The Complexity of Computing the MCD-Estimator*

Thorsten Bernholt Paul Fischer
Lehrstuhl Informatik 2
Universität Dortmund, Germany
bernholt/paulf@ls2.cs.uni-dortmund.de

November 6, 2001

Abstract

In modern statistics the robust estimation of parameters is a central problem, i. e., an estimation that is not or only slightly affected by outliers in the data. The Minimum Covariance Determinant estimator (MCD) [8] is probably one of the most important robust estimators of location and scatter. It's complexity of computing the MCD, however, was unknown and generally thought to be exponential even if the dimensionality of the data is fixed.

Here we present a polynomial time algorithm for MCD for fixed dimension of the data. In contrast we show that computing the MCD-estimator is NP-hard if the dimension varies.

Keywords: Computational statistics, efficient algorithms, NP-completeness, combinatorial geometry

1 Introduction

In modern mathematical statistics and data analysis one fundamental problem is that of constructing statistical methods which are *robust* against model deviations. For example, it is well known that the standard estimates of location and scatter – sample mean and sample variance – are not robust. A single data point which is moved far out will change these quantities arbitrarily. In general one assumes that the observed data is mainly generated by some process or distribution which one would like to analyse. We shall call the part of the data coming from the distribution of interest the data from the *true population.*

*The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

The rest of the data, however, might come from other sources or is altered by noise, we call this the *outliers*. The goal is to nevertheless estimate statistical quantities of the true population. This is clearly impossible if the majority of the data consists of outliers, thus we shall assume that the majority of the data comes from the true population.

One possible approach to tackle the problem of robust estimation is to find a sufficiently large subset of the data mainly consisting of elements of the true population and to base the estimation on this subset. Several authors follow this approach, e. g. [3, 8]. One of the most popular methods in this context is to select a subset with a minimum value of the covariance determinant (MCD estimator, [2, 4, 8]). Heuristic search algorithms for the MCD can be found in [5, 6, 9, 10, 11]. A comparison of the MCD, MVE and S-estimator is presented in [1].

More precisely, given N observations, a subset of size h, for some $h > N/2$, is selected for which the determinant of the empirical covariance matrix is minimal over all subsets of size h. We shall now formally define MCD and then discuss some of its properties.

Let $\mathcal{X} = \mathbf{x}_1, \ldots, \mathbf{x}_h$ be a set[1] of points in $\mathbb{R}^d$ for some constant d. Let $\mathbf{x_i} = (x_{i1}, \ldots, x_{id})^T$. The *(empirical) covariance matrix* $C = C(\mathcal{X}) = (c_{ab})_{1\leq a,b\leq d}$ of $\mathcal{X}$ is the $(d \times d)$-matrix defined by

$$c_{ab} = \frac{1}{h}\sum_{i=1}^{h}(x_{ia} - t_a)\cdot(x_{ib} - t_b) \text{ where } \quad t_j = \frac{1}{h}\sum_{i=1}^{h} x_{ij} \ ,$$

or in matrix notation

$$C(\mathcal{X}) = \frac{1}{h}\sum_{i=1}^{h}\mathbf{x_i}\,\mathbf{x_i}^T - \mathbf{t}\,\mathbf{t}^T \ .$$

The covariance matrix is positive definite. For a $d\times d$-matrix M, let $\det(M)$ denote its determinant. For the determinant of a covariance matrix C we write $\det(\mathcal{X}) = \det(C(\mathcal{X}))$ and we shall call it *covariance determinant*. Let us now define the problem:

Definition 1.1 (MCD) *Let $d < N/2$. Let $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be a set of N points in $\mathbb{R}^d$. Let h be a natural number, $N/2 < h < N$. The* minimum covariance determinant problem *for $\mathcal{X}$ and h, MCD for short, is the problem to find an h-element set $\mathcal{X}' = \{\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_h}\} \subset \mathcal{X}$ such that* $\det(\mathcal{X}')$ *is minimal, over all h-element sets.*
For the decision version *of MCD, MCDd, a positive real number B is given in addition. The problem is to decide whether there exists an h-element set $\mathcal{X}' = \{\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_h}\} \subset \mathcal{X}$ such that* $\det(\mathcal{X}') \leq B$.

[1] Strictly speaking we are considering multisets here, i. e., we allow multiple occurrences of the same element. We shall nevertheless use the term set as is the practice in statistics. One may as well think of weighted points, where the weight indicates the multiplicity.

Another robust estimator of location and shape is the Minimum Volume Ellipoid and our results can be easily adapted to this estimator:

Definition 1.2 (MVE) *The Minimum Volume Ellipsoid is the problem to find a subset of size h, for that the enclosing ellipsoid has the minimal volume.*

The empirical covariance matrix $C(\mathcal{X}')$ with minimal determinant yields a robust estimator S of scatter, $S = S(\mathcal{X}') = c_0 \cdot C(\mathcal{X}')$, where c_0 is a suitably chosen constant to achieve consistency. As an estimator for the location one uses the mean (or center of gravity)

$$\mathbf{t} = \mathbf{t}(\mathcal{X}') = \frac{1}{h} \sum_{\mathbf{x} \in \mathcal{X}'} \mathbf{x}$$

of the h points in the set $\mathcal{X}'$. The pair $(\mathbf{t}, S)$ is called *MCD-estimator* with respect to $\mathcal{X}$.

There is a nice geometric interpretation of the MCD. The inverse $C^{-1}(\mathcal{X}')$ of the minimum covariance matrix $C(\mathcal{X}')$ and the mean $\mathbf{t}(\mathcal{X}')$ define an ellipsoid in $\mathbb{R}^d$, for details see Section 2. This ellipsoid nicely matches the points $\mathcal{X}'$, see Figure 1 for an example in two dimensions. The determinant is a measure of volume. Hence a small determinant corresponds to an ellipsoid of small volume. If the extensions of the ellipsoid in all dimensions are small then the set $\mathcal{X}'$ is quite compact. Another way to get a small volume is that the ellipsoid is somewhat "flat", i. e., it might have a large extension in some directions but only small ones in others. This indicates that the set $\mathcal{X}'$ is "essentially lower dimensional".

In this paper we address the complexity of computing the MCD-estimator. Obviously, computing $\det(\mathcal{X}')$ for all $\binom{N}{h}$ subsets $\mathcal{X}'$ of $\mathcal{X}$ of size h solves the problem, though it might take exponential time in h. It was not clear whether the estimator itself has this complexity independent of the dimensionality d of the data. Here we show that the complexity of MCD is polynomial if the dimension is fixed. This is achieved by avoiding to consider all subsets of size h. Exploiting geometric properties of the estimator, we have been able to design an algorithm which enumerates a sequence of subsets of size h of the input data set $\mathcal{X}$ in polynomial time. We show that one of the sets enumerated has minimum covariance determinant. The running time of our algorithm is $O\left(N^{d^2}\right)$.

On the other hand it is possible to show that the decision version of the MCD problem is NP-complete if the dimension varies. This is achieved by reducing CLIQUE to MCDd. The reduction combines combinatorial and algebraic methods in a clever way and is of its own interest. The main problem in constructing a reduction is that one cannot control the entries of the covariance matrix directly but only through the data points. Moreover, changing a data point might alter *all* entries of the covariance matrix. The continuous nature of the MCD-estimator introduces further difficulties.

The next section states some properties of the covariance determinant and the related ellipsoid which will be helpful in proving our results.

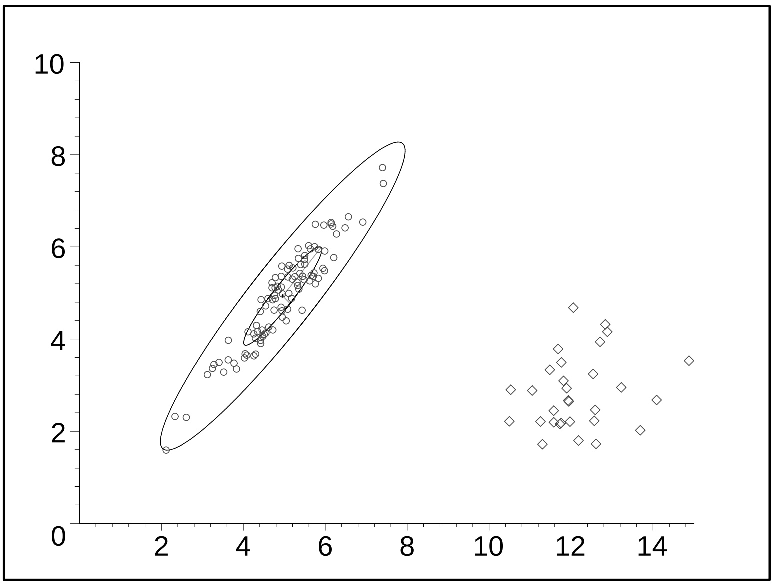

Figure 1: The figure shows the ellipsoid which corresponds to the covariance matrix with minimum determinant. The ellipsoid is plotted for two different radii. The points at right are outside the ellipsoid even for a very large radius.

2 MCD and Ellipses

Fix d and let $v = d(d+3)/2$. A *quadric* Q in $\mathbb{R}^d$ is a $(d-1)$-dimensional manifold determined by a second order expression which depends on $v+1$ real parameters $a_0, a_1, \ldots, a_d$ and a_{ij} for $1 \le i \le j \le d$. Every point $\mathbf{z} = (z_1, \ldots, z_d)^T \in Q$ satisfies the condition

$$a_0 + a_1 z_1 + \cdots + a_d z_d + a_{11} z_1^2 + 2a_{12} z_1 z_2 + \cdots + 2a_{d,d-1} z_d z_{d-1} + a_{dd} z_d^2 = 0 \ . \quad (1)$$

Note that there are only v degrees of freedom because equation (1) can be multiplied by any non-zero constant without changing the quadric. Equation (1) can be rewritten in matrix form as follows. Let the symmetric matrix A and the vector $\mathbf{b}$ be defined by

$$A := \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{12} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1d} & a_{2d} & \cdots & a_{dd} \end{pmatrix} \ , \quad \mathbf{b} := \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix} \ .$$

Then Equation (1) is equivalent to

$$\mathbf{z}^T A \mathbf{z} + \mathbf{z}^T \mathbf{b} + a_0 = 0 \ . \quad (2)$$

We say that quadric Q *selects* a subset $\mathcal{X}' \subset \mathcal{X}$ if

$$\mathbf{x}^T A \mathbf{x} + \mathbf{x}^T \mathbf{b} + a_0 \begin{cases} \le 0 & \forall \mathbf{x} \in \mathcal{X}' \\ > 0 & \forall \mathbf{x} \in \mathcal{X} \setminus \mathcal{X}' \ . \end{cases}$$

If the quadric surface is the surface of an ellipsoid then Equation (2) can be rewritten as

$$(\mathbf{z}-\mathbf{t})^T M (\mathbf{z}-\mathbf{t}) = r^2 \ , \tag{3}$$

where M is a positive definite $(d \times d)$-matrix, $\mathbf{t} \in \mathbb{R}^d$ is the *center point* and $r \in \mathbb{R}$ is the radius. Given M, $\mathbf{t}$ and r we denote by $E(M, \mathbf{t}, r)$ the solid ellipsoid defined by Formula (3) with equality replaced by "less than or equal".

Selection by quadrics in d dimensions is equivalent to linear separation in v dimensions. To this end consider the mapping $\widehat{\ }: \mathbb{R}^d \mapsto \mathbb{R}^v$ defined by

$$\widehat{\mathbf{z}} = (z_1, \ldots, z_d, z_1 z_1, \ldots, z_i z_j, \ldots, z_d z_d)^T \ , \quad 1 \leq i \leq j \leq d \ .$$

For a set $\mathcal{Z} \subset \mathbb{R}^d$ let $\hat{\mathcal{Z}} := \{\widehat{\mathbf{z}} \mid \mathbf{z} \in \mathcal{Z}\}$. Now the parameters a_i, a_{ij} in (1) define a hyperplane in $\mathbb{R}^v$ which separates the points of $\hat{\mathcal{X}}'$ from those in $\hat{\mathcal{X}} \setminus \hat{\mathcal{X}}'$.

As mentioned in Section 1, the covariance matrix $C(\mathcal{X}')$ of a point set $\mathcal{X}'$ is positive definite. Its inverse C^{-1} is also positive definite and the ellipsoid $E(C^{-1}, \mathbf{t}(\mathcal{X}'), r)$ is an ellipsoid which "fits" the point set $\mathcal{X}'$ for a suitably chosen radius r.

Throughout this paper we assume that the points of $\mathcal{X}$ are in *general quadric position*, i. e., no hyperplane in $\mathbb{R}^v$ contains more than $v+1$ points of $\hat{\mathcal{X}}$.

The following result of Rousseeuw [9] shows that the fit is even better for the set defining the minimum covariance determinant.

Lemma 2.1 *Let $d < h < N$ and let $\mathcal{X} \subset \mathbb{R}^d$ be a set of N points. Let $\mathcal{X}_{opt} \subseteq \mathcal{X}$, $|\mathcal{X}_{opt}| = h$ be such that $\det(\mathcal{X}_{opt})$ is minimal for all subsets of $\mathcal{X}$ of cardinality h. Let $C_{opt} = C(\mathcal{X}_{opt})$ be the corresponding covariance matrix and $\mathbf{t}_{opt} = \mathbf{t}(\mathcal{X}_{opt})$ be the center of gravity. Then there exists a radius $r > 0$ such that*

$$\mathcal{X}_{opt} = \mathcal{X} \cap E(C_{opt}^{-1}, \mathbf{t}_{opt}, r) \ ,$$

that is, $E(C_{opt}^{-1}, \mathbf{t}_{opt}, r)$ selects $\mathcal{X}_{opt}$.

Given a set $\mathcal{S} \subseteq \mathcal{X}$, $|\mathcal{S}| = v$ then (by our assumption on the position of the points) there is a unique quadric $Q(\mathcal{S})$ through the points of $\mathcal{S}$; we call it the quadric *defined by* $\mathcal{S}$. It can be computed by writing an equation of the form (1) for every point $\mathbf{z} \in \mathcal{S}$ and solving the resulting system of linear equations for $a_0, a_1, \ldots, a_{dd}$. As mentioned above the value of one parameter, e. g., a_0, can be chosen arbitrarily.

The next lemma shows that a set of points selectable by an ellipsoid is (almost) selectable by a quadric defined by a set of v points.

Lemma 2.2 *Given $\mathcal{X} \subset \mathbb{R}^d$ in general quadric position and an ellipsoid E let $\mathcal{X}' = E \cap \mathcal{X}$. Then there exists a set $\mathcal{S} \subseteq \mathcal{X}$, $|\mathcal{S}| = v := d(d+3)/2$ such* *that for the quadric $Q(\mathcal{S})$ the following holds: Let A, $\mathbf{b}$ and a_0 define $Q(\mathcal{S})$ as in (2) then*

$$\begin{array}{ll}
\mathbf{x}^T A\mathbf{x} + \mathbf{x}^T\mathbf{b} + a_0 \le 0 & \mathbf{x} \in \mathcal{X}' \\
\mathbf{x}^T A\mathbf{x} + \mathbf{x}^T\mathbf{b} + a_0 \ge 0 & \mathbf{x} \in \mathcal{X} \setminus \mathcal{X}' \\
\mathbf{x}^T A\mathbf{x} + \mathbf{x}^T\mathbf{b} + a_0 = 0 & \textit{for at most } v \textit{ points } \mathbf{x} \in \mathcal{X} \setminus \mathcal{X}'
\end{array}$$

Proof. Let an ellipsoid $E = E(M, \mathbf{t}, r)$ be given which selects a set $\mathcal{X}' \subseteq \mathcal{X}$, i. e.,

$$(\mathbf{x} - \mathbf{t})^T M (\mathbf{x} - \mathbf{t}) \begin{cases} \le r^2 & \text{if } \mathbf{x} \in \mathcal{X}' \ , \\ > r^2 & \text{if } \mathbf{x} \in \mathcal{X} \setminus \mathcal{X} \ . \end{cases}$$

Expanding the matrix equation into the form (1) with $a_0 = -r^2$ we arrive at

$$a_1x_1 + \cdots + a_dx_d + a_{11}x_1^2 + 2a_{12}x_1x_2 + \cdots + a_{dd}x_d^2 \begin{cases} \le r^2 \text{ if } \mathbf{x} \in \mathcal{X}' \ , \\ > r^2 \text{ if } \mathbf{x} \in \mathcal{X} \setminus \mathcal{X}' \ . \end{cases} \tag{4}$$

The hyperplane

$$a_1x_1 + \cdots + a_dx_d + a_{11}x_1^2 + 2a_{12}x_1x_2 + \cdots + a_{dd}x_d^2 = r^2$$

separates the points of $\hat{\mathcal{X}}'$ and $\hat{\mathcal{X}} \setminus \hat{\mathcal{X}}'$ in $\mathbb{R}^v$. This hyperplane is now moved in such a way that it contains v points of $\hat{\mathcal{X}}$ but no point has passed through it. By our assumption made on the position of the points it follows that the hyperplane contains exactly v points. This means that the inequality or strict inequality (4) becomes an equality for the points $\mathbf{x}$ on the hyperplane. Let a_i' and a_{ij}' denote the parameters of the resulting hyperplane. Clearly a_i' and a_{ij}' define a quadric in $\mathbb{R}^d$ but not necessarily an ellipsoid. Altogether there are at most v new points $\mathbf{x} \in \mathcal{X} \setminus \mathcal{X}'$ such that

$$a_1'x_1 + \cdots + a_d'x_d + a_{11}'x_1^2 + 2a_{12}'x_1x_2 + \cdots + 2a_{d,d-1}'x_dx_{d-1} + a_{dd}'x_d^2 = Fa_0' \ .$$

$\square$

3 An Efficient Algorithm for Fixed Dimension

Using the results from the previous section, we show how to list all subsets selectable by ellipsoids in polynomial time. Actually we shall list a polynomial collection of sets which contains all those selectable by ellipsoids. There are in general infinitely many ellipsoids which select the same subset. Given $\mathcal{X}' \subseteq \mathcal{X}$ let $\mathcal{E}(\mathcal{X}')$ denote the set of all ellipsoids selecting $\mathcal{X}'$. Next we show how to select a representative from every $\mathcal{E}(\mathcal{X}')$, $\mathcal{X}' \subseteq \mathcal{X}$.

Let $E \in \mathcal{E}(\mathcal{X}')$. According to Lemma 2.2, there is hyperplane $H(E)$ in $\mathbb{R}^v$ such that for all points of $\mathcal{X}'$ the inequality for $H(E)$ is satisfied with "less or equal"and there are at most v points of $\mathcal{X} \setminus \mathcal{X}'$ satisfying it with equality.

The algorithm loops through all subsets $\mathcal{S}$ of $\mathcal{X}$ of cardinality v. For every $\mathcal{S}$ it computes the hyperplane $H(\mathcal{S})$ defined by $\mathcal{S}$, which is possible in time $O(v^3)$. We then compute the set

$$\mathcal{T} = \{\mathbf{x} \in \mathcal{X} \mid \mathbf{x} \text{ satisfies (4) with "strictly less"}\} \quad .$$

The time to compute $\mathcal{T}$ is $O(n)$. Finally, for every $\mathcal{S}' \subseteq \mathcal{S}$ let $\mathcal{T}' = \mathcal{S}' \cup \mathcal{T}$.

The sets $\mathcal{T}'$ enumerated as above contain all subsets of $\mathcal{X}$ selectable by ellipsoids and possibly other sets. We are only interested in sets of size h. Let $\mathcal{T}'_1, \ldots, \mathcal{T}'_k$ be the sequence sets constructed as above with $|\mathcal{T}_i| = h$. For each $\mathcal{T}'_i$ the covariance determinant is computed and the overall minimum is selected. By Lemma 2.1, any set defining an optimal covariance determinant is selectable by an ellipsoid, hence, an optimal set appears in the sequence.

The number of enumerated sets $\mathcal{T}'$ is at most $\binom{N}{v}2^v = O(N^v)$, where $v = d(d+3)/2$. For every set time $O(N)$ is spent. The following theorem summarizes this result.

Theorem 3.1 *For fixed dimension d the MCD-problem can be solved in polynomial time $O(N^{v+1})$ where $v = d(d+3)/2$.*

4 The Hardness Result

In this section we show that the decision version MCDd of the Minimum Covariance Determinant problem is NP-complete if the dimension varies. To indicate this we shall use n to denote the dimension in this section.

Definition 4.1 (Mahalanobis distance) *Let C be a positive definite $n \times n$-matrix and let $\mathbf{t}$ be a n-vector. Then the* Mahalanobis distance $\mathrm{md}(\mathbf{x}; C, \mathbf{t})$ *of vector $\mathbf{x}$ w. r. t. C and $\mathbf{t}$ is defined by*

$$\mathrm{md}(\mathbf{x}; C, \mathbf{t}) := (\mathbf{x} - \mathbf{t})^T \; C^{-1} \; (\mathbf{x} - \mathbf{t}) \quad .$$

The following Lemma is a particular case of the Theorem 1 from [9]:

Lemma 4.2 (Exchange Lemma) *Let $\mathbf{x}$ and $\mathbf{y}$ be points from $\mathbb{R}^n$ and let $\mathcal{X}$ be a set of points from $\mathbb{R}^n$ not containing $\mathbf{x}$ or $\mathbf{y}$. Let $C = C(\mathcal{X})$ be the covariance matrix of $\mathcal{X}$ and $\mathbf{t} = \mathbf{t}(\mathcal{X})$ be the center of gravity of the points in $\mathcal{X}$. If*

$$\mathrm{md}(\mathbf{x}; C, \mathbf{t}) > \mathrm{md}(\mathbf{y}; C, \mathbf{t})$$

then

$$\det(\mathcal{X} \cup \{\mathbf{x}\}) > \det(\mathcal{X} \cup \{\mathbf{y}\}) \quad .$$

Intuitively, exchanging a distant point with a closer one decreases the determinant. We show that the decision version of MCD is NP-complete by reducing the maximum clique problem $n/2$-CLIQUE to it. For the sake of completeness let us repeat the definition of the latter problem.

Definition 4.3 ($n/2$-CLIQUE) *Let a graph G with n vertices and m edges be given. The problem is to decide whether G contains a complete subgraph on $\lceil n/2 \rceil$ vertices, i. e., a subset of $\lceil n/2 \rceil$ vertices in which all edges are present.*

The parameters used in the following theorem and in the rest of the paper are summarized in Table 1 of the appendix. Now we are ready to state the main theorem:

Theorem 4.4 *MCDd is NP-complete.*

Proof. Let $G = (V, E)$ be a graph with vertex set V, $|V| = n$ and edge set E. For a reduction we construct an input for the MCD-estimator by mapping the vertices and edges of G into points in $\mathbb{R}^n$ and by choosing the appropriate constant $B = \frac{1}{h^n} \left(k + 2wz^2\right)^k \cdot \left(2wz^2\right)^{(n-k)}$. The dimension of the resulting input is the number of vertices of the graph. Let v_i, $i = 1, \ldots, n$, be the vertices and let e_{ij} denote the edge between v_i and v_j. Let $\mathbf{e}_i$ denote the i-th unit vector in $\mathbb{R}^n$. Vectors and points in $\mathbb{R}^n$ are identified as usual.

We use k to denote $\lceil n/2 \rceil$ in the following. Three types of points in $\mathbb{R}^n$ are used in the reduction: Vertex points (v-points), edge-points (e-points), and auxiliary points (a-points). Let $\mathcal{X}$ consist of the following points:

- For every vertex v_i add the point $\mathbf{e}_i$ on the i-th coordinate axis.
- For every edge e_{ij} add the point $\mathbf{e}_i + \mathbf{e}_j$ on the diagonal of the 2-dimensional subspace in dimensions i and j.
- For every $i \in \{1, \ldots, n\}$ add $k^4/2$ times the point $k^{-2k} \cdot \mathbf{e_i}$ and $k^4/2$ times the point $-(k^{-2k}) \cdot \mathbf{e_i}$. These points are on the i-th coordinate axis very close to the origin.

Altogether $\mathcal{X}$ contains $N := n+m+nk^4$ points. MCD selects a subset $\mathcal{X}' \subset \mathcal{X}$ of cardinality $h < N$ such that $\det(\mathcal{X}')$ is minimal. We set $h := k + \binom{k}{2} + nk^4$ which is the number of a-points plus the number of edges and vertices of a k-clique.

The a-points serve two purposes: By choice of h, at least $k^4 - \binom{k}{2} - k$ copies of every a-point have to be selected. This ensures that the covariance determinant for $k \geq 2$ is not zero, because the a-points span $\mathbb{R}^n$. Second their large number ensures that the center of any covariance ellipsoid defined by h points is very close to the origin.

The a-points are close to the origin and do not contribute much to the covariance determinant resp. to the volume of the associated ellipsoid. We shall show that one has to select all of them for a minimum ellipsoid. Still $\binom{k}{2} + k$ points are missing and we have to select them from the v- and e-points. These points are far away from the origin (and the center of any ellipsoid defined by h points), hence they contribute much to the covariance determinant. In order to keep the determinant small the vectors they represent should span a low-dimensional space.

If G contains a k-clique then the set $\mathcal{X}'$ can be completed by adding the points corresponding to the edges and vertices of a clique. We shall call this a *clique configuration*. The vectors of the v- and e-points of $\mathcal{X}'$ span a space of only k dimensions, which bounds their influence on the covariance determinant. Altogether the covariance determinant $\det(\mathcal{X}')$ is small. In Section 4.1 we will show that $\det(\mathcal{X}')$ is not larger than B.

If G does not contain a k-clique, we will be forced to add v- and e-points to $\mathcal{X}'$ such that the corresponding vectors span at least $k+1$ dimensions. This results in a much larger value of $\det(\mathcal{X}')$. All such configurations are called *non-clique configurations*.

In order to lower bound the determinant we will construct in Section 4.3 an arrangement of h points which cannot be realized by the reduction. It consists of all a-points, exactly $k+1$ v-points and $h-(k+1)-nk^4$ copies of the origin. We call this a *minimal* $(k+1)$*-configuration*. We will show in Section 4.4 that this minimal $(k+1)$-configuration has a smaller determinant than a non-clique configuration, but is still greater than B.

To show that the reduction works in polynomial time, we look at the number of points and the bit-length of the numbers. The number of points is bounded by $O(k^5)$. The numbers itself are described by rational numbers. B is the largest number, and nominator and denominator are bounded by $2k^{(4k+1)n}$. Therefore the bit-length is less than $O(k^2 \log k)$. □

The proofs of the facts in the following sections have to cope with many technical problems. MCD is a continuous problem, not a combinatorial one. The main difficulty, however, is that we cannot control the entries of the covariance matrix directly. In general a point in $\mathcal{X}'$ influences all entries in the covariance matrix $C(\mathcal{X}')$ as well as the center of gravity.

4.1 An Upper Bound on the Determinant of the Clique Configuration

Let $\mathcal{X}'$ constitute a clique configuration. The center of gravity $\mathbf{t}$ of $\mathcal{X}'$ is

$$\mathbf{t} = \frac{1}{h}\,(k,\dots,k,0,\dots,0)^T\,,$$

where the transition of the entries from k to 0 occurs after position k. The covariance matrix C of the clique configuration has the following form:

$$C = \frac{1}{h}\left[\begin{array}{cccc|cccc} a & b & \cdots & b & 0 & \cdots & \cdots & 0 \\ b & \ddots & \ddots & \vdots & \vdots & & & \vdots \\ \vdots & \ddots & \ddots & b & \vdots & & & \vdots \\ b & \cdots & b & a & 0 & \cdots & \cdots & 0 \\ \hline 0 & \cdots & \cdots & 0 & c & 0 & \cdots & 0 \\ \vdots & & & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 & c \end{array}\right]\,,$$

where the upper left submatrix is $k \times k$ and

$$\begin{aligned} a &= k + 2wz^2 - k^2/h \\ b &= 1 - k^2/h \\ c &= 2wz^2 \ . \end{aligned}$$

By Hadarmard's determinant inequality, see e. g. Horn and Johnson [7], the determinant of a positive definte matrix is bounded by the product of its diagonal elements. Hence we arrive at

$$\begin{aligned} h^n \cdot \det(C) \leq a^k \cdot c^{n-k} &= \left(k + 2wz^2 - \frac{k^2}{h}\right)^k \cdot \left(2wz^2\right)^{(n-k)} \\ &\leq \left(k + 2wz^2\right)^k \cdot \left(2wz^2\right)^{(n-k)} = h^n \cdot B \end{aligned}$$

4.2 All a-Points Have to be Selected

Let $\mathcal{X}' \subset \mathcal{X}$ be a set of h points. We want to show that for any choice of $\mathcal{X}'$ every a-point is closer to the center of gravity of the set $\mathcal{X}'$ than any v- or e-point. Closeness here is measured with respect to the Mahalanobis distance. More precisely:

Lemma 4.5 *Let* $\mathbf{a}$ *be an arbitrary a-point and let* $\mathbf{x}$ *be an arbitrary v- or e-point,* $\mathbf{a}, \mathbf{x} \in \mathcal{X}$. *Then the following relation holds for any set* $\mathcal{X}' \subset \mathcal{X}$ *with* $|\mathcal{X}'| = h$

$$\mathrm{md}(\mathbf{a}; C(\mathcal{X}'), \mathbf{t}(\mathcal{X}')) < \mathrm{md}(\mathbf{x}; C(\mathcal{X}'), \mathbf{t}(\mathcal{X}')) \ .$$

Proof. We try to construct $\mathcal{X}'$ in such a way that the difference

$$\mathrm{md}(\mathbf{x}; C(\mathcal{X}'), \mathbf{t}(\mathcal{X}')) - \mathrm{md}(\mathbf{a}; C(\mathcal{X}'), \mathbf{t}(\mathcal{X}'))$$

is minimized and show that it is always larger than 0. It then follows from Lemma 4.2 that a set $\mathcal{X}'$ defining a minimal covariance determinant has to contain all a-points.

In order to maximize the Mahalanobis distance of an a-point and simultaneously minimize that of a v-point w. r. t. C and $\mathbf{t}$, we even allow configurations of points which are not realizable by our reduction. We allow that a v- or e-point is chosen multiple times, in order to "pull the ellipsoid towards it".

The analysis distinguishes several cases. We describe the analysis of one case in detail, namely that of maximizing the Mahalanobis distance of an a-point and simultaneously minimizing that of a v-point in a different dimension. The arguments for the other cases follow the same line.[2]

The influence of a point on the covariance matrix – and hence the Mahalanobis distance – is maximal in the direction from the point to the center

[2]Maple worksheets for the cases not treated here can be found on the following website: http://ls2-www.cs.uni-dortmund.de/~bernholt/mcd/index.html

of gravity and is minimal in orthogonal directions. We thus consider a two-dimensional sub-scenario of a d-dimensional one. Let us, w.l.o.g., devote dimension 1 to minimize the distance to a v-point and dimension 2 to maximize that of an a-point.

As we have moved all degrees of freedom into dimensions 1 and 2, the a-points in dimensions $3, 4, \ldots, n$ are symmetrically placed. These a-points do not influence the upper left (2×2)-submatrix of the covariance matrix, they do however affect the center of gravity. As they are symmetrical with respect to the origin, they sum up to the origin. There are at most $n + n(n-1)/2$ v- and e-points. In the case we are looking at now, we allow u copies of the v-point $\mathbf{v}_1 := (1, 0, 0, , \ldots, , 0)^T$ and $w - u$ copies of the a-point $(0, z, 0, , \ldots, , 0)^T$. We then compute the (2-dimensional) covariance matrix C and the center of gravity $\mathbf{t}$ of this arrangement of points.

$$\mathbf{t} = (t_1, t_2)^T = \frac{1}{h}\Big(u, (w-u)z + w(-z)\Big)^T = \frac{1}{h}\Big(u, -zu\Big)^T \ ,$$

$$C = \begin{bmatrix} a & b \\ b & c \end{bmatrix} = \frac{1}{h}\begin{bmatrix} u + 2w\,z^2 - \frac{u^2}{h} & \frac{u^2\,z}{h} \\ \frac{u^2\,z}{h} & (2w-u)\,z^2 - \frac{u^2\,z^2}{h} \end{bmatrix} \ .$$

Let $\mathbf{a}_2 := (0, z, 0, , \ldots, , 0)^T$ be the a-point in dimension 2. Let $d(\mathbf{v}_1) = \mathrm{md}(\mathbf{v}_1; C, \mathbf{t})$ be the Mahalanobis distance of $\mathbf{v}_1$ w.r.t. C and $\mathbf{t}$, and let $d(\mathbf{a}_2) = \mathrm{md}(\mathbf{a}_2; C, \mathbf{t})$ be the corresponding value for $\mathbf{a}_2$. We consider the inequality $d(\mathbf{v}_1) - d(\mathbf{a}_2) > 0$. Multiplying with $\det(C)$ we find

$$\left(2\,t_2\,z - z^2\right)\,a + \left(2\,t_2 - 2\,t_1\,z\right)b + \left(-2\,t_1 + 1\right)c > 0 \ .$$

Substituting a, b, c, t_1, t_2 and multiplying with h/z^2 yields

$$\left(-2 - \frac{1+z^2}{k}\right)\,u + \left(1 - z^2\right)\,k^4 > 0 \ .$$

In order to minimize the left-hand side, one has to choose u as large as possible, i.e., $u = 2k + 2k(2k-1)/2$:

$$k^4 - 4\,k^2 + \left(-k^4 - 2\,k - 1\right)\,z^2 - 4\,k - 1 > 0 \ .$$

The term k^4 is the dominant one, and the left-hand side increases with k and thus the inequality is true for $k \geq 3$.

The other cases that one has to consider include distributing the u missing a-points in other possible ways and the consideration of two v-points and mixtures of e- and v-points. The arguments are along the same line and always establish that the corresponding difference of the Mahalanobis distance is larger than 0 for $k \geq 3$. Altogether it follows that the Mahalanobis distance of a v- or e-point is always larger than that of any a-point. $\square$

The following lemma is an immediate consequence of Lemma 4.5 and the fact that the origin is contained in the convex hull of the a-points.

Lemma 4.6 *Let* $\mathbf{0}$ *be the origin and let* $\mathbf{x}$ *be an arbitrary v- or e-point,* $\mathbf{x} \in \mathcal{X}$*. Then the following inequation holds for any set* $\mathcal{X}' \subset \mathcal{X}$ *with* $|\mathcal{X}'| = h$*:*

$$\mathrm{md}(\mathbf{0}; C(\mathcal{X}'), \mathbf{t}(\mathcal{X}')) < \mathrm{md}(\mathbf{x}; C(\mathcal{X}'), \mathbf{t}(\mathcal{X}')) \ .$$

4.3 Constructing a Minimal Configuration

Assume that the graph G of our clique problem does not contain a k-clique. Let $\mathcal{X}$ be the set of points of the corresponding MCDd problem and let $\mathcal{X}' \subseteq \mathcal{X}$ be any set of h points. We now show that $\det(\mathcal{X}')$ is at least as large as the determinant of the minimal $(k+1)$-configuration. To this end we show how $\mathcal{X}'$ can be transformed into the minimal $(k+1)$-configuration, whithout increasing the determinant.

Lemma 4.7 *Let* G *be a graph on* n *vertices without a* k*-clique,* $k = \lceil n/2 \rceil$*. Let* $\mathcal{X}$ *be the set of points of the corresponding MCDd problem. Let* $\mathcal{X}' \subseteq \mathcal{X}$ *be any set of* h *points. Let* D *be the determinant of a minimal* $(k+1)$*-configuration. Then*

$$\det(\mathcal{X}') \geq D \ .$$

Proof. For the proof let us introduce some notation. Given a set of $\mathcal{Y}$ consisting of e- and v-points we say that it *spans* k dimensions if the subspace spanned by the corresponding vectors is k-dimensional. We say that $\mathcal{Y}$ *touches* k dimensions if there are at least k positions in which some member of $\mathcal{Y}$ has a 1-entry. For example the set $\{(1,1,0,0,0),(0,0,1,1,0)\}$ spans 2 dimensions and touches 4 dimensions. Adding the vector $(0,1,1,0,0)$ does not increase the number of dimensions touched but increases the dimension of the span to 3.

From Section 4.2 we know that any h-element set with minimal covariance determinant has to contain all a-points. Consequently it has to contain exactly $m := (k-1)k + k$ e- or v-points $\mathbf{y}_1, \ldots, \mathbf{y}_m$. Let $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ be the set of these points. As G does not contain a k-clique, the vectors in $\mathcal{Y}$ span a space of at least $k+1$ dimensions.

If there are $k+1$ v-points in $\mathcal{Y}$ then we can achieve the minimal $(k+1)$-configuration directly by moving all but $k+1$ v-points to the origin. By Lemma 4.6 and the Exchange Lemma 4.2, the covariance determinant of the resulting configuration is less than or equal to that of the original configuration.

Otherwise we have to replace some e-points by v-points in addition, without increasing the determinant. In order to control the change of the determinant during the replacement, one has to carefully select which e- and v-points to keep and which to move into the origin. Therefore, the location of the points in $\mathcal{Y}$ relative to each other is important. We represent this structure as an undirected graph $H = H(\mathcal{Y})$. For every v-point $\mathbf{v}_i \in \mathcal{Y}$ there is a vertex i in H. For every e-point $\mathbf{e}_{ij} \in \mathcal{Y}$ the vertices i and j are in H as is the edge $\{i, j\}$. In order to distinguish between vertices which are solely introduced by e-points and those for which the corresponding v-point is in $\mathcal{Y}$ we call a vertex

i of H *marked* if $\mathbf{v}_i \in \mathcal{Y}$. The resulting graph H is isomorphic to a subgraph of the original graph G, but has two types of vertices, marked and unmarked ones. The marked vertices correspond to v-points really present in $\mathcal{Y}$ while the unmarked vertices of H do not have a corresponding v-point in $\mathcal{Y}$. They are merely induced by an e-point in $\mathcal{Y}$.

We now show how a set $\mathcal{Y}' \subset \mathcal{Y}$ can be constructed such that the resulting graph $H' = H(\mathcal{Y}')$ is cycle-free and that $\mathcal{Y}'$ spans $k+1$ dimensions.

Definition 4.8 *Let B be a tree with marked vertices as described above. If B has m edges the* value $w(B)$ *of B is defined by*

$$w(B) = \begin{cases} m+1 & \textit{if at least one vertex is marked,} \\ m & \textit{otherwise.} \end{cases}$$

Note that if the tree B is defined by a set $\mathcal{Y}$ of v- and e-points, i. e. $B = H(\mathcal{Y})$, and $w(B) = s$, then $\mathcal{Y}$ spans s dimensions, but might touch more. In contrast, a cyclic graph defined by a even number s of e-points, e. g. $\mathbf{e}_{12}, \mathbf{e}_{23}, \ldots, \mathbf{e}_{s1}$ only spans $s-1$ dimensions.

Claim 4.9 *If a graph H has at least $\binom{k}{2}$ edges and does not contain a k-clique then there are vertex-disjoint trees $B_1, \ldots, B_r$ in G with $\sum_{i=1}^{r} w(B_i) \geq k$.*

Proof. Assume that for all choices of r and vertex-disjoint trees $B_1, \ldots, B_r$ the equation $\sum_{i=1\ldots r} w(B_i) \leq k-1$ holds true. Then there are at most $k-1+r$ vertices in the trees $B_1, \ldots, B_r$.

Let $B_1, \ldots, B_r$ any r trees such that all edges of H lie within these trees. Let k_i be the number of vertices in tree B_i. We want to establish an upper bound for the number of edges in the connected components induced by the trees.

To this end let $B'_1, \ldots, B'_r$ be the vertex disjoint graphs induced by the trees $B_1, \ldots, B_r$. A graph B'_i has k_i vertices and at most $\binom{k_i}{2}$ edges. The number of edges in the graphs $B'_1, \ldots, B'_r$ is at most $Z = \binom{k_1}{2} + \cdots + \binom{k_r}{2}$ and the B'_i then contain $\sum_{i=1\ldots r} k_i = k-1+r$ vertices.

If $r = 1$, the graph B'_1 is identical to the graph H. Moreover, $w(B_1) \leq k-1$ hence B_1 (and H) has at most k vertices. As H has at least $\binom{k}{2}$ edges, it contains a k-clique contrary to our assumption.

Otherwise, if $r \geq 2$, the edges are distributed over two or more graphs and due to the convexity of W there are fewer edges in the graph G than $\binom{k}{2}$. So this leads to a contradiction. Thus there must be vertex disjoint trees $B_1, \ldots, B_r$ with $\sum_{i=1\ldots r} w(B_i) \geq k$. □

Claim 4.10 *Let H be a graph with M marked vertices, $\binom{k}{2} + k - M$ edges and let H contain no k-clique. Then there are vertex-disjoint trees $B_1, \ldots, B_r$ in H with $\sum_{i=1\ldots r} w(B_i) \geq k+1$.*

Proof. We prove this claim by constructing trees with the desired property:

- Case 1: $M \geq k+1$
  There are $k+1$ trees each consisting of a single marked vertex.

- Case 2: $1 \leq M \leq k$
  The graph has at least $\binom{k}{2}$ edges. Claim 4.9 shows that there exists some r and trees $B_1, \ldots, B_r$ such that $\sum_{i=1\ldots r} W(B_i) \geq k$. Take one marked vertex additionally.

- Case 3: $M = 0$
  The graph has at least $\binom{k}{2} + k = \binom{k+1}{2}$ edges. By Claim 4.9 it follows that there exists some r and trees $B_1, \ldots, B_r$ such that $\sum_{i=1\ldots r} w(B_i) \geq k+1$. $\square$

For the construction of the set $\mathcal{Y}'$ apply Claim 4.10 to the graph $H(\mathcal{Y})$. Let $B_1, \ldots, B_r$ be the resulting trees. Now move all e-points in $\mathcal{Y}$ corresponding to edges that are not present in some B_i into the origin. By Lemma 4.6 the covariance determinant is only decreased by this operation.

In the following, the e-points in $\mathcal{Y}'$ will be replaced by suitably chosen v-points. We then end up with $\mathcal{Y}'$ consisting of at least $k+1$ v-points and no e-point.

Let us consider a single tree B_i and the corresponding v- and e-points. The formula below shows a (3×3)-submatrix of corresponding rows and columns of the covariance matrix. For technical reasons the covariance matrix is split into the sum of the pure covariance part and the offset resulting from the fact that the center of gravity is not the origin. The first row is a prototype of a dimension which is touched by exactly q v- or e-points. The second row is a dimension which is solely touched by a single e-point that also touches the dimension of the third row. The third row represents a dimension which is either unmarked and is touched by p e-points, $p \geq 2$, or which is marked and touched by $p-1$ e-points, $p \geq 2$.

$$\begin{array}{l} \left(\begin{array}{ccc} q & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1+p \end{array}\right) + \left(\begin{array}{ccc} -\frac{q^2}{h} & -\frac{q}{h} & -\frac{(1+p)q}{h} \\ -\frac{q}{h} & -\frac{1}{h} & -\frac{1+p}{h} \\ -\frac{(1+p)q}{h} & -\frac{1+p}{h} & -\frac{(1+p)^2}{h} \end{array}\right) \end{array} \begin{array}{l} \text{any other row} \\ \text{a leaf of the tree} \\ \text{a node with } p \text{ v/e-points} \end{array}$$

We track the effect of replacing an e-point touching the dimension of row 2 and 3 by a v-point on the matrix. The replacement is reflected by the following operations on the rows and columns.

column 3 := column 3 - column 2
row 3 := row 3 - row 2

The resulting matrix is

$$\left(\begin{array}{ccc} q & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & p \end{array}\right) + \left(\begin{array}{ccc} -\frac{q^2}{h} & -\frac{q}{h} & -\frac{pq}{h} \\ -\frac{q}{h} & -\frac{1}{h} & -\frac{p}{h} \\ -\frac{pq}{h} & -\frac{p}{h} & -\frac{p^2}{h} \end{array}\right)$$

The determinant of the matrix has not been changed in this process. Thus after successively applying this process to all edges of the tree B_i, all points corresponding to the tree are replaced by isolated vertices. Claim 4.10 ensures that there are at least $k+1$ vertices. If we move the superfluous vertices into the origin, we obtain a minimal $(k+1)$-configuration and Lemma 4.7 has been proved. □

4.4 A Lower Bound on the Determinant of a Minimal $(k+1)$-Configuration

In this section we compute the covariance determinant of the minimal $(k+1)$-configuration as constructed in the previous section. The center of gravity $\mathbf{t}$ of the clique configuration is

$$\mathbf{t} = \frac{1}{h}(1, \ldots, 1, 0, \ldots, 0)^T,$$

where the transition of the entries from 1 to 0 occurs after position $k+1$. The covariance matrix C_m of the minimal $(k+1)$-configuration has the following form:

$$C_m = \frac{1}{h}\left[\begin{array}{cccc|cccc} a & b & \cdots & b & 0 & \cdots & \cdots & 0 \\ b & \ddots & \ddots & \vdots & \vdots & & & \vdots \\ \vdots & \ddots & \ddots & b & \vdots & & & \vdots \\ b & \cdots & b & a & 0 & \cdots & \cdots & 0 \\ \hline 0 & \cdots & \cdots & 0 & c & 0 & \cdots & 0 \\ \vdots & & & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 & c \end{array}\right]$$

where the upper left submatrix is $(k+1) \times (k+1)$ and

$$\begin{aligned} a &= 1 + 2wz^2 - 1/h \\ b &= -1/h \\ c &= 2wz^2 \; . \end{aligned}$$

According to Geršcorin's Disc Theorem, see e. g. [7], all eigenvalues of a matrix $M = [m_{ij}]$ are located in the union of the discs $|m_{ii} - \gamma| \leq \sum_{j=1, j\neq i}^{n} |m_{ij}|$ for

$\gamma \in \mathbb{C}$ such that the determinant is lower bounded for $k \geq 2$ as follows:

$$\begin{aligned}
&h^n \cdot \det(C_m) \\
&\quad \geq (a - k \cdot |b|)^{k+1} \cdot c^{n-k-1} \\
&\quad = \left(1 + 2wz^2 - \frac{k+1}{h}\right)^{k+1} \cdot (2wz^2)^{n-k-1} \\
&\quad \geq \left(\frac{9}{10}\right)^{k+1} \cdot (2wz^2)^{n-k-1} \\
&\quad \geq (k + 2wz^2)^k \cdot (2wz^2) \cdot (2wz^2)^{n-k-1} \\
&\quad = h^n \cdot B
\end{aligned}$$

We used the following relations:

$$\begin{aligned}
1 + 2wz^2 - (k+1)/h > 9/10 \text{ for } k \geq 2 \\
9/10 > (k + 2wz^2)^k \cdot (2wz^2) \text{ for } k \geq 2
\end{aligned}$$

And that completes the proof of the NP-completeness of MCD.

5 Summary

We have presented a polynomial-time algorithm for the minimum covariance determinant problem for fixed dimensions of the data. On the other hand we have shown that the problem is NP-hard for varying dimension.

The running time of our algorithm on N d-dimensional data points is $O\left(N^{d(d+3)/2}\right)$. The hardness result suggests that any uniform algorithm for the MCD problem has a running time where d appears more than poly-logarithmic in the exponent. It is, however, possible that algorithms exist which have a running time of $N^{\mathcal{O}(d)}$.

A paper concerning implementation details and empirical results gained on benchmark data is in preparation.

Let us also remark that the algorithm can be easily adapted for the Minimum Volume Ellipsoid problem and that our result implies that this problem is NP-complete for varying dimension as well.

Acknowledgement

We would like to thank Claudia Becker, Thomas Fender, Ursula Gather for introducing us to the questions of robust statistics in general and in particular to the MCD-problem. We are also grateful for suggestions concerning the presentation of the statistical results. We thank Thomas Hofmeister for pointing out a simpler proof of Claim 4.9.

A Appendix

n	number of vertices of graph G
m	number of edges of graph G
N	number of points of the MCD-problem (here $N = n + m + nk^4$)
k	clique size (here $k = \lceil n/2 \rceil$)
h	selection size (here $h = k + \binom{k}{2} + nk^4$)
z	distance of an a-point from the origin (here $z = k^{-2k}$)
w	weight of an a-point (here $w = k^4/2$)
B	Bound for MCDd here $B = \frac{1}{h^n}\left(k + 2wz^2\right)^k \cdot \left(2wz^2\right)^{(n-k)}$
$\mathcal{X}$	the set of points in $\mathbb{R}^n$ constructed in the reduction
$\mathcal{X}'$	the subset of h points for which the covariance determinant is computed

Table 1: The table summarizes the parameters used in the paper.

References

[1] C. Becker and U. Gather. The largest nonidentifiable outlier: A comparison of multivariate simultaneous outlier identification rules. *Computational Statistics and Data Analysis*, 36:119–217, 2000.

[2] R. W. Buttler, P. L. Davies, and M. Jhun. Asymptotics for the minimum covariance determinant estimator. *Annals of Statistics*, 21:1385–1400, 1993.

[3] P. L. Davies. Asymptotic behaviour of s-estimates of multivariate location parameters and dispersion matrices. *Annals of Statistics*, 15:1269–1292, 1987.

[4] R. Grübel. A minimal characterization of the covariance matrix. *Metrika*, 35:49–52, 1988.

[5] D. M. Hawkins. A feasible solution algorithms for the minimum covariance determinant estimator in multivariate data. *Computational Statistics and Data Analysis*, 17:197–210, 1994.

[6] D. M. Hawkins and D. J. Olive. Improved feasible solution algorithms for high breakdown estimation. *Computational Statistics and Data Analysis*, 30:1–11, 1999.

[7] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[8] P. J. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association*, 79:871–880, 1984.

[9] P. J. Rousseeuw and K. Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223, 1999.

[10] D. Woodruff and D. Rocke. Computable robust estimation of multivariate location and shape in high dimension using compound estimators. *Journal of the American Statistical Association*, 89:888–896, 1994.

[11] D. Woodruff and D. Rocke. Identification of outliers in multivariate data. *Journal of the American Statistical Association*, 91:1047–1061, 1996.