Available online at www.sciencedirect.com
ScienceDirect

ELSEVIER

Linear Algebra and its Applications 419 (2006) 93–106

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

Bounds for the largest Mahalanobis distance

Eugene G. Gath *, Kevin Hayes

Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland

Received 30 March 2006; accepted 14 April 2006
Available online 16 June 2006
Submitted by I. Olkin

Abstract

Upper and lower bounds for the magnitude of the largest Mahalanobis distance, calculated from n multivariate observations of length p, are derived. These bounds are multivariate extensions of corresponding bounds that arise for the most deviant Z-score calculated from a univariate sample of size n. The approach taken is to pose optimization problems in a mathematical context and to employ variational methods to obtain solutions. The attainability of the bounds obtained is demonstrated. Bounds for related quantities (elements of the "hat matrix") are also derived.

AMS classification: 26D15; 65C60; 62G15; 15A45

Keywords: Mahalanobis distance; Lagrange; Multivariate outlier; Bound; Inequality

1. Contexts: Statistical and general

Let $y_1, \dots, y_n$ be points in $\mathbb{R}$. Define Z-scores

$$z_i = \frac{(y_i - \bar{y})}{s}$$

for $i = 1, \dots, n$, where the sample mean is calculated as

$$\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$$

* Corresponding author.
E-mail addresses: eugene.gath@ul.ie (E.G. Gath), kevin.hayes@ul.ie (K. Hayes).

0024-3795/$ - see front matter
doi:10.1016/j.laa.2006.04.007

and the standard deviation is calculated as

$$s = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n-1}}.$$

Shiffler [9] showed that $(n-1)/\sqrt{n}$ is an upper bound for $\max_i |z_i|$, the magnitude of the most deviant Z-score in a univariate sample of size n. Olkin [6] noted that although this result is often attributed to Samuelson [8] historically the upper bound should identified with Thompson [7]. Various generalizations of this result have been considered. Hayes [4] showed that the magnitude of the most deviant Z-score has the lower bound $\sqrt{(n-1)/n}$, and that attaining this lower bound necessarily requires that Shiffler and Harsha's [10] upper bound for the standard deviation be also attained. Gray and Woodall [3] discussed the extension of the upper bound to residuals obtainable from the linear model with independent error terms of common variance.

In statistical analysis, the row vectors x_i' in $\mathbb{R}^p$, for $i = 1, \ldots, n$, of an $n \times p$ matrix X are often interpreted as independent realizations from a multivariate normal density $\mathrm{MVN}_p(\mu, \Sigma)$. The maximum likelihood estimates of the p dimensional vector mean μ and $p \times p$ variance matrix Σ are $\hat{\mu} = \bar{x} = X'u/n$, and $\hat{\Sigma} = X'CX/n$, respectively, where u is the vector of ones in $\mathbb{R}^n$ and the $n \times n$ matrix $C = I_n - \frac{1}{n}uu'$, where I_n is the $n \times n$ identity matrix. Observe that the matrix C is idempotent and that the matrix CX has rows $x_i' - \bar{x}'$, for $i = 1, \ldots, n$. Typically, the unbiased estimate $S = X'CX/(n-1)$ of Σ is used in place of $\hat{\Sigma}$.

If S is non-singular, the Mahalanobis distance of the p dimensional vector x_i is defined as $D_i^2 = (x_i - \bar{x})'S^{-1}(x_i - \bar{x})$. This is a generalization to the multivariate setting of the quantities z_i^2 of the univariate case. Two standard texts on outlier detection Barnett and Lewis [2], and Hawkins [5] provide a thorough discussion on the role of the Mahalanobis distance in the outlier literature.

Using a useful matrix identity, Olkin [6] verified that $(n-1)^2/n$ is an upper bound for the maximum Mahalanobis distance in the special case of bivariate data. This result follows from a more general upper bound for bivariate data derived by Olkin [6] who also noted that the same upper bound follows as a consequence of Corollary 1 to Theorem 1 of Arnold and Groeneveld [1] for higher dimensional data.

In this paper, an upper bound for the largest Mahalanobis distance obtainable from a multivariate sample is derived by posing the problem in a general mathematical context and it is then solved as a Lagrange optimization problem. The largest Mahalanobis distance obtainable in a sample of n multivariate p length vectors has upper bound $(n-1)^2/n$, and when $n > p+1$, the upper bound is attainable for k vectors, for each $k = 1, \ldots, p$. Lower bounds for the largest Mahalanobis distance are also obtained in various cases, using inter alia a *dual approach* to the optimization problems; the attainability of these bounds is demonstrated. Finally, the maximum and minimum values of $(x_i - \bar{x})'S^{-1}(x_j - \bar{x})$, for $i \neq j$, are obtained.

The assumption of multivariate normality was introduced to motivate the estimates $\bar{x}$ and S, but it is not essential and so is dispensed with for the remainder of the paper; so henceforth we need not regard the x_is as random variables. We conclude this section by framing the problems in a generalized mathematical setting as follows. Let $x_1, \ldots, x_n$ be n vectors (data points, observations) in $\mathbb{R}^p$. Define the vector

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.1}$$

and the $p \times p$ matrix

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'. \tag{1.2}$$

We assume that the matrix S thus obtained is invertible and positive definite (thus requiring that $n \geqslant p$). The $n \times n$ matrix H is defined by its components

$$h_{ij} = \frac{1}{n-1}(x_i - \bar{x})' S^{-1}(x_j - \bar{x}). \tag{1.3}$$

H is a non-negative, symmetric, idempotent matrix with trace p. Also observe that each row and column of H sums to zero, that is

$$\sum_{j=1}^{n} h_{ij} = \sum_{j=1}^{n} h_{ji} = 0.$$

For vector x_i define the Mahalanobis distance D_i by

$$D_i^2 = (n-1)h_{ii}.$$

Define, for a given set of points $x_1, \ldots, x_n$, the maximum Mahalanobis distance

$$M = \max_{i=1,\ldots,n} D_i^2.$$

Our goal is to optimize M and the h_{ij}s over points in $\mathbb{R}^p$, subject to the constraints that $\bar{x}$ and S are fixed. This is achieved using the attractive and natural framework of Lagrangian methods.

2. The maximum Mahalanobis distance

First we find the maximum[1] of h_{kk}, for a given k subject to the mean and variance, as defined in (1.1) and (1.2), being fixed. Without loss of generality, we take the mean $\bar{x} = 0$, following an overall translation of all the (data) points. The rank of S must be p, since it is invertible. This fact, along with the mean constraint, requires that the number of points must satisfy $n \geqslant p + 1$. We will show that the maximum value of h_{kk} is $(n-1)/n$ and that a necessary and sufficient condition for this to be achieved is that the vectors $x_2, \ldots, x_n$ be coplanar (subject to S being invertible). We also construct an example, using a suitable choice of data points, which demonstrates that it is possible to make the p largest h_{kk}s all equal to $(n-1)/n$; that is the p largest Mahalanobis distances are all equal to $(n-1)^2/n$.

Theorem 2.1. *The maximum value of*

$$h_{kk} = \frac{x_k' S^{-1} x_k}{n-1}$$

is

$$\frac{n-1}{n},$$

where $x_1, \ldots, x_n$ are points in $\mathbb{R}^p$, constrained such that

$$\sum_{i=1}^{n} x_i = 0$$

[1] Technically, we find the supremum $\sup h_{kk}$ over sets of n vectors in $\mathbb{R}^p$, with $\bar{x}$ and S fixed. Using the crude estimate from the trace of H that $h_{kk} \leqslant p$, the h_{kk}s are bounded above and hence the supremum exists.

and the $p \times p$ *matrix*

$$S = \frac{1}{n-1}\sum_{i=1}^{n} x_i x_i'$$

is invertible.

Proof. Without loss of generality, we take $k = 1$. For convenience, instead of S, we use

$$\sigma = \frac{n-1}{n}S = \frac{1}{n}\sum_{i=1}^{n} x_i x_i'. \tag{2.1}$$

The problem is posed as a Lagrange optimization problem, with the Lagrangian function

$$L = x_1'\sigma^{-1}x_1 + \lambda'\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\left(x_i'\Omega x_i - \operatorname{trace}(\Omega\sigma)\right),$$

where λ and Ω are Lagrange multipliers and Ω is a $p \times p$ symmetric matrix. Taking derivatives with respect to x_i gives

$$(\sigma^{-1} + \Omega)x_1 = -\frac{1}{2}\lambda \tag{2.2}$$

and

$$\Omega x_i = -\frac{1}{2}\lambda \tag{2.3}$$

for $i = 2, \ldots, n$. The derivative with respect to λ returns the constraint $\bar{x} = 0$. Combining these equations gives

$$\Omega x_1 = \frac{(n-1)}{2}\lambda \tag{2.4}$$

and

$$\sigma^{-1}x_1 = -\frac{n}{2}\lambda. \tag{2.5}$$

Using the definition of σ, multiplied on the right by Ω, along with Eqs. (2.3), (2.4) and the constraint $\bar{x} = 0$, gives

$$\sigma\Omega = \frac{1}{2}x_1\lambda'. \tag{2.6}$$

It follows from this that, for $p \geqslant 2$, $\det(\sigma\Omega) = 0$ and consequently, Ω is not invertible.

Define the $n \times p$ matrix X with row vectors x_i', so that $n\sigma = X'X$ and $x_i'\sigma^{-1}x_i = nh_{ii}$, where h_{ij} is the generic entry of the $n \times n$ matrix

$$H = X(X'X)^{-1}X'.$$

H corresponds to the *hat matrix* in the language of statistical outliers. From Eq. (2.5), we obtain

$$h_{11} = \frac{1}{n}x_1'\sigma^{-1}x_1 = -\frac{1}{2}x_1'\lambda.$$

Multiplying Eq. (2.6) on the left by σ^{-1} and on the right by x_1 gives

$$\Omega x_1 = \frac{1}{2}\sigma^{-1}x_1(\lambda' x_1).$$

Eqs. (2.4) and (2.5) can be used to substitute for Ωx_1 and $\sigma^{-1}x_1$ in terms of λ and together yield

$$\frac{(n-1)}{2}\lambda = -\frac{n}{4}\lambda(\lambda' x_1).$$

If the trivial case of $\lambda = 0$ is excluded (which with Eq. (2.5) implies $x_1 = 0$), we get

$$\lambda' x_1 = -\frac{2(n-1)}{n}$$

and hence $h_{11} = (n-1)/n$ is the maximum possible value of a diagonal element of the matrix H. So

$$M = \frac{(n-1)^2}{n} \tag{2.7}$$

is the maximum possible Mahalanobis distance. □

Corollary 2.2. *The points $x_2, \dots, x_n$ must be coplanar in $\mathbb{R}^p$ for the maximum* (2.7) *to be achieved.*

Proof. From Eq. (2.6), it follows that

$$\sigma \Omega x_i = (x_i' \lambda)\frac{x_1}{2}$$

for $i = 2, \dots, n$. Substituting from Eqs. (2.3) and (2.5) then yields

$$\sigma\left(-\frac{1}{2}\lambda\right) = \left(-\frac{2}{n}x_i'\sigma^{-1}x_1\right)\frac{x_1}{2}.$$

Again using (2.5) on the left side and assuming $x_1 \neq 0$, we find that $x_i'\sigma^{-1}x_1 = -1$ and so $x_i' S^{-1} x_1 = -(n-1)/n$, for $i = 2, \dots, n$. This implies that the elements of the first column (and row) of H satisfy $h_{i1} = -1/n$, for $i = 2, \dots, n$. Furthermore, we deduce that since $S^{-1}x_1 \neq 0$, the points $x_2, \dots, x_n$ must be *coplanar* in $\mathbb{R}^p$ for the maximum h_{11} to be achieved. (In the case of $p = 1$ the remaining $(n-1)$ observations must be coincident and $h_{jj} = \frac{1}{n(n-1)}$ for $j > 1$.) □

Corollary 2.3. *The condition of Corollary* 2.2 *is not only necessary for achieving the maximum value* (2.7), *but is also sufficient.*

Proof. Assume that for $i = 2, \dots, n$, there exists a vector $v \neq 0$ and a scalar ρ such that

$$v' x_i = \rho, \tag{2.8}$$

i.e., the points $x_2, \dots, x_n$ lie in a common plane in $\mathbb{R}^p$. Using the assumption $\bar{x} = 0$, we also obtain $v' x_1 = -(n-1)\rho$. Then

$$(n-1)Sv = \sum_{n=1}^{n} x_i x_i' v = -(n-1)\rho x_1 + \sum_{i=2}^{n} \rho x_i = -n\rho x_1.$$

Since S is invertible, the right side is non-zero and in particular $\rho \neq 0$. Hence,

$$v = -\frac{n}{n-1}\rho S^{-1} x_1 \tag{2.9}$$

and

$$(n-1)\rho = -v'x_1 = \frac{n}{(n-1)}\rho x_1' S^{-1} x_1.$$

Dividing by ρ, it follows that

$$x_1' S^{-1} x_1 = \frac{(n-1)^2}{n} = M$$

as defined in Eq. (2.7), and $h_{11} = (n-1)/n$. From (2.9) we see that $S^{-1}x_1$ is a normal vector to the plane. Also, by combining Eqs. (2.8) and (2.9), we find, as before, that $x_i' S^{-1} x_1 = -(n-1)/n$ for $i = 2, \dots, n$. □

We now demonstrate that the maximum is achievable for any given n and p with $n > p+1$. In so doing, we address a more general question viz. what is the maximum number of the diagonal elements of the matrix H that can simultaneously achieve the maximum value $(n-1)/n$? From the property $\operatorname{trace}(H) = \sum_{i=1}^n h_{ii} = p$, it follows that, if $n > p+1$, then at most p of the h_{ii}s can be equal to the maximum value $(n-1)/n$. The example below illustrates that this is achievable.

Example 2.1. Let e_α be the standard basis of $\mathbb{R}^p$, i.e., $(e_\alpha)_\beta = \delta_{\alpha\beta}$. Define

$$x_i = \begin{cases} e_i, & i = 2, \dots, p, \\ e_1, & i = p+1, \dots, n. \end{cases}$$

Note that the points $x_2, \dots, x_n$ are all coplanar, lying in the plane $\xi_1 + \cdots + \xi_p = 1$. From the zero mean property x_1 is found to be

$$x_1 = -(n-p)e_1 - \sum_{j=2}^{p} e_j.$$

Then $S = (n-1)^{-1} \sum_{i=1}^n x_i x_i'$ gives

$$S = \frac{1}{(n-1)}\left[(n-p)(n-p+1)e_1e_1' + (n-p)\sum_{i=2}^{p}(e_1e_i' + e_ie_1') + \sum_{i=2}^{p}\sum_{j=2}^{p} e_ie_j' + \sum_{i=2}^{p} e_ie_i'\right].$$

From this, one may verify that the inverse is

$$S^{-1} = \frac{(n-1)}{n}\left[\frac{p}{n-p}e_1e_1' - \sum_{i=2}^{p}(e_1e_i' + e_ie_1') - \sum_{i=2}^{p}\sum_{j=2}^{p} e_ie_j' + n\sum_{i=2}^{p} e_ie_i'\right].$$

Using this in definition (1.3) of the generic entry h_{ij} of the matrix H, we obtain

$$H = \left(\begin{array}{cccc|ccc}
\frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\
-\frac{1}{n} & \frac{n-1}{n} & \cdots & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
-\frac{1}{n} & -\frac{1}{n} & \cdots & \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\
\hline
-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{p}{n(n-p)} & \cdots & \frac{p}{n(n-p)} \\
-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{p}{n(n-p)} & \cdots & \frac{p}{n(n-p)} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{p}{n(n-p)} & \cdots & \frac{p}{n(n-p)}
\end{array}\right),$$

where the top left block is $p \times p$. One can verify that H has all the desired properties listed above (symmetric, each row sums to zero, it is idempotent and $\text{trace}(H) = p$). Observe that $h_{11} = h_{22} = \cdots = h_{pp} = (n-1)/n$, the maximum possible number of maximum diagonal values. (Indeed if $n = p + 1$ then $h_{(p+1)(p+1)} = (n-1)/n$ also.) This proves that

$$\max\{(h_{11} + \cdots + h_{kk})\} = \frac{k(n-1)}{n}$$

for each $k = 1, \ldots, p$.

3. Lower bounds for the maximum Mahalanobis distance

Theorem 3.1. *A lower bound for the quantity* $M = \max_i\{x_i' S^{-1} x_i\}$, *is*

$$m_g = \frac{(n-1)p}{n}, \tag{3.1}$$

where $x_1, \ldots, x_n$ *are points in* $\mathbb{R}^p$ *that are subject to the constraint*

$$\bar{x} = \sum_{i=1}^{n} x_i = 0$$

and to the $p \times p$ *matrix* $S = (n-1)^{-1} \sum_{i=1}^{n} x_i x_i'$ *being invertible.*

We call m_g the *global lower bound* for M.

Proof. The sum of the Mahalanobis distances for the vectors $x_1, \ldots, x_n$ is

$$\sum_{i=1}^{n} x_i' S^{-1} x_i = \text{trace}((n-1)H) = (n-1)p, \tag{3.2}$$

where the $n \times n$ matrix H defined in Eq. (1.3). Since each term in the sum is non-negative, the maximum term $M \geqslant \frac{(n-1)p}{n}$. $\square$

Corollary 3.2. $M = m_g$ *if and only if*

$$x_i' S^{-1} x_i = \frac{(n-1)p}{n} \tag{3.3}$$

for all $i = 1, \ldots, n$.

This occurs if and only if $h_{ii} = p/n$ for all i. We emphasize that the infinimum of M will equal the global lower bound m_g, only for certain values of n and p. We explore this in some detail in the following examples.

Example 3.1. When $p = 1$, Eq. (3.3) implies that $M = m_g$, the global lower bound, if and only if

$$x_i^2 = \frac{(n-1)S}{n}$$

for all i. Taking this together with the mean constraint $\sum_{i=1}^{n} x_n = 0$, we reach the following two conclusions; when n is even, the global lower bound is achieved when exactly half of the x_is equal

$\sqrt{(n-1)S/n}$ and the other half equal $-\sqrt{(n-1)S/n}$ (and hence $\inf M = m_g = (n-1)/n$ for $p = 1$ and n even), as was derived in Hayes [4]; when n is odd, the global lower bound is *never achieved* as there is no combination of $\pm\sqrt{(n-1)S/n}$ with an odd number of terms that sums to zero. Note that this does not preclude the infinimum of M being equal to m_g, although we do not show this. However, it is straightforward to put an upper bound on $\inf M$ by taking the vector of x_is of the type $(a, -a, a, -a, \ldots, a, -a, 0)$ thus obtaining the bounds

$$\frac{n-1}{n} \leqslant \inf M \leqslant 1$$

for $p = 1$ and n odd.

Example 3.2. Next we examine the case of $p = 2$. Consider n (data) points uniformly distributed on the circle of radius a; we take $x_j = (a\cos\omega_j, a\sin\omega_j)$, where $\omega_j = 2\pi(j-1)/n$ for $j = 1, \ldots, n$. Then $\sum_{j=1}^{n} x_j = 0$, for $n > 2$ since, using complex exponential notation,

$$\sum_{j=1}^{n} \mathrm{e}^{\frac{2\pi\imath(j-1)}{n}} = 0.$$

Also

$$x_j x_j' = a^2 \begin{pmatrix} \cos^2\omega_j & \cos\omega_j \sin\omega_j \\ \cos\omega_j \sin\omega_j & \sin^2\omega_j \end{pmatrix} = \frac{a^2}{2} \begin{pmatrix} 1+\cos 2\omega_j & \sin 2\omega_j \\ \sin 2\omega_j & 1-\cos 2\omega_j \end{pmatrix}.$$

Now

$$\sum_{j=1}^{n} \mathrm{e}^{2\imath\omega_j} = \sum_{j=1}^{n} \mathrm{e}^{\frac{4\pi\imath(j-1)}{n}} = 0$$

for $n > 2$, giving

$$(n-1)S = \sum_{j=1}^{n} x_j x_j' = \frac{na^2}{2} I_2,$$

and so

$$S = \frac{na^2}{2(n-1)} I_2,$$

where I_2 is the 2×2 identity matrix. Hence,

$$x_j' S^{-1} x_j = \frac{2(n-1)}{na^2} \|x_j\|^2 = \frac{2(n-1)}{n}$$

for all j. This is m_g of Theorem 3.1 with $p = 2$. In conclusion, for $p = 2$ and all $n > 2$, $\inf M = m_g$ and the bound is achievable.

Example 3.3. Next we examine the general case, assuming that $n \geqslant 2p$. Define $R = n \bmod 2p$, so $n = 2dp + R$, where $d \geqslant 1$, and $0 \leqslant R \leqslant 2p - 1$. Define

$$x_i = \begin{cases} (-1)^{i+1} e_{\{\frac{i+1}{2} \bmod p\}} & \text{if } i \leqslant 2dp, \\ 0 & \text{if } 2dp < i \leqslant n, \end{cases}$$

where we define $e_0 = e_p$. Then $x_1 = e_1$, $x_2 = -e_1, \ldots, x_{2p-1} = e_p$, $x_{2p} = -e_p$, $x_{2p+1} = e_1$, $x_{2p+2} = -e_1, \ldots, x_{2dp-1} = e_p, x_{2dp} = -e_p$, $x_{2dp+1} = x_{2dp+2} = \cdots = x_n = 0$. Clearly, $\sum_{i=1}^{n} x_i = 0$ and also

$$(n-1)S = 2d\sum_{\alpha=1}^{p} e_\alpha e'_\alpha = 2dI_p,$$

so

$$\frac{1}{n-1}S^{-1} = \frac{1}{2d}I_p = \frac{1}{2\lfloor\frac{n}{2p}\rfloor}I_p,$$

where I_p is the $p \times p$ identity matrix. The diagonal elements of the H matrix are

$$h_{ii} = \begin{cases} \frac{1}{2\lfloor\frac{n}{2p}\rfloor} & \text{if } 1 \leqslant i \leqslant 2p\lfloor\frac{n}{2p}\rfloor, \\ 0 & \text{if } 2p\lfloor\frac{n}{2p}\rfloor < i \leqslant n. \end{cases}$$

These values satisfy the condition $\sum_{i=1}^{n} h_{ii} = p$, and it follows that

$$\max_i\{h_{ii}\} = \frac{1}{2d} = \frac{1}{2\lfloor\frac{n}{2p}\rfloor}.$$

So the maximum Mahalanobis distance for these data is

$$M_1 = \frac{n-1}{2\lfloor\frac{n}{2p}\rfloor}. \tag{3.4}$$

Clearly, $M_1 = m_g$, the global lower bound, when $2p$ divides evenly into n.

4. A dual approach to optimization problems

A dual approach to the above problems and specifically the infinimum of M (the minimax) problem is presented. This will also illustrate the fact that these problems are fundamentally algebraic and geometric in nature, though their origins are rooted in statistics.

We assume that S is invertible and positive definite. There then exists a rotation matrix $\widetilde{R}$ such that $(n-1)S = \widetilde{R}'\Delta\widetilde{R}$, where Δ is a diagonal matrix with diagonal elements $\delta_{\alpha\alpha} > 0$. Define $y_i = \widetilde{R}x_i$ for $i = 1, \ldots, n$. The zero-mean condition is preserved, i.e., $\sum_{i=1}^{n} y_i = 0$ and

$$\Delta = \sum_{i=1}^{n} y_i y'_i. \tag{4.1}$$

The elements of the H matrix, defined in (1.3) become $h_{ij} = y'_i\Delta^{-1}y_j$ and the various problems then consist of finding n vectors in $\mathbb{R}^p$ with certain desired optimization properties. We now transpose the problems by defining the vectors z_α ($\alpha = 1, \ldots, p$) in $\mathbb{R}^n$ by

$$(z_\alpha)_i = \frac{(y_i)_\alpha}{\sqrt{\delta_{\alpha\alpha}}}.$$

Then (4.1) becomes $z'_\alpha z_\beta = \delta_{\alpha\beta}$, that is $\{z_\alpha\}$ forms an orthonormal set in $\mathbb{R}^n$. The zero mean condition becomes $z'_\alpha u = 0$ where u is the vector of ones in $\mathbb{R}^n$. The H matrix elements are

$$h_{ij} = \sum_{\alpha=1}^{p}(f'_i z_\alpha)(f'_j z_\alpha),$$

where $\{f_i\}$ is the standard basis of $\mathbb{R}^n$ and the Mahalanobis distances are

$$D_i = \sqrt{(n-1)\sum_{\alpha=1}^{p}(f_i' z_\alpha)^2}.$$

As before, we define, for a given set of vectors $x_1, \ldots, x_n$ in $\mathbb{R}^p$, $M = \max_i D_i^2$ which in the transposed version is defined for a given set of vectors $z_1, \ldots, z_p$ in $\mathbb{R}^n$. We now define, for given n and p,

$$M_{n,p} = \sup M \tag{4.2}$$

and

$$m_{n,p} = \inf M, \tag{4.3}$$

where the supremum and infinimum are taken over sets of p orthonormal vectors z_α in $\mathbb{R}^n$, which are orthogonal to u. It follows that $M_{n,p} = \sup D_1^2 = (n-1)\sup h_{11}$. So the previous problems are transformed into optimization of the M over orthonormal sets of p vectors in the orthogonal space of $u = (1, 1, \ldots, 1)$ in $\mathbb{R}^n$. These problems depend only on n and p and not on the specific nature of any given matrix S (such as its eigenvalues). We have shown that

$$M_{n,p} = \frac{(n-1)^2}{n}.$$

We will not present a similar general formula for $m_{n,p}$ but rather present some bounds; specifically we have the global lower bound $m_{n,p} \geqslant m_g$ for all n, p and the upper bound M_1 in (3.4). We now use the dual approach to obtain another upper bound. Both of these bounds give rise to an infinite number of instances when $m_{n,p} = m_g$.

We assume here that $n \geqslant 2p$ and define

$$\mu = \lfloor \log_2 p \rfloor + 1. \tag{4.4}$$

Then $p < 2^\mu \leqslant n$. We write $n = \kappa 2^\mu + \phi$, where $\kappa = \left\lfloor \frac{n}{2^\mu} \right\rfloor$ with $\kappa \geqslant 1$ and $0 \leqslant \phi < 2^\mu$. We now construct a $2^\mu \times 2^\mu$ orthogonal matrix Q by defining the (a, b) element, for $a, b \in \{0, 1, \ldots, 2^\mu - 1\}$ as follows: we write the binary expansions of a and b, respectively, as $a = \sum_{l=1}^{\mu} 2^{l-1} a_l$ and $b = \sum_{l=1}^{\mu} 2^{l-1} b_l$ (with $a_k, b_k \in \{0, 1\}$) and then we define

$$Q_{ab} = 2^{-\frac{\mu}{2}}(-1)^{\sum_{k=1}^{\mu} a_k b_k}.$$

Note that the rows of $2^{\frac{\mu}{2}} Q$ consist of ± 1s; the first row is $Q_{0\star} = 2^{-\frac{\mu}{2}}(1, 1, \ldots, 1)$ and all other rows are orthogonal to this. We now choose the p vectors z_α in $\mathbb{R}^n$ as follows:

$$z_\alpha = \frac{1}{\sqrt{\kappa}}\,(\underbrace{Q_{\alpha\star}, Q_{\alpha\star}, \ldots, Q_{\alpha\star}}_{\kappa \text{ times}}, \underbrace{0, \ldots, 0}_{\phi \text{ times}})$$

for $\alpha = 1, \ldots, p$, where $Q_{\alpha\star}$ consists of the 2^μ elements of row α of the matrix Q defined above. These vectors form an orthonormal set in $\mathbb{R}^n$ and each one is orthogonal to $Q_{0\star}$ and hence to u, the vector of ones in $\mathbb{R}^n$. For each $\alpha = 1, \ldots, p$, observe that $(f_i' z_\alpha)^2 = \frac{1}{\kappa 2^\mu}$ for $i = 1, \ldots, \kappa 2^\mu$ and $(f_i' z_\alpha)^2 = 0$ for $i = \kappa 2^\mu + 1, \ldots, n$. So the corresponding Mahalanobis distances D_i satisfy

$$D_i^2 = \begin{cases} \frac{(n-1)p}{\kappa 2^\mu} & \text{if } 1 \leqslant i \leqslant \kappa 2^\mu, \\ 0 & \text{if } \kappa 2^\mu < i \leqslant n. \end{cases}$$

So the maximum Mahalanobis distance $M_2 = \max_i D_i^2$ for these data is given by

$$M_2 = \frac{(n-1)p}{\kappa 2^{\mu}} = \frac{(n-1)p}{2^{\mu}\left\lfloor \frac{n}{2^{\mu}} \right\rfloor} \tag{4.5}$$

with μ as given in (4.4). Note that this produces an infinite number of instances for which $M_2 = m_g$, namely if there exists ν such that $2^{\nu} \geqslant 2p$ and 2^{ν} divides evenly into n. Note also that the bounds M_1 in (3.4) and M_2 in (4.5) are independent, as illustrated by the following examples: when $n = 6$ and $p = 3$, $M_1 = \frac{5}{2}$ and $M_2 = \frac{15}{4}$, so $M_1 < M_2$; when $n = 8$ and $p = 3$, $M_1 = \frac{7}{2}$ and $M_2 = \frac{21}{8}$ so $M_1 > M_2$. In conclusion, if $n \geqslant 2p$, then

$$m_g \leqslant m_{n,p} \leqslant \min\left(\frac{n-1}{2\left\lfloor \frac{n}{2p} \right\rfloor}, \frac{(n-1)p}{2^{\mu}\left\lfloor \frac{n}{2^{\mu}} \right\rfloor}\right),$$

where μ is given in (4.4).

5. Optimization of the elements of the H matrix

We have already shown that for any given diagonal element h_{kk} of H,

$$0 \leqslant h_{kk} \leqslant \frac{n-1}{n}$$

and that both bounds are achievable (the lower bound being trivial).

It remains to address the optimization of the off-diagonal elements. Without loss of generality, we search for the optima of

$$h_{12} = \frac{1}{n-1} x_1 S^{-1} x_2.$$

We again introduce the variable $\sigma = \frac{n}{n-1} S$, so that $h_{12} = (x_1 \sigma^{-1} x_2)/n$. We treat the problem as a constrained optimization with the Lagrangian function

$$L = x_1' \sigma^{-1} x_2 + \lambda' \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} (x_i' \Omega x_i - \operatorname{trace}(\Omega\sigma)),$$

where again Ω is symmetric. Taking derivatives with respect to x_1, x_2 and x_i (for $i \geqslant 3$) gives

$$\begin{aligned} &\sigma^{-1} x_2 + 2\Omega x_1 + \lambda = 0, \\ &\sigma^{-1} x_1 + 2\Omega x_2 + \lambda = 0, \\ &2\Omega x_i + \lambda = 0, \quad i = 3, \ldots, n. \end{aligned} \tag{5.1}$$

Adding these equations gives

$$\sigma^{-1}(x_1 + x_2) = -n\lambda. \tag{5.2}$$

From the definition of σ, as in (2.1), multiplied on the right by Ω, and using the fact that $\bar{x} = 0$ along with Eqs. (5.1), we find that

$$-2n\sigma\Omega = x_1(\sigma^{-1}x_2)' + x_2(\sigma^{-1}x_1)'.$$

Hence,

$$-2\sigma\Omega x_i = h_{2i} x_1 + h_{1i} x_2 \tag{5.3}$$

for $i = 1, 2, 3, \dots, n$. Now multiplying Eqs. (5.1) on the left by $(-n\sigma)$, and eliminating λ using (5.2), gives

$$\begin{aligned}
&x_1 + (1-n)x_2 - 2n\sigma\Omega x_1 = 0,\\
&x_2 + (1-n)x_1 - 2n\sigma\Omega x_2 = 0,\\
&x_1 + x_2 - 2n\sigma\Omega x_i = 0, \quad i = 3, \dots, n.
\end{aligned}$$

Combining these with (5.3), we obtain

$$\begin{aligned}
&x_1(1+nh_{12}) + x_2(nh_{11} - (n-1)) = 0,\\
&x_1(nh_{22} - (n-1)) + x_2(1+nh_{12}) = 0,\\
&x_1(1+nh_{2i}) + x_2(1+nh_{1i}) = 0, \quad i = 3, \dots, n.
\end{aligned} \tag{5.4}$$

We assume that $x_1 \neq 0$ and $x_2 \neq 0$; otherwise $h_{12} = 0$. If x_1 and x_2 are linearly independent, then these equations give $h_{11} = \frac{n-1}{n} = h_{22}$, $h_{12} = -\frac{1}{n}$ and $h_{1i} = h_{2i} = -\frac{1}{n}$ for $i \geqslant 3$. This case does not correspond to an extremum of h_{12}. Otherwise, there exists γ such that $x_2 = \gamma x_1$. Then $h_{12} = \gamma h_{11}$, $h_{22} = \gamma^2 h_{11}$ and $h_{2i} = \gamma h_{1i}$. So, Eqs. (5.4) become

$$\begin{aligned}
&1 + n\gamma h_{11} + \gamma(nh_{11} - n + 1) = 0,\\
&n\gamma^2 h_{11} - (n-1) + \gamma(1 + n\gamma h_{11}) = 0,\\
&1 + n\gamma h_{1i} + \gamma(1 + nh_{1i}) = 0.
\end{aligned} \tag{5.5}$$

Taking γ times the first equation minus the second equation gives $\gamma^2 = 1$.

If $\gamma = 1$, solving the equations gives $h_{11} = \frac{n-2}{2n} = h_{12} = h_{22}$ and $h_{1i} = h_{2i} = -\frac{1}{n}$ for $i \geqslant 3$; if $\gamma = -1$, then $h_{11} = \frac{1}{2} = h_{22}$, $h_{12} = -\frac{1}{2}$ and $h_{1i} = h_{2i} = 0$ for $i \geqslant 3$. So the maximum value of h_{12} is $\frac{1}{2} - \frac{1}{n}$ and the minimum is $-\frac{1}{2}$. In conclusion we find that for all $n > 2$,

$$-\frac{1}{2} \leqslant h_{12} \leqslant \frac{1}{2} - \frac{1}{n}.$$

We will now demonstrate how these extreme values of h_{12} are achieved by presenting actual data that reproduce them. It is illustrative to consider first the case of one dimension, $p = 1$. In that instance, the $\gamma = 1$ case corresponds to n points: $(n-2)r, (n-2)r, -2r, -2r, \dots, -2r$ where $r^2 = \frac{\sigma}{2n-4}$; the $\gamma = -1$ case corresponds to n points: $r, -r, 0, 0, \dots, 0$ where $r^2 = \frac{n\sigma}{2}$.

We now consider the higher dimensional cases, where $p \geqslant 2$. Observe that for σ to be invertible, i.e., of full rank, it is necessary that $n \geqslant p + 2$, since we now have the additional constraint on the x_is that $x_2 = \pm x_1$.

Example 5.1. We present an example corresponding to $\gamma = 1$, in which one can observe that $h_{12} = \frac{1}{2} - \frac{1}{n}$. Let e_α be the standard basis of $\mathbb{R}^p$ and define

$$x_i = \begin{cases} -\frac{1}{2}\left((n-p-1)e_p + \sum_{j=1}^{p-1} e_j\right), & i = 1, 2,\\ e_{i-2}, & i = 3, \dots, p+1,\\ e_p, & i = p+2, \dots, n. \end{cases}$$

Then, Eq. (2.1) gives

$$\sigma = \frac{1}{2n}\left[2\sum_{i=1}^{p-1} e_i e_i' + \sum_{i=1}^{p-1}\sum_{j=1}^{p-1} e_i e_j' + (n-p-1)\sum_{i=1}^{p-1}(e_p e_i' + e_i e_p') + (n-p-1)(n-p+1)e_p e_p'\right].$$

From this, one may verify that the inverse is

$$\sigma^{-1} = n\sum_{i=1}^{p-1} e_i e_i' - \sum_{i=1}^{p-1}\sum_{j=1}^{p-1} e_i e_j' - \sum_{i=1}^{p-1}(e_p e_i' + e_i e_p') + \frac{p+1}{n-p-1} e_p e_p'.$$

Using this, we then calculate the matrix

$$H = \left(\begin{array}{cc|cccc|ccc}
\frac{n-2}{2n} & \frac{n-2}{2n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\
\frac{n-2}{2n} & \frac{n-2}{2n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\
\hline
\frac{1}{n} & -\frac{1}{n} & \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\
-\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \frac{n-1}{n} & \cdots & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\
-\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\
\hline
\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{p+1}{n(n-p-1)} & \cdots & \frac{p+1}{n(n-p-1)} \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\
-\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{p+1}{n(n-p-1)} & \cdots & \frac{p+1}{n(n-p-1)}
\end{array}\right),$$

where the square diagonal blocks have sides respectively of size 2, $p-1$ and $n-p-1$.

Example 5.2. We present an example corresponding to $\gamma = -1$, in which one can observe that $h_{12} = -\frac{1}{2}$. Define

$$x_i = \begin{cases} (-1)^{i-1} e_1, & i = 1, 2, \\ e_{i-1}, & i = 3, \ldots, p, \\ -\sum_{j=2}^{p-1} e_j - (n-p-1)e_p, & i = p+1, \\ e_p, & i = p+2, \ldots, n, \end{cases}$$

where again we take $n \geqslant p+2$ and also assume $p \geqslant 3$. The case $p = 2$ is obtained in a similar analysis by omitting the second line of this definition and the sum term from the third line. Then, Eq. (2.1) gives

$$\sigma = \frac{1}{n}\left[2e_1e_1' + \sum_{i=2}^{p-1} e_i e_i' + \sum_{i=2}^{p-1}\sum_{j=2}^{p-1} e_i e_j' + (n-p-1)\sum_{i=2}^{p-1}(e_p e_i' + e_i e_p') + (n-p)(n-p-1)e_p e_p'\right].$$

From this, one may verify that the inverse is

$$\sigma^{-1} = \frac{n}{2}e_1e_1' + n\sum_{i=2}^{p-1} e_i e_i' - \frac{n}{n-2}\sum_{i=2}^{p-1}\sum_{j=2}^{p-1} e_i e_j' - \frac{n}{n-2}\sum_{i=2}^{p-1}(e_p e_i' + e_i e_p') + \frac{n(p-1)}{(n-2)(n-p-1)} e_p e_p'.$$

Using this, we then calculate the matrix

$$H=\left(\begin{array}{cc|cccc|ccc}
\frac{1}{2} & -\frac{1}{2} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0\\
-\frac{1}{2} & \frac{1}{2} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0\\
\hline
0 & 0 & \frac{n-3}{n-2} & -\frac{1}{n-2} & \cdots & -\frac{1}{n-2} & -\frac{1}{n-2} & \cdots & -\frac{1}{n-2}\\
0 & 0 & -\frac{1}{n-2} & \frac{n-3}{n-2} & \cdots & -\frac{1}{n-2} & -\frac{1}{n-2} & \cdots & -\frac{1}{n-2}\\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots\\
0 & 0 & -\frac{1}{n-2} & -\frac{1}{n-2} & \cdots & \frac{n-3}{n-2} & -\frac{1}{n-2} & \cdots & -\frac{1}{n-2}\\
\hline
0 & 0 & -\frac{1}{n-2} & -\frac{1}{n-2} & \cdots & -\frac{1}{n-2} & \frac{p-1}{(n-2)(n-p-1)} & \cdots & \frac{p-1}{(n-2)(n-p-1)}\\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots\\
0 & 0 & -\frac{1}{n-2} & -\frac{1}{n-2} & \cdots & -\frac{1}{n-2} & \frac{p-1}{(n-2)(n-p-1)} & \cdots & \frac{p-1}{(n-2)(n-p-1)}
\end{array}\right),$$

where the square diagonal blocks have sides respectively of size 2, $p-1$ and $n-p-1$.

Acknowledgments

E. Gath would like to thank the Department of Finance, McCombs School, University of Texas, Austin at which much of this research was conducted. He would also like to acknowledge useful correspondence with Prof. Finbarr Holland, University College Cork, Ireland concerning the $p=1$ minimax problem.

References

[1] B.C. Arnold, R.A. Groeneveld, Bounds for deviations between sample and population statistics, Biometrika 61 (1974) 387–389.
[2] V. Barnett, T. Lewis, Outliers in Statistical Data, third ed., John Wiley and Sons, Chichester, 1994.
[3] J.B. Gray, W.H. Woodall, The maximum size of standardized and internally studentized residuals in regression analysis, The Amer. Statist. 48 (1994) 111–113.
[4] K. Hayes, A lower bound for the most deviant Z-score, Teach. Statist. 26 (3) (2004) 89–91.
[5] D.M. Hawkins, Identification of Outliers, Chapman and Hall, London, 1980.
[6] I. Olkin, A matrix formulation on how deviant an observation can be, The Amer. Statist. 46 (1992) 205–209.
[7] W.R. Thompson, On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation, The Ann. Math. Statist. 6 (1935) 214–219.
[8] P.A. Samuelson, How deviant can you be? J. Amer. Statist. Assoc. 63 (1968) 1522–1525.
[9] R.E. Shiffler, Maximum Z-scores and outliers, The Amer. Statist. 42 (1988) 79–80.
[10] R.E. Shiffler, P.D. Harsha, Upper and lower bounds for the sample standard deviation, Teach. Statist. 2 (1980) 84–86.