambda_q)(\Sigma(Q_k \ominus Q_k))$ are even more restrictive. In particular, a necessary condition is that

$$|\mathbf{y}_k| \;\to_{\rm p}\; \infty \quad (k \to \infty) \quad \text{if } \mathbf{y}_k \sim H_k.$$

This observation is important, because coplanar contamination "at infinity" is easier to detect than arbitrary coplanar contamination. This leads to the question about breakdown caused by "tight" contamination. That means, we replace the neighborhood $\mathcal{U}(P, \epsilon)$ with

$$\begin{aligned}\mathcal{U}(P, \epsilon \,|\, \Lambda) \;:=\; \Big\{ &(1-\epsilon)P + \epsilon H : H \text{ some distribution on } \mathbf{R}^q \\ &\text{such that } H\{x : |x| > r\} \le \Lambda(r) \text{ for all } r > 0 \Big\},\end{aligned}$$

where Λ is some continuous function from $[0, \infty]$ into $[0, 1]$ with $\Lambda(0) = 1$ and $\Lambda(\infty) = 0$. In case of $P \mapsto \Sigma(P)$ replacing $\mathcal{U}(P, \epsilon)$ with $\mathcal{U}(P, \epsilon \,|\, \Lambda)$ does not alter the breakdown point. However, let $\epsilon_*^{\rm s}(P \,|\, \Lambda)$ be defined as $\epsilon_*^{\rm s}(P)$ with $\mathcal{U}(P, \epsilon \,|\, \Lambda)$ in place of $\mathcal{U}(P, \epsilon)$. Then it turns out that for smooth P,

$$\epsilon_*^{\rm s}(P \,|\, \Lambda) \;=\; \sqrt{\frac{1}{q}}.$$

All proofs are deferred to Section 4.

2 The breakdown properties of $P \mapsto \Sigma(P)$

Condition (1.1) is equivalent to $\beta(P) > 0$, where

$$\beta(P) \;:=\; \min_{V \in \mathcal{V}} \frac{\Big(\dim(V)/q - P(V)\Big)^+}{1 - P(V)} \;\in\; \Big[0, \frac{1}{q}\Big]$$

with $0/0 := 0$. That this minimum is well-defined follows from Lemma 4.1 in Section 4. The set of all $V \in \mathcal{V}$ such that $[\dim(V)/q - P(V)]^+/(1 - P(V))$ equals $\beta(P)$ is denoted by $\mathcal{V}(P)$. Another useful abbreviation is

$$\Pi P \ := \ \frac{1}{2}\Big(\mathcal{L}\Big(|\mathbf{x}|^{-1}\mathbf{x}\,\Big|\,\mathbf{x}\neq 0\Big)+\mathcal{L}\Big(-|\mathbf{x}|^{-1}\mathbf{x}\,\Big|\,\mathbf{x}\neq 0\Big)\Big) \quad \text{with } \mathbf{x}\sim P.$$

This is a symmetric distribution on the unit sphere $\mathbf{S}^{q-1}$ of $\mathbf{R}^q$. Note that $G(P,\cdot)\equiv G(\Pi P,\cdot)$ and thus $\Sigma(P)=\Sigma(\Pi P)$.

Theorem 2.1 *Suppose that $\Sigma(P)\in\mathbf{M}^+$. Let $P=P\{0\}\delta_0+(1-P\{0\})P_o$, where δ_x denotes Dirac measure in $x\in\mathbf{R}^q$ and P_o is a distribution on $\mathbf{R}^q\setminus\{0\}$. Then*

$$\epsilon_*(P) \ = \ \begin{cases} \dfrac{(1-P\{0\})\beta(P_o)}{1-P\{0\}\beta(P_o)} & \textit{in general,} \\[2ex] \beta(P) & \textit{if } P\{0\}=0, \\[2ex] \dfrac{1}{q} & \textit{if P is smooth in the sense of (1.3).} \end{cases}$$

Suppose that $\epsilon=\epsilon_*(P)$. For any $Q=(1-\epsilon)P+\epsilon H$ in $\mathcal{U}(P,\epsilon)$, one has $\Sigma(Q)=0$ if, and only if, $H\{0\}=0$ and $H(V)=1$ for some $V\in\mathcal{V}(P_o)$. Moreover, for $k\geq 1$ let $Q_k=(1-\epsilon)P+\epsilon H_k\in\mathcal{U}(P,\epsilon)$ such that $\Sigma(Q_k)\in\mathbf{M}^+$. Then $\lim_{k\to\infty}(\lambda_1/\lambda_q)(\Sigma(Q_k))=\infty$ if, and only if, the following two conditions are satisfied:

$$\lim_{k\to\infty} H_k\{0\} \ = \ 0; \tag{2.1}$$

(2.2) *any cluster point $\widetilde{H}$ of $(\Pi H_k)_k$ is supported by some $V\in\mathcal{V}(P_o)$.*

An interesting fact is that the M-estimators introduced by Maronna (1976) have breakdown point at most $1/(q+1)$; cf. Stahel (1981) or Tyler (1986).

3 The breakdown properties of $P \mapsto \Sigma(P \ominus P)$

Theorem 3.1 *Suppose that $\Sigma(P \ominus P) \in \mathbf{M}^+$. Then*

$$\epsilon_*^{\mathrm{s}}(P) = \begin{cases} 1 - \sqrt{\dfrac{1-\beta(\Pi(P\ominus P))}{1 - P\ominus P\{0\}\beta(\Pi(P\ominus P))}} & \textit{in general}, \\ 1 - \sqrt{1-\beta(\Pi(P\ominus P))} & \textit{if } P \textit{ has no atoms}, \\ 1 - \sqrt{1 - \dfrac{1}{q}} & \textit{if } P \textit{ is smooth in the sense of (1.3).} \end{cases}$$

Suppose that $\epsilon = \epsilon_*^{\mathrm{s}}(P)$. Then $\Sigma(Q \ominus Q) \in \mathbf{M}^+$ for any Q in $\mathcal{U}(P, \epsilon)$. Moreover, for $k \geq 1$ let $Q_k = (1-\epsilon)P + \epsilon H_k \in \mathcal{U}(P,\epsilon)$. Then $\lim_k (\lambda_1/\lambda_q)(\Sigma(Q_k \ominus Q_k)) = \infty$ if, and only if, the following three conditions are satisfied:

$$\lim_{k\to\infty} \max_{x \in \mathbf{R}^q} H_k\{x\} = 0; \tag{3.1}$$

$$|\mathbf{y}_k| \to_{\mathrm{p}} \infty \quad (k \to \infty) \quad \textit{where } \mathbf{y}_k \sim H_k; \tag{3.2}$$

$$\begin{array}{c} \textit{for any cluster point } (\widetilde{H}_1, \widetilde{H}_2) \textit{ of } \Big((\Pi H_k, \Pi(H_k \ominus H_k))\Big)_k \textit{ there is} \\ \textit{a space } V \in \mathcal{V}(\Pi(P \ominus P)) \textit{ such that } \widetilde{H}_1(V) = \widetilde{H}_2(V) = 1. \end{array} \tag{3.3}$$

This theorem shows that symmetrization lowers the breakdown point of the M-functional. However, the type of contamination required in order to cause breakdown of the functional $P \mapsto \Sigma(P \ominus P)$ is far more special than in case of $P \mapsto \Sigma(P)$. The quantity $\beta(\Pi(P \ominus P))$ difficult to compute. However, for $V \in \mathcal{V}$,

$$\begin{aligned} \Pi(P\ominus P)(V) &= \frac{P\ominus P(V) - P\ominus P\{0\}}{1 - P\ominus P\{0\}} \\ &\leq P \ominus P(V) \\ &\leq \max_{x\in\mathbf{R}^q} P(x+V), \end{aligned}$$

so that

$$\beta(\Pi(P\ominus P)) \geq \beta^{\mathrm{s}}(P) := \min_{x \in \mathbf{R}^q, V \in \mathcal{V}} \frac{\Big(\dim(V)/q - P(x+V)\Big)^+}{1 - P(x+V)}.$$

Here is the result on tight contamination mentioned in the introduction.

Theorem 3.2 *Suppose that $\Sigma(P \ominus P) \in \mathbf{M}^+$. Then*

$$\epsilon_*^{\mathrm{s}}(P \,|\, \Lambda) \begin{cases} \geq \sqrt{\beta^{\mathrm{s}}(P)} & \text{in general,} \\ = \sqrt{\dfrac{1}{q}} & \text{if } P \text{ is smooth in the sense of (1.3).} \end{cases}$$

Suppose that P satisfies (1.3). Then $\Sigma(Q \ominus Q) = 0$ for $Q = (1-\epsilon)P + \epsilon H \in \mathcal{U}(P, \epsilon \,|\, \Lambda)$ if, and only if, H has no atoms and H is supported by some one-dimensional affine subspace of $\mathbf{R}^q$. Similarly, for $k \geq 1$ let $Q_k = (1-\epsilon)P + \epsilon H_k \in \mathcal{U}(P, \epsilon)$ such that $\Sigma(Q_k \ominus Q_k) \in \mathbf{M}^+$. Then $\lim_{k\to\infty}(\lambda_1/\lambda_q)(\Sigma(Q_k \ominus Q_k)) = \infty$ if, and only if, the following two conditions are satisfied:

(3.4) $$\lim_{k\to\infty} \max_{x \in \mathbf{R}^q} H_k\{x\} \;=\; 0;$$

(3.5) *any cluster point of $(\Pi(H_k \ominus H_k))_k$ is supported by some $V \in \mathcal{V}$ with $\dim(V) = 1$.*

One can easily show that Condition (3.5) implies that any cluster point of $(H_k)_k$ is supported by some one-dimensional affine subspace of $\mathbf{R}^q$.

4 Proofs

Lemma 4.1 *For $0 \leq d < q$ let $\mathcal{V}(d)$ be the set of all d-dimensional linear subspaces of $\mathbf{R}^q$. Then both*

$$\max_{V \in \mathcal{V}(d)} Q(V) \quad \text{and} \quad \max_{x \in \mathbf{R}^q, V \in \mathcal{V}(d)} Q(x + V)$$

are well-defined and upper semicontinuous in Q.

Proof of Lemma 4.1. Let $(Q_k)_k$ be any sequence of distributions converging weakly to some Q. Let $V_k \in \mathcal{V}(d)$ and $x_k \in \mathbf{R}^q$ such that either

(4.1) $$x_k \;=\; 0 \quad \text{and} \quad Q_k(V_k) \;>\; \sup_{V \in \mathcal{V}(d)} Q_k(V) - k^{-1},$$

or

(4.2) $$x_k \;\in\; V_k^{\perp} \quad \text{and} \quad Q_k(x_k + V_k) \;>\; \sup_{x \in \mathbf{R}^q, V \in \mathcal{V}(d)} Q_k(x + V) - k^{-1}.$$

Let $M_k \in \mathbf{M}$ describe the orthogonal projection from $\mathbf{R}^q$ onto V_k. After replacing $(Q_k)_k$ with a subsequence if necessary, one may assume that $(M_k)_k$ converges to some projection matrix M, and we define $V := M\mathbf{R}^q$. Further one may assume that

$$\lim_{k\to\infty} |x_k| \ = \ \infty \quad \text{or} \quad \lim_{k\to\infty} x_k \ = \ x \in \mathbf{R}^q.$$

Since $x_k + V_k \subset \{y : |y| \geq |x_k|\}$ one easily deduces from $\lim_k Q_k = Q$ and $\lim_k |x_k| = \infty$ that $\lim_k Q_k(x_k + V_k) = 0$. If $\lim_k x_k = x$, then for any $R > 0$,

$$\begin{aligned} \limsup_{k\to\infty} Q_k(x_k + V_k) &\leq \lim_{k\to\infty} \int \Big(1 - R|y - M_k y - x_k|\Big)^+ Q_k(dy) \\ &= \int \Big(1 - R|y - My - x|\Big)^+ Q(dy) \\ &\to Q(x + V) \quad (R \to \infty). \end{aligned}$$

These considerations show that $\sup_{V\in\mathcal{V}(d)} Q(V)$ and $\sup_{x\in\mathbf{R}^q, V\in\mathcal{V}(d)} Q(x+V)$ are upper semicontinuous in Q. In the special case $(Q_k)_k \equiv Q$ one realizes that both suprema are attained. □

Propositions 1.1 and 1.2 entail the following two facts:

Lemma 4.2 (a) *Let $\mathcal{Q}$ be a familiy of nondegenerate distributions on $\mathbf{R}^q$ such that $\Sigma(Q) \in \mathbf{M}^+$ for all $Q \in \mathcal{Q}$ and let $\{\Pi Q : Q \in \mathcal{Q}\}$ be closed. Then*

$$\sup_{Q\in\mathcal{Q}} \frac{\lambda_1}{\lambda_q}(\Sigma(Q)) \ < \ \infty.$$

(b) *Let $(Q_k)_k$ be a sequence of nondegenerate distributions on $\mathbf{R}^q$ such that $\Sigma(Q_k) \in \mathbf{M}^+$ for all k and*

$$\lim_{k\to\infty} \Pi Q_k \ = \ \widetilde{Q}, \quad \lim_{k\to\infty} \frac{\lambda_1}{\lambda_q}(\Sigma(Q_k)) \ = \ \rho \in [1, \infty].$$

If $\rho = \infty$, then $\widetilde{Q}(V) \geq \dim(V)/q$ for some $V \in \mathcal{V}$. If $\rho < \infty$ but $\widetilde{Q}(V) \geq \dim(V)/q$ for some space $V \in \mathcal{V}$, then there is a second space $W \in \mathcal{V}$ such that $V \cap W = \{0\}$ and $\widetilde{Q}(V \cup W) = 1$.

Proof of Lemma 4.2. As for part (a), Prohorov's Theorem implies that $\{\Pi Q : Q \in \mathcal{Q}\}$ is even compact. Since $\Sigma(Q) = \Sigma(\Pi Q) \in \mathbf{M}^+$ for all $Q \in \mathcal{Q}$, Proposition 1.2 yields

$$\sup_{Q\in\mathcal{Q}} \frac{\lambda_1}{\lambda_q}(\Sigma(Q)) \;=\; \max_{Q\in\mathcal{Q}} \frac{\lambda_1}{\lambda_q}(\Sigma(\Pi Q)) \;<\; \infty.$$

In part (b) suppose first that $\widetilde{Q}(V) < \dim(V)/q$ for all $V \in \mathcal{V}$. Then $\Sigma(\widetilde{Q}) \in \mathbf{M}^+$ by Proposition 1.1, and $\Sigma(\widetilde{Q}) = \lim_k \Sigma(Q_k)$ by Proposition 1.2, whence $\rho = (\lambda_1/\lambda_q)(\Sigma(\widetilde{Q})) < \infty$.

Now suppose that $\rho < \infty$. After replacing $(Q_k)_k$ with a subsequence if necessary, one may assume that $\lim_k \Sigma(Q_k) = M \in \mathbf{M}^+$. But then

$$I \;=\; \lim_{k\to\infty} G(\Pi Q_k, \Sigma(Q_k)) \;=\; G(\widetilde{Q}, M),$$

because $G(\cdot, \Sigma(Q_k))$ converges uniformly to $G(\cdot, M)$ as $k \to \infty$. Thus if $\widetilde{Q}(V) \geq \dim(V)/q$ for some $V \in \mathcal{V}$, then the second part of Proposition 1.1 says that $V \cap W = \{0\}$ and $\widetilde{Q}(V \cup W) = 1$ for some $W \in \mathcal{V}$. □

Proof of Theorem 2.1. Note first that $\Big\{\Pi Q : Q \in \mathcal{U}(P,\epsilon)\Big\}$ is equal to the closed set

$$\Big\{(1-\epsilon_o)\Pi P + \epsilon_o \widetilde{H} : \widetilde{H} \text{ any symmetric distribution on } \mathbf{S}^{q-1}\Big\},$$

where

$$\epsilon_o \;:=\; \frac{\epsilon}{1-(1-\epsilon)P\{0\}}.$$

For if $Q = (1-\epsilon)P + \epsilon H \in \mathcal{U}(P,\epsilon)$, then

$$\Pi Q \;=\; \frac{(1-\epsilon)(1-P\{0\})\Pi P + \epsilon(1-H\{0\})\Pi H}{(1-\epsilon)(1-P\{0\}) + \epsilon(1-H\{0\})} \;=\; (1-\epsilon')\Pi P + \epsilon' \widetilde{H}$$

for some symmetric distribution $\widetilde{H}$ on $\mathbf{S}^{q-1}$ and

$$\epsilon' \;:=\; \frac{\epsilon(1-H\{0\})}{(1-\epsilon)(1-P\{0\}) + \epsilon(1-H\{0\})} \;\leq\; \epsilon_o.$$

Further,

$$\Pi Q(V) \;\leq\; (1-\epsilon_o)\Pi P(V) + \epsilon_o \;=\; (1-\epsilon_o)P_o(V) + \epsilon_o$$

with equality if, and only if, $H\{0\} = 0$ and $H(V) = 1$. This is strictly smaller than $\dim(V)/q$ if, and only if,

$$\epsilon_o \ < \ \frac{\dim(V)/q - P_o(V)}{1 - P_o(V)}.$$

Hence we can concude the following: If $\epsilon_o < \beta(P_o)$ then $\Sigma(\cdot) \in \mathbf{M}^+$ on $\mathcal{U}(P, \epsilon)$, and Lemma 4.2 (a) yields that $(\lambda_1/\lambda_q)(\Sigma(\cdot))$ is bounded on $\mathcal{U}(P, \epsilon)$. If $\epsilon_o = \beta(P_o)$, then $\Sigma(Q) = 0$ for $Q = (1-\epsilon)P + \epsilon H \in \mathcal{U}(P, \epsilon)$ if, and only if, $H\{0\} = 0$ and $H(V) = 1$ for some $V \in \mathcal{V}(P_o)$. Since ϵ_o is strictly increasing in ϵ, inverting the equation $\epsilon_o = \beta(P_o)$ yields

$$\epsilon_*(P) \ = \ \frac{(1 - P\{0\})\beta(P_o)}{1 - P\{0\}\beta(P_o)}.$$

Let $\epsilon = \epsilon_*$ and $Q_k = (1-\epsilon)P + \epsilon H_k \in \mathcal{U}(P, \epsilon)$ as stated in the theorem. After replacing $(Q_k)_k$ with a subsequence if necessary, one may assume that $\lim_k H_k\{0\} = a \in [0, 1]$, $\lim_k \Pi H_k = \widetilde{H}$ (where $\Pi\delta_0$ may be defined arbitrarily) and $\lim_k (\lambda_1/\lambda_q)(\Sigma(Q_k)) = \rho \in [1, \infty]$. This implies that

$$\lim_{k\to\infty} \Pi Q_k \ = \ \widetilde{Q} := \frac{(1-\epsilon)(1 - P\{0\})\Pi P + \epsilon(1-a)\widetilde{H}}{(1-\epsilon)(1 - P\{0\}) + \epsilon(1-a)}.$$

Since $\Sigma(P) \in \mathbf{M}^+$, $\Pi P(V \cup W) < 1$ for arbitrary $V, W \in \mathcal{V}$ with $V \cap W = \{0\}$. The limit distribution $\widetilde{Q}$ inherits this property. Thus one can apply Lemma 4.2 (b) and conclude that $\rho = \infty$ if, and only if, $\widetilde{Q}(V) \geq \dim(V)/q$ for some $V \in \mathcal{V}$. But for any $V \in \mathcal{V}$,

$$\begin{aligned}
\widetilde{Q}(V) \ &= \ \frac{(1-\epsilon)(1 - P\{0\})P_o(V) + \epsilon(1-a)\widetilde{H}(V)}{(1-\epsilon)(1 - P\{0\}) + \epsilon(1-a)} \\
&\leq \ \frac{(1-\epsilon)(1 - P\{0\})P_o(V) + \epsilon(1-a)}{(1-\epsilon)(1 - P\{0\}) + \epsilon(1-a)} \\
&\leq \ \frac{(1-\epsilon)(1 - P\{0\})P_o(V) + \epsilon}{(1-\epsilon)(1 - P\{0\}) + \epsilon} \\
&= \ (1-\epsilon_o)P_o(V) + \epsilon_o \\
&\leq \ (1-\epsilon_o)\frac{\dim(V)/q - \beta(P_o)}{1 - \beta(P_o)} + \epsilon_o \\
&= \ \dim(V)/q
\end{aligned}$$

with equality if, and only if, $\widetilde{H}(V) = 1$, $a = 0$ and $V \in \mathcal{V}(P_o)$. □

The following preliminary result for the proof of Theorems 3.1 and 3.2 describes the possible limits of a sequence $(\Pi(P \ominus H_k))_k$.

Proposition 4.3 *Let $(H_k)_{k\geq 1}$ be a sequence of distributions on $\mathbf{R}^q$. A pair $(a, \widetilde{B})$ is cluster point for the sequence $\Big((P \ominus H_k\{0\}, \Pi(P \ominus H_k))\Big)_{k\geq 1}$ if, and only if, it can be represented as follows:*

$$a \;=\; \sum_{x\in\mathbf{R}^q} P\{x\}a_x$$

and

$$\widetilde{B} \;=\; \frac{\eta\widetilde{B}_\infty + \sum_{x\in\mathbf{R}^q} P\{x\}\Big((1-\eta)H\{x\} - a_x\Big)\widetilde{B}_x + (1-\eta)(1 - P\ominus H\{0\})\Pi(P\ominus H)}{1 - \sum_{x\in\mathbf{R}^q} P\{x\}a_x}$$

with

$$\begin{aligned}
&\eta \;:=\; \lim_{r\to\infty}\ \liminf_{k\to\infty}\ H_k\{x : |x| > r\},\\
&\textit{some distribution } H \textit{ on } \mathbf{R}^q,\\
&\textit{numbers } a_x \in [0, (1-\eta)H\{x\}] \textit{ and}\\
&\textit{symmetric distributions } \widetilde{B}_\infty \textit{ and } \widetilde{B}_x \textit{ on } \mathbf{S}^{q-1}.
\end{aligned}$$

Proof of Proposition 4.3. We compactify $\mathbf{R}^q$ via the mapping

$$x \;\mapsto\; \psi(x) \;:=\; (1+|x|)^{-1}x \;\in\; U(0,1),$$

where $U(y,\delta)$ and $B(y,\delta)$ denote, respectively, the open and closed ball around $y \in \mathbf{R}^q$ with radius $\delta \geq 0$. Without loss of generality one may assume that

$$\lim_{k\to\infty} H_k \circ \psi^{-1} \;=\; D \quad \text{and} \quad \eta \;=\; D(\mathbf{S}^{q-1}).$$

Even if D is concentrated on $U(0,1)$ the Continuous Mapping Theorem is not applicable to $\Pi(P \ominus H_k)$, because the points in $\mathbf{X} := \Big\{x \in \mathbf{R}^q : D\{\psi(x)\} > 0\Big\}$ require special attention. Since

$$D\{\psi(x)\} \;=\; \lim_{\delta\downarrow 0}\ \liminf_{k\to\infty}\ H_k U(x,\delta) \;=\; \lim_{\delta\downarrow 0}\ \limsup_{k\to\infty}\ H_k B(x,\delta)$$

for any $x \in \mathbf{X}$ and

$$\eta \;=\; \lim_{r\downarrow\infty}\ \liminf_{k\to\infty}\ H_k(\mathbf{R}^q \setminus B(0,r)) \;=\; \lim_{r\downarrow\infty}\ \limsup_{k\to\infty}\ H_k(\mathbf{R}^q \setminus U(0,r)),$$

one can find numbers $\delta_{x,k} \geq 0$ and $r_k > 0$ such that with $U_{x,k} := U(x, \delta_{x,k})$ and $U_{\infty,k} := \mathbf{R}^q \setminus B(0, r_k)$ the following requirements are met:

$$\begin{aligned}
\lim_{k\to\infty} \delta_{x,k} &= 0 \quad \text{and} \quad \lim_{k\to\infty} H_k U_{x,k} = D\{\psi(x)\} \quad \text{for } x \in \mathbf{X}, \\
\lim_{k\to\infty} r_k &= \infty \quad \text{and} \quad \lim_{k\to\infty} H_k U_{\infty,k} = \eta, \\
U_{x,k} \cap U_{y,k} &= \emptyset \quad \text{for different } x, y \in \mathbf{X} \cup \{\infty\}.
\end{aligned}$$

After replacing $(H_k)_k$ with a suitable subsequence if necessary, one may assume further that for any $x \in \mathbf{X}$,

$$\begin{aligned}
\lim_{k\to\infty} H_k\{x\} &= a_x \in [0, D\{\psi(x)\}], \\
\lim_{k\to\infty} \Pi\mathcal{L}\Big(x - \mathbf{y}_k \,\Big|\, \mathbf{y}_k \in U_{x,k} \setminus \{x\}\Big) &= \widetilde{B}_x \quad \text{if } \mathbf{y}_k \sim H_k.
\end{aligned}$$

Since $\lim_k H_k\{x\} = 0$ whenever $D\{\psi(x)\} = 0$, this implies that

$$\lim_{k\to\infty} P \ominus H_k\{0\} = \sum_{x\in\mathbf{X}} P\{x\} a_x. \tag{4.3}$$

Further we write $D = \eta\widetilde{B}_\infty + (1-\eta)H \circ \psi^{-1}$ with distributions $\widetilde{B}_\infty$ on $\mathbf{S}^{q-1}$ and H on $\mathbf{R}^q$. Now let

$$f(x) := \begin{cases} g(|x|^{-1}x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

for some even, continuous function g on $\mathbf{S}^{q-1}$, and let $\mathbf{x} \sim P$, $\mathbf{y}_k \sim H_k$ and $\mathbf{y} \sim H$ be independent. Then, as $k \to \infty$,

$$\begin{aligned}
\mathbb{E} f(\mathbf{x} - \mathbf{y}_k) &= \mathbb{E}\, 1\{\mathbf{y}_k \in U_{\infty,k}\} f(\mathbf{x} - \mathbf{y}_k) + \mathbb{E}\, 1\{\mathbf{y}_k \notin U_{\infty,k}\} f(\mathbf{x} - \mathbf{y}_k) \\
&= \eta \int g\, d\widetilde{B}_\infty + \mathbb{E}\, 1\{\mathbf{y}_k \notin U_{\infty,k}\} f(\mathbf{x} - \mathbf{y}_k) + o(1) \\
&= \eta \int g\, d\widetilde{B}_\infty + \sum_{x\in\mathbf{X}} P\{x\}\, \mathbb{E}\, 1\{\mathbf{y}_k \in U_{x,k} \setminus \{x\}\} f(x - \mathbf{y}_k) \\
&\quad + \sum_{x\in\mathbf{X}} P\{x\}\, \mathbb{E}\, 1\{\mathbf{y}_k \notin U_{x,k} \cup U_{\infty,k}\} f(x - \mathbf{y}_k) \\
&\quad + \mathbb{E}\, 1\{\mathbf{x} \notin \mathbf{X}, \mathbf{y}_k \notin U_{\infty,k}\} f(\mathbf{x} - \mathbf{y}_k) + o(1) \\
&= \eta \int g\, d\widetilde{B}_\infty + \sum_{x\in\mathbf{X}} P\{x\} \Big(D\{\psi(x)\} - a_x\Big) \int g\, d\widetilde{B}_x \\
&\quad + (1-\eta) \sum_{x\in\mathbf{X}} P\{x\}\, \mathbb{E}\, 1\{\mathbf{y} \neq x\} f(x - \mathbf{y})
\end{aligned}$$

$$
\begin{aligned}
& + (1-\eta)\,\mathbb{E}\,1\{\mathbf{x}\notin\mathbf{X}\}f(\mathbf{x}-\mathbf{y}) + o(1) \\
= \;& \eta\int g\,d\widetilde{B}_\infty + \sum_{x\in\mathbf{X}} P\{x\}\Big(D\{\psi(x)\} - a_x\Big)\int g\,d\widetilde{B}_x \\
& + (1-\eta)(1-P\ominus H\{0\})\int g\,d\Pi(P\ominus H).
\end{aligned}
$$

Together with (4.3) this shows that $(\Pi(P\ominus H_k))_k$ converges weakly to a distribution $\widetilde{B}$ as stated in the proposition. □

Proof of Theorem 3.1. The clue to the proof is a detailed study of the closure of the set $\Big\{\Pi(Q\ominus Q) : Q\in\mathcal{U}(P,\epsilon)\Big\}$. For $k\geq 1$ let $Q_k = (1-\epsilon)P + \epsilon H_k \in \mathcal{U}(P,\epsilon)$ such that

$$
\lim_{k\to\infty} \Pi(Q_k\ominus Q_k) \;=\; \widetilde{Q}.
$$

By compactness arguments one may assume without loss of generality that

$$
\begin{aligned}
\lim_{k\to\infty} P\ominus H_k\{0\} &= a_{PH}, \\
\lim_{k\to\infty} \Pi(P\ominus H_k) &= \widetilde{B}_{PH}, \\
\lim_{k\to\infty} H_k\ominus H_k\{0\} &= a_{HH}, \\
\lim_{k\to\infty} \Pi(H_k\ominus H_k) &= \widetilde{B}_{HH}.
\end{aligned}
$$

With $a_{PP} := P\ominus P\{0\}$ and $\widetilde{B}_{PP} := \Pi(P\ominus P)$ one obtains the representation

$$
\widetilde{Q} \;=\; \frac{(1-\epsilon^2)(1-a_{PP})\widetilde{B}_{PP} + 2\epsilon(1-\epsilon)(1-a_{PH})\widetilde{B}_{PH} + \epsilon^2(1-a_{HH})\widetilde{B}_{HH}}{(1-\epsilon^2)(1-a_{PP}) + 2\epsilon(1-\epsilon)(1-a_{PH}) + \epsilon^2(1-a_{HH})}.
$$

Note first that $\widetilde{Q} = (1-\epsilon')\widetilde{B}_{PP} + \epsilon'\widetilde{H}$ for some symmetric distribution $\widetilde{H}$ on $\mathbf{S}^{q-1}$ and

$$
\begin{aligned}
\epsilon' \;&:=\; \frac{2\epsilon(1-\epsilon)(1-a_{PH}) + \epsilon^2(1-a_{HH})}{(1-\epsilon^2)(1-a_{PP}) + 2\epsilon(1-\epsilon)(1-a_{PH}) + \epsilon^2(1-a_{HH})} \\
&\leq\; \frac{2\epsilon-\epsilon^2}{(1-\epsilon^2)(1-a_{PP}) + 2\epsilon-\epsilon^2} \;=\; \frac{1-(1-\epsilon)^2}{1-(1-\epsilon)^2 a_{PP}} \\
&=:\; \epsilon_o.
\end{aligned}
$$

Thus $\Big\{\Pi(Q\ominus Q) : Q\in\mathcal{U}(P,\epsilon)\Big\}$ is *contained* in the closed set

$$
\Big\{(1-\epsilon_o)\Pi(P\ominus P) + \epsilon_o\widetilde{H} : \widetilde{H} \text{ any symmetric distribution on } \mathbf{S}^{q-1}\Big\}.
$$

Consequently, $\Sigma(Q \ominus Q) \in \mathbf{M}^+$ for all $Q \in \mathcal{U}(P,\epsilon)$ with $\sup_{Q\in\mathcal{U}(P,\epsilon)}(\lambda_1/\lambda_q)(\Sigma(Q\ominus Q))$ being finite, provided that $\epsilon_o < \beta(\Pi(P\ominus P))$, which is equivalent to

$$\epsilon \ < \ \epsilon_{**}^{\mathrm{s}}(P) := 1 - \sqrt{\frac{1-\beta(\Pi(P\ominus P))}{1-P\ominus P\{0\}\beta(\Pi(P\ominus P))}}.$$

Now suppose that $\epsilon = \epsilon_{**}^{\mathrm{s}}(P)$, that means, $\epsilon_o = \beta(\widetilde{B}_{PP})$. Then $\widetilde{Q}(V) \ge \dim(V)/q$ for some $V \in \mathcal{V}$ if, and only if, $a_{PH} = a_{HH} = 0$, $\widetilde{B}_{PH}(V) = \widetilde{B}_{HH}(V) = 1$ and $V \in \mathcal{V}(\Pi(P\ominus P))$. These equations cannot hold if $\widetilde{Q} = \Pi(Q\ominus Q)$ for some $Q = (1-\epsilon)P + \epsilon H \in \mathcal{U}(P,\epsilon)$. For then

$$\widetilde{B}_{PH}(V) \ = \ P\ominus H(V) \ \le \ \max_{x\in\mathbf{R}^q} P(x+V) \ < \ 1,$$

because otherwise $P(x+V) = 1$ for some $x \in \mathbf{R}^q$, so that $P\ominus P(V) = 1$ and $\Sigma(P\ominus P) = 0$.

The equation $a_{HH} = 0$ is equivalent to Condition (3.1) and entails that $a_{PH} = 0$ as well. Moreover, Proposition 4.3 implies that

$$\widetilde{B}_{HP} \ = \ \eta\widetilde{B}_\infty + (1-\eta)\sum_{x\in\mathbf{R}^q} P\{x\}H\{x\}\widetilde{B}_x + (1-\eta)(1-P\ominus H\{0\})\Pi(P\ominus H)$$

for some distribution H on $\mathbf{R}^q$, some number $\eta \in [0,1]$ and symmetric distributions $\widetilde{B}_y$, $y \in \mathbf{R}^q \cup \{\infty\}$, on $\mathbf{S}^{q-1}$. This representation shows that $\widetilde{B}_{PH}(V) = 1$ for some $V \in \mathcal{V}(\Pi(P\ominus P))$ if, and only if, $\eta = 1$ and $\widetilde{B}_\infty = \widetilde{B}_{PH} = \lim_k \Pi H_k$ is concentrated on V. Together with the requirement $\widetilde{B}_{HH}(V) = 1$ we end up with Conditions (3.2, 3.3) about the sequence $(H_k)_k$. All requirements (3.1, 3.2, 3.3) are satisfied, for instance, by $H_k := \mathcal{L}(k\mathbf{y})$, where $\mathbf{y}$ is some random vector whose distribution is concentrated on V but has no atoms. Thus $\epsilon_{**}^{\mathrm{s}}(P) = \epsilon_*^{\mathrm{s}}(P)$. □

Proof of Theorem 3.2. Let $Q_k = (1-\epsilon)P + \epsilon H_k \in \mathcal{U}(P,\epsilon\,|\,\Lambda)$, and let $\widetilde{Q}$, a_{PP}, a_{PH}, a_{HH}, $\widetilde{B}_{PP}$, $\widetilde{B}_{PH}$, $\widetilde{B}_{HH}$ be as in the proof of Theorem 3.1. Since the sequence $(H_k)_k$ is tight by definition of $\mathcal{U}(P,\epsilon\,|\,\Lambda)$, Proposition 4.3 yields that $a_{PH} = \sum_{x\in\mathbf{R}^q} P\{x\}a_x$ and

$$\widetilde{B}_{PH} \ = \ \frac{\sum_{x\in\mathbf{R}^q} P\{x\}(H\{x\}-a_x)\widetilde{B}_x + (1-P\ominus H\{0\})\Pi(P\ominus H)}{1-\sum_{x\in\mathbf{R}^q} P\{x\}a_x}$$

for some distribution H on $\mathbf{R}^q$, numbers $a_x \in [0, H\{x\}]$ and symmetric distributions B_y, $y \in \mathbf{R}^q \cup \{\infty\}$, on $\mathbf{S}^{q-1}$. Thus for any $V \in \mathcal{V}$,

$$(1-a_{PH})\widetilde{B}_{PH}(V) \ \ \le \ \ \sum_{x\in\mathbf{R}^q} P\{x\}(H\{x\}-a_x) + (1-P\ominus H\{0\})\Pi(P\ominus H)(V)$$

$$
\begin{aligned}
&= P \ominus H(V) - a_{PH}, \\
&\leq \max_{x \in \mathbf{R}^q} P(x+V) - a_{PH}, \\
(1 - a_{PP})\widetilde{B}_{PP}(V) &= P \ominus P(V) - a_{PP}, \\
&\leq \max_{x \in \mathbf{R}^q} P(x+V) - a_{PP},
\end{aligned}
$$

whence

$$
\begin{aligned}
\widetilde{Q}(V) &\leq \frac{(1-\epsilon^2)\max_{x\in\mathbf{R}^q} P(x+V) + \epsilon^2 - (1-\epsilon)^2 a_{PP} - 2\epsilon(1-\epsilon)a_{PH} - \epsilon^2 a_{HH}}{1 - (1-\epsilon)^2 a_{PP} - 2\epsilon(1-\epsilon)a_{PH} - \epsilon^2 a_{HH}} \\
&\leq (1-\epsilon^2)\max_{x\in\mathbf{R}^q} P(x+V) + \epsilon^2.
\end{aligned}
$$

This shows that

$$
\epsilon_*^{\mathrm{s}}(P \mid \Lambda) \;\leq\; \sqrt{\beta^{\mathrm{s}}(P)}.
$$

In case of P being smooth,

$$
\widetilde{Q}(V) \;=\; \frac{\epsilon^2(1-a_{HH})\widetilde{B}_{HH}(V)}{1-\epsilon^2+\epsilon^2(1-a_{HH})} \;\leq\; \epsilon^2
$$

with equality if, and only if, $a_{HH} = 0$ and $\widetilde{B}_{HH}(V) = 1$. □

References

Duembgen, L. (1996). On Tyler's M-functional of scatter in high dimension. Preprint, tentatively accepted for publication in *Ann. Inst. Statist. Math.*

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

Kent, J.T. and D.E. Tyler (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *J. Appl. Statist.* **15**, 247-254

Kent, J.T. and D.E. Tyler (1991). Redescending M-estimates of multivariate location and scatter. *Ann. Statist.* **19**, 2102-2119

Maronna, R.A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4**, 51-67

Stahel, W.A. (1981). Breakdown of covariance estimators. Research report 31, Fachgruppe Statistik, ETH Zürich

Tyler, D.E. (1986). Breakdown properties of the M-estimators of multivariate scatter. Technical report, Rutgers University

Tyler, D.E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.* **15**, 234-251

Lutz Duembgen
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
D-69120 Heidelberg
Germany
lutz@statlab.uni-heidelberg.de

David Tyler
Department of Statistics
Rutgers University
Hill Center, Busch Campus
Piscataway, NJ 08854
U.S.A.
dtyler@rci.rutgers.edu