Filomat 37:23 (2023), 7993–8011
https://doi.org/10.2298/FIL2323993R

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Generalized Mahalanobis distance and its application in detecting matrix outliers

Amir Rezaei[a], Kambiz Ahmadi[b]

[a] *Ph.D. graduate in Statistics from the University of Birjand, Birjand, Iran*
[b] *Department of Computer Science, Shahrekord University, Shahrekord, Iran*

Abstract. In this paper, a new distance for matrix observations called generalized Mahalanobis distance is introduced, some of its properties are studied, and its distribution is obtained for the observations of the matrix variate elliptically contoured distributions. Also, as a significant application, the introduced distance is used in detecting matrix outliers, and its method is described. Finally, some examples are provided for illustrative purposes, and the performance of the presented approach of detecting outliers is investigated by a simulation study.

1. Introduction

Many methods in multivariate analysis such as hypothesis testing, clustering and classification methods, outlier detection, and goodness-of-fit tests are usually based on different distances defined for observations. These distances measure and calculate in different ways the similarity or the amount of difference between observations that are usually vectors. One of the most important distances in multivariate analysis methods is the Mahalanobis distance, which was first introduced by [12]. According to his definition, if $\boldsymbol{x}$ is an observation of a multivariate distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, then its Mahalanobis distance is calculated as

$$\mathcal{MD}(\boldsymbol{x}) = \sqrt{(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}.$$

Mahalanobis distance appears naturally in multivariate analysis methods and is used for different purposes. For example, the Mahalanobis distance is the basis for multivariate outlier detection such that observations having a large Mahalanobis distance are considered as multivariate outliers. Among the new research related to outlier detection using the Mahalanobis distance, the reader is referred to [4], [1], and [11].

Matrix observations and their distributions play an important role in multivariate analysis and are useful for describing and modeling repeated measurements in multivariate variables. For example, suppose n blood tests are taken from N patients where in each test, p variables are measured. In this case, for each patient, a matrix observation of size $p \times n$ is observed. Most of the distances in multivariate analysis methods are similar to the Mahalanobis distance related to vector observations, and among them, a

2020 *Mathematics Subject Classification*. 65F35; 62H10
Keywords. Mahalanobis distance, Matrix variate distribution, Outlier detection, Minimum covariance determinant estimator
Received: 07 February 2023; Revised: 09 March 2023; Accepted: 06 April 2023
Communicated by Biljana Č. Popović
Email addresses: amir.rezaei.st@gmail.com (Amir Rezaei), K.Ahmadi@sku.ac.ir (Kambiz Ahmadi)

few, including the Frobenius distance, are related to matrix observations; see [6] for more information about the Frobenius distance. Accordingly, we introduce, in the next section, a new distance for matrix observations called the generalized Mahalanobis distance and examine some of its properties. In Section 3, as a significant application, we describe the detection of matrix outliers by using the introduced distance. As an illustration of the suitability and applicability of the introduced distance in detecting matrix outliers, we find outliers in simulated and real examples in Section 4. In Section 5, we present a simulation study to examine how the presented approach of detecting outliers performs. Finally, we provide a summary discussion in Section 6.

2. Definition and some properties

In this section, we introduce a new distance for matrix observations and investigate some of its properties. For this purpose, suppose that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are observations of size $p \times n$ from a matrix variate distribution with the mean matrix $\boldsymbol{M}$ of size $p \times n$ and positive definite matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ of sizes $n \times n$ and $p \times p$, respectively, such that $\boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}$ is the covariance matrix of $\mathsf{vec}(\boldsymbol{X}^T)$ and $\mathsf{vec}(\boldsymbol{Y}^T)$, where $\mathsf{vec}(\cdot)$ denotes the vectorization operator. We define the distance between two matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ as

$$\mathcal{D}(\boldsymbol{X},\boldsymbol{Y}) = \sqrt{\operatorname{tr}\big(\boldsymbol{\Psi}^{-1}(\boldsymbol{X}-\boldsymbol{Y})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{Y})\big)}, \tag{1}$$

where $\operatorname{tr}(\boldsymbol{A})$ is the trace of a square matrix $\boldsymbol{A}$ and $\boldsymbol{B}^T$ denotes the transpose of $\boldsymbol{B}$. The following corollary provides another form of the distance $\mathcal{D}$.

Corollary 2.1. *The distance $\mathcal{D}$ can be written as follows*

$$\mathcal{D}(\boldsymbol{X},\boldsymbol{Y}) = \sqrt{\big(\mathsf{vec}(\boldsymbol{X}^T-\boldsymbol{Y}^T)\big)^T(\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi})^{-1}\,\mathsf{vec}(\boldsymbol{X}^T-\boldsymbol{Y}^T)}.$$

Proof. By using Theorem 1.17 (ii) of [7], we have

$$\operatorname{tr}\big(\boldsymbol{\Psi}^{-1}(\boldsymbol{X}-\boldsymbol{Y})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{Y})\big) = \big(\mathsf{vec}\big((\boldsymbol{X}-\boldsymbol{Y})^T\big)\big)^T(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Psi}^{-1})\mathsf{vec}(\boldsymbol{X}^T-\boldsymbol{Y}^T).$$

The result follows from $\mathsf{vec}\big((\boldsymbol{X}-\boldsymbol{Y})^T\big) = \mathsf{vec}(\boldsymbol{X}^T-\boldsymbol{Y}^T)$ and $(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Psi}^{-1}) = (\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi})^{-1}$. □

Some properties of the distance $\mathcal{D}$ can be found in the following proposition.

Proposition 2.2. *Consider the matrices $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$. Then*

(i) $\mathcal{D}(\boldsymbol{X},\boldsymbol{Y}) \geq 0$,

(ii) $\boldsymbol{X}=\boldsymbol{Y}$ *if and only if* $\mathcal{D}(\boldsymbol{X},\boldsymbol{Y}) = 0$,

(iii) $\mathcal{D}(\boldsymbol{X},\boldsymbol{Y}) = \mathcal{D}(\boldsymbol{Y},\boldsymbol{X})$,

(iv) $\mathcal{D}(\boldsymbol{X},\boldsymbol{Z}) \leq \mathcal{D}(\boldsymbol{X},\boldsymbol{Y}) + \mathcal{D}(\boldsymbol{Y},\boldsymbol{Z})$.

Proof. The properties *(ii)* and *(iii)* are clear and we only prove the properties *(i)* and *(iv)*.
(i) Consider the form of $\mathcal{D}$ in Corollary 2.1. Because $\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}$ is a positive definite matrix, we know

$$\big(\mathsf{vec}(\boldsymbol{X}^T-\boldsymbol{Y}^T)\big)^T(\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi})^{-1}\mathsf{vec}(\boldsymbol{X}^T-\boldsymbol{Y}^T) > 0,$$

when $X \neq Y$ and the equality happens if $X = Y$. Hence, $\mathcal{D}(X, Y) \geq 0$.
(iv) Let $\|a\| = \sqrt{a^T a}$ is the Euclidean norm of the vector a. By the triangle inequality, we can write

$$\begin{aligned}
\mathcal{D}(X, Z) &= \sqrt{\left(\mathrm{vec}(X^T - Z^T)\right)^T (\Sigma \otimes \Psi)^{-1} \mathrm{vec}(X^T - Z^T)} \\
&= \|(\Sigma \otimes \Psi)^{-\frac{1}{2}} \mathrm{vec}(X^T - Z^T)\| \\
&= \|(\Sigma \otimes \Psi)^{-\frac{1}{2}} \mathrm{vec}(X^T - Y^T + Y^T - Z^T)\| \\
&= \|(\Sigma \otimes \Psi)^{-\frac{1}{2}} \mathrm{vec}(X^T - Y^T) + (\Sigma \otimes \Psi)^{-\frac{1}{2}} \mathrm{vec}(Y^T - Z^T)\| \\
&\leq \|(\Sigma \otimes \Psi)^{-\frac{1}{2}} \mathrm{vec}(X^T - Y^T)\| + \|(\Sigma \otimes \Psi)^{-\frac{1}{2}} \mathrm{vec}(Y^T - Z^T)\| \\
&= \sqrt{\left(\mathrm{vec}(X^T - Y^T)\right)^T (\Sigma \otimes \Psi)^{-1} \mathrm{vec}(X^T - Y^T)} \\
&\quad + \sqrt{\left(\mathrm{vec}(Y^T - Z^T)\right)^T (\Sigma \otimes \Psi)^{-1} \mathrm{vec}(Y^T - Z^T)} \\
&= \mathcal{D}(X, Y) + \mathcal{D}(Y, Z).
\end{aligned}$$

□

The following proposition shows that in a particular case, the distance $\mathcal{D}$ becomes the Euclidean distance.

Proposition 2.3. *If Σ and Ψ are identity matrices, then*

$$\mathcal{D}(X, Y) = \mathcal{ED}(\mathrm{vec}(X^T), \mathrm{vec}(Y^T)),$$

where $\mathcal{ED}(a, b)$ is the Euclidean distance between two vectors a and b.

Proof. It follows from Corollary 2.1 that

$$\begin{aligned}
\mathcal{D}(X, Y) &= \sqrt{\left(\mathrm{vec}(X^T - Y^T)\right)^T \mathrm{vec}(X^T - Y^T)} \\
&= \sqrt{\left(\mathrm{vec}(X^T) - \mathrm{vec}(Y^T)\right)^T \left(\mathrm{vec}(X^T) - \mathrm{vec}(Y^T)\right)} \\
&= \|\mathrm{vec}(X^T) - \mathrm{vec}(Y^T)\| \\
&= \mathcal{ED}(\mathrm{vec}(X^T), \mathrm{vec}(Y^T)).
\end{aligned}$$

□

Based on distance $\mathcal{D}$, we define the generalized Mahalanobis distance as follows.

Definition 2.4. *Let $X \in \mathbb{R}^{p \times n}$ be a sample from a known matrix variate distribution with the mean matrix M and positive definite matrices Ψ and Σ. The generalized Mahalanobis distance (GMD) of X is a continuous function from $\mathbb{R}^{p \times n}$ to $[0, \infty)$ that compute as*

$$\mathcal{GMD}(X) = \sqrt{\mathrm{tr}\left(\Psi^{-1}(X - M)^T \Sigma^{-1}(X - M)\right)}. \tag{2}$$

In the following, we present some properties of the GMD related to a widely used class of matrix variate distributions.

One of the most important classes of matrix variate distributions is the class of matrix variate elliptically contoured distributions. The comprehensive collections of the most important results on the matrix variate elliptically contoured distributions can be found in [7] and [2]. A random matrix X of dimension $p \times n$ is said to have a matrix variate elliptically contoured distribution and is denoted by $\boldsymbol{X} \sim E_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}; g)$, if its probability density function (pdf) is given by

$$f_g(\boldsymbol{X}; \boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}) = |\boldsymbol{\Sigma}|^{-\frac{n}{2}} |\boldsymbol{\Psi}|^{-\frac{p}{2}} g\left(\mathsf{tr}\big(\boldsymbol{\Psi}^{-1}(\boldsymbol{X}-\boldsymbol{M})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\big)\right),$$

where $\boldsymbol{M}$ is a $p \times n$ matrix corresponding to the mean of $\boldsymbol{X}$, while $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ are positive definite matrices of dimension $p \times p$ and $n \times n$, respectively, such that $\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$ is the covariance matrix of $\mathsf{vec}(\boldsymbol{X}^T)$ and g is a density generator function such that $\int_0^\infty u^{\frac{pn}{2}-1} g(u)du = \Gamma(\frac{pn}{2})/\pi^{\frac{pn}{2}}$ with the gamma function Γ.

The class of matrix variate elliptically contoured distributions includes a wild range of symmetric matrix variate distributions, such that by considering different density generator functions, different symmetric matrix distributions are obtained. Some special cases of the matrix variate elliptically contoured distributions are the following;

(a) **Matrix variate normal distribution**: This distribution is obtained by considering $g(x) = (2\pi)^{-\frac{pn}{2}} e^{-\frac{x}{2}}$, $x \in \mathbb{R}$ and is denoted by $N_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi})$.

(b) **Matrix variate generalized Pearson type II (GPII) distribution**: This case, which is denoted here by $GPII_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}; \omega, s)$, follows by considering

$$g(x) = \frac{\Gamma\left(\omega + 1 + \frac{pn}{2}\right)}{\Gamma(\omega+1)\, \pi^{\frac{pn}{2}}} s^{-(\omega + \frac{pn}{2})} (s-x)^\omega, \quad 0 < x < s, \quad \omega > -1, \quad s > 0.$$

(c) **Matrix variate generalized t (GT) distribution**: It is obtained by considering

$$g(x) = \frac{\Gamma\left(\frac{\nu+pn}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \pi^{\frac{pn}{2}}} \tau^{\frac{\nu}{2}} (\tau + x)^{-\frac{(\nu+pn)}{2}}, \quad x > 0, \quad \nu, \tau > 0.$$

We denote this distribution by $GT_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}; \nu, \tau)$.

The following proposition shows that the GMD related to a matrix variate elliptically contoured distribution is invariant under linear transformations.

Proposition 2.5. *Suppose that $\mathcal{H}$ is the set of all transformations $h(\boldsymbol{X}) = \boldsymbol{AXB} + \boldsymbol{C}$ for some invertible matrices $\boldsymbol{A}_{p\times p}$ and $\boldsymbol{B}_{n\times n}$, and some matrix $\boldsymbol{C}_{p\times n}$. If $\boldsymbol{X} \sim E_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}; g)$, then*

$$\mathcal{GMD}\big(h(\boldsymbol{X})\big) = \mathcal{GMD}(\boldsymbol{X}).$$

In other words, the GMD is invariant under the transformations

$$\mathcal{H} = \Big\{h : h(\boldsymbol{X}) = \boldsymbol{AXB} + \boldsymbol{C}, \textit{ for some invertible matrices } \boldsymbol{A} \textit{ and } \boldsymbol{B} \textit{ and some matrix } \boldsymbol{C}\Big\}.$$

Proof. From properties of the matrix variate elliptically contoured distributions, we have

$$h(\boldsymbol{X}) = \boldsymbol{AXB} + \boldsymbol{C} \sim E_{p\times n}\big(\boldsymbol{AMB} + \boldsymbol{C}, (\boldsymbol{A\Sigma A}^T) \otimes (\boldsymbol{B}^T\boldsymbol{\Psi B}); g\big).$$

Therefore

$$\begin{aligned}
\mathcal{GMD}\big(h(\boldsymbol{X})\big) &= \sqrt{\mathsf{tr}\left((\boldsymbol{B}^T\boldsymbol{\Psi B})^{-1}(h(\boldsymbol{X}) - \boldsymbol{AMB} - \boldsymbol{C})^T (\boldsymbol{A\Sigma A}^T)^{-1} (h(\boldsymbol{X}) - \boldsymbol{AMB} - \boldsymbol{C})\right)} \\
&= \sqrt{\mathsf{tr}\left(\boldsymbol{B}^{-1}\boldsymbol{\Psi}^{-1}(\boldsymbol{B}^T)^{-1}\big(\boldsymbol{A}(\boldsymbol{X}-\boldsymbol{M})\boldsymbol{B}\big)^T (\boldsymbol{A}^T)^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}^{-1}\big(\boldsymbol{A}(\boldsymbol{X}-\boldsymbol{M})\boldsymbol{B}\big)\right)} \\
&= \sqrt{\mathsf{tr}\big(\boldsymbol{B}^{-1}\boldsymbol{\Psi}^{-1}(\boldsymbol{X}-\boldsymbol{M})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\boldsymbol{B}\big)} \\
&= \sqrt{\mathsf{tr}\big(\boldsymbol{B}\boldsymbol{B}^{-1}\boldsymbol{\Psi}^{-1}(\boldsymbol{X}-\boldsymbol{M})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\big)} \\
&= \sqrt{\mathsf{tr}\big(\boldsymbol{\Psi}^{-1}(\boldsymbol{X}-\boldsymbol{M})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\big)} = \mathcal{GMD}(\boldsymbol{X}).
\end{aligned}$$

□

The distribution of the GMD related to matrix variate elliptically contoured distributions is provided in the following theorem. Let the notation $X \stackrel{d}{=} Y$ means that X and Y have the same distribution.

Theorem 2.6. *If $\boldsymbol{X} \sim E_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}; g)$, then*

$$\mathcal{GMD}^2(\boldsymbol{X}) \stackrel{d}{=} R^2,$$

where R is a positive random variable with the pdf

$$f_R(r) = \frac{2\pi^{\frac{pn}{2}}}{\Gamma\left(\frac{pn}{2}\right)} r^{pn-1} g(r^2), \quad r > 0. \tag{3}$$

Proof. By Corollary 2.1,

$$\begin{aligned}\mathcal{GMD}^2(\boldsymbol{X}) &= \mathrm{tr}\big(\boldsymbol{\Psi}^{-1}(\boldsymbol{X}-\boldsymbol{M})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{M})\big)\\ &= \big(\mathrm{vec}(\boldsymbol{X}^T) - \mathrm{vec}(\boldsymbol{M}^T)\big)^T(\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi})^{-1}\big(\mathrm{vec}(\boldsymbol{X}^T) - \mathrm{vec}(\boldsymbol{M}^T)\big).\end{aligned}$$

From [7], we know that $\boldsymbol{X} \sim E_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}; g)$ if and only if $\mathrm{vec}(\boldsymbol{X}^T) \sim E_{pn}\left(\mathrm{vec}(\boldsymbol{M}^T), \boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}; g\right)$. Hence, by [2],

$$\big(\mathrm{vec}(\boldsymbol{X}^T) - \mathrm{vec}(\boldsymbol{M}^T)\big)^T(\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi})^{-1}\big(\mathrm{vec}(\boldsymbol{X}^T) - \mathrm{vec}(\boldsymbol{M}^T)\big) \stackrel{d}{=} R^2,$$

such that R has the pdf (3). □

Corollary 2.7. *If $\boldsymbol{X} \sim N_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi})$, then $\mathcal{GMD}^2(\boldsymbol{X})$ has the chi-squared distribution with pn degrees of freedom, i.e. $\mathcal{GMD}^2(\boldsymbol{X}) \sim \chi^2_{pn}$.*

Proof. By using the density generator function of the matrix variate normal distribution, the pdf (3) becomes

$$f_R(r) = \frac{r^{pn-1}}{\Gamma\left(\frac{pn}{2}\right)}\left(\frac{1}{2}\right)^{\frac{pn}{2}-1} e^{-\frac{r^2}{2}}, \quad r > 0.$$

By substituting $U = R^2$ we have

$$f_U(u) = \frac{u^{\frac{pn}{2}-1}}{\Gamma\left(\frac{pn}{2}\right)}\left(\frac{1}{2}\right)^{\frac{pn}{2}} e^{-\frac{u}{2}}, \quad u > 0,$$

that is the pdf of the distribution χ^2_{pn}. □

Corollary 2.8. *If $\boldsymbol{X} \sim GPII_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}; \omega, s)$, then $\frac{1}{s}\mathcal{GMD}^2(\boldsymbol{X})$ has the beta distribution with parameters $\frac{pn}{2}$ and $\omega + 1$, i.e. $\frac{1}{s}\mathcal{GMD}^2(\boldsymbol{X}) \sim Beta\left(\frac{pn}{2}, \omega+1\right)$.*

Proof. From the density generator function of the matrix variate GPII distribution, we have

$$f_R(r) = \frac{\Gamma\left(\omega+1+\frac{pn}{2}\right)}{\Gamma(\omega+1)\,\Gamma\left(\frac{pn}{2}\right)} 2r^{pn-1}s^{-\left(\omega+\frac{pn}{2}\right)}(s-r^2)^{\omega}, \quad r^2 < s.$$

Hence, the pdf of $U = \frac{R^2}{s}$ can be obtained as follows;

$$\frac{\Gamma\left(\omega+1+\frac{pn}{2}\right)}{\Gamma(\omega+1)\,\Gamma\left(\frac{pn}{2}\right)} u^{\frac{pn}{2}-1}(1-u)^{\omega}, \quad 0 < u < 1,$$

which is the pdf of the distribution $Beta\left(\frac{pn}{2}, \omega+1\right)$. □

Corollary 2.9. *If* $\boldsymbol{X} \sim GT_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}; \nu, \tau)$*, then* $\frac{\mathcal{GMD}^2(\boldsymbol{X})}{\tau+\mathcal{GMD}^2(\boldsymbol{X})} \sim Beta\left(\frac{pn}{2}, \frac{\nu}{2}\right)$.

Proof. By Theorem 2.6, it must be proved that $\frac{R^2}{\tau+R^2} \sim Beta\left(\frac{pn}{2}, \frac{\nu}{2}\right)$. By using the density generator function of the matrix variate GT distribution, we have

$$f_R(r) = \frac{\Gamma\left(\frac{\nu+pn}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{pn}{2}\right)} 2r^{pn-1}\tau^{\frac{\nu}{2}}\left(\tau+r^2\right)^{-\frac{(\nu+pn)}{2}}, \quad r>0.$$

Therefore the pdf of $U = \frac{R^2}{\tau+R^2}$ is

$$f_U(u) = \frac{\Gamma\left(\frac{\nu+pn}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{pn}{2}\right)} u^{\frac{pn}{2}-1}(1-u)^{\frac{\nu}{2}-1}, \quad 0<u<1,$$

which is the pdf of the distribution $Beta\left(\frac{pn}{2}, \frac{\nu}{2}\right)$. □

3. Application in detecting outliers

In this section, we describe detecting matrix outliers by using the GMD as its significant application in the case of the matrix variate normal distribution.

Suppose that $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$ is an independent random sample from the distribution $N_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma}\otimes\boldsymbol{\Psi})$. By using the sample mean matrix

$$\bar{\boldsymbol{X}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{X}_i,$$

and the sample covariance matrix

$$(\widehat{\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}}) = \frac{1}{N-1}\sum_{i=1}^{N}\mathrm{vec}\left(\boldsymbol{X}_i^T - \bar{\boldsymbol{X}}^T\right)\left(\mathrm{vec}\left(\boldsymbol{X}_i^T - \bar{\boldsymbol{X}}^T\right)\right)^T, \tag{4}$$

an estimator for $\mathcal{GMD}(\boldsymbol{X}_i)$, $i = 1, 2, \ldots, N$ can be obtained as

$$\widehat{\mathcal{GMD}}(\boldsymbol{X}_i) = \sqrt{\left(\mathrm{vec}(\boldsymbol{X}_i^T - \bar{\boldsymbol{X}}^T)\right)^T(\widehat{\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}})^{-1}\mathrm{vec}(\boldsymbol{X}_i^T - \bar{\boldsymbol{X}}^T)}.$$

The distribution of $\widehat{\mathcal{GMD}}(\boldsymbol{X}_i)$ is obtained in the following theorem which its proof is clear by [5] and using Corollary 2.1 and properties of the matrix variate normal distributions. Let *iid* is the abbreviated form of independent and identically distributed.

Theorem 3.1. *If* $\boldsymbol{X}_i \overset{iid}{\sim} N_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma}\otimes\boldsymbol{\Psi})$*, for* $i = 1, 2, \ldots, N$*, then*

$$\frac{N}{(N-1)^2}\widehat{\mathcal{GMD}}^2(\boldsymbol{X}_i) \sim Beta\left(\frac{pn}{2}, \frac{N-pn-1}{2}\right).$$

Outliers are observations that have deviation from the pattern of the majority of the data. Since the GMD measures how far is an observation from the center $\boldsymbol{M}$ taking into account $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$, the matrix outliers can simply be defined as observations having a large GMD. In practice the values of $\boldsymbol{M}$ and $\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}$ are unknown, hence their estimators should be considered and the estimator of the GMD should be used for detecting.

For a sample $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$ from the distribution $N_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma}\otimes\boldsymbol{\Psi})$, based on the distribution in **Theorem 3.1**, a criterion can be provided for detecting the matrix outliers as follows:

$$\boldsymbol{X}_i \quad \text{is an outlier if} \quad \widehat{\mathcal{GMD}}(\boldsymbol{X}_i) > (N-1)\sqrt{\frac{B_{\frac{pn}{2},\frac{N-pn-1}{2}:\alpha}}{N}},$$

where $B_{a,b:\alpha}$ is the α-quantile (for example, the 97.5th percentile) of the distribution $Beta\,(a, b)$.

3.1. Robust estimators

Unfortunately, the sample mean matrix $\bar{\boldsymbol{X}}$ and the sample covariance matrix $(\widehat{\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}})$ are very sensitive to outliers, which make the presented approach suffers from the masking effect. To fix this problem, using their robust estimators is necessary for obtaining robust generalized Mahalanobis distance (RGMD). One of the highly robust estimators for multivariate data is the minimum covariance determinant (MCD) estimator introduced by [15]. Hubert et al.[9] gives a more detailed overview of the MCD estimator and its properties. The MCD version of $\bar{\boldsymbol{X}}$ and $(\widehat{\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}})$ can be obtained based on a subsample of size k, which has the lowest sample covariance determinant, by

$$\bar{\boldsymbol{X}}_{mcd} = \frac{1}{k}\sum_{j\in\mathcal{K}} \boldsymbol{X}_j,$$

and

$$(\widehat{\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}})_{mcd} = \frac{1}{k-1}\sum_{j\in\mathcal{K}} \mathrm{vec}\left(\boldsymbol{X}_j^T - \bar{\boldsymbol{X}}_{mcd}^T\right)\left(\mathrm{vec}\left(\boldsymbol{X}_j^T - \bar{\boldsymbol{X}}_{mcd}^T\right)\right)^T,$$

where $\mathcal{K}$ is a set containing the indices of the subsample whose its covariance matrix has the lowest possible determinant. Therefore, the MCD version of $\widehat{\mathcal{GMD}}(\boldsymbol{X}_i)$ for $i = 1, 2, \ldots, N$, is

$$\widehat{\mathcal{GMD}}_{mcd}(\boldsymbol{X}_i) = \sqrt{\left(\mathrm{vec}(\boldsymbol{X}_i^T - \bar{\boldsymbol{X}}_{mcd}^T)\right)^T (\widehat{\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}})_{mcd}^{-1}\mathrm{vec}(\boldsymbol{X}_i^T - \bar{\boldsymbol{X}}_{mcd}^T)}.$$

Using the robust estimators $\bar{\boldsymbol{X}}_{mcd}$ and $(\widehat{\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}})_{mcd}$ makes the distribution of $\widehat{\mathcal{GMD}}_{mcd}(\boldsymbol{X}_i)$ unknown (as it happens in the case of the multivariate robust Mahalanobis distance). Hence, considering the α-quantile of beta distribution in the critical value of detecting outliers might not be the best. In the case of the multivariate robust Mahalanobis distance, some related works provide different critical values for detecting outliers. For example, Filzmoser [3] has developed an adjusted quantile as a critical value in detecting outliers.

In [8], an approximate distribution for the multivariate robust Mahalanobis distance based on the MCD estimators is given. Since

$$\boldsymbol{X} \sim N_{p\times n}(\boldsymbol{M}, \boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}) \iff \mathrm{vec}(\boldsymbol{X}^T) \sim N_{pn}\left(\mathrm{vec}(\boldsymbol{M}^T), \boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}\right), \tag{5}$$

we can generalize the obtained results in [8] for observations of a matrix variate normal distribution and present an approximate distribution for $\widehat{\mathcal{GMD}}_{mcd}(\boldsymbol{X}_i)$ as

$$\frac{c(m-pn+1)}{mpn}\widehat{\mathcal{GMD}}^2_{mcd}(\boldsymbol{X}_i) \approx F_{pn,m-pn+1}, \tag{6}$$

where $X \approx F_{\nu_1,\nu_2}$ means that the random variable X approximately follows the F distribution with ν_1 and ν_1 degrees of freedom, and m and c are the unknown parameters that need to be estimated. The parameters m and c can be estimated in three ways: using simulations, using an asymptotic expression, or using an adjustment to the asymptotic expression. In this paper, we use simulations and the asymptotic expression to estimate m and c; see [8] for more details.

By the approximate distribution (6), based on the MCD estimators, we can provide a criterion for detecting the matrix outliers as follows:

$$\boldsymbol{X}_i \quad \text{is an outlier if} \quad \widehat{\mathcal{GMD}}_{mcd}(\boldsymbol{X}_i) > \sqrt{\frac{\hat{m}pn}{\hat{c}(\hat{m}-pn+1)}F_{pn,\hat{m}-pn+1:\alpha}},$$

where $\hat{m}$ and $\hat{c}$ are the estimations of m and c, respectively, and $F_{\nu_1,\nu_2:\alpha}$ is the α-quantile of the distribution F_{ν_1,ν_2}.

3.2. Breakdown value

A breakdown value is the maximum proportion of outliers that can safely be tolerated by an estimator. Similarly, the breakdown value for an outlier detection method can be defined as the maximum proportion of outliers that the procedure can detect successfully. According to the property (5) of the matrix variate normal distributions, the MCD estimators in the previous section have a breakdown value $\frac{N-k+1}{N}$, hence the number k determines the robustness of the estimators. In the case of $k = \lfloor \frac{(N+pn+1)}{2} \rfloor$ where $\lfloor . \rfloor$ denotes the integer part, the MCD estimators can achieve the highest possible breakdown value. A value close to $0.5N$ for k is recommended when a large proportion of contamination is expected. Otherwise, an intermediate value for k, such as $0.75N$, should be chosen to obtain a higher finite-sample efficiency. See [10] for more information about the breakdown value of different types of estimators.

3.3. Affine equivariant estimators

In addition to the robustness property, we are particularly interested in affine equivariant estimators of $\boldsymbol{M}$ and $\boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}$, see [16]. Any estimators $\hat{\boldsymbol{M}}$ and $(\widehat{\boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}})$ are called affine equivariant if they behave satisfactorily under affine transformations of the observations. This means that for all invertible matrices $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ and $\boldsymbol{B} \in \mathbb{R}^{n\times n}$, and any matrix $\boldsymbol{C} \in \mathbb{R}^{p\times n}$,

$$\hat{\boldsymbol{M}}(\boldsymbol{A}\boldsymbol{X}_1\boldsymbol{B} + \boldsymbol{C}, \dots, \boldsymbol{A}\boldsymbol{X}_N\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A}\hat{\boldsymbol{M}}(\boldsymbol{X}_1, \dots, \boldsymbol{X}_N)\boldsymbol{B} + \boldsymbol{C},$$

and

$$(\widehat{\boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}})(\boldsymbol{A}\boldsymbol{X}_1\boldsymbol{B} + \boldsymbol{C}, \dots, \boldsymbol{A}\boldsymbol{X}_N\boldsymbol{B} + \boldsymbol{C}) = \Big(\big(\boldsymbol{A}\widehat{\boldsymbol{\Sigma}}\boldsymbol{A}^T\big) \otimes \big(\boldsymbol{B}^T\widehat{\boldsymbol{\Psi}}\boldsymbol{B}\big)\Big)(\boldsymbol{X}_1, \dots, \boldsymbol{X}_N).$$

The affine equivariance property of the estimators of $\boldsymbol{M}$ and $\boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}$ makes the estimator of the GMD affine invariant (this can be easily shown similar to the proof of **Proposition 2.5**), and also the detection of outliers to be independent of the measurement scales of the variables, and the translation or rotation of the observations. Since the estimator of the GMD is affine equivariant, the properties and the procedures that are based on it can be calculated for standardized distributions without losing generality. According to this, for the procedures under normality, we can use the matrix variate standard normal distribution, i.e. $N_{p\times n}(\boldsymbol{0}_{p\times n}, \boldsymbol{I}_p \otimes \boldsymbol{I}_n)$ where $\boldsymbol{0}_{p\times n}$ is the $p \times n$ matrix of zeros and $\boldsymbol{I}_m$ is the m-dimensional identity matrix.

In the following proposition, we examine the affine equivariance property of the estimators presented for $\boldsymbol{M}$ and $\boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}$.

Proposition 3.2. *The sample mean matrix $\bar{\boldsymbol{X}}$ and the sample covariance matrix $(\widehat{\boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}})$, and also their MCD version, i.e. $\bar{\boldsymbol{X}}_{mcd}$ and $(\widehat{\boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}})_{mcd}$, are affine equivariant.*

Proof. We only prove the affine equivariance property of $\bar{\boldsymbol{X}}$ and $(\widehat{\boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}})$, the proof for $\bar{\boldsymbol{X}}_{mcd}$ and $(\widehat{\boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}})_{mcd}$ is similar. Considering $\boldsymbol{Y}_i = \boldsymbol{A}\boldsymbol{X}_i\boldsymbol{B} + \boldsymbol{C}$, $i = 1, 2, \dots, N$, we have

$$\bar{\boldsymbol{Y}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Y}_i = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{A}\boldsymbol{X}_i\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A}\Big(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{X}_i\Big)\boldsymbol{B} + \boldsymbol{C} = \boldsymbol{A}\bar{\boldsymbol{X}}\boldsymbol{B} + \boldsymbol{C},$$

which shows the affine equivariance property of $\bar{\boldsymbol{X}}$. For the sample covariance matrix, from (4) we can write

$$\begin{aligned}
(\widehat{\boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}})_{\boldsymbol{Y}} &= \frac{1}{N-1}\sum_{i=1}^{N} \mathsf{vec}\left(\boldsymbol{Y}_i^T - \bar{\boldsymbol{Y}}^T\right)\left(\mathsf{vec}\left(\boldsymbol{Y}_i^T - \bar{\boldsymbol{Y}}^T\right)\right)^T \\
&= \frac{1}{N-1}\sum_{i=1}^{N} \mathsf{vec}\left(\boldsymbol{B}^T(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T\boldsymbol{A}^T\right)\left(\mathsf{vec}\left(\boldsymbol{B}^T(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T\boldsymbol{A}^T\right)\right)^T \\
&= \frac{1}{N-1}\sum_{i=1}^{N} \mathsf{vec}\left((\boldsymbol{A}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})\boldsymbol{B})^T\right)\left(\mathsf{vec}\left((\boldsymbol{A}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})\boldsymbol{B})^T\right)\right)^T.
\end{aligned}$$

By Theorem 1.17 (i) of [7], we know

$$\mathrm{vec}\left((\boldsymbol{A}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})\boldsymbol{B})^T\right)=(\boldsymbol{A}\otimes\boldsymbol{B}^T)\mathrm{vec}\left((\boldsymbol{X}_i-\bar{\boldsymbol{X}})^T\right).$$

Hence,

$$\begin{aligned}
(\widehat{\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}})_{Y}&=\frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{A}\otimes\boldsymbol{B}^T)\mathrm{vec}\left((\boldsymbol{X}_i-\bar{\boldsymbol{X}})^T\right)\left(\mathrm{vec}\left((\boldsymbol{X}_i-\bar{\boldsymbol{X}})^T\right)\right)^T(\boldsymbol{A}\otimes\boldsymbol{B}^T)^T\\
&=(\boldsymbol{A}\otimes\boldsymbol{B}^T)\left[\frac{1}{N-1}\sum_{i=1}^{N}\mathrm{vec}\left((\boldsymbol{X}_i-\bar{\boldsymbol{X}})^T\right)\left(\mathrm{vec}\left((\boldsymbol{X}_i-\bar{\boldsymbol{X}})^T\right)\right)^T\right](\boldsymbol{A}\otimes\boldsymbol{B}^T)^T\\
&=(\boldsymbol{A}\otimes\boldsymbol{B}^T)(\widehat{\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}})(\boldsymbol{A}\otimes\boldsymbol{B}^T)^T.
\end{aligned}$$

Consider the positive definite matrices $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}$ and $\widehat{\boldsymbol{\Psi}}_{\boldsymbol{X}}$ of dimension $p\times p$ and $n\times n$, respectively, such that $(\widehat{\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}})=\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}\otimes\widehat{\boldsymbol{\Psi}}_{\boldsymbol{X}}$. By properties of the Kronecker product, we have

$$\begin{aligned}
(\widehat{\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}})_{Y}&=(\boldsymbol{A}\otimes\boldsymbol{B}^T)(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}\otimes\widehat{\boldsymbol{\Psi}}_{\boldsymbol{X}})(\boldsymbol{A}\otimes\boldsymbol{B}^T)^T\\
&=(\boldsymbol{A}\otimes\boldsymbol{B}^T)(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}\otimes\widehat{\boldsymbol{\Psi}}_{\boldsymbol{X}})(\boldsymbol{A}^T\otimes\boldsymbol{B})\\
&=(\boldsymbol{A}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}\otimes\boldsymbol{B}^T\widehat{\boldsymbol{\Psi}}_{\boldsymbol{X}})(\boldsymbol{A}^T\otimes\boldsymbol{B})\\
&=(\boldsymbol{A}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{X}}\boldsymbol{A}^T\otimes\boldsymbol{B}^T\widehat{\boldsymbol{\Psi}}_{\boldsymbol{X}}\boldsymbol{B}).
\end{aligned}$$

Thus $(\widehat{\boldsymbol{\Sigma}\otimes\boldsymbol{\Psi}})$ is affine equivariant. □

4. Illustrative examples

In this section, we find outliers in simulated examples and real ones, to illustrate the applicability of the introduced distance in detecting outliers. For this purpose, in all examples, we consider $k=\lfloor\frac{(N+pn+1)}{2}\rfloor$ in obtaining the MCD estimators and use simulations for estimating m and c.

4.1. Simulated examples

Here, we present three simulated examples which show the GMD is well suited for detecting matrix outliers. For this purpose, we first generate a sample from a matrix variate normal distribution, and then we contaminate the generated sample to have outliers by replacing a number of observations having the largest GMD with a sample from a specified distribution.

4.1.1. Example 1

In this example, we contaminate a generated sample of size $N=137$ from the distribution

$$N_{3\times2}\left(\begin{pmatrix}-4&5\\3&-3\\-5&4\end{pmatrix},\begin{pmatrix}5&1&-3\\1&5&1\\-3&1&5\end{pmatrix}\otimes\begin{pmatrix}1.5&-1.5\\-1.5&4\end{pmatrix}\right),$$

to have outliers by replacing a sample of 17 from the distribution $N_{3\times2}(\boldsymbol{0}_{3\times2},2.5\boldsymbol{I}_3\otimes\boldsymbol{I}_2)$. The plot of the GMD versus the RGMD (distance-distance plot) for the contaminated sample is shown in Figure 1. Since in this example, the critical values for detecting outliers are

$$136\sqrt{\frac{B_{3,65:0.975}}{137}}=3.741827,$$

and

$$\sqrt{\frac{86.01934 \times 6}{0.8887568(81.01934)} F_{6,81.01934:0.975}} = 4.290827,$$

the horizontal and vertical lines are drawn respectively at values **4.290827** and **3.741827** in Figure 1. It can be seen that by considering the GMD only one of the real outliers is detected while all real outliers are detected by the RGMD.

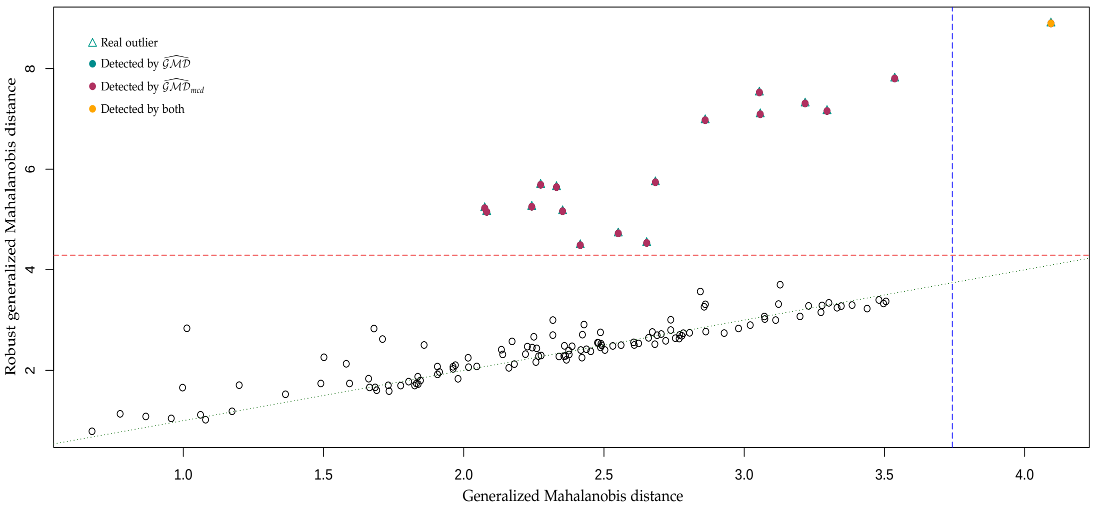

Figure 1: Distance-distance plot of example 1.

4.1.2. *Example 2*

In this example, we generate a sample of size $N = 689$ from the distribution $N_{4\times 6}(\mathbf{0}_{4\times 6}, \boldsymbol{I}_4 \otimes \boldsymbol{I}_6)$, and contaminate it by replacing a sample of size 101 from the distribution

$$GT_{4\times 6}\left(-1.725\boldsymbol{1}_{4\times 6}, \begin{pmatrix} 1 & -1 & 2 & 0 \\ -1 & 3.5 & -1 & 1 \\ 2 & -1 & 6 & 0 \\ 0 & 1 & 0 & 4 \end{pmatrix} \otimes 1.75\boldsymbol{I}_6; 4.01, 4.01\right),$$

where $\boldsymbol{1}_{4\times 6}$ is the 4×6 matrix of ones. Here, the critical values are

$$688\sqrt{\frac{B_{12,332:0.975}}{689}} = 6.238839,$$

and

$$\sqrt{\frac{583.3614(24)}{0.9620942(560.3614)} F_{24,560.3614:0.975}} = 6.576866.$$

Figure 2 presents the distance-distance plot for the contaminated sample. In this figure, the horizontal and vertical lines are drawn at values **6.576866** and **6.238839**, respectively. It is observed that all real outliers are detected by the RGMD, while some of them are not detected by the GMD.

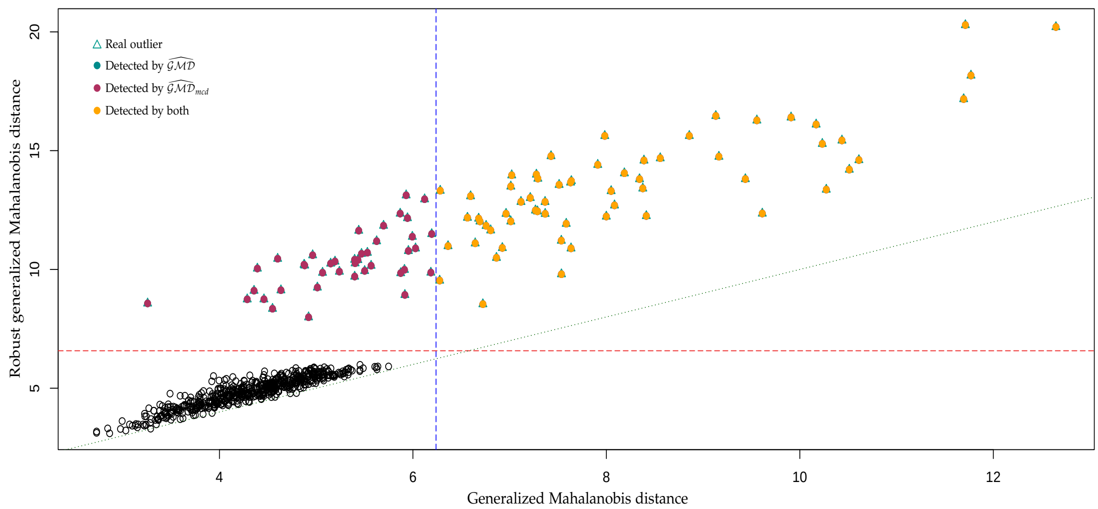

Figure 2: Distance-distance plot of example 2.

4.1.3. *Example 3*

To have a contaminated sample, in this example, we consider a sample of size $N = 1000$ from the distribution $N_{14\times14}(\mathbf{0}_{14\times14}, \boldsymbol{I}_{14} \otimes \boldsymbol{I}_{14})$ and contaminate it by replacing a sample of size 100 from the distribution $N_{14\times14}(2.5\boldsymbol{1}_{14\times14}, \boldsymbol{I}_{14} \otimes \boldsymbol{I}_{14})$. The critical values are

$$999\sqrt{\frac{B_{98,401.5:0.975}}{1000}} = 15.21992,$$

and

$$\sqrt{\frac{537.6831(196)}{0.9363094(342.6831)}F_{196,342.6831:0.975}} = 20.48051.$$

Figure 3 is the distance-distance plot for the contaminated sample which the horizontal and vertical lines are drawn respectively at values **20.48051** and **15.21992**. As can be seen from Figure 3, all real outliers are detected by the RGMD, while seven real outliers and four non-outliers are detected as outliers by the GMD.

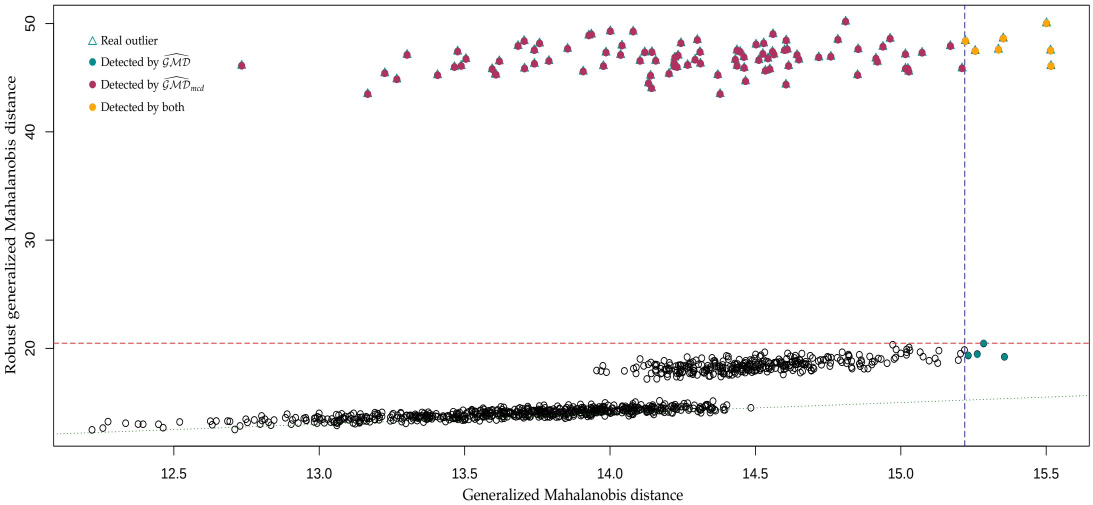

Figure 3: Distance-distance plot of example 3.

4.2. Real examples

To show the applicability of the obtained results in detecting outliers, we find outliers of two real datasets Dow-Jones Dividends data and Ford Motor Company data.

4.2.1. Dow-Jones Dividends data

We find outliers of Dow-Jones Dividends data which is a real dataset. Dow-Jones Dividends data consists of two components dividend and divisor of Dow-Jones Industrial Common Stocks for quarters from 1920 to 1934 and the dimension of matrices is 2×4; for more details see [13].

Recently, Rezaei et al.[14] have fitted some matrix variate distributions such as normal to this dataset. By considering the matrix variate normal distribution, we calculated the GMD and the RGMD for this dataset. The calculated distances are given in Table 1. According to the dimension of matrices, the critical values for detecting outliers are

$$14\sqrt{\frac{B_{4,3:0.975}}{15}} = 3.394594,$$

and

$$\sqrt{\frac{9.644833(8)}{0.8531358(2.644833)}F_{8,2.644833:0.975}} = 25.43599.$$

Based on the critical values, the matrix data related to the years 1928 and 1932 are detected as outliers by both types of distances. The detected outliers of Dow-Jones Dividends data can be seen in Figure 4 where the horizontal and vertical lines are drawn respectively at values **25.43599** and **3.394594**.

Table 1: GMD and RGMD of Dow-Jones Dividends data. The distances related to the outlier data detected by GMD and RGMD were colored dark cyan and maroon, respectively. Also, the years related to the outlier data detected by both were marked with a light orange box.

Year	GMD	RGMD	Year	GMD	RGMD	Year	GMD	RGMD
1920	1.781426	1.613546	1925	2.680427	2.624501	1930	2.934475	2.815863
1921	2.430803	3.188700	1926	3.119204	3.098863	1931	3.126866	9.916637
1922	2.489488	2.936198	1927	2.395432	3.026937	1932	3.696810	38.912858
1923	2.870990	2.920636	1928	3.682438	25.459493	1933	2.902705	2.780853
1924	2.257682	3.316625	1929	3.076219	3.015911	1934	2.285657	2.160099

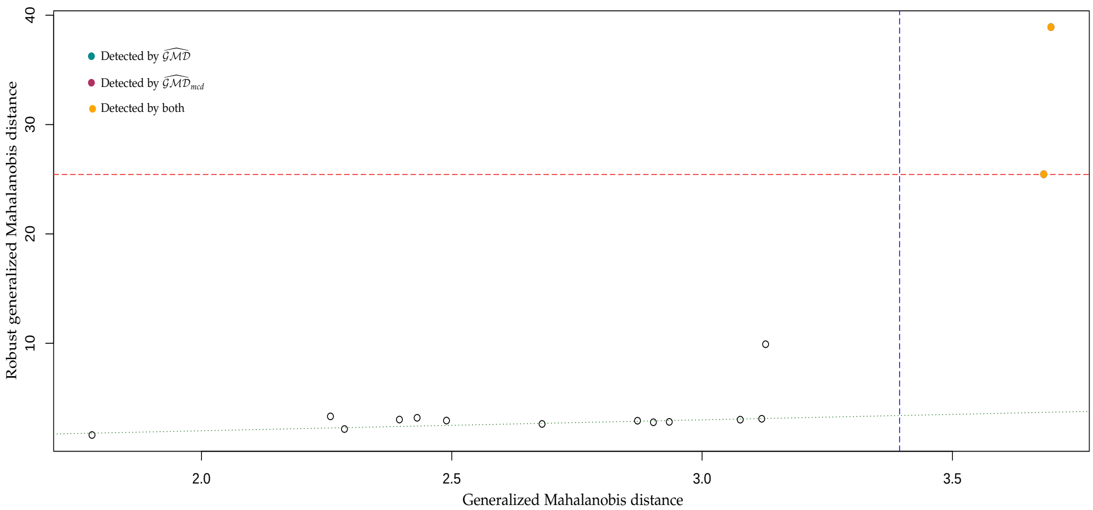

Figure 4: Distance-distance plot of Dow-Jones Dividends data.

4.2.2. *Ford Motor Company data*

Ford Motor Company data contains the average values of open, high, low, and close prices for quarters from 1992 to 2021. These data are derived from stock prices of Ford Motor Company in https://finance.yahoo.com/quote/F?p=F and are presented in Tables 2–4.

We calculated the GMD and the RGMD for this dataset, given in Tables 2–4, by considering the matrix variate normal distribution for the data. To detect outliers in this dataset, the critical values are

$$29\sqrt{\frac{B_{8,6.5:0.975}}{30}} = 4.701158,$$

and

$$\sqrt{\frac{19.92367(16)}{0.9047772(4.923671)}F_{16,4.923671:0.975}} = 21.60489.$$

By comparing the calculated distances in Tables 2–4 with the critical values, the matrix data of the years 1998, 1999, 2000, and 2015 are detected as outliers by both types of distances, while the matrix data of the years 1997, 2003, and 2021 are detected as outliers by the RGMD and the matrix data of the year 2001 is detected by the GMD. Figure 5 shows the related distance-distance plot where the horizontal and vertical lines are drawn respectively at values **21.60489** and **4.701158**.

Table 2: Ford Motor Company's quarterly average stock prices and their distances (1992-2001). The distances related to the outlier data detected by GMD and RGMD were colored dark cyan and maroon, respectively. Also, the years related to the outlier data detected by GMD, RGMD, and both were marked with dark cyan, maroon, and light orange boxes, respectively.

Year	Quarter	Open	High	Low	Close	GMD	RGMD
1992	Q1	5.78	6.92	5.66	6.42	3.107039	4.006377
	Q2	7.76	8.60	7.54	8.25		
	Q3	8.04	8.20	7.13	7.66		
	Q4	7.22	7.69	6.71	7.38		
1993	Q1	8.25	9.31	8.13	8.76	2.274481	3.022121
	Q2	9.69	10.24	9.17	9.67		
	Q3	9.51	10.13	8.99	9.67		
	Q4	10.89	11.81	10.63	11.38		
1994	Q1	11.76	12.50	10.98	11.43	2.482584	2.857500
	Q2	10.64	11.25	10.06	10.65		
	Q3	11.11	11.56	10.19	10.81		
	Q4	10.25	10.71	9.56	10.28		
1995	Q1	9.63	10.22	9.02	9.52	2.333578	3.106487
	Q2	10.13	10.84	9.75	10.48		
	Q3	10.89	11.74	10.51	11.06		
	Q4	10.78	11.21	10.26	10.45		
1996	Q1	10.93	11.86	10.61	11.57	3.000183	3.259710
	Q2	13.02	13.52	12.12	12.74		
	Q3	11.95	12.33	11.22	11.82		
	Q4	11.59	12.29	11.21	11.71		
1997	Q1	11.83	12.45	11.42	11.72	4.380136	33.861659
	Q2	12.62	13.72	12.29	13.41		
	Q3	14.86	16.25	14.68	15.69		
	Q4	16.17	17.79	15.45	16.45		
1998	Q1	19.05	21.15	18.09	20.97	5.189126	77.127494
	Q2	26.18	29.81	25.06	28.70		
	Q3	29.16	30.90	25.68	27.23		
	Q4	28.56	31.20	26.16	30.77		
1999	Q1	32.65	34.60	30.96	32.50	4.950197	32.131579
	Q2	32.71	35.41	30.13	32.51		
	Q3	29.17	30.13	26.19	27.64		
	Q4	28.69	30.60	27.15	29.07		
2000	Q1	26.42	27.92	23.58	25.15	5.145706	71.271911
	Q2	27.49	30.00	25.08	27.12		
	Q3	25.26	28.31	24.12	25.44		
	Q4	24.92	26.19	22.54	24.10		
2001	Q1	26.51	29.70	25.86	28.04	4.986016	4.581256
	Q2	26.97	29.00	24.95	26.13		
	Q3	23.21	24.00	19.56	20.90		
	Q4	17.28	18.77	15.56	16.90		

Table 3: Ford Motor Company's quarterly average stock prices and their distances (2002-2011). The distances related to the outlier data detected by GMD and RGMD were colored dark cyan and maroon, respectively. Also, the years related to the outlier data detected by GMD, RGMD, and both were marked with dark cyan, maroon, and light orange boxes, respectively.

Year	Quarter	Open	High	Low	Close	GMD	RGMD
2002	Q1	15.15	16.74	14.28	15.56	4.156131	4.086770
	Q2	16.53	17.61	15.24	16.55		
	Q3	13.62	13.76	10.47	11.68		
	Q4	9.92	11.13	8.17	9.71		
2003	Q1	9.00	9.51	7.83	8.32	3.588619	21.898671
	Q2	9.42	11.08	9.00	10.60		
	Q3	11.24	11.95	10.51	11.13		
	Q4	12.10	14.38	11.73	13.78		
2004	Q1	14.69	15.46	13.46	13.95	3.193159	4.013046
	Q2	14.53	16.12	13.85	15.29		
	Q3	14.81	15.18	13.85	14.29		
	Q4	13.80	14.59	13.13	13.95		
2005	Q1	13.50	13.67	12.16	12.38	4.417089	4.491116
	Q2	10.28	11.21	9.33	9.78		
	Q3	10.30	10.81	9.77	10.19		
	Q4	8.86	8.99	7.73	8.06		
2006	Q1	8.08	8.65	7.67	8.17	3.020631	3.283652
	Q2	7.45	7.58	6.61	7.01		
	Q3	7.36	8.31	6.71	7.71		
	Q4	8.18	8.67	7.55	7.97		
2007	Q1	7.83	8.61	7.49	7.98	3.358693	3.414312
	Q2	8.10	9.02	7.89	8.60		
	Q3	8.61	9.17	7.63	8.27		
	Q4	8.28	8.58	7.22	7.70		
2008	Q1	6.61	6.76	5.49	6.30	4.595666	4.512595
	Q2	6.97	8.13	5.61	6.62		
	Q3	4.73	5.84	4.24	4.82		
	Q4	3.33	3.92	1.61	2.39		
2009	Q1	2.02	2.67	1.65	2.17	4.315896	4.426430
	Q2	4.83	6.35	4.11	5.93		
	Q3	7.43	8.21	6.37	7.60		
	Q4	7.97	9.16	7.54	8.63		
2010	Q1	11.04	12.85	10.76	11.72	4.247945	4.175596
	Q2	12.51	13.40	10.75	11.61		
	Q3	11.61	13.06	10.81	12.10		
	Q4	14.27	16.31	14.14	15.62		
2011	Q1	16.09	16.89	14.56	15.30	4.313366	4.294219
	Q2	15.11	15.59	13.83	14.73		
	Q3	12.60	12.77	10.38	11.00		
	Q4	10.50	11.92	9.55	11.01		

Table 4: Ford Motor Company's quarterly average stock prices and their distances (2012-2021). The distances related to the outlier data detected by GMD and RGMD were colored dark cyan and maroon, respectively. Also, the years related to the outlier data detected by GMD, RGMD, and both were marked with dark cyan, maroon, and light orange boxes, respectively.

Year	Quarter	Open	High	Low	Close	GMD	RGMD
2012	Q1	12.07	13.03	11.66	12.43	4.153089	4.114499
	Q2	11.36	11.70	10.19	10.48		
	Q3	9.39	10.05	9.00	9.48		
	Q4	10.88	11.95	10.37	11.85		
2013	Q1	12.96	13.69	12.38	12.90	4.159697	4.423241
	Q2	14.19	15.30	13.27	14.95		
	Q3	16.36	17.65	15.83	16.65		
	Q4	17.10	17.49	15.99	16.54		
2014	Q1	15.21	16.03	14.78	15.32	4.034922	4.539254
	Q2	16.11	16.80	15.78	16.61		
	Q3	17.29	17.84	16.07	16.41		
	Q4	14.88	15.60	13.68	15.11		
2015	Q1	15.56	16.34	14.91	15.73	4.818559	31.180452
	Q2	15.80	15.92	15.21	15.33		
	Q3	14.69	15.07	12.56	14.09		
	Q4	14.29	15.12	13.56	14.41		
2016	Q1	12.93	13.50	11.65	12.65	4.354036	3.948091
	Q2	13.46	13.85	12.47	13.21		
	Q3	12.72	13.14	12.01	12.44		
	Q4	12.02	12.74	11.58	11.94		
2017	Q1	12.44	12.93	11.92	12.18	2.802329	3.685898
	Q2	11.46	11.59	10.89	11.26		
	Q3	11.22	11.70	10.85	11.41		
	Q4	12.34	12.64	12.01	12.43		
2018	Q1	11.37	11.99	10.41	10.89	3.741338	4.508144
	Q2	11.33	11.87	10.90	11.29		
	Q3	10.20	10.43	9.43	9.59		
	Q4	9.56	9.77	8.18	8.87		
2019	Q1	8.38	9.00	7.99	8.78	3.214257	3.922030
	Q2	9.65	10.44	9.21	10.07		
	Q3	9.68	9.94	9.05	9.29		
	Q4	8.97	9.35	8.63	8.98		
2020	Q1	8.42	8.67	6.45	6.87	4.379906	4.489953
	Q2	5.13	6.53	4.79	5.63		
	Q3	6.51	7.28	6.23	6.70		
	Q4	7.90	9.11	7.66	8.53		
2021	Q1	10.44	12.72	10.14	11.49	4.644665	30.520030
	Q2	12.85	14.83	12.27	13.64		
	Q3	13.99	14.70	12.60	13.71		
	Q4	17.13	19.95	16.52	19.01		

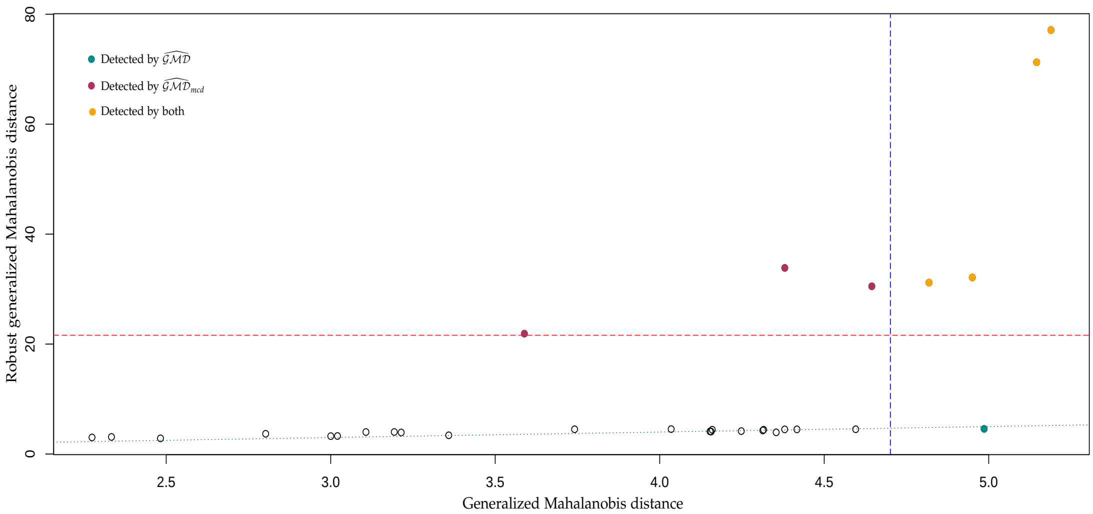

Figure 5: Distance-distance plot of Ford Motor Company data.

5. A Simulation Study

In this section, we present a simulation study to investigate how the presented approach of detecting outliers performs as well as its computation time with increasing data dimensions. For this purpose, we consider $p, n = 6, 18, 30, 42$, and for each dimension from $p \times n = 6 \times 6$ to $p \times n = 42 \times 42$, we generate 15 random samples of size 8900 from the distribution $N_{p\times n}(\mathbf{0}_{p\times n}, \boldsymbol{I}_p \otimes \boldsymbol{I}_n)$ and contaminate each generated sample to have outliers by replacing 890 observations having the largest GMD with a sample of size 890 from the distribution $N_{p\times n}(-10\mathbf{1}_{p\times n}, 1.5\boldsymbol{I}_p \otimes 0.75\boldsymbol{I}_n)$. After being contaminated with outlier data, for each contaminated sample of size 8900, we find outliers by the proposed method, obtain the number of the real outliers that are not detected (false negatives) and the number of the non-outliers that are detected as outliers (false positives), and compute the related computation times (in seconds). It should be noted that in outlier detection, we consider $k = \lfloor \frac{(8900+pn+1)}{2} \rfloor$ to obtain the MCD estimators and use the asymptotic expression to estimate m and c.

The obtained results are reported in Table 5. It contains the averages of the percentage of false negatives (PFN), the percentage of false positives (PFP), and the computation times ($Time$) of the 15 contaminated samples for different dimensions. All the computations are performed in R software package (R x64 4.2.1) using a machine equipped with an Intel Core i5-3230M 2.60 GHz processor and 4 GB RAM. The R code can be obtained on request from the authors.

Table 5 shows that by increasing the dimension ($p \times n$), the average of PFN decreases, and the average of PFP increases for the GMD. This means that as the dimension of matrix observations increases, by the GMD, more real outliers are detected and at the same time more non-outliers are detected as outliers. In addition, the results show that for the RGMD, the averages of PFN and PFP are equal to zero, which seems that increasing the dimension (as much as was considered here) does not affect the detection of outlier data based on the RGMD.

Based on the results obtained for computing time in Table 5, it can be said that the computation time of the GMD is always less than the computation time of the RGMD, and with the increase of the dimension, the computation times of both types of distances increase.

Table 5: Averages of PFN, PFP, and $Time$ (in seconds) based on 15 contaminated samples.

p	n	GMD			RGMD		
		PFN	PFP	$Time$	PFN	PFP	$Time$
6	6	71.38577	0	0.02496	0	0	1.33120
	18	69.33333	0	0.15704	0	0	7.16428
	30	64.92884	0	0.41601	0	0	16.67331
	42	60.82397	0	0.80392	0	0	30.53965
18	6	69.03371	0	0.15600	0	0	6.98881
	18	56.95131	0	1.30520	0	0	47.73608
	30	44.74906	0.05390	3.56096	0	0	131.27320
	42	35.19850	0.11617	7.46647	0	0	270.08350
30	6	65.03371	0	0.42267	0	0	16.74084
	18	44.56180	0.08128	3.75597	0	0	139.12780
	30	30.30712	0.20374	10.02146	0	0	372.36430
	42	20.89888	0.50792	20.62218	0	0	778.62360
42	6	60.95880	0	0.81251	0	0	30.27416
	18	34.86142	0.04592	7.02001	0	0	259.35530
	30	20.72659	0.30223	20.41971	0	0	783.47220
	42	13.01873	0.69151	40.88713	0	0	1657.25500

6. Conclusion

In this paper, a new distance for matrix observations called generalized Mahalanobis distance (GMD) was introduced. For observations of the matrix variate elliptically contoured distributions, some properties of the GMD were investigated and its distribution was presented. As one of its significant applications, the method of detecting outliers in observations having matrix variate normal distribution was described. Finally, outliers of three simulated examples and two real datasets were detected by using the described method to show the applicability of the GMD in detecting outliers, and a simulation study was presented to examine how the presented approach of detecting outliers performs. The presented approach of detecting outliers only is for the case of the matrix variate normal distribution. Providing a general method to detect outliers among a sample of any matrix variate distribution can be considered as future research.

Acknowledgments

The authors would like to thank an anonymous referee and the Co-Editor Prof. Biljana Č. Popović for providing helpful comments and suggestions, which improved this paper significantly.

References

[1] E. Cabana, R. E. Lillo, H. Laniado, Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators, Statistical papers 62(2021) 1583–1609.
[2] K. T. Fang, Y. T. Zhang, Generalized Multivariate Analysis, Springer, New York, 1990.
[3] P. Filzmoser, A multivariate outlier detection method, Proceedings of the Seventh International Conference on Computer Data Analysis and Modeling (2004) 18–22.
[4] H. Ghorbani, Mahalanobis distance and its application for detecting multivariate outliers, Facta Universitatis Series Mathematics and Informatics 34(2019) 583–595.
[5] R. Gnanadesikan, J. R. Kettenring, Robust estimates, residuals, and outlier detection with multiresponse data, Biometrics (1972) 81–124.
[6] G. H. Golub, C. F. Van Loan, Matrix computations, (3rd edition), Johns Hopkins University Press, London, 2013.
[7] A. K. Gupta, T. Varga, T. Bodnar, Elliptically Contoured Models in Statistics and Portfolio Theory, Springer, New York, 2013.
[8] J. Hardin, D. M. Rocke, The distribution of robust distances, Journal of Computational and Graphical Statistics 14(2005) 928–946.

[9] M. Hubert, M. Debruyne, P. J. Rousseeuw, Minimum covariance determinant and extensions, Wiley Interdisciplinary Reviews: Computational Statistics 10(2018) e1421.
[10] H. P. Lopuhaa, P. J. Rousseeuw, Breakdown points of affine equivariant estimators of multivariate location and covariance matrices, The Annals of Statistics (1991) 229–248.
[11] B. Magyar, A. Kenyeres, S. Tóth, I. Hajdu, R. Horváth, Spatial outlier detection on discrete GNSS velocity fields using robust Mahalanobis-distance-based unsupervised classification, GPS Solution 26(2022) 1–11.
[12] P. C. Mahalanobis, On the generalised distance in statistics, Proceedings of the National Institute of Sciences of India 2(1936) 49–55.
[13] P. Rappaport, E. N. White, Was There a Bubble in the 1929 Stock Market?, Journal of Economic History 53(1993) 549–574.
[14] A. Rezaei, F. Yousefzadeh, R. B. Arellano-Valle, Scale and shape mixtures of matrix variate extended skew normal distributions, Journal of Multivariate Analysis 179(2020) 104649.
[15] P. J. Rousseeuw, Multivariate estimation with high breakdown point, Mathematical statistics and applications 8(1985) 283–297.
[16] P. J. Rousseeuw, A. M. Leroy, Robust Regression and Outlier Detection, John Wiley & Sons, New York, 1987.