# Computing the Minimum Covariance Determinant

Jiaqi Wang
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA, jwang3737@gatech.edu

Weijun Xie
Department of Data Science, City University of Hong Kong, Hong Kong, wj.xie@cityu.edu.hk

**Abstract**

The minimum covariance determinant (MCD) problem selects a fixed-cardinality subset of data points to minimize its sample covariance determinant. We develop approximation and exact algorithms and apply them to prediction residuals, using the selected mean and covariance to define ellipsoidal conformal regions. For rational data, quadratic-form selection gives a deterministic approximation scheme for the geometric mean of the covariance eigenvalues. Its running time is polynomial for fixed dimension and accuracy, with an exponent linear in the dimension at fixed accuracy. An affine-hull test solves instances with optimal value zero exactly. Zonotope enumeration gives an exact algorithm with polynomial bit complexity in fixed dimension, without a general-position assumption, and permits repeated observations and affine dependencies. Sampled versions of both constructions provide a hybrid initialization for branch-and-bound with spectral and completion lower bounds. We also derive two-sided bounds relating the optimal MCD value to minimum covering ellipsoid volume. Computational comparisons under common search rules and time limits show that the proposed node bounds increase the number of instances solved within numerical tolerances. Prediction experiments on synthetic, weather, and river-flow data show smaller calibrated region volumes with empirical coverage evaluated alongside volume.

**Keywords.** Minimum covariance determinant, subset selection, approximation algorithms, fixed-dimensional optimization, branch-and-bound, conformal prediction.

## 1 Introduction

### 1.1 Problem formulation

Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{Q}^d$ be data points, where $d \geq 1$, and let $h \in \{1, \ldots, n\}$ be the fixed subset size. Write $[n] = \{1, \ldots, n\}$. For each nonempty $S \subseteq [n]$, define the sample mean and covariance matrix by

$$\bar{\boldsymbol{v}}_S = \frac{1}{|S|} \sum_{i \in S} \boldsymbol{v}_i, \qquad \boldsymbol{C}_S = \frac{1}{|S|} \sum_{i \in S} (\boldsymbol{v}_i - \bar{\boldsymbol{v}}_S)(\boldsymbol{v}_i - \bar{\boldsymbol{v}}_S)^\top.$$

Let $v(S) = \sqrt[d]{\det(\boldsymbol{C}_S)}$ be the objective value of $S$. The minimum covariance determinant (MCD) problem is formulated as

$$v^\star = \min_{\substack{S \subseteq [n] \\ |S| = h}} \sqrt[d]{\det(\boldsymbol{C}_S)}. \tag{1}$$

Taking the $d$th root preserves the minimizing subsets, so MCD (1) is equivalent to the determinant minimization formulation. The sample mean and covariance of an optimal subset estimate the center and spread of the data. We write $k = n - h$ for the number of excluded data points. The objective $v(S)$ is the geometric mean of the eigenvalues of $\boldsymbol{C}_S$. Both the sample mean and covariance depend on the selected subset. Since $\boldsymbol{C}_S$ is positive semidefinite, $v^\star \geq 0$. The optimal value is zero if and only if some $h$ data points lie in a proper affine subspace of $\mathbb{R}^d$. In particular, $h \leq d$ implies $v^\star = 0$.

### 1.2 Applications

The MCD estimator provides robust estimates of the mean vector and covariance matrix, with $h$ specifying the number of data points used in the estimates [1, 2]. Covariance matrices are inputs to several operations research models. In portfolio optimization, they determine portfolio variance [3], and robust estimates have been used in mean–variance models with integer transaction lots [4]. In process monitoring and outlier detection, they define multivariate control statistics and Mahalanobis distances, which measure deviations from the mean using the inverse covariance [5, 6]. In robust optimization, they specify the shape of ellipsoidal uncertainty sets [7]. These applications motivate algorithms for solving MCD (1) with approximation guarantees and computable bounds on the optimal value.

We apply the MCD algorithms to conformal prediction, which constructs prediction regions with finite-sample coverage guarantees under exchangeability [8, 9]. Exchangeability means that the joint distribution is invariant under permutations of the observations. The coverage guarantee is marginal: the probability that the region contains a future response, averaged over the observed data and the future observation, is at least the target coverage level. In split-conformal prediction, the prediction function and nonconformity score are specified using data separate from the calibration set. Once both are fixed, the calibration data determine a threshold through an order statistic of the error values, with a finite-sample correction. The prediction region contains all responses whose nonconformity score does not exceed this threshold. Its geometry and volume depend on both the nonconformity score and the calibrated threshold.

For multivariate regression, we apply MCD to residual vectors, defined as the differences between observed and predicted responses. The selected subset determines a residual mean and covariance. When the covariance is positive definite, its Mahalanobis distance defines ellipsoidal prediction regions. The residual mean adjusts the prediction center. The covariance specifies the shape and orientation, and calibration determines the radius in the Mahalanobis metric. The region volume depends on both the covariance determinant and the radius. In our numerical study, we compare volume and empirical coverage on synthetic, weather, and river-flow data.

### 1.3 Relevant literature

Hubert et al. [2] review the statistical properties and reweighting procedures of the MCD estimator. Related developments include robust principal component analysis (ROBPCA) [10] and the minimum regularized covariance determinant estimator [11]. We study the optimization of the MCD criterion through approximation algorithms, exact enumeration, and branch-and-bound.

Computational methods include Hawkins's feasible solution algorithm [12], Woodruff's ghost-image method [13], FAST-MCD [1], and RelaxMCD [14]. These methods use multistart search, concentration steps, or continuous optimization. We provide a deterministic worst-case approximation guarantee for $v(S)$ to complement these feasible-solution procedures. For fixed multiplicative accuracy, our running-time exponent is linear in $d$, compared with the quadratic dependence in the exact algorithm of Bernholt and Fischer [15]. To our knowledge, this is the first MCD approximation–complexity tradeoff with this dimension dependence for the determinant-root objective. The guarantee applies when all index–weight tuples in the specified finite family are examined.

Our approximation proof draws on the inverse-moment analysis of leveraged volume sampling, a determinant-weighted sampling method for least-squares regression [16]. A trace-product inequality and discretization of bounded weights convert the inverse-moment argument into an approximation guarantee for $v(S)$. Related results for D-optimal design and data fusion concern information

determinant maximization [17, 18]. Our analysis addresses covariance determinant minimization, including the dependence of the center on the selected subset.

For exact optimization, Bernholt and Fischer [15] impose a general-position condition on the data points relative to quadratic surfaces. We remove this condition and give an $n^{O(d^2)}\,\mathrm{poly}(B)$-time algorithm for arbitrary rational data points, where $B$ is the binary input encoding length. The result includes repeated data points, affine dependencies, and both $v^\star = 0$ and $v^\star > 0$. We specialize the projected-edge method of Onn and Rothblum [19], which uses images of polytope edge directions to enumerate linear objectives, to the centered MCD criterion and provide an exact rational implementation.

For branch-and-bound, Agulló [20] develops an order-statistic bound on completing a partially selected subset, while Pesch [21, 22] uses determinant monotonicity and geometric pruning. Our completion bound dominates Pesch's scalar monotonicity bound and complements Agulló's bound. We also derive spectral bounds based on the possible exclusions. These bounds can be informative even when the covariance of the data points fixed for inclusion has determinant zero. The construction uses cardinality-dependent Gershgorin bounds, quadratic-knapsack relaxations, and spectral moment inequalities [23–25]. Comparisons within a common solver show that the bounds increase the number of searches completed under the same time limit and numerical tolerances.

Conformal uncertainty sets and ellipsoidal prediction regions are studied by Johnstone and Cox [26] and Messoudi et al. [27]. Braun et al. [28] directly optimize covering-set volume. We optimize the residual subset defining the ellipsoid's mean and covariance, with approximation guarantees and computable bounds on the MCD objective. We also derive a two-sided comparison between $v^\star$ and the minimum covering ellipsoid volume. Coverage follows from separate conformal calibration under exchangeability. The experiments identify contamination settings in which the resulting regions have smaller calibrated volumes than the comparison methods.

### 1.4 Summary of contributions

In this paper, we develop approximation and exact algorithms for solving the MCD (1), derive bounds on covering ellipsoid volume, and evaluate the algorithms and prediction regions computationally. The contributions are as follows.

1. We develop an approximation algorithm based on quadratic-form selection. An affine-hull test solves the case $v^\star = 0$ exactly. For $v^\star > 0$, the algorithm constructs auxiliary matrices from weighted outer products of augmented data points. The inverse of each positive definite matrix defines the objective coefficients of a cardinality-constrained linear selection problem. For any integer $t \geq d+1$ and rational grid parameter $\delta \in (0, 1]$, enumerating all index tuples of length $t$ and their grid weights returns a feasible subset $S_{\mathrm{alg}}$ satisfying $\sqrt[d]{\det(\boldsymbol{C}_{S_{\mathrm{alg}}})} \leq (1+\delta)\frac{t}{t-d}v^\star$. The approximation ratio can be made arbitrarily close to one, with polynomial running time in the input encoding length for fixed dimension and accuracy.

2. We analyze an exact algorithm with bit complexity $n^{O(d^2)}\,\mathrm{poly}(B)$. It partitions a parameter space of dimension at most $d(d+3)/2$ by hyperplanes on which two linear selection costs coincide, and solves one linear selection problem in each full-dimensional region. Different from the literature, the algorithm applies to repeated data points and affine dependencies, with no position assumption.

3. We develop a branch-and-bound algorithm with spectral and completion lower bounds. The spectral bounds limit the joint determinant decrease from the remaining exclusions, while the

completion bounds account for the additional observations required for inclusion. A hybrid initialization combines sampled quadratic-form selection and zonotope walks, using the best generated subset as the initial incumbent.

4. We derive two-sided bounds relating the optimal MCD value to the minimum volume of an ellipsoid containing at least $h$ data points. For $v^\star > 0$, a feasible subset satisfying $v(S) \le \rho v^\star$, with $\rho \ge 1$, yields an ellipsoid covering the selected data points with volume at most $\big(\rho(h-1)/d\big)^{d/2}$ times the minimum covering volume.

**Organization.** Section 2 introduces the approximation algorithm and analyzes $v^\star = 0$ and $v^\star > 0$ separately. Section 3 presents exact enumeration. Section 4 develops branch-and-bound with hybrid initialization and spectral bounds. Section 5 derives the covering-volume bounds. Section 6 reports the computational results, and Section 7 concludes.

**Notation.** For a positive integer $m$, let $[m] = \{1, \dots, m\}$. Bold lowercase and uppercase letters denote vectors and matrices. We write $\boldsymbol{I}_m$ for the identity matrix, $\mathbf{1}$ for an all-ones vector, and $\boldsymbol{e}_j$ for a coordinate unit vector of the required dimension. The sets $\mathbb{S}^m$, $\mathbb{S}^m_+$, and $\mathbb{S}^m_{++}$ contain symmetric, positive semidefinite, and positive definite real matrices, respectively. We write $\boldsymbol{A} \succeq \boldsymbol{B}$ if $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite, and $\boldsymbol{A} \succ \boldsymbol{B}$ if $\boldsymbol{A} - \boldsymbol{B}$ is positive definite. The matrix inner product is $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \operatorname{tr}(\boldsymbol{A}^\top \boldsymbol{B})$. We write $\det(\boldsymbol{A})$, $\operatorname{tr}(\boldsymbol{A})$, and $\operatorname{rank}(\boldsymbol{A})$ for the determinant, trace, and rank. The symbols aff, conv, and span denote affine hull, convex hull, and linear span. For a symmetric matrix, $\lambda_{\min}$ and $\lambda_{\max}$ denote its smallest and largest eigenvalues, respectively. The constant $\omega_d$ is the volume of the Euclidean unit ball in $\mathbb{R}^d$. Probability laws are denoted by $\mathbb{P}$ and $\mathbb{Q}$, expectations by $\mathbb{E}$, and random quantities by a tilde. All logarithms are natural. In complexity bounds, poly denotes a polynomial in its displayed arguments.

## 2 Approximation by quadratic-form selection

We construct feasible subsets by solving cardinality-constrained linear selection problems. An augmented moment matrix represents the covariance objective, and the inverse of a positive definite auxiliary matrix defines the selection coefficients as quadratic forms in the augmented data points. We treat the cases $v^\star = 0$ and $v^\star > 0$ separately.

First, let $r = d + 1$. Append a final coordinate equal to one to each data point, and define

$$\begin{aligned} \boldsymbol{z}_i &= (\boldsymbol{v}_i^\top, 1)^\top \in \mathbb{Q}^r, && \forall\, i \in [n], \\ \boldsymbol{Z}_S &= \sum_{i \in S} \boldsymbol{z}_i \boldsymbol{z}_i^\top, && \forall\, S \subseteq [n]. \end{aligned}$$

We call $\boldsymbol{z}_i$ the augmented data point and $\boldsymbol{Z}_S$ the subset moment matrix. By convention, the empty sum is $\boldsymbol{Z}_\varnothing = \mathbf{0}$. For $S \subseteq [n]$ with $|S| = h$, we have

$$\boldsymbol{Z}_S = \begin{pmatrix} \sum_{i \in S} \boldsymbol{v}_i \boldsymbol{v}_i^\top & h\bar{\boldsymbol{v}}_S \\ h\bar{\boldsymbol{v}}_S^\top & h \end{pmatrix}.$$

The Schur complement of the bottom-right entry is $\sum_{i \in S} \boldsymbol{v}_i \boldsymbol{v}_i^\top - h\bar{\boldsymbol{v}}_S \bar{\boldsymbol{v}}_S^\top = h\boldsymbol{C}_S$. The block determinant identity yields

$$\det(\boldsymbol{Z}_S) = h^r \det(\boldsymbol{C}_S), \qquad v(S) = h^{-r/d} \sqrt[d]{\det(\boldsymbol{Z}_S)}. \tag{2}$$

Both identities hold when the determinant is zero. The subset moment matrix accounts for the subset-dependent mean and is linear in the selection indicators: if $x_i = 1$ for $i \in S$ and $x_i = 0$ otherwise, then $\boldsymbol{Z}_S = \sum_{i=1}^n x_i \boldsymbol{z}_i \boldsymbol{z}_i^\top$.

### 2.1 The case $v^\star = 0$

The affine hull is the smallest affine subspace containing the data points. Its dimension equals the rank of their covariance matrix. The following criterion gives an exact test for $v^\star = 0$.

**Proposition 1** (Affine-hull criterion)**.** *For every nonempty $S \subseteq [n]$,*

$$\operatorname{rank}(\boldsymbol{C}_S) = \dim \operatorname{aff}\{\boldsymbol{v}_i : i \in S\}. \tag{3}$$

*If $h \le d$, then $v^\star = 0$. If $h > d$, then $v^\star = 0$ if and only if the affine hull of some $d$ data points contains at least $h$ data points, counted by their indices.*

*Proof.* Let $\boldsymbol{X}_S$ have columns $\boldsymbol{v}_i - \bar{\boldsymbol{v}}_S$, indexed by $i \in S$. Since $\boldsymbol{C}_S = |S|^{-1} \boldsymbol{X}_S \boldsymbol{X}_S^\top$, its rank equals the dimension of the span of the centered data points, which proves (3).

If $h \le d$, this affine dimension is at most $h - 1 < d$ for every feasible subset. Suppose $h > d$ and $v^\star = 0$. Choose a feasible subset $S$ with $\det(\boldsymbol{C}_S) = 0$. An affine basis, that is, an affinely independent set spanning its affine hull, uses at most $d$ indices. Add indices from $S$ until the resulting set has cardinality $d$. Its affine hull contains every data point indexed by $S$. Conversely, the affine hull of $d$ data points has dimension at most $d - 1$, so any $h$ data points in that hull have objective value zero. □

When $h \le d$, any feasible subset is optimal. When $h > d$, we enumerate the $\binom{n}{d}$ index sets and test which data points belong to each affine hull. If one hull contains at least $h$ data points, any $h$ corresponding indices give an optimal solution. If no such subset exists, then $v^\star > 0$ and every feasible covariance matrix is positive definite.

### 2.2 The case $v^\star > 0$

Let $\boldsymbol{A} \in \mathbb{S}^r_{++}$. We call $\boldsymbol{A}$ an auxiliary matrix since it is used only to define selection costs. For each data point, let

$$q_i(\boldsymbol{A}) = \boldsymbol{z}_i^\top \boldsymbol{A}^{-1} \boldsymbol{z}_i, \qquad \forall\, i \in [n].$$

With $\boldsymbol{A}$ fixed, solve the linear selection problem

$$\min_{\boldsymbol{x} \in \{0,1\}^n} \left\{ \sum_{i=1}^n q_i(\boldsymbol{A}) x_i : \sum_{i=1}^n x_i = h \right\}. \tag{4}$$

Let $S(\boldsymbol{A})$ contain the indices of the $h$ smallest costs $q_i(\boldsymbol{A})$, with ties resolved by increasing index. Equivalently, $S(\boldsymbol{A})$ minimizes $\operatorname{tr}(\boldsymbol{A}^{-1} \boldsymbol{Z}_S)$ over $S \subseteq [n]$ with $|S| = h$. The auxiliary matrix $\boldsymbol{A}$ determines the selection costs. The selected subset is evaluated using its sample covariance through the MCD objective $v(S(\boldsymbol{A})) = \sqrt[d]{\det(\boldsymbol{C}_{S(\boldsymbol{A})})}$.

Fix an integer $t \ge r$ and a rational $\delta \in (0, 1]$. The integer $t$ is the number of terms used to form an auxiliary matrix. It can differ from the selected cardinality $h$. Define the geometric grid

$$K = \min\{j \in \mathbb{Z}_{\ge 0} : (1+\delta)^j \ge h\}, \qquad \mathcal{G}_\delta = \{(1+\delta)^j : j \in \{0, \dots, K\}\}.$$

Successive grid values differ by the factor $1+\delta$. For an ordered tuple of indices $(i_1,\dots,i_t)\in[n]^t$, with repetitions allowed, and weights $(g_1,\dots,g_t)\in\mathcal{G}_\delta^t$, form

$$\boldsymbol{A}=\frac{r}{t}\sum_{j=1}^{t}g_j\boldsymbol{z}_{i_j}\boldsymbol{z}_{i_j}^\top. \tag{5}$$

For each index tuple $(i_1,\dots,i_t)\in[n]^t$, enumerate all weight vectors $(g_1,\dots,g_t)\in\mathcal{G}_\delta^t$. The weight $g_j$ is assigned to the $j$th tuple entry and multiplies $\boldsymbol{z}_{i_j}\boldsymbol{z}_{i_j}^\top$ in (5). Since the weights are positive, $\boldsymbol{A}\succ\boldsymbol{0}$ exactly when the tuple vectors $\{\boldsymbol{z}_{i_j}\}_{j=1}^t$ span $\mathbb{R}^r$. Skip tuples that do not satisfy this condition.

For each remaining index–weight tuple, compute $q_i(\boldsymbol{A})=\boldsymbol{z}_i^\top\boldsymbol{A}^{-1}\boldsymbol{z}_i$ for all $i\in[n]$, and let $S(\boldsymbol{A})$ contain the $h$ indices with the smallest costs, breaking ties by index. The selection is made from all $n$ observations, with no requirement to include the indices used to construct $\boldsymbol{A}$. Compute the mean and covariance of $S(\boldsymbol{A})$ using equal weights $1/h$. After examining all index–weight tuples, return the candidate with the smallest covariance determinant. Thus candidates from different tuples are compared by the MCD objective. We call this procedure quadratic-form selection (QFS).

Note that the weights determine the selection costs through a weighted mean and covariance. For the current index–weight tuple, define

$$G=\sum_{j=1}^{t}g_j,\qquad \boldsymbol{\mu}_g=\frac{1}{G}\sum_{j=1}^{t}g_j\boldsymbol{v}_{i_j},\qquad \boldsymbol{C}_g=\frac{1}{G}\sum_{j=1}^{t}g_j(\boldsymbol{v}_{i_j}-\boldsymbol{\mu}_g)(\boldsymbol{v}_{i_j}-\boldsymbol{\mu}_g)^\top.$$

If $\boldsymbol{A}\succ\boldsymbol{0}$, then $\boldsymbol{C}_g\succ\boldsymbol{0}$, and block inversion leads to

$$q_i(\boldsymbol{A})=\frac{t}{rG}\left[1+(\boldsymbol{v}_i-\boldsymbol{\mu}_g)^\top\boldsymbol{C}_g^{-1}(\boldsymbol{v}_i-\boldsymbol{\mu}_g)\right],\qquad \forall\, i\in[n].$$

For any fixed indices $i_1,\dots,i_t$ and weights $g_1,\dots,g_t$, the quantities $G$, $\boldsymbol{\mu}_g$, and $\boldsymbol{C}_g$ appear in the formula of evaluating $q_i(\boldsymbol{A})$ for all $i\in[n]$. Adding one and multiplying by the positive factor $t/(rG)$ preserves the ordering of the squared Mahalanobis distances. For a center $\boldsymbol{\mu}$ and a positive definite covariance matrix $\boldsymbol{C}$, the Mahalanobis distance of $\boldsymbol{v}$ from $\boldsymbol{\mu}$ is $\sqrt{(\boldsymbol{v}-\boldsymbol{\mu})^\top\boldsymbol{C}^{-1}(\boldsymbol{v}-\boldsymbol{\mu})}$. Thus, $S(\boldsymbol{A})$ consists of the $h$ observations with smallest distances under $(\boldsymbol{\mu}_g,\boldsymbol{C}_g)$, with ties resolved by index. Hence, selecting the $h$ smallest costs is equivalent to selecting the $h$ smallest Mahalanobis distances under $(\boldsymbol{\mu}_g,\boldsymbol{C}_g)$. These weighted quantities define the selection rule only. The selected subset is evaluated using its own equally weighted mean and covariance.

Algorithm 1 combines the affine-hull test and the finite enumeration. It compares ordinary determinants, which are rational and order feasible subsets in the same way as $v(S)$. The variable $D_{\text{best}}$ in the algorithm represents the determinant of the best subset found.

**Algorithm 1** Quadratic-form selection for MCD

**Require:** Data points $\boldsymbol{v}_i \in \mathbb{Q}^d$ , $i \in [n]$, cardinality $1 \le h \le n$, integer $t \ge d+1$, and rational $\delta \in (0, 1]$.
**Ensure:** A feasible subset $S_{\text{alg}}$.

**if** $h \le d$ **then**
  **return** $\{1, \dots, h\}$.
**end if**
**for** each $T \subseteq [n]$ with $|T| = d$ **do**
  $H_T \leftarrow \{i \in [n] : \boldsymbol{v}_i \in \text{aff}\{\boldsymbol{v}_j : j \in T\}\}$.
  **if** $|H_T| \ge h$ **then**
    **return** the $h$ smallest indices in $H_T$.
  **end if**
**end for**
Set $r \leftarrow d+1$ and $\boldsymbol{z}_i \leftarrow (\boldsymbol{v}_i^\top, 1)^\top$, for each $i \in [n]$.
Form the grid $\mathcal{G}_\delta$ defined above.
$S_{\text{best}} \leftarrow \{1, \dots, h\}$ and $D_{\text{best}} \leftarrow \det(\boldsymbol{C}_{S_{\text{best}}})$.
**for** each tuple $(i_1, \dots, i_t, g_1, \dots, g_t) \in [n]^t \times \mathcal{G}_\delta^t$ **do**
  Form $\boldsymbol{A}$ using (5).
  **if** $\boldsymbol{A} \succ \boldsymbol{0}$ **then**
    Solve $\boldsymbol{A}\boldsymbol{y}_i = \boldsymbol{z}_i$ and set $q_i \leftarrow \boldsymbol{z}_i^\top \boldsymbol{y}_i$, for each $i \in [n]$.
    Let $S$ contain the first $h$ indices ordered by $(q_i, i)$.
    Compute $\bar{\boldsymbol{v}}_S$, $\boldsymbol{C}_S$, and $D \leftarrow \det(\boldsymbol{C}_S)$.
    **if** $D < D_{\text{best}}$ **then**
      $S_{\text{best}} \leftarrow S$ and $D_{\text{best}} \leftarrow D$.
    **end if**
  **end if**
**end for**
**return** $S_{\text{best}}$, with objective value $\sqrt[d]{D_{\text{best}}}$.

**Theorem 1** (Quadratic-form selection)**.** *Algorithm 1 returns a feasible subset $S_{\text{alg}}$ satisfying*

$$\sqrt[d]{\det(\boldsymbol{C}_{S_{\text{alg}}})} \le (1+\delta)\frac{t}{t-d}\, v^\star. \tag{6}$$

*It returns an optimal subset whenever $v^\star = 0$. For $h > d$, let $B$ be the binary encoding length of the data points, $h$, and $\delta$. Its bit complexity is bounded by*

$$\left[\binom{n}{d} + \big(n(K+1)\big)^t\right] \text{poly}(B, t, K). \tag{7}$$

Note that increasing $t$ reduces the factor $t/(t-d)$, and decreasing $\delta$ reduces the error from rounding the weights. For example, setting $t = 2d$ and $\delta = 1/2$ guarantees a factor 3 for $v(S)$ in $n^{O(d)}\,\text{poly}(B)$ time. The next subsection proves Theorem 1.

### 2.3 Proof of Theorem 1

We first prove the following lemma that compares the objective of the feasible subset $S(\boldsymbol{A})$ with that of any feasible subset $P$. The bound is expressed as a product of two traces.

**Lemma 1** (Trace-product bound)**.** *Let $P \subseteq [n]$ satisfy $|P| = h$ and $\boldsymbol{Z}_P \succ \boldsymbol{0}$. For every $\boldsymbol{A} \succ \boldsymbol{0}$, we have*

$$\frac{\sqrt[d]{\det(\boldsymbol{C}_{S(\boldsymbol{A})})}}{\sqrt[d]{\det(\boldsymbol{C}_P)}} \le \frac{\text{tr}(\boldsymbol{Z}_P^{-1}\boldsymbol{A})\,\text{tr}(\boldsymbol{A}^{-1}\boldsymbol{Z}_P)}{r^2}.$$

*Proof.* By optimality in (4), we have

$$\operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{Z}_{S(\boldsymbol{A})}) = \sum_{i\in S(\boldsymbol{A})} q_i(\boldsymbol{A}) \le \sum_{i\in P} q_i(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{Z}_P).$$

Set $\boldsymbol{Y} = \boldsymbol{Z}_P^{-1/2}\boldsymbol{A}\boldsymbol{Z}_P^{-1/2}$, $a = \operatorname{tr}(\boldsymbol{Y})$, and $b = \operatorname{tr}(\boldsymbol{Y}^{-1})$. Let $\boldsymbol{M} = \boldsymbol{A}^{-1/2}\boldsymbol{Z}_{S(\boldsymbol{A})}\boldsymbol{A}^{-1/2} \succeq \boldsymbol{0}$. Since $\boldsymbol{Y}^{-1} = \boldsymbol{Z}_P^{1/2}\boldsymbol{A}^{-1}\boldsymbol{Z}_P^{1/2}$, cyclic invariance of the trace and the preceding selection inequality imply

$$\operatorname{tr}(\boldsymbol{M}) = \operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{Z}_{S(\boldsymbol{A})}) \le \operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{Z}_P) = \operatorname{tr}(\boldsymbol{Y}^{-1}) = b.$$

Let $\lambda_1, \dots, \lambda_r$ be the eigenvalues of $\boldsymbol{M}$, with $\lambda_j \ge 0$, for all $j \in [r]$. The arithmetic–geometric mean inequality implies

$$\det(\boldsymbol{M}) = \prod_{j=1}^{r} \lambda_j \le \left(\frac{1}{r}\sum_{j=1}^{r}\lambda_j\right)^r = \left(\frac{\operatorname{tr}(\boldsymbol{M})}{r}\right)^r \le \left(\frac{b}{r}\right)^r.$$

Using $\det(\boldsymbol{M}) = \det(\boldsymbol{Z}_{S(\boldsymbol{A})})/\det(\boldsymbol{A})$ and $\det(\boldsymbol{Y}) = \det(\boldsymbol{A})/\det(\boldsymbol{Z}_P)$, we obtain

$$\frac{\det(\boldsymbol{Z}_{S(\boldsymbol{A})})}{\det(\boldsymbol{Z}_P)} = \det(\boldsymbol{Y})\det(\boldsymbol{M}) \le \det(\boldsymbol{Y})\left(\frac{b}{r}\right)^r.$$

Let $e_{r-1}(\boldsymbol{Y})$ be the sum of all products of $r-1$ distinct eigenvalues of $\boldsymbol{Y}$. Maclaurin's inequality bounds this term by

$$\det(\boldsymbol{Y})b = e_{r-1}(\boldsymbol{Y}) \le r\left(\frac{a}{r}\right)^{r-1}.$$

Combining the inequalities, we obtain

$$\frac{\det(\boldsymbol{Z}_{S(\boldsymbol{A})})}{\det(\boldsymbol{Z}_P)} \le \left(\frac{ab}{r^2}\right)^{r-1}.$$

By (2), the determinant ratio on the left is $\big(v(S(\boldsymbol{A}))/v(P)\big)^d$. Taking the $d$th root and using $r-1=d$ proves the claim. Cyclic invariance of the trace gives $a = \operatorname{tr}(\boldsymbol{Z}_P^{-1}\boldsymbol{A})$ and $b = \operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{Z}_P)$. □

We prove the existence of a suitable weighted auxiliary matrix by determinant reweighting. Starting from independently sampled indices, we multiply each tuple probability by the determinant of its associated matrix and normalize, as in leveraged volume sampling [16]. The resulting distribution is supported on positive definite matrices. We bound the expected trace of the inverse matrix under this distribution, so at least one realization satisfies the same upper bound.

**Lemma 2** (Auxiliary matrix with bounded weights)**.** *Let $P \subseteq [n]$ satisfy $|P| = h$ and $\boldsymbol{Z}_P \succ \boldsymbol{0}$. For every integer $t \ge r$, there exist tuples $(i_1, \dots, i_t) \in P^t$ and $(\gamma_1, \dots, \gamma_t) \in [1, h]^t$ such that*

$$\boldsymbol{A} = \frac{r}{t}\sum_{j=1}^{t}\gamma_j \boldsymbol{z}_{i_j}\boldsymbol{z}_{i_j}^\top \succ \boldsymbol{0}, \qquad \operatorname{tr}(\boldsymbol{Z}_P^{-1}\boldsymbol{A}) = r, \qquad \operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{Z}_P) \le \frac{rt}{t-r+1}.$$

*Proof.* The proof has three steps.

**Step 1.** Define the leverage values relative to $P$ by

$$\ell_i = \boldsymbol{z}_i^\top \boldsymbol{Z}_P^{-1}\boldsymbol{z}_i, \qquad \forall\, i \in P.$$

Taking a trace gives $\sum_{i\in P}\ell_i = r$. The inequality $\boldsymbol{z}_i\boldsymbol{z}_i^\top \preceq \boldsymbol{Z}_P$ implies $\ell_i \le 1$, for all $i \in P$. For the last coordinate vector $\boldsymbol{e}_r$, Cauchy–Schwarz gives

$$1 = (\boldsymbol{e}_r^\top \boldsymbol{z}_i)^2 \le (\boldsymbol{e}_r^\top \boldsymbol{Z}_P \boldsymbol{e}_r)(\boldsymbol{z}_i^\top \boldsymbol{Z}_P^{-1}\boldsymbol{z}_i) = h\ell_i, \qquad \forall\, i \in P.$$

Thus $1/\ell_i \in [1, h]$ for all $i \in P$.

Set

$$p_i = \frac{\ell_i}{r}, \qquad \boldsymbol{u}_i = \frac{\boldsymbol{Z}_P^{-1/2}\boldsymbol{z}_i}{\sqrt{p_i}}, \qquad \forall\, i \in P.$$

Under a product distribution $\mathbb{P}$, draw $\widetilde{I}_1, \dots, \widetilde{I}_t$ independently from $P$ with

$$\mathbb{P}(\widetilde{I}_j = i) = p_i, \qquad \forall\, i \in P,\ j \in [t],$$

and define

$$\widetilde{\boldsymbol{Y}} = \frac{1}{t}\sum_{j=1}^{t} \boldsymbol{u}_{\widetilde{I}_j}\boldsymbol{u}_{\widetilde{I}_j}^\top.$$

Then

$$\mathbb{E}_{\mathbb{P}}[\boldsymbol{u}_{\widetilde{I}_j}\boldsymbol{u}_{\widetilde{I}_j}^\top] = \boldsymbol{I}_r, \quad \forall\, j \in [t],$$
$$\|\boldsymbol{u}_i\|_2^2 = r, \quad \forall\, i \in P.$$

In particular, $\operatorname{tr}(\widetilde{\boldsymbol{Y}}) = r$ for every realization.

**Step 2.** For $s \in [r]$, write $(t)_s = t(t-1)\cdots(t-s+1)$. For $J \subseteq [r]$ with $|J| = s$, the notation $\widetilde{\boldsymbol{Y}}[J, J]$ denotes the principal submatrix with row and column indices in $J$. Fix $s \in [r]$ and $J = \{j_1, \dots, j_s\} \subseteq [r]$, with $j_1 < \cdots < j_s$. Restrict each sampled vector to the coordinates in $J$ and define

$$\widetilde{\boldsymbol{w}}_k = \big((\boldsymbol{u}_{\widetilde{I}_k})_{j_1}, \dots, (\boldsymbol{u}_{\widetilde{I}_k})_{j_s}\big)^\top \in \mathbb{R}^s, \qquad \forall\, k \in [t].$$

The vectors $\widetilde{\boldsymbol{w}}_1, \dots, \widetilde{\boldsymbol{w}}_t$ are independent. Since $\mathbb{E}_{\mathbb{P}}[\boldsymbol{u}_{\widetilde{I}_k}\boldsymbol{u}_{\widetilde{I}_k}^\top] = \boldsymbol{I}_r$, their second-moment matrices satisfy

$$\mathbb{E}_{\mathbb{P}}[\widetilde{\boldsymbol{w}}_k\widetilde{\boldsymbol{w}}_k^\top] = \boldsymbol{I}_s, \qquad \forall\, k \in [t].$$

Set $\widetilde{\boldsymbol{W}} = [\,\widetilde{\boldsymbol{w}}_1\ \cdots\ \widetilde{\boldsymbol{w}}_t\,] \in \mathbb{R}^{s\times t}$. Then

$$\widetilde{\boldsymbol{Y}}[J, J] = \frac{1}{t}\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{W}}^\top.$$

For $T \subseteq [t]$ with $|T| = s$, let $\widetilde{\boldsymbol{W}}_T$ be the square submatrix consisting of the columns indexed by $T$, ordered increasingly. Cauchy–Binet gives

$$\begin{aligned}\det(\widetilde{\boldsymbol{Y}}[J, J]) &= \frac{1}{t^s}\det(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{W}}^\top) \\ &= \frac{1}{t^s}\sum_{\substack{T\subseteq[t]\\|T|=s}} \det(\widetilde{\boldsymbol{W}}_T)^2.\end{aligned}$$

To compute the expectation of each summand, fix $T$ and write $\widetilde{\boldsymbol{B}} = \widetilde{\boldsymbol{W}}_T$. The columns of $\widetilde{\boldsymbol{B}}$ are independent, and

$$\mathbb{E}_{\mathbb{P}}[\widetilde{B}_{aj}\widetilde{B}_{bj}] = \begin{cases} 1, & a = b, \\ 0, & a \ne b, \end{cases} \qquad \forall\, a, b, j \in [s].$$

Let $\mathfrak{S}_s$ denote the set of permutations of $[s]$, and let $\operatorname{sgn}(\sigma)$ denote the sign of a permutation. Expanding the determinant by columns and using column independence gives

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[\det(\widetilde{\boldsymbol{B}})^2] &= \sum_{\sigma,\tau\in\mathfrak{S}_s} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)\mathbb{E}_{\mathbb{P}}\left[\prod_{j=1}^{s} \widetilde{B}_{\sigma(j),j}\widetilde{B}_{\tau(j),j}\right] \\ &= \sum_{\sigma,\tau\in\mathfrak{S}_s} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)\prod_{j=1}^{s} \mathbb{E}_{\mathbb{P}}\left[\widetilde{B}_{\sigma(j),j}\widetilde{B}_{\tau(j),j}\right] \\ &= s!.\end{aligned}$$

If $\sigma \neq \tau$, at least one factor vanishes. If $\sigma = \tau$, all factors and the sign product equal one. The sum therefore contains exactly $s!$ nonzero terms, each equal to one. This argument uses independence across columns and the second-moment identity.

There are $\binom{t}{s}$ column subsets of cardinality $s$. By linearity of expectation,

$$\mathbb{E}_{\mathbb{P}}[\det(\widetilde{\boldsymbol{Y}}[J,J])] = \frac{1}{t^s}\sum_{\substack{T\subseteq[t]\\|T|=s}} \mathbb{E}_{\mathbb{P}}[\det(\widetilde{\boldsymbol{W}}_T)^2] = \frac{\binom{t}{s}s!}{t^s} = \frac{(t)_s}{t^s}.$$

Since $J$ is arbitrary, the identity holds for every $J \subseteq [r]$ with $|J| = s$. Let $\operatorname{adj}(\boldsymbol{M})$ denote the adjugate matrix, whose entries are the transposed cofactors of $\boldsymbol{M}$. Taking $s = r$ and summing the principal minors of order $r-1$, we have

$$\mathbb{E}_{\mathbb{P}}\left[\det(\widetilde{\boldsymbol{Y}})\right] = \frac{(t)_r}{t^r} > 0, \qquad \mathbb{E}_{\mathbb{P}}\left[\operatorname{tr}(\operatorname{adj}(\widetilde{\boldsymbol{Y}}))\right] = r\frac{(t)_{r-1}}{t^{r-1}}.$$

**Step 3.** Define $\mathbb{Q}$ by assigning each tuple $(i_1,\dots,i_t) \in P^t$ the probability

$$\mathbb{Q}(\widetilde{I}_1 = i_1,\dots,\widetilde{I}_t = i_t) = \frac{\det\left(\frac{1}{t}\sum_{j=1}^{t}\boldsymbol{u}_{i_j}\boldsymbol{u}_{i_j}^{\top}\right)}{\mathbb{E}_{\mathbb{P}}[\det(\widetilde{\boldsymbol{Y}})]}\prod_{j=1}^{t} p_{i_j}.$$

Every matrix in the numerator is positive semidefinite, so the assigned probabilities are nonnegative. By the definition of expectation under the product distribution, the sum of determinants weighted by $\prod_{j=1}^{t} p_{i_j}$ equals $\mathbb{E}_{\mathbb{P}}[\det(\widetilde{\boldsymbol{Y}})]$. The denominator is this same positive sum, so the probabilities sum to one. Thus $\mathbb{Q}$ is a probability distribution. Tuples with determinant zero have probability zero under $\mathbb{Q}$. Since a positive semidefinite matrix has positive determinant if and only if it is positive definite,

$$\mathbb{Q}(\widetilde{\boldsymbol{Y}} \succ \boldsymbol{0}) = 1.$$

On the positive definite realizations,

$$\det(\widetilde{\boldsymbol{Y}})\operatorname{tr}(\widetilde{\boldsymbol{Y}}^{-1}) = \operatorname{tr}(\operatorname{adj}(\widetilde{\boldsymbol{Y}})).$$

The trace of the adjugate is nonnegative for every positive semidefinite realization, including those with rank below $r$. It follows that

$$\mathbb{E}_{\mathbb{Q}}\left[\operatorname{tr}(\widetilde{\boldsymbol{Y}}^{-1})\right] \leq \frac{\mathbb{E}_{\mathbb{P}}\left[\operatorname{tr}(\operatorname{adj}(\widetilde{\boldsymbol{Y}}))\right]}{\mathbb{E}_{\mathbb{P}}\left[\det(\widetilde{\boldsymbol{Y}})\right]} = \frac{rt}{t-r+1}.$$

Under probability $\mathbb{Q}$, at least one positive definite realization has inverse trace no greater than its expectation. Hence there exists a tuple $(i_1, \dots, i_t) \in P^t$ whose associated matrix

$$\boldsymbol{Y} = \frac{1}{t} \sum_{j=1}^{t} \boldsymbol{u}_{i_j} \boldsymbol{u}_{i_j}^\top$$

satisfies

$$\boldsymbol{Y} \succ \boldsymbol{0}, \qquad \operatorname{tr}(\boldsymbol{Y}^{-1}) \leq \mathbb{E}_{\mathbb{Q}}[\operatorname{tr}(\widetilde{\boldsymbol{Y}}^{-1})] \leq \frac{rt}{t - r + 1}.$$

Fix such a tuple. Define its weights and the corresponding auxiliary matrix by

$$\gamma_j = \frac{1}{\ell_{i_j}}, \qquad \forall\, j \in [t], \qquad \boldsymbol{A} = \boldsymbol{Z}_P^{1/2} \boldsymbol{Y} \boldsymbol{Z}_P^{1/2}.$$

Since $1/h \leq \ell_i \leq 1$ for all $i \in P$, the weights satisfy $\gamma_j \in [1, h]$ for all $j \in [t]$.

The matrix $\boldsymbol{A}$ has the required weighted-sum representation. Since $\boldsymbol{u}_{i_j} = \boldsymbol{Z}_P^{-1/2} \boldsymbol{z}_{i_j} / \sqrt{p_{i_j}}$ and $p_{i_j} = \ell_{i_j} / r$, we have

$$\boldsymbol{A} = \frac{1}{t} \sum_{j=1}^{t} \boldsymbol{Z}_P^{1/2} \boldsymbol{u}_{i_j} \boldsymbol{u}_{i_j}^\top \boldsymbol{Z}_P^{1/2} = \frac{1}{t} \sum_{j=1}^{t} \frac{1}{p_{i_j}} \boldsymbol{z}_{i_j} \boldsymbol{z}_{i_j}^\top = \frac{r}{t} \sum_{j=1}^{t} \gamma_j \boldsymbol{z}_{i_j} \boldsymbol{z}_{i_j}^\top.$$

Moreover, $\boldsymbol{A} \succ \boldsymbol{0}$ since $\boldsymbol{Y} \succ \boldsymbol{0}$ and $\boldsymbol{Z}_P^{1/2}$ is invertible.

Cyclic invariance of the trace and $\|\boldsymbol{u}_i\|_2^2 = r$ for all $i \in P$, give

$$\operatorname{tr}(\boldsymbol{Z}_P^{-1} \boldsymbol{A}) = \operatorname{tr}(\boldsymbol{Z}_P^{-1/2} \boldsymbol{A} \boldsymbol{Z}_P^{-1/2}) = \operatorname{tr}(\boldsymbol{Y}) = \frac{1}{t} \sum_{j=1}^{t} \|\boldsymbol{u}_{i_j}\|_2^2 = r.$$

For the inverse-trace bound, the definition of $\boldsymbol{A}$ implies

$$\boldsymbol{A}^{-1} = \boldsymbol{Z}_P^{-1/2} \boldsymbol{Y}^{-1} \boldsymbol{Z}_P^{-1/2}.$$

Another application of cyclic invariance yields

$$\operatorname{tr}(\boldsymbol{A}^{-1} \boldsymbol{Z}_P) = \operatorname{tr}(\boldsymbol{Z}_P^{1/2} \boldsymbol{A}^{-1} \boldsymbol{Z}_P^{1/2}) = \operatorname{tr}(\boldsymbol{Y}^{-1}) \leq \frac{rt}{t - r + 1}.$$

These identities prove the two trace bounds and complete the proof. □

Since the last coordinate of $\boldsymbol{z}_i = (\boldsymbol{v}_i^\top, 1)^\top$ equals one, the preceding Cauchy–Schwarz inequality reduces to $\ell_i \geq 1/h$, $\forall\, i \in P$. Together with $\ell_i \leq 1$, this yields $\gamma_j = 1/\ell_{i_j} \in [1, h]$, $\forall\, j \in [t]$. This weight interval is independent of the coordinate magnitudes. Each weight can therefore be rounded upward to a power of $1 + \delta$ using a grid with $O(1 + \delta^{-1} \log h)$ values. To complete the proof of Theorem 1, we bound the effect of this rounding on the trace product.

*Proof of Theorem 1.* Proposition 1 proves the claim when $v^\star = 0$. Suppose $v^\star > 0$, and let $P \subseteq [n]$ be an optimal solution of the MCD problem (1). Thus $|P| = h$ and $\sqrt[d]{\det(\boldsymbol{C}_P)} = v^\star$. By (2), $\det(\boldsymbol{Z}_P) = h^r (v^\star)^d > 0$. Since $\boldsymbol{Z}_P \succ \boldsymbol{0}$, we can apply Lemma 2 to this subset. Lemma 2 supplies indices $(i_1, \dots, i_t) \in P^t$ and weights $\gamma_j \in [1, h]$, for all $j \in [t]$, such that

$$\boldsymbol{A} = \frac{r}{t} \sum_{j=1}^{t} \gamma_j \boldsymbol{z}_{i_j} \boldsymbol{z}_{i_j}^\top \succ \boldsymbol{0}.$$

Round each weight upward to the smallest admissible grid value and we obtain

$$g_j = \min\{g \in \mathcal{G}_\delta : g \geq \gamma_j\}, \qquad \forall\, j \in [t].$$

This grid value exists since $\gamma_j \leq h$ and the grid extends to at least $h$. Define the auxiliary matrix using these grid weights and the same indices:

$$\widehat{\boldsymbol{A}} = \frac{r}{t} \sum_{j=1}^{t} g_j \boldsymbol{z}_{i_j} \boldsymbol{z}_{i_j}^\top.$$

Since successive grid values differ by the factor $1 + \delta$,

$$\gamma_j \leq g_j \leq (1+\delta)\gamma_j, \qquad \forall\, j \in [t].$$

Each outer product $\boldsymbol{z}_{i_j} \boldsymbol{z}_{i_j}^\top$ is positive semidefinite. Hence,

$$\mathbf{0} \prec \boldsymbol{A} \preceq \widehat{\boldsymbol{A}} \preceq (1+\delta)\boldsymbol{A}.$$

The matrix $\widehat{\boldsymbol{A}}$ belongs to the family enumerated by Algorithm 1, since its indices belong to $[n]$ and its weights belong to $\mathcal{G}_\delta$. Thus $\operatorname{tr}(\boldsymbol{Z}_P^{-1}\widehat{\boldsymbol{A}}) \leq (1+\delta)r$. Inversion reverses the positive definite order, so $\operatorname{tr}(\widehat{\boldsymbol{A}}^{-1}\boldsymbol{Z}_P) \leq rt/(t-r+1)$. Applying Lemma 1 gives

$$\frac{v(S(\widehat{\boldsymbol{A}}))}{v^\star} \leq (1+\delta)\frac{t}{t-r+1} = (1+\delta)\frac{t}{t-d}.$$

The indices and grid weights defining $\widehat{\boldsymbol{A}}$ belong to the enumerated family. The algorithm therefore constructs $\widehat{\boldsymbol{A}}$ and computes $S(\widehat{\boldsymbol{A}})$ by selecting the $h$ smallest costs $q_i(\widehat{\boldsymbol{A}})$. Since it returns the best subset found according to the MCD objective, its output satisfies $v(S_{\text{alg}}) \leq v(S(\widehat{\boldsymbol{A}}))$.

The affine-hull test examines at most $\binom{n}{d}$ subsets of cardinality $d$. For each subset, it tests membership of all $n$ data points in the corresponding affine hull using rational linear algebra. These tests require polynomially many bit operations in $B$. The grid $\mathcal{G}_\delta$ is computed once. Since $\delta$ is rational, each value $(1+\delta)^j$ has encoding length $O((K+1)B)$, for all $j \in \{0, \ldots, K\}$. The enumeration examines $n^t(K+1)^t$ index–weight tuples. For each tuple, the algorithm forms $\boldsymbol{A}$ and tests positive definiteness. If $\boldsymbol{A} \succ \mathbf{0}$, it solves $\boldsymbol{A}\boldsymbol{y}_i = \boldsymbol{z}_i$, for all $i \in [n]$, computes and orders the costs $\boldsymbol{z}_i^\top \boldsymbol{y}_i$, selects $h$ indices, and evaluates the resulting covariance determinant. All matrices and costs used in these computations are rational, with encoding lengths polynomial in $B$, $t$, and $K$. Exact rational linear algebra and comparisons therefore require $\operatorname{poly}(B, t, K)$ bit operations per tuple. This includes the computations defining $S(\boldsymbol{A})$ in Lemma 1. Comparisons of $v(S)$ use $\det(\boldsymbol{C}_S)$, so extracting $d$th roots is unnecessary. Since $B$ is the total input encoding length, the polynomial factors in $n$ and $d$ are also included. The total bit complexity is $[\binom{n}{d} + n^t(K+1)^t]\operatorname{poly}(B, t, K)$, which proves (7). □

### 2.4 Accuracy and implementation

The following parameter choice specifies a desired accuracy for the MCD objective (1).

**Corollary 1** (Desired accuracy)**.** *For rational $\varepsilon \in (0, 1]$, set $t = d + \lceil 2d/\varepsilon \rceil$ and $\delta = \varepsilon/2$. Algorithm 1 returns a feasible subset with*

$$\sqrt[d]{\det(\boldsymbol{C}_{S_{\text{alg}}})} \leq e^{\varepsilon} v^\star.$$

*Its bit complexity is*

$$\left[n(1 + \varepsilon^{-1}\log h)\right]^{O(d/\varepsilon)} \operatorname{poly}(B_\varepsilon),$$

*where $B_\varepsilon$ includes the encoding of $\varepsilon$. The case $v^\star = 0$ is solved exactly.*

*Proof.* The case $v^\star = 0$ follows from the theorem. For $v^\star > 0$, use

$$\log \frac{t}{t-d} \le \frac{d}{t-d} \le \frac{\varepsilon}{2}, \qquad \log(1+\delta) \le \delta = \frac{\varepsilon}{2}$$

in (6). Since $\log(1+\delta) \ge \delta/2$, the grid has $K = O(1 + \varepsilon^{-1} \log h)$. Substitution into (7) proves the complexity bound. □

The approximation guarantee requires checking all index–weight tuples in $[n]^t \times \mathcal{G}_\delta^t$ and choosing the best subset $S(\boldsymbol{A})$ generated by a positive definite auxiliary matrix. To accelerate the solution process, the computational implementation in our numerical study instead samples these tuples under a specified time or trial limit. Since sampling may omit the matrix used in the proof, the sampled feasible solutions for branch-and-bound may not satisfy the theorem's approximation guarantee.

We improve each candidate by the following concentration steps. For a current subset $S$ with $\boldsymbol{C}_S \succ \boldsymbol{0}$, compute the score $\eta_i(S) = (\boldsymbol{v}_i - \bar{\boldsymbol{v}}_S)^\top \boldsymbol{C}_S^{-1} (\boldsymbol{v}_i - \bar{\boldsymbol{v}}_S)$ for each $i \in [n]$. Replace $S$ by the indices of the $h$ smallest scores, resolving ties by index, and recompute the mean and covariance on the new subset. Each update is nonincreasing in $\det(\boldsymbol{C}_S)$ and therefore in $v(S)$ [1]. If $v(S) = 0$, the subset is globally optimal and no further update is needed. Section 6 specifies the sampling and iteration limits and compares the resulting initial solutions for the branch-and-bound method.

# 3 Exact optimization based on a hyperplane arrangement

For fixed $d$, we reduce the worst-case arithmetic bound for exact MCD from $O_d(n^{d(d+3)/2+1})$ to $O_d(n^{d(d+3)/2})$ and remove the general-quadric-position assumption of Bernholt and Fischer [15, Theorem 3.1]. Their algorithm determines the points selected by each candidate quadric by scanning all $n$ observations. Our algorithm orders the computation so that moving between adjacent regions adds or removes one group of identical observations. This updates the subset moment matrix in constant time for fixed $d$, including when observations coincide.

## 3.1 Exact optimization by zonotope enumeration

Let $\boldsymbol{A}_i = \boldsymbol{z}_i \boldsymbol{z}_i^\top$ for each $i \in [n]$, using the augmented data points of Section 2. The cardinality polytope and its image under the moment map are

$$\Delta(n,h) = \left\{ \boldsymbol{x} \in [0,1]^n : \sum_{i=1}^n x_i = h \right\}, \qquad P_h = \left\{ \sum_{i=1}^n x_i \boldsymbol{A}_i : \boldsymbol{x} \in \Delta(n,h) \right\}.$$

Since the vertices of $\Delta(n,h)$ are the indicators of $h$-element subsets, the projected cardinality polytope satisfies

$$P_h = \operatorname{conv}\{\boldsymbol{Z}_S : S \subseteq [n],\ |S| = h\} \subseteq \mathbb{S}_+^r.$$

**Lemma 3.** *Some optimal MCD subset $S^\star$ has $\boldsymbol{Z}_{S^\star}$ at a vertex of $P_h$.*

*Proof.* For $\boldsymbol{M} \in \mathbb{S}_+^r$,

$$\sqrt[r]{\det(\boldsymbol{M})} = \inf_{\substack{\boldsymbol{W} \succ \boldsymbol{0} \\ \det(\boldsymbol{W}) = 1}} \frac{1}{r} \operatorname{tr}(\boldsymbol{W}\boldsymbol{M}). \tag{8}$$

The arithmetic–geometric mean inequality for the eigenvalues of $\boldsymbol{W}^{1/2}\boldsymbol{M}\boldsymbol{W}^{1/2}$ establishes the lower bound on the trace. If $\boldsymbol{M} \succ \boldsymbol{0}$, equality holds at $\boldsymbol{W} = \sqrt[r]{\det(\boldsymbol{M})}\,\boldsymbol{M}^{-1}$. If $\boldsymbol{M}$ is singular, use the minimizing $\boldsymbol{W}$ for $\boldsymbol{M} + \varepsilon \boldsymbol{I}_r$, where $\varepsilon > 0$. Then $\operatorname{tr}(\boldsymbol{W}\boldsymbol{M})/r \leq \sqrt[r]{\det(\boldsymbol{M} + \varepsilon \boldsymbol{I}_r)} \to 0$ as $\varepsilon \downarrow 0$.

According to (8), the determinant root is concave. Its value at any convex combination of vertices of $P_h$ is at least its smallest value at those vertices. Each vertex equals $\boldsymbol{Z}_S$ for some $|S| = h$, and (2) shows that $\sqrt[r]{\det(\boldsymbol{Z}_S)}$ and $v(S)$ have the same minimizing subsets. □

The last coordinate of every $\boldsymbol{z}_i$ equals one. Thus, the bottom-right entry of a subset moment matrix counts the selected observations. To use this fact, first allow the cardinality to vary and form the sum of line segments

$$\mathcal{Z} = \left\{ \sum_{i=1}^{n} x_i \boldsymbol{A}_i : 0 \leq x_i \leq 1,\ \forall\, i \in [n] \right\} = \sum_{i=1}^{n} [\boldsymbol{0}, \boldsymbol{A}_i].$$

This polytope is a zonotope, with generators $\boldsymbol{A}_i$. Let $\mathcal{V} = \operatorname{span}\{\boldsymbol{A}_i : i \in [n]\}$. For $\boldsymbol{M} = \sum_{i=1}^{n} x_i \boldsymbol{A}_i$, we have $M_{rr} = \sum_{i=1}^{n} x_i$. Thus

$$P_h = \mathcal{Z} \cap \{\boldsymbol{M} \in \mathcal{V} : M_{rr} = h\}. \tag{9}$$

Every vertex of this intersection lies at a zonotope vertex or on an edge, as we show below. Consider maximizing $\langle \boldsymbol{W}, \boldsymbol{M} \rangle$ over $\mathcal{Z}$, with $\boldsymbol{W} \in \mathcal{V}$. The coefficient of $x_i$ is $\langle \boldsymbol{W}, \boldsymbol{A}_i \rangle$. Its sign can change only on the hyperplane

$$\{\boldsymbol{W} \in \mathcal{V} : \langle \boldsymbol{W}, \boldsymbol{A}_i \rangle = 0\}, \qquad \forall\, i \in [n]. \tag{10}$$

The complement of these hyperplanes in $\mathcal{V}$ consists of full-dimensional regions, called cells. Within each cell $C$, the coefficients $\langle \boldsymbol{W}, \boldsymbol{A}_i \rangle$ have fixed nonzero signs. Every $\boldsymbol{W} \in C$ therefore exposes the same vertex $\boldsymbol{M}_C$ of $\mathcal{Z}$, obtained by summing the generators with positive coefficients. This correspondence follows the zonotope construction of Onn and Rothblum [19]. Together with (9), it permits enumeration using one hyperplane per distinct generator. A construction based on pairwise cost orderings uses up to $\binom{n}{2}$ hyperplanes. The reduction follows from the fact that linear maximization over $\mathcal{Z}$ requires only the signs of the generator coefficients, not their pairwise ordering.

Two adjacent cells share a face of codimension one, called a facet. On crossing it, the coefficient signs change for one group $J$ of identical generators. Indeed, coincident hyperplanes have parallel generators, and $(\boldsymbol{A}_i)_{rr} = 1$ forces parallel generators to be identical. Write $\boldsymbol{A}_i$ for their common matrix, with $i \in J$, and $\boldsymbol{M}^-$ for the vertex on the side where that group's coefficient is negative. The other endpoint is $\boldsymbol{M}^+ = \boldsymbol{M}^- + |J|\boldsymbol{A}_i$. If $M^-_{rr} \leq h \leq M^+_{rr}$, their edge meets $P_h$ at

$$\boldsymbol{M}_h = \boldsymbol{M}^- + (h - M^-_{rr})\boldsymbol{A}_i. \tag{11}$$

Take the observations selected at $\boldsymbol{M}^-$ and add any $h - M^-_{rr}$ indices from $J$. This gives exactly $h$ observations with moment matrix $\boldsymbol{M}_h$, even when $h$ lies strictly between the two endpoint cardinalities.

Let $m$ be the number of distinct generators. Define

$$\mathcal{L} = \operatorname{span}\{\boldsymbol{A}_i - \boldsymbol{A}_1 : i \in [n]\}, \qquad s = \dim \mathcal{L}.$$

All matrices in $\mathcal{L}$ have bottom-right entry zero, whereas $(\boldsymbol{A}_1)_{rr} = 1$. Hence $\mathcal{V} = \mathcal{L} \oplus \operatorname{span}\{\boldsymbol{A}_1\}$, $\dim \mathcal{V} = s+1$, and $s \leq r(r+1)/2 - 1 = d(d+3)/2$. Choose a basis $\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{s+1}$ from the generators and write $\boldsymbol{W} = \sum_{k=1}^{s+1} w_k \boldsymbol{B}_k$. Positive scaling of $\boldsymbol{W}$ preserves every coefficient sign. We therefore

enumerate cells and shared facets on the $2(s+1)$ affine slices $w_j = \epsilon$, $(j, \epsilon) \in [s+1] \times \{-1, 1\}$. Each slice has $s$ free coordinates.

Algorithm 2 processes one such slice. It stores the best determinant and a cell or edge identifying the candidate. Copying the selected indices at each improvement can take $O(n)$ time. Algorithm 3 recovers those indices once, after the minimum is known.

**Algorithm 2** Candidate moment matrices on one affine slice

**Require:** The cells and shared facets on $w_j = \epsilon$, the generator groups, and $h$.
**Ensure:** The smallest candidate determinant on this slice and its cell or edge, if one exists.

Set $D \leftarrow +\infty$. For a root cell $C$, sum its positive-coefficient generators to obtain $\boldsymbol{M}_C$, counting multiplicities.
**for** each parent–child pair $(C, C')$ of a spanning tree of the cell adjacency graph, parent first **do**
    Let $J$ be the group whose sign changes and choose $i \in J$.
    Set $\boldsymbol{M}_{C'} \leftarrow \boldsymbol{M}_C \pm |J| \boldsymbol{A}_i$, adding if $J$ becomes positive.
**end for**
**for** each cell $C$ with $(\boldsymbol{M}_C)_{rr} = h$ **do**
    If $\det(\boldsymbol{M}_C) < D$, store this determinant in $D$ and record $C$.
**end for**
**for** each shared facet, including those outside the spanning tree **do**
    Order its endpoint matrices by $M^-_{rr} < M^+_{rr}$.
    **if** $M^-_{rr} < h < M^+_{rr}$ **then**
        Compute $\boldsymbol{M}_h$ by (11).
        If $\det(\boldsymbol{M}_h) < D$, store this determinant in $D$ and record the edge.
    **end if**
**end for**
**return** $D$ and the recorded cell or edge if $D < +\infty$.

**Algorithm 3** Exact MCD by zonotope enumeration

**Require:** Data points $\boldsymbol{v}_i \in \mathbb{Q}^d$, $i \in [n]$ and an integer $1 \le h \le n$.
**Ensure:** A subset $S^\star$ of cardinality $h$ attaining $v^\star$ in (1).

Form $\boldsymbol{A}_i = \boldsymbol{z}_i \boldsymbol{z}_i^\top$, for each $i \in [n]$. Group identical matrices and retain the original indices of each group.
**if** there is only one group **then**
    **return** $\{1, \dots, h\}$.
**end if**
Choose a basis of $\mathcal{V}$ from the generators. Set $D_{\text{best}} \leftarrow +\infty$.
**for** each $(j, \epsilon) \in [s+1] \times \{-1, 1\}$ **do**
    Construct the arrangement (10) on $w_j = \epsilon$.
    Apply Algorithm 2. If $D < D_{\text{best}}$, store $D$ in $D_{\text{best}}$ and record the slice and its winning cell or edge.
**end for**
Recover the positive groups of the winning cell, or of the lower endpoint of the winning edge. For an edge, add $h - M^-_{rr}$ indices from its crossing group.
**return** the resulting subset $S^\star$.

**Theorem 2** (Exact optimization by zonotope enumeration)**.** *For fixed $d$, Algorithm 3 returns a subset $S^\star \subseteq [n]$ satisfying*

$$|S^\star| = h, \qquad \sqrt[d]{\det(\boldsymbol{C}_{S^\star})} = v^\star.$$

*It evaluates $O_d(m^s)$ candidate moment matrices and uses $O_d(n \log n + m^s)$ arithmetic operations and comparisons, with constants depending only on $d$. For rational input of total binary encoding length $B$, the bit complexity is $(n \log n + m^s)\,\mathrm{poly}(B)$.*

Theorem 2 improves the worst-case exponent and adapts the complexity bound to the affine dimension of the generators. For $d = 2$, the general arithmetic bound is $O(n^5)$, compared with $O(n^6)$ in Bernholt and Fischer [15, Theorem 3.1] under their general-quadric-position assumption. For observations on the unit circle, the additional equality $(\boldsymbol{A}_i)_{11} + (\boldsymbol{A}_i)_{22} = 1$, $\forall\, i \in [n]$, implies $s \leq 4$ and reduces our bound to $O(n^4)$. This reduction does not require rank-deficient covariances. For distinct observations on the circle and $h \geq 3$, every feasible subset contains three noncollinear points, so its covariance matrix is positive definite.

### 3.2 Proof of Theorem 2

*Proof.* If $s = 0$, all the generators are identical and every feasible subset has the same moment matrix. Assume $s \geq 1$. Next, we split the proof into two steps.

**Step 1. Coverage of the section.** Let $\boldsymbol{M}$ be a vertex of $P_h$ and let $F$ be the smallest face of $\mathcal{Z}$ containing it. The point $\boldsymbol{M}$ lies in the relative interior of $F$. If $\dim F \geq 2$, its direction space contains a nonzero $\boldsymbol{H}$ with $H_{rr} = 0$. For sufficiently small $\delta > 0$, both $\boldsymbol{M} + \delta\boldsymbol{H}$ and $\boldsymbol{M} - \delta\boldsymbol{H}$ belong to $F \cap P_h$, contradicting extremality. Thus $\boldsymbol{M}$ is a zonotope vertex of cardinality $h$ or lies in the interior of a zonotope edge that meets $M_{rr} = h$.

For any objective $\boldsymbol{W}$, the maximizing face of $\mathcal{Z}$ fixes $x_i = 1$ at positive coefficients and $x_i = 0$ at negative coefficients. The zero-coefficient generators may vary freely, so their span is the direction space of that face. A vertex therefore corresponds to a cell with all coefficients nonzero. For an edge, the zero-coefficient generators span one dimension and hence form one group of identical matrices. Perturbing an objective exposing that edge to either side of the group's hyperplane identifies its two endpoint cells. Conversely, an objective in the relative interior of a shared facet exposes precisely the edge joining the adjacent vertices.

Each cell and each shared facet contains a nonzero objective in its relative interior since $s \geq 1$. Choose $j$ with $w_j \neq 0$ and divide the coordinates by $|w_j|$. This preserves all signs, including zeros, and puts the objective on $w_j = 1$ or $w_j = -1$. Thus Algorithm 2 examines every vertex of $P_h$ on at least one slice. Its cell candidates are feasible by their cardinality. Its edge candidates are feasible by (11). Optimality now follows from Lemma 3 and (2).

**Step 2. Arithmetic and bit complexity.** Sorting the upper-triangular entries of the generators lexicographically groups duplicates in $O_d(n \log n)$ operations. A basis of $\mathcal{V}$ can be extracted in $O_d(m)$ operations by testing each distinct matrix for independence against the current basis. Substitution of $w_j = \epsilon$ into (10) gives at most $m$ affine hyperplanes in dimension $s$. An equation with zero left-hand coefficients has a nonzero constant term and can be discarded. For fixed $s \geq 2$, the cells and their facet incidences can be constructed in $O_s(m^s)$ operations, including for degenerate arrangements [29, 30]. For $s = 1$, sorting the intersection points takes $O(m \log m)$ operations and produces $O(m)$ cells and shared facets.

On each slice, the initial sign tests and moment sum take $O_d(m)$ operations. Each tree update, edge intersection, and determinant evaluation or comparison costs $O_d(1)$ since the matrices have order $d + 1$. There are $2(s + 1)$ slices and $O_d(m^s)$ cells and shared facets in total. To recover the winning subset, replay the spanning-tree path on its slice, updating one sign per group crossing, then list the selected indices once. The path costs $O_d(m^s)$ and listing costs $O(n)$. This proves the arithmetic bound.

For rational data, the basis consists of input generators. Arrangement incidence and orientation tests use determinants of fixed order in their rational coefficients. Every candidate matrix is a sum of at most $n$ generators whose entries are products of input coordinates. A common denominator for those entries is the square of the product of all input denominators, which has $O(B)$ bits. Thus

all required signs and determinant comparisons can be computed with poly($B$) bit operations per arithmetic step. □

The worst-case exponent $d(d+3)/2$ still grows quadratically with dimension. Section 4 uses selected cells of the same arrangement to initialize a branch-and-bound search.

# 4 Branch and bound with a hybrid warm start and spectral bounds

We develop a customized branch-and-bound method with hybrid initialization and spectral and completion lower bounds. The warm start combines sampled QFS and concentration steps in Section 2 with a randomized search over zonotope vertices in Section 3.1. The subsequent search branches on the selection or exclusion of data points. The spectral bounds use cardinality-dependent Gershgorin bounds [23], row-wise bounds for quadratic knapsack [24], and spectral moment inequalities [25]. The bounds limit the determinant decrease attainable by the remaining exclusions and the determinant increase required by the remaining inclusions.

## 4.1 Hybrid warm start

We initialize the search using QFS from Section 2 and a random search through the generator hyperplane arrangement in Section 3.1. Both components apply the concentration steps in Section 2 to their candidate subsets. The best feasible subset becomes the initial incumbent.

**Sampled QFS.** Each QFS trial samples an ordered tuple $(i_1, \ldots, i_t) \in [n]^t$ uniformly with replacement and independently samples each weight $g_j$ uniformly from the geometric grid $\mathcal{G}_\delta$ defined in Section 2. Form

$$\boldsymbol{A} = \frac{r}{t} \sum_{j=1}^{t} g_j \boldsymbol{z}_{i_j} \boldsymbol{z}_{i_j}^\top.$$

If $\boldsymbol{A}$ is not positive definite, reject the trial. Otherwise, compute $q_i(\boldsymbol{A}) = \boldsymbol{z}_i^\top \boldsymbol{A}^{-1} \boldsymbol{z}_i$ for every data point and select the $h$ smallest costs, resolving ties by index.

**Random zonotope search.** Random zonotope search traverse selected cells without constructing the full arrangement. Let $S^{\text{best}}$ be the incumbent subset, with $|S^{\text{best}}| = h < n$. If $v(S^{\text{best}}) = 0$, it is globally optimal and the procedure terminates. Otherwise, define

$$q_i = \boldsymbol{z}_i^\top \boldsymbol{Z}_{S^{\text{best}}}^{-1} \boldsymbol{z}_i, \quad \forall\, i \in [n].$$

Order the quadratic costs as $q_{(1)} \le \cdots \le q_{(n)}$, and let

$$\tau = \frac{q_{(h)} + q_{(h+1)}}{2}, \qquad \boldsymbol{W}_0 = \tau \boldsymbol{e}_r \boldsymbol{e}_r^\top - \boldsymbol{Z}_{S^{\text{best}}}^{-1}. \tag{12}$$

Then $\langle \boldsymbol{W}_0, \boldsymbol{A}_i \rangle = \tau - q_i$. When $q_{(h)} < q_{(h+1)}$, the positive inner products identify exactly the $h$ smallest costs, as in Section 3.1. We also evaluate the subset defined by the $h$ smallest costs directly, with ties resolved by index. We use the positive-sign convention of Section 3.1. The objective matrices may lie in $\mathbb{S}^r$: only their orthogonal projections onto set $\mathcal{V}$ affect their inner products with the generators.

Draw a random symmetric direction $\boldsymbol{D}$ and follow the ray

$$\boldsymbol{W}(\alpha) = \boldsymbol{W}_0 + \alpha \boldsymbol{D}, \qquad \alpha \ge 0.$$

Along the ray, each generator coefficient is affine:

$$g_i(\alpha) = \langle \boldsymbol{W}_0, \boldsymbol{A}_i \rangle + \alpha \langle \boldsymbol{D}, \boldsymbol{A}_i \rangle. \tag{13}$$

For each $i \in [n]$ with $\langle \boldsymbol{D}, \boldsymbol{A}_i \rangle \neq 0$, solving $g_i(\alpha) = 0$ gives $\alpha_i = -\frac{\langle \boldsymbol{W}_0, \boldsymbol{A}_i \rangle}{\langle \boldsymbol{D}, \boldsymbol{A}_i \rangle}$. If $\alpha_i > 0$, the ray crosses the hyperplane $\langle \boldsymbol{W}, \boldsymbol{A}_i \rangle = 0$ at $\boldsymbol{W}(\alpha_i)$, and $g_i(\alpha)$ changes sign as $\alpha$ passes through $\alpha_i$. Negative values lie outside the ray, while $\alpha_i = 0$ means that $\boldsymbol{W}_0$ already lies on the hyperplane.

Sort the finite positive crossing parameters. The coefficient signs are constant between consecutive crossings, identifying the cells visited by the ray without constructing the full arrangement. After processing each group of crossings, choose an interior point of the next interval and update

$$S = \{i : g_i(\alpha) > 0\}, \qquad \boldsymbol{M} = \sum_{i \in S} \boldsymbol{A}_i$$

by the corresponding generator additions or deletions. Simultaneous crossings are processed together. A visited cell supplies a candidate when every coefficient is nonzero and $|S| = h$.

Evaluate the sign pattern at an interior point of each open interval, including the intervals before the first crossing and after the last. For each previously unevaluated candidate of cardinality $h$, recompute the covariance determinant and apply concentration steps. A concentration step replaces the current subset with the $h$ data points with the smallest Mahalanobis distances under its mean and covariance. Update $S^{\text{best}}$ whenever a subset evaluated within the time limit improves the incumbent objective. After completing a ray, construct another ray from the latest incumbent and continue until the time allocated to this component is exhausted.

### 4.2 Node bounds

A node is described by $F \subseteq A \subseteq [n]$. The data points in $F$ are fixed for inclusion, those outside $A$ are excluded, and those in $A \setminus F$ are undecided. The node bound depends on these fixed inclusion and exclusion decisions. Assume $|F| \leq h \leq |A|$. Every feasible completion has the form $S = A \setminus R$ for $R \subseteq A \setminus F$ with $|R| = |A| - h$.

If $\det(\boldsymbol{Z}_A) = 0$, then $\boldsymbol{0} \preceq \boldsymbol{Z}_S \preceq \boldsymbol{Z}_A$ implies $\det(\boldsymbol{Z}_S) = 0$ for every completion. Any such subset solves MCD globally. Otherwise, factor $\boldsymbol{Z}_A = \boldsymbol{L}\boldsymbol{L}^\top$ and define

$$\boldsymbol{a}_i = \boldsymbol{L}^{-1}\boldsymbol{z}_i, \quad \ell_i = \|\boldsymbol{a}_i\|_2^2, \quad g_{ij} = \boldsymbol{a}_i^\top \boldsymbol{a}_j, \quad \forall\, i, j \in A \setminus F.$$

The value $\ell_i$ is the leverage of data point $i$ relative to $A$. For an exclusion set $R$, put $\boldsymbol{H}_R = \sum_{i \in R} \boldsymbol{a}_i \boldsymbol{a}_i^\top$. Using $\boldsymbol{z}_i = \boldsymbol{L}\boldsymbol{a}_i$, we obtain

$$\boldsymbol{Z}_{A \setminus R} = \boldsymbol{Z}_A - \sum_{i \in R} \boldsymbol{z}_i \boldsymbol{z}_i^\top = \boldsymbol{L} \left( \boldsymbol{I}_r - \sum_{i \in R} \boldsymbol{a}_i \boldsymbol{a}_i^\top \right) \boldsymbol{L}^\top = \boldsymbol{L}(\boldsymbol{I}_r - \boldsymbol{H}_R)\boldsymbol{L}^\top.$$

Therefore,

$$\log\det(\boldsymbol{Z}_{A \setminus R}) = \log\det(\boldsymbol{Z}_A) + \log\det(\boldsymbol{I}_r - \boldsymbol{H}_R), \qquad \forall\, R \subseteq A \setminus F \text{ with } |R| = q. \tag{14}$$

**Spectral lower bounds.** We seek a lower bound valid for every completion $S$ of the node $(A, F)$. By (14), it suffices to bound the decrease $-\log\det(\boldsymbol{I}_r - \boldsymbol{H}_R)$ uniformly over all admissible exclusion sets.

**Proposition 2** (Spectral lower bounds)**.** *Suppose that for every admissible $R$, we have*

$$\operatorname{tr}(\boldsymbol{H}_R) \le T_1, \qquad \operatorname{tr}(\boldsymbol{H}_R^2) \le T_2, \qquad \lambda_{\max}(\boldsymbol{H}_R) \le \widehat{u}$$

*for some $T_1, T_2 \ge 0$ and $\widehat{u} \in [0, 1)$. Then, every completion $S$ of the node satisfies*

$$\log\det(\boldsymbol{Z}_S) \ge \log\det(\boldsymbol{Z}_A) - T_1 - \psi(\widehat{u})T_2,$$

*where*

$$\psi(t) := \frac{-\log(1-t) - t}{t^2}, \qquad 0 < t < 1, \qquad \psi(0) := \frac{1}{2}.$$

*Proof.* For $0 \le \lambda \le \widehat{u} < 1$, the logarithm series implies

$$-\log(1-\lambda) \le \lambda + \psi(\widehat{u})\lambda^2.$$

Applying this inequality to the eigenvalues of $\boldsymbol{H}_R$ yields

$$-\log\det(\boldsymbol{I}_r - \boldsymbol{H}_R) \le \operatorname{tr}(\boldsymbol{H}_R) + \psi(\widehat{u})\operatorname{tr}(\boldsymbol{H}_R^2) \le T_1 + \psi(\widehat{u})T_2.$$

Substitution into (14) proves the bound. □

The next lemma provides the values of $T_1, T_2$ and $\hat{u}$.

**Lemma 4.** *Let $\operatorname{Top}_q(\boldsymbol{c})$ be the sum of its $q$ largest entries,*

$$T_1 = \operatorname{Top}_q((\ell_i)_{i\in A\setminus F}), \quad b_i = \operatorname{Top}_q((g_{ij}^2)_{j\in A\setminus F}), \quad T_{2,\mathrm{row}} = \operatorname{Top}_q((b_i)_{i\in A\setminus F}), \quad \forall\, i \in A \setminus F.$$

*For a positive scaling vector $\boldsymbol{p} = (p_i)_{i\in A\setminus F}$, define*

$$u(\boldsymbol{p}) = \max_{i\in A\setminus F} \frac{\operatorname{Top}_q((|g_{ij}|p_j)_{j\in A\setminus F})}{p_i}, \quad u_F = 1 - \lambda_{\min}(\boldsymbol{L}^{-1}\boldsymbol{Z}_F(\boldsymbol{L}^{-1})^\top).$$

*Choose a finite family $\mathcal{P}$ of positive scaling vectors containing $\mathbf{1}$, for example, $\mathcal{P} = \{\mathbf{1}\}$. Set*

$$\widehat{u} = \min\left\{1, T_1, \sqrt{T_{2,\mathrm{row}}}, u_F, \min_{\boldsymbol{p}\in\mathcal{P}} u(\boldsymbol{p})\right\}, \quad T_2 = \min\{T_1^2, T_{2,\mathrm{row}}, \widehat{u}T_1\}.$$

*Then for every $R \subseteq A \setminus F$ with $|R| = q$, we have*

$$\operatorname{tr}(\boldsymbol{H}_R) \le T_1, \qquad \operatorname{tr}(\boldsymbol{H}_R^2) \le T_2, \qquad \lambda_{\max}(\boldsymbol{H}_R) \le \widehat{u}.$$

*Proof.* Fix an exclusion set $R$ of cardinality $q$. Then

$$\operatorname{tr}(\boldsymbol{H}_R) = \sum_{i\in R} \ell_i \le T_1, \qquad \operatorname{tr}(\boldsymbol{H}_R^2) = \sum_{i,j\in R} g_{ij}^2 \le \sum_{i\in R} b_i \le T_{2,\mathrm{row}}.$$

The factorization of the moment matrix implies $\mathbf{0} \preceq \boldsymbol{H}_R \preceq \boldsymbol{I}_r$. Its largest eigenvalue is also at most its trace and at most $\sqrt{\operatorname{tr}(\boldsymbol{H}_R^2)}$. These prove the bounds 1, $T_1$, and $\sqrt{T_{2,\mathrm{row}}}$.

The nonzero eigenvalues of $\boldsymbol{H}_R$ equal those of $(g_{ij})_{i,j\in R}$. The matrix $(g_{ij}p_j/p_i)_{i,j\in R}$ is diagonally similar to this Gram matrix. By Gershgorin's circle theorem [31, Section 6.1], we have

$$\lambda_{\max}(\boldsymbol{H}_R) \le \max_{i\in R} \frac{\sum_{j\in R} |g_{ij}|p_j}{p_i} \le u(\boldsymbol{p}), \qquad \forall\, \boldsymbol{p} \in \mathcal{P}.$$

Since $g_{ii} = \ell_i \geq 0$, the absolute row sum includes the diagonal term. Since $S = A \setminus R$ contains $F$,

$$\boldsymbol{I}_r - \boldsymbol{H}_R = \boldsymbol{L}^{-1}\boldsymbol{Z}_S(\boldsymbol{L}^{-1})^\top \succeq \boldsymbol{L}^{-1}\boldsymbol{Z}_F(\boldsymbol{L}^{-1})^\top.$$

Taking smallest eigenvalues proves $\lambda_{\max}(\boldsymbol{H}_R) \leq u_F$. Also $\boldsymbol{0} \preceq \boldsymbol{L}^{-1}\boldsymbol{Z}_F(\boldsymbol{L}^{-1})^\top \preceq \boldsymbol{I}_r$, so $u_F \in [0,1]$. These bounds imply $\lambda_{\max}(\boldsymbol{H}_R) \leq \widehat{u}$. Since the eigenvalues are nonnegative, we have

$$\operatorname{tr}(\boldsymbol{H}_R^2) \leq [\operatorname{tr}(\boldsymbol{H}_R)]^2 \leq T_1^2, \qquad \operatorname{tr}(\boldsymbol{H}_R^2) \leq \widehat{u}\operatorname{tr}(\boldsymbol{H}_R) \leq \widehat{u}T_1,$$

which proves the bound $T_2$. □

Note that when $\widehat{u} = 1$, this spectral construction does not provide a finite lower bound.

**Completion lower bounds.** Every completion $S$ of the node $(A, F)$ adds exactly $m = h - |F|$ data points to $F$. Thus, we have $\boldsymbol{Z}_A \succeq \boldsymbol{Z}_F$. When $\boldsymbol{Z}_F \succ \boldsymbol{0}$, define

$$s_i = \boldsymbol{z}_i^\top \boldsymbol{Z}_F^{-1}\boldsymbol{z}_i, \qquad i \in A \setminus F,$$

and write $s_{(1)} \leq \cdots \leq s_{(|A\setminus F|)}$ for these quadratic costs in non-decreasing order.

**Proposition 3** (Completion lower bounds)**.** *If $\boldsymbol{Z}_F \succ \boldsymbol{0}$, every completion $S$ satisfies*

$$\log\det(\boldsymbol{Z}_S) \geq \log\det(\boldsymbol{Z}_F) + \log\left(1 + \sum_{j=1}^{m} s_{(j)}\right). \tag{15}$$

*Proof.* Define

$$\boldsymbol{Q} = \boldsymbol{Z}_F^{-1/2}(\boldsymbol{Z}_S - \boldsymbol{Z}_F)\boldsymbol{Z}_F^{-1/2} \succeq \boldsymbol{0}.$$

For any positive semidefinite matrix $\boldsymbol{Q}$, $\det(\boldsymbol{I}_r + \boldsymbol{Q}) \geq 1 + \operatorname{tr}(\boldsymbol{Q})$. It follows that

$$\det(\boldsymbol{Z}_S) = \det(\boldsymbol{Z}_F)\det(\boldsymbol{I}_r + \boldsymbol{Q}) \geq \det(\boldsymbol{Z}_F)\left(1 + \sum_{i\in S\setminus F} s_i\right) \geq \det(\boldsymbol{Z}_F)\left(1 + \sum_{j=1}^{m} s_{(j)}\right).$$

Taking logarithms proves the bound. □

When $\boldsymbol{Z}_F$ is singular, the completion bound is not finite. The node lower bound is the maximum of the completion and spectral bounds.

For a nonempty subset $S$, let $\boldsymbol{\Sigma}_S = |S|\boldsymbol{C}_S$. Suppose $\boldsymbol{\Sigma}_F \succ \boldsymbol{0}$, and put $f = |F|$ and $m = h - f \geq 1$. Define

$$B_0 = \frac{\sqrt[d]{\det(\boldsymbol{\Sigma}_F)}}{h}, \qquad \tau_i = (\boldsymbol{v}_i - \bar{\boldsymbol{v}}_F)^\top \boldsymbol{\Sigma}_F^{-1}(\boldsymbol{v}_i - \bar{\boldsymbol{v}}_F), \quad \forall\, i \in A \setminus F.$$

Let $\tau_{(j)}$ denote the $j$th smallest optional value, $j \in [|A \setminus F|]$. Pesch's scalar monotonicity bound [21], Agulló's order-statistic bound [20], and (15), in units of $v(S)$, are

$$B_0, \qquad B_0\sqrt[d]{1 + \frac{f}{f+1}\tau_{(m)}}, \qquad B_0\sqrt[d]{1 + \frac{f}{h}\sum_{j=1}^{m}\tau_{(j)}}.$$

The second and the third bounds do not dominate each other while both dominate the third bound.

**Algorithm 4** Branch-and-bound with hybrid initialization for MCD

**Require:** Data points $\boldsymbol{v}_i \in \mathbb{Q}^d$, $i \in [n]$, cardinality $h$, initialization budget, gap tolerance $\varepsilon_{\text{gap}} \geq 0$, and stopping limits.

**Ensure:** A feasible subset $\widehat{S}$, objective bounds $v_L, v_U$, and termination status.

Start the timer and form the augmented data points.
Solve $h \leq d$, $\det(\boldsymbol{Z}_{[n]}) = 0$, and $n - h \leq 2$ directly when applicable, returning equal objective bounds.
Obtain $\widehat{S}$ by the hybrid warm start within its budget.
If $v(\widehat{S}) = 0$, return it with both objective bounds zero.
Set $U \leftarrow \log\det(\boldsymbol{Z}_{\widehat{S}})$, $L_{\text{closed}} \leftarrow +\infty$, and $\mathcal{Q} \leftarrow \{([n], \varnothing, -\infty)\}$.
**while** $\mathcal{Q} \neq \varnothing$ and the stopping limits permit **do**
    Update $L$ according to newly computed lower bounds.
    If $L$ is finite and $U - L \leq \varepsilon_{\text{gap}}$, stop the search.
    Pop $(A, F, b_{\text{parent}})$. Discard the node if $|F| > h$ or $|A| < h$.
    **if** $b_{\text{parent}} \geq U$ **then**
        $L_{\text{closed}} \leftarrow \min\{L_{\text{closed}}, b_{\text{parent}}\}$. Continue.
    **end if**
    If $\det(\boldsymbol{Z}_A) = 0$, return any completion with both objective bounds zero.
    Set $q \leftarrow |A| - h$ and factor $\boldsymbol{Z}_A$ as needed.
    **if** $q = 0$, $q = |A \setminus F|$, or $q \in \{1, 2\}$ **then**
        Solve the node directly, obtaining $S$ and $w = \log\det(\boldsymbol{Z}_S)$.
        Update $(\widehat{S}, U)$ if $w < U$.
        If $U = -\infty$, return objective bounds zero.
        $L_{\text{closed}} \leftarrow \min\{L_{\text{closed}}, w\}$. Continue.
    **end if**
    Set $b \leftarrow \max\{b_{\text{parent}}, b(A, F)\}$.
    **if** $b \geq U$ **then**
        $L_{\text{closed}} \leftarrow \min\{L_{\text{closed}}, b\}$. Continue.
    **end if**
    Choose $j \in \arg\max_{i \in A \setminus F} \ell_i$, with ties resolved by index.
    Push $(A, F \cup \{j\}, b)$, then $(A \setminus \{j\}, F, b)$.
**end while**
Update $L$ according to newly computed lower bounds.
Return $\widehat{S}$, $v_L = \exp((L - r\log h)/d)$, $v_U = \exp((U - r\log h)/d)$, and termination status.

### 4.3 Branching rules and computational cost

At a node requiring further branching, choose $j \in \arg\max_{i \in A \setminus F} \ell_i$, with ties resolved by index. The children $(A \setminus \{j\}, F)$ and $(A, F \cup \{j\})$ partition the feasible completions according to whether data point $j$ is excluded or selected. We use depth-first search and visit the exclusion child first.

For a fixed number of supplied scaling vectors, a nonterminal node requires $O(|A|r^2+r^3+|A\setminus F|^2 r)$ arithmetic operations and $O(|A \setminus F|^2)$ optional Gram-matrix storage. Each sampled initialization ray has at most $n$ crossings and requires $O(nr^2 + n \log n)$ work to form and order them, in addition to candidate evaluations and concentration steps. The search tree can be exponential. The computational study reports solution times and optimality gaps.

## 5 Comparison with minimum-volume ellipsoids

The minimum-volume ellipsoid (MVE) problem minimizes the volume of an ellipsoid containing at least $h$ data points. We compare its optimal value with $v^\star$ in (1) and obtain a covering-volume guarantee for approximate MCD solutions.

Let vol denote $d$-dimensional volume, and recall

$$\omega_d = \mathrm{vol}\big(\{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\|_2 \le 1\}\big),$$

the volume of the Euclidean unit ball. For a center $\boldsymbol{\mu} \in \mathbb{R}^d$ and a positive definite matrix $\boldsymbol{H} \in \mathbb{S}^d_{++}$, define

$$E(\boldsymbol{\mu}, \boldsymbol{H}) = \left\{\boldsymbol{v} \in \mathbb{R}^d : (\boldsymbol{v}-\boldsymbol{\mu})^\top \boldsymbol{H}(\boldsymbol{v}-\boldsymbol{\mu}) \le 1\right\}.$$

This ellipsoid is the image of the unit ball under $\boldsymbol{u} \mapsto \boldsymbol{\mu} + \boldsymbol{H}^{-1/2}\boldsymbol{u}$, so

$$\mathrm{vol}(E(\boldsymbol{\mu}, \boldsymbol{H})) = \frac{\omega_d}{\sqrt{\det(\boldsymbol{H})}}.$$

The MVE problem jointly optimizes the subset, center, and quadratic-form matrix [32]:

$$V_{\mathrm{MVE}} = \inf_{S, \boldsymbol{\mu}, \boldsymbol{H}} \left\{ \frac{\omega_d}{\sqrt{\det(\boldsymbol{H})}} : (\boldsymbol{v}_i - \boldsymbol{\mu})^\top \boldsymbol{H}(\boldsymbol{v}_i - \boldsymbol{\mu}) \le 1, \forall\, i \in S \subseteq [n], |S| = h, \boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{H} \in \mathbb{S}^d_{++} \right\}. \tag{16}$$

The constraint $|S| = h$ requires the ellipsoid to contain at least $h$ data points. It may contain additional data points. The center $\boldsymbol{\mu}$ is a decision variable and may not be equal $\bar{\boldsymbol{v}}_S$. In MCD, the selected subset determines its mean and covariance. We use an infimum in (16) since, when a feasible subset lies in a proper affine subspace, the volumes of feasible ellipsoids can approach zero without attaining the infimum.

The next theorem bounds the MVE optimal value in terms of the MCD optimal value. The proof constructs a covering ellipsoid from any feasible subset with positive definite covariance.

**Theorem 3** (MCD and minimum-volume ellipsoids)**.** *Let $v^\star$ and $V_{\mathrm{MVE}}$ be the optimal values of MCD Problem* (1) *and MVE problem* (16)*, respectively. For $d+1 \le h \le n$,*

$$\omega_d (d v^\star)^{d/2} \le V_{\mathrm{MVE}} \le \omega_d \big((h-1)v^\star\big)^{d/2}. \tag{17}$$

*Moreover, $V_{\mathrm{MVE}} = 0$ if and only if $v^\star = 0$. When $h = d+1$, both inequalities hold with equality. For $1 \le h \le d$, both optimal values are zero.*

*Proof.* We first prove the lower bound. Let $(S, \boldsymbol{\mu}, \boldsymbol{H})$ be any feasible solution of (16). Expanding the quadratic terms about $\bar{\boldsymbol{v}}_S$ and using $\sum_{i\in S}(\boldsymbol{v}_i - \bar{\boldsymbol{v}}_S) = \boldsymbol{0}$, we obtain

$$\operatorname{tr}(\boldsymbol{H}\boldsymbol{C}_S) = \frac{1}{h}\sum_{i\in S}(\boldsymbol{v}_i - \boldsymbol{\mu})^\top \boldsymbol{H}(\boldsymbol{v}_i - \boldsymbol{\mu}) - (\bar{\boldsymbol{v}}_S - \boldsymbol{\mu})^\top \boldsymbol{H}(\bar{\boldsymbol{v}}_S - \boldsymbol{\mu}) \le 1.$$

The matrix $\boldsymbol{H}^{1/2}\boldsymbol{C}_S\boldsymbol{H}^{1/2}$ is positive semidefinite. Applying the arithmetic–geometric mean inequality to its $d$ nonnegative eigenvalues yields

$$\det(\boldsymbol{H})\det(\boldsymbol{C}_S) = \det(\boldsymbol{H}^{1/2}\boldsymbol{C}_S\boldsymbol{H}^{1/2}) \le \left[\frac{\operatorname{tr}(\boldsymbol{H}^{1/2}\boldsymbol{C}_S\boldsymbol{H}^{1/2})}{d}\right]^d = \left[\frac{\operatorname{tr}(\boldsymbol{H}\boldsymbol{C}_S)}{d}\right]^d \le d^{-d}.$$

Therefore,

$$\operatorname{vol}(E(\boldsymbol{\mu}, \boldsymbol{H})) = \frac{\omega_d}{\sqrt{\det(\boldsymbol{H})}} \ge \omega_d d^{d/2}\sqrt{\det(\boldsymbol{C}_S)} = \omega_d\big(dv(S)\big)^{d/2} \ge \omega_d(dv^\star)^{d/2}.$$

Taking the infimum over all feasible ellipsoids proves the lower bound.

To construct a covering ellipsoid, fix any $S = \{i_1, \ldots, i_h\} \subseteq [n]$ with $|S| = h$ and $\boldsymbol{C}_S \succ \boldsymbol{0}$. Let $\boldsymbol{Y} \in \mathbb{R}^{d\times h}$ have $j$th column $(\boldsymbol{v}_{i_j} - \bar{\boldsymbol{v}}_S)/\sqrt{h}$, for each $j \in [h]$. Then

$$\boldsymbol{Y}\boldsymbol{Y}^\top = \boldsymbol{C}_S, \qquad \boldsymbol{Y}\boldsymbol{1} = \boldsymbol{0}.$$

Define $\boldsymbol{P} = \boldsymbol{Y}^\top\boldsymbol{C}_S^{-1}\boldsymbol{Y}$. The identities $\boldsymbol{P}^\top = \boldsymbol{P}$, $\boldsymbol{P}^2 = \boldsymbol{P}$, and $\boldsymbol{P}\boldsymbol{1} = \boldsymbol{0}$ show that $\boldsymbol{P}$ is an orthogonal projection onto a subspace of $\{\boldsymbol{x} \in \mathbb{R}^h : \boldsymbol{1}^\top\boldsymbol{x} = 0\}$. The projection onto this latter space is $\boldsymbol{I}_h - \boldsymbol{1}\boldsymbol{1}^\top/h$. Hence

$$\boldsymbol{0} \preceq \boldsymbol{P} \preceq \boldsymbol{I}_h - \frac{1}{h}\boldsymbol{1}\boldsymbol{1}^\top, \qquad P_{jj} \le 1 - \frac{1}{h}, \quad \forall\, j \in [h].$$

By the definition of the columns of $\boldsymbol{Y}$,

$$(\boldsymbol{v}_{i_j} - \bar{\boldsymbol{v}}_S)^\top\boldsymbol{C}_S^{-1}(\boldsymbol{v}_{i_j} - \bar{\boldsymbol{v}}_S) = hP_{jj} \le h - 1, \qquad \forall\, j \in [h].$$

Thus $E(\bar{\boldsymbol{v}}_S, \boldsymbol{C}_S^{-1}/(h-1))$ contains every data point indexed by $S$, and its volume is

$$\operatorname{vol}\big(E(\bar{\boldsymbol{v}}_S, \boldsymbol{C}_S^{-1}/(h-1))\big) = \omega_d(h-1)^{d/2}\sqrt{\det(\boldsymbol{C}_S)} = \omega_d\big((h-1)v(S)\big)^{d/2}.$$

If $v^\star > 0$, choose $S$ to be an optimal solution of (1), so that $v(S) = v^\star$. This proves the upper bound.

Suppose now that $v^\star = 0$, and choose a feasible subset $S$ with $\det(\boldsymbol{C}_S) = 0$. There is a unit vector $\boldsymbol{a} \in \mathbb{R}^d$ satisfying $\boldsymbol{C}_S\boldsymbol{a} = \boldsymbol{0}$. Since

$$0 = \boldsymbol{a}^\top\boldsymbol{C}_S\boldsymbol{a} = \frac{1}{h}\sum_{i\in S}\big[\boldsymbol{a}^\top(\boldsymbol{v}_i - \bar{\boldsymbol{v}}_S)\big]^2,$$

we have $\boldsymbol{a}^\top(\boldsymbol{v}_i - \bar{\boldsymbol{v}}_S) = 0$, $\forall\, i \in S$. Choose $R > 0$ with $R > \max_{i\in S}\|\boldsymbol{v}_i - \bar{\boldsymbol{v}}_S\|_2$. For each $\tau > 0$, define

$$\boldsymbol{H}_\tau = \frac{1}{R^2}(\boldsymbol{I}_d + \tau\boldsymbol{a}\boldsymbol{a}^\top) \succ \boldsymbol{0}.$$

Every selected data point satisfies

$$(\boldsymbol{v}_i - \bar{\boldsymbol{v}}_S)^\top\boldsymbol{H}_\tau(\boldsymbol{v}_i - \bar{\boldsymbol{v}}_S) = \frac{\|\boldsymbol{v}_i - \bar{\boldsymbol{v}}_S\|_2^2}{R^2} \le 1, \qquad \forall\, i \in S.$$

Thus $E(\bar{\boldsymbol{v}}_S, \boldsymbol{H}_\tau)$ is feasible for every $\tau > 0$. Since $\|\boldsymbol{a}\|_2 = 1$, the eigenvalues of $\boldsymbol{I}_d + \tau \boldsymbol{a}\boldsymbol{a}^\top$ are $1+\tau$ and $d-1$ copies of 1. Its volume therefore satisfies

$$\mathrm{vol}\big(E(\bar{\boldsymbol{v}}_S, \boldsymbol{H}_\tau)\big) = \frac{\omega_d R^d}{\sqrt{1+\tau}} \longrightarrow 0 \qquad \text{as } \tau \to \infty.$$

This proves $V_{\mathrm{MVE}} = 0$. The lower bound proves the converse. If $h \le d$, every feasible covariance has rank at most $h - 1 < d$, so $v^\star = 0$ and the same construction establishes $V_{\mathrm{MVE}} = 0$. Finally, when $h = d + 1$, the two bounds in (17) coincide. □

The covering construction converts an approximate MCD solution into a feasible MVE solution. Specifically, suppose $v^\star > 0$ and a subset $S$ of cardinality $h$ satisfies $v(S) \le \rho v^\star$, where $\rho \ge 1$. Combining its covering volume with the lower bound in (17) proves

$$\frac{\mathrm{vol}\big(E(\bar{\boldsymbol{v}}_S, \boldsymbol{C}_S^{-1}/(h-1))\big)}{V_{\mathrm{MVE}}} \le \left(\rho \frac{h-1}{d}\right)^{d/2}.$$

The factor $\rho^{d/2}$ converts the approximation ratio for $v(S)$ into a volume ratio. The remaining factor accounts for covering all selected data points with an ellipsoid defined by their sample mean and covariance. The dependence on $h$ in (17) is necessary. Suppose that $h = n \ge 2d$, and consider the data points $\{\pm \boldsymbol{e}_j\}_{j=1}^d$ together with $h - 2d$ copies of the origin. There is only one feasible subset, with mean zero and covariance $(2/h)\boldsymbol{I}_d$, so $v^\star = 2/h$. The unit ball contains all data points. Conversely, if $E(\boldsymbol{\mu}, \boldsymbol{H})$ contains them, averaging the containment inequalities for $\boldsymbol{e}_j$ and $-\boldsymbol{e}_j$ yields

$$H_{jj} + \boldsymbol{\mu}^\top \boldsymbol{H} \boldsymbol{\mu} \le 1, \qquad \forall\, j \in [d].$$

Summing over $j$ yields $\mathrm{tr}(\boldsymbol{H}) \le d$. The arithmetic–geometric mean inequality then gives

$$\det(\boldsymbol{H}) \le \left(\frac{\mathrm{tr}(\boldsymbol{H})}{d}\right)^d \le 1.$$

Every feasible ellipsoid therefore has volume at least $\omega_d$, and the unit ball attains this value. It follows that

$$\frac{V_{\mathrm{MVE}}}{\omega_d (v^\star)^{d/2}} = \left(\frac{h}{2}\right)^{d/2}.$$

For fixed $d$, this has the same order in $h$ as the upper bound in (17). At $h = 2d$, the example also attains the lower bound.

The volume bounds concern ellipsoids covering the data points in (16). In split-conformal prediction, a separate calibration sample determines the radius. The prediction-region volume therefore depends on both the selected covariance and the calibrated radius. The computational study evaluates that volume together with empirical coverage.

# 6 Computational experiments

We compare the MCD subsets returned by branch-and-bound (B&B), FastMCD [1], and a medoid-distance heuristic, which selects the $h$ points nearest to the sample medoid. We then measure how the node lower bound affects search completion and how the initialization affects the best feasible determinant. The prediction experiments evaluate the calibrated volume and empirical coverage of regions fitted to multivariate regression residuals.

### 6.1 Algorithms and computational setting

The optimization tables report ratios of covariance determinants. For two feasible subsets $S_1$ and $S_2$ with positive determinants, the corresponding ratio for the MCD objective is

$$R_{\det} = \frac{\det(\boldsymbol{C}_{S_1})}{\det(\boldsymbol{C}_{S_2})}, \qquad R_v = \frac{v(S_1)}{v(S_2)} = \sqrt[d]{R_{\det}}.$$

For instances of a common dimension $d$, the $d$th root of the geometric-mean determinant ratio gives the geometric-mean objective ratio. When dimensions vary, we take the root for each instance using its own dimension and then compute the geometric mean. A determinant ratio below one favors the method in the numerator.

In the comparison tables, boldface marks the better result for the stated metric. We compare completion counts within each instance group and strict wins within each pair of methods. We bold ties for the best completion count, and the tie count when all paired cases agree. A bold ratio at or below one indicates that the numerator method has the smaller or equal geometric mean. For empirical coverage, boldface marks the estimate closest to the nominal level. Ordinary multistart B&B initializes the search from multiple feasible subsets. Hybrid B&B uses the sampled quadratic-form selection (QFS) and zonotope searches of Section 4.1. The experiments use one thread on the same ARM64 laptop under macOS 15.6.1, Python 3.13.5, and NumPy 2.1.3.

### 6.2 Comparison with FastMCD and medoid-distance selection

We tested 60 instances with $n = 18$, $d = 8$, and $h = 13$ against the optimum obtained by exhaustive enumeration. Within the five-second limit, B&B attained the optimum on all 60 instances. FastMCD attained it on 21 and the medoid-distance heuristic on 17. Table 1 reports the resulting determinant ratios to the enumerated optimum.

Table 1: Heuristic and B&B solutions on the same 60 eight-dimensional instances. The ratio is the geometric mean of the method's determinant divided by the enumerated optimum.

| Method | Optima attained | Determinant ratio |
|---|---:|---:|
| B&B | **60/60** | **1.000** |
| FastMCD | 21/60 | 1.779 |
| Medoid distance | 17/60 | 9.884 |

### 6.3 Comparison of lower bounds

We evaluated three lower bounds at the same 640 nodes in the branch-and-bound tree from 120 instances with $n = 18$, $h = 13$, and $d \in \{2, 8\}$. The combined bound takes the larger of the spectral and completion bounds in Section 4. It exceeds Pesch's determinant-monotonicity bound [21] at 496 nodes and ties it at the other 144. Against Agulló's order-statistic bound [20], it wins at 403 nodes and loses at seven (Table 2). Taking the maximum therefore improves on the combined bound at seven of the 640 nodes.

We also used each bound as the pruning rule in the same B&B solver on 49 archived instances. The suite contains 13 small instances, 24 instances with $n = 64$, $d = 8$, and $h \in \{32, 48\}$, and 12 larger Gaussian or principal-component instances. With the reported 60-second budget, all four rules completed all 13 small searches within numerical tolerances. The combined bound and the maximum of all three bounds each completed 15 medium and 10 large searches. Agulló's bound completed one medium search, and Pesch's completed none of the medium or large searches (Table 3).

Table 2: Pairwise comparisons of node lower bounds on 640 fixed nodes. A strict win means that the method computes the larger lower bound at that node.

| Benchmark | Combined wins | Ties | Benchmark wins |
|---|---|---|---|
| Pesch | **496** | 144 | 0 |
| Agulló | **403** | 230 | 7 |

Table 3: Numerically completed B&B searches within each size group. The last column sums the three groups. Maximum uses the largest of Pesch's, Agulló's, and the combined bound at each node.

| Pruning bound | Small | Medium | Large | Total |
|---|---|---|---|---|
| Pesch | **13/13** | 0/24 | 0/12 | 13/49 |
| Agulló | **13/13** | 1/24 | 0/12 | 14/49 |
| Spectral and completion | **13/13** | **15/24** | **10/12** | **38/49** |
| Maximum | **13/13** | **15/24** | **10/12** | **38/49** |

### 6.4 Hybrid initialization

We allowed random starts, sampled QFS, random zonotope searches, and the hybrid initialization for two seconds each. For each of two Gaussian-mixture families, we generated three independent data matrices at each of $(n, d) = (96, 4)$ and $(128, 8)$. On each of the 12 matrices we used three retention fractions, $h/n \approx 0.4$, 0.6, and 0.8, rounding $h$ to an integer. Thus, there are 36 cases. Each method ran twice per case.

Each QFS trial draws $2d$ observation indices with replacement and weights from the geometric grid with $\delta = 1/2$. When the resulting auxiliary matrix is positive definite, QFS selects the $h$ observations with the smallest quadratic scores. A random zonotope search proposes feasible subsets as it crosses generator hyperplanes in Section 3.1. After the common starts, the hybrid splits the remaining time equally between QFS and zonotope searches. All four procedures refine each candidate by at most 40 concentration steps. We average the two log determinants within each method and case before making paired comparisons.

Table 4: Initialization at two seconds on 36 paired cases. The ratio is the geometric mean of the hybrid determinant divided by the benchmark determinant.

| Benchmark | Hybrid wins | Ties | Benchmark wins | Determinant ratio |
|---|---|---|---|---|
| Random starts | **17** | 19 | 0 | **0.978712** |
| QFS | **4** | 32 | 0 | **0.998657** |
| Zonotope searches | **4** | 32 | 0 | **0.994468** |

The hybrid achieved a smaller determinant than random starts on 17 of 36 cases and tied on the remaining 19. It improved on QFS alone in four cases and on zonotope searches alone in four, with no losses in either comparison. The geometric-mean reductions were 2.129% against random starts, 0.134% against QFS, and 0.553% against zonotope searches (Table 4). Figure 1 tracks candidate quality during initialization.

A separate eight-second comparison measures the effect of adding sampled QFS proposals to ordinary multistart before the same B&B search. The eight seconds include two seconds of initialization. At $h/n \approx 0.4$, the QFS warm start improved the final determinant in eight of 12 cases

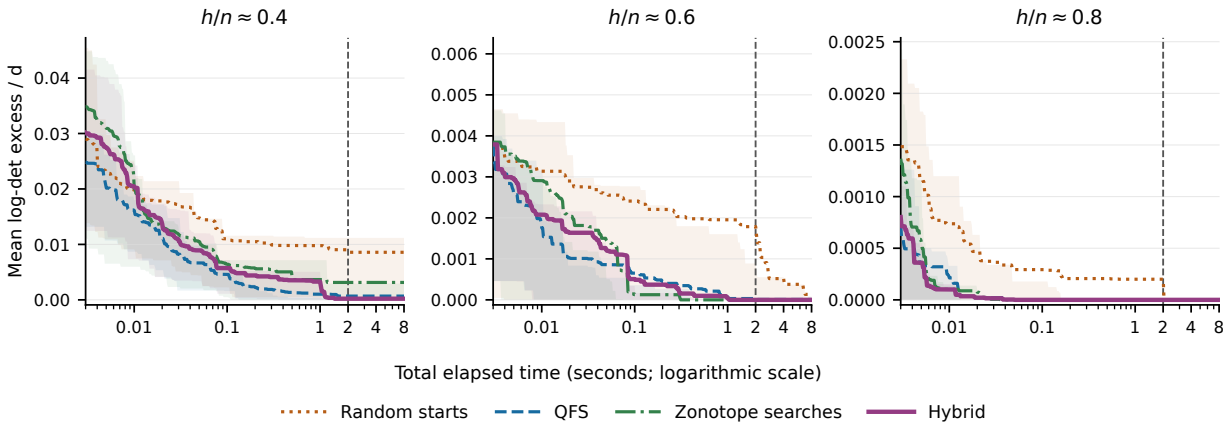



Figure 1: Candidate quality over the two-second initialization. For each case, the vertical axis is the difference between the current log determinant and the smallest eligible endpoint log determinant across the four methods and two repeats, divided by $d$. Shading shows the interquartile range across cases.

and tied in four. The geometric-mean reduction was 4.6%, and the largest case-level reduction was 19.8%. At $h/n \approx 0.6$, the geometric-mean reduction was 0.09%. At $h/n \approx 0.8$, all 12 cases tied. Before averaging the two repeats within each case, the QFS warm start won 22 of the 72 paired runs and tied in 50 (Table 5).

Table 5: Final incumbent determinants at eight seconds, including initialization. Each row contains 12 paired cases. The ratio compares the QFS warm start with ordinary multistart after averaging the two log determinants for each method and case.

| Target $h/n$ | QFS warm wins | Ties | Multistart wins | Determinant ratio |
|---|---|---|---|---|
| 40% | **8** | 4 | 0 | **0.9541** |
| 60% | **3** | 9 | 0 | **0.9991** |
| 80% | 0 | **12** | 0 | **1.0000** |

## 6.5 Larger instances

Table 6 reports a small Gaussian reference and seven larger instances. For Breast Cancer and Digits, we standardized the original coordinates and projected them onto the first $d$ principal components, using $(d, k) = (10, 10)$ and $(20, 20)$ for each dataset, where $k = n - h$. Breast Cancer uses all 569 observations and Digits uses a fixed sample of 1,000. The other large instances are two clean Gaussian samples and a 20-dimensional sample whose last 20 observations receive independent Gaussian perturbations.

Hybrid B&B closed the numerical log gap on all eight reported instances within the 1,800-second total limit. The seven larger searches took 2.11–130.38 seconds including initialization. The longest was the clean 20-dimensional Gaussian instance. These times describe individual runs.

Table 6: Hybrid B&B running times, including initialization, and final log gaps. The log gap is the log of the ratio of the upper to the lower determinant bound. C denotes numerical completion.

| Instance | $n$ | $d$ | $h$ | Time (s) | Status | Log gap |
|---|---|---|---|---|---|---|
| Small Gaussian | 12 | 3 | 9 | 2.00 | C | 0 |
| Breast Cancer | 569 | 10 | 559 | 2.11 | C | 0 |
| Breast Cancer | 569 | 20 | 549 | 9.98 | C | 0 |
| Digits | 1000 | 10 | 990 | 2.53 | C | 0 |
| Digits | 1000 | 20 | 980 | 23.14 | C | 0 |
| Gaussian | 1000 | 10 | 990 | 3.45 | C | 0 |
| Gaussian | 1000 | 20 | 980 | 130.38 | C | 0 |
| Perturbed Gaussian | 1000 | 20 | 980 | 35.53 | C | 0 |

### 6.6 Conformal prediction application

The prediction study uses 440 residual samples and six subset cardinalities per sample, for a total of 2,640 MCD fits. We reused fits with a closed numerical gap and ran hybrid B&B on the remaining instances, with a 900-second limit per fit including initialization.

**Comparison methods.** At fixed $h$, we compared B&B and raw FastMCD on the determinant of the covariance matrix of the selected residuals. Both methods evaluated the covariance on $h$ residuals with equal weights. The untrimmed estimator used all 64 residuals. For prediction, we also compared with the minimum-volume covering sets (MVCS) method of Braun et al. [28]. MVCS fits a center, a linear transformation, and a shape exponent to minimize empirical region volume at a target coverage level. We calibrated its fitted score on the same separate sample used for the MCD score.

**Data and regression functions.** Let $\boldsymbol{\xi}$ denote the explanatory variables, $\boldsymbol{y} \in \mathbb{R}^d$ the response, and $\boldsymbol{f}(\boldsymbol{\xi})$ the regression function. For the Gaussian instances, $d = 8$, the residual mean is zero, the covariance satisfies $(\boldsymbol{C}_0)_{ij} = 0.65^{|i-j|}$, $\forall\, i, j \in [d]$, and the regression function is identically zero. The Seoul weather responses are next-day maximum and minimum temperatures [33]. The River RF1 responses are eight river-flow measurements 48 hours ahead [34]. For each dataset, we fit 400 extremely randomized regression trees with at least five observations per terminal node [35]. We fit the regression function once and hold it fixed across methods and replications. We impute missing predictor values with regression-sample coordinate medians. Let $\boldsymbol{D}$ be the diagonal matrix of residual standard deviations from that sample. The standardized residual is $\boldsymbol{r}(\boldsymbol{\xi}, \boldsymbol{y}) = \boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\xi}))$, with $\boldsymbol{D} = \boldsymbol{I}_d$ for the Gaussian instances.

Table 7 separates the samples used to estimate the regression, fit covariance matrices, select $h$, calibrate the final score, and evaluate the regions. For weather, we train the regression on 2013–2015 data, group the 2016 station records by date, and evaluate on 2017 data. RF1 uses a fixed random partition.

**Contamination experiments.** Let $\boldsymbol{\mu}_{\text{ref}}$ and $\boldsymbol{C}_{\text{ref}}$ denote the reference mean and covariance. For Gaussian data, they are $\mathbf{0}$ and $\boldsymbol{C}_0$. For real data, they are the empirical mean and covariance of the residual pool before artificial contamination. Let $c$ be the number of replaced residual vectors, possibly distinct from the exclusion count $k = n - h$. Each instance uses $n = 64$ residual vectors, of which $c \in \{0, 1, 2, 3, 6, 13\}$ are replaced. Let $J \subseteq [n]$ be the set of replaced indices, with

Table 7: Sample sizes for each stage of prediction. Each covariance fit uses 64 residuals drawn from the indicated pool. The Gaussian regression function is known.

| Dataset | $d$ | Regression | Covariance | Validation | Calibration | Evaluation |
|---|---|---|---|---|---|---|
| Gaussian | 8 | – | 512 | 512 | 1,024 | 4,096 |
| Seoul weather | 2 | 4,638 | 500 | 250 | 547 | 1,538 |
| River RF1 | 8 | 4,106 | 1,369 | 456 | 913 | 1,825 |

$|J| = c$. For each $i \in J$, the scattered replacement is $\boldsymbol{\mu}_{\text{ref}} + 5\boldsymbol{L}\boldsymbol{g}_i$, and the clustered replacement is $\boldsymbol{\mu}_{\text{ref}} + \boldsymbol{L}(8\boldsymbol{u} + 0.5\boldsymbol{g}_i)$, where $\boldsymbol{g}_i \sim N(\boldsymbol{0}, \boldsymbol{I}_d)$, $\forall\, i \in J$, $\boldsymbol{L}\boldsymbol{L}^\top = \boldsymbol{C}_{\text{ref}}$, and $\boldsymbol{u}$ is a random unit vector shared by the cluster. Within each instance, all methods receive the same replacements in the covariance sample. The other samples retain their original observations. We use 20 Gaussian replications and ten replications per real dataset for each condition, sharing the uncontaminated case across the two mechanisms. Real-data replications use the same regression function and the same validation, calibration, and evaluation samples.

**Prediction regions and negative log-likelihood.** For a selected residual subset $S$ with mean $\bar{\boldsymbol{r}}_S$ and positive definite covariance $\boldsymbol{C}_S$, the nonconformity score is the Mahalanobis distance

$$s_S(\boldsymbol{\xi}, \boldsymbol{y}) = \sqrt{(\boldsymbol{r}(\boldsymbol{\xi}, \boldsymbol{y}) - \bar{\boldsymbol{r}}_S)^\top \boldsymbol{C}_S^{-1} (\boldsymbol{r}(\boldsymbol{\xi}, \boldsymbol{y}) - \bar{\boldsymbol{r}}_S)}.$$

For nominal coverage $1 - \alpha$, $\alpha \in (0, 1)$, let $m_{\text{cal}}$ be the number of calibration scores. The radius $\widehat{t}_S$ is their $\lceil (m_{\text{cal}} + 1)(1 - \alpha) \rceil$th order statistic after appending $+\infty$. The region contains the responses with score at most $\widehat{t}_S$. In the original response coordinates, its volume is

$$V_S = \omega_d |\det(\boldsymbol{D})| v(S)^{d/2}\, \widehat{t}_S^{\,d}. \tag{18}$$

Empirical coverage is the fraction of evaluation responses contained in their prediction regions. The split-conformal guarantee requires exchangeability of the calibration and future nonconformity scores [9]. For real data, we report observed coverage.

For density estimation, the Gaussian negative log-likelihood at $(\boldsymbol{\xi}, \boldsymbol{y})$ is

$$\ell_S(\boldsymbol{\xi}, \boldsymbol{y}) = \frac{d}{2} \log(2\pi) + \log |\det(\boldsymbol{D})| + \frac{d}{2} \log v(S) + \frac{1}{2} s_S(\boldsymbol{\xi}, \boldsymbol{y})^2.$$

We averaged this loss over the validation sample when choosing $h$ for density estimation.

**Choice of subset cardinality.** The fixed rule used $h = 64 - c$ and thus used the known replacement count. The validation rule fit each $h \in \mathcal{H} = \{51, 58, 61, 62, 63, 64\}$ on the same residual sample. For density estimation, we chose the fit with the smallest mean validation negative log-likelihood. For prediction regions, we computed a provisional radius on the validation sample and chose the $h$ with the smallest provisional volume separately at each nominal level. We then computed the final radius on the calibration sample. The 900-second limit applies to each MCD fit. Selecting $h$ from the grid required all six fits and the validation calculations.

**Summary statistics.** We report geometric means of paired volume ratios. Each pointwise 95% interval uses 2,000 paired bootstrap resamples with common replication indices for the two methods. For full-sample real-data negative log-likelihood, we also resampled evaluation blocks of seven weather dates or 32 original RF1 rows. The real-data intervals condition on the fitted regression and data partition. We selected the displayed comparisons after examining the results.

**Equal-cardinality comparison on prediction instances.** Across the ten conditions with $c > 0$, B&B and FastMCD selected the same number of residuals, $h = 64 - c$. There are 200 Gaussian, 100 weather, and 100 RF1 comparisons. For each original sample of residuals, we construct ten contaminated instances.

Table 8: Fixed-cardinality B&B versus raw FastMCD. A strict win means a smaller determinant. The last column is the geometric-mean determinant ratio, B&B divided by FastMCD. Absolute log-determinant differences of at most $10^{-8}$ count as ties.

| Dataset | B&B wins | Ties | FastMCD wins | Determinant ratio |
|---|---|---|---|---|
| Gaussian | **33** | 167 | 0 | **0.969** |
| Weather | **2** | 98 | 0 | **0.992** |
| River RF1 | **26** | 74 | 0 | **0.864** |

B&B attained a smaller determinant on 61 of the 400 paired instances and tied on the other 339 (Table 8). The largest geometric-mean reduction was 13.6% on RF1, compared with 3.1% on Gaussian data and 0.8% on weather data.

**Calibrated region volume and coverage.** The validation-selected MCD ellipsoid and the MVCS region used the same regression function, residual samples, and nominal coverage level. Both methods calibrated the chosen score on the separate calibration sample and measure coverage on the same evaluation responses.

With six scattered Gaussian replacements and a 99% target, the validation-selected MCD region had 77.5% lower geometric-mean volume than MVCS. Empirical coverage was 99.12% for MCD and 99.08% for MVCS. Figure 2 also illustrates the clustered replacements.

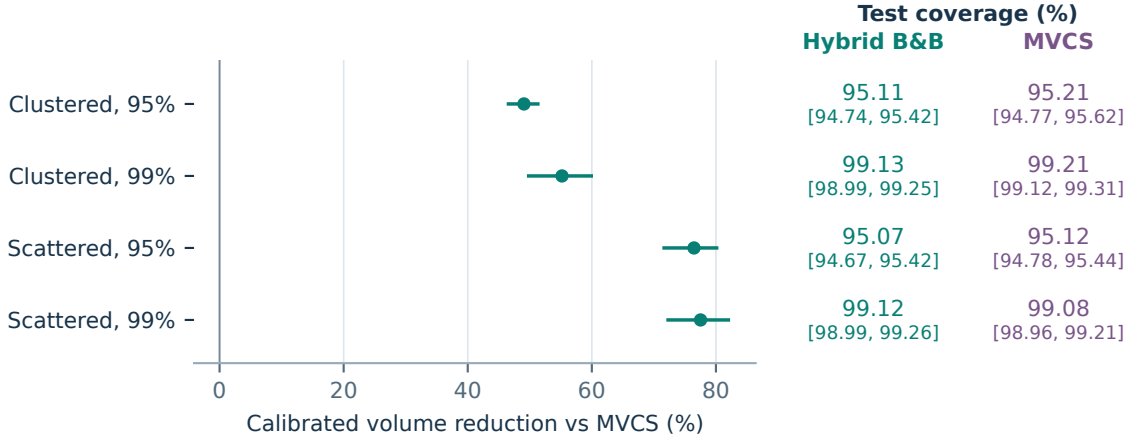



Figure 2: Gaussian prediction regions with six replacements among the 64 fitting residuals. At each target, B&B used validation volume to select $h$. The left panel shows geometric-mean volume reductions relative to MVCS with paired 95% bootstrap intervals. The right panel shows empirical coverage with pointwise intervals. The plot labels the MCD fit Hybrid B&B. Both methods used the same replacements across 20 replications.

The validation-selected MCD regions had smaller geometric-mean volumes in all six real-data comparisons in Table 9. For RF1 at the 90% target with six clustered replacements, the MCD

Table 9: Validation-selected MCD versus MVCS on real data. The ratio is MCD-calibrated volume divided by MVCS-calibrated volume, with a paired 95% bootstrap interval. For coverage, boldface marks the estimate closer to the nominal target.

| Data and replacements | Target | Volume ratio | 95% interval for ratio | MCD coverage (%) | MVCS coverage (%) |
|---|---|---|---|---|---|
| Weather, 13 clustered | 90% | **0.828** | [0.768, 0.887] | 85.76 | **86.55** |
| | 95% | **0.888** | [0.772, 0.988] | **95.42** | 94.30 |
| RF1, 6 scattered | 90% | **0.370** | [0.275, 0.504] | **90.04** | 89.87 |
| | 95% | **0.421** | [0.320, 0.567] | **94.54** | 93.93 |
| RF1, 6 clustered | 90% | **0.364** | [0.262, 0.508] | **90.00** | 89.62 |
| | 95% | **0.446** | [0.327, 0.615] | **94.48** | 94.00 |

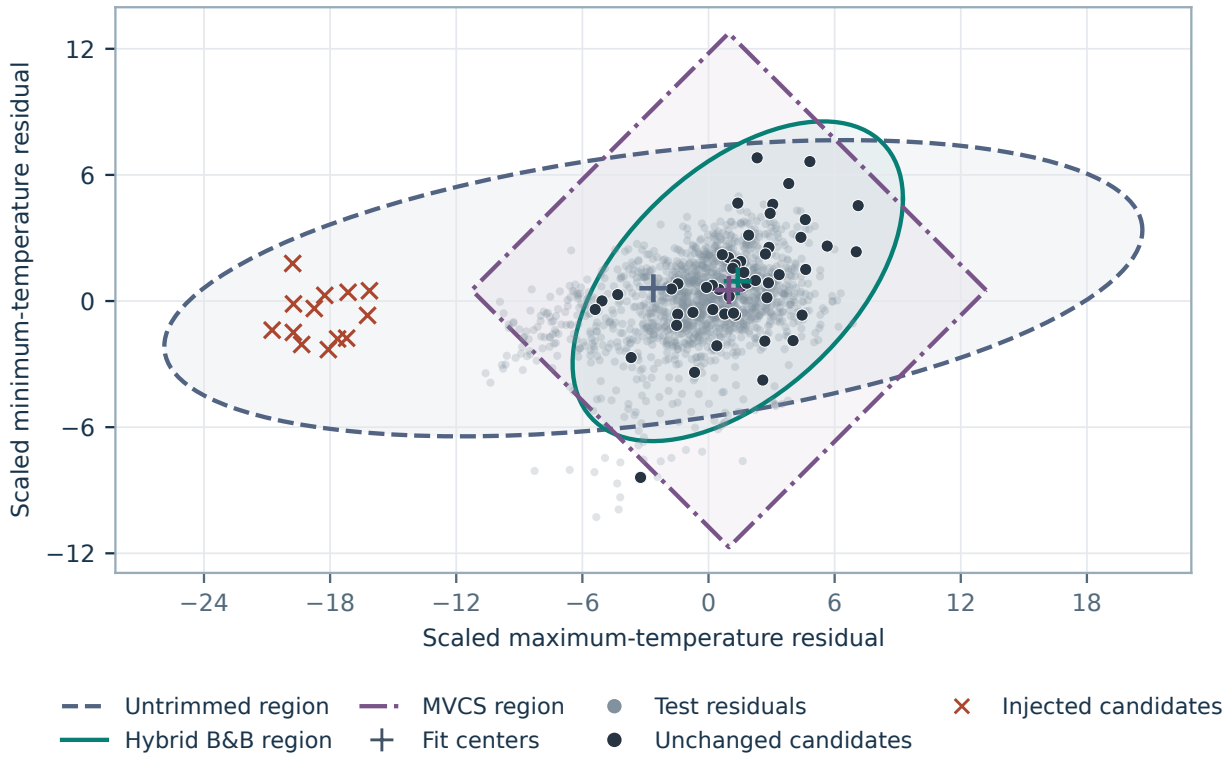



Figure 3: One weather replication with 13 clustered replacements at the 95% target. The MCD region, labeled Hybrid B&B, uses $h = 51$, selected by the validation step, and has 33.9% of the untrimmed area and 53.7% of the MVCS area.

volume was 36.4% of the MVCS volume, with empirical coverage of 90.00% and 89.62%, respectively. For weather at the 90% target, coverage fell to 85.76% for MCD and 86.55% for MVCS. At the 95% weather target, MCD covered 95.42% with a volume ratio of 0.888. Figure 3 illustrates the validation-selected cardinality rule for one weather replication. Equation (18) shows why both the fitted determinant and the calibration radius enter the volume.

# 7 Conclusion and research directions

The minimum covariance determinant problem selects a subset of fixed size whose sample covariance matrix has minimum determinant. We develop approximation and exact algorithms for this problem. Quadratic-form selection provides a deterministic approximation scheme for the geometric-mean objective, and an affine-hull test solves instances with optimal value zero. Zonotope enumeration provides a polynomial-time exact algorithm in fixed dimension that permits repeated observations and affine dependencies. The covering-ellipsoid analysis converts an approximate MCD solution into a covering-volume guarantee. In the computational study, spectral and completion bounds improve search completion within numerical tolerances, with the largest gains on medium and large instances. Hybrid initialization improves starting solutions over random starts, and branch-and-bound finds smaller covariance determinants than FastMCD on some contaminated prediction instances at equal cardinality. The prediction experiments also show that smaller region volume can coincide with empirical coverage below the target level, so both measures are needed to assess prediction regions. Future research can use dependencies among zonotope generators to strengthen node bounds and develop criteria for solving residual subproblems by exact enumeration. Computational comparisons may account for the cost of generating inequalities and enumerating candidates. Further questions concern subset-cardinality selection with limited validation data and prediction performance when calibration and evaluation observations come from different periods.

# Acknowledgments

The authors acknowledge the use of OpenAI's ChatGPT to assist with polishing the exposition and refining some of the results presented in this paper.

# References

[1] P. J. Rousseeuw and K. Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223, 1999. doi:10.1080/00401706.1999.10485670.

[2] M. Hubert, M. Debruyne, and P. J. Rousseeuw. Minimum covariance determinant and extensions. *WIREs Computational Statistics*, 10:e1421, 2018. doi:10.1002/wics.1421.

[3] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7:77–91, 1952. doi:10.1111/j.1540-6261.1952.tb01525.x.

[4] D. Rosadi, E. P. Setiawan, M. Templ, and P. Filzmoser. Robust covariance estimators for mean–variance portfolio optimization with transaction lots. *Operations Research Perspectives*, 7:100154, 2020. doi:10.1016/j.orp.2020.100154.

[5] H. Hotelling. Multivariate quality control—illustrated by the air testing of sample bombsights. In C. Eisenhart, M. W. Hastay, and W. A. Wallis, editors, *Techniques of Statistical Analysis*, pages 111–184. McGraw–Hill, New York, 1947.

[6] P. J. Rousseeuw and B. C. van Zomeren. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85:633–639, 1990. doi:10.1080/01621459.1990.10474920.

[7] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23:769–805, 1998. doi:10.1287/moor.23.4.769.

[8] V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic Learning in a Random World*. Springer, New York, 2005. doi:10.1007/b106715.

[9] J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113:1094–1111, 2018. doi:10.1080/01621459.2017.1307116.

[10] M. Hubert, P. J. Rousseeuw, and K. Vanden Branden. ROBPCA: A new approach to robust principal component analysis. *Technometrics*, 47:64–79, 2005. doi:10.1198/004017004000000563.

[11] K. Boudt, P. J. Rousseeuw, S. Vanduffel, and T. Verdonck. The minimum regularized covariance determinant estimator. *Statistics and Computing*, 30:113–128, 2020. doi:10.1007/s11222-019-09869-x.

[12] D. M. Hawkins. The feasible solution algorithm for the minimum covariance determinant estimator in multivariate data. *Computational Statistics & Data Analysis*, 17:197–210, 1994. doi:10.1016/0167-9473(92)00071-X.

[13] D. L. Woodruff. Ghost image processing for minimum covariance determinants. *ORSA Journal on Computing*, 7:468–473, 1995. doi:10.1287/ijoc.7.4.468.

[14] M. Schyns, G. Haesbroeck, and F. Critchley. RelaxMCD: Smooth optimisation for the minimum covariance determinant estimator. *Computational Statistics & Data Analysis*, 54:843–857, 2010. doi:10.1016/j.csda.2009.11.005.

[15] T. Bernholt and P. Fischer. The complexity of computing the MCD-estimator. *Theoretical Computer Science*, 326:383–398, 2004. doi:10.1016/j.tcs.2004.08.005.

[16] M. Dereziński, M. K. Warmuth, and D. Hsu. Leveraged volume sampling for linear regression. In *Advances in Neural Information Processing Systems*, volume 31, 2018. URL `https://proceedings.neurips.cc/paper/2018/hash/2ba8698b79439589fdd2b0f7218d8b07-Abstract.html`.

[17] M. Singh and W. Xie. Approximation algorithms for D-optimal design. *Mathematics of Operations Research*, 45:1512–1534, 2020. doi:10.1287/moor.2019.1041.

[18] Y. Li, M. Fampa, J. Lee, F. Qiu, W. Xie, and R. Yao. D-optimal data fusion: Exact and approximation algorithms. *INFORMS Journal on Computing*, 36:97–120, 2024. doi:10.1287/ijoc.2022.0235.

[19] S. Onn and U. G. Rothblum. Convex combinatorial optimization. *Discrete & Computational Geometry*, 32:549–566, 2004. doi:10.1007/s00454-004-1138-y.

[20] J. Agulló Candela. *Computación de estimadores con alto punto de ruptura*. PhD thesis, Universidad de Alicante, 1997. URL `https://hdl.handle.net/10045/3146`.

[21] C. Pesch. Fast computation of the minimum covariance determinant estimator. Technical Report MIP-9806, Universität Passau, 1998. URL https://webdoc.sub.gwdg.de/ebook/e/2001/mip/MIP-9806.html.

[22] C. Pesch. Computation of the minimum covariance determinant estimator. In W. Gaul and H. Locarek-Junge, editors, *Classification in the Information Age*, pages 225–232. Springer, Berlin, 1999. doi:10.1007/978-3-642-60187-3_22.

[23] D. Bertsimas, R. Cory-Wright, and J. Pauphilet. Solving large-scale sparse PCA to certifiable (near) optimality. *Journal of Machine Learning Research*, 23(13):1–35, 2022. URL https://jmlr.org/papers/v23/20-1188.html.

[24] D. Pisinger, A. B. Rasmussen, and R. Sandvik. Solution of large-sized quadratic knapsack problems through aggressive reduction. *INFORMS Journal on Computing*, 19:280–290, 2007. doi:10.1287/ijoc.1050.0172.

[25] Z. Bai and G. H. Golub. Bounds for the trace of the inverse and the determinant of symmetric positive definite matrices. *Annals of Numerical Mathematics*, 4:29–38, 1997.

[26] C. Johnstone and B. Cox. Conformal uncertainty sets for robust optimization. In *Proceedings of the Tenth Symposium on Conformal and Probabilistic Prediction and Applications*, volume 152 of *Proceedings of Machine Learning Research*, pages 72–90. PMLR, 2021.

[27] S. Messoudi, S. Destercke, and S. Rousseau. Ellipsoidal conformal inference for multi-target regression. In *Conformal and Probabilistic Prediction with Applications*, volume 179 of *Proceedings of Machine Learning Research*, pages 1–13. PMLR, 2022.

[28] S. Braun, L. Aolaritei, M. I. Jordan, and F. Bach. Minimum volume conformal sets for multivariate regression. *Journal of the American Statistical Association*, 2026. doi:10.1080/01621459.2026.2692078.

[29] H. Edelsbrunner, J. O'Rourke, and R. Seidel. Constructing arrangements of lines and hyperplanes with applications. *SIAM Journal on Computing*, 15:341–363, 1986. doi:10.1137/0215024.

[30] H. Edelsbrunner, R. Seidel, and M. Sharir. On the zone theorem for hyperplane arrangements. *SIAM Journal on Computing*, 22:418–429, 1993. doi:10.1137/0222031.

[31] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 2 edition, 2013.

[32] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. Wiley, New York, 1987.

[33] UCI Machine Learning Repository. Bias correction of numerical prediction model temperature forecast. Dataset, 2020.

[34] Mulan Project. Multi-target regression datasets: RF1. Dataset catalog. URL https://mulan.sourceforge.net/datasets-mtr.html.

[35] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine Learning*, 63: 3–42, 2006. doi:10.1007/s10994-006-6226-1.