Statistica Sinica **14**(2004), 297-316

INFLUENCE FUNCTION AND ASYMPTOTIC EFFICIENCY OF THE AFFINE EQUIVARIANT RANK COVARIANCE MATRIX

Esa Ollila, Christophe Croux and Hannu Oja

Helsinki University of Technology, University of Leuven and University of Jyväskylä

Abstract: Visuri, Koivunen and Oja (2003) proposed and illustrated the use of the affine equivariant rank covariance matrix (RCM) in classical multivariate inference problems. The RCM was shown to be asymptotically multinormal but explicit formulas for the limiting variances and covariances were not given. In this paper the influence functions and the limiting variances and covariances of the RCM and the corresponding scatter estimate are derived in the multivariate elliptical case. Limiting efficiencies are given in the multivariate normal and t distribution cases. The estimates based on the RCM are highly efficient in the multinormal case, and for heavy-tailed distribution, perform better than those based on the regular covariance matrix. Finite-sample and asymptotic efficiency comparisons to a selected redecending M-estimator and S-estimator are reported.

Key words and phrases: Elliptical distribution, influence function, multivariate analysis, multivariate rank, scatter matrix.

1. Introduction

Ranks and signs are frequently used in statistical analysis to obtain procedures which are less sensitive to the model assumptions. Computing statistical quantities based on ranks instead of on the original observations can result in more robust methods. When observations are multivariate, it is not so obvious how “sign” and “rank” are to be defined. In this article, the affine equivariant multivariate extension of the concept of rank as proposed by Brown and Hettmansperger (1987) is considered. This concept of rank is based on the Oja (1983) median, and has been successfully applied to multivariate one-sample, two-sample and multisample location problems. For example, nonparametric and robust competitors of MANOVA have been developed in Hettmansperger, Möttönen and Oja (1998). For a review of statistical methods based on ranks we refer to Hettmansperger and McKean (1998) and Oja (1999). For a recent review of other techniques of multivariate ordering, we refer to Serfling (2002) and the references therein.

The approach based on multivariate signs and ranks has recently been extended to other classical multivariate inference problems, such as principal component analysis, canonical correlation analysis and multivariate regression analysis. These developments are based on the affine equivariant multivariate sign and rank covariance matrices, as defined in Visuri, Koivunen and Oja (2000). For the affine equivariant sign covariance matrix (SCM), the asymptotic distribution and asymptotic variances were obtained by Ollila, Oja and Croux (2003). Knowledge of the limit distribution of the SCM allowed to obtain asymptotic results for multivariate regression based on the SCM (Ollila, Hettmansperger and Oja (2002)). Multivariate inference based on the affine equivariant rank covariance matrix (RCM) was outlined and illustrated in Visuri et al. (2003). Their simulation studies and examples showed that the estimates based on the RCM enjoy very good efficiency properties, at the price of not being robust to extreme outliers. The asymptotic variances of the RCM, however, were not derived.

The main contribution of this paper are that the limiting variances of the RCM. Moreover, we also obtain an expression for the influence function of the RCM. This influence function is seen to be approximately linear, in contrast with the influence function of the regular covariance matrix, the latter being quadratic. The RCM is therefore more robust than the classical covariance matrix, but still has an unbounded influence function. Despite that, we show that the RCM remains quite efficient at heavy-tailed distributions.

In Section 2 the concept and properties of the affine equivariant rank based on the Oja objective function (1983) are briefly reviewed. The RCM and corresponding scatter matrix estimator are defined in Section 3. Influence functions of the estimators at elliptical model distributions are given in Section 4. The limiting variances and covariances of the estimates in the elliptical case are presented in Section 5. A comparison with the limiting efficiencies of a chosen redescending M-estimator and an S-estimator is included. Finite-sample efficiencies based on simulations are reported in Section 6. The paper is closed with some final comments in Section 7.

2. Affine Equivariant Ranks

Let $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ be a k-variate data set. Then the volume of the k-variate simplex determined by $k+1$ vertices $\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_k}$ (k data points) and $\boldsymbol{x}$ is a constant $(1/k!)$ times

$$V(\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_k}, \boldsymbol{x}) = \text{abs}\left\{ \det \begin{pmatrix} 1 & \ldots & 1 & 1 \\ \boldsymbol{x}_{i_1} & \ldots & \boldsymbol{x}_{i_k} & \boldsymbol{x} \end{pmatrix} \right\}, \tag{1}$$

and the affine equivariant Oja (1983) median minimizes the criterion function

$$V(\boldsymbol{x}; X) = \text{ave}\{V(\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_k}, \boldsymbol{x})\},$$

where the average is taken over all possible k-subsets $\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_k}$ with $1 \le i_1 < \ldots < i_k \le n$. The multivariate centered *rank function* is defined as the gradient

$$\boldsymbol{R}(\boldsymbol{x}; X) = \nabla_{\boldsymbol{x}} V(\boldsymbol{x}; X).$$

Note that in the univariate case $V(x; X) = \text{ave}|x - x_i|$, the mean deviation, and leads to the univariate median and the univariate centered rank function $R(x; X) = \text{ave}\{\text{sign}(x - x_i)\}$. The *observed ranks* $\boldsymbol{R}_i = \boldsymbol{R}(\boldsymbol{x}_i; X)$, $i = 1, \ldots, n$, are centered, so $\sum_i \boldsymbol{R}_i = \boldsymbol{0}$. They are affine equivariant in the sense that if the ranks $\boldsymbol{R}_i^*$ are calculated from the transformed observations $\boldsymbol{x}_i^* = A\boldsymbol{x}_i + \boldsymbol{b}$, with a nonsingular $k \times k$-matrix A and a k-vector $\boldsymbol{b}$, then $\boldsymbol{R}_i^* = \text{abs}\,\{\det(A)\}(A^{-1})^T \boldsymbol{R}_i$. The population counterparts are as follows. If $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ is a random sample from a k-variate distribution with cdf F with finite first-order moments, then the expected volume of the simplex is a constant $(1/k!)$ times $V(\boldsymbol{x}; F) = E_F[V(\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_k}, \boldsymbol{x})]$. The multivariate centered *population rank function* is then

$$\boldsymbol{R}(\boldsymbol{x}; F) = \nabla_{\boldsymbol{x}} V(\boldsymbol{x}; F). \tag{2}$$

Naturally the population rank function is also affine equivariant. The empirical rank function $\boldsymbol{R}(\boldsymbol{x}; X)$ converges uniformly in probability to the population rank function $\boldsymbol{R}(\boldsymbol{x}; F)$. In the univariate case $R(x; F) = 2F(x) - 1$. See Hettmansperger, Möttönen and Oja (1998), Oja (1999) and Visuri et al. (2003).

A k-dimensional random vector $\boldsymbol{x}$ with a cdf $F_{\mu,\Sigma}$ has an *elliptically symmetric distribution* if its density function is of the form

$$f(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = \{\det(\Sigma)\}^{-1/2} f_0\{(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\} \tag{3}$$

for some $\boldsymbol{\mu} \in \mathbb{R}^k$, a positive definite $k \times k$-matrix Σ, and a nonnegative function f_0 which is independent of $\boldsymbol{\mu}$ and Σ. If $\boldsymbol{\mu} = \boldsymbol{0}$ and $\Sigma = I_k$ (the $k \times k$ identity matrix) then $\boldsymbol{x}$ is said to have a *spherically symmetric distribution* centered at zero. Throughout the paper we denote $G = F_{\boldsymbol{0}, I_k}$. The parameter $\boldsymbol{\mu}$ is the symmetry center of the distribution of $\boldsymbol{x}$ and it equals the expected value $E(\boldsymbol{x})$ when it exists, whereas the parameter Σ is proportional to the covariance matrix of $\boldsymbol{x}$ when it exists, $\text{Cov}\,(\boldsymbol{x}) = (E_G(\|\boldsymbol{x}\|^2)/k)\Sigma$. Elliptical distributions are often used for studying robustness of multivariate statistics. For this purpose the k-variate t-distribution with $\nu > 0$ degrees of freedom, $t_{k,\nu}$, is particularly interesting as it yields distributions with varying heaviness of the tails. To be more specific, if $\boldsymbol{x} \sim t_{k,\nu}$, then f_0 at (3) is of the form $f_0(t) = c_{k,\nu}(1 + t/\nu)^{-(k+\nu)/2}$, where

$c_{k,\nu}$ is a normalizing constant. The value $\nu = 1$ corresponds to k-variate Cauchy distribution whereas the limiting value $\nu \to \infty$ yields the multivariate standard normal density. For $\nu > 2$ the covariance matrix of the k-variate t-distribution is $\text{Cov}(\boldsymbol{x}) = \{\nu/(\nu-2)\}\Sigma$. In the multivariate normal distribution case, $f_0(t) = (2\pi)^{-k/2}\exp(-t/2)$ and $\text{Cov}(\boldsymbol{x}) = \Sigma$. A good review of elliptical distribution is given in Muirhead (1982).

Consider now the population rank function at a spherical model distribution G. If $\boldsymbol{x}$ follows a spherically symmetric distribution G, then its radius $r = \|\boldsymbol{x}\|$ and direction $\boldsymbol{u} = \boldsymbol{x}/\|\boldsymbol{x}\|$ are independent and $\boldsymbol{u}$ is uniformly distributed on the unit sphere (Muirhead (1982, Theorem 1.5.6)). Now $V(\boldsymbol{x}; G)$ depends on $\boldsymbol{x} = r\boldsymbol{u}$ only through r. Hence, in this case, we may write $V(\boldsymbol{x}; G) = V_0(r; G)$. Then the population rank function at G is simply

$$\boldsymbol{R}(\boldsymbol{x}; G) = \nabla_x V_0(r; G) = q(r; G)\,\boldsymbol{u} \tag{4}$$

with $q(r; G) = V_0'(r; G)$. Expressions for the functions $V_0(r; G)$ and $q(r; G)$ at spherical normal and t-distributions are given in Lemma 1 of the Appendix. In the elliptical case, thanks to affine equivariance, the population rank function of F at $\boldsymbol{x} = \Sigma^{1/2}\boldsymbol{z} + \boldsymbol{\mu}$ is

$$\boldsymbol{R}(\boldsymbol{x}; F) = \text{abs}\{\det(\Sigma^{1/2})\}\Sigma^{-1/2}\boldsymbol{R}(\boldsymbol{z}; G).$$

3. The Rank Covariance Matrix

Let $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$ be the observed ranks for a k-variate data set $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$. The *rank covariance matrix* (RCM) is then $\hat{D} = \text{ave}\{\boldsymbol{R}_i\boldsymbol{R}_i^T\}$. Since the ranks are centered, the RCM is nothing more than the covariance matrix computed from ranks instead of from the original observations. It is affine equivariant in the sense that if the RCM $\hat{D}^*$ is calculated from the transformed observations $\boldsymbol{x}_i^* = A\boldsymbol{x}_i + \boldsymbol{b}$, with nonsingular $k \times k$-matrix A and k-vector $\boldsymbol{b}$, then $\hat{D}^* = \det(A)^2(A^{-1})^T\hat{D}A^{-1}$. Visuri et al. (2003) showed that if $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ is a random sample from a k-variate distribution F with finite first-order moments, then the rank covariance matrix converges in probability to the *population rank covariance matrix*,

$$D(F) = E_F[\boldsymbol{R}(\boldsymbol{x}; F)\boldsymbol{R}^T(\boldsymbol{x}; F)]. \tag{5}$$

Consider now a spherical distribution G. Then (4) together with $E_G[\boldsymbol{u}\boldsymbol{u}^T] = I_k/k$ yields

$$D(G) = (c_G^2/k)\,I_k, \tag{6}$$

where $c_G^2 = E_G[q^2(r;G)]$. For values of c_G^2 in multivariate spherical normal and t distribution cases, see Lemma 2 in the Appendix. If $\boldsymbol{z} \sim G$, then $\boldsymbol{x} = \Sigma^{1/2}\boldsymbol{z} + \boldsymbol{\mu}$ has elliptical cdf $F = F_{\mu,\Sigma}$. Using (6) and the affine equivariance, we get

$$D(F) = \det(\Sigma)\Sigma^{-1/2}\, D(G)\, \Sigma^{-1/2} = (c_G^2/k)\det(\Sigma)\Sigma^{-1}. \tag{7}$$

Thus, at elliptical models, the RCM is proportional to the inverse of Σ (and also to the inverse of the covariance matrix $\mathrm{Cov}(\boldsymbol{x})$ when it exists). (This is also true in the wider class of location-scale models, Visuri et al. (2003).) The inverse of the RCM is therefore an estimator of a multiple of the scatter matrix Σ and we may say that it estimates the shape of Σ.

But RCM also carries information about the size of the data cloud, and one can construct an affine equivariant *scatter matrix functional* $C(F)$ based on the RCM:

$$C(F) = (k\det\{D(F)\}/c_G^2)^{1/(k-1)}D(F)^{-1}. \tag{8}$$

It is immediate to check that $C(F) = \Sigma$ at elliptical distributions $F = F_{\mu,\Sigma}$, so C is Fisher consistent for Σ at elliptical models. The *scatter matrix estimate* based on the RCM, $\hat{C} = (k\det(\hat{D})/c_G^2)^{1/(k-1)}\hat{D}^{-1}$, is then a consistent estimator of Σ of the elliptical population $F = F_{\mu,\Sigma}$. Moreover, the above estimator is affine equivariant in the sense that $\hat{C}^*$ computed from the transformed observations $\boldsymbol{x}_i^* = A\boldsymbol{x}_i + \boldsymbol{b}$ satisfies $\hat{C}^* = A\hat{C}A^T$. Therefore we call $\hat{C}$ a scatter matrix estimator, which can be compared with other estimators of multivariate scatter (see Maronna and Yohai (1998) for an overview of different scatter matrix estimators).

4. Influence Functions in the Elliptical Case

Next we derive the influence functions of the RCM functional D and the associated scatter functional C at elliptical models. For that, consider the contaminated distribution $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_{\boldsymbol{x}}$, where $\Delta_{\boldsymbol{x}}$ is a distribution putting all its mass at $\boldsymbol{x}$. Then the influence function is defined by (see Hampel, Ronchetti, Rousseeuw and Stahel (1986))

$$\mathrm{IF}\,(\boldsymbol{x}; T, F) = \lim_{\varepsilon\downarrow 0}\frac{T(F_\varepsilon) - T(F)}{\varepsilon} = \frac{\partial}{\partial\varepsilon}T(F_\varepsilon)|_{\varepsilon=0}.$$

The next theorem gives an expression for the influence function of the RCM at a spherical model distribution. The proof is quite technical and can be found in the Appendix.

Theorem 1. *The influence function of the RCM functional D at a k-variate spherical distribution G is*

$$\begin{aligned}\mathrm{IF}\,(\boldsymbol{x}; D, G) &= \{q^2(r;G) + \gamma(r;G) - \eta(r;G)\}\boldsymbol{u}\boldsymbol{u}^T - \Big\{2k + 1 - \frac{\eta(r;G)}{c_G^2/k}\Big\}D(G)\\ &= \alpha(r;G)\boldsymbol{u}\boldsymbol{u}^T - \beta(r;G)D(G),\end{aligned}$$

where $r = \|\boldsymbol{x}\|$, $\boldsymbol{u} = \|\boldsymbol{x}\|^{-1}\boldsymbol{x}$, $D(G) = (c_G^2/k)I_k$ *and* $q(r;G)$ *is defined by* (4). *Furthermore*

$$\gamma(r;G) = 2kE_G\left[\frac{(z_1-r)z_1}{\|\boldsymbol{z}_r\|\,\|\boldsymbol{z}\|}\Big\{V_0(\rho_z;G_{k-1}) - \rho_z\, q(\rho_z;G_{k-1})\Big\}q(\|\boldsymbol{z}\|;G)\right],$$

$$\eta(r;G) = 2kE_G\left[\frac{z_2^2}{\|\boldsymbol{z}_r\|\,\|\boldsymbol{z}\|}\Big\{V_0(\rho_z;G_{k-1}) + \frac{(r^2-\rho_z^2)}{\rho_z}\, q(\rho_z;G_{k-1})\Big\}q(\|\boldsymbol{z}\|;G)\right],$$

where $\boldsymbol{z} = (z_1,\ldots,z_k)^T \sim G$, G_{k-1} *is the* $(k-1)$*-variate spherical distribution of* $(z_1,\ldots,z_{k-1})$, $\boldsymbol{z}_r = (z_1-r, z_2,\ldots,z_k)^T$ *and* $\rho_z^2 = r^2\{1-(z_1-r)^2/\|\boldsymbol{z}_r\|^2\}$.

The influence function of scatter functional $C(F)$ is obtained next. The influence function of any affine equivariant scatter estimator at a spherical model G may be expressed as

$$\text{IF}\,(\boldsymbol{x};C,G) = \tilde{\alpha}(\|\boldsymbol{x}\|;G)\boldsymbol{u}\boldsymbol{u}^T - \tilde{\beta}(\|\boldsymbol{x}\|;G)I_k, \tag{9}$$

with again $\boldsymbol{u} = \boldsymbol{x}/\|\boldsymbol{x}\|$, for some real-valued functions $\tilde{\alpha}$ and $\tilde{\beta}$ depending on the estimator and the model. See Hampel et al. (1986) and Croux and Haesbroeck (2000). Using Theorem 1 and (8), the functions $\tilde{\alpha}$ and $\tilde{\beta}$ of the RCM scatter functional C are easily obtained after using some matrix differentiation rules for the determinant and the inverse of a non-singular matrix (see Magnus and Neudecker (1988, Chap. 8, Theorems 1 and 3)).

Corollary 1. *The influence function of the scatter functional* C *at a* k*-variate spherical distribution* G *is determined by setting, in* (9),

$$\tilde{\alpha}(r;G) = -\frac{\alpha(r;G)}{(c_G^2/k)} \quad \textit{and} \quad \tilde{\beta}(r;G) = \frac{1}{k-1}\left\{2k+1-\frac{q^2(r;G)}{c_G^2/k} - \frac{\gamma(r;G)}{c_G^2/k}\right\}.$$

In Figure 1 we picture the functions $\tilde{\alpha}(r;G)$ and $\tilde{\beta}(r;G)$ of the regular covariance matrix estimator and the scatter estimator C based on the RCM at the bivariate normal model. Note that estimators are comparable; a correction factor is not needed for $\text{Cov}(\boldsymbol{x})$ in the normal case. For non-normal elliptical populations the regular covariance matrix estimator must be multiplied by a consistency factor $\sigma = k/E_G(\|\boldsymbol{x}\|^2)$ in order to estimate the scatter matrix Σ of the elliptical population at hand. Also plotted are the corresponding functions of Kent and Tyler's (1991) redescending M-estimator based on the t-distibution with $\nu = 5$ degrees of freedom, denoted by $M(t_5)$. In other words, $M(t_5)$ is the maximum likelihood estimator (MLE) for the $t_{k,5}$-distribution. It needs to be multiplied by a consistency factor (except at a $t_{k,5}$-distibution) in order to consistently estimate Σ. To get further insight in to the gains/losses of using RCM scatter estimator instead of high breakdown robust estimator, also plotted are the corresponding functions for the S-estimator (Rousseeuw and Leroy (1984), Davies (1987) and Lopuhaä (1989)) using Tukey's biweight ρ-function. The aforementioned S-estimator is defined so that the estimator has 25% breakdown point and consistency for Σ at the multivariate normal distribution. This estimator will be referred to as $S(25)$. Again $S(25)$ needs to be multiplied by a consistency factor (except at the normal). Expressions for the influence functions of M-estimators and S-estimators can be found in Maronna (1976), Huber (1981, Sec. 8.7) and Lopuhaä (1989).

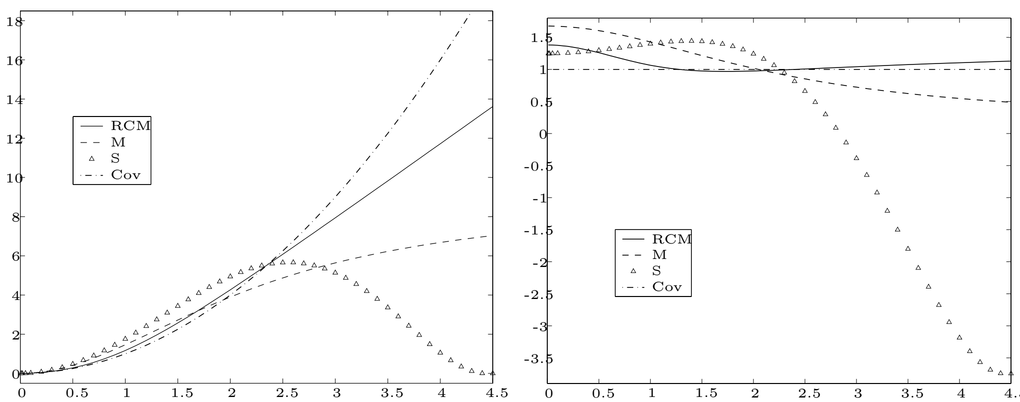

Figure 1. Functions $\tilde{\alpha}(r;G)$ (left panel) and $\tilde{\beta}(r;G)$ (right panel) of the RCM scatter estimator, the regular covariance estimator and the $M(t_5)$- and $S(25)$-estimator at the bivariate normal model ($G = \Phi_2$).

The function $\tilde{\alpha}$ may be interpreted as a weight function since the influence of $\boldsymbol{x}$ on an off-diagonal element C_{ij}, $i \neq j$, of C is $\text{IF}(\boldsymbol{x}; C_{ij}, G) = \tilde{\alpha}(r;G)u_i u_j$, where u_i and u_j are the ith and jth components of $\boldsymbol{u} = \boldsymbol{x}/\|\boldsymbol{x}\|$. This means that for boundedness of the influence function at an off-diagonal element, a necessary and sufficient condition is that the weight function $\tilde{\alpha}(r;G)$ be bounded. As we see from Figure 1, the $\tilde{\alpha}$-function of the regular covariance matrix is quadratic in the radius r, whereas that of $C(F)$ is approximately linear for large r. This suggests that the RCM scatter functional will give more protection to moderate outliers than the regular covariance matrix but it is not robust in the strict sense. The RCM resembles an L_1-based method: it is more robust than an L_2-based approach, very efficient (as we will see in the next section), but not highly robust. We also see from Figure 1, that the $\tilde{\alpha}$-function of the $M(t_5)$ is bounded whereas that of $S(25)$ is even redescending to zero, meaning that outliers with r larger than a certain rejection point will be given a zero weight.

The $\tilde{\beta}$-function is not as important, since it does not intervene in the influence function of the off-diagonal elements of C. For example, the influence function of the correlation matrix estimator associated with C will depend solely on $\tilde{\alpha}(r;G)$ (Croux and Haesbroeck (2000), Ollila, Oja and Croux (2003)). The influence function of a diagonal element C_{ii} of C will also depend on $\tilde{\alpha}$ because $\text{IF}\,(\boldsymbol{x};C_{ii},G) = \tilde{\alpha}(r;G)u_i^2 - \tilde{\beta}(r;G)$. Therefore boundedness of $\tilde{\beta}(r;G)$ is a necessary but not sufficient condition for boundedness of the influence function of a diagonal element. From Figure 1 we see that the $\tilde{\beta}$-functions for all the considered estimators are bounded, but the $\tilde{\beta}$-function of $S(25)$ differs from that of other estimators by giving a large negative weight to perturbation vectors $\boldsymbol{x}$ with large radius.

Finally, we note that the influence functions of the RCM functional D and the associated scatter matrix functional C at an elliptical distribution $F = F_{\mu,\Sigma}$ are given by

$$\text{IF}\,(\boldsymbol{x};D,F) = \det(\Sigma)\Sigma^{-1/2}\,\text{IF}\,(\Sigma^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu});D,G)\Sigma^{-1/2},$$
$$\text{IF}\,(\boldsymbol{x};C,F) = \Sigma^{1/2}\,\text{IF}\,(\Sigma^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu});C,G)\Sigma^{1/2},$$

due to affine equivariance of the functionals.

5. Limiting Variances and Covariances in the Elliptical Case

We start this section with some notations. We use "vec" as an operator working on matrices: $\text{vec}\,(A)$ vectorizes matrix A by stacking the columns of the matrix on top of each other. A commutation matrix, $I_{k,k}$, is a $k^2 \times k^2$ block matrix with (i,j)-block a $k \times k$ matrix having 1 at entry (j,i) and zero elsewhere. Finally, the Kronecker product of two $k \times k$ matrices A and B, denoted by $A \otimes B$, is a $k^2 \times k^2$-block matrix with $k \times k$-blocks, the (i,j)-block equal to $a_{ij}B$. For properties on Kronecker products, commutation matrices and the vec-operator, the reader is referred to Magnus and Neudecker (1988).

First, we note that exists a kernel $h(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{2k+1})$ such that $E[h(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{2k+1})] = D(F)$; see equation (11) in the Appendix. Equivalently, $D(F) = E[h^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{2k+1})]$, where $h^*(\cdot)$ is a kernel that is symmetric in its arguments, i.e.,

$$h^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{2k+1}) = \{(2k+1)!\}^{-1}\sum h(\boldsymbol{x}_{\pi(1)},\ldots,\boldsymbol{x}_{\pi(2k+1)}),$$

where the summation is over all permutations π of $\{1,\ldots,2k+1\}$. Thereby,

$$\text{IF}\,(\boldsymbol{x};D,F) = (2k+1)\{E[h^*(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{2k+1})\Gamma\boldsymbol{x}_{2k+1} = \boldsymbol{x}] - D(F)\}.$$

Visuri et al. (2003) showed that, if the observations come from a k-variate distribution with finite second-order moments, then $\hat{D}$ is asymptotically equivalent to a U-statistic U_n with kernel $h^*(\cdot)$. This means that $\sqrt{n}(\hat{D} - D) = \sqrt{n}(U_n - D) + R_n = \sqrt{n}\{(1/n)\sum_{i=1}^n \mathrm{IF}\,(\boldsymbol{x}_i; D, F)\} + R_n^*$, where the remainder terms satisfy $R_n, R_n^* \to_p 0$. It then follows that the limiting distribution of $\sqrt{n}\,\mathrm{vec}\,(\hat{D} - D)$ at spherical G is multivariate normal with zero mean and asymptotic variance-covariance matrix $\mathrm{ASV}\,(\hat{D}; G) = E[\mathrm{vec}\,\{\mathrm{IF}\,(\boldsymbol{x}; D, G)\}\,\mathrm{vec}\,\{\mathrm{IF}\,(\boldsymbol{x}; D, G)\}^T]$ if the second-order moments of G exist.

The structure of the influence function of D at spherical distributions, and the symmetry properties of G, imply that $\mathrm{ASV}\,(\hat{D}; G)$ will depend only on two numbers: $\mathrm{ASV}\,(\hat{D}_{11}; G)$ and $\mathrm{ASV}\,(\hat{D}_{12}; G)$. The asymptotic covariances between diagonal elements are all $\mathrm{ASC}\,(\hat{D}_{11}, \hat{D}_{22}; G) = \mathrm{ASV}\,(\hat{D}_{11}; G) - 2\,\mathrm{ASV}\,(\hat{D}_{12}; G)$, while all the other limiting covariances of the elements of $\hat{D}$ are zero. Similar developments naturally also hold true for $\hat{C}$ and the following holds.

Corollary 2. *The covariance matrices of the limiting distribution of* $\sqrt{n}\,\mathrm{vec}\,(\hat{D} - D)$ *and* $\sqrt{n}\,\mathrm{vec}\,(\hat{C} - C)$ *at a spherical distribution* G *with finite second-order moments are given by* $\mathrm{ASV}\,(\hat{D}_{12}; G)(I_{k^2} + I_{k,k}) + \mathrm{ASC}\,(\hat{D}_{11}, \hat{D}_{22}; G)\,\mathrm{vec}\,(I_k)\,\mathrm{vec}\,(I_k)^T$ *and* $\mathrm{ASV}\,(\hat{C}_{12}; G)(I_{k^2} + I_{k,k}) + \mathrm{ASC}\,(\hat{C}_{11}, \hat{C}_{22}; G)\mathrm{vec}(I_k)\mathrm{vec}(I_k)^T$, *respectively.*

In fact, the above developments for the asymptotic covariance matrices are valid for any asymptotically normal affine equivariant scatter matrix estimate $\hat{C}$ at spherical models, not only for the RCM-based one; See Tyler (1982, Theorem 1). Using the affine equivariance (and properties of vec-operator, commutation matrix and Kronecker product), the limiting covariance matrix of $\sqrt{n}\,\mathrm{vec}\,(\hat{D} - D)$ at an elliptical distribution $F = F_{\mu,\Sigma}$ is given by

$$\frac{k^2}{c_G^4}\{\mathrm{ASV}\,(\hat{D}_{12}; G)(I_{k^2} + I_{k,k})(D \otimes D) + \mathrm{ASC}\,(\hat{D}_{11}, \hat{D}_{22}; G)\,\mathrm{vec}\,(D)\,\mathrm{vec}\,(D)^T\}$$

and the limiting covariance matrix of $\sqrt{n}\,\mathrm{vec}\,(\hat{C} - \Sigma)$ is $\mathrm{ASV}\,(\hat{C}_{12}; G)(I_{k^2} + I_{k,k})(\Sigma \otimes \Sigma) + \mathrm{ASC}\,(\hat{C}_{11}, \hat{C}_{22}; G)\mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)^T$. See also Tyler (1982, Corollary 1).

Therefore we measure the asymptotic relative efficiency (ARE) of a scatter matrix estimate $\hat{C}$ with respect to the sample covariance matrix $\hat{\mathrm{Cov}}$ at an elliptical distribution by the two ratios

$$\mathrm{ARE}(\hat{C}_{11}, \hat{\mathrm{Cov}}_{11}; G) = \frac{\sigma^2\,\mathrm{ASV}\,(\hat{\mathrm{Cov}}_{11}; G)}{\mathrm{ASV}\,(\hat{C}_{11}; G)} \qquad \text{and}$$
$$\mathrm{ARE}(\hat{C}_{12}, \hat{\mathrm{Cov}}_{12}; G) = \frac{\sigma^2\,\mathrm{ASV}\,(\hat{\mathrm{Cov}}_{12}; G)}{\mathrm{ASV}\,(\hat{C}_{12}; G)},$$

being the ARE of diagonal and off-diagonal elements of the scatter matrix estimate $\hat{C}$ with respect to the sample covariance matrix $\hat{\mathrm{Cov}}$. Here $\sigma = k/E_G(|\boldsymbol{x}\|^2)$

is the consistency factor for Cov. In our case, the limiting variances are readily obtained from Theorem 1 and Corollary 1. For example, $\mathrm{ASV}(\hat{C}_{12};G) = (k^2/c_G^4)\,\mathrm{ASV}(\hat{D}_{12};G)$, where

$$\mathrm{ASV}(\hat{D}_{12};G) = E_G[\mathrm{IF}(\boldsymbol{x};D_{12},G)^2] = E_G[\alpha^2(\|\boldsymbol{x}\|;G)u_1^2u_2^2] = \frac{E_G[\alpha^2(\|\boldsymbol{x}\|;G)]}{k(k+2)},$$

which can be calculated using numerical integration or Monte-Carlo techniques.

In Table 1 the diagonal and off-diagonal ARE of the RCM-based scatter matrix estimator are obtained for the multivariate $t_{k,\nu}$-distribution. Dimensions $k = 2, 3, 5, 8$ and degrees of freedom $\nu = 5, 6, 8, 15$ are considered. Efficiencies for multivariate normal distributions ($\nu \to \infty$) are also given. We also compared with the corresponding efficiencies for the scatter estimator of Ollila, Oja and Croux (2003) based on the sign covariance matrix (SCM). This is interesting, as the multivariate signs are closely related to multivariate ranks and the resulting scatter estimator has similar robustness properties, e.g., unbounded influence function and zero breakdown point. Furthermore, the corresponding efficiencies for $M(t_5)$ and $S(25)$ were also calculated using efficient numerical integration whenever analytical expression could not be derived for their asymptotical variances. Asymptotic theory (limiting normality, asymptotic variance and covariance expressions, etc.) for the scatter estimator based on SCM, M-estimators and S-estimators will not be elaborated upon here; the reader is referred to Maronna (1976), Huber (1981), Tyler (1982), Hampel, Ronchetti, Rousseeuw and Stahel (1986), Davies (1987), Lopuhaä (1989), Kent and Tyler (1991) and Ollila, Oja and Croux (2003).

First we note from Table 1 that the AREs for the diagonal and off-diagonal elements of the RCM-based scatter matrix estimator are surprisingly high at the normal distribution. There is almost no loss in efficiency, all AREs being above 96%. The efficiencies of $M(t_5)$ and the high breakdown $S(25)$ increase with dimension, but for $k = 2$ and 3, the RCM estimator (and also the SCM estimator) has significantly better performance. For multivariate t-distributions, the RCM-based scatter matrix estimator outperforms the sample covariance matrix. The gain gets large when the degrees of freedom decrease, i.e., when the distribution gets heavier tails. This is true also for the SCM estimator of scatter and for $M(t_5)$ and $S(25)$. However, in the case of RCM and SCM scatter estimators, the efficiencies are decreasing with the dimension whereas they are increasing for $M(t_5)$ and $S(25)$. The heavier the tails of the t-distribution and the larger the dimension, the better is the performance of $M(t_5)$ and $S(25)$ when compared to sign and rank based estimators. The efficiencies for the RCM differ only slightly from those for the SCM estimator, the biggest difference is at the bivariate normal distribution. Note that $M(t_5)$ clearly outperform $S(25)$ in the cases $\nu = 5, 6$ and 8. (Recall that the $M(t_5)$ is the MLE for $\nu = 5$ and therefore most efficient in this case.)

Table 1. ARE of off-diagonal (panel A) and diagonal (panel B) elements of the RCM-based and SCM-based scatter estimators, $M(t_5)$ and $S(25)$ with respect to the sample covariance matrix at multivariate $t_{k,\nu}$-distributions. The column $\nu = \infty$ corresponds to the normal distribution.

		A.	Degrees of freedom ν				**B.**	Degrees of freedom ν			
k		5	6	8	15	∞	5	6	8	15	∞
2	RCM	2.050	1.470	1.200	1.060	0.990	2.330	1.610	1.270	1.080	0.980
	SCM	2.000	1.447	1.184	1.031	0.956	2.286	1.589	1.250	1.044	0.935
	$M(t_5)$	2.333	1.597	1.237	1.018	0.902	2.593	1.710	1.279	1.018	0.882
	$S(25)$	2.058	1.418	1.110	0.932	0.850	2.402	1.605	1.222	1.001	0.899
3	RCM	2.000	1.440	1.180	1.050	0.990	2.260	1.570	1.240	1.070	0.980
	SCM	1.960	1.429	1.179	1.038	0.973	2.227	1.563	1.243	1.054	0.960
	$M(t_5)$	2.400	1.634	1.257	1.027	0.904	2.667	1.749	1.301	1.028	0.886
	$S(25)$	2.229	1.539	1.207	1.015	0.924	2.455	1.648	1.262	1.042	0.941
5	RCM	1.920	1.380	1.140	1.020	0.970	2.140	1.500	1.200	1.040	0.960
	SCM	1.905	1.400	1.167	1.040	0.987	2.148	1.522	1.225	1.057	0.981
	$M(t_5)$	2.500	1.690	1.290	1.042	0.908	2.778	1.810	1.336	1.046	0.894
	$S(25)$	2.339	1.616	1.267	1.064	0.967	2.498	1.683	1.293	1.072	0.971
8	RCM	1.880	1.370	1.140	1.040	0.990	2.110	1.480	1.200	1.060	0.990
	SCM	1.861	1.375	1.153	1.037	0.994	2.085	1.486	1.206	1.053	0.991
	$M(t_5)$	2.600	1.748	1.326	1.062	0.916	2.891	1.873	1.373	1.067	0.906
	$S(25)$	2.386	1.645	1.288	1.079	0.983	2.517	1.696	1.305	1.082	0.985

6. Finite-Sample Efficiency

Before explaining our simulation study, we describe the computation of the RCM scatter estimator $\hat{C}$ in some detail. Let $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ be a k-variate data set and $I = (i_1, \ldots, i_k)$, $1 \leq i_1 < \ldots < i_k \leq n$ refer to a k-subset of observations with indices listed in I. Let $\mathcal{I}$ denote the set of all possible k-subsets I. Then $\boldsymbol{R}(\boldsymbol{x}; X) = \text{ave}_I[\text{sign}\{d_0(I) + \boldsymbol{d}^T(I)\boldsymbol{x}\}\boldsymbol{d}(I)]$, where $d_0(I)$ and $\boldsymbol{d}(I)$ satisfy

$$\det \begin{pmatrix} 1 & \ldots & 1 & 1 \\ \boldsymbol{x}_{i_1} & \ldots & \boldsymbol{x}_{i_k} & \boldsymbol{x} \end{pmatrix} = d_0(I) + \boldsymbol{d}^T(I)\boldsymbol{x},$$

meaning that $(d_0(I), \boldsymbol{d}^T(I))$ is the vector of cofactors of the last column of the matrix above. The calculation of the ranks $\boldsymbol{R}_i = \boldsymbol{R}(\boldsymbol{x}_i; X)$ is explicit but requires the calculation of subdeterminants $d_0(I)$ and $\boldsymbol{d}(I)$ for all $|\mathcal{I}| = \binom{n}{k}$ possible k-subsets I. Thus the calculation of the ranks is a computing intensive task even for moderate sample size n when k is large. If this is the case we approximate the ranks by $\tilde{\boldsymbol{R}}_i = \text{ave}_j[\text{sign}\{d_0(I_j) + \boldsymbol{d}^T(I_j)\boldsymbol{x}\}\boldsymbol{d}(I_j)]$, where $I_1, \ldots, I_M$ is a random sample from $\mathcal{I}$. An approximation of the rank covariance matrix $\hat{D}$ is then $\tilde{D} = \text{ave}_i(\tilde{\boldsymbol{R}}_i \tilde{\boldsymbol{R}}_i^T)$. Next, recall that the scatter estimator $\hat{C}$ is obtained from

$$\hat{C} = (k \det(\hat{D})/c_G^2)^{1/(k-1)} \hat{D}^{-1}, \tag{10}$$

where the scalar c_G is a constant dependent on the assumed underlying model distribution G. Explicit expressions to calculate c_G are given in Appendix A.1 in the case of the normal distribution $G = \Phi_k$ and the t-distribution, $G = t_{k,\nu}$. When the calculation of the exact value of $\hat{D}$ (and thus $\hat{C}$) is not feasible (i.e., for large $|\mathcal{I}|$) we approximate $\hat{C}$ by $\tilde{C}$ obtained by substituting $\tilde{D}$ for $\hat{D}$ in (10). Presumably, with sufficiently large M, finite-sample and large-sample performance of $\hat{C}$ and $\tilde{C}$ should be similar.

Next, the finite-sample efficiencies for $\hat{C}$, $M(t_5)$, $S(25)$ and the sample covariance matrix $\hat{\text{Cov}}$ are estimated by means of the following modest simulation study. For $m = 5,000$ samples of sizes $n = 50, 150, 250$, observations were generated from a bivariate ($k = 2$) and trivariate ($k = 3$) $t_{k,\nu}$-distribution with $\nu = 5, 8$ and $\nu = \infty$ degrees of freedom. Naturally, $M(t_5)$, $S(25)$ and the sample covariance matrix are multiplied by corresponding consistency factors (when needed) so that they estimate the true underlying scatter matrix Σ. To calculate $M(t_5)$ we used the novel iterative algorithm described in Kent and Tyler (1991). The computation of the S-estimator is a bit trickier. To calculate it we used the SURREAL algorithm of Ruppert (1992), which computes an approximation to the actual value of the estimate based on a user-specified number of random $(k+1)$-subsets. In our simulations 100 $(k+1)$-subsets were considered for each computation of the S-estimate of scatter. Note also that, unlike the RCM-based estimator $\hat{C}$, all the other estimators of scatter considered here require a simultaneous location estimate $\hat{\boldsymbol{\mu}}$. For $k = 3$ and sample sizes $n = 150$ and $n = 250$, it takes about 5 and 40 seconds respectively to calculate $\hat{C}$ in a personal computer running with the AMD athlon 704MHz processor. Thus for these data sets, computation of $\hat{C}$ is close to "interactive". Note also that the AMD Athlon 704MHz machine is rather slow by today's standards.

Denote by $\hat{\Sigma}_{ij}^l$ the (i,j) element of the scatter estimate obtained from the lth sample, with $1 \leq l \leq m$. The accuracy of (i,j)th element is measured by the mean squared error (MSE), $\text{MSE}(\hat{\Sigma}_{ij}) = (n/m)\sum_{l=1}^m (\hat{\Sigma}_{ij}^l - \Sigma_{ij})^2$. The finite-sample relative efficiency (FSRE) of the off-diagonal and diagonal element of the scatter estimate is then

$$\text{FSRE}(\hat{\Sigma}_{ij}) = \frac{(\nu-2)/(\nu-4)}{\text{ave}_{i \neq j}\{\text{MSE}(\hat{\Sigma}_{ij})\}} \quad \text{and} \quad \text{FSRE}(\hat{\Sigma}_{ii}) = \frac{(2\nu-2)/(\nu-4)}{\text{ave}_i\{\text{MSE}(\hat{\Sigma}_{ii})\}},$$

since $(\nu-2)/(\nu-4)$ and $(2\nu-2)/(\nu-4)$ are the asymptotic variances for the off-diagonal and diagonal elements of the sample covariance matrix (standardized by the consistency factor) at $t_{k,\nu}$-distribution. The simulation results are reported in Table 2, 3 and 4. For the case $k=3$ we also calculated the finite sample efficiencies for the scatter estimator $\tilde{C}$ with $M=10000$ subsets.

First of all, note that the finite-sample relative efficiencies of $\hat{C}$ approximate well the asymptotic ones listed under $n=\infty$ in the tables. Similar performance is seen also for the $S(0.25)$ and $M(t_5)$ estimators. The finite-sample efficiencies of the sample covariance matrix approximate well the asymptotic values except for the case $\nu=5$, in which case the finite sample efficiencies are substantially higher than 1. One also notices that $\tilde{C}$ with $M=10000$ random subsets does not quite achieve the performance of $\hat{C}$ for sample sizes $n=150$ and $n=250$. However, the number of subsets M was kept constant independent of the sample size n and dimension k. Our simulations showed that increasing M with n and k increases the finite sample performance of $\tilde{C}$. It would be useful to find the value of M for a given sample size n and dimension k so that a modest loss in performance would be experienced compared to $\hat{C}$. This would need extensive theoretical and simulation studies and would be a subject of a paper of its own. It is our intention to study properties of the approximation algorithm in a separate forthcoming paper. To conclude: if one is willing to use $\tilde{C}$ as an approximation of $\hat{C}$, the computation time is reduced and depends essentially on M, the number of subsets. Then the computation time of the RCM will be comparable to that of high breakdown affine equivariant scatter matrix estimators, which also often use approximation schemes based on subsets, but larger than the computation times of iterative algorithms for solving M-equations.

Table 2. FSRE of off-diagonal elements of the scatter estimators at k-variate normal distribution.

		Off-diagonal FSRE				Diagonal FSRE			
k	$n=$	50	150	250	∞	50	150	250	∞
2	$\hat{C}$	0.976	0.983	0.989	0.990	0.960	0.942	0.941	0.980
	$M(t_5)$	0.897	0.898	0.902	0.902	0.875	0.858	0.856	0.882
	$S(25)$	0.835	0.845	0.840	0.850	0.896	0.897	0.912	0.899
	$\hat{\text{Cov}}$	0.975	0.993	0.998	1.000	0.968	0.965	0.968	1.000
3	$\hat{C}$	0.972	0.976	0.990	0.990	0.943	0.952	0.966	0.980
	$\tilde{C}$	0.957	0.917	0.916		0.924	0.920	0.894	
	$M(t_5)$	0.911	0.894	0.907	0.904	0.872	0.868	0.878	0.886
	$S(25)$	0.924	0.936	0.920	0.924	0.929	0.939	0.935	0.941
	$\hat{\text{Cov}}$	0.982	0.985	0.997	1.000	0.961	0.970	0.982	1.000

Table 3. FSRE of off-diagonal elements of the scatter estimators at multivariate $t_{k,\nu}$-distributions.

		$\nu = 5$				$\nu = 8$			
k	$n =$	50	150	250	∞	50	150	250	∞
2	$\hat{C}$	1.852	1.954	1.935	2.050	1.155	1.153	1.211	1.200
	$M(t_5)$	2.197	2.284	2.323	2.333	1.188	1.191	1.245	1.237
	$S(25)$	1.880	1.991	2.051	2.058	1.075	1.104	1.084	1.110
	$\hat{\text{Cov}}$	1.099	1.050	1.086	1.000	0.975	0.979	1.003	1.000
3	$\hat{C}$	1.799	1.898	1.904	2.000	1.155	1.170	1.170	1.180
	$\tilde{C}$	1.769	1.830	1.772		1.138	1.119	1.082	
	$M(t_5)$	2.281	2.370	2.354	2.400	1.202	1.217	1.220	1.257
	$S(25)$	2.092	2.184	2.218	2.229	1.166	1.211	1.192	1.207
	$\hat{\text{Cov}}$	1.236	1.206	1.168	1.000	1.009	0.998	0.981	1.000

Table 4. FSRE of diagonal elements of the scatter estimators at multivariate $t_{k,\nu}$-distributions.

		$\nu = 5$				$\nu = 8$			
k	$n =$	50	150	250	∞	50	150	250	∞
2	$\hat{C}$	2.014	2.174	2.188	2.330	1.218	1.212	1.230	1.270
	$M(t_5)$	2.438	2.552	2.566	2.593	1.256	1.270	1.270	1.279
	$S(25)$	2.314	2.374	2.396	2.402	1.189	1.222	1.226	1.222
	$\hat{\text{Cov}}$	1.037	1.114	1.160	1.000	0.991	0.966	0.964	1.000
3	$\hat{C}$	1.979	2.067	2.183	2.260	1.186	1.191	1.227	1.240
	$\tilde{C}$	1.958	1.992	2.080		1.175	1.124	1.170	
	$M(t_5)$	2.554	2.591	2.694	2.667	1.239	1.228	1.250	1.301
	$S(25)$	2.301	2.446	2.418	2.455	1.220	1.253	1.245	1.262
	$\hat{\text{Cov}}$	1.248	1.140	1.239	1.000	0.987	0.971	0.983	1.000

7. Final Comments

Classical multivariate analysis is based on the sample mean vector and sample covariance matrix. To robustify the inference procedures, the mean vector and covariance matrix have often been replaced by robust affine equivariant location vector and scatter matrix estimates. Influence functions are then used for robustness considerations and derivations of the limiting variances and covariances of the estimates.

At elliptical models, efficiencies of diagonal and off-diagonal elements characterize the efficiency properties of an affine equivariant scatter matrix. In a multivariate multiple regression problem, for example, the off-diagonal efficiency gives the efficiency of the regression coefficient estimate based on the scatter matrix estimate; in principal component analysis, the diagonal and off-diagonal efficiencies yield the efficiencies of the corresponding eigenvalue and eigenvector estimates; in canonical correlation analysis, the efficiency of the canonical correlations is given by the off-diagonal efficiency, and the efficiency of the canonical vectors depend on both diagonal and off-diagonal efficiencies. See e.g., Croux and Haesbroeck (2000), Van Aelst, Van Driessen and Rousseeuw (2000), Croux, Dehon, Rousseeuw and Van Aelst (2001) and Taskinen et al. (2002). The asymptotic efficiencies of the diagonal and off-diagonal elements of the RCM have been obtained in this paper.

Comparisons between different robust estimators of multivariate scatter is a difficult job. The attraction of the rank covariance matrix can be found in its high efficiency, even at heavy tailed distribution, and in its close relationship to existing rank concepts which gives it a non-parametric flavor. Of course, the RCM is not meant to be a competitor with high breakdown scatter matrices in terms of robustness. It is remarkable that no location estimate is needed to construct the RCM. This makes the RCM approach different from the SCM approach of Ollila, Oja and Croux (2003).

There is over recent years an increased activity in the development of approaches to multivariate inference based on different kinds of multivariate quantiles. A comprehensive review is given by Serfling (2002), who recommends median-oriented quantile functions. Besides vector valued quantiles as in Chaudhuri (1996), data depth measures (see Zuo and Serfling (2000)) are useful here. For example, Liu, Parelius and Singh (1999) introduce depth weighted L-type location and scatter matrix estimators. Zuo, Cui and He (2001) have shown consistency and asymptotic normality for a large class of weighted L-type multivariate location estimators. For the estimation of multivariate scatter, much less seems to be known. To our best knowledge, no asymptotic efficiency results are available yet for estimation of multivariate scatter based on multivariate quantiles. Moreover, computation of most versions of multivariate quantiles and data depth measures is prohibitive in large samples or high dimensions.

Appendix

A.1. Expressions for $q(r;G)$, $V_0(r;G)$ and c_G^2 at Spherical Normal and t Distributions

Before stating the expressions, we recall the following definition.

Definition 1. A generalized hypergeometric series is defined as

$${}_pF_q(a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; z) = \sum_{i=0}^{\infty} \frac{(a_1)_i (a_2)_i \ldots (a_p)_i}{(b_1)_i (b_2)_i \ldots (b_q)_i} \frac{z^i}{i!},$$

where $(c)_i = c(c+1)\cdots(c+i-1) = \Gamma(c+i)/\Gamma(c)$.

Lemma 1. *In the k-variate standard normal case, $G = \Phi_k$,*

$$V_0(r;\Phi_k) = \frac{2^{k/2}\Gamma(\frac{k+1}{2})}{\sqrt{\pi}} \exp(-kr^2/2)\, {}_1F_1\left(\frac{k+1}{2};\frac{k}{2};\frac{kr^2}{2}\right),$$
$$q(r;\Phi_k) = r\,\frac{2^{k/2}\Gamma(\frac{k+1}{2})}{\sqrt{\pi}} \exp(-kr^2/2)\, {}_1F_1\left(\frac{k+1}{2};\frac{k+2}{2};\frac{kr^2}{2}\right).$$

In the k-variate spherical t-distribution with ν degrees of freedom, $G = t_{k,\nu}$,

$$V_0(r;t_{k,\nu}) = \frac{c_{k,\nu}}{(1+r^2/\nu)^{k(\nu-1)/2}}\, {}_2F_1\left(\frac{k+1}{2},\frac{k(\nu-1)}{2};\frac{k}{2};\frac{r^2/\nu}{1+r^2/\nu}\right),$$
$$q(r;t_{k,\nu}) = \frac{c_{k,\nu} r(\nu-1)}{\nu(1+r^2/\nu)^{k(\nu-1)/2+1}}\, {}_2F_1\left(\frac{k+1}{2},\frac{k(\nu-1)+2}{2};\frac{k+2}{2};\frac{r^2/\nu}{1+r^2/\nu}\right),$$
$$c_{k,\nu} = \frac{\nu^{k/2}\Gamma(\frac{k+1}{2})\Gamma^k(\frac{\nu-1}{2})}{\Gamma^k(\frac{\nu}{2})\sqrt{\pi}}.$$

Lemma 2. *The constant c_G^2 in (6) for the standard normal distribution $G = \Phi_k$ and for t-distributions $G = t_{k,\nu}$ is given by*

$$c^2_{\Phi_k} = \frac{k2^k\Gamma^2(\frac{k+1}{2})}{\pi(k+1)}\, {}_2F_1\left(\frac{1}{2},\frac{1}{2};\frac{k+2}{2};\frac{k^2}{(k+1)^2}\right),$$
$$c^2_{t_{k,\nu}} = d_{k,\nu}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{\Gamma(\frac{k+1}{2}+i)\Gamma(\frac{k(\nu-1)+2}{2}+i)\Gamma(\frac{k+1}{2}+j)\Gamma(\frac{k(\nu-1)+2}{2}+j)\Gamma(\frac{k+2}{2}+i+j)}{\Gamma(\frac{k+2}{2}+i)\Gamma(\frac{k+2}{2}+j)\Gamma(\frac{2k\nu-k+\nu+4}{2}+i+j)i!j!},$$
$$d_{k,\nu} = \frac{\nu^{k-1}\Gamma(\frac{k}{2})\Gamma^{2k}(\frac{\nu-1}{2})\Gamma(\frac{k+\nu}{2})\Gamma(\frac{2k(\nu-1)+\nu+2}{2})}{\Gamma^2(\frac{k(\nu-1)}{2})\Gamma^{2k+1}(\frac{\nu}{2})\pi}.$$

For example, $c^2_{\Phi_k} = 0.712, 7.681, 203.749$ for dimensions $k = 2,4,6$ respectively. These results are as in Möttönen, Hettmansperger, Oja and Tienari (1998), but slightly simplified.

A.2. Proofs and Additional Lemmas

To prove Theorem 1 we need the following Lemma.

Lemma 3. *Let r be a constant scalar and write $\boldsymbol{v} = (1,0,\ldots,0)^T$ for a unit k-vector. Let G denote a cdf of a k-variate spherical random vector $\boldsymbol{x}_i = (x_{i1},\ldots,x_{ik})^T$ and let G_{k-1} denote a cdf of a $(k-1)$-variate spherical random (sub)vector*

$\boldsymbol{x}_i' = (x_{i2}, \ldots, x_{ik})^T$. *Then* $E_G[V(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{x})|\boldsymbol{x}_1 = r\boldsymbol{v}, \boldsymbol{x}] = \|\boldsymbol{x} - r\boldsymbol{v}\| V_0(\rho_x; G_{k-1})$, *where* $\rho_x^2 = r^2\{1 - (x_1 - r)^2/\|\boldsymbol{x} - r\boldsymbol{v}\|^2\}$, *and* $V(\cdot)$ *and* $V_0(\cdot)$ *are defined by* (1) *and* (4), *respectively.*

Proof of Lemma 3. Let P be a $k \times k$ *orthogonal* (rotation) matrix (hence $PP^T = P^TP = I_k$ and $\text{abs}\{\det(P^T)\} = 1$) such that $P(\boldsymbol{x} - r\boldsymbol{v}) = (\|\boldsymbol{x} - r\boldsymbol{v}\|, 0, \ldots, 0)^T$. Then

$$P = \begin{pmatrix} \boldsymbol{p}_1^T \\ P_2 \end{pmatrix} = \begin{pmatrix} \|\boldsymbol{x} - r\boldsymbol{v}\|^{-1}(\boldsymbol{x} - r\boldsymbol{v})^T \\ P_2 \end{pmatrix},$$

where P_2 is a $(k-1) \times k$-matrix. By symmetry, it is equivalent to solve the expectation

$$\begin{aligned}
& E_G[V(\boldsymbol{x}_1, P^T\boldsymbol{x}_2 \ldots, P^T\boldsymbol{x}_k, \boldsymbol{x})|\boldsymbol{x}_1 = r\boldsymbol{v}, \boldsymbol{x}] \\
&= E_G\left[\text{abs}\left\{\det\left(P^T\boldsymbol{x}_2 - \boldsymbol{x}_1 \ldots P^T\boldsymbol{x}_k - \boldsymbol{x}_1 \; \boldsymbol{x} - \boldsymbol{x}_1\right)\right\} |\boldsymbol{x}_1 = r\boldsymbol{v}, \boldsymbol{x}\right] \\
&= E_G\left[\text{abs}\left\{\det\left(\boldsymbol{x}_2 - P\boldsymbol{x}_1 \ldots \boldsymbol{x}_k - P\boldsymbol{x}_1 \; P(\boldsymbol{x} - \boldsymbol{x}_1)\right)\right\} |\boldsymbol{x}_1 = r\boldsymbol{v}, \boldsymbol{x}\right] \text{abs}\{\det(P^T)\} \\
&= E_G\left[\text{abs}\left\{\det\begin{pmatrix} x_{21} - \boldsymbol{p}_1^T\boldsymbol{x}_1 \ldots x_{k1} - \boldsymbol{p}_1^T\boldsymbol{x}_1 & \|\boldsymbol{x} - \boldsymbol{x}_1\| \\ \boldsymbol{x}_2' - P_2\boldsymbol{x}_1 \; \ldots \; \boldsymbol{x}_k' - P_2\boldsymbol{x}_1 & \boldsymbol{0} \end{pmatrix}\right\} \Big|\boldsymbol{x}_1 = r\boldsymbol{v}, \boldsymbol{x}\right] \\
&= \|\boldsymbol{x} - r\boldsymbol{v}\| \, E_G\left[\text{abs}\left\{\det\left(\boldsymbol{x}_2' - P_2\boldsymbol{x}_1 \ldots \boldsymbol{x}_k' - P_2\boldsymbol{x}_1\right)\right\} |\boldsymbol{x}_1 = r\boldsymbol{v}\right] \\
&= \|\boldsymbol{x} - r\boldsymbol{v}\| \, E_G\left[\text{abs}\left\{\det\begin{pmatrix} 1 \; \ldots \; 1 & 1 \\ \boldsymbol{x}_2' \ldots \boldsymbol{x}_k' & P_2\boldsymbol{x}_1 \end{pmatrix}\right\} \Big|\boldsymbol{x}_1 = r\boldsymbol{v}\right] \\
&= \|\boldsymbol{x} - r\boldsymbol{v}\| \, V_0(\|P_2r\boldsymbol{v}\|; G_{k-1}).
\end{aligned}$$

From $\|Pr\boldsymbol{v}\|^2 = r^2$, we obtain the relation $(\boldsymbol{p}_1^Tr\boldsymbol{v})^2 + \|P_2r\boldsymbol{v}\|^2 = r^2$ and, as $\boldsymbol{p}_1^Tr\boldsymbol{v} = \|\boldsymbol{x} - r\boldsymbol{v}\|^{-1}(x_1 - r)r$, it follows that $\|P_2r\boldsymbol{v}\|^2 = r^2\{1 - (x_1 - r)^2/\|\boldsymbol{x} - r\boldsymbol{v}\|^2\} = \rho_x^2$, which completes the proof.

Proof of Theorem 1. Let $X = \{\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_k}, \boldsymbol{x}_{j_1}, \ldots, \boldsymbol{x}_{j_k}, \boldsymbol{x}_i\}$ denote the data set of $2k+1$ i.i.d. observations from the k-variate spherical distribution G. Further, write $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_k)$ for the k-sets of indices. Then let the the scalar $d_0(I)$ and the k-vector $\boldsymbol{d}(I) = (d_1(I), \ldots, d_k(I))^T$ be as in Section 6 and write $S_I(\boldsymbol{x}_i) = \text{sign}\{d_0(I) + \boldsymbol{d}(I)^T\boldsymbol{x}_i\}$. By reversing the order of the expectation and differentiation, (2) can be rewritten as $\boldsymbol{R}(\boldsymbol{x}_i; G) = E_G[S_I(\boldsymbol{x}_i)\boldsymbol{d}(I)|\boldsymbol{x}_i]$. Then note that

$$\begin{aligned}
D(G) = E_{x_i}[\boldsymbol{R}(\boldsymbol{x}_i; G)\boldsymbol{R}(\boldsymbol{x}_i; G)] &= E_{x_i}\left[E_G[S_I(\boldsymbol{x}_i)S_J(\boldsymbol{x}_i)\boldsymbol{d}(I)\boldsymbol{d}(J)^T|\boldsymbol{x}_i]\right] \\
&= E_G[S_I(\boldsymbol{x}_i)S_J(\boldsymbol{x}_i)\boldsymbol{d}(I)\boldsymbol{d}(J)^T]. \qquad (11)
\end{aligned}$$

It is now straightforward to show that the influence function of D at G is

IF $(\boldsymbol{x}; D, G)$

$$= \frac{\partial}{\partial \varepsilon} \int \cdots \int \int S_I(\boldsymbol{x}_i) S_J(\boldsymbol{x}_i) \boldsymbol{d}(I) \boldsymbol{d}(J)^T \, dG_\varepsilon(\boldsymbol{x}_{i_1}) \cdots dG_\varepsilon(\boldsymbol{x}_{j_k}) dG_\varepsilon(\boldsymbol{x}_i) \Big|_{\varepsilon=0}$$
$$= \boldsymbol{R}(\boldsymbol{x}; G) \boldsymbol{R}^T(\boldsymbol{x}; G) + 2k E_G[S_I(\boldsymbol{x}_i) S_J(\boldsymbol{x}_i) \boldsymbol{d}(I) \boldsymbol{d}(J)^T | \boldsymbol{x}_{i_1} = \boldsymbol{x}] - (2k+1) D(G), \quad (12)$$

where $G_\varepsilon = (1-\varepsilon)G + \varepsilon \Delta_x$ is the contaminated distribution.

We derive the influence function (12) of D for a point in the direction of the first axis, that is, we set $\boldsymbol{x} = r\boldsymbol{v}$, where $\boldsymbol{v} = (1, 0, \ldots, 0)^T$. Using the fact that for a spherical random variable $\boldsymbol{x} = (x_i)_{1 \le i \le k}$, $(s_i x_{\pi(i)})_{1 \le i \le k}$ has the same distribution for arbitrary $s_i \in \{-1, 1\}$ and permutation π of $\{1, \ldots, k\}$, one immediately finds that

$$2k \, E_G[S_I(\boldsymbol{x}_i) S_J(\boldsymbol{x}_i) \boldsymbol{d}(I) \boldsymbol{d}(J)^T | \boldsymbol{x}_{i_1} = r\boldsymbol{v}] = \begin{pmatrix} \gamma(r; G) & \boldsymbol{0}^T \\ \boldsymbol{0} & \eta(r; G) I_{k-1} \end{pmatrix},$$

where

$$\gamma(r; G) = 2k \, E_G[S_I(\boldsymbol{x}_i) S_J(\boldsymbol{x}_i) d_1(I) d_1(J) | \boldsymbol{x}_{i_1} = r\boldsymbol{v}]$$
$$= 2k \, E_{x_i} \left[E_G[S_I(\boldsymbol{x}_i) d_1(I) | \boldsymbol{x}_{i_1} = r\boldsymbol{v}, \boldsymbol{x}_i] \, q(\|\boldsymbol{x}_i\|; G) \frac{x_{i1}}{\|\boldsymbol{x}_i\|} \right]$$

as $E_G[S_J(\boldsymbol{x}_i) \boldsymbol{d}(J) | \boldsymbol{x}_i] = \boldsymbol{R}(\boldsymbol{x}_i; G) = q(\|\boldsymbol{x}_i\|; G) \boldsymbol{x}_i / \|\boldsymbol{x}_i\|$ by (4). Similarly, we obtain

$$\eta(r; G) = 2k E_{x_i} \left[E_G[S_I(\boldsymbol{x}_i) d_2(I) | \boldsymbol{x}_{i_1} = r\boldsymbol{v}, \boldsymbol{x}_i] \, q(\|\boldsymbol{x}_i\|; G) \frac{x_{i2}}{\|\boldsymbol{x}_i\|} \right].$$

Next, note that

$$E_G[S_I(\boldsymbol{x}_i) d_1(I) | \boldsymbol{x}_{i_1} = r\boldsymbol{v}, \boldsymbol{x}_i] = \frac{\partial}{\partial x_{i1}} E_G[V(\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_k}, \boldsymbol{x}_i) | \boldsymbol{x}_{i_1} = r\boldsymbol{v}, \boldsymbol{x}_i]$$
$$= \frac{\partial}{\partial x_{i1}} \|\boldsymbol{x}_i - r\boldsymbol{v}\| V_0(\rho_{x_i}; G_{k-1})$$
$$= \frac{x_{i1} - r}{\|\boldsymbol{x}_i - r\boldsymbol{v}\|} \{V_0(\rho_{x_i}; G_{k-1}) - \rho_{x_i} \, q(\rho_{x_i}; G_{k-1})\},$$

where the first equality follows from reversing the order of expectation and differentiation, the second equality follows from Lemma 3, and the third equality follows by simple differentiation rules (use the chain rule to obtain $\partial V_0(\rho_{x_i}; G_{k-1}) / \partial x_{i1} = q(\rho_{x_i}; G_{k-1}) \partial \rho_{x_i} / \partial x_{i1}$ (as $V_0' = q$) and $\partial \rho_{x_i} / \partial x_{i1} = -(x_{i1} - r) \rho_{x_i} \|\boldsymbol{x}_i - r\boldsymbol{v}\|^{-2}$). Similarly, one can show that

$$E_G[S_I(\boldsymbol{x}_i) d_2(I) | \boldsymbol{x}_{i_1} = r\boldsymbol{v}, \boldsymbol{x}_i]$$
$$= \frac{x_{i2}}{\|\boldsymbol{x}_i - r\boldsymbol{v}\|} \left\{ V_0(\rho_{x_i}; G_{k-1}) + \frac{(r^2 - \rho_{x_i}^2)}{\rho_{x_i}} q(\rho_{x_i}; G_{k-1}) \right\}.$$

This then yields the stated expressions for $\gamma(r;G)$ and $\eta(r;G)$.

Then, as $\boldsymbol{R}(r\boldsymbol{v};G) = q(r;G)\boldsymbol{v}$, we may now write the influence function (12) of D for a point $\boldsymbol{x} = r\boldsymbol{v}$ as

$$\begin{aligned}\mathrm{IF}\,(r\boldsymbol{v};D,G) &= q^2(r;G)\boldsymbol{v}\boldsymbol{v}^T + \begin{pmatrix}\gamma(r;G) & \mathbf{0}^T \\ \mathbf{0} & \eta(r;G)I_{k-1}\end{pmatrix} - (2k+1)D(G) \\ &= \{q^2(r;G)+\gamma(r;G)-\eta(r;G)\}\boldsymbol{v}\boldsymbol{v}^T + \eta(r;G)I_k - (2k+1)D(G). \quad (13)\end{aligned}$$

An influence point in an arbitrary direction is obtained by setting $\boldsymbol{x} = Pr\boldsymbol{v} = r\boldsymbol{u}$ for a well chosen orthogonal ($PP^T = I_k$) rotation matrix $P = [\boldsymbol{u} \cdots \boldsymbol{u}_k]$. The influence function is then given by $\mathrm{IF}\,(\boldsymbol{x};D,G) = P\,\mathrm{IF}\,(r\boldsymbol{v};D,G)P^T$ which, by (13) and relations $P\boldsymbol{v} = \boldsymbol{u}$, $PP^T = I_k$ and $D(G) = (c_G^2/k)I_k$, reduces to

$$\mathrm{IF}\,(\boldsymbol{x};D,G) = \left\{q^2(r;G) + \gamma(r;G) - \eta(r;G)\right\}\boldsymbol{u}\boldsymbol{u}^T - \left\{2k+1-\frac{\eta(r;G)}{c_G^2/k}\right\}D(G),$$

which completes the proof.

References

Brown, B. and Hettmansperger, T. P. (1987). Affine invariant rank methods in the bivariate location model. *J. Roy. Statist. Soc. Ser. B* **49**, 301-310.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91**, 862-872.

Croux, C. and Haesbroeck, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix : influence functions and efficiencies. *Biometrika* **87**, 603-618.

Croux, C., Dehon, C., Rousseeuw, P. J. and Van Aelst, S. (2001). Robust estimation of the conditional median at elliptical models. *Statist. Probab. Lett.* **51**, 361-368.

Davies, P. L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15**, 1269-1292.

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

Hampel, F., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

Hettmansperger, T. P. and McKean, J. (1998). *Robust Nonparametric Statistical Methods*. Arnold, London.

Hettmansperger, T. P., Möttönen, J. and Oja, H. (1998). Affine invariant multivariate rank tests for several samples. *Statist. Sinica* **8**, 765-800.

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

Liu, R. Y., Parelius, J. M. and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann. Statist.* **27**, 783-858.

Lopuhaä, H. P. (1989). On the relation between S-estimators and M-estimators of multivariate location and scatter. *Ann. Statist.* **17**, 1662-1683.

Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, New York.

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

Maronna, R. A. and Yohai, V. (1998). Robust estimation of multivariate location and scatter, In *Encyclopedia of Statistical Sciences Update* **2** (Edited by S. Kotz, C. Read and D. Banks), 589-596. Wiley, New York.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

Möttönen, J., Hettmansperger, T. P., Oja, H. and Tienari, J. (1998). On the efficiency of affine invariant multivariate rank tests. *J. Multivariate Anal.* **66**, 118-132.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1**, 327-332.

Oja, H. (1999). Affine equivariant multivariate sign and rank tests and corresponding estimates: a review. *Scand. J. Statist.* **26**, 319-343.

Ollila, E., Oja, H. and Croux, C. (2003). The Affine equivariant sign covariance matrix: asymptotic behavior and efficiencies. *J. Multivariate Anal.* **87**, 328-355.

Ollila, E., Hettmansperger, T. P. and Oja, H. (2002). Estimates of regression coefficients based on the sign covariance matrix. *J. Roy. Statist. Soc. Ser. B* **64**, 447-466.

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

Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Statist. Neerlandica* **56**, 214-232.

Taskinen, S., Croux, C., Kankainen, A., Ollila, E. and Oja, H. (2002). Canonical analysis based on scatter and shape matrices. Submitted.

Tyler, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika* **70**, 411-420.

Van Aelst, S., Van Driessen, K. and Rousseeuw, P. (2000). A robust method for multivariate regression. In *Data Analysis, Classification and Related Methods* (Edited by H. A. L. Kiers et al.), Springer, Heidelberg.

Visuri, S., Koivunen, V. and Oja, H. (2000). Sign and rank covariance matrices. *J. Statist. Plann. Inference* **91**, 557-575.

Visuri, S., Ollila, E., Koivunen, V., Möttönen, J. and Oja, H. (2003). Affine equivariant multivariate rank methods. *J. Statist. Plann. Inference* **114**, 161-185.

Zuo, Y., Cui, H. and He, X. (2001). On the Stahel-Donoho estimator and depth-weighted means of multivariate data. *Ann. Statist.* To appear.

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28**, 461-482.

Signal Processing Laboratory, Helsinki University of Technology, P.O. Box 3000, FIN-02015 HUT, Finland.

E-mail: esollila@wooster.hut.fi

Department of Applied Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.

E-mail: christophe.croux@econ.kuleuven.ac.be

University of Jyväskylä, P.O. Box 35, FIN-40351 Jyväskylä, Finland.

E-mail: Hannu.oja@maths.jyu.fi

(Received January 2002; accepted May 2003)